Departamento de Informática y Sistemas

Métodos Variacionales para la Estimación del Flujo Óptico y Mapas de Disparidad

Optical Flow and Disparity Maps Estimation using Variational Methods

Tesis Doctoral

Agustín Javier Salgado de la Nuez

Las Palmas de Gran Canaria
Enero 2010

A mis padres

Agradecimientos

El desarrollo de una tesis doctoral supone un gran esfuerzo en el que han intervenido un gran número de personas. Aprovecho este momento para agradecer la ayuda y el apoyo recibido durante este periodo.

En primer lugar, quisiera agradecer a mis padres el esfuerzo económico realizado durante estos años para que pudiera estudiar una carrera universitaria y posteriormente continuar con los estudios de doctorado.

Hacer una mención especial a Luis Alvarez por ofrecerme la posibilidad de realizar los estudios de doctorado en el seno del grupo AMI y continuar la relación iniciada en el proyecto fin de carrera. Además agraceder al tutor de mi tesis, Javier Sánchez su tiempo, apoyo y dedicación durante el desarrollo de la tesis.

Durante todos estos años he tenido la oportunidad de compartir momentos de trabajo y distracción con compañeros en el laboratorio de investigación AMI de la Universidad de Las Palmas de Gran Canaria. Quisiera recordarlos: Karl, Carlos, David, Jesús, Pedro, Antonio, Miguel y Laura. Agraceder a los miembros del grupo AMI la ayuda y comentarios recibidos: Agustín Trujillo, Carmelo Cuenca, Julio Esclarín, Luis Mazorra y Miguel Alemán.

Quisiera agraceder a Joachim Weickert por permitirme realizar una estancia de investigación en el seno de su grupo de investigación. Durante este período tuve la posibilidad de asistir a sus clases, ampliar mis conocimientos sobre métodos variacionales para la estimación del flujo óptico e intercambiar ideas con él durante muchísimas horas. Quisiera recordar a Andres Bruhn, Luis Pizarro e Irena Galic por su colaboración, consejos y las ayudas recibidas durante dicha estancia.

Por último, quisiera agradecer a las instituciones que han financiado parte de los trabajos realizados en el contexto de esta tesis. En primer lugar, al Departamento de Informática y Sistemas y al servicio de investigación de la Universidad de Las Palmas *de Gran Canaria por poner los medios suficientes para llevar a cabo este trabajo. A otras instituciones como el DAAD (Deutscher Akademischer Austausch Dienst) que me concedieron una beca para el desarrollo de un proyecto de investigación en el grupo del Profesor Joachim Weickert. A la fundación de la Universidad de Las Palmas de Gran Canaria que gracias a la beca Innova, financiada por Unelco, me dió la posibilidad de adquirir material utilizado en el desarrollo de esta tesis. Algunos trabajos de esta tesis forman parte de proyectos financiados por Consejería de Educación Cultura y Deportes del gobierno de Canarias (PI2002/193), Ministerio de Ciencia y Tecnología y FEDER (TIC2003-08957). Agredecer también a la empresa MediaPro la cesión y el permiso para la utilización de una secuencia de fútbol utilizada en un método de estimación del mapa de disparidad.*

Índice general

Introducción

La visión por ordenador es un amplio campo de investigación que trata de extraer propiedades del mundo real a partir de la información captada por una serie de sensores. Esta tesis se centra en el estudio del movimiento registrado en una escena a través del análisis de una secuencia de imágenes. Para ello, se han utilizado las imágenes captadas por una cámara fotográfica. Cuando se toman imágenes de una escena se pueden presentar distintas configuraciones, dependiendo del número de cámaras y de la disposición de éstas, que se pueden resumir en dos casos: (i) una secuencia de imágenes tomadas por una sola cámara en distintos instantes de tiempo, y (ii) una secuencia de imágenes captadas por varias cámaras en el mismo instante. Estas dos situaciones representan dos problemas claves dentro de la visión por ordenador y que han sido objeto de un amplísimo estudio durante décadas. Estos dos problemas son: (i) la *estimación del flujo óptico* y (ii) la *estimación del mapa de disparidad*. Ambos se reducen a una búsqueda de correspondencias.

El flujo óptico se define como el desplazamiento existente entre los píxeles de imágenes bidimensionales que han sido tomadas por una sola cámara en distintos instantes de tiempo. En principio, se desconoce si ese desplazamiento se debe a los objetos o a la propia cámara. Las imágenes no son más que las proyecciones del movimiento de los objetos en escenas tridimensionales. Con una sola proyección no es posible determinar de forma unívoca la profundidad de los puntos. Por ese motivo, no siempre el movimiento descrito en el flujo óptico se corresponde con el de la escena real.

En la visión estereoscópica disponemos de dos vistas de la escena en el mismo instante de tiempo. La estimación del mapa de disparidad consiste en el cálculo del desplazamiento de los píxeles de una vista a la otra. Este problema se reduce a una búsqueda de correspondencias entre las dos vistas, por lo que en cierto modo se asemeja bastante a la estimación del flujo óptico. La visión estereoscópica dispone de una herramienta muy útil como es la *geometría epipolar*. La geometría epipolar permite relacionar los puntos en correspondencias entre varias imágenes acotando el área de búsqueda a una recta. Esto sólo se cumple cuando el sistema de cámaras está perfectamente calibrado.

En la literatura se han propuesto distintos tipos de técnicas para resolver estos dos problemas. Entre todas estas técnicas los métodos variacionales destacan por la precisión de sus soluciones. Se trata de un técnica basada en la minimización de energías en la que la solución debe cumplir una serie de restricciones impuestas en el modelo. Los métodos variacionales ofrecen una serie de ventajas que comentamos a continuación:

- *Campos de desplazamientos densos*: Las soluciones de los métodos variacionales son densas, es decir, que disponemos de un valor del desplazamiento para todos los píxeles de la imagen. Se tratan de métodos globales, de forma que si la información local no es suficiente para estimar la solución se utiliza la información de los vecinos. Otro tipo de técnicas, como puede ser una basada en correlación, calculan la solución en determinados puntos de la imagen y la única forma de obtener soluciones densas es a través de un proceso de interpolación.

- *Restricciones reflejadas en el modelo de energía*: Todas las restricciones que se imponen en el método están reflejadas en el modelo de energía. No existe ninguna restricción que se implemente y que no aparezca en dicho modelo.

- *Base matemática sólida*: El modelado de la energía y el proceso de minimización se sustenta sobre una sólida base matemática que asegura, si se cumple una serie de condiciones, la existencia y unicidad de la solución, así como la convergencia del método.

Los modelos de energía definidos en esta tesis se componen de dos partes: (1) el término de ligadura y (2) el término de suavizado. El término de ligadura asume que cierta propiedad en la imagen no varía a lo largo del tiempo; mientras que el de suavizado supone cierta restricción de suavidad en el flujo. La minimización del modelo de energía genera un sistema de ecuaciones diferenciales en derivadas parciales (EDP's, PDE's en inglés). Para la resolución de este sistema de ecuaciones se suele recurrir a un método de descenso por gradiente. Éste tiene la ventaja de ser fácil de implementar y la convergencia es relativamente rápida. Existen otras técnicas como SOR (*Sucessive Over-Relaxation*) o multigrid que convergen con un número menor de iteraciones. La implementación del sistema de ecuaciones se realiza mediante esquemas numéricos explícitos, semi-implícitos o implícitos. Ofrecen gran estabilidad y la complejidad de la discretización es relativamente baja; supone un compromiso entre estabilidad y eficiencia.

En secuencias de imágenes reales los desplazamientos de los objetos suelen ser grandes, superiores a varios píxeles. Los métodos variacionales se basan en la información de las derivadas para hacer sus cálculos. Cuando los desplazamientos son grandes la condición de derivabilidad no se cumple por lo que es necesario reflejar esta circunstancia en el modelo de energía. Esto origina la aparición de términos no-lineales. Para simplificar estas expresiones se recurre a la linealización de estos términos *"conflictivos"* mediante la expansión de Taylor suprimiendo los elementos de orden superior. Para evitar que los métodos se queden atrapados en mínimos locales irrelevantes se suele embeber el esquema numérico en una estrategia multiescala. De esta forma se acelera la convergencia y se permite la detección de los desplazamientos largos. El uso de los enfoques multiescalas han sido ampliamente utilizados en la literatura [Anandan89, Battiti91, Luettgen94, Bornemann96], [Enkelmann88, Mémin02].

Motivación y alcance del trabajo

El problema de la estimación del flujo óptico y del mapa de disparidad ha sido objeto de estudio durante más de veinte años. A pesar de la amplia y profunda investigación realizada durante este tiempo, hoy en día se siguen haciendo numerosas aportaciones que han conseguido incrementar la precisión y robustez de las estimaciones. Los primeros métodos utilizaban únicamente dos imágenes. La necesidad de aumentar la precisión de las estimaciones hizo que se empleara información de los frames vecinos surgiendo los primeros métodos espaciotemporales. Estos métodos demostraron la ventaja que suponía el uso de la información temporal. Los métodos espaciales perdieron su hegemonía en la literatura dando paso a otros que incluían información temporal.

El contenido de esta tesis se ha centrado en el estudio y desarrollo de métodos variacionales espaciotemporales para el cálculo del *flujo óptico* en secuencias de imágenes y el cálculo de la *carta de disparidad* en pares estéreos. Los métodos propuestos utilizan información procedente de varias imágenes, ya sea del mismo o de varios canales. Algunos trabajos, como el de [Papenberg06], incluyen un término de regularización que asume un flujo suave tanto en el dominio espacial como en el temporal. Esta suposición incorpora una serie de limitaciones que tratamos de superar con la división del término del regularización en uno espacial y otro temporal, permitiendo los desplazamientos largos en el dominio temporal sin afectar a la regularización espacial.

En aplicaciones de meteorología el uso de imágenes satélites multiespectrales es muy frecuente. El análisis de los fenómenos registrados en esas secuencias es fundamental para predecir el tiempo. El desarrollo de métodos para el seguimiento de las estructuras nubosas presentes en la atmósfera resulta una herramienta muy interesante en este campo. Hemos querido desarrollar un método que utilice imágenes multiespectrales para mejorar la estimación del desplazamiento de las nubes. Se trata de una modificación del método de Alvarez/Weickert/Sánchez (2000) [Alvarez00] para la estimación del flujo óptico en el que se ha añadido información multicanal en el término de ligadura y un gradiente combinado (de todos los canales) en el término de regularización.

En la literatura se han propuesto innumerables modelos de energías que incorporan diversas restricciones en el término de ligadura y que ofrecen distintos comportamientos en el término de regularización. Sería deseable el diseño de un modelo de energía genérico que aglutinase todas esas ideas bajo una misma estructura configurable, fácilmente adaptable y extensible ante cualquier nueva invarianza que quisiésemos añadir. Algunas de las aportaciones realizadas en los últimos años permite la acumulación de varias invarianzas en una notación compacta (*tensor de movimiento*, [Bruhn06a]), o proporcionan robustez a los métodos ante la presencia de ruido en la secuencia (*funciones de robustificación*, [Black91, Black96b]). Como parte de esta tesis se propone un innovador framework que se ayuda de alguna de esas técnicas para la definición un modelo de energía genérico que establece una interconexión *explícita* entre el término de ligadura y el de regularización.

Los métodos estéreos que estiman el mapa de disparidad utilizan únicamente la información procedente de un par estéreo. Estos métodos cuentan con la geometría epipolar como una potente herramienta para delimitar la zona de búsqueda de las correspondencias. Cuando los pares estéreos se ven afectados por el ruido no se dispone de información suficiente que nos indique cuando un píxel es ruidoso o una característica de la imagen. Por ello, la incorporación de información del flujo óptico permitiría detectar estas situaciones y mejorar fácilmente la precisión de las estimaciones. En uno de los trabajos presentados en esta tesis se combina en el modelo de energía la información del flujo óptico y estéreo para incrementar la calidad del mapa de disparidad. Este método necesita una secuencia de pares estéreos como dato de entrada.

Los métodos graph–cuts ofrecen muy buenos resultados a la hora de estimar el mapa de disparidad. Una limitación de esta técnica está en la precisión entera de sus soluciones. Los métodos variacionales permiten obtener soluciones en precisión subpíxel pero para asegurar su convergencia se suele recurrir a algún tipo de inicialización. Normalmente la técnica más utilizada es la correlación. Sin embargo, vamos a demostrar que la combinación de un método variacional y otro de graph–cuts parece más ventajosa que un variacional con una técnica de correlación.

Contenido de la tesis

Este documento está dividido en cuatro capítulos. En el primero de ellos se hace un recorrido por el estado del arte de los dos problemas tratados en esta tesis. Se dedicará un capítulo completo a cada problema describiendo exhaustivamente los métodos desarrollados así como las principales aportaciones a la literatura. En el último capítulo se comentará las principales conclusiones obtenidas de los trabajos realizados y las tareas pendientes que se han dejado como trabajo futuro. A continuación, se describe con mayor nivel de detalle el contenido de cada capítulo y los distintos anexos que podemos encontrar al final de este documento.

- Capítulo 1, *Estado del arte*:

  En este capítulo se describe el *estado del arte* relativo al problema de la estimación del flujo óptico y de la carta de disparidad. Se hace mención a los métodos más relevantes de la literatura, describiendo algunas de las técnicas más innovadoras surgidas en los últimos años que han servido de referencia a otros métodos desarrollados con posterioridad. Al final de este capítulo se enumeran las secuencias reales y sintéticas utilizadas en los experimentos de esta tesis, así como las métricas empleadas para la obtención de resultados cuantitativos.

- Capítulo 2, *Estimación del Flujo Óptico en Secuencias de Imágenes*:

  El contenido del capítulo 2 está destinado íntegramente al desarrollo de métodos para la estimación del flujo óptico con información procedente de varias imágenes. Se quiere poner de manifiesto que la combinación de la información de varias imágenes, ya sean del mismo canal o multicanal, permiten aumentar la precisión de las estimaciones respecto a su homólogo espacial o incluso respecto otros métodos similares en la literatura. Al comienzo de este capítulo, se describe la notación utilizada para la definición de un modelo de energía general que aglutina todos los casos que pueden darse a la hora de definir un funcional de energía. Para ello, tenemos en cuenta la experiencia recogida en la literatura y las distintas aportaciones realizadas en esta tesis.

  En el primer trabajo se plantea un método variacional para la detección de masas nubes en secuencias satélites multiespectrales. Se basa en el modelo monocanal propuesto por Alvarez/Weickert/Sánchez (2000) [Alvarez00]. La versión monocanal será modificada añadiéndole información de varias imágenes de distintos canales. El término de ligadura se compone de una suma ponderada de la información de cada canal mientras que el término de regularización se trata del operador de Nagel–Enkelmann con un gradiente combinado (de todos los canales). A partir del modelo de energía se derivarán las ecuaciones de Euler-Lagrange y se justificará la discretización y el esquema numérico empleado. En los experimentos realizados con secuencias del satélite Meteosat se comprueba la mejora obtenida por la combinación de la información multicanal.

El segundo trabajo consiste en una modificación de un método espacial al que se le ha añadido un término de regularización temporal que permite la detección de los desplazamientos largos. Se comenzará introduciendo el modelo de energía justificando la división de la regularización en dos términos distintos, uno espacial y otro temporal. A partir de este funcional de energía se deducirá las ecuaciones de Euler-Lagrange y se describirá con detalle el esquema numérico utilizado. Con el objetivo de evaluar la mejora obtenida se presentarán una serie de experimentos con secuencias reales y sintéticas demostrando que la diferenciación temporal en el término de regularización ofrece estimaciones más robustas.

En el último trabajo de este tema se propone un *framework* que establece una relación de *complementariedad* entre el término de ligadura y el de regularización del modelo de energía. En primer lugar, se presentan las técnicas en las que nos apoyamos para definir dicha complementariedad. A continuación, se expone el modelo de energía de Nagel–Enkelmann cuyo término de regularización suaviza en la dirección ortogonal al gradiente. A partir de este modelo e incluyendo diversas técnicas, como el tensor de movimiento o funciones de robustificación, definimos la estructura general de nuestro *framework*. La descomposición del tensor de movimiento en autovalores y autovectores permite la identificación de las direcciones dominantes del flujo. A partir de esta información se construiría un tensor inverso que sería utilizado en término de regularización para el guiado del proceso de difusión. De esta forma, se crearía una interconexión *explícita* entre la información del término de ligadura y regularización. La flexibilidad del tensor de movimiento permite la inclusión de múltiples invarianzas dentro de la misma estructura. La incorporación de las funciones de robustificación proporciona insensibilidad frente al ruido. Una vez descrito las distintas variantes del modelo general se minimizará el caso más completo y complejo, desarrollando la discretización y el esquema numérico. En este trabajo se ha utilizado un esquema multipiramidal para la detección de los desplazamientos largos. Para demostrar que las ventajas que ofrece este nuevo framework se presentarán los resultados experimentales obtenidos utilizando secuencias reales y sintéticas. Al mismo tiempo, se compararán dichos resultados con los mejores métodos de la literatura.

Al final de este capítulo se expone las conclusiones de cada trabajo destacando aquellos aspectos que han supuesto una importante aportación a la literatura.

- Capítulo 3, *Estimación del Mapa de Disparidad en Pares Estéreo*:

  Los trabajos presentados en este capítulo tienen como objetivo la mejora de la estimación del mapa de disparidad mediante la combinación de varios métodos.

  En la primera parte de este capítulo se presenta un método que combina la información del flujo óptico y estéreo para obtener una estimación más robusta. Se trata de la continuación de los trabajos de Alvarez/Weickert/Sánchez

(2000) [Alvarez00] y Alvarez/Deriche/Sánchez/Weickert (2002) [Alvarez02b]. Ambos métodos se basan en una técnica de minimización de energía que ofrecen resultados precisos y densos. Se comenzará repasando algunos conceptos sobre la geometría epipolar para ir profundizando en ellos de forma que se vayan fusionando las ideas del método del flujo óptico y del estéreo. La notación definida en esta introducción se utilizará en la presentación del modelo de energía. Se continua con la minimización y el esquema numérico incluyendo la descripción del enfoque multipiramidal que permite detectar los desplazamientos largos. El esquema numérico utiliza la técnica del descenso del gradiente. Finalizaremos presentando los resultados experimentales obtenidos utilizando secuencias sintéticas y reales.

En el segundo trabajo se propone un método para la estimación del mapa de disparidad que combina otros dos: uno variacional descrito en Alvarez/Deriche/Sánchez/Weickert (2002) [Alvarez02b] y uno de graph–cuts, [Boykov04]. En primer lugar, haremos mención al proceso de rectificación. Cuando la disposición de las cámaras es frontoparalela la formulación de los modelos de energía se simplifica ya que el desplazamiento de los píxeles queda expresado como un escalar (en una única dirección). La convergencia del método variacional se acelera si se dispone de una buena aproximación inicial. En este trabajo se describen dos técnicas utilizadas para el cálculo de la aproximación inicial: la *correlación* y *graph–cuts*. Seguidamente se presenta las características del método variacional y cómo se ha combinado con el de graph–cuts para calcular el mapa de disparidad. Se introduce el método variacional en un enfoque multipiramidal para la detección de largos desplazamientos y evitar su convergencia mínimos locales irrelevantes. Por último, se muestran los resultados experimentales obtenidos estableciendo una comparación entre el método variacional con las dos aproximaciones iniciales. Se quiere poner de manifiesto la mejora que supone la inclusión del método de graph–cuts. En una segunda batería de pruebas que quiere poner a prueba la estabilidad de las distintas técnicas ante la presencia de ruido en los pares estéreo.

Al finalizar este capítulo se presentan las conclusiones valorando los hechos más significativos de los trabajos presentados en este capítulo.

- Capítulo 4, *Conclusiones*:

  En la sección 4.1 de este capítulo se exponen las conclusiones finales de la tesis comentando los problemas encontrados durante su desarrollo y destacando las aportaciones realizadas a la literatura en todos los trabajos recogidos en este documento. En la sección 4.2 se enumera algunas tareas que por su complejidad y falta de tiempo se ha dejado pendiente para el futuro.

- Anexo 4.2, *Notación*:

  En este anexo se unifica toda la notación utilizada en la tesis. Se trata de una guía rápida en la que el lector puede familiarizarse con la nomenclatura empleada en el documento.

Las principales aportaciones

En este documento se presentan una serie de trabajos que representan las aportaciones novedosas a la literatura en el campo de la estimación del flujo óptico y estéreo realizadas en esta tesis. A continuación, se describen brevemente estas contribuciones.

- **Método Variacional Espacial Multicanal: Extensión del Modelo de Alvarez/Weickert/Sánchez (2000) [Alvarez00]:**

  Basándonos en el funcional de energía propuesto por [Alvarez00] se le ha aplicado una serie de modificaciones para la *inclusión* de la *información multicanal*. Los cambios introducidos en el modelo de energía son los siguientes: (1) la definición de un término de ligadura como una suma ponderada de la suposición lambertiana de cada uno de los canales y (2) la modificación del operador de Nagel-Enkelmann para que su gradiente aglutine información de varias imágenes. Se han definido diversas estrategias para la construcción de dicho gradiente.

  Este método se ha utilizado para la detección de masas nubes en secuencias satélites multiespectrales.

- **Método Variacional que incorpora un Término de Regularización exclusivamente Temporal no Continuo:**

  La principal contribución de este trabajo ha sido el desarrollo de un método variacional espaciotemporal que incorpora un término de regularización exclusivamente temporal. El desacoplamiento del término de suavizado en dos, uno espacial y otro temporal, pretende evitar algunas de las limitaciones presentes en trabajos con una regularización temporal continua basada en derivadas parciales temporales, como en [Papenberg06]. Nuestro método realiza convenientemente la regularización temporal teniendo en cuenta la presencia de desplazamientos largos sin afectar a la regularización espacial.

- **Método Variacional basado en el Análisis Espectral del Tensor de Movimiento:**

  La principal aportación de este trabajo ha sido el desarrollo de un método variacional espaciotemporal que intercambia información complementaria entre el término de ligadura y el de suavizado. El proceso de difusión sólo se aplica en las direcciones ortogonales al flujo dominante. La combinación de varias técnicas existentes posibilita la creación de un *framework* capaz de definir cualquier modelo de una forma compacta, sencilla, fácilmente extensible y adaptable.

  El *tensor de movimiento* permite representar cualquier invarianza en una notación compacta. La descomposición espectral de este tensor asegura la fácil identificación de las direcciones dominantes del flujo al mismo tiempo que define una notación independiente de las invarianzas definidas en el modelo de energía.

  La inclusión de términos no-cuadráticos no sólo aporta robustez a las estimaciones frente al ruido sino que la combinación de las funciones de robustificación junto a la descomposición espectral del tensor de movimiento permite la construcción de cuatro prototipos capaces de mejorar las estimaciones de métodos de características similares.

- **Método para la estimación del mapa de disparidad utilizando una secuencia de pares estéreo:**

  La principal aportación de este trabajo es la creación de un método que combina la información del flujo estéreo (Alvarez/Deriche/Sánchez/Weickert (2002), [Alvarez02b]) con la del flujo óptico (Alvarez/Weickert/Sánchez (2000), [Alvarez00]) aumentando la robustez de las estimaciones gracias a la fusión, dentro del mismo modelo de energía, de las ideas plasmadas en dos métodos de gran precisión.

  La estimación del mapa de disparidad se apoya en la geometría epipolar para poder establecer las correspondencias de cada par estéreo. Esas mismas correspondencias se pueden obtener mediante el cálculo del flujo óptico a lo largo de la secuencia de ambas cámaras. Por lo tanto, en una secuencia de pares estéreo disponemos de hasta cuatro vistas de un punto 3D en dos instantes de tiempo distintos.

  Este método aprovecha la información suministrada por el flujo estéreo y óptico, estimados de forma conjunta, para así poder lograr un mapa de disparidad más preciso incluso en situaciones donde existen oclusiones. Para asegurar la coherencia del resultado final se ha incluido en el modelo de energía una restricción que fuerza la congruencia de ambos flujos.

- **Combinación de un método variacional espacial y uno de graph–cuts para la estimación del mapa de disparidad:**

  La segunda aportación en el campo de la estimación del mapa de disparidad consiste en la combinación de un método variacional (Alvarez/Deriche/Sánchez/Weickert (2002), [Alvarez02b]) y una técnica de graph–cuts ([Kolmogorov01, Boykov04]). El método de graph–cuts se ha utilizado para obtener una aproximación inicial del mapa de disparidad que el método variacional se encargaría de refinar. Normalmente, la técnica más empleada para estimar la inicialización es una basada en la correlación a ventanas.

  Dado los buenos resultados que ofrece algunos métodos de graph–cuts en el campo de la estimación del mapa de disparidad parece una buena alternativa a los tradicionales métodos basados en correlación a ventanas.

Publicaciones realizadas

En esta sección se hará una breve descripción de las publicaciones realizadas en el contexto de esta tesis.

- **Multi-Channel Satellite Image Analysis Using a Variational Approach** Alvarez/Castaño/García/Krissian/Mazorra/Salgado/Sánchez (2008) [Alvarez08]:

  En este trabajo se presenta un método variacional multicanal para hacer frente a algunos de los problemas más habituales en el análisis de imágenes por satélite, como es la estimación del movimiento de las estructuras nubosas presentes en la atmósfera. El modelo de energía propuesto se basa en el trabajo de Alvarez/Weickert/Sánchez (2000) [Alvarez00] y combina la información de varios canales del satélite. Las ventajas y mejoras en las estimaciones que ofrece este nuevo método se ponen de manifiesto en los experimentos realizadas con dos secuencias de satélite.

- **Optical Flow Estimation with Large Displacements: A Temporal Regularizer** Salgado/Sánchez (2006, 2007) [Salgado06a], [Salgado06b], [Salgado07b]:

  En este trabajo se presenta un modelo variacional para la estimación del flujo óptico en una secuencia de imágenes. Basándonos en el modelo espacial propuesto por Alvarez/Weickert/Sánchez (2000) [Alvarez00] se ha añadido un nuevo regularizador temporal que desacopla la información espaciotemporal al mismo tiempo que es capaz de detectar los desplazamientos grandes. La disociación entre la regularización espacial y temporal es necesaria con el objetivo de evitar o resolver algunas incongruencias detectadas en otros trabajos.

- **3D Geometry Reconstruction from a Stereoscopic Video Sequence** Salgado/Sánchez (2005) [Salgado05a], [Salgado05d]:

  En este trabajo se propone un método que estima la geometría 3D de una escena a partir de una secuencia de vídeo tomada desde un par de cámaras estéreo. Las cámaras están rígidamente situadas en una posición fija, en posición frontoparalela, y hay una serie de objetos moviéndose por la escena. El método propuesto calcula el desplazamiento de los objetos y la estructura 3D de la escena a través de la estimación del flujo óptico de la secuencia de cada cámara y de la estimación del mapa de disparidad de cada par estéreo de dicha secuencia. Para relacionar esta información se ha establecido una restricción temporal que relaciona e impone cierta coherencia entre el flujo óptico y estéreo. Esta restricción se justifica haciendo uso de la formulación matemática común que existe entre estos problemas.

- **Combining two Methods to Accurately Estimate Dense Disparity Maps** Salgado/Sánchez (2005, 2007) [Salgado05c], [Salgado07a], [Salgado05b]:

  En este trabajo se combinan dos métodos con el objeto de mejorar las estimaciones de la geometría 3D de una escena. Para ello, se apoya en par de imágenes estereoscópicas donde la disposición de las cámaras puede ser arbitraria. Para simplificar la complejidad de los cálculos se recurre a un proceso de rectificación. La estimación del mapa de disparidad se realiza empleando el método variacional propuesto en Alvarez/Deriche/Sánchez/Weickert (2002), [Alvarez02b]. A partir de un modelo de energía se obtiene un PDE cuya solución determina el desplazamiento registrado en cada par estéreo. Uno de los problemas de este tipo de técnicas es que depende en gran medida de la primera aproximación, es decir, de su inicialización. En este trabajo se compara la influencia sobre el resultado final de dos técnicas que se utilizan para calcular esta aproximación inicial. Por un lado, tenemos una técnica basada en la correlación y, por otro lado, una de *graph–cuts*.

Abstract

The overall objective of this thesis is to contribute originally to the development of the a set of variational methods for estimating optical flow and disparity maps from image sequences. Depending on the number of cameras in a scene, we can have two different settings: (i) a sequence of images taken by one camera at different time step and (ii) a sequence of images taken from several cameras at the same time. These two cases represent two key problems in computer vision and have been investigated for decades: (i) the *optical flow* and (ii) the *disparity map* estimation. The optic flow is the apparent motion of pixels between images. Given two images, the aim is the retrieval of the pixel displacements from one image to the other. This problem is the base for a broad number of applications such as 3D reconstruction, driver assistance systems, video compression and surveillance systems. There are still open research issues regarding this topic and new important contributions often appear. In order to solve the optic flow problem many techniques have been proposed. Among them, the variational methods have demonstrated to be one of the best techniques to obtain accurate solutions. In the stereo problem, we have two views of the scene taken at the same time. The disparity map is the computation of the pixels displacement from one view to the other. The stereo problem has a very useful tool such as the epipolar geometry. The epipolar geometry allows us to limit the search area of the points in correspondence along a line.

The first part of this thesis is focused on the variational optic flow methods using image sequences. In this topic we have developed:

1. a mathematical model, based on a variational approach, to estimate the cloud structure motion by combining information from various satellite channels. We include information of all the channels in a single variational motion estimation model. The initial optical flow technique [Alvarez00] is the base of our multichannel sequential motion tracking algorithm. We extend this variational optical flow method to deal with multichannel sequential data;

2. a model for computing the optical flow in a sequence of images with a spatial–temporal regularizer explicitly designed for large displacements. We study the

introduction of a temporal regularizer that expands the information beyond two consecutive frames. We propose to decouple the spatial and temporal regularising terms to avoid an incongruous formulation between the data and smoothness term. Our model is based on an energy functional that yields a partial differential equation (PDE). This PDE is embedded into a multi-pyramidal strategy to recover large displacements;

3. an energy model for the optic flow estimation using a sequence of images. It takes advantage of the recently introduced motion tensor within an energy functional. Through a principal component analysis of the motion tensor we construct novel energy functionals. We define a new framework that nicely combines two complementary tensors. This provides a direct mechanism to switch between smoothing the solution along the prominent directions and attracting objects with similar intensity values. Each one is derived from the eigenvalues and eigenvectors of the motion tensor. The main contribution of this work is the inclusion of a modified tensor to steer the diffusion process and the motion tensor decomposition. We use non-quadratic functionals improve the method's robustness with respect to outliers. This energy model can handle large displacements through the use of warping techniques.

The second part of this thesis is dedicated to the variational stereo flow methods using pairs of images. The main contributions in this topic are:

1. a novel method for the reconstruction of the 3D geometry of a scene from a stereoscopic video sequence. There are two video–cameras pointing to the same scene and recording frames at the same time. For every stereoscopic pair of images in the sequence we may compute a disparity map independently from the other pairs, to obtain a set of independent disparity maps. The problem is that, in general, the continuity of the solution is not preserved and it is very sensitive to the presence of noise. If we want to overcome this problem, we have to relate the estimation of disparity maps through the sequence. One way to do it is to compute the displacement of objects on both video–cameras and use this information to constraint the computation of the disparity maps in time. This work is a continuation of previous works on optical flow [Alvarez00] and disparity map estimation [Alvarez02b]. These two methods were also based on energy minimization techniques and proved to be reliable and accurate;

2. a method that combines two techiques for computing disparity maps using a pair of images. The first one is based on graph–cut energy minimization [Kolmogorov01,

Boykov04]. This method has demonstrated good results in integer precision which is enough for some applications. If we need a better accuracy, then it is necessary to use a different technique. In this case, we propose to use the variational method described in [Alvarez02b] as a suitable complement. Both methods are based on energy minimization approach. One of the problems of these methods is that they need a good initial approximation in order to obtain an accurate solution. We normally use a cross-correlation technique to compute this initialization.

As a conclusion, we have presented in this thesis different approaches to deal with the optical and stereo flow problems. As a result of our research, we have developed new methods that give solutions for different situations where displacements may be large and sub-pixel accuracy is needed.

Capítulo 1

Estado del Arte

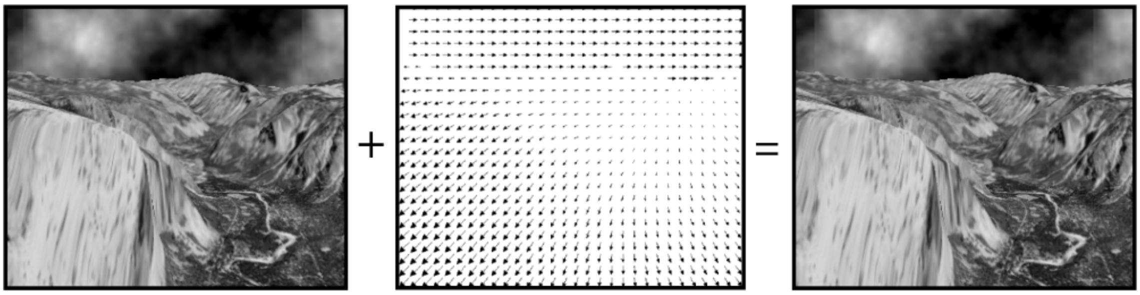
En los últimos treinta años se ha desarrollado una intensa actividad investigadora en el campo de la visión por computador. En sus orígenes la visión por computador estaba estrechamente relacionada con la robótica pero gracias al acercamiento de los ordenadores al gran público y el incremento de la capacidad de cómputo de las máquinas surgieron nuevos problemas a los que la visión por computador podía dar respuesta. Entre estos problemas destacamos la *estimación del flujo óptico* en secuencias de imágenes y el *cálculo del mapa de disparidad* en pares estéreo. Durante todo este tiempo se han desarrollado métodos que ofrecen soluciones de gran precisión. Sin embargo, todavía siguen existiendo algunas cuestiones sin resolver a la que intentaremos dar respuesta en esta tesis.

En este capítulo se describe el estado del arte referente al problema de la estimación del flujo óptico y de la carta de disparidad. Se hace un recorrido por los métodos más importantes de la literatura comentando algunas de las técnicas más exitosas utilizadas en los últimos años y que han supuesto una fuente de inspiración para los métodos desarrollados con posterioridad. Por último, se describe las secuencias de imágenes utilizadas en los distintos experimentos y las métricas de error en las que nos apoyamos para evaluar cuantitativamente las soluciones de los métodos.

1.1. Flujo Óptico

El cálculo del flujo óptico consiste en la estimación del movimiento aparente de los objetos en una secuencia de imágenes. Dado un conjunto de imágenes, el objetivo es calcular el desplazamiento de los píxeles entre las distintas imágenes. Disponemos de una cámara (o videocámara) que capta imágenes de una escena. En la escena podemos encontrar una serie de objetos estáticos o dinámicos que, por lo general, se verán influenciados por condiciones variables del entorno, tales como fuentes de iluminación, sombras, reflejos y otros efectos luminosos, así como por otras dificultades asociadas a la aparición y desaparición de objetos en la escena o la oclusión de unos objetos con otros.

El problema del flujo óptico se ha convertido en uno de los más importantes a resolver en el campo de la visión por ordenador. Su importancia radica fundamentalmente en el



I_1 Flujo óptico I_2

Cuadro 1.1: Ejemplo de flujo óptico ideal entre dos imágenes. Si a cada píxel de la primera imagen (I_1) le sumamos el desplazamiento descrito por el flujo óptico obtendríamos exactamente la segunda imagen (I_2).

gran número de aplicaciones que tiene, como por ejemplo, reconstrucción 3D, compresión de vídeo, segmentación, detección de objetos, sistemas de navegación robótica, etc.

El desplazamiento de los píxeles no es más que la proyección en una imagen del movimiento tridimensional. Se trata de un *problema inverso* en donde los valores de algunos parámetros del modelo deben ser obtenidos de los datos observados. En nuestro caso, las imágenes son los datos observados y queremos conocer el movimiento que se registra en la escena. Esto provoca que la estimación del flujo óptico sea un problema *mal condicionado* ([Bertero88]) ya que puede haber varias soluciones para un mismo desplazamiento, lo que da lugar a cierta ambigüedad. Según Jacques Hadamard, para convertir este tipo de problemas en *bien condicionado* se tiene que cumplir tres condiciones: (i) existencia, (ii) unicidad y (iii) estabilidad de la solución. Para cumplir con estos requisitos se suele recurrir a la incorporación de un elemento de estabilización, normalmente en forma de término de regularización o suavizado.

En el cuadro 1.1, vemos un ejemplo del flujo óptico entre dos imágenes. En un caso ideal, si a cada píxel de la primera imagen le sumamos el desplazamiento descrito por el flujo óptico obtendríamos exactamente la segunda imagen. En principio, cabría pensar que cualquier desplazamiento de píxeles en la imagen implica necesariamente movimiento en la escena; pero existen situaciones en las que se produce movimiento de píxeles pero no de objetos y viceversa. Existen muchos factores, que comentaremos en este capítulo, que dificultan la estimación exacta del flujo óptico.

Si representamos una imagen como una aplicación $I : (x, y, t) \rightarrow I(x, y, t)$ donde (x, y) representa la coordenada espacial de la imagen y t el tiempo, se puede ver una secuencia de imágenes como la variación de la intensidad en las coordenadas de la imagen a lo largo del tiempo. El vector desplazamiento se define como la función $\mathbf{h}(x, y, t) = (u(x, y, t), v(x, y, t))^\top$ y representa el movimiento horizontal y vertical de los píxeles a través de la secuencia de imágenes.

Para detectar las correspondencias de los píxeles entre dos imágenes se suele suponer que alguna propiedad de la imagen no varía a lo largo del tiempo. Esta suposición se puede representar tal que,

$$f(x+u, y+v, t+1) - f(x, y, t) = 0, \tag{1.1}$$

donde f es algún tipo de propiedad en la imagen, y $t+1$ y t representan dos imágenes de la escena en distintos instantes de tiempo.

La intensidad de los píxeles de una imagen es un valor que indica la cantidad de radiación luminosa reflejada por la superficie de los objetos. Existen muchos modelos para representar los distintos tipos de superficies. Uno de los más simples es el modelo *lambertiano*. En una superficie lambertiana el brillo aparente es el mismo para todas las direcciones de vista. Un píxel perteneciente a este tipo de superficie mantendrá el mismo valor de intensidad en todas las secuencias de imágenes. Por este motivo, si se asume que las superficies de los objetos son lambertianas bastaría con sustituir f por I en la ecuación (1.1) para representar esta invarianza.

Si analizamos la expresión (1.1) observamos que no es *lineal*. Sin embargo, esta *no linealidad* se puede evitar si realizamos un desarrollo de Taylor de dicha expresión y desechamos los términos de orden superior. De esta forma, se obtendría una ecuación que ha sido bautizada como la *ecuación de restricción del flujo óptico* (1.2) (sólo cuando $f := I$).

$$I_x u + I_y v + I_t = 0. \tag{1.2}$$

donde los subíndices indican derivadas parciales. A partir de la expresión (1.2) no es posible determinar el vector desplazamiento de forma unívoca ya que tenemos dos incógnitas y una sola ecuación. Tan sólo se puede estimar la componente en la dirección normal a la curva de nivel, en la dirección del gradiente. Esto es lo que se conoce como el *problema de apertura*, [Ullman79]. Para solventar este problema es necesario la inclusión de una segunda restricción; normalmente se suele imponer una restricción sobre el flujo. En la figura 1.1 se muestra gráficamente en qué consiste el problema de apertura. Vemos una elipse en cuyo interior de izquierda a derecha se aprecia una línea en los instantes de tiempo t y $t+1$, respectivamente. A priori, sólo podemos conocer el desplazamiento de la componente normal en la dirección del gradiente de los píxeles de la línea. Desconocemos si esa línea se ha desplazado también en la dirección de la componente ortogonal a la dirección del gradiente.

1.1.1. Clasificación de los Métodos

A pesar de las muchas contribuciones que ha habido en el campo del flujo óptico todo el trabajo realizado se podría resumir en unos pocos métodos de referencia. El resto de contribuciones supone la aportación de nuevos elementos que combinados con los anteriores han permitido mejorar la precisión de las estimaciones. Dado el gran volumen de aportaciones en este campo algunos autores han propuesto una serie de clasificaciones para establecer relaciones entre los distintos métodos.

En esta sección vamos a realizar un breve recorrido por el estado del arte basándonos en la clasificación hecha por Barron et al. [Barron94]. Sólo se comentará las principales técnicas y los métodos más representativos de cada uno de los grupos que guardan cierta

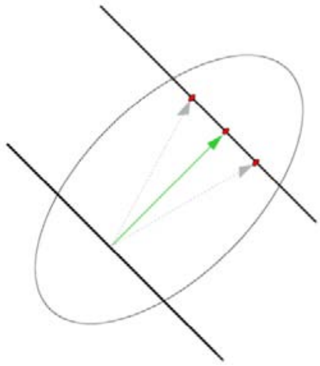


Figura 1.1: Problema de apertura.

relación con esta tesis. En la literatura también existen otras clasificaciones propuestas por autores como Beauchemin-Barron [Beauchemin95], Galvin [Galvin98], Mitiche-Bouthemy [Mitiche96] y Fleet [Fleet92]. Haremos algunas referencias entre las distintas clasificaciones.

Métodos Diferenciales

Los métodos diferenciales calculan el desplazamiento de los píxeles a partir de las derivadas espaciales (o espaciotemporales) de las intensidades de la imagen. Una limitación que tiene este tipo de métodos es que obliga a que las las derivadas sean computables en el dominio de la imagen.

En función de cómo se use la información de las derivadas Beauchemin-Barron [Beauchemin95] estableció la siguiente subcategoría para los *métodos diferenciales*: *locales*, *globales*, *de superficie*, *de contornos* y *multirestricciones*. Los métodos *locales* y *globales* se basan en la ecuación de restricción del flujo óptico, ec. (1.2). En la sección anterior se explicó que únicamente con la expresión (1.2) no era posible estimar el flujo óptico (*problema de apertura*) y, que por ello, era necesario la aportación de una restricción adicional. Los métodos *globales* usan directamente la ecuación (1.2) y añaden como restricción global un término de regularización sobre el flujo. Esta restricción supone que el campo de desplazamiento es suave. Con este tipo de métodos se obtienen campos de desplazamientos densos. El método más representativo es el de Horn-Schunck [Horn81]. El campo de desplazamiento se obtiene gracias a la minimización del funcional de energía descrito en (1.3).

$$\int_{\Omega} \left((I_x u + I_y v + I_t)^2 + \alpha \left(\|\nabla u\|^2 + \|\nabla v\|^2 \right) \right) dx. \tag{1.3}$$

A partir de la idea propuesta por Horn-Schunck han surgido innumerables aportaciones, en forma de nuevos términos de regularización como los de Nagel-Enkelmann [Nagel86], Alvarez/Esclarín/Lefébure/Sánchez (1999) [Alvarez99], Uras et al. [Uras88], Schnörr [Schnörr94b] o la inclusión de varias invarianzas en la *ecuación de restricción del flujo óptico*, Papenberg et al. [Papenberg06].

Los métodos *locales* utilizan la información en una vecindad alrededor de un píxel para estimar su movimiento. El método más representativo de esta familia es el de Lucas-Kanade [Lucas81]. Este método calcula el desplazamiento a partir de la minimización de la ecuación del flujo óptico alrededor de una ventana centrado en un píxel, ec. (1.4).

$$\sum_{(x,y)\in N} W^2(x,y)\left(I_x u + I_y v + I_t\right)^2, \tag{1.4}$$

donde $W(x,y)$ es la ventana centrada en el píxel (x,y) y N es la vecindad.

El mayor inconveniente de este tipo de métodos es que sólo es posible detectar el movimiento en aquellas zonas donde exista variaciones en la imagen. En zonas homogéneas donde puede haber movimiento éste no es detectable. Por ello, los campos de desplazamiento no son densos. Algunos autores han utilizado más funciones que las intensidades de los píxeles para evitar que el problema está mal condicionado, como Mitiche et al. [Mitiche87], Wohn et al. [Wohn83]. Otros, han utilizado operadores diferenciales como Tretiak-Pastor [Tretiak84] o Campani-Verri [Campani90].

Los métodos *de superficie* y *de contornos* realizan una segmentación del flujo óptico de los objetos (o superficies) que se mueven independientemente. Los métodos *de contornos* utilizan la información de los bordes de los objetos para detectar el desplazamiento. Aplican técnicas diferenciales para la extracción de determinadas estructuras en la imagen para luego establecer correspondencias entre estas estructuras. Entre los métodos más relevantes pertenecientes a este grupo destacamos [Hildreth84] y [Buxton84]. Hildreth propone aplicar una restricción de suavizado sobre el flujo estimado a partir de los contornos extraídos de las imágenes en la dirección temporal. En el método de Buxton-Buxton el flujo óptico se calcula en base a un modelo del movimiento de los bordes a lo largo de una secuencia de imágenes. En los métodos de *superficie* destacamos el de Longuet-Higgins y Prazdny [LH80].

Métodos basados en la Correlación

Los métodos basados en la correlación realizan la búsqueda de correspondencias utilizando ventanas o patrones alrededor de cada píxel. La idea que subyace a estos métodos es que es mucho más fácil encontrar las correspondencias entre los píxeles a través de la comparación de regiones entre las imágenes por maximización de alguna medida de similaridad. Una ventaja que tienen estos métodos es que al usar mayor información la búsqueda de las correspondencias es más efectiva.

$$C(\omega) = \int_{\Omega} f(x+\omega)g(x)dx \tag{1.5}$$

donde ω es el desplazamiento del píxel x y Ω es el dominio de la imagen. En el caso discreto, la medida de la correlación en un punto que se suele tomar es la siguiente:

$$C(\omega) = \frac{\displaystyle\sum_{\delta\omega=-(a,b)}^{(a,b)} \left(f(x+\delta\omega) - \overline{f(x)}\right)\left(g(x+\omega+\delta\omega) - \overline{g(x+\omega)}\right)}{(2a+1)(2b+1)\sigma_f(x)\sigma_g(x+\omega)} \tag{1.6}$$

donde (a, b) representa las dimensiones de la ventana de correlación, $\overline{f(x)}$ la media de la imagen f en esa ventana y $\sigma_f(x)$ la desviación estándar.

Dentro de este tipo de métodos podemos destacar, por su relevancia, los métodos de Kories-Zimmerman [Kories86], Sutton et al. [Sutton83], Kalivas-Sawchunk [Kalivas91] y Little et al. [Little88].

Métodos basados en la Frecuencia

Los métodos basados en la frecuencia utilizan la transformada de Fourier para calcular el flujo óptico a través del dominio de la frecuencia. Si adaptamos la expresión que describe cualquier invarianza en el dominio de la imagen, ec. (1.1), al dominio de la frecuencia obtenemos

$$\hat{I}(k, f) - \hat{I}_0(k)\delta(\omega^\top k + f) = 0, \tag{1.7}$$

donde $\hat{I}_0(k)$ es la transformada de Fourier de $I(x, y, 0)$, δ es la delta de Dirac y k, f es la frecuencia espacio–temporal.

Según esta expresión, cualquier patrón que se mueva de una imagen a otra por una simple traslación, se manifiesta en el dominio de Fourier como un cambio de fase. Este tipo de métodos suele resultar muy útil para la detección del movimiento de objetos que son difíciles de capturar, como es el caso de puntos aleatorios. Castro et al. [Castro87] definió una generalización para incluir las rotaciones. Este tipo de métodos siguen presentando el problema de apertura en el dominio de la frecuencia, ec. (1.8), por lo que requieren de algún tipo de término de regularización para solventar el *mal condicionamiento.*

$$\omega^\top k + f = 0. \tag{1.8}$$

Alguno de los trabajos más destacados son Adelson-Bergen [Adelson85], Fleet-Jepson [Fleet90], Heeger [Heeger88], Watson-Ahumada [Watson85].

1.1.2. Métodos Variacionales

La estimación del movimiento en secuencias de imágenes es un problema clave en el campo de visión por ordenador. Como hemos visto en el apartado anterior para resolver este problema se han propuesto un gran número de técnicas. Los métodos variacionales han demostrado ser una de las mejores técnicas y de mayor precisión que existen. La idea que subyace a este tipo de métodos es la definición de una energía que penaliza las desviaciones respecto a las restricciones impuestas en el modelo. Una de las ventajas que ofrecen es que todas estas restricciones están presentes en la energía; no existe ningún tipo de suposición adicional que no se refleje en el modelo y sí en la implementación. A diferencia de otro tipo de técnicas las estimaciones son densas por lo que no es necesario realizar ningún proceso de interpolación. La solución se obtiene directamente mediante la minimización de esta energía. En base a la teoría matemática existente si esta energía cumple una serie de condiciones el proceso de minimización permitirá alcanzar el mínimo global y asegurará la unicidad de dicha solución. Hasta hace unos años, una de las críticas que recibían los métodos variacionales era que carecían de alta precisión y que no eran aptos para aplicaciones en tiempo real. Sin embargo, en los últimos años se han propuesto algunos métodos [Bruhn06b] que ofrecen resultados de altísima precisión con un rendimiento cercano al tiempo real.

La aplicación de los métodos variacionales para la estimación del flujo óptico comenzó con la aportación de Horn-Schunck. En [Horn81] se propuso la minimización de una expresión basada en la ecuación de restricción del flujo óptico al que añadirían un término de regularización que imponía una suavidad en el flujo, ec. (1.3). Con este término de regularización el cálculo del flujo óptico se convertía en un *problema bien condicionado*.

Basándose en las dos aportaciones hechas en [Horn81], (1) ecuación de restricción del flujo óptico y (2) término de regularización, han aparecido en la literatura una serie de contribuciones con el objetivo de mejorar las estimaciones. La *ecuación de restricción del flujo óptico* asume la invarianza de las intensidades de los píxeles entre imágenes. Sin embargo, esta suposición falla en muchas situaciones. Algunos autores proponen mejoras en esta ecuación de forma que se pueda detectar cambios en la iluminación en la escena [Schnörr94a] o la posibilidad de combinar varias invarianzas bajo una misma estructura [Brox02, Bruhn05c, Papenberg06].

El proceso de difusión genera una propagación del flujo óptico entre los píxeles vecinos. Sería deseable que esta propagación se detuviese exactamente en los bordes de los objetos. En el caso de [Horn81] se asume que el campo de desplazamiento varía de forma suave en el espacio. Sin embargo, este término es poco preciso ya que la difusión no se detiene en los bordes de los objetos. Los resultados mejoran considerablemente cuando se fuerza a que el campo de desplazamiento sea suave por trozos. Uno de los primeros trabajos en este sentido fue el desarrollado por Nagel [Nagel83]. Una extensión de [Nagel83] se convirtió en el ampliamente conocido operador *Nagel-Enkelmann* [Nagel86]. En la literatura se han propuesto distintas formas de controlar el proceso de difusión para permitir discontinuidades en el flujo, como [Black96b, Aubert99, Papenberg06], o trabajos basados en los *Markov random fields* [Heitz93, Shulman89], o basados en funciones de robustificación [Nesi93, Schnörr94b, Weickert01a]. Otras contribuciones a mencionar son [Cohen93, Deriche95] que preservan las discontinuidades basándose en la información del flujo. Otras contribuciones dignas de mención son [Proesmans94] y [Anandan89].

En respuesta a los diferentes tipos de regularizadores que se habían propuesto en la literatura Weickert et al. [Weickert01a] sugirieron clasificar los regularizadores en cinco tipos: i) *homogenous* [Horn81], la difusión es la misma en todas las direcciones; ii) *isotropic image–driven* [Alvarez99], el suavizado se aplica en todas las direcciones pero preservando los bordes de los objetos en la imagen; iii) *anisotropic image–driven* [Nagel86], este tipo de regularizadores inhibe el suavizado a través de las discontinuidades en la imagen pero sí lo aplica a lo largo de ellas; iv) *isotropic flow-driven* [Schnörr94b, Weickert98], el suavizado se aplica en todas las direcciones pero preservando las discontinuidades en el flujo; v) *anisotropic flow–driven* [Schnörr94b], suaviza en todas las direcciones excepto a través de las discontinuidades en el flujo.

El ruido es uno de los problemas que más suele afectar a las imágenes ya que altera la información presente en ellas, dificultando la detección de las correspondencias. Las funciones de robustificación se han convertido en una de las herramientas más útiles para para minimizar los problemas que genera el ruido ya que los *outliers* son penalizados en menor grado que una función cuadrática. El uso de estadísticos robustos en los métodos de estimación del flujo óptico los introdujo [Black91, Black96b]. Basándose en la investigación realizada por [Hampel86, Huber81] acerca de los estadísticos robustos, Black et al. propuso el uso de *M-estimators* como funciones de robustificación. En este sentido, cabe destacar el trabajo realizado por [Mémin98a] en el cual el problema de optimización *no-convexa* se resolvió haciendo un ajuste por mínimos cuadrados ponderados e iterativo. Existen otras contribuciones como [Black92, Black96a] en la que también se proponen otras funciones para atenuar el efecto de los *outliers*. En el trabajo [Haussecker01] se propone un método para tratar las variaciones de intensidad de las imágenes. En [Wells96, Viola97, Hermosillo02] se proponen métodos más robustos basados en la estimación de correspondencias multimodal entre imágenes para tratar con secuencias donde existe una transformación compleja en las intensidades de los píxeles. En [Bruhn05c, Papenberg06] se ha demostrado que el uso de funciones de robustificación reduce significativamente el error mejorando la precisión de las estimaciones.

Muchos de los trabajos que hemos comentado utilizan únicamente información espacial para calcular el flujo óptico. En una secuencia de imágenes el movimiento de los objetos se propaga más allá de dos frames. Por ello, la información temporal procedente de los frames anteriores o posteriores nos permite la corrección de las estimaciones y mejorar sustancialmente la precisión de las mismas. Uno de los primeros trabajos que incluyeron información espaciotemporal fue el de Nagel [Nagel90], en donde se propone una extensión temporal de su operador, [Nagel86]. Black y Anandan [Black91] propusieron un método, que asume la existencia de una aceleración en el tiempo, en el que el movimiento se calcula incrementalmente. Para hacer este método más robusto frente al ruido se calcula una media de la aceleración en la componente temporal. Pese a que hace más de quince años se propusieron estos métodos no ha sido hasta hace unos pocos años cuando la información temporal no ha sido ampliamente incluida en los métodos. Algunas contribuciones interesantes son [Elad98, Weickert01b, Farnebäck01]. Weickert y Schnörr [Weickert01b] propusieron un modelo continuo con término de suavizado espaciotemporal convexo nolineal. En este trabajo tanto las derivadas espaciales como temporales se formularon de una forma homogénea. En Brox et al. [Weickert04], la función de robustificación conocida como *Total Variation* se aplica tanto en el término de ligadura como en el de suavizado. Este trabajo permite detectar los largos desplazamientos e incluye en el término de ligadura dos invarianzas: la ya tradicional suposición lambertiana y el gradiente constante. En el término de suavizado se asume que el flujo es suave tanto en la dirección espacial como temporal. Recientemente, Papenberg et al. [Papenberg06] ha obtenido unos resultados excelentes gracias a la combinación de múltiples elementos, como son el uso de funciones de robustificación, la inclusión de varias invarianzas dentro del término de ligadura y la información espaciotemporal.

Los métodos más sofisticados hasta ahora combinan las técnicas variacionales con los *level sets*. Amiaz et al. [Amiaz06, Amiaz07] propone un método que incluye la técnica de segmentación en el modelo de energía variacional. Brox et al. [Brox06] han propuesto un trabajo muy similar que combina el método de Papenberg et al. [Papenberg06] e incluyen información de level sets. Uno de los últimos trabajos más relevantes en el campo de la estimación del flujo óptico es el de [Zimmer09]. En él se propone un modelo de energía que introduce el concepto de complementariedad entre el término de ligadura y suavizado al mismo tiempo que utiliza alguna de las técnicas más innovadoras de la literatura. Otras de las estrategias que se están investigando es la inclusión de técnicas para la detección de oclusiones.

Los métodos actuales han llegado a un alto nivel de sofisticación y complejidad. Cada nuevo elemento que se incluye en el modelo de energía requiere un nuevo parámetro. Esto conlleva que la optimización de los parámetros sea una tarea que consume mucho tiempo. En los últimos años han surgido algunos métodos conocidos como *learning optic flow* [Roth07, Sun08, Li08] cuya característica principal es la utilización de un único vector de parámetros para cualquier secuencia. Para obtener este único vector de parámetros el método se *entrena* con una base de datos de secuencias. En cierto modo, ese vector de parámetros aglutina la mejor optimización para cualquier tipo de movimiento recogido en dicha base de datos. Estos parámetros no son los óptimos para una determinada secuencia pero de forma genérica son los mejores para cualquier secuencia.

A la hora de formular el modelo de energía se asume que el flujo óptico es asimétrico, es decir, que el desplazamiento se produce en una sola dirección, de una imagen a otra. En la literatura se han realizado una serie de trabajos [Cachier00, Christensen01], Alvarez/Deriche/Sánchez/Weickert (2002, 2007) [Alvarez02a, Alvarez07a], Alvarez/Castaño/García/Krissian/Mazorra/Salgado/Sánchez (2007) [Alvarez07b] en el que se describen modelos simétricos del flujo óptico donde la solución se obtiene como una combinación del flujo en ambas direcciones.

Como se ha visto en secciones anteriores la *ecuación de restricción del flujo (OFC)* se basa en la idea de la existencia de las derivadas. Esta suposición no es válida cuando los desplazamientos son largos ya que la condición de derivabilidad no se cumple. Para poder abordar este problema y utilizar las técnicas existentes para desplazamientos cortos, en la literatura han surgido lo que se conoce como *estrategias multiescala*. La idea consiste en crear distintas escalas (versiones) de la secuencia de imágenes de entrada, de forma que los desplazamientos en cada una de ellas se vaya reduciendo llegando a un punto en el que la *OFC* es computable. Este desplazamiento corto en la escala inferior representa el máximo desplazamiento en la secuencia inicial. El flujo óptico calculado en la escala inferior será utilizado como inicialización en la escala superior. Dado que en la escala inferior no existe ninguna estimación previa se aplica el método variacional directamente. Existen varias estrategias para crear cada una de las escalas, entre las que destacamos el enfoque piramidal y gaussiano.

La estrategia multipiramidal consiste en crear por cada imagen una familia de subimágenes, aplicándole un factor de escalado en cada nivel. Este enfoque ha sido utilizado en gran número de métodos como en [Anandan89, Battiti91, Luettgen94, Bornemann96], [Enkelmann88, Mémin02]. Esta estrategia ofrece dos importantes ventajas. Por un lado, este tipo de métodos mejora el rendimiento ya que el flujo óptico se calculará más rápidamente en las escalas inferiores y con la inicialización en las superiores la convergencia del método será más rápida, [Bruhn05b]. Por otro lado, en los funcionales de energía no-convexo la convergencia al mínimo global no está asegurada. En las escalas inferiores del esquema multipiramidal la posibilidad de alcanzar mínimos locales indeseados desaparece, creando buenas inicializaciones que disminuye el riesgo de alcanzar un mínimo local irrelevante en las escalas superiores y mejorando así, las estimaciones obtenidas ([Black96b, Mémin98b, Bruhn05c, Papenberg06]).

Cuando se aplica una gaussiana a una imagen se genera un efecto de suavizado difuminando los bordes. Si aplicásemos sucesivas gaussianas aumentaríamos ese efecto sobre la imagen. En cierta forma, los desplazamientos largos se van reduciendo por el suavizado. En esta estrategia las escalas se crean a partir de sucesivas aplicaciones de gaussianas con desviación σ; el tamaño de las imágenes no varía. Un trabajo que utiliza esta estrategia es Alvarez/Weickert/Sánchez (2000) [Alvarez00].

1.2. Visión Estereoscópica

En la visión estereoscópica tenemos dos vistas de la misma escena en el mismo instante de tiempo. Cada punto 3D (M) se proyecta en ambas cámaras, de forma que, existen dos proyecciones m y m' de cada punto 3D. Se define como *disparidad* al desplazamiento entre el punto 2D m situado en la imagen izquierda (I_l) y su punto en correspondencia m' en la imagen derecha (I_r). Por lo tanto, la visión estereoscópica se reduce a un problema de correspondencias. En este sentido guarda una gran relación con el cálculo del flujo óptico. Ambos problemas tienen muchas similitudes en cuanto a las características y dificultades, de forma que muchas de las ideas que se aplican en un problema se adaptan al otro.

El objetivo de la visión estereoscópica es el de reconstruir la escena a partir de la proyección de un par estéreo. Para poder recuperar la información 3D es necesario: (1) calibrar el sistema de cámaras, (2) estimar las correspondencias entre las imágenes y, por último, (3) calcular los puntos 3D a partir de las correspondencias. Salvo este último paso los dos anteriores han sido objeto de amplio estudio y existen muchísimos trabajos que proponen distintas técnicas para su resolución.

1. **Calibración de cámaras**: Para la estimación de los mapas de disparidad es necesario que las cámaras estén *calibradas*. En la literatura se definen dos tipos de calibraciones: *fuerte*, que consiste en hallar las matrices de proyección asociadas a cada cámara y *débil*, cuyo objetivo es la estimación de la matriz fundamental, que relaciona la información entre las cámaras mediante una transformación lineal. La matriz fundamental ha sido utilizada en trabajos como [Faugeras93a, Faugeras01, Hartley03].

2. **Estimación de las correspondencias**: La búsqueda de correspondencias, como se ha visto en el caso del cálculo del flujo óptico, se trata de un problema bastante complejo. Sin embargo, es posible simplificarlo basándonos en la *geometría epipolar*. Dado que la *geometría epipolar* ya ha sido descrita en varios trabajos ([Faugeras01, Hartley03]) en el apartado 1.2.1 sólo se mencionará los conceptos básicos para que el lector no tenga que recurrir a documentación adicional.

3. **Reconstrucción de los puntos 3D**: Dado que las imágenes del par estéreo se toman en el mismo instante de tiempo, el desplazamiento de los píxeles nos da información acerca de la profundidad de los objetos respecto a la cámara. Píxeles situados cerca de la cámara tendrán un desplazamiento mayor que los situados a mayor distancia. Los mapas de disparidad son unas imágenes, normalmente expresadas en niveles de grises, que nos indican el desplazamiento de cada uno de los píxeles de la imagen. Esta distancia se suele expresar como la norma del vector desplazamiento $\sqrt{u^2 + v^2}$, donde u y v representan el desplazamiento horizontal y vertical respectivamente.

   Si dispusiésemos de un sistema de cámaras calibradas fuertemente, sería posible reconstruir el punto 3D a partir de las de las correspondencias de los puntos 2D. Sin embargo, en la práctica observamos que existen muchos factores que dificultan la correcta reconstrucción de los puntos 3D. Estos factores se comentarán en el apartado 1.3.2.

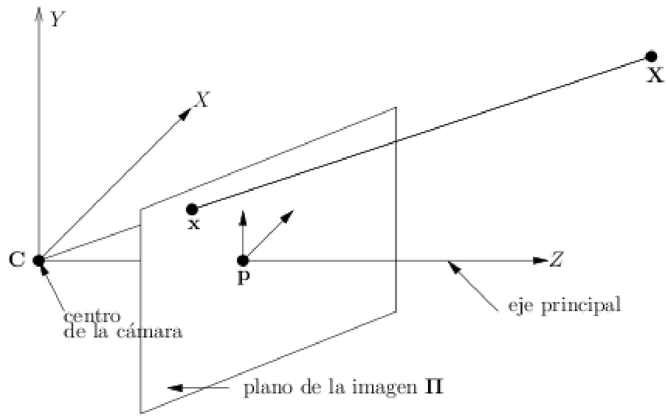


Figura 1.2: Geometría de una cámara proyectiva. **C** es el centro de la cámara y **p** es el punto principal.

1.2.1. Geometría Epipolar

El problema estéreo relaciona la información captada por dos cámaras en el mismo instante de tiempo. Esta tarea no es sencilla dado que se utiliza dispositivos independientes situados en posiciones distintas. Aunque dos cámaras sean exactamente iguales, mismo modelo, marca y configuración, las imágenes que toman no son idénticas ya que hay que tener en cuenta que se tratan de dispositivos físicos y el desgaste de sus piezas no es exactamente el mismo.

Cualquier cámara fotográfica se puede representar mediante el modelo de *pin-hole* (figura 1.2) de forma que los puntos de la escena se proyectan a través del foco de la cámara (C) en el plano proyectivo o imagen, $\mathbf{\Pi}$. La línea desde el centro de la cámara hasta el plano de la imagen recibe el nombre de *eje principal* o *rayo principal*, y el punto donde el eje principal corta al plano de la imagen recibe el nombre de *punto principal*. El plano que contiene al centro óptico y paralelo al plano de la imagen es nominado *plano principal*. En [Faugeras93a] se introduce de manera formal la geometría proyectiva y se describe la cámara proyectiva. Una cámara se puede modelar como una matriz de proyección $\tilde{\mathbf{P}}$, de tamaño 3×4, que transforma un punto 3D, $M = [X, Y, Z]^t$, en unas coordenadas de imagen 2D, $m = [x, y]^t$. La matriz de proyección incluye información interna y externa sobre la cámara. Los *parámetros extrínsecos* lo forman el vector de traslación y la matriz de rotación de la cámara respecto al sistema de referencia global. Los *parámetros intrínsecos* representan ciertos parámetros que determina cómo los objetos de la escena son proyectados en el plano de la imagen y lo conforman: origen de la imagen, distancia focal, tamaño del píxel y desviación de los ejes.

En el problema estéreo intervienen dos cámaras, cada una de ellas dispone de una matriz de proyección. Cada punto 3D se proyectará generando dos puntos, m y m', en las dos imágenes; estableciendo la siguiente relación entre las matrices de proyección y los puntos

$$m = \tilde{\mathbf{P}}M \quad , \quad m' = \tilde{\mathbf{P}}'M. \tag{1.9}$$

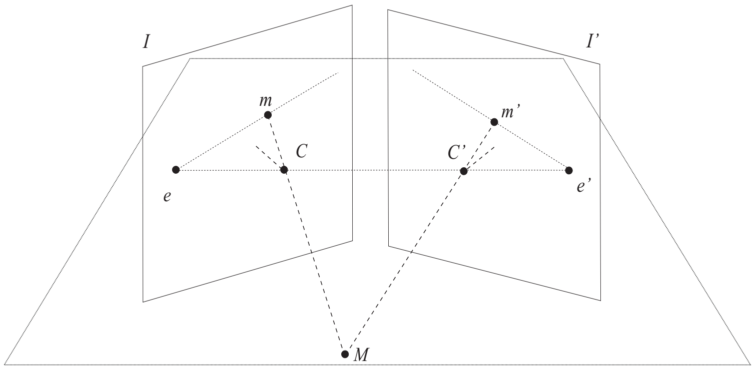


Figura 1.3: Geometría epipolar. I e I' son las imágenes del par estéreo. C y C' son los focos de las cámaras. e y e' los epipolos. m y m' son las proyecciones del punto 3D M.

La intersección de la recta que une los focos de las cámaras C y C' con los planos de imagen de cada cámara crea dos puntos que se denominan epipolos, e y e'.

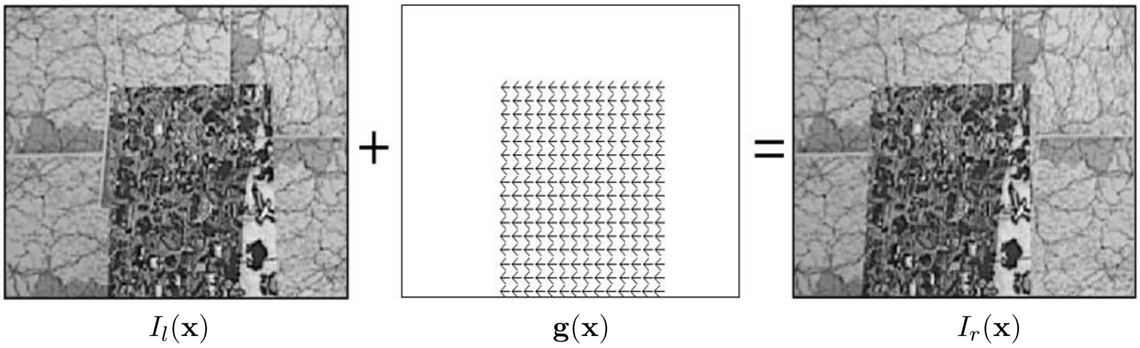
La proyección de la recta que pasa por el punto m y el epipolo de su cámara genera una línea sobre el plano de la imagen de la otra cámara que se conoce como *línea epipolar* (figura 1.3). La característica más relevante es que el punto *m'*, proyección del punto 3D en la otra cámara, pertenece a dicha línea. En principio, el área de búsqueda de correspondencias se reduce a una recta y el desplazamiento de cada píxel se podría expresar en función de un escalar dentro de la línea epipolar. Esto a priori supone una gran ventaja pero requiere un sistema de cámaras correctamente calibrado. Si no fuera así, la geometría epipolar no sería válida y la zona de búsqueda no se reduciría a dicha línea.

No siempre es posible disponer de un sistema de cámaras *fuertemente* calibrado. Por ello, se recurre a una configuración más sencilla que se conoce como *calibrado débil*. En [Luong96] se define la restricción epipolar a través de una matriz, denominada *matriz fundamental*, mostrando la relación que existe entre las matrices de proyección de las cámaras y la matriz fundamental. A partir de la matriz fundamental no se puede calcular las matrices de proyección de las cámaras. Ésta sólo ofrece información relativa de una cámara a la otra, no respecto al sistema de referencia global. En la matriz fundamental se encuentran representados los parámetros intrínsecos – origen de la imagen, distancia focal, tamaño del píxel y desviación de los ejes – y los parámetros extrínsecos – matriz de rotación y vector de traslación – de las cámaras. A continuación, se describe formalmente cómo expresar este desplazamiento basándonos en la geometría epipolar y en la matriz fundamental. Vamos a suponer que el sistema de cámaras está calibrado débilmente y que las cámaras están alineadas horizontalmente. Con esta configuración de cámaras la *matriz fundamental* sería muy simple, ec. (1.10).

$$F = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix}. \tag{1.10}$$

Se define un conjunto de pares de píxeles A, ec. (1.11), donde puede encontrar la correspondencia.

$$A = \left\{ \langle m, m' \rangle \, | m_y = m'_y \text{ y } 0 \leq m'_x - m_x < k \right\}, \tag{1.11}$$



Cuadro 1.2: $I_l(\mathbf{x})$ y $I_r(\mathbf{x})$ son las imágenes izquierda y derecha de un par estéreo. $\mathbf{g}(\mathbf{x})$ representa el mapa de disparidad calculado a partir del par estéreo.

donde (x, y) las componentes horizontal y vertical de cada píxel. k es un umbral que define la zona de búsqueda de la correspondencia dentro del área A. En el caso simple, la línea epipolar.

La ecuación de restricción epipolar $\mathbf{m}'^t\mathbf{F}\mathbf{m} = 0$ establece que dos puntos están en correspondencia, $\mathbf{m} = (\mathbf{x}, 1) = (x, y, 1)$ y $\mathbf{m}'^t = (\mathbf{x}', 1) = (x', y', 1)$, en dos imágenes están definidas por la matriz fundamental, $\mathbf{F}$ [Faugeras01, Hartley03]. Esta definición permite estimar el flujo estéreo solamente sobre ciertas líneas. De algún modo reduce la zona de búsqueda de las correspondencias siguiendo las siguientes ecuaciones:

$$\begin{aligned} a(\mathbf{x}) &= f_{11}x + f_{12}y + f_{13} \\ b(\mathbf{x}) &= f_{21}x + f_{22}y + f_{23} \\ c(\mathbf{x}) &= f_{31}x + f_{32}y + f_{33} \end{aligned}$$

Utilizando esta notación, la línea epipolar Δ se puede escribir como

$$a(x, y)x' + b(x, y)y' + c(x, y) = 0. \tag{1.12}$$

En el cuadro 1.2, llamamos al flujo estéreo como $\mathbf{g}(\mathbf{x}) = (u(\mathbf{x}), v(\mathbf{x}))^t$. El flujo estéreo depende de una función escalar $\lambda(\mathbf{x})$ y define una distancia en la línea epipolar de la siguiente forma:

$$\begin{aligned} u(\mathbf{x}) &= \frac{-\lambda(\mathbf{x})b(\mathbf{x})}{\sqrt{a^2(\mathbf{x})+b^2(\mathbf{x})}} - \frac{a(\mathbf{x})x+b(\mathbf{x})y+c(\mathbf{x})}{a^2(\mathbf{x})+b^2(\mathbf{x})}a(\mathbf{x}) \\ v(\mathbf{x}) &= \frac{\lambda(\mathbf{x})a(\mathbf{x})}{\sqrt{a^2(\mathbf{x})+b^2(\mathbf{x})}} - \frac{a(\mathbf{x})x+b(\mathbf{x})y+c(\mathbf{x})}{a^2(\mathbf{x})+b^2(\mathbf{x})}b(\mathbf{x}) \end{aligned}$$

En Alvarez/Deriche/Sánchez/Weickert (2002) [Alvarez02b] podemos encontrar descrito con más detalle la parametrización anterior.

1.2.2. Clasificación de los Métodos

Los métodos de estéreo se pueden agrupar en cuatro categorías: (i) basados en *características*, (ii) basados en *áreas*, (iii) basados en *frecuencia* y (iv) basados en *energías*.

En [Brown03] podemos encontrar un resumen de las distintas técnicas existentes y los últimos avances en visión estereoscópica.

Basados en *Características*:

Los métodos que pertenecen a este grupo establecen las correspondencias basándose en algún tipo de característica extraída de las imágenes [Grimson85], como curvas [Brint90, Robert91, Nasrabadi92], líneas [Medioni85, Ayache87, McIntosh88] o los bordes de los objetos [Ohta85, Pollard85]. Por un lado, con este tipo de métodos es posible estimar mapas de disparidad con muy poca información (características extraídas de la imagen). Por otro lado, las soluciones obtenidas no son densas.

Basados en *Áreas*:

Los mapas de disparidad se calculan a partir de la correlación de ciertas zonas de la imagen asumiendo que existe algún tipo de similaridad [Scharstein02, Devernay94, Faugeras93b, Fua93, Nishihara84]. Con este tipo de métodos se obtienen muy buenos resultados si se aplican sobre pares estéreo altamente texturados.

Basados en *Frecuencias*:

Los métodos basados en frecuencia utilizan la información de las imágenes en el dominio de Fourier [Froehlinghaus96, Barron94, Fleet90, Fleet93, Jenkin94, Kuglin75, Wiklund92].

Basados en la *Energía*:

En este tipo de métodos la disparidad se calcula a partir de la minimización de una energía que penaliza las desviaciones respecto a las restricciones impuestas en el modelo. El mapa de disparidad se obtiene tras la resolución de un sistema de ecuaciones diferenciales parciales (EDP's), normalmente del tipo difusión-reacción. Este tipo de ecuaciones diferenciales se componen de dos términos: uno de reacción (lo que hasta hora hemos llamado término de ligadura) y otro de difusión (de regularización o suavizado).

Algunos autores proponen distintas formas de clasificar los métodos basados en energías. Al igual que Barron et al. [Barron94] para el caso del flujo óptico unos autores agrupan los métodos en función de la densidad de los mapas de disparidad: *locales* y *globales*. Otros lo hacen en función del tipo de información que utilizan para realizar las estimaciones: *probabilísticos* y *variacionales*.

Los mapas de disparidad en los métodos locales no son densos mientras en los globales sí. Al igual que ocurría con los métodos que estimaban el flujo óptico, los métodos locales calculaban la disparidad de cada píxel a partir de la información dentro de una ventana centrada en dicho píxel. Para ello, utilizaban alguna característica de la imagen ya sea en color o en escala de grises. Los métodos más representativos son [Kanade94, Yoon05]. Sin embargo, en los métodos globales es necesario la inclusión de un término de regularización para convertir el problema en bien condicionado. A su vez, estos métodos se pueden dividir en varios subtipos: graph–cuts [Kolmogorov01, Boykov04], programación dinámica [Lei06], *belief propagation* [Klaus06] o métodos variacionales [Shah93, Robert96, Mansouri98, Alvarez02b, Kim03, Slesareva05].

Los métodos probabilísticos calculan las estimaciones en función de la disparidad más *probable*, de forma que ésta surge de la minimización de una energía discreta ([Kolmogorov01, Lei06, Klaus06]). Las soluciones obtenidas por este tipo de métodos ofrecen muy buenos resultados, aunque debido a su naturaleza discreta, éstas son en precisión entera. Aunque la precisión de las soluciones puede ser suficientemente para muchas aplicaciones como segmentación o la estimación del mapa de disparidad, no lo es tanto para otras como reconstrucción 3D donde se requiere una precisión mayor. Li y Zucker [Li06] comentan que uno de los inconvenientes que tiene este tipo de técnicas es cuando no se cumple la suposición de la disparidad constante por trozos, hecho que ocurre cuando la profundidad en la escena varía suavemente. Un ejemplo de este fenómeno lo podemos observar en los experimentos hechos en esta tesis con la secuencia del *pasillo*. En los últimos años se han propuesto nuevos métodos que se basan en la teoría de *graph–cuts* [Roy98, Ishikawa98, Kolmogorov02, Boykov04, Kolmogorov04]. Todos ellos son una extensión de la idea *máximo–mínimo* presentada originalmente en [Greig89, Wu93, Roy99, Bobick99].

Métodos Variacionales

Los métodos variacionales no tienen algunas de las limitaciones que ofrecen los probabilísticos. Por un lado, dada la naturaleza continua de los funcionales de energía que definen, las soluciones obtenidas son densas y con precisión subpíxel. Por otro lado, todas las restricciones impuestas en el modelo están presentes en la definición de la energía y existe una sólida teoría matemática subyacente.

El término de ligadura asume que una determina propiedad en las imágenes del par estéreo no varía. En trabajos como [Robert96, Mansouri98, Alvarez02b] se ha utilizado la suposición lambertiana. En otros, como [Slesareva05], se ha incluido varias invarianzas dentro de ese término. [Slesareva05] es una adaptación del método de [Papenberg06] para el caso estéreo. Otra contribución a destacar ha sido la de [Ari07] que ha utilizado en el término de ligadura la norma L1, a diferencia del tradicional término cuadrático ampliamente usado en la literatura.

Las similitudes entre las ideas propuestas para el problema del flujo óptico y estéreo se reflejan también en la clasificaciones hechas en los términos de regularización. Los términos de regularización se pueden agrupar en función de la información utilizada en la difusión (*image–driven*, *disparity–driven*) o en función de la dirección en el que se aplica (*isotrópico*, *anisotrópico*). Los regularizadores anisotrópicos siempre han ofrecido mejores resultados que los isotrópicos ya que son capaces de ajustar la difusión en los bordes de los objetos. En la literatura nos encontramos algunos ejemplos de regularizadores *anisotrópicos image–driven* como [Alvarez02b, Mansouri98], en los que se preservan los bordes utilizando la información de la imagen. Se basan en la idea que los píxeles con gradiente grandes reflejan una discontinuidad en la escena (objetos distintos). Sin embargo, en el caso de imágenes texturadas esta suposición no siempre se cumple. El segundo tipo de regularizadores son los *disparity–driven*. En trabajos como [Slesareva05, Ari07] se utilizan este tipo de regularizadores que, con un comportamiento similar a los *flow-driven* en flujo óptico, suavizan en función de la información del mapa de disparidad. [Shah93] propuso el uso de una restricción de suavizado que preservara las discontinuidades basándose en la disparidad. Posteriormente, [Robert96] describió un framework para la definición de regularizadores *disparity-driven*. En [Zimmer08] se ha propuesto un método que incluye un regularizador anisotrópico *disparity–driven* que ofrece dos ventajas. Por un lado, se aprovecha de la precisión del anisotrópico y, por otro lado, el *disparity-driven* evita la sobresegmentación del mapa disparidad como ocurre en los regularizadores *image-driven*.

Las funciones de robustificación se han utilizado en el problema del flujo óptico con dos fines: para atenuar los efectos del ruido en las imágenes ([Black91, Black96b, Bruhn05c]) o para controlar el proceso de difusión ([Nesi93, Schnörr94b, Weickert01a]). Tradicionalmente, en el caso estéreo, se ha utilizado como término de regularización la función de Tikhonov, [Shah93]. En otros trabajos, [Shah91, Kumar97, Slesareva05], se empleado la función conocida como *Total Variation (TV)* o la propuesta por Perona-Malik, [Perona90]. La función de *Total Variation* tiene la ventaja que carece de parámetros y ha demostrado ser muy efectiva ante la presencia de *outliers*. Sin embargo, la función de Perona-Malik crea discontinuidades más nítidas en el flujo.

Las oclusiones son unos fenómenos que se producen por la existencia de píxeles que sólo son visibles en una imagen del par estéreo. Hasta hace unos años los métodos no incluían ningún tipo de técnica para la detección de las oclusiones [Robert96, Kumar97, Mansouri98, Alvarez02b, Slesareva05]. En [Ari07] se propone un método que ofrece muy buenos resultados y que aglutina varias técnicas muy exitosas. Está basado en el framework de Mumford-Shah [Mumford89], tiene un término de regularización *disparity–driven* y maneja las oclusiones de forma similar a [Shah93].

Las técnicas multiescala utilizadas para la detección de desplazamientos largos desarrolladas para la estimación del flujo óptico siguen siendo válidas en el problema estéreo. De hecho, existen trabajos como Alvarez/Deriche/Sánchez/Weickert (2002) [Alvarez02b] en el que se describe un método para la estimación de la disparidad para desplazamientos largos que utiliza un enfoque similar a Alvarez/Weickert/Sánchez (2000), [Alvarez00].

1.3. Características de las Imágenes

En este apartado se describen las características de las imágenes y la complejidad que puede tener una escena. Conviene tenerlas en cuenta a la hora de diseñar los modelos de energía ya que los métodos variacionales propuestos para la estimación del flujo óptico y del mapa de disparidad se basan en (i) la idea de invarianza sobre alguna propiedad de las imágenes y en (ii) una restricción sobre el desplazamiento de los píxeles.

1.3.1. Secuencias de Imágenes

El sensor de una cámara captura la luz procedente de una escena. Esta energía puede proceder directamente de fuentes de luz o reflejada por los objetos, pudiendo alterar la apariencia de la escena. Teniendo en cuenta esto, podemos analizar la información de la imagen desde dos puntos de vista distintos: *fotométrico* o *geométrico*. Los *aspectos fotométricos* hacen referencia a la incidencia de la luz en la escena, ya sea directamente sobre la cámara o indirectamente a través de la energía reflejada por los objetos. Aquí interviene de forma especial algunas propiedades de los objetos, como puede ser la absorción de luz, reflectancia, etc. Los *aspectos geométricos* hacen referencia a la incidencia que puede ocasionar la presencia e interacción de los objetos entre sí. Cuando varios objetos interactúan pueden darse situaciones de impacto entre ellos, ocultación parcial o total, o simplemente que aparezca y desaparezcan del plano de la cámara. Por este motivo, conviene repasar tanto los aspectos fotométricos como geométricos para llegar a comprender el porqué de algunas de las aportaciones que se han hecho en la literatura.

Aspectos Fotométricos

Los aspectos fotométricos están relacionados con la luz presente en la escena, su incidencia en los objetos y otras características ópticas. Aquí juega un papel fundamental tanto las propiedades de la luz como las de la superficie de los objetos. Si atendemos a la superficie de los objetos podemos reducirlas a dos casos: *superficies lambertianas* o *superficies especulares*. Las *superficies lambertianas* reflejan de manera uniforme la luz, independientemente del ángulo de incidencia. Por el contrario, las *superficies especulares* reflejan la energía en función del ángulo de incidencia. La mayoría de los métodos que calculan el flujo óptico asumen que la superficie de los objetos es *lambertiana*. Sin embargo, esta suposición falla cuando tratamos con secuencias reales. El cambio de posición de los objetos o de la fuente de luz modifica el ángulo de incidencia lo que provoca cambios en los brillos de los objetos.

En el trabajo de [Mileva07] se ha hecho un amplio estudio sobre distintas invarianzas fotométricas que se pueden utilizar en los métodos variacionales para la estimación del flujo óptico. Se investigó el uso de tres estrategias diferentes: (i) técnicas de normalización, (ii) diferenciación de canales a través de logaritmos y (iii) la transformación de los canales RGB a otros espacios esféricos/cónicos, como HSV. Estas invarianzas pretendían analizar la robustez de las métodos frente a sombras y cambios de iluminación en secuencias de imágenes tanto reales como sintéticas.

A continuación, se describen los principales aspectos fotométricos que nos podemos encontrar en una escena real.

- **Fuentes de luz:**

  Las fuentes de luz presentes en una escena determinan la apariencia de los objetos. La posición, tipo e intensidad de las mismas puede cambiar la intensidad de los píxeles, generar sombras, alterar los colores, generar brillos y variar notablemente la información tomada por las cámaras. En la visión por computador el control de la iluminación es fundamental para asegurar la correcta resolución de un problema. Para ello, existen dos estrategias: (1) controlar la iluminación de una escena; (2) utilizar en los métodos técnicas para la atenuación o eliminación de las perturbaciones presentes en la escena, ya sean luminosas o de otra naturaleza.

En situaciones reales no siempre es posible controlar la iluminación de la escena. Actualmente se han propuesto en la literatura una serie de técnicas para incrementar la robustez de los métodos para la estimación del flujo óptico pero, como veremos en esta sección, existen muchas perturbaciones en la escena difíciles de detectar.

Las fuentes de luz se pueden catalogar: según su *naturaleza* (natural, artificial) o por la *forma en la que produce la emisión* (térmico, luminiscente, fotoluminiscencia).

Si tenemos en cuenta la naturaleza de la luz podemos clasificarla como luz natural o artificial. En la luz natural tiene mucha importancia la hora del día, las condiciones meteorológicas y la época del año, ya que pueden alterar notablemente la percepción de los colores y los brillos capturados por la cámara. La luz artificial, ya sea fluorescente, halógena, o de otro tipo suele alterar las tonalidades de los colores y crear brillo en la superficie de los objetos.

Si analizamos las fuentes de luz respecto a la orientación del foco de luz nos podemos encontrar con dos casos: omnidireccional o puntual. Las fuentes de luz puntual, dependiendo de su intensidad, suelen crean brillos sobre la superficie de los objetos, sombras y alterar las tonalidades de los objetos. Por el contrario, las fuentes de luz omnidireccional crean tonalidades homogéneas y según su intensidad: sombras débiles o ausencia de brillo.

- **Sombras:**

  La sombra se produce en la parte posterior de un objeto cuando la luz incide sobre él. Las sombras también se pueden crear por la ocultación parcial o total de un objeto respecto a una fuente de luz. El grado de oscuridad de la sombra dependerá de la intensidad de la luz, de la presencia de otras fuentes de luz en la escena o de las propias características del objeto.

  Las sombras alteran la percepción de los objetos y su presencia en una escena puede generar la estimación de falsos movimientos (crean desplazamiento de píxeles que no se corresponden con movimiento de objetos), además de dificultar la detección de correspondencias.

- **Propiedades de los objetos:**

  Aparte de las fuentes de luz las propiedades de los objetos determinan la apariencia de la escena que captura la cámara. Como ya se ha dicho, en la estimación del flujo óptico se asume que ciertas propiedades en la imagen no varían. Si estas propiedades modifican en exceso la apariencia del objeto complicará la detección de correspondencias y nos podemos encontrar ante situaciones en las que la percepción del objeto varía notablemente de una imagen a otra. Dos ejemplos claros son las *transparencias* y los *reflejos.*

  En el caso de las transparencias, la superficie de los objetos deja pasar parte de la luz que recibe permitiendo ver los objetos que están detrás de éste. Este tipo de objetos tienen un alto comportamiento especular y su correcta identificación por parte de los métodos es compleja. Los reflejos son productos de la luz rebotada por un objeto. La intensidad del reflejo depende de la capacidad de reflectancia del objeto.

Como se ha visto en este apartado son muchos los factores que modifican la apariencia de los objetos presentes en una escena. Esto dificulta la detección de correspondencias y, con ello, la correcta estimación del flujo óptico.

Aspectos Geométricos

Los aspectos geométricos están relacionados con la interacción de los objetos en la escena. Esta interacción modifica la apariencia geométrica de los objetos. A continuación, comentaremos las situaciones más comunes que suelen aparecer en las secuencias de imágenes.

- **Oclusiones:**

  La oclusión es un fenómeno que se produce por presencia de un objeto que oculta parcial o totalmente a otro. Cuando se produce un movimiento en la escena, ya sea de la cámara o de cualquiera de los objetos, puede darse el caso que la forma de los objetos cambie debido a que parte quede oculto respecto a la vista de la cámara. Las oclusiones crean discontinuidades en el flujo óptico lo que genera la violación de algunas de las restricciones impuestas. La detección de las oclusiones mejora la estimación del flujo óptico en los bordes de los objetos.

- **Plano de la imagen:**

  La cámara sólo es capaz de capturar la información que está enfrente del objetivo y que se refleja en el plano de proyección. Dado las dimensiones del plano de la imagen sólo es posible captar parcialmente la escena por lo que habrá objetos que queden recortados. A medida que estos objetos o la propia cámara se muevan es posible que aparezcan o desaparezcan del plano de la imagen. Esto es un tipo de oclusión provocado por las dimensiones del plano de la imagen.

- **Propiedades físicas de los objetos:**

  Cuando pensamos en los objetos solemos imaginar como figuras geométricas rígidas. Sin embargo, en el mundo real es mucho más complejo y existen objetos de formas y composición variable. Por ello, se puede dar casos de deformación, pérdida de volumen, elasticidad, que modifiquen la apariencia del objeto y compliquen su seguimiento a lo largo de la escena.

- **Ruido:**

  Cuando una cámara capta una escena real es frecuente que en las imágenes aparezcan defectos que alteran ciertas propiedades en la imagen, ya sea por imperfecciones de la lentes, polvo en el aire, u otros factores. Dado que los métodos asumen la constancia de ciertas propiedades de la imagen la presencia de ruido genera una violación de dichas suposiciones lo que dificulta la correcta estimación del movimiento. Algunos métodos incorporan técnicas que corrigen o atenúan los efectos del ruido sobre las imágenes.

1.3.2. Pares Estéreo

Debido a la gran similitud que existen entre el problema del flujo óptico y el estéreo muchas de las características descritas en el apartado anterior son extrapolables a éste. Sin embargo, como es lógico, existen diferencias dada la naturaleza estática o dinámica de la información que captura/n la/s cámara/s en ambos problemas.

A continuación, describiremos brevemente algunas de las peculiaridades de los pares estéreos.

- **Uso de dos cámaras:**

  En la visión estereoscópica el uso de varias cámaras implica que las imágenes obtenidas por cada una de ellas no sean iguales. Aunque las cámaras sean idénticas, misma marca/modelo y parámetros, existirán pequeñas diferencias que harán que los parámetros *intrínsecos* no sean exactamente los mismo provocando que la forma de proyectar de las cámaras sea distinta. Estas diferencias pueden ser corregidas durante el proceso de calibración.

- **Geometría Epipolar:**

  Gracias a la geometría epipolar es posible acotar la búsqueda de correspondencias a una recta, pero requiere disponer de un sistema de cámaras correctamente calibrado.

1.4. Secuencias de Prueba

1.4.1. Flujo Óptico

En esta sección se hace una breve descripción de las distintas secuencias tanto reales como sintéticas que se han utilizado en esta tesis para la evaluación de los métodos de flujo óptico. La mayoría de estas secuencias han sido ampliamente utilizadas en la literatura. Por un lado, tienen por objetivo cubrir todos los tipos de situaciones que se pueden dar en cualquier secuencia y, por otro lado, la estandarización de las secuencias facilita las comparaciones entre los distintos métodos.

Las secuencias sintéticas permiten conocer con total exactitud el desplazamiento de los objetos presentes en la escena. De este modo, podemos evaluar la diferencia entre el movimiento exacto y el estimado. Sin embargo, en este tipo de secuencias no suelen aparecer determinadas perturbaciones que sí hay en el mundo real. Por este motivo, también se han utilizado secuencias reales para analizar el comportamiento de los métodos ante distinto tipo de perturbaciones.

Secuencias Sintéticas

Secuencia del *Cuadrado*

La secuencia denominada *Cuadrado* contiene diez imágenes de tamaño 185×128 píxeles y ha sido creada por el grupo AMI (ULPGC). En ella aparece un cuadrado negro desplazándose horizontalmente de izquierda a derecha sobre un fondo texturado. La velocidad de este cuadrado es uniforme y de quince píxeles. La textura utilizada es una hoja de papel arrugada y también se desplaza unos tres píxeles por frame. En la figura 1.4 podemos ver algunos de los frames que componen esta secuencia.

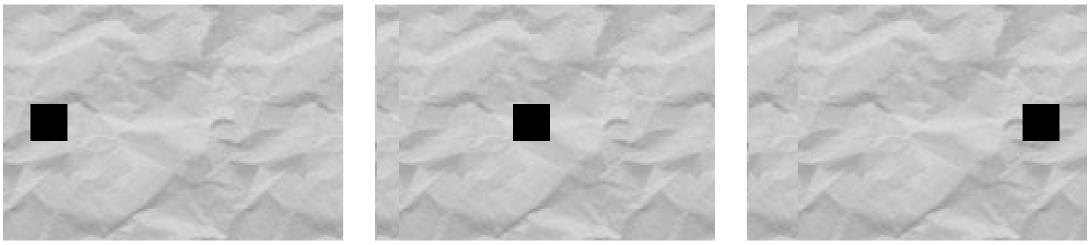


Figura 1.4: Frames 0, 4 y 9 de la *Secuencia0*.

Las características más notables de esta secuencia son el movimiento uniforme del objeto, el largo desplazamiento entre frames y la utilización de una imagen texturada como fondo de la escena. El único objeto presente en la secuencia es el cuadrado negro y no existe ninguna oclusión salvo la del propio objeto con el fondo.

Marble Blocks

La secuencia conocida como *Marble Blocks*, figura 1.5, está compuesta por treinta imágenes de tamaño 512 × 512 píxeles. Fue creada por H.H. Nagel (KOGS/IAKS, Universidad de Karlsruhe, Alemania, `http://i21www.ira.uka.de/ image_sequences/`) y en ella, aparecen cinco torres de mármoles. Cuatro de estas torres, las oscuras, permanecen estáticas en la escena y la de color claro se desplaza hacia la izquierda.

La cámara que capta la escena se desplaza hacia la izquierda por lo que todos los objetos estáticos aparecen con un movimiento hacia la derecha, excepto, el bloque de mármol blanco que también se desplaza hacia la izquierda pero a una velocidad mayor a lo que lo hace la cámara. Por ello, si nos fijamos en el *ground truth* (figuras 1.6 y 1.7) los objetos que más se mueven son los más próximos a la cámara. El menor desplazamiento lo registra el bloque de mármol blanco que sigue el movimiento de la cámara.

En principio, el desplazamiento de los objetos en cada frame son superiores a un píxel y constantes. Sin embargo, hemos detectado que en los frames 4, 8, 13, 18, 23 y 28 existen unas aceleraciones de los objetos respecto a los frames contiguos. Estos cambios en la velocidad pueden ser percibidos sutilmente a simple vista pero cuando calculamos la velocidad media entre frames, utilizando el flujo real suministrado por Nagel, observamos que este valor se incrementa notablemente. La velocidad media en los distintos frames de la secuencia es de $1,33$ con una desviación típica de $0,03$. En los frames especificados anteriormente, la velocidad media se incrementa hasta los $1,58$ con una desviación típica de $0,02$.

Yosemite con Nubes

La secuencia conocida como *Yosemite* fue creada por Lynn Quam. Representa una imagen sintética del parque nacional Yosemite en Sierra Nevada, California. La

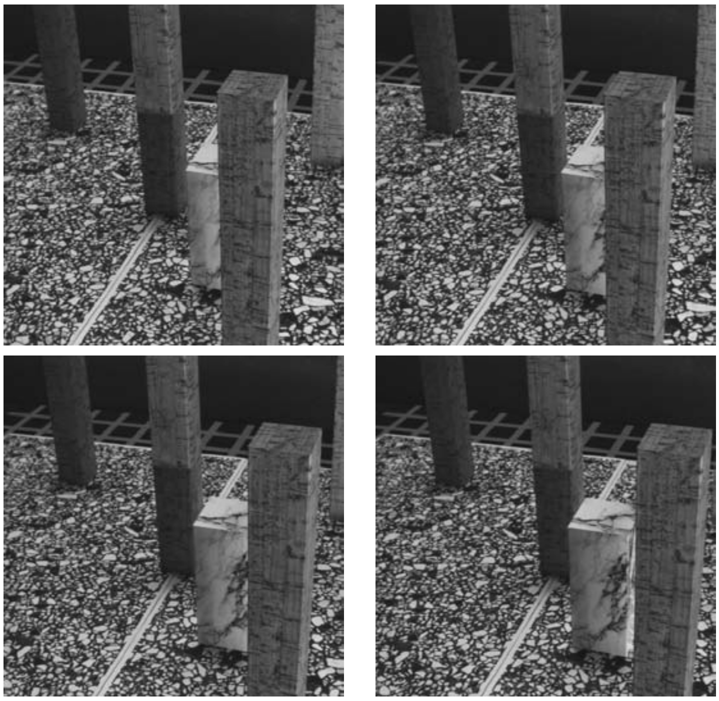


Figura 1.5: Frames 5, 10, 15, y 20 de la secuencia *Marble Blocks*.

cámara simula una navegación aérea que se adentra en el parque. Existen dos versiones de esta secuencia: *Yosemite CON nubes* y *Yosemite SIN nubes*. La diferencia entre ellas está en la presencia o no de nubes en el cielo. En la secuencia donde aparecen las nubes éstas se desplazan en sentido horizontal a una velocidad de dos píxeles por frame. También hay un cambio de iluminación en las nubes. La versión *con nubes* está disponible en `ftp://ftp.csd.uwo.ca/pub/vision` y la de *sin nubes* puede encontrarse en `http://www.cs.brown.edu/people/black/images.html`.

Yosemite está compuesta por quince imágenes de tamaño 316×252 píxeles. En esta secuencia se combinan movimientos traslacionales y divergentes. En la figura 1.8 podemos ver tres frames de la secuencia de Yosemite con nubes en las que se puede observar el movimiento de las nubes y cómo éstas cambian de intensidad a lo largo de la secuencia. En esta zona de la imagen el desplazamiento es traslacional, continuo y constante. En el resto de la imagen, el movimiento es divergente pero muy suave creándose un efecto de zoom que simula un acercamiento al valle. En las figuras 1.9 y 1.10 se muestran el *ground truth* para los frames 8 y 9, en mapas de niveles de grises y en campos de vectores, respectivamente.

En la figura 1.10, se observa cómo el movimiento en la parte central de la imagen es prácticamente nulo. En la parte inferior izquierda donde el movimiento divergente se acentúa. En las flechas de la figura se observan distintas longitudes (velocidades) producto del movimiento divergente reinante. A medida que la cámara sobrevuela *Yosemite* se producen oclusiones en los bordes de las montañas y desaparece terreno debido al efecto *zoom-in*.

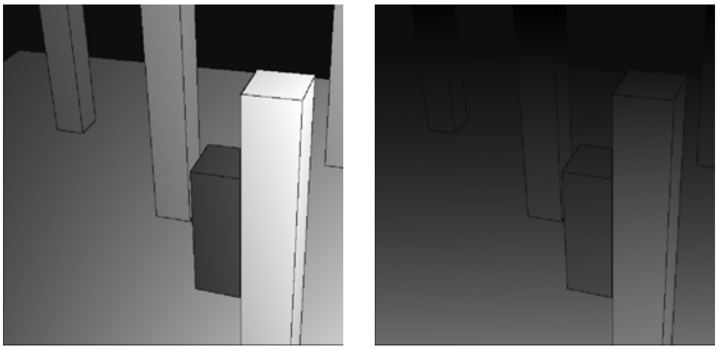


Figura 1.6: De izquierda a derecha, el desplazamiento horizontal y vertical que se registra entre los frames 9 y 10. Los tonos claros representan desplazamientos positivos y los oscuros, negativos. Los píxeles en negro no se tienen en cuenta y corresponden con el fondo de la escena y las aristas de las torres de mármol.

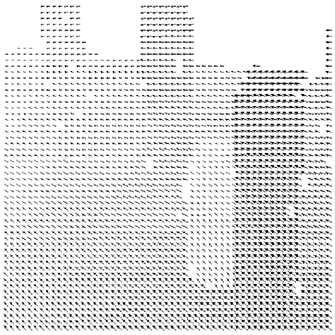


Figura 1.7: Campo de desplazamiento entre los frames 9 y 10 expresado en campo de vectores.

Esta secuencia ha sido ampliamente utilizada en la literatura debido a la diversidad de movimientos presentes en la misma. En la evaluación de los métodos se utiliza el flujo óptico calculado entre los frames 8 y 9. Sólo se dispone del campo de desplazamiento con bastante precisión entre estos dos frames. En el resto de la secuencia los campos de desplazamientos disponibles tienen una precisión menor.

Secuencias Reales

En este apartado se describe las secuencias reales utilizadas en los experimentos. A diferencia de las secuencias sintéticas, éstas presentan una serie de perturbaciones, como *efecto de entrelazado*, *cambios de iluminación*, *sombras*, etc., que dificultan la correcta estimación del flujo óptico. Otro de los inconvenientes es la ausencia de *ground truth* por lo que no es posible obtener resultados cuantitativos.

Taxi

La secuencia del *Taxi* de Hamburgo se trata de una de las más famosas y más utilizadas

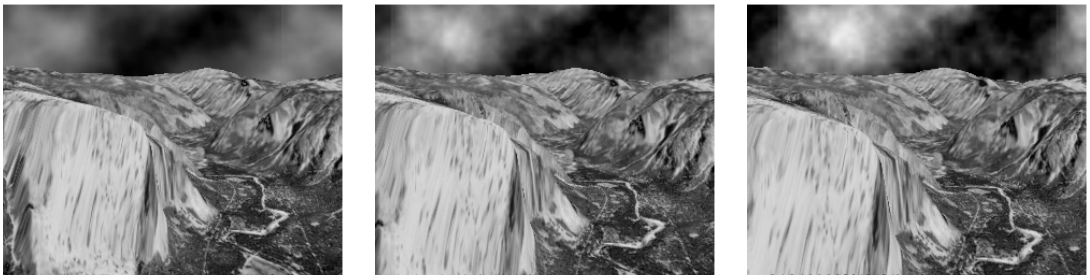


Figura 1.8: Frames 0, 7 y 14 de la secuencia de Yosemite con nubes.

Figura 1.9: Mapas de desplazamiento horizontal y vertical que se registra entre los frames 8 y 9. Los tonos claros representan desplazamientos positivos y los oscuros, negativos.

en los trabajos sobre el cálculo del flujo óptico. Fue creada por H.H. Nagel (KOGS/IAKS, Universidad de Karlsruhe, Alemania, `http://i21www.ira.uka.de/image_sequences/`). Está compuesta por cuarenta frames de tamaño 256×190 píxeles y, en ella, se puede observar el efecto del *entrelazado*, *cambios de iluminación* y *oclusiones*.

En la figura 1.11 podemos ver tres frames que componen esta secuencia. En ella, se aprecia como un taxi ejecuta un giro hacia la derecha para entrar en una calle y un coche oscuro situado en la esquina inferior izquierda circula hacia la derecha a gran velocidad. En la esquina inferior derecha, circula un camión justo detrás del taxi y, por último, en la esquina superior izquierda un peatón se desplaza por la acera a una velocidad muy inferior a la que lo hacen los vehículos. El resto de objetos presentes en la secuencia no se mueven.

Rheinhafen

Otra de las secuencias reales utilizadas en los experimentos es la conocida como *Rheinhafen*. Fue creada también por Nagel (`http://i21www.ira.uka.de/image_sequences`). Se trata de una secuencia en escala de grises compuesta por mil frames de tamaño 688×565 píxeles. Una cámara situada a cierta altura captura el tráfico rodado que circula por una vía. Se observan varios vehículos que circulan en distintas direcciones y velocidades. Próximo a la cámara se capta una furgoneta que circula a gran velocidad. Al fondo, se ven varios vehículos que se mueven a menor velocidad y que se disponen a cambiar de dirección. En esta secuencia se percibe con total claridad el efecto de *entrelazado* en los contornos de la furgoneta situada cerca de la cámara. En la figura 1.12 se muestra dos frames no consecutivos en los que se aprecia con mayor claridad el movimiento registrado en la escena.

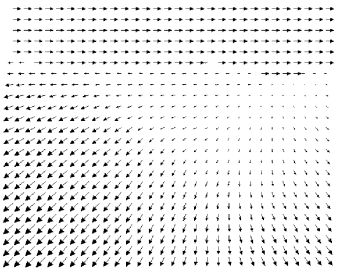


Figura 1.10: Mapa de desplazamiento entre los frames 8 y 9 expresado en campo de vectores.

Figura 1.11: Frames 0, 10 y 19 de la secuencia del *Taxi* de Hamburgo.

Secuencias del Meteosat

Como parte de esta tesis se ha desarrollado un método variacional multicanal para la estimación del movimiento de estructuras nubosas. Para evaluar este nuevo método se han utilizado dos secuencias satélites captadas por el Meteosat. El Meteosat es un satélite de meteorología provisto de una serie de sensores que captan la energía reflejada por la Tierra en distintos rangos de frecuencia. Estas secuencias han sido tomadas por la segunda generación de satélites Meteosat que son capaces de capturar imágenes de diez bits de cuantificación cada 15 minutos, con un tamaño de píxel de tres kilómetros cuadrados para cada uno de los once canales con lo que dispone, desde el infrarrojo hasta el visible.

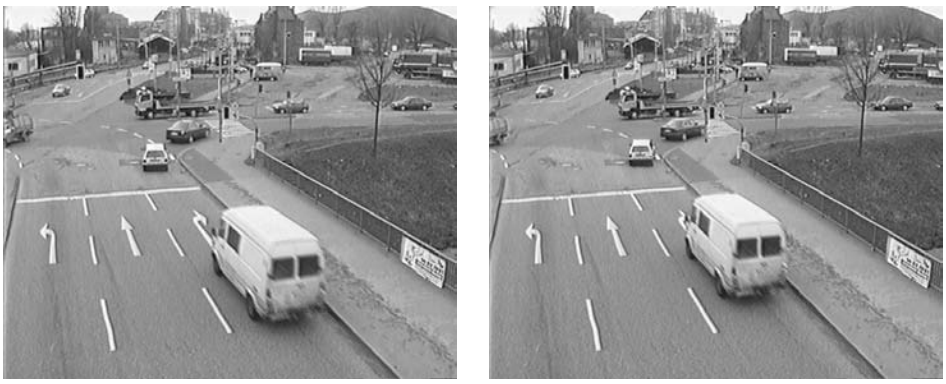


Figura 1.12: Dos frames no consecutivos de la secuencia *Rheinhafen*.

ID del Canal	Longitud de Onda	Aplicación Principal
VIS 0.6	0.63 μm	Detección y seguimiento de nubes, identificación de superficie
VIS 0.8	0.81 μm	Detección y seguimiento de nubes, identificación de superficie
NIR 1.6	1.64 μm	Discriminación entre nieve, hielo y masas nubosas
IR 3.9	3.92 μm	Detección de nubes bajas y nieblas durante la noche
WV 6.2	6.25 μm	Masas de vapor de agua situadas a media altura
WV 7.3	7.35 μm	Masas de vapor de agua situadas a baja altura
IR 8.7	8.70 μm	Distinción entre agua e hielo
IR 9.7	9.66 μm	Detección de ozono en la parte inferior de la estratosfera
IR 10.8	10.80 μm	Estimación de la temperatura de las nubes y de la superficie terrestre
IR 12	12.00 μm	Estimación de la temperatura de las nubes y de la superficie terrestre
IR 13.4	13.40 μm	Estimación de nubes a gran altura
HRV	0.75 μm	Alta resolución espacial

Cuadro 1.3: Características y principales aplicaciones de los canales del Meteosat.

En el canal visible cuenta con un sensor de alta resolución que captura imágenes con un tamaño de píxel de un kilómetro cuadrado [Schmetz02]. Cada uno de los once sensores que el satélite dispone suministran información que es utilizada en distintas aplicaciones. En la tabla 1.3, podemos ver resumido la aplicación asociada a cada canal. Sin embargo, la mayoría de las aplicaciones meteorológicas combinan la información de varios canales, principalmente de estos cuatro, VIS 0.8, WV 6.2, WV 7.3 y IR 10.8 [Schmetz02].

El canal VIS 0.8 capta información del espectro visible. Este canal permite la identificación y seguimiento de las nubes en la atmósfera, la monitorización de la vegetación y de la superficie terrestre. Los canales WV 6.2 y WV 7.3 permiten observar el vapor de agua presente en la atmósfera así como las corrientes de aire. También es posible la localización de nubes semitransparentes situadas a gran altura [Schmetz93]. Por último, el canal IR 10.8 capta información referente a la temperatura de la superficie terrestre, los océanos y de la parte superior de las nubes [Inoue87].

Secuencia del Huracán Vince

Esta secuencia capta la presencia del huracán Vince en el océano Atlántico. Este huracán supone un efecto insólito ya que se formó en una zona situada demasiada al Este de donde se suele producir habitualmente. Este huracán categoría 1 (según la escala Saffir-Simpson) se formó a las 18:00 el 9 Octubre del 2005 al noroeste de Funchal (Islas Madeira). A partir de entonces, empezó a perder fuerza hasta convertirse en una tormenta tropical casi seis horas después [Franklin06]. En la figura 1.13, se puede observar las imágenes

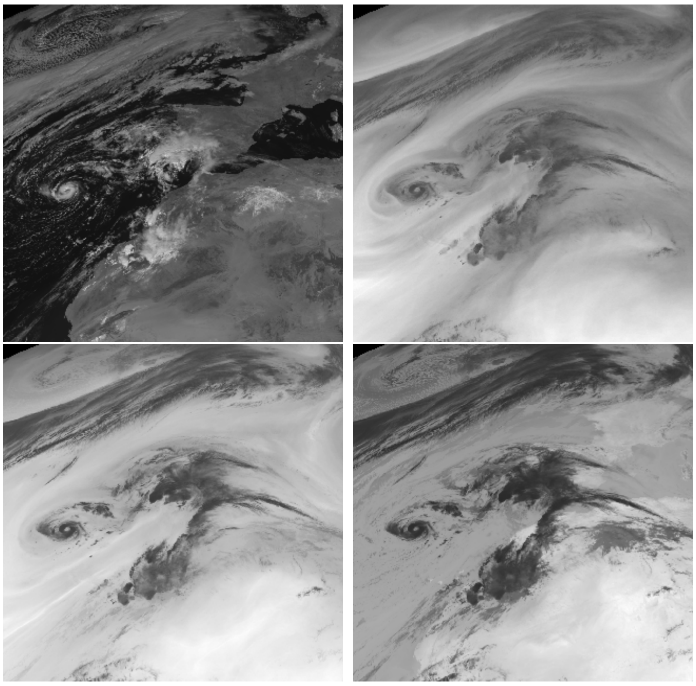


Figura 1.13: Secuencia de *Vince*. De izquierda a derecha y de arriba hacia abajo, se muestra el canal visible 0,81μm, los canales de vapor de agua 6,25μm y 7,35μm y a la derecha el canal de infrarrojo, 10,80μm.

de cuatro canales tomadas por el Meteosat. Las zonas de interés de esta secuencia son dos: el huracán *Vince* y la masa nubosa presente en el Atlántico Norte (próxima a las Islas Británicas). Los movimientos que principalmente se registran en esta secuencia son rotacionales.

Secuencia del Atlántico Norte(June 5th, 2004)

Esta secuencia fue tomada por el Meteosat el 5 de Junio del 2004. En ella se muestra los fenómenos atmosféricos acontecidos aquel día en el hemisferio norte en la franja europeo-norteafricana. En la figura 1.14, se puede observar las imágenes de la atmósfera de cuatro canales captadas por el Meteosat. En esta secuencia cabe destacar las dos grandes masas nubosas. Por un lado, la procedente del Atlántico Norte aproximándose hacia las islas Británicas y, por otro lado, la gran masa nube que se aproxima desde el oeste hacia la península Ibérica. El movimiento predominante en estas dos grandes masas nubosas es rotacional. En el resto de la secuencia el movimiento es menos severo y principalmente traslacional.

1.4.2. Visión Estereoscópica

Al igual que hacíamos en el apartado anterior vamos a comentar las distintas secuencias de pares estéreo que se han utilizado en los experimentos. Durante las pruebas se han

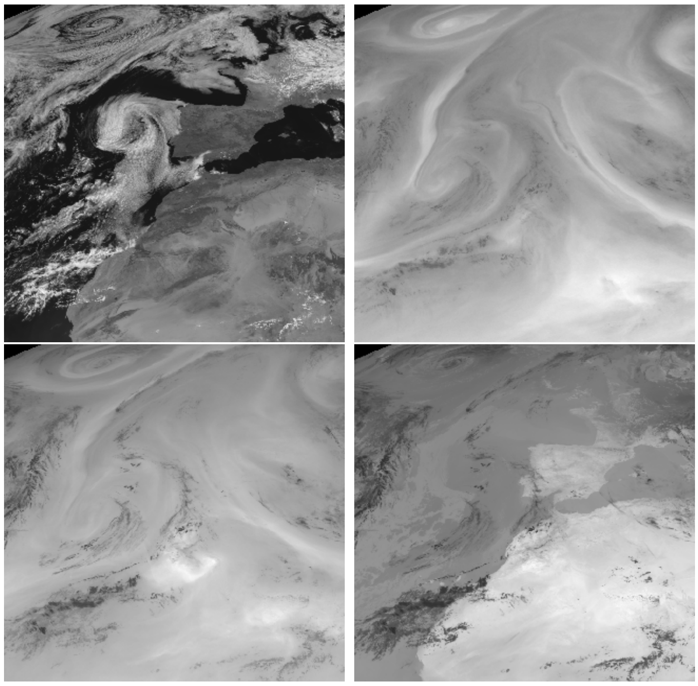


Figura 1.14: Secuencia del *Atlántico Norte*. De izquierda a derecha y de arriba hacia abajo, se muestra el canal visible 0,81μm, los canales de vapor de agua 6,25μm y 7,35μm y a la derecha el canal de infrarrojo, 10,80μm.

empleado tanto imágenes reales como sintéticas. Algunas de ellas han sido ampliamente utilizadas en la literatura y otras han sido creadas por miembros del grupo AMI (ULPGC).

Un par estéreo se compone de dos imágenes tomadas en el mismo instante de tiempo. Para establecer las correspondencias entre los puntos es necesario que el sistema de cámaras está calibrado. Existen múltiples configuraciones pero, normalmente, para simplificar el problema de la estimación de los mapas de disparidad se suele situar ambas cámaras de forma frontoparalela. De este modo, los focos están situados en el mismo plano de proyección y las líneas epipolares son paralelas. En las cámara se ha seleccionado una configuración horizontal. El desplazamiento de los objetos sólo ocurre en una sola dirección. La matriz fundamental asociada a las cámaras en este caso es

$$F = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \tag{1.13}$$

En todas las secuencias sintéticas que hemos utilizado en los experimentos los focos de las cámaras están situadas en el mismo plano horizontal. En las secuencias reales se ha recurrido a un proceso de rectificación para los desplazamientos se produjesen en un sólo eje.

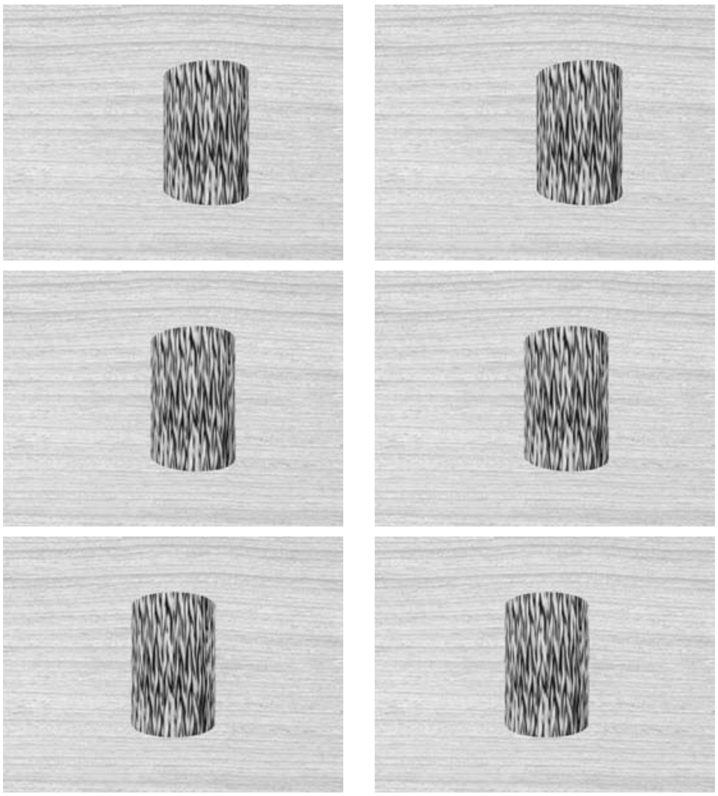


Figura 1.15: Distintos pares estéreos de la secuencia del *Cilindro*. En la primera columna, las imágenes de la cámara izquierda. En la segunda columna, las imágenes de la cámara derecha.

Secuencias Sintéticas

En nuestros experimentos se han utilizado seis secuencias sintéticas: (1) *Cilindro*, (2) *Cilindro y Esfera*, (3) *Venus*, (4) *Map*, (5) *Sawtooth* y (6) *Corridor*. Las dos primeras han sido creadas por el grupo AMI, las tres siguientes forman parte de la base de datos 2001 de Middlebury (`http://vision.middlebury.edu/stereo/data/scenes2001/`) y, por último, el par estéreo del pasillo que ha sido creado por el grupo de visión por ordenador de la Universidad de Bonn.

Cilindro

La secuencia conocida como *Cilindro* consta de trece pares de imágenes de tamaño 800×600 píxeles. En ella, se observa cómo se desplaza de izquierda a derecha un objeto cilíndrico. Este objeto tiene una textura compuesta por unas hebras de lana entrelazadas. En el fondo de la imagen se ha planchado una textura de tono claro para que cause contraste con el cilindro. La velocidad del cilindro es de unos ocho píxeles por imagen. En la figura 1.15 se muestra distintos pares estéreo de la secuencia.

En el mapa de disparidad (fig. 1.16) se observa que el único movimiento está en el cilindro, el fondo es estático. Dado que los píxeles del cilindro no están a la misma

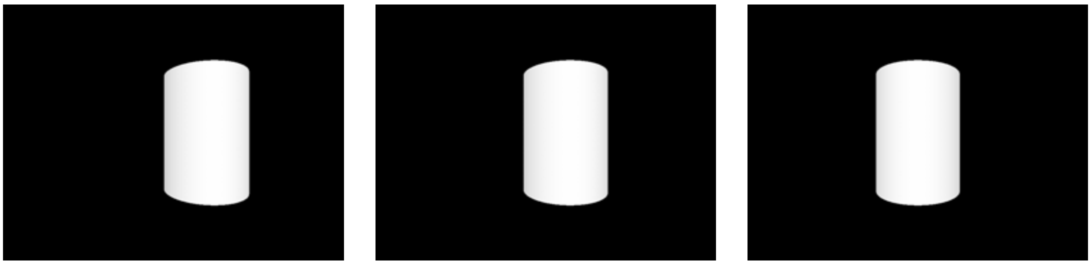


Figura 1.16: De izquierda a derecha los mapa de disparidad de los pares estéreo mostrados en la figura 1.15. Los píxeles que registren un mayor desplazamiento tendrán un tono claro y los que se muevan a menor velocidad un tono más oscuro (grisáceo). En el fondo de la escena no se registra movimiento alguno por lo que el tono de los píxeles es negro.

profundidad el desplazamiento varía en función de la distancia a la cámara (menor distancia mayor desplazamiento y viceversa).

Cilindro y Esfera

La secuencia del *Cilindro y Esfera* se trata de una versión algo más compleja que la del *Cilindro*. En ella, se observan dos objetos en movimiento a distinta velocidad: un cilindro y una esfera (figura 1.17). El desplazamiento se produce de derecha a izquierda. La superficie de los objetos está recubierta por una serie de texturas. En primer plano encontramos el cilindro que se desplaza a una velocidad de once píxeles por imagen. Justo detrás del cilindro se observa una esfera. El movimiento de este objeto es de alrededor de seis píxeles. Cerca del fondo de la escena se aprecia un panel estático con una textura clara que ocupa casi la totalidad de la imagen.

En la figura 1.18 se muestra los mapas de disparidad de la secuencia estéreo. En esas imágenes se aprecia, en función de la distancia a la cámara, los tres objetos con distinta tonalidad.

Venus

La base de datos de Middlebury incluye una serie de secuencias de imágenes y pares estéreo tanto reales como sintéticas (`http://vision.middlebury.edu/stereo/data/`). *Venus* es un par estéreo sintético incluido en dicha base de datos. En la figura 1.19 podemos ver las imágenes de este par. En ellas se observa un lienzo al fondo y dos texturas situadas en primer plano: una página de un periódico de deportes y otra de un póster. En la figura 1.20, se aprecia el mapa de disparidad de *Venus*. Como es de esperar el desplazamiento de los píxeles de los objetos más próximos a las cámaras es mayor.

Map

El par estéreo conocido como *Map* también forma parte de la base de datos de Middlebury. En este par se observa un objeto rectangular situado en primer plano sobre un fondo. El objeto tiene una textura oscura mientras que el fondo está compuesto por un

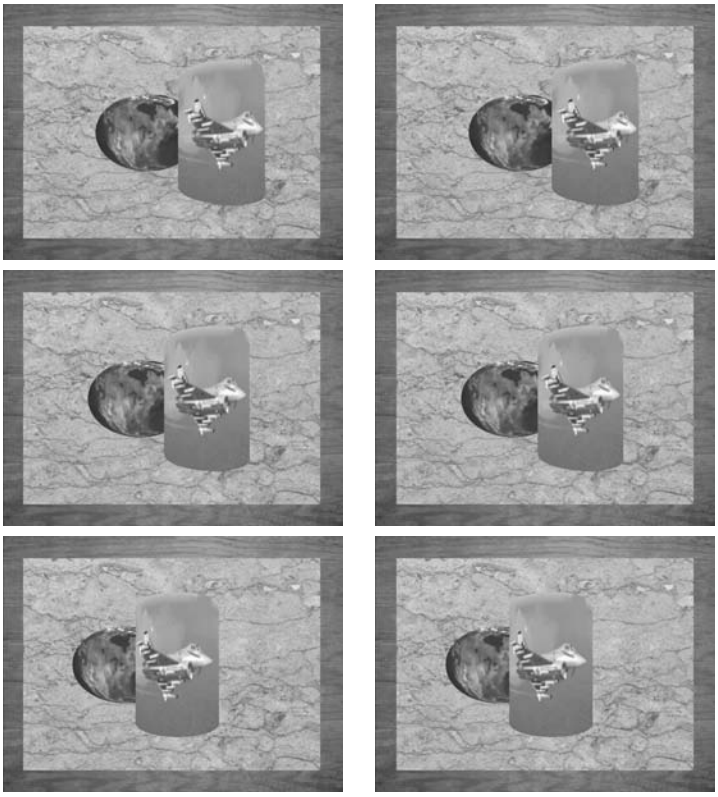


Figura 1.17: Distintos pares estéreos de la secuencia del *Cilindro* y la *Esfera*. En la primera columna, las imágenes de la cámara izquierda. En la segunda columna, las imágenes de la cámara derecha.

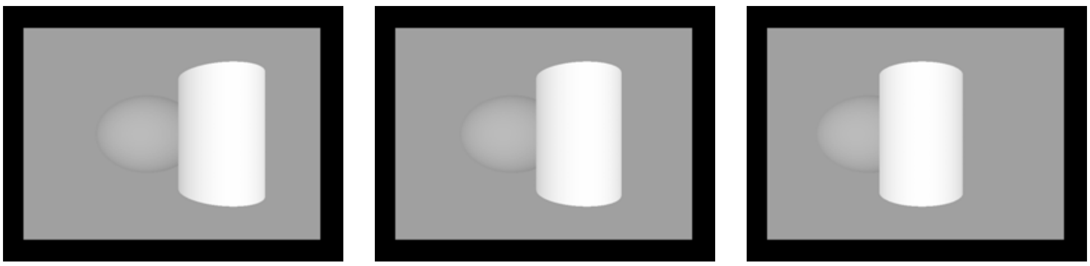


Figura 1.18: De izquierda a derecha los mapa de disparidad de los pares estéreo mostrados en la figura 1.17. Los píxeles que registren un mayor desplazamiento tendrán un tono claro y los que se muevan a menor velocidad un tono más oscuro (grisáceo). En los mapas de disparidad se aprecia en distinta tonalidad los tres objetos en movimiento: el cilindro (el más próximo a la cámara), la esfera (en una posición intermedia) y el panel (situado detrás de los dos objetos anteriores). La parte del fondo de la imagen que no queda oculta por el panel no se aprecia ningún desplazamiento por lo que el tono de los píxeles es negro.

Figura 1.19: Par estéreo de Venus.

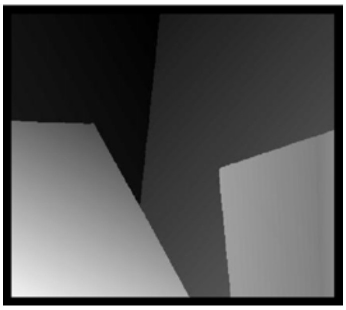


Figura 1.20: El mapa de disparidad de *Venus*.

patrón que se repite pero con distintas orientaciones. En las figuras 1.21 y 1.22 tenemos el par estéreo de *Map* y su mapa de disparidad, respectivamente.

Sawtooth

El par estéreo conocido como *Sawtooth* está integrado en Middlebury. Dos lienzos (uno en tonos claros y otro oscuro) conforman el fondo de la escena. En primer plano, tenemos una textura grisácea en cuya parte superior se asemeja a dientes de sierra, de ahí el nombre del par estéreo. En las figuras 1.23 y 1.24 podemos observar el par estéreo de *Sawtooth* y el mapa de disparidad asociado.

Secuencia del Pasillo

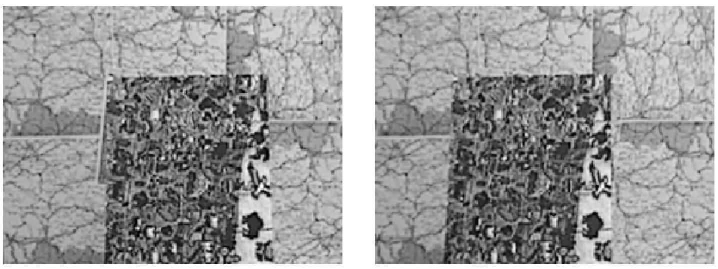


Figura 1.21: Par estéreo de *Map*.

Figura 1.22: El mapa de disparidad de *Map*.

Figura 1.23: Par estéreo de *Sawtooth*.

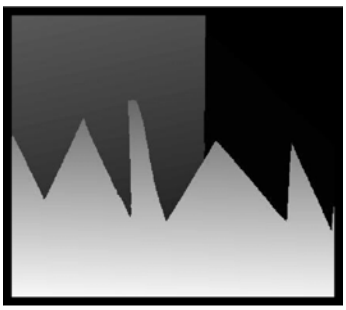


Figura 1.24: El mapa de disparidad de *Sawtooth*.

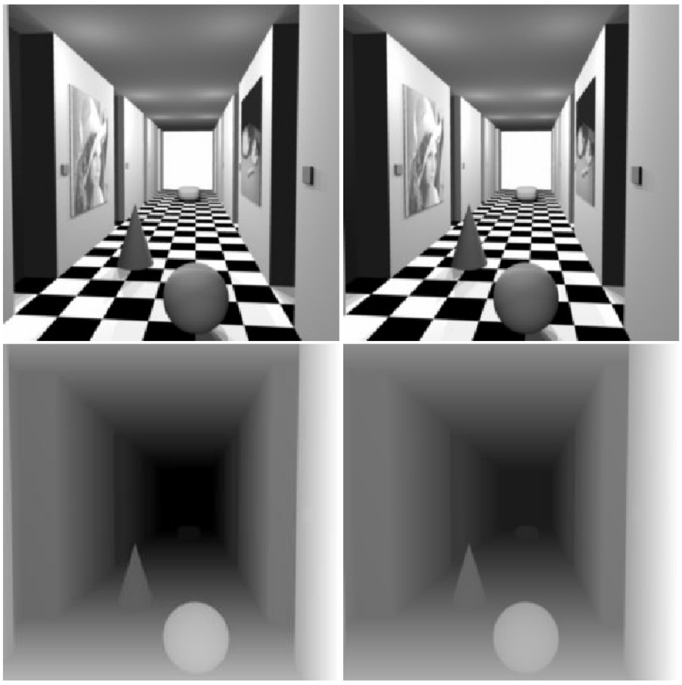


Figura 1.25: En la parte superior, el par estéreo *Corridor*. En la parte inferior, los mapas de disparidad en ambos sentidos.

El par estéreo conocido como *Corridor* ha sido generado de forma sintética a través de un programa de ray tracing. Este par puede descargarse en `http://www-student.informatik.uni-bonn.de/~gerdes/MRTStereo/`). En esta escena nos encontramos un pasillo en donde hay una esfera, un cono y un objeto al fondo. Todos situados a profundidades distintas con respecto a las cámaras.

En la figura 1.25 se muestra el par estéreo y los mapas de disparidad en ambos sentidos. En los experimentos se han utilizado tres variantes de este par con ruido de distinta intensidad: (i) sin ruido, (ii) con un ruido de varianza 10 y (iii) de varianza 100. En la sección de resultados se mostrarán las imágenes de los pares con ruido.

Secuencias Reales

Tsukuba

El par estéreo conocido como *Tsukuba* también pertenece a la base de datos de Middlebury. Al igual que las secuencias sintéticas *Tsukuba* fue tomada con cámaras situadas en posición frontoparalela. Respecto a la posición de la cámara se observa una serie de objetos: en primer plano una lámpara naranja, un busto blanco sobre una cajón de cartón ligeramente más retrasado y, al fondo, una videocámara sobre un trípode y unas estanterías con libros. A pesar de tratarse de una secuencia real se dispone de una mapa de disparidad que representa en precisión entera los desplazamientos de los principales objetos de la escena. En las figuras 1.26 y 1.27 se muestran el par estéreo y el mapa de disparidad, respectivamente.

Figura 1.26: Par estéreo de *Tsukuba*.

Figura 1.27: El mapa de disparidad del par estéreo *Tsukuba*.

Secuencia del Fútbol

La secuencia de fútbol corresponde a un partido de fútbol celebrado en el MiniEstadi del F.C. Barcelona. Ha sido suministrada por la empresa *Mediapro* y, su uso se ha restringido a algunos experimentos realizados en el contexto de esta tesis. En ella se observa desde dos cámaras situadas en la grada, un área del campo y parte de la grada (figura 1.28). El juego se desarrolla en la parte del campo que no es visible. En las imágenes se puede ver al portero caminando desde el punto de penalti hacia el borde del área. Dada las dimensiones de las imágenes (1900×1080 píxeles) se ha seleccionado una región más pequeña, de tamaño 501×171 píxeles, donde se percibe el único movimiento de interés de la secuencia, el caminar del portero en el área.

Desgraciadamente la posición y orientación de las cámaras no es frontoparalela. Se ha realizado un proceso de calibración y rectificación de las imágenes de ambas cámaras (figura 1.29). Al tratarse de una secuencia real no disponemos de mapa de disparidad real pero dado que el movimiento del portero es puramente traslacional, es muy fácil verificar visualmente el correcto funcionamiento del método en esta situación.

1.5. Medidas de Error

Para evaluar la calidad de los métodos descritos en esta tesis es necesario introducir una serie de métricas que nos permitan cuantificar el error entre la solución exacta y la estimada por un determinado método. En [Barron94] se proponen dos métricas conocidas como *Average Euclidean Error* (AEE, ec. (1.14)) y *Average Angular Error* (AAE, ec.

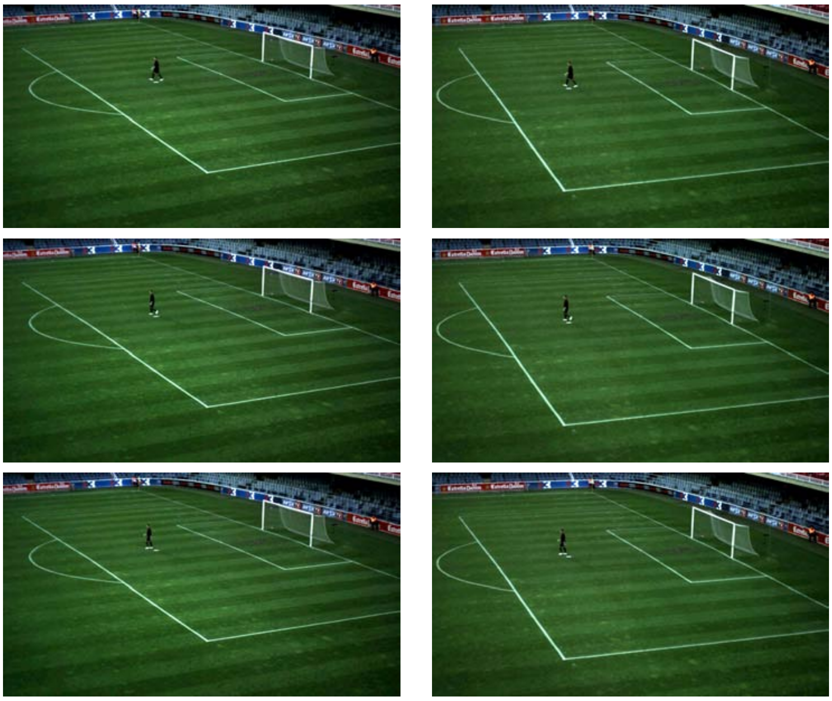


Figura 1.28: Secuencia del MiniEstadi. En la primera columna, los frames 0, 5 y 9 tomados por la cámara izquierda. En la segunda columna, los mismos frames pero captados por la cámara derecha.

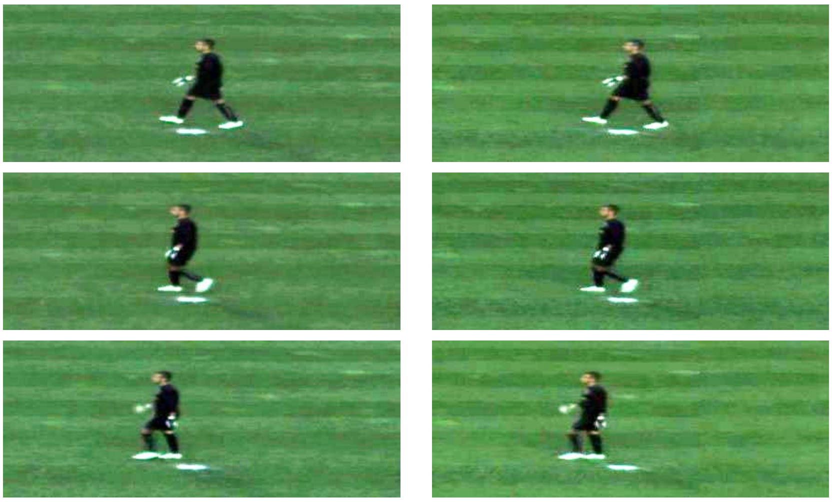


Figura 1.29: Imágenes rectificadas de las cámaras izquierda y derecha correspondientes a los frames mostrados en la figura 1.28. Las imágenes han sido modificadas para reducir el desplazamiento del portero a unos pocos píxeles. Simplemente se ha recortado un trozo del césped de la imagen correspondiente a la parte izquierda de la cámara 4 y pegado en la parte derecha.

(1.15)) que han sido ampliamente utilizadas en la literatura.

$$AEE = \frac{1}{N}\sum_{i=1}^{N} \left|\mathbf{h}_i - \mathbf{h}_{ref\ i}\right| , \tag{1.14}$$

$$AAE = \frac{180}{N\pi}\sum_{i=1}^{N} \arccos\left(\frac{\mathbf{h}_i\mathbf{h}_{ref\ i} + 1}{\sqrt{\|h_i\|^2 + 1}\sqrt{\|h_{ref\ i}\|^2 + 1}}\right) , \tag{1.15}$$

donde N es el número de píxeles de la imagen, $\mathbf{h}_i$ y $\mathbf{h}_{ref\ i}$ son los valores del campo de desplazamiento en el píxel i de la imagen estimada y la de referencia, respectivamente.

Recientemente se ha publicado un trabajo [Baker07] en el que se propone una actualización de las secuencias de referencia así como de las métricas a utilizar. Se sugiere el uso de estas cuatro métricas:

- *Angular error*. Se trata de la misma métrica descrita en la ecuación (1.15).
- *End-point error*. Se trata del error euclídeo, ec. (1.14).
- *Interpolation error* y *Normalized interpolation error*. Son métricas que cuantifican la precisión de la interpolación realizada a flujos no densos. Nuestros métodos generan soluciones densas por lo que no se emplea este tipo de métricas.

Capítulo 2

Estimación del Flujo Óptico en Secuencias de Imágenes

2.1. Introducción

Normalmente el movimiento de los objetos no se refleja en sólo dos frames. El desplazamiento se percibe a lo largo de una secuencia de imágenes. Por lo tanto, es natural la utilización de algo más de dos frames para estimar el movimiento registrado en la escena. Además, las imágenes se suelen ver afectadas por una serie de perturbaciones que no son fáciles de identificar cuando se dispone de dos frames. Estas perturbaciones no influyen por igual y de forma constante a los mismos píxeles de la secuencia, lo que posibilita su reconocimiento. El reto consiste en averiguar cuándo el valor de un píxel se trata de una anomalía producto de alguna perturbación o se trata una característica propia de la imagen. Cuando se disponen de varias vistas de la escena esa *anomalía* es más sencilla de identificar ya que no afecta de la misma forma a todas las vistas.

Los primeros métodos variacionales para la estimación del flujo óptico utilizaban únicamente dos imágenes. A medida que fueron evolucionando las técnicas quedó patente la necesidad de utilizar información temporal con el objetivo de incrementar la precisión de las estimaciones. Pese al importante avance que han supuesto todas estas técnicas todavía quedan algunas lagunas por resolver. Los métodos variacionales consiguen soluciones muy precisas y ofrecen una serie de ventajas respecto a otro tipo de técnicas. Entre las distintas ventajas podemos destacar: (i) todas las restricciones se plasman en el modelo de energía y, (ii) la minimización de esta energía se apoya sobre una sólida base matemática; por lo que se asegura la existencia de una solución si se cumple una serie de condiciones. Sin embargo, a la hora de implementar los modelos de energía existen distintos tipos de discretizaciones. Empíricamente se ha demostrado que se producen diferencias considerables entre las estimaciones obtenidas por el mismo método cuando se emplean distintas discretizaciones. Esto puede provocar que la idea reflejada en el modelo de energía no se vea potenciada por el tipo de discretización utilizada. En los trabajos desarrollados en esta tesis nos hemos centrado en el estudio de los modelos de energía sin entrar en demasiado detalle en los tipos de discretizaciones. Se ha recurrido a discretizaciones estándar.

Como parte principal de esta tesis hemos querido dedicar un capítulo completo al problema de la estimación del flujo óptico. Siguiendo la tendencia de los últimos años todos los métodos propuestos en este capítulo incluyen información procedente de varias imágenes y manejo de largos desplazamientos. Antes de entrar en detalle en cada uno de ellos se comentarán las contribuciones hechas a la literatura y se mostrará una generalización de la estructura del modelo de energía que es compartida por los tres métodos descritos en este capítulo.

2.1.1. Contribuciones de este Capítulo

Las contribuciones de los métodos descritos en este capítulo son las siguientes:

- **Método Variacional Multicanal: Extensión del Modelo de Alvarez/Weickert/Sánchez (2000) [Alvarez00]:**

  Basándonos en el funcional de energía propuesto por [Alvarez00] se ha realizado una serie de modificaciones para la inclusión de la información multicanal. Este método se ha utilizado para la detección de masas nubes en secuencias satélites multiespectrales.

  Los cambios introducidos en el modelo de energía son los siguientes. El término de ligadura se ha definido como una suma ponderada de la suposición lambertiana de cada uno de los canales y, el término de suavizado, el operador de Nagel-Enkelmann modificado cuyo gradiente aglutina información de todos los canales. Se han definido diversas estrategias para la construcción de dicho gradiente.

  Con las modificaciones realizadas se ha constatado una mejora en las estimaciones respecto al modelo original.

- **Método Variacional que incorpora un Término de Regularización exclusivamente Temporal no Continuo:**

  En este trabajo se ha desarrollado un método variacional espaciotemporal que desacopla la regularización temporal de la espacial. El desacoplamiento del término de suavizado en dos, uno espacial y otro temporal, pretende evitar algunas de las limitaciones presentes en trabajos con una regularización temporal continua basada en derivadas parciales temporales, como en [Papenberg06]. Nuestro método realiza convenientemente la regularización temporal teniendo en cuenta la presencia de desplazamientos largos sin afectar a la regularización espacial.

- **Método Variacional basado en el Análisis Espectral del Tensor de Movimiento:**

  En este trabajo se ha desarrollado un método variacional espaciotemporal que intercambia información complementaria entre el término de ligadura y el de suavizado. El proceso de difusión sólo se aplica en las direcciones ortogonales al flujo dominante. La combinación de varias técnicas existentes posibilita la creación de un *framework* capaz de representar cualquier modelo de una forma compacta, sencilla, fácilmente extensible y adaptable.

  El *tensor de movimiento* permite representar cualquier invarianza en una notación compacta. La descomposición espectral de este tensor asegura la fácil identificación de las direcciones dominantes del flujo al mismo tiempo que define una notación independiente de las invarianzas definidas en el modelo de energía.

La inclusión de términos no-cuadráticos no sólo aporta robustez a las estimaciones frente al ruido sino que la combinación de las funciones de robustificación junto a la descomposición espectral del tensor de movimiento permite la construcción de cuatro prototipos capaces de mejorar las estimaciones de métodos de características similares.

Todas las aportaciones hechas en este capítulo hacen uso intensivo de la información procedente de múltiples imágenes, ya sean del mismo o de distintos canales. Por ello, antes de entrar en detalle con las distintas aportaciones conviene definir un modelo de energía general que aglutine todas las ideas multicanal descritas en cada uno de los métodos.

2.2. Generalización de los Modelos de Energía

Los métodos variacionales se basan en la idea de que la solución se obtiene mediante la minimización de un funcional de energía. En esta sección se describe un modelo de energía que recoge prácticamente todas las variantes presentadas en la literatura. A partir de la *ecuación de restricción de flujo óptico* vamos generalizando dicho modelo gracias a la incorporación de una serie de técnicas ampliamente utilizadas en la literatura. Para facilitar la comprensión al lector comenzaremos con la obtención de un modelo de energía en el dominio continuo que luego extenderemos al dominio discontinuo.

2.2.1. Modelos de Energía Continuos

Los modelos de energía empleados en el problema de la estimación del flujo óptico normalmente se componen de dos términos: (1) el de *ligadura* y (2) el de *regularización* (o suavizado). Como ya se comentó en la sección 1.1 el término de ligadura asume que cierta propiedad en la imagen no varía a lo largo del tiempo mientras que el de suavizado impone cierta restricción de suavidad en el flujo.

La suposición más simple expresada por el término de ligadura se puede representar tal que

$$f(\mathbf{x}) - f(\mathbf{x} + \mathbf{h}(\mathbf{x})) \quad = \quad 0, \tag{2.1}$$

donde f es una propiedad en la imagen, $\mathbf{x} := (x, y, t)$ las coordenadas en la secuencia de imágenes y $\mathbf{h}(\mathbf{x}) := (u(\mathbf{x}), v(\mathbf{x}), 1)^\top$ es una función que representa el desplazamiento de los píxeles en la secuencia de imágenes.

En la ecuación (2.1) se aprecia una no linealidad que se suele sortear con del desarrollo de Taylor y la eliminación de los términos de orden superior

$$\begin{aligned} f(\mathbf{x} + \mathbf{h}(\mathbf{x})) = f(x + \delta x, y + \delta y, t + \delta t) \quad &= \quad f(x, y, t) + \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y + \frac{\partial f}{\partial t}\delta t \\ &\quad\; + O(\|\mathbf{h}(\mathbf{x})^n\|), \end{aligned} \tag{2.2}$$

Sustituyendo la ecuación (2.2) en la expresión (2.1) obtendríamos la *ecuación de restricción del flujo óptico* (ec. (2.3)). Esta expresión sólo es válida cuando el desplazamiento de los objetos es suave, o dicho de otra forma, las derivadas de la expresión son computables.

$$f(\mathbf{x}) - f(\mathbf{x} + \mathbf{h}(\mathbf{x})) \approx \frac{\partial f}{\partial x}\frac{\delta x}{\delta t} + \frac{\partial f}{\partial y}\frac{\delta y}{\delta t} + \frac{\partial f}{\partial t}\frac{\delta t}{\delta t} = 0,$$

que también se suele representar como

$$f_x(\mathbf{x})u(\mathbf{x}) + f_y(\mathbf{x})v(\mathbf{x}) + f_t(\mathbf{x}) = 0, \tag{2.3}$$

donde $u(\mathbf{x}) := \dfrac{\delta x}{\delta t}$ $v(\mathbf{x}) := \dfrac{\delta y}{\delta t}$ y f_x, f_y, f_t indican derivadas parciales.

La solución de la expresión anterior se puede estimar mediante mínimos cuadrados

$$(f_x u + f_y v + f_t)^2 = (\mathbf{h}(\mathbf{x})\nabla f)^2 = 0,$$

Una forma de extender la idea expresada en la ecuación anterior consistiría en la inclusión de varias invarianzas en el término de ligadura con el objetivo de detectar ciertos movimientos que sólo son perceptibles para algunas invarianzas. De esta forma, aunque se viole la suposición de una invarianza el desplazamiento puede ser identificado por otra. En (2.4) expresamos el término de ligadura como un conglomerado ponderado de varias invarianzas

$$\sum_{c=1}^{N_c} \gamma_c \left(f_x^c(\mathbf{x})u(\mathbf{x}) + f_y^c(\mathbf{x})v(\mathbf{x}) + f_t^c(\mathbf{x}) \right)^2 = \sum_{c=1}^{N_c} \gamma_c \left(\mathbf{h}(\mathbf{x}) \nabla f_c(\mathbf{x}) \right)^2 , \tag{2.4}$$

N_c el número total de invarianzas definidas en el término de ligadura y c el índice que especifica cada una de ellas. γ_c es un peso que indica la importancia de cada invarianza.

En principio, hemos supuesto la forma cuadrática como la única forma de minimizar el término de ligadura pero existen otras, como la norma L1. Por ello, para generalizar aún más la ecuación anterior (2.4) la expresaremos como

$$\int_\Omega \sum_{c=1}^{N_c} \gamma_c \, \mathcal{D}_c \left(\nabla f_c(\mathbf{x}), \mathbf{h}(\mathbf{x}) \right) , \tag{2.5}$$

donde $\mathcal{D}(.)$ podría variar dependiendo de la naturaleza de las imágenes y Ω el dominio de la imagen. Por ejemplo, las funciones de robustificación (propuestas por algunos autores [Black92, Black96a, Mémin98a]) serían una alternativa a la forma cuadrática para atenuar el efecto de los *outliers*.

Normalmente, los funcionales de energía están compuestos por un segundo término denominado de *regularización* o *suavizado*. Los primeros términos de regularización utilizaban únicamente información espacial. Unos de los primeros fue el de Horn-Schunck [Horn81]. Supone que el campo de desplazamiento es suave y este término suaviza en función del módulo del gradiente del flujo al cuadrado, ec. (2.6).

$$\|\nabla_2 \mathbf{h}\|^2 = \|\nabla_2 u\|^2 + \|\nabla_2 v\|^2 . \tag{2.6}$$

A partir de este término de regularización surgieron otros que se pueden representar mediante la ecuación (2.7)

$$\int_\Omega \mathcal{R} \left(\nabla I(\mathbf{x}), \nabla \mathbf{h}(\mathbf{x}) \right) . \tag{2.7}$$

Todavía es posible extender la ecuación anterior considerando, al igual que se hizo con el término de ligadura, ec. (2.5), la posibilidad de usar otras propiedades de la imagen distintas a la intensidad de los píxeles, ec.(2.8).

$$\int_\Omega \sum_{c=1}^{N_c} \alpha_c \mathcal{R}_c \left(\nabla f_c(\mathbf{x}), \nabla \mathbf{h}(\mathbf{x}) \right) , \tag{2.8}$$

donde c se trata de un índice asociado a cada característica de la imagen, α_c un peso de cada término de suavizado y $\mathcal{R}(.)$ una función de robustificación.

La ecuación (2.9) resume para el caso continuo el funcional de energía genérico que estima el flujo óptico para una secuencia de imágenes teniendo en cuenta la información de los frames vecinos.

$$E(\mathbf{h}) = \int_\Omega \sum_{c=1}^{N_c} \gamma_c \mathcal{D}_c \left(\nabla f_c, \mathbf{h}\right) \, d\mathbf{x} + \int_\Omega \sum_{c=1}^{N_c} \alpha_c \mathcal{R}_c \left(\nabla f_c, \nabla \mathbf{h}\right) \, d\mathbf{x}. \tag{2.9}$$

2.2.2. Modelos de Energía No Continuos Secuencial

En este apartado vamos a tratar de definir un modelo de energía genérico que englobe todas las situaciones para el caso discontinuo en el que se tenga en cuenta la información aportada por los frames vecinos.

La *ecuación de restricción del flujo óptico* sólo es válida cuando el desplazamiento de los objetos es pequeño. Cuando los desplazamientos entre frames son grandes esta ecuación deja de ser válida. Por lo tanto, es necesario formular una nueva expresión que sea válida para cualquier situación, tanto para desplazamientos cortos como largos. En el caso discontinuo podríamos aplicar la siguiente expresión

$$f_1(\tilde{\mathbf{x}}) - f_2(\tilde{\mathbf{x}} + \mathbf{h}_1(\tilde{\mathbf{x}})) = 0, \tag{2.10}$$

donde $\tilde{\mathbf{x}} := (x, y)$, $\mathbf{h}_t(\tilde{\mathbf{x}}) := \mathbf{h}(\tilde{\mathbf{x}}, t) := (u(\tilde{\mathbf{x}}, t), v(\tilde{\mathbf{x}}, t), 1)^\top$ y $f_i := f(\tilde{\mathbf{x}}, i)$. En la ecuación (2.10) se asume que cierta propiedad de la imagen en dos píxeles en correspondencia en dos imágenes consecutivas es constante.

En algunos trabajos como [Papenberg06] se han incluido varias invarianzas en el término de ligadura. Al igual que se desarrolló en el caso continuo podemos expresar el término de ligadura como un conglomerado ponderado de varias invarianzas ahora en el plano discontinuo.

$$\sum_{c=1}^{N_c} \gamma_c \left(f_1^c(\tilde{\mathbf{x}}) - f_2^c(\tilde{\mathbf{x}} + \mathbf{h}_1(\tilde{\mathbf{x}})\right)^2.$$

Ahora debemos utilizar una secuencia de imágenes y el flujo óptico es calculado entre cada par de imágenes de la secuencia. En vez de utilizar esta invarianza sólo entre dos frames podemos ampliar esta idea a toda la secuencia

$$\sum_{i=1}^{N-1} \sum_{c=1}^{N_c} \gamma_c \left(f_i^c(\tilde{\mathbf{x}}) - f_{i+1}^c(\tilde{\mathbf{x}} + \mathbf{h}_i(\tilde{\mathbf{x}})\right)^2. \tag{2.11}$$

Un término de ligadura genérico debería quedar expresado por la definición de múltiples invarianzas y la función de robustificación que las englobe, ec. (2.12).

$$\sum_{i=1}^{N-1} \sum_{c=1}^{N_c} \gamma_c \mathcal{D}_c \left(f_{i,c}(\tilde{\mathbf{x}}), f_{i+1,c}(\tilde{\mathbf{x}}), \mathbf{h}_i(\tilde{\mathbf{x}})\right). \tag{2.12}$$

En la literatura podemos observar cómo los distintos regularizadores propuestos tienen en común la utilización del gradiente del flujo para suavizar la estimación hecha por el término de ligadura. La ecuación (2.13) representa un término de regularización genérico que utiliza el gradiente del flujo y de la imagen de toda la secuencia

$$\sum_{i=1}^{N-1} \mathcal{R}\left(\nabla f_i(\tilde{\mathbf{x}}), \nabla \mathbf{h}_i(\tilde{\mathbf{x}})\right). \tag{2.13}$$

La ecuación anterior se puede ampliar considerando la posibilidad de usar cualquier propiedad de la imagen, ec. (2.14).

$$\sum_{i=1}^{N-1} \sum_{c=1}^{N_c} \alpha_c \mathcal{R}_c\left(\nabla f_i^c(\tilde{\mathbf{x}}), \nabla \mathbf{h}_i(\tilde{\mathbf{x}})\right). \tag{2.14}$$

La ecuación (2.15) resume el funcional de energía que estima el flujo óptico para una secuencia de imágenes teniendo en cuenta la información de los frames vecinos.

$$E(\mathbf{h}_1, \ldots, \mathbf{h}_{N-1}) = \int_\Omega \sum_{i=1}^{N-1} \sum_{c=1}^{N_c} \gamma_c \, \mathcal{D}_c\left(f_{i,c}, f_{i+1,c}, \mathbf{h}_i\right) \, d\tilde{\mathbf{x}} + \int_\Omega \sum_{i=1}^{N-1} \sum_{c=1}^{N_c} \alpha_c \, \mathcal{R}_c\left(\nabla f_i^c, \nabla \mathbf{h}_i\right) \, d\tilde{\mathbf{x}}. \tag{2.15}$$

2.2.3. Modelos de Energía No Continuos Aleatorio

En la definición de los modelos de energía no continuos se asume ímplicitamente que los flujos ópticos se deben estimar sólo a partir de la información de los frames vecinos. Esto es una limitación que se impone al modelo sólo con el fin de simplificar la notación y reducir la complejidad del funcional de energía. En este apartado vamos a describir la estructura de un modelo de energía no continuo que tiene en cuenta la información de cualquier frame de la secuencia de imágenes para la estimación del flujo óptico.

En un modelo no continuo secuencial el término de ligadura se definía como

$$\sum_{i=1}^{N-1} \sum_{c=1}^{N_c} \gamma_c \, \mathcal{D}_c\left(f_{i,c}(\tilde{\mathbf{x}}), f_{i+1,c}(\tilde{\mathbf{x}}), \mathbf{h}_i(\tilde{\mathbf{x}})\right). \tag{2.16}$$

La generalización del término de ligadura se puede llevar a un paso más allá, de forma, que las invarianzas no se tengan que cumplir únicamente entre frames consecutivos sino para cualquier frame de la secuencia. En el caso más simple el término de ligadura quedaría expresado como

$$\sum_{c=1}^{N_c} \gamma_c \left(f_i^c(\tilde{\mathbf{x}}) - f_j^c(\tilde{\mathbf{x}} + \mathbf{h}_{ij}(\tilde{\mathbf{x}}))\right)^2. \tag{2.17}$$

donde $\mathbf{h}_{ij}(\tilde{\mathbf{x}})$ es el flujo óptico calculado entre dos frames, i, j, cualesquiera de una secuencia de imágenes. Si ampliamos esta idea para cualquier par de imágenes de la secuencia obtendríamos la ecuación

$$\sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \sum_{c=1}^{N_c} \gamma_c \, \mathcal{D}_c\left(f_{i,c}(\tilde{\mathbf{x}}), f_{j,c}(\tilde{\mathbf{x}}), \mathbf{h}_{ij}(\tilde{\mathbf{x}})\right). \tag{2.18}$$

Como es de esperar el incremento de los grados de libertad en el término de ligadura acarrea un aumento de las incógnitas a resolver en el sistema de ecuaciones derivado a partir del modelo de energía.

Con esta generalización hemos querido poner de manifiesto las múltiples combinaciones existentes en la definición del término de ligadura en los funcionales de energía. Aunque a efectos prácticos, por simplicidad, se recurrirá a versiones más simples de (2.18), como puede ser (2.11). Si nos fijamos en los términos de ligadura de los métodos propuestos en la literatura son casos particulares de la ecuación (2.18).

En un modelo no continuo secuencial el término de suavizado se definía como

$$\sum_{i=1}^{N-1}\sum_{c=1}^{N_c}\alpha_s\mathcal{R}_c\left(\nabla f_i^c(\tilde{\mathbf{x}}),\nabla\mathbf{h}_i(\tilde{\mathbf{x}})\right).$$

Si ampliamos la ecuación anterior considerando la posibilidad de regularizar el flujo con información de frames no consecutivos obtendríamos

$$\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sum_{c=1}^{N_c}\alpha_c\mathcal{R}_c\left(\nabla f_i^c(\tilde{\mathbf{x}}),\nabla\mathbf{h}_{ij}(\tilde{\mathbf{x}})\right). \tag{2.19}$$

A partir de las generalizaciones de los términos de ligadura, ec. (2.18), y suavizado, ec. (2.19), se ha definido un modelo de energía que a diferencia de (2.15) utiliza la información procedente de frames cualesquiera.

$$\begin{aligned} E(\mathbf{h}_{12},\ldots,\mathbf{h}_{N(N-1)}) &= \int_\Omega\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sum_{c=1}^{N_c}\gamma_c\,\mathcal{D}_c\left(f_{i,c},f_{j,c},\mathbf{h}_{ij}\right)\;d\tilde{\mathbf{x}} \\ &\quad+\int_\Omega\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\sum_{c=1}^{N_c}\alpha_c\,\mathcal{R}_c\left(\nabla f_i^c,\nabla\mathbf{h}_{ij}\right)\;d\tilde{\mathbf{x}}. \end{aligned} \tag{2.20}$$

2.3. Análisis de Imágenes Satélite Multicanal usando Métodos Variacionales

El análisis de las imágenes satélites se ha convertido en un campo de estudio muy activo en los últimos años. Los avances en la tecnología de los sensores ha hecho posible la obtención de imágenes de mayor resolución y en un mayor rango de frecuencias. Esto ha impulsado el campo de la teledetección y de la meteorología. La estimación del movimiento de las nubes a partir de las imágenes satélites tiene mucha aplicación en la climatología y meteorología ([Hasler90]).

En la literatura se han propuesto distintas clases de técnicas para la estimación del movimiento de las nubes, como es *local cross-correlation* ([Leese71, Phillips72, Schmetz93]) y *cross-correlation* combinado con *relaxation labeling* ([Wu95, Evans06]), estimación de movimiento usando imágenes estéreo ([Young90, Kambhamettu95]), redes neuronales ([Côté95]) técnicas *block-matching* ([Brad02]), *ajuste local* ([LZ01]) o métodos variacionales ([Corpetti02]). A pesar de la variedad de las técnicas utilizadas predomina el uso de la correlación. Tiene la ventaja de ser robusta frente a cambios de intensidad global, pero el inconveniente del alto coste computacional que requieren las estimaciones y la no integración de una restricción de regularización global. A la hora de buscar las correspondencias de los píxeles se busca la similaridad entre ventanas o patrones alrededor de un píxel. Para obtener un resultado denso sería necesario la aplicación de esta operación en cada píxel de la imagen. Dado lo costoso computacionalmente hablando de este tipo de técnicas se selecciona un conjunto de puntos en la imagen en los que se aplica la correlación. A priori se desconoce cuáles son los puntos de interés en la imagen por lo que se seleccionada un conjunto de puntos equidistantes entre ellos alineados representando una rejilla (ver figura 2.1). Para obtener un mapa de desplazamiento denso es necesario aplicar alguna técnica de interpolación a partir de los valores obtenidos en la rejilla. Por un lado, la densidad de la rejilla debe ser lo suficientemente pequeña para que todos los desplazamientos sean detectados. Por otro lado, esta densidad debe ser lo suficientemente grande para que la solución sea obtenida en el menor tiempo posible. Sería deseable que este tipo de técnicas incluyese algún tipo de restricción que asegurase la coherencia de los resultados, es decir, que el desplazamiento sea una función suave. El *local cross-correlation* puede resultar útil cuando los movimientos son rígidos pero no funciona tan bien en el caso de desplazamientos rápidos no rígidos.

Los satélites captan la radiación luminosa en distintos rangos de frecuencias creando así un amplio abanico de imágenes. Tradicionalmente para el seguimiento de las nubes se han utilizado las imágenes del canal infrarrojo (IR: 10.5-12.5 μm, [Leese71, Schmetz93]). El canal visible permite la detección y seguimiento de nubes de baja altura ([Ottenbacher97, LZ01]). Dada la riqueza de información multiespectral ofrecida por los satélites, sería lógico pensar que la combinación de toda esta información permitiría mejorar el seguimiento de las masas nubosas. Recientemente, se han propuesto algunos métodos que combinan datos multicanal, como es una técnica *cross-correlation* multicanal que usa el canal visible e infrarrojo ([Evans06]) o una estimación multicapa de las nubes tomando información de varios canales ([Héas06]).

Los métodos variacionales han tenido bastante éxito en el campo de la estimación

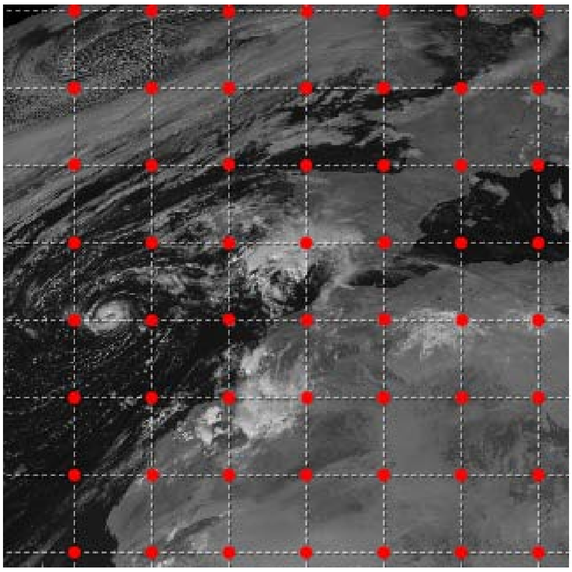


Figura 2.1: Un ejemplo de una imagen satélite en la que se ha resaltado en rojo los puntos de la rejilla en los que se aplica la correlación.

del flujo óptico. Estos métodos imponen una restricción de suavizado global creando resultados coherentes. Las soluciones son densas por lo que no es necesario ningún tipo de interpolación una vez hecho los cálculos. La combinación de varios canales para la estimación del flujo óptico no es nada novedoso en los métodos variacionales. En la literatura existe un gran número de métodos que utilizan secuencias de color (tres canales, rojo, verde y azul). Sin embargo, el método que presentamos en esta sección combina la información de los canales de un modo diferente a como se suele hacer en el caso de color. En secuencias de color la información está repartida por los tres canales por lo que es imprescindible la combinación de todos ellos para calcular los desplazamientos. En las secuencias multicanal de los satélites la información se concentra en cada canal, es posible estimar el movimiento de las estructuras nubosas en cada canal independiente. Pero hay determinadas estructuras que se detectan mejor en unos canales que en otros. Por ese motivo, la combinación de los canales satélites enriquece las estimaciones, de forma que cuando un canal no ofrece una buena estimación, ésta sería aportada por otro donde el movimiento se detectara con mayor nitidez.

En esta sección presentamos un método variacional multicanal aplicado al análisis de imágenes satélites multiespectrales. Este método combina la información de varios canales para mejorar la estimación del desplazamiento de las nubes. Se trata de una variante del modelo de Nagel-Enkelmann, [Nagel86]. Comenzaremos, en el apartado 2.3.1, presentando un método variacional monocanal [Alvarez00], al que le aplicaremos una serie de pequeñas modificaciones para que pueda utilizar información multicanal, apartado 2.3.2. Dado que la información de relevancia no se reparte equitativamente entre los distintos canales sería deseable establecer algunas estrategias para priorizar unos canales frente a otros. Por último, mostraremos en el apartado 2.3.5 los resultados cuantitativos y cualitativos obtenidos con las dos secuencias de satélites comentadas en el apartado 1.4.1. En estos experimentos se refleja la importancia y mejora que ofrece el método multicanal frente a monocanal.

2.3.1. Método Variacional Monocanal

La estimación del movimiento de las masas nubosas por la atmósfera se trata de un problema de cálculo del flujo óptico. Disponemos de unas imágenes tomadas por un satélite donde las estructuras nubosas son unos píxeles que se desplazan por las imágenes. Para estimar este movimiento hemos utilizado un método variacional. Los métodos variacionales más simples utilizan únicamente un sólo canal, por ejemplo una imagen en escala de grises. Otros más sofisticados combinan la información de varios canales.

En este apartado se presenta el modelo de energía utilizado por el método monocanal. Este modelo corresponde con el descrito en [Alvarez00]. Se trata de un método 2D compuesto por dos términos: uno de ligadura y otro de suavizado. El término de ligadura incluye la suposición lambertiana mientras que el de suavizado el operador de Nagel-Enkelmann [Nagel86] con algunas mejoras.

El funcional de energía a minimizar es

$$E(\mathbf{h}) = \int_{\Omega} (I_1(\tilde{\mathbf{x}}) - I_2(\tilde{\mathbf{x}} + \mathbf{h}(\tilde{\mathbf{x}})))^2 \, d\tilde{\mathbf{x}} + \alpha \int_{\Omega} tr(\nabla \mathbf{h}^t D \nabla \mathbf{h}) d\tilde{\mathbf{x}}, \tag{2.21}$$

donde $\tilde{\mathbf{x}}$ es un punto que pertenece al dominio Ω, α es una constante que pondera el término de suavizado, ∇ es el operador del gradiente y, D es una matriz de proyección regularizada en la dirección ortogonal a ∇I_1.

Operador de Nagel–Enkelmann

La matriz de difusión D se define como:

$$D(\nabla I_1) = \frac{1}{\|\nabla I_1\|^2 + 2\zeta^2} \left(\xi\xi^t + \zeta^2 Id \right), \tag{2.22}$$

donde $\xi = (\frac{\partial I_1}{\partial y}, -\frac{\partial I_1}{\partial x})^t$ es un vector ortogonal a ∇I_1, $\|.\|$ indica la norma, Id es la matriz identidad y, ζ es un coeficiente que determina el comportamiento isotrópico del suavizado e inhibe los bordes borrosos cuando la magnitud del gradiente es alta: $\|\nabla I_1\| \gg \zeta$.

Los parámetros de entrada de este método son C y $\lambda \in (0,1)$. A partir de estas variables se calculan dos pesos utilizados en el modelo de energía como son C y ζ donde

$$\alpha = \frac{C}{max\left(|(\nabla G_\sigma * I_1)(\overline{x})|^2\right)}, \tag{2.23}$$

$$\lambda = \int_0^{\zeta} H_{|\nabla G_\sigma * I_1|}(z) dz, \tag{2.24}$$

donde $G_\sigma * I_1$ representa la convolución de I_1 con una gaussiana de desviación estándar σ, $H_{|\nabla G_\sigma * I_1|}(z)$ representa el histograma normalizado de $|\nabla G_\sigma * I_1|$. λ se conoce como *fracción isotrópica*. Cuando $\lambda \to 0$, la difusión aplicada es anisotrópica mientras $\lambda \to 1$, lo hace isotrópico. Esta normalización de α y ζ permite a la energía ser invariante frente a cambios de intensidad del tipo $(I_1, I_2) \to (kI_1, kI_2)$.

2.3.2. Modelo de Energía

Los satélites disponen de gran variedad de sensores que captan imágenes en distintos rangos de frecuencias. Esta información resulta fundamental para el análisis de todos los fenómenos que ocurren en la atmósfera. Las nubes generan cambios en la concentración del vapor de agua, en la presión del aire y en la radiación térmica reflejada por la Tierra. Combinando esta información es posible obtener una estimación más robusta del movimiento de las nubes.

En esta sección presentamos un método variacional multicanal para estimar el flujo óptico. Este método se trata de una extensión del descrito en el apartado 2.3.1. Se ha modificado el modelo de energía original para dar cabida a la información multicanal manteniendo una formulación parecida a la original. A continuación, se describe el modelo de energía y se desarrolla el esquema numérico resultante de la minimización del funcional de energía.

Los datos de entrada de nuestro método son imágenes satélites procedente de distintos canales. Por cada canal disponemos de una secuencia de imágenes. El modelo de energía es una adaptación del descrito en [Alvarez00] por lo que nuestro método no incorpora información temporal. La estimación del flujo óptico se hace utilizando solamente dos frames de cada canal.

El modelo de energía multicanal es un caso particular de la ecuación (2.15). Se compone de dos términos: el de ligadura y el de suavizado. Dada la naturaleza de las imágenes y el término de regularización espacial que incorpora el modelo, la ecuación(2.15) se transforma en la (2.25).

$$E(\mathbf{h}) \;=\; \int_{\Omega} \sum_{c=1}^{N_c} \rho_c\, \mathcal{D}_c(I_1^c, I_2^c, \mathbf{h})\; d\tilde{\mathbf{x}} + \alpha \int_{\Omega} \sum_{c=1}^{N_c} \gamma_c\, \mathcal{R}(\nabla I_c, \nabla \mathbf{h})\; d\tilde{\mathbf{x}}. \tag{2.25}$$

donde N_c es el número de canales, $c \in [1, N_c]$ es un identificador de cada canal.

El término de ligadura utiliza la información de dos frames (I_1^c, I_2^c) de cada canal. Se trata de una suma ponderada de la suposición lambertiana aplicada a cada canal por separado. En vez de utilizar distintas invarianzas, el término de ligadura incluye la misma invarianza aplicada a distintas imágenes. Cuando $N_c = 1$, versión monocanal, nos encontramos ante la suposición lambertiana aplicada a un único canal.

En la versión monocanal, el operador Nagel-Enkelmann usa el gradiente de la imagen para determinar la cantidad de difusión. La versión multicanal mantiene esa idea combinando la información de los diferentes canales de acuerdo a una estrategia preestablecida. En este apartado se comenta las dos estrategias utilizadas para calcular el gradiente empleado en el término de regularización.

El modelo de energía del método multicanal es

$$E(\mathbf{h}) = \int_{\Omega} \sum_{c=1}^{N_c} \rho_c (I_1^c(\tilde{\mathbf{x}}) - I_2^c(\tilde{\mathbf{x}} + \mathbf{h}))^2 d\tilde{\mathbf{x}} + \alpha \int_{\Omega} tr(\nabla \mathbf{h}^t D \nabla \mathbf{h}) d\tilde{\mathbf{x}}, \tag{2.26}$$

donde I_1^c y I_2^c son la primera y segunda imágenes del canal c. El resto de la notación utilizada es la misma que la definida en el apartado 2.3.1.

En el método monocanal el tensor de difusión D se definía como:

$$D(\nabla I) = \frac{1}{\|\nabla I\|^2 + 2\zeta^2} \left(\xi\xi^t + \zeta^2 Id\right),$$

donde $\xi = (\frac{\partial \nabla I}{\partial y}, -\frac{\partial \nabla I}{\partial x})^t$ es un vector ortogonal a ∇I. Para introducir información multicanal en este tensor de difusión es necesario definir un vector $\bar{g}$ que juegue el mismo papel que ∇I.

Las dos estrategias utilizadas para estimar el gradiente multicanal $\bar{g}$ son:

- *Gradiente Máximo.* En cada píxel, $\bar{g}$ se calcula como el gradiente mayor en magnitud de los gradientes de todos los canales:

$$argmax \left\{\|\overrightarrow{v}\|, \overrightarrow{v} \in \{\nabla I^c, c \in [1, N_c]\}\right\}.$$

- *Gradiente Medio.* $\bar{g}$ se calcula como la dirección dominante de los gradientes de todos los canales. Esta orientación se calcula a partir del *tensor de estructuras*. Este *tensor de estructuras* se define como

$$\sum_{c=1}^{N_c} \rho_c \left(\nabla I_i^c\right) \left(\nabla I_i^c\right)^t,$$

$\bar{e}_{max}$ es el autovector normalizado asociado al autovalor máximo λ_{max} de la matriz anterior y $\bar{g}$ se puede definir como:

$$\bar{g} = \sqrt{\frac{\lambda_{max}}{\sum_{c=1}^{N_c} \rho_c}}\bar{e}_{max}.$$

De hecho, $\bar{g}$ es el mínimo si se cumple la restricción $\|\bar{g}\| = \sqrt{\lambda_{max}/\sum_{c=1}^{N_c} \rho_c}$, de la siguiente energía:

$$E(\bar{g}) = -\sum_{c=1}^{N_c} \rho_c \left(\bar{g}^t \cdot \nabla I_i^c\right)^2.$$

2.3.3. Minimización de la Energía

La solución de los métodos variacionales se obtiene a través de la minimización del modelo de energía. Para obtener dicha solución basta con derivar las ecuaciones de Euler–Lagrange a partir de dicha energía, que en nuestro caso se trata de la ecuación (2.26).

Las ecuaciones de Euler-Lagrange asociadas al funcional de energía (2.26) son:

$$\begin{aligned}
\alpha\, div(D\nabla u) + \sum_{c=1}^{N_c} \rho_c (I_1^c(\tilde{\mathbf{x}}) - I_2^c(\tilde{\mathbf{x}} + \mathbf{h})) \frac{\partial I_2^c}{\partial x}(\tilde{\mathbf{x}} + \mathbf{h}) &= 0 \\
\alpha\, div(D\nabla v) + \sum_{c=1}^{N_c} \rho_c (I_1^c(\tilde{\mathbf{x}}) - I_2^c(\tilde{\mathbf{x}} + \mathbf{h})) \frac{\partial I_2^c}{\partial y}(\tilde{\mathbf{x}} + \mathbf{h}) &= 0.
\end{aligned} \tag{2.27}$$

2.3.4. Esquema Numérico

Para discretizar el sistema de ecuaciones (2.27) se ha utilizado un esquema implícito de diferencias finitas. El esquema implícito ha demostrado tener estabilidad y converger más rápidamente que un esquema explícito.

La matriz D, se define en cada píxel i como:

$$D_i = \begin{pmatrix} a_i & b_i \\ b_i & c_i \end{pmatrix}.$$

La discretización del término de la divergencia en cada píxel i se realiza de la siguiente manera:

$$div(D_i \nabla u) = \begin{pmatrix} a_i \partial_x u + b_i \partial_y u \\ b_i \partial_x u + c_i \partial_y u \end{pmatrix} = \partial_x \left(a_i \partial_x u \right) + \partial_x \left(b_i \partial_y u \right) + \partial_y \left(b_i \partial_x u \right) + \partial_y \left(c_i \partial_y u \right),$$

N_i^* se define como los vecinos alrededor del píxel i. Utilizando un esquema de diferencias estándar la divergencia se puede expresar como:

$$div(D_i \nabla u_i) = \sum_{n \in N_i^*} (d_n u_n) + d_i u_i, \tag{2.28}$$

para los coeficientes adecuados d_n. Para el caso de $div(D_i \nabla v)$ se procede de la misma manera.

Los componentes del flujo óptico (u_i, v_i) se obtienen resolviendo el sistema utilizando el algoritmo Gauss-Seidel tal y como se ha descrito en [Alvarez00], donde k representa el número de la iteración. Los términos $I(\tilde{\mathbf{x}} + \mathbf{h}^{k+1})$ han sido linealizados mediante mediante una expansión de Taylor:

$$\begin{aligned} I_1^c(\tilde{\mathbf{x}}) - I_2^c(\tilde{\mathbf{x}} + \mathbf{h^{k+1}}) \quad \simeq \quad & I_1^c(\tilde{\mathbf{x}}) - I_2^c(\tilde{\mathbf{x}} + \mathbf{h^k}) \\ & - \frac{\partial I_2^c}{\partial x}(\tilde{\mathbf{x}} + \mathbf{h^k})(u^{k+1} - u^k) - \frac{\partial I_2^c}{\partial y}(\tilde{\mathbf{x}} + \mathbf{h^k})(v^{k+1} - v^k), \end{aligned}$$

donde $I_{2,i,x}^{c,k} = \dfrac{\partial I_2^c}{\partial x}(\tilde{\mathbf{x}}_\mathbf{i} + \mathbf{h^k})$ y $I_{2,i,y}^{c,k} = \dfrac{\partial I_2^c}{\partial y}(\tilde{\mathbf{x}}_\mathbf{i} + \mathbf{h^k})$.

Una vez comentado cómo se ha discretizado individualmente cada uno de los términos de las ecuaciones de Euler–Lagrange vamos a mostrar el esquema iterativo resultante que define los desplazamientos horizontales y verticales de cada píxel:

$$\begin{aligned} u_i^{k+1} \quad = \quad & \frac{u_i^k + dt\, \alpha \displaystyle\sum_{n \in N_i^*} \left(d_n u_n^k \right)}{1 + dt \left(\alpha\, d_i + \displaystyle\sum_{c=1}^{N_c} \rho_c \left(I_{2,i,x}^{c,k} \right)^2 \right)} \\ + \quad & \frac{dt \displaystyle\sum_{c=1}^{N_c} \rho_c \left(I_1^c(\tilde{\mathbf{x}}_\mathbf{i}) - I_2^c(\tilde{\mathbf{x}}_\mathbf{i} + \mathbf{h^k}) + u_i^k I_{2,i,x}^{c,k} - (v_i^{k+1} - v_i^k) I_{2,i,y}^{c,k} \right) I_{2,i,x}^{c,k}}{1 + dt \left(\alpha\, d_i + \displaystyle\sum_{c=1}^{N_c} \rho_c \left(I_{2,i,x}^{c,k} \right)^2 \right)}, \end{aligned} \tag{2.29}$$

$$
\begin{aligned}
v_i^{k+1} &= \frac{v_i^k + dt\,\alpha \displaystyle\sum_{n\in N_i^*}\left(d_n v_n^k\right)}{1+dt\left(\alpha\, d_i + \displaystyle\sum_{c=1}^{N_c}\rho_c\left(I_{2,i,y}^{c,k}\right)^2\right)} \\
&+ \frac{dt\displaystyle\sum_{c=1}^{N_c}\rho_c\left(I_1^c(\tilde{\mathbf{x}}_\mathbf{i}) - I_2^c(\tilde{\mathbf{x}}_\mathbf{i}+\mathbf{h}^\mathbf{k}) + v_i^k I_{2,i,y}^{c,k} - (u_i^{k+1}-u_i^k) I_{2,i,x}^{c,k}\right) I_{2,i,y}^{c,k}}{1+dt\left(\alpha\, d_i + \displaystyle\sum_{c=1}^{N_c}\rho_c\left(I_{2,i,y}^{c,k}\right)^2\right)}.
\end{aligned}
\tag{2.30}
$$

Para poder detectar los desplazamientos largos y asegurar que la solución del sistema no converge a un mínimo local las ecuaciones (2.29) y (2.30) se introducirán en un enfoque multipiramidal.

2.3.5. Resultados Experimentales

En este apartado vamos a mostrar los resultados experimentales obtenidos por el método variacional multicanal que ha sido presentado en 2.3.2. Para establecer el porcentaje de mejora que ofrece este nuevo método se comparará con su homólogo monocanal. En los experimentos se han utilizado dos secuencias satélites del Meteosat descritas en el apartado 1.4.1. Han sido suministradas por EUMESAT y captan la región del Atlántico Norte. La primera secuencia corresponde con el huracán Vince (producido el 8 de Octubre del 2005); y la segunda secuencia data del 5 de Junio del 2004. Para hacer un estudio más profundo se han generado dos nuevas secuencias a partir de los datos reales en los que se conoce el desplazamiento de los píxeles. De este modo, será posible hacer una comparación más exacta al disponer de datos cuantitativos.

Antes de presentar los resultados conviene comentar cómo se han generado las secuencias satélites pseudo-reales. En el proceso de creación se ha tenido especial cuidado en mantener la independencia respecto al método variacional, de forma, que este proceso no favorezca las estimaciones hechas por este método. Las secuencias satélites tienen un tamaño de 1024×1024 píxeles. El método multicanal sólo necesita dos frames por canal. Para reflejar de una forma más realista la evolución del movimiento de la escena se ha utilizado cuatro frames durante el proceso de creación. Los pasos seguidos se describen a continuación:

1. Extracción del movimiento de la secuencia real.

   Se ha calculado el flujo óptico de cuatro frames consecutivos de cada canal usando un método de correlación básico. Con un algoritmo de interpolación bilineal se ha obtenido un campo de desplazamiento denso. Una vez finalizado este paso disponemos de doce campos de desplazamiento (tres por cada canal).

2. Cálculo del desplazamiento de la nueva secuencia.

El desplazamiento de cada píxel de la imagen se calcula como la media de ese píxel en los doce campos obtenidos en el paso anterior. Esta nueva secuencia sigue conservando el realismo ya que el movimiento de los píxeles es la combinación de varios canales en distintos instantes de tiempo. Además conserva la independencia del método variacional.

3. Generación de la nueva secuencia pseudo-real.

   Tomando la primera imagen de la secuencia se creará una segunda mediante un warping utilizando el mapa de desplazamiento calculado en el paso 2. Esta nueva secuencia tiene una apariencia realista y, al mismo tiempo, permite disponer del mapa de desplazamiento exacto. Ahora será posible obtener resultados cuantitativos y establecer comparaciones entre los métodos.

La nueva secuencia ofrece una serie de ventajas respecto a la original

- el desplazamiento de los píxeles se conoce,
- el desplazamiento simulado no está asociado a ningún canal en particular ya que se trata de una media de todos los canales por lo que no se le otorga ninguna preferencia a ningún canal,
- el desplazamiento simulado ha sido estimado con una técnica de correlación por lo que no existe ningún tipo de relación con el método variacional.

La información captada por los distintos canales de los satélites es de gran importancia y se utiliza en un gran número de aplicaciones, como podemos ver en la tabla 1.3. Aunque es en el canal visible donde se registra la mayor parte del movimiento de las nubes, no cabe duda que hay cierta información, como puede ser las nubes de baja altura, cuya presencia no es tan nítida en el canal visible y sí en el de vapor de agua.

Dado que la aportación de cada canal no es equitativa se ha introducido en el modelo de energía, ec. (2.26), unos pesos ρ_c que determinan la importancia de cada canal. Para establecer un ranking de la contribución hecha por cada canal a la solución global se ha calculado el flujo óptico en cada uno de ellos. De esta forma, aquellos canales cuya estimación se aproxime más al flujo real deberán tener más peso respecto al resto.

En los experimentos se ha querido comparar la mejora que ofrece un método variacional multicanal (MC, *Multi-Channel*) frente a su versión monocanal (SF, *Simple Flow*). La principal diferencia entre ellos radica en los datos de entrada que utilizan para estimar el flujo óptico. En las tablas 2.1 y 2.2 podemos encontrar los resultados cuantitativos obtenidos de la comparación de ambos métodos. El método SF se aplica a cada canal por separado mientras que el MC se aplica sobre los cuatro canales conjuntamente. Las cuatro primeras estimaciones corresponden al SF aplicado a cada canal y, las dos siguientes, al método multicanal con las dos estrategias comentadas: *Gradiente Máximo* y *Gradiente Medio*.

En las figuras 2.3 y 2.7, se muestran las soluciones obtenidas con cada uno de los canales para cada secuencia. Dado que gran parte de la información se percibe en el canal visible se ha asignado mayores valores al peso de este canal. Los pesos utilizados en la secuencia de Vince han sido 0,5 para el *visible*, 0,3 para el *infrarrojo* y 0,1 para el resto de canal. En la secuencia del Atlántico Norte éstos han sido $[0{,}5, 0{,}1, 0{,}2, 0{,}2]$ =(VIS, VP1, VP2, IR). Los canales de vapor de agua no ofrecen mucha información y los desplazamientos detectados en estos canales no se asemejan a los detectados en el visible. Por eso, para que los desplazamientos sean coherentes no conviene ponderar en exceso estos canales. Sin embargo, el canal infrarrojo contribuye a una mejora en la estimación del canal visible.

En el modelo de energía, los parámetros α y ζ definidos como pesos en el término de suavizado y como parámetro del operador de Nagel–Enkelmann han sido calculados a partir de las ecuaciones (2.23) y (2.24). Estas ecuaciones dependen de C y λ que serán dos parámetros de entrada al método. En [Alvarez00] podemos encontrar más detalles acerca de cómo estos parámetros se ajustan automáticamente al rango de la imagen. En los experimentos, la mejor configuración de estos parámetros para ambas secuencias han sido $C = 6{,}67$ y $\lambda = 0{,}5$ para el método multicanal y $C = 10$ y $\lambda = 0{,}5$ para el monocanal. La estabilidad de C y λ es alta, de forma que pequeñas variaciones en sus valores no ofrecen cambios significativos en los errores obtenidos en las estimaciones.

Para facilitar la interpretación del campo de desplazamiento hemos mostrado los datos mediante flechas superpuestas sobre una imagen de la secuencia. Estas flechas están dispuestas a una distancia de 18 píxeles y en color verde o azul. Se han utilizado estos dos colores debido al contraste que provocan con el fondo de la imagen. En el caso que el fondo sea claro se han coloreado las flechas en azul y, en color verde, si el fondo es oscuro.

Para evaluar la calidad de los resultados se ha utilizado las dos métricas presentadas en 1.5, como son el *Average Euclidean Error* (AEE, ec. (1.14)) y el *Average Angular Error* (AAE, ec. (1.15)).

Secuencia 1. Atlántico Norte (8 Octubre 2005), Huracán Vince

En la figura 2.2 se muestran las imágenes de cuatro canales del satélite Meteosat (MSG) que captan el huracán Vince. Para analizar con más detalle los fenómenos más interesantes de esta secuencia hemos seleccionado una región de 642×559 píxeles (figura 2.4). En esta región podemos observar el huracán *Vince* y una gran masa nubosa sobre el noroeste de la costa africana. Los movimientos dominantes registrados son (1) *rotacional* debido a la influencia de Vince y (2) *traslacional* por parte de las nubes situadas sobre África. En la figura 2.3 se muestra los campos de desplazamiento de los cuatro canales calculados con el método SF.

En la figura 2.4 podemos observar el flujo óptico obtenido con el método multicanal usando los cuatro canales. En esta imagen se percibe cómo la influencia de Vince es bastante alta originando en las nubes próximas el típico movimiento rotacional característico de los huracanes. La fuerza de los vientos de Vince genera, como es lógico, fuertes vorticidades en el campo de desplazamiento. Este movimiento se puede ver perfectamente en el canal visible. El segundo fenómeno de interés se localiza en la costa africana. Esta masa nubosa incrementa su velocidad cuando se aproxima a la zona de influencia de Vince. Este movimiento se registra en los canales infrarrojo y visible.

Analizando los datos de la tabla 2.1 se desprende que la contribución de los cuatro canales tratados conjuntamente ofrece mejores estimaciones que si utilizamos un único

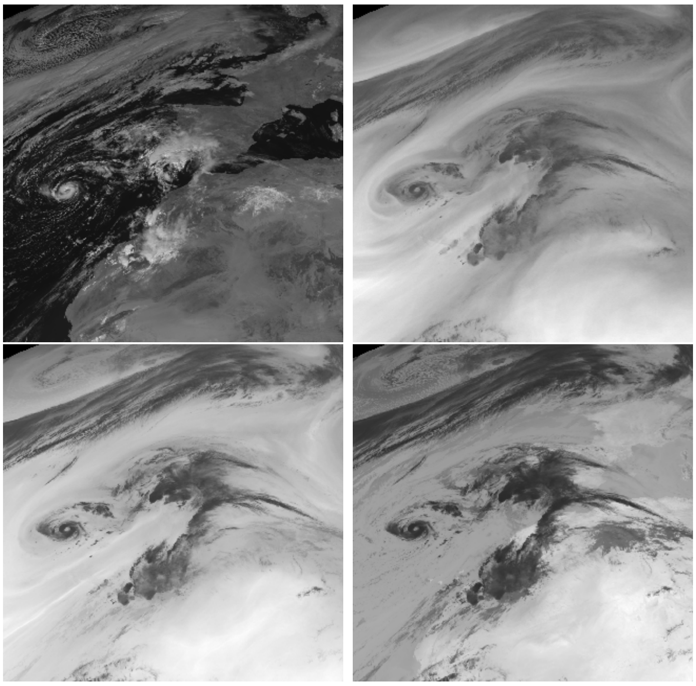


Figura 2.2: Secuencia de *Vince*. De izquierda a derecha y de arriba hacia abajo, se muestra el canal visible 0,81μm, los canales de vapor de agua 6,25μm y 7,35μm y a la derecha el canal de infrarrojo, 10,80μm.

canal. La mejora ha sido de 27.34 % y 27.58 % en el AEE y AAE, respectivamente. Existe una pequeña diferente de 1.61 % (AEE) y 1.92 % (AAE) entre las dos estrategias del método multicanal. En la figura 2.5 se muestra una imagen a color que representa el AEE entre el *ground truth* y la mejor estimación del método multicanal. Como se puede observar la solución obtenida es bastante buena excepto en la costa africana y en los bordes de la imagen.

Método/Canal	**AEE**	**AAE**
SF, VIS 0.8	0.2608	5.2055
SF, WV 6.2	0.5280	12.3891
SF, WV 7.3	0.4992	11.9396
SF, IR 10.8	0.3413	7.8710
MC Avg.	0.1926	3.8432
MC Max.	**0.1895**	**3.7696**

Cuadro 2.1: Secuencia de Vince: AEE y AAE obtenidos por las distintas versiones del método monocanal y multicanal.

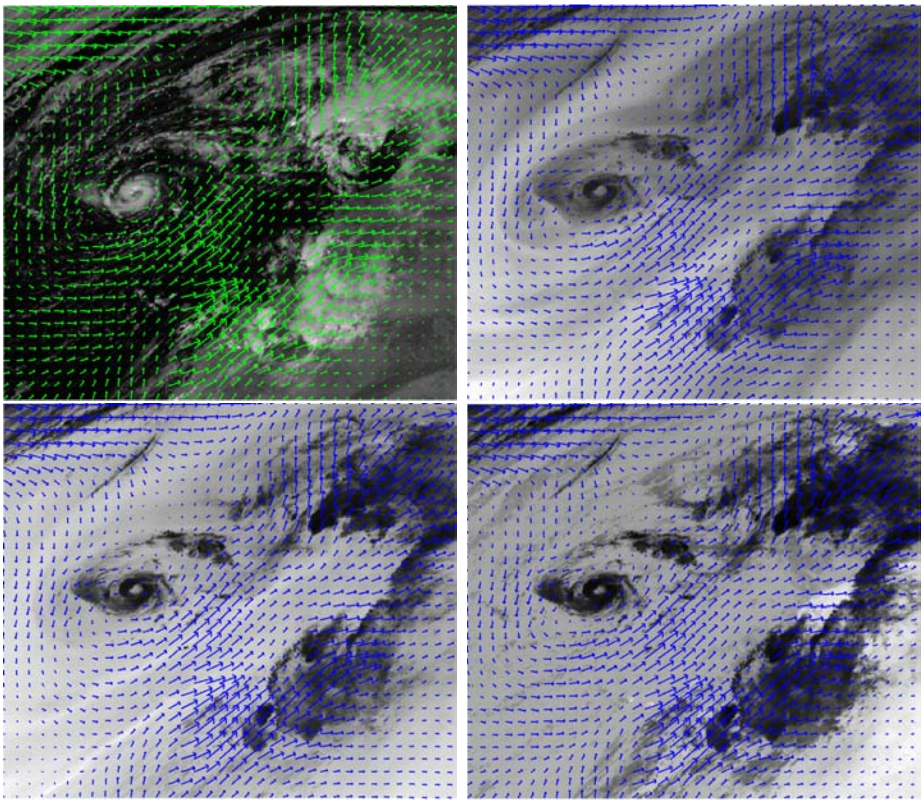


Figura 2.3: Secuencia 1. Arriba: el campo desplazamiento de los canales VIS 0.8 y WV 6.2. Abajo: el campo desplazamiento de los canales WV 7.3 y IR 10.8.

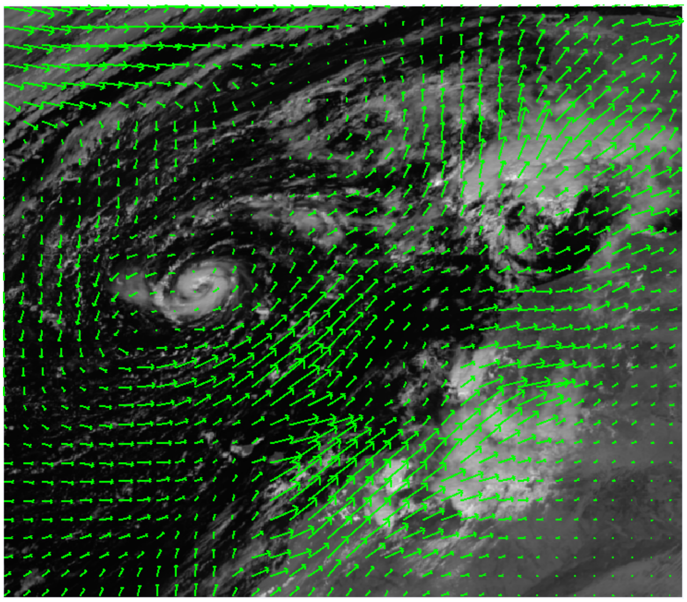


Figura 2.4: Secuencia 1. Campo de desplazamiento obtenido con el método multicanal (estrategia *gradiente Máximo*).

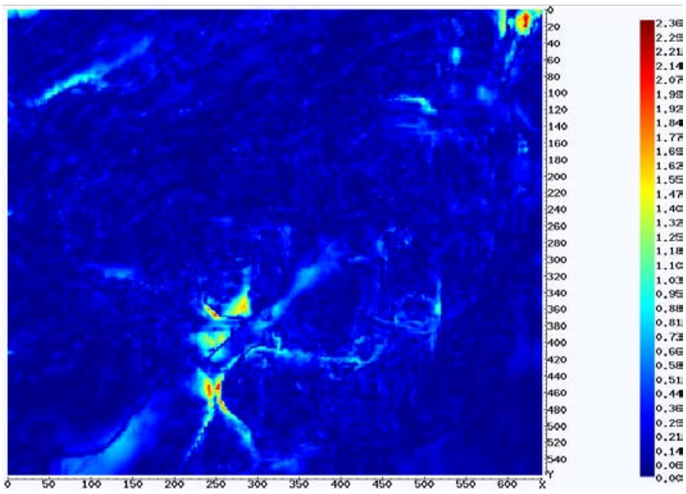


Figura 2.5: El Error Euclídeo Medio (AEE) entre el *ground truth* y la mejor estimación del método multicanal.

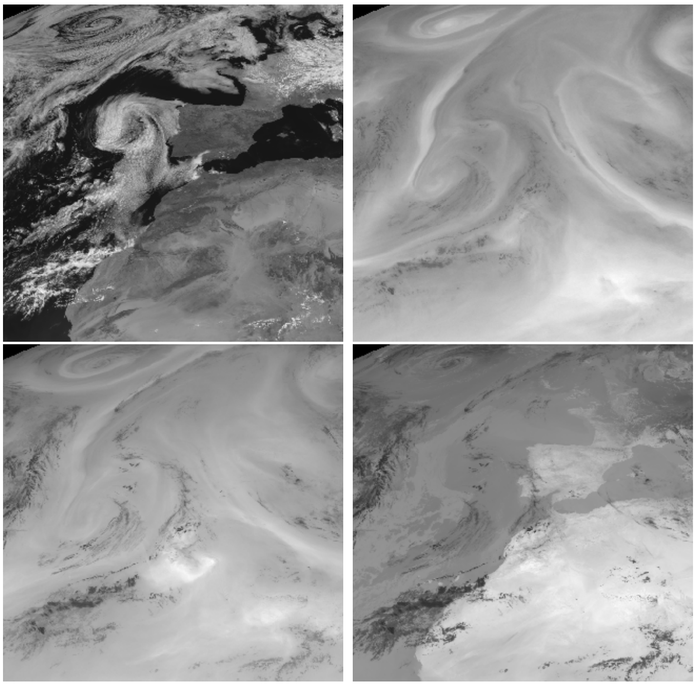


Figura 2.6: Secuencia del *Atlántico Norte*. De izquierda a derecha y de arriba hacia abajo, se muestra el canal visible 0,81μm, los canales de vapor de agua 6,25μm y 7,35μm y a la derecha el canal de infrarrojo, 10,80μm.

Secuencia 2. Atlántico Norte (5 Junio 2004)

La segunda secuencia utilizada en los experimentos corresponde a unos datos del Meteosat del día 5 de Junio del 2004. Se trata de una vista de la zona afroeuropea del hemisferio norte. En la figura 2.6 podemos observar las imágenes de cuatro canales del satélite. Dado el tamaño de las imágenes se ha seleccionado una región de 559×575 píxeles en donde, a nuestro entender, se registran los fenómenos atmosféricos más importantes (figura 2.8). En esa región se localizan dos grandes masas nubosas. Una de ellas es una borrasca que se aproxima del Atlántico hacia las islas británicas por la parte norte. La segunda masa nubosa se sitúa cerca de la península Ibérica. El movimiento predominante en ambas masas nubosas es rotacional. En la figura 2.7 se muestra la estimación del desplazamiento de las nubes en cada uno de los canales del satélite. Estos resultados son muy parecidos en magnitud. Sin embargo, en el canal visible se detecta con más claridad las vorticidades presentes en la imagen; mientras que en el resto de canales ese fenómeno se detecta como casi traslacional. Por ello, la contribución de los datos de este canal debe ser mayor.

En la figura 2.8 se representa el campo de desplazamiento obtenido con el método multicanal utilizando los pesos con los valores comentados anteriormente. Estos pesos otorgan un valor predominante al canal visible. En las figuras 2.7 y 2.8 se han mostrado los distintos campos de desplazamiento calculados con el método variacional monocanal y multicanal. Visualmente no es posible percibir las pequeñas diferencias entre los resultados. Por este motivo, en la tabla 2.2 se recogen unos datos cuantitativos que permiten comparar con más exactitud los distintos resultados. La mejora de la estimación hecha por el método multicanal frente a las del monocanal es de al menos 20.77 % para el AEE y de 23.47 % para el AAE. Al igual que ocurría en la secuencia anterior las dos estrategias utilizadas en el método multicanal ofrecen resultados parecidos. En la figura 2.9 se muestra una imagen a color que representa el AEE entre el *ground truth* y la mejor estimación del método multicanal.

Método/Canal	**AEE**	**AAE**
SF, VIS 0.8	0.1704	4.4748
SF, WV 6.2	0.5813	15.1845
SF, WV 7.3	0.4776	12.6222
SF, IR 10.8	0.3064	7.7821
MC Avg.	0.1593	4.1831
MC Max.	**0.1350**	**3.4246**

Cuadro 2.2: Secuencia del Atlántico Norte: AEE y AAE obtenidos por las distintas versiones del método monocanal y multicanal.

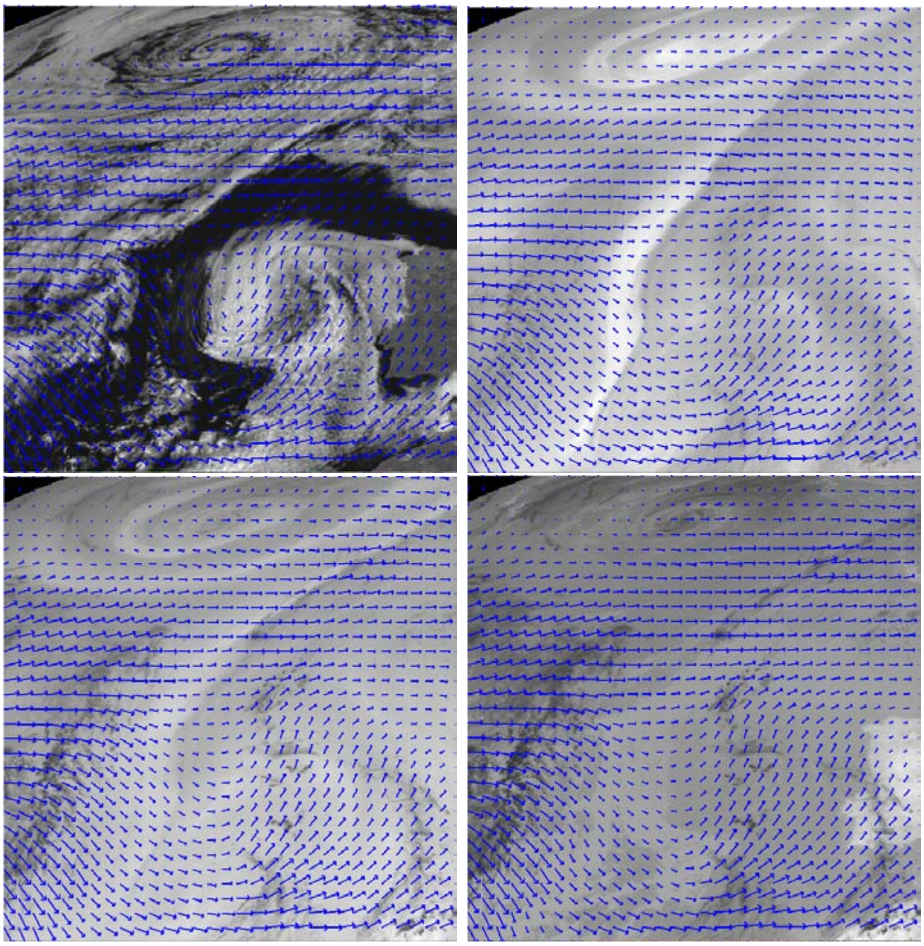


Figura 2.7: Secuencia 2. Arriba: el campo desplazamiento de los canales VIS 0.8 y WV 6.2. Abajo: el campo desplazamiento de los canales WV 7.3 y IR 10.8.

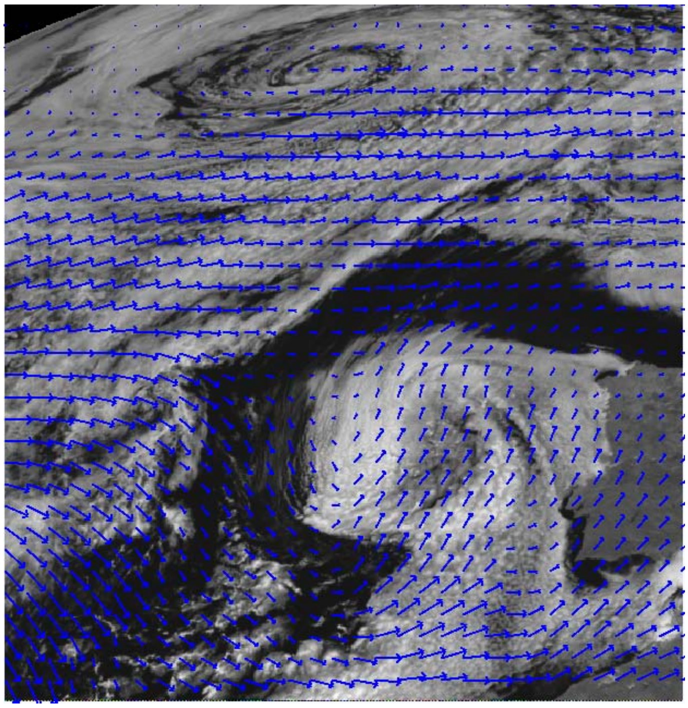


Figura 2.8: Secuencia 2. Campo de desplazamiento obtenido con el método multicanal (estrategia *gradiente Máximo*).

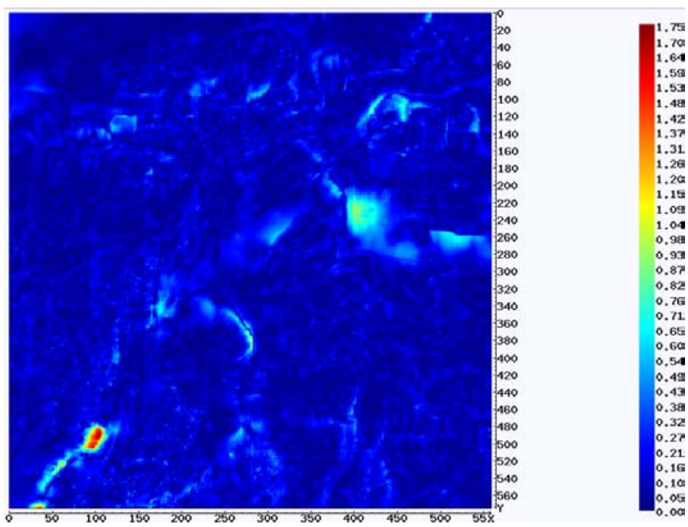


Figura 2.9: El Error Euclídeo Medio (AEE) entre el *ground truth* y la mejor estimación del método multicanal.

2.4. Método Variacional que incorpora un Término de Regularización exclusivamente Temporal no Continuo

En esta sección se describe un método para la estimación del flujo óptico en una secuencia de imágenes. Se trata de una modificación del modelo de energía propuesto por Nagel-Enkelmann [Nagel86] al que se le ha añadido un término de regularización temporal.

En la ecuación (2.31) podemos observar el modelo de energía propuesto en [Papenberg06]. En ese trabajo se analizó el impacto de la combinación de varias invarianzas en el término de ligadura. El término de ligadura utilizado se compone de dos invarianzas: gradiente constante y la suposición lambertiana, y el de suavizado se trata de una versión espaciotemporal robustificada del propuesto por Horn-Schunck.

$$\begin{aligned} E(\mathbf{h}) &= \int_{\Omega} \Psi\left((I_1(\mathbf{x}) - I_2(\mathbf{x}+\mathbf{h}))^2 + \gamma\left(\nabla I_1(\mathbf{x}) - \nabla I_2(\mathbf{x}+\mathbf{h})\right)^2 \right) d\mathbf{x} \\ &\quad + \alpha \int_{\Omega} \Psi\left(\|\nabla_3 u\|^2 + \|\nabla_3 v\|^2 \right) d\mathbf{x}. \end{aligned} \tag{2.31}$$

Si tomamos como referencia el trabajo de [Papenberg06], su modelo de energía incluye un término de regularización basado en el gradiente espaciotemporal del flujo. Este término impone una restricción de continuidad en el flujo en todas direcciones, es decir, que el desplazamiento sea suave tanto en la dirección espacial como temporal. Este método ha demostrado ser uno de los más precisos que existen. En la evaluación de los métodos se utilizan secuencias sintéticas que omiten parte de la complejidad del mundo real. En situaciones reales la restricción de Papenberg et al. no es un modelo satisfactorio para la estimación de ciertos tipos de flujos. Pese a que el tratamiento de forma conjunta de la información espacial y temporal ofrece una serie de ventajas, obliga a una continuidad en el dominio temporal que no es siempre posible. Por ello, como parte de esta tesis se ha desarrollado un método que añade un término de regularización temporal independiente del espacial. De esta forma, se desacopla la información del flujo espacial del temporal evitando la limitación anteriormente comentada.

Antes de entrar en detalle en la descripción de nuestro modelo de energía conviene repasar alguno de los términos de regularización más conocidos en la literatura. Uno de los primeros métodos variacionales que se propusieron para la estimación del flujo óptico fue el de Horn–Schunck [Horn81]. El funcional de energía se definía mediante dos términos: (1) la ecuación de restricción del flujo óptico, ec. (1.2), y (2) un término que consiste en la norma del gradiente del flujo óptico.

$$\|\nabla \mathbf{h}\|^2 = \|\nabla u\|^2 + \|\nabla v\|^2 , \tag{2.32}$$

donde $\|.\|$ es el la norma de un vector y ∇ es el operador gradiente. Uno de los inconvenientes que tiene este término de regularización es que no preserva las discontinuidades en el flujo.

El modelo de energía propuesto por Nagel-Enkelmann [Nagel86] supone otra de las grandes aportaciones en el campo del flujo óptico. Este modelo, similar al definido por Horn-Schunck, incluye un término de suavizado anisotrópico. Este término varía la cantidad de difusión aplicada en función del gradiente de la imagen. En las regiones donde el gradiente es pequeño actúa de forma isotrópica y en aquéllas donde el gradiente es alto de forma anisotrópica, suavizando a lo largo de los contornos y no a través de ellos. Inicialmente, el modelo de Nagel-Enkelmann [Nagel86] se diseñó para dos frames, ec. (2.33); pero luego se creó una versión espaciotemporal del mismo, [Nagel90]. El método descrito en esta sección se basa en la versión espacial.

$$E(\mathbf{h}) = \int_\Omega \left(I_1(\tilde{\mathbf{x}}) - I_2(\tilde{\mathbf{x}} + \mathbf{h})\right)^2 d\tilde{\mathbf{x}} + \alpha \int_\Omega tr(\nabla \mathbf{h}^\top D \nabla \mathbf{h}) d\tilde{\mathbf{x}}, \tag{2.33}$$

donde $\tilde{\mathbf{x}}$ es un punto en el dominio Ω, $(.)^\top$ se trata del operador de trasposición, I_1 y I_2 son las dos imágenes de entrada, $tr(.)$ es el operador conocido como *trace*, α es un peso constante que pondera al término de suavizado y D es una matriz de proyección regularizada en la dirección ortogonal al ∇I_1.

La matriz D se define como

$$D(\nabla I_1) = \frac{1}{\|\nabla I_1\|^2 + 2\zeta^2} \left(\xi\xi^\top + \zeta^2 Id\right), \tag{2.34}$$

donde $\xi = (\frac{\partial I_1}{\partial y}, -\frac{\partial I_1}{\partial x})^\top$ es el vector ortogonal a ∇I_1, Id es la matriz identidad, y ζ es un coeficiente que determina el comportamiento isotrópico aplicado en el suavizado e inhibe la difusión a través de los bordes: $\|\nabla I_1\| \gg \zeta$.

La estructura de esta sección se divide de la siguiente forma. En la sección 2.4.1 se introducirá el modelo de energía y se justificará la división de la regularización en dos términos distintos, uno espacial y otro temporal. En la sección 2.4.2, se describe con detalle el esquema numérico utilizado a partir de la minimización del funcional de energía. Así mismo, en la sección 2.4.5, se mostrará una serie de experimentos que justificarán la inclusión del término de regularización temporal. Para ello, se ha evaluado nuestro método utilizando secuencias tanto reales como sintéticas.

2.4.1. Modelo de Energía

Los primeros métodos variacionales propuestos para la estimación del flujo óptico incluían términos de regularización espacial. Estos términos suavizan la solución utilizando la información procedente de dos imágenes. Cuando se dispone de una secuencia de imágenes éstos no hacen uso de la información de los frames vecinos. Para aumentar la precisión de las estimaciones y aprovechar la transferencia de información de los frames consecutivos se han diseñado nuevos términos de regularización espaciotemporales. Muchos de ellos, sólo son una extensión temporal de su homólogo espacial ([Weickert01b, Nagel90]).

En el trabajo de Weickert-Schnörr [Weickert01b] se propone un método que no es más que una extensión espaciotemporal del método de Horn-Schunck. El término de regularización es de la forma $\mathcal{R}(|\nabla_3 u|^2 + |\nabla_3 v|^2)$, donde $\mathcal{R}$ se trata de una función de robustificación que penaliza en menor grado las desviaciones en el flujo. Este término trata de forma conjunta las derivadas espaciales y temporal. Por un lado, la formulación del modelo queda expresada de una forma homogénea. Por otro lado, esta formulación continua está condicionada a que el desplazamiento de los píxeles sea pequeño.

A partir del método [Weickert01b] se creó una extensión para desplazamientos largos, [Weickert04]. El nuevo modelo de energía propuesto incluía como término de ligadura, la ecuación de restricción de flujo óptico para largos desplazamientos, ec. (2.10) y, para el de suavizado se utilizó –$\mathcal{R}(|\nabla_3 u|^2 + |\nabla_3 v|^2)$–, el mismo que en [Weickert01b]. Este funcional de energía tiene dos limitaciones. En primer lugar, cuando se producen desplazamientos largos la regularización temporal puede influir negativamente en la regularización espacial debido al acoplamiento que existe entre las derivadas espaciotemporales. En segundo lugar, se da una ligera incongruencia entre el término de ligadura y el de suavizado. Uno permite discontinuidades en el tiempo y, el otro, exige que el flujo sea continuo en el tiempo.

Una forma de evitar estas dos limitaciones consiste en reemplazar el término $\mathcal{R}(|\nabla_3 u|^2 + |\nabla_3 v|^2)$ por $\mathcal{R}(|\nabla u|^2 + |\nabla v|^2) + \mathcal{T}(|u_t|^2 + |v_t|^2)$ en el que las derivadas espaciales y temporales estén separadas. Todavía se sigue teniendo el segundo inconveniente, la continuidad del flujo y la imposibilidad de manejar largos desplazamientos. Esto se soluciona sustituyendo la derivada temporal por estimación que tenga en cuenta los desplazamientos largos. El término $\mathcal{T}(.)$ debe permitir la detección de los desplazamientos largos en la dirección temporal por lo que la derivada temporal del flujo no es válida. Si se asume que la velocidad del objeto es suave en toda la secuencia, los flujos ópticos entre frames vecinos deben ser parecidos. Esta semejanza se puede expresar como $\mathbf{h}_i(\tilde{\mathbf{x}})$ y $\mathbf{h}_{i+1}(\tilde{\mathbf{x}} + \mathbf{h}_i(\tilde{\mathbf{x}}))$ para el flujo en el frame i y el consecutivo, respectivamente. Por ello, el término $\mathcal{T}(.)$ estaría expresado en función de $\mathcal{T}(\mathbf{h}_i(\tilde{\mathbf{x}}), \mathbf{h}_{i+1}(\tilde{\mathbf{x}} + \mathbf{h}_i(\tilde{\mathbf{x}})))$.

De acuerdo a la generalización descrita en la sección 2.2, nuestro modelo de energía se ajusta a un caso particular de la ecuación (2.15). Se trata de un método espaciotemporal que utiliza la información de toda la secuencia. Se compone de dos partes: un término de ligadura muy simple que sólo incluye una sola invarianza y un término de regularización algo más complejo que suaviza por separado el flujo espacial y temporal. El modelo de energía propuesto tiene la estructura siguiente

$$
\begin{aligned}
E(\mathbf{h}) \quad = \quad & \sum_{i=1}^{N-1} \int_{\Omega} \mathcal{D}(I_i, I_{i+1}, \mathbf{h}_i)\; d\tilde{\mathbf{x}} \\
& + \alpha \sum_{i=1}^{N-1} \int_{\Omega} \mathcal{R}(\nabla I_i, \nabla \mathbf{h}_i)\; d\tilde{\mathbf{x}} \\
& + \beta \sum_{i=1}^{N-2} \int_{\Omega} \mathcal{T}(\mathbf{h}_i, \mathbf{h}_{i+1})\; d\tilde{\mathbf{x}}. \qquad (2.35)
\end{aligned}
$$

Este funcional se compone de tres términos: (i) la ecuación de restricción del flujo óptico para largos desplazamientos (ec. (2.11) con una sola invarianza, la suposición lambertiana), (ii) el operador de Nagel-Enkelmann y (iii) un término temporal $T(\mathbf{h}_i, \mathbf{h}_{i+1}) =$

$\Phi(\|\mathbf{h}_i - \mathbf{h}_{i+1}(\tilde{\mathbf{x}} + \mathbf{h}_i)\|^2)$.

$$\begin{aligned} E(\mathbf{h}) &= \sum_{i=1}^{N-1} \int_\Omega (I_i - I_{i+1}(\tilde{\mathbf{x}} + \mathbf{h}_i))^2 \; d\tilde{\mathbf{x}} \\ &+\alpha \sum_{i=1}^{N-1} \int_\Omega trace(\nabla \mathbf{h}_i^T \mathbf{D}(\nabla I_i) \nabla \mathbf{h}_i^T) \; d\tilde{\mathbf{x}} \\ &+\beta \sum_{i=1}^{N-2} \int_\Omega \Phi(\|\mathbf{h}_i - \mathbf{h}_{i+1}(\tilde{\mathbf{x}} + \mathbf{h}_i)\|^2) \; d\tilde{\mathbf{x}}, \end{aligned} \tag{2.36}$$

donde $\Phi(x^2) = 1 - \gamma e^{-\frac{x^2}{\gamma}}$. Cuando el desplazamiento es pequeño, $\mathbf{h}_i - \mathbf{h}_{i+1}(\tilde{\mathbf{x}} + \mathbf{h}_i)$ es una aproximación a la derivada temporal. Un inconveniente que tiene este término es que la transferencia de información se hace desde los últimos frames a los primeros. Existe una fuerte dependencia respecto al último flujo de la secuencia; si éste es malo, esa estimación se propagará negativamente al resto de la secuencia. Una forma de compensar este problema consiste en la inclusión de otro término, también temporal, que favorezca la transferencia inversa desde los primeros frames a los últimos, ec. (2.37). De esta forma, nuestro modelo de energía incluye información temporal del flujo sin premiar/penalizar a ninguno en especial. La transferencia inversa del flujo ($\mathbf{h}^*$) desde los primeros frames a los últimos se explica con detalle en el apartado 2.4.4.

$$\begin{aligned} E(\mathbf{h}) &= \sum_{i=1}^{N-1} \int_\Omega (I_i - I_{i+1}(\tilde{\mathbf{x}} + \mathbf{h}_i))^2 \; d\tilde{\mathbf{x}} \\ &+\alpha \sum_{i=1}^{N-1} \int_\Omega trace(\nabla \mathbf{h}_i^T \mathbf{D}(\nabla I_i) \nabla \mathbf{h}_i^T) \; d\tilde{\mathbf{x}} \\ &+\beta \sum_{i=1}^{N-2} \int_\Omega \Phi(\|\mathbf{h}_i - \mathbf{h}_{i+1}(\tilde{\mathbf{x}} + \mathbf{h}_i)\|^2) \; d\tilde{\mathbf{x}} \\ &+\beta \sum_{i=2}^{N-1} \int_\Omega \Phi(\left\|\mathbf{h}_i - \mathbf{h}_{i-1}(\tilde{\mathbf{x}} + \mathbf{h}_{i-1}^*)\right\|^2) \; d\tilde{\mathbf{x}} \end{aligned} \tag{2.37}$$

Los dos últimos términos de la ecuación (2.37) asumen un modelo en la velocidad de los objetos. Esto obliga a que las funciones $\mathbf{h}_i$ sean similares en dirección y magnitud. Por ello, este modelo de energía es más adecuado cuando los desplazamientos de los objetos por la escena son constantes y en la misma dirección. Cuando el desplazamiento no se ajuste a esta suposición, $\Phi(.)$ decrecerá por el efecto de $\mathcal{T}(.)$ perdiendo importancia la regularización temporal convirtiendo al modelo en puramente espacial.

Definir un modelo genérico para $\mathcal{T}(.)$ que tenga en cuenta todas las posibles situaciones que pueden darse en secuencias reales no es tarea sencilla. Otras alternativas para el término temporal que podrían favorecer un movimiento constante independientemente de la dirección podría ser $\Phi(\|\mathbf{h}_i\|^2 - \|\mathbf{h}_{i+1}(\tilde{\mathbf{x}} + \mathbf{h}_i)\|^2) + \Phi(\|\mathbf{h}_i\|^2 - \left\|\mathbf{h}_{i-1}(\tilde{\mathbf{x}} + \mathbf{h}_{i-1}^*)\right\|^2)$. La minimización de la diferencia de la magnitud de los flujos permitiría la detección de las rotaciones o traslaciones. Con el término temporal se favorecen determinados patrones de movimiento frente a otros.

2.4.2. Minimización de la Energía

Los métodos variacionales son técnicas basadas en la minimización de energía. El flujo óptico es la solución de la minimización de un funcional de energía. Para obtener dicha solución basta con derivar las ecuaciones de Euler–Lagrange a partir de dicha energía, que en nuestro caso, se trataría de la ecuación (2.37).

Las ecuaciones de Euler–Lagrange asociadas vienen dadas por el sistema de EDP

$$\begin{aligned}
\mathbf{0} \;=\; & -\left(I_i(\tilde{\mathbf{x}}) - I_{i+1}(\tilde{\mathbf{x}}+\mathbf{h}_i)\right) \nabla I_{i+1}(\tilde{\mathbf{x}}+\mathbf{h}_i) \\
& -\alpha \left(\mathbf{div}(\mathbf{D}\left(\nabla \mathbf{h}_i\right) \nabla u_i), \mathbf{div}(\mathbf{D}\left(\nabla \mathbf{h}_i\right) \nabla v_i)\right)^T \\
& +\beta\, \Phi'(\|\mathbf{h}_i - \mathbf{h}_{i+1}(\tilde{\mathbf{x}}+\mathbf{h}_i)\|^2) \\
& \cdot \Big((\mathbf{h}_i - \mathbf{h}_{i+1}(\tilde{\mathbf{x}}+\mathbf{h}_i))^T \left(\mathbf{Id} - \nabla \mathbf{h}_{i+1}^T(\tilde{\mathbf{x}}+\mathbf{h}_i)\right) \Big) \\
& +\beta \Phi'(\left\|\mathbf{h}_i - \mathbf{h}_{i-1}(\tilde{\mathbf{x}}+\mathbf{h}_{i-1}^*)\right\|^2) \\
& \cdot \left(\mathbf{h}_i - \mathbf{h}_{i-1}(\tilde{\mathbf{x}}+\mathbf{h}_{i-1}^*)\right)
\end{aligned} \tag{2.38}$$

En la ecuación (2.38) se ha utilizado una notación unificada para expresar mediante la función $\mathbf{h}$ los desplazamientos horizontales y verticales de píxeles.

Si aplicamos la técnica de descenso del gradiente alcanzamos la solución del sistema de ecuaciones (2.38).

$$\begin{aligned}
\frac{\partial \mathbf{h}_i}{\partial t} \;=\; & \left(I_i(\tilde{\mathbf{x}}) - I_{i+1}(\tilde{\mathbf{x}}+\mathbf{h}_i)\right) \frac{\partial I_{i+1}(\tilde{\mathbf{x}}+\mathbf{h}_i)}{\partial x} \\
& +\alpha \left(\mathbf{div}(\mathbf{D}\left(\nabla \mathbf{h}_i\right) \nabla u_i), \mathbf{div}(\mathbf{D}\left(\nabla \mathbf{h}_i\right) \nabla v_i)\right)^T \\
& -\beta\, \Phi'(\|\mathbf{h}_i - \mathbf{h}_{i+1}(\tilde{\mathbf{x}}+\mathbf{h}_i)\|^2) \\
& \cdot \Big((\mathbf{h}_i - \mathbf{h}_{i+1}(\tilde{\mathbf{x}}+\mathbf{h}_i))^T \left(\mathbf{Id} - \nabla \mathbf{h}_{i+1}^T(\tilde{\mathbf{x}}+\mathbf{h}_i)\right) \Big) \\
& -\beta\, \Phi'(\left\|\mathbf{h}_i - \mathbf{h}_{i-1}(\tilde{\mathbf{x}}+\mathbf{h}_{i-1}^*)\right\|^2) \\
& \cdot \left(\mathbf{h}_i - \mathbf{h}_{i-1}(\tilde{\mathbf{x}}+\mathbf{h}_{i-1}^*)\right)
\end{aligned} \tag{2.39}$$

donde $i = 1, \ldots, N-1$. El sistema se compone de $N-1$ pares de ecuaciones que corresponden con los flujos ópticos de la secuencia completa.

Para poder detectar los desplazamientos largos las ecuaciones (2.39) se introducirán en un enfoque multipiramidal. Se trata de una estrategia *coarse–to–fine* ampliamente utilizada en la literatura ([Enkelmann88, Anandan89, Mémin02, Weickert04]). En el enfoque multipiramidal se crean un cierto número de escalas $s_1, s_2, \ldots s_n$, donde cada una representa imágenes de distinto tamaño. Normalmente el factor de escala aplicado es de 0,5 pero se puede usar valores reales entre $(0, 1)$. En [Weickert04] se recomienda que el factor de escala η esté dentro del rango [0,5, 0,95]. En cada escala se resuelve el sistema de ecuaciones anterior para todo el conjunto de incógnitas $\{u_i^s, v_i^s\}$.

2.4.3. Esquema Numérico

En esta sección se describe el esquema número utilizado para implementar nuestro modelo de energía. Para ello discretizamos el sistema de ecuaciones parabólico (2.39) por diferencias finitas. Las derivadas espaciales se han aproximado por diferencias centradas y para la discretización en la dirección de t utilizamos un esquema lineal implícito. Si se utiliza la siguiente notación para las componentes de la matriz $\mathbf{D}\left(\nabla I_{i,j,k}\right) = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ el esquema numérico para los desplazamientos horizontales y verticales quedaría expresado como

$$
\begin{aligned}
\frac{u_{i,j,k}^{n+1} - u_{i,j,k}^{n}}{\tau} = \; & \left(I_{i,j,k} - I_{i,j,k+1}^{\mathbf{h}_{i,j,k}^{n}} + u_{i,j,k}^{n+1} I_{x_{i,j,k+1}}^{\mathbf{h}_{i,j,k}^{n}} + v_{i,j,k}^{n+1} I_{y_{i,j,k+1}}^{\mathbf{h}_{i,j,k}^{n}}\right) I_{x_{i,j,k+1}}^{\mathbf{h}_{i,j,k}^{n}} \\
& - u_{i,j,k}^{n+1} I_{x_{i,j,k+1}}^{2,\mathbf{h}_{i,j,k}^{n}} - v_{i,j,k}^{n+1} I_{y_{i,j,k+1}}^{\mathbf{h}_{i,j,k}^{n}} I_{x_{i,j,k+1}}^{\mathbf{h}_{i,j,k}^{n}} \\
& + \alpha \left(\frac{a_{i+1,j,k} + a_{i,j}}{2} \frac{u_{i+1,j,k}^{n+1} - u_{i,j,k}^{n}}{h_1^2} + \frac{a_{i-1,j,k} + a_{i,j}}{2} \frac{u_{i-1,j,k}^{n+1} - u_{i,j,k}^{n+1}}{h_1^2} \right. \\
& + \frac{c_{i,j+1,k} + c_{i,j}}{2} \frac{u_{i,j+1,k}^{n+1} - u_{i,j,k}^{n+1}}{h_2^2} + \frac{c_{i,j-1,k} + c_{i,j}}{2} \frac{u_{i,j-1,k}^{n+1} - u_{i,j,k}^{n+1}}{h_2^2} \\
& + \frac{b_{i+1,j+1,k} + b_{i,j}}{2} \frac{u_{i+1,j+1,k}^{n+1} - u_{i,j,k}^{n+1}}{2h_1h_2} + \frac{b_{i-1,j-1,k} + b_{i,j}}{2} \frac{u_{i-1,j-1,k}^{n+1} - u_{i,j,k}^{n+1}}{2h_1h_2} \\
& \left. + \frac{b_{i+1,j-1,k} + b_{i,j}}{2} \frac{u_{i+1,j-1,k}^{n+1} - u_{i,j,k}^{n+1}}{2h_1h_2} + \frac{b_{i-1,j+1,k} + b_{i,j}}{2} \frac{u_{i-1,j+1,k}^{n+1} - u_{i,j,k}^{n+1}}{2h_1h_2} \right) \\
& - \beta \Phi' \left[\left(u_{i,j,k}^{n+1} - u_{i,j,k+1}^{n,\mathbf{h}_{i,j,k}^{n}} + u_{i,j,k}^{n+1} u_{x_{i,j,k+1}}^{n,\mathbf{h}_{i,j,k}^{n}} + v_{i,j,k}^{n+1} u_{y_{i,j,k+1}}^{n,\mathbf{h}_{i,j,k}^{n}} \right) \left(1 - u_{x_{i,j,k+1}}^{n,\mathbf{h}_{i,j,k}^{n}}\right) \right. \\
& - u_{i,j,k}^{n+1} u_{x_{i,j,k+1}}^{n,\mathbf{h}_{i,j,k}^{n}} \left(1 - u_{x_{i,j,k+1}}^{n,\mathbf{h}_{i,j,k}^{n}}\right) - v_{i,j,k}^{n+1} u_{y_{i,j,k+1}}^{n,\mathbf{h}_{i,j,k}^{n}} \left(1 - u_{x_{i,j,k+1}}^{n,\mathbf{h}_{i,j,k}^{n}}\right) \\
& + \left(v_{i,j,k}^{n+1} - v_{i,j,k+1}^{n,\mathbf{h}_{i,j,k}^{n}} + u_{i,j,k}^{n+1} v_{x_{i,j,k+1}}^{n,\mathbf{h}_{i,j,k}^{n}} + v_{i,j,k}^{n+1} v_{y_{i,j,k+1}}^{n,\mathbf{h}_{i,j,k}^{n}} \right) v_{x_{i,j,k+1}}^{n,\mathbf{h}_{i,j,k}^{n}} \\
& \left. - u_{i,j,k}^{n+1} v_{x_{i,j,k+1}}^{2,n,\mathbf{h}_{i,j,k}^{n}} - v_{i,j,k}^{n+1} v_{y_{i,j,k+1}}^{n,\mathbf{h}_{i,j,k}^{n}} v_{x_{i,j,k+1}}^{n,\mathbf{h}_{i,j,k}^{n}} \right] \\
& - \beta \Phi' \left[u_{i,j,k}^{n+1} - u_{i,j,k+1}^{n,\mathbf{h}_{i,j,k}^{*,n}} \right] \qquad (2.40)
\end{aligned}
$$

$$
\begin{aligned}
\frac{v^{n+1}_{i,j,k} - v^{n}_{i,j,k}}{\tau} = {} & \left(I_{i,j,k} - I^{\mathbf{h}^n_{i,j,k}}_{i,j,k+1} + u^{n+1}_{i,j,k} I^{\mathbf{h}^n_{i,j,k}}_{x_{i,j,k+1}} + v^{n+1}_{i,j,k} I^{\mathbf{h}^n_{i,j,k}}_{y_{i,j,k+1}}\right) I^{\mathbf{h}^n_{i,j,k}}_{y_{i,j,k+1}} \\
& - u^{n+1}_{i,j,k} I^{\mathbf{h}^n_{i,j,k}}_{x_{i,j,k+1}} I^{\mathbf{h}^n_{i,j,k}}_{y_{i,j,k+1}} - v^{n+1}_{i,j,k} I^{2,\mathbf{h}^n_{i,j,k}}_{y_{i,j,k+1}} \\
& + \alpha \left(\frac{a_{i+1,j,k} + a_{i,j}}{2} \frac{v^{n+1}_{i+1,j,k} - v^{n+1}_{i,j,k}}{h_1^2} + \frac{a_{i-1,j,k} + a_{i,j}}{2} \frac{v^{n+1}_{i-1,j,k} - v^{n+1}_{i,j,k}}{h_1^2} \right. \\
& + \frac{c_{i,j+1,k} + c_{i,j}}{2} \frac{v^{n+1}_{i,j+1,k} - v^{n+1}_{i,j,k}}{h_2^2} + \frac{c_{i,j-1,k} + c_{i,j}}{2} \frac{v^{n+1}_{i,j-1,k} - v^{n+1}_{i,j,k}}{h_2^2} \\
& + \frac{b_{i+1,j+1,k} + b_{i,j}}{2} \frac{v^{n+1}_{i+1,j+1,k} - v^{n+1}_{i,j,k}}{2h_1h_2} + \frac{b_{i-1,j-1,k} + b_{i,j}}{2} \frac{v^{n+1}_{i-1,j-1,k} - v^{n+1}_{i,j,k}}{2h_1h_2} \\
& \left. + \frac{b_{i+1,j-1,k} + b_{i,j}}{2} \frac{v^{n+1}_{i+1,j-1,k} - v^{n+1}_{i,j,k}}{2h_1h_2} + \frac{b_{i-1,j+1,k} + b_{i,j}}{2} \frac{v^{n+1}_{i-1,j+1,k} - v^{n+1}_{i,j,k}}{2h_1h_2} \right) \\
& - \beta\Phi' \left[\left(v^{n+1}_{i,j,k} - v^{n,\mathbf{h}^n_{i,j,k}}_{i,j,k+1} + u^{n+1}_{i,j,k} v^{n,\mathbf{h}^n_{i,j,k}}_{x_{i,j,k+1}} + v^{n+1}_{i,j,k} v^{n,\mathbf{h}^n_{i,j,k}}_{y_{i,j,k+1}} \right) \left(1 - v^{n,\mathbf{h}^n_{i,j,k}}_{y_{i,j,k+1}}\right) \right. \\
& - u^{n+1}_{i,j,k} v^{n,\mathbf{h}^n_{i,j,k}}_{x_{i,j,k+1}} \left(1 - v^{n,\mathbf{h}^n_{i,j,k}}_{y_{i,j,k+1}}\right) - v^{n+1}_{i,j,k} v^{n,\mathbf{h}^n_{i,j,k}}_{y_{i,j,k+1}} \left(1 - v^{n,\mathbf{h}^n_{i,j,k}}_{y_{i,j,k+1}}\right) \\
& + \left(u^{n+1}_{i,j,k} - u^{n,\mathbf{h}^n_{i,j,k}}_{i,j,k+1} + u^{n+1}_{i,j,k} u^{n,\mathbf{h}^n_{i,j,k}}_{x_{i,j,k+1}} + v^{n+1}_{i,j,k} u^{n,\mathbf{h}^n_{i,j,k}}_{y_{i,j,k+1}} \right) u^{n,\mathbf{h}^n_{i,j,k}}_{y_{i,j,k+1}} \\
& \left. - u^{n+1}_{i,j,k} u^{n,\mathbf{h}^n_{i,j,k}}_{x_{i,j,k+1}} u^{n,\mathbf{h}^n_{i,j,k}}_{y_{i,j,k+1}} - v^{n+1}_{i,j,k} u^{2,n,\mathbf{h}^n_{i,j,k}}_{y_{i,j,k+1}} \right] \\
& - \beta\Phi' \left[v^{n+1}_{i,j,k} - v^{n,\mathbf{h}^{*,n}_{i,j,k}}_{i,j,k+1} \right]
\end{aligned}
\tag{2.41}
$$

donde j, k indican la posición del píxel e $I^{\mathbf{h}^{n+1}_{i,j,k}}_{i+1,j,k} = I_{i+1}(\tilde{\mathbf{x}}_{j,k} + \mathbf{h}^{n+1}_{i,j,k})$. El subíndice i representa la dimensión temporal, τ es el paso temporal (un parámetro utilizado en el descenso del gradiente) y $v_{y_{i,j,k+1}}$ representa la derivada de v en la dirección vertical en el píxel $(j, k+1)$ del frame i. Estas derivadas también se aproximan por diferencias finitas.

En el esquema implícito las variables del nuevo instante de tiempo aparecen en ambas partes de la ecuación. Este tipo de esquema ha demostrado ser bastante estable. En sistemas de ecuaciones *no lineales* tienen un alto coste computacional, pero en el caso *lineal* son relativamente más rápidos. Se puede obtener muy buenas aproximaciones de términos donde la *no linealidad* está presente mediante su desarrollo de Taylor.

Los sistemas de ecuaciones lineales (2.40) y (2.41) se resuelven utilizando el algoritmo de Gauss–Seidel.

2.4.4. Cálculo del Flujo Inverso, $\mathbf{h}^*$

En esta sección se describe cómo se calcula el flujo *inverso* a partir del flujo *directo.* En la ecuación (2.42) es fácil de ver la relación entre ambos flujos.

$$\mathbf{h}(\tilde{\mathbf{x}}) = -\mathbf{h}^*(\tilde{\mathbf{x}} + \mathbf{h}(\tilde{\mathbf{x}})) \tag{2.42}$$

El problema está en cómo tratar este cálculo cuando tenemos imágenes discretas. El flujo $\mathbf{h}(\tilde{\mathbf{x}})$ tiene precisión real pero la posición de los píxeles $(\tilde{\mathbf{x}})$ están expresadas en precisión entera. $\mathbf{h}^*(\tilde{\mathbf{x}})$ tiene valores únicamente en algunas posiciones discretas. En la figura 2.10 se observa que la correspondencia no se ajusta exactamente a la posición de un píxel por lo que es necesario hacer un promediado de las porciones del flujo que caen sobre cada píxel.

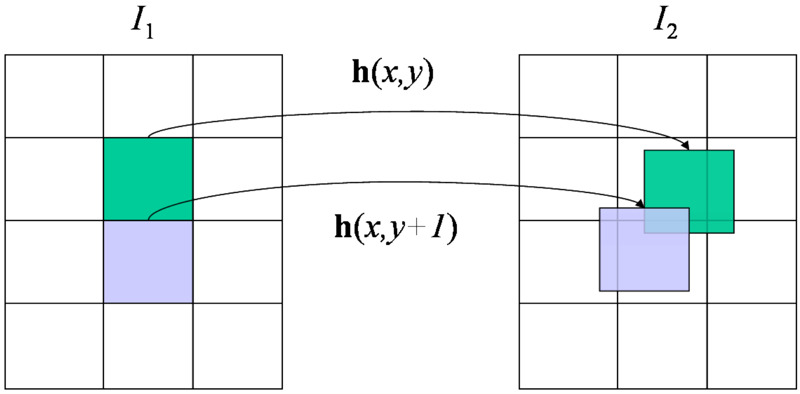


Figura 2.10: La suma de la posición entera de un píxel, $\tilde{\mathbf{x}}$, y el flujo asociado a esa posición, $\mathbf{h}(\tilde{\mathbf{x}})$, no coincide con las posiciones del flujo inverso, $\mathbf{h}^*(\tilde{\mathbf{x}})$. Para calcular el valor de la función discreta $\mathbf{h}^*(\tilde{\mathbf{x}})$ es necesario dividir cada correspondencia en cuatro estimaciones, uno para cada píxel.

Dada la naturaleza discreta la función $\mathbf{h}^*(\tilde{\mathbf{x}})$ se podría estimar según la ecuación (2.43)

$$\mathbf{h}^*(\tilde{\mathbf{x}}_i) = -\frac{\displaystyle\sum_{j=1}^{N} \mathbf{h}(\tilde{\mathbf{x}}_j) p_{i,j}\left(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j + \mathbf{h}(\tilde{\mathbf{x}}_j)\right)}{\displaystyle\sum_{j=1}^{N} p_{i,j}\left(\tilde{\mathbf{x}}_i, \tilde{\mathbf{x}}_j + \mathbf{h}(\tilde{\mathbf{x}}_j)\right)} \tag{2.43}$$

donde N es el tamaño de la imagen y $p_{i,j}$ indica el área del píxel que cubre el flujo óptico (Figura 2.10). Cada $\tilde{\mathbf{x}}_j + \mathbf{h}(\tilde{\mathbf{x}}_j)$ generará cuatro diferentes $p_{i,j}$ que hace referencia a los píxeles vecinos. Los pesos $p_{i,j}$ se calculan como

$$p_{i,j}(\mathbf{a}, \mathbf{b}) = \begin{cases} (1-(b_x-a_x))\cdot(1-(b_y-a_y)) & si\ a_x < b_x\ y\ a_y < b_y \\ (1-(b_x-a_x))\cdot(b_y-a_y) & si\ a_x < b_x\ y\ a_y > b_y \\ (b_x-a_x)\cdot(1-(b_y-a_y)) & si\ a_x > b_x\ y\ a_y < b_y \\ (b_x-a_x)\cdot(b_y-a_y) & si\ a_x > b_x\ y\ a_y > b_y \\ 0 & si\ distancia\,(\mathbf{a},\mathbf{b}) \geq 1 \end{cases}$$

donde $\mathbf{a} = (a_x, a_y) = \tilde{\mathbf{x}}_i$ y $\mathbf{b} = (b_x, b_y) = \tilde{\mathbf{x}}_j + \mathbf{h}(\tilde{\mathbf{x}}_j)$.

El cálculo de $p_{i,j}$ no es más que la estimación de la parte proporcional del área del flujo que cae en los cuatro vecinos. Hay que tener en cuenta que los píxeles vacíos, en los que no existe correspondencia en la otra imagen, se tratan posiblemente de oclusiones.

Una forma de mejorar las estimaciones del flujo consiste en la identificación y tratamiento diferenciado de estos puntos conflictivos. Se rellenan los huecos del flujo mediante un promediado con la información de los píxeles vecinos.

2.4.5. Resultados Experimentales

En este apartado queremos mostrar las mejoras que ofrece nuestro término de regularización temporal. Para ello vamos a realizar una comparación de tres modelos: (1) *Spatial*, un modelo de energía puramente espacial, ec. (2.33); (2) *Temporal*, con un modelo que incluye el término $\mathcal{T}(x) = \Phi(\|\mathbf{h}_i - \mathbf{h}_{i+1}(\tilde{\mathbf{x}} + \mathbf{h}_i)\|^2)$, ec. (2.36) y, (3) *Bidirectional–Temporal* con un doble término temporal, $\mathcal{T}(x) = \Phi(\|\mathbf{h}_i - \mathbf{h}_{i+1}(\tilde{\mathbf{x}} + \mathbf{h}_i)\|^2) + \Phi(\left\|\mathbf{h}_i - \mathbf{h}_{i-1}(\tilde{\mathbf{x}} + \mathbf{h}^*_{i-1})\right\|^2)$, ec. (2.37).

En la evaluación de los modelos se ha utilizado tres secuencias que han sido descritas en la sección 1.4.1. Estas secuencias corresponden con Secuencia del *Cuadrado* (Fig. 1.4) y *Marble Blocks* (Fig. 1.5) y el *Taxi de Hamburgo* (Fig. 1.11).

Antes de comentar los resultados experimentales obtenidos conviene recordar los distintos parámetros que intervienen en nuestro método.

- Los parámetros de regularización α y β determinan el peso del término de suavizado espacial y temporal, respectivamente. Cuando el valor de este parámetro es alto genera campos de desplazamiento más suaves.

- El factor de isotropía ζ se calcula a través del histograma acumulativo de las magnitudes de los gradientes de las imágenes. Este parámetro especifica el comportamiento isotrópico o anisotrópico del término de regularización. Tanto α como ζ son parámetros que se calculan siguiendo las ecuaciones descritas en el apartado 2.3.1. Por lo tanto los parámetros de entrada serán λ y C. Valores altos de λ provoca que el término sea isotrópico y, anisotrópico en caso contrario. El rango de valores oscila entre $[0, 1]$.

- γ. Se trata de una constante de configuración de la función de robustificación ($\Phi(x^2) = 1 - \gamma e^{-\frac{x^2}{\gamma}}$) que define el comportamiento del término de regularización temporal (ver figura 2.11).

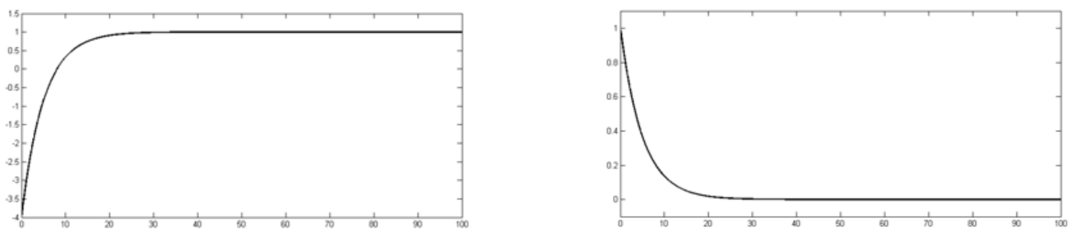


Figura 2.11: A la izquierda, la función de robustificación $\Phi(x^2)$. A la derecha, su derivada $\Phi'(x^2)$. $\gamma = 5$.

- El intervalo de tiempo τ se trata de un parámetro puramente numérico que se utiliza en el algoritmo del descenso del gradiente. Hemos fijado $\tau = 10$; para valores de τ más pequeños la convergencia del algoritmo requiere un mayor número de iteraciones.

- El tiempo de parada T especifica el número de iteraciones del algoritmo que se llevan a cabo en cada escala del enfoque piramidal. En los experimentos se ha utilizado $T = 500$. Valores mayores no aseguran la mejora en la calidad de las soluciones ya que el algoritmo alcanza un estado asintótico en una etapa temprana ($T < 500$).
- El ratio de descenso $\eta \in (0, 1)$ para el cálculo de las escalas. En [Weickert04] se recomienda que el factor de escala η esté dentro del rango $[0{,}5, 0{,}95]$. Hemos utilizado $\eta = 0{,}5$.
- n. Indica el número de escalas utilizadas en la estrategia multipiramidal. El número de escalas depende del desplazamiento máximo presente en la secuencia de imágenes. A priori, este dato no se tiene porqué conocer y, se suele optar por utilizar un valor alto para este parámetro.

Secuencia del *Cuadrado*

La primera secuencia utilizada en estos experimentos es la del *Cuadrado*. En ella se observa un cuadrado negro desplazándose horizontalmente de izquierda a derecha sobre un fondo texturado (figura 2.12). La velocidad del objeto es uniforme y de quince píxeles por frame. La textura del fondo también se desplaza, a una velocidad de tres píxeles. Para más información acerca de esta secuencia podemos ir al apartado 1.4.1 de la tesis.

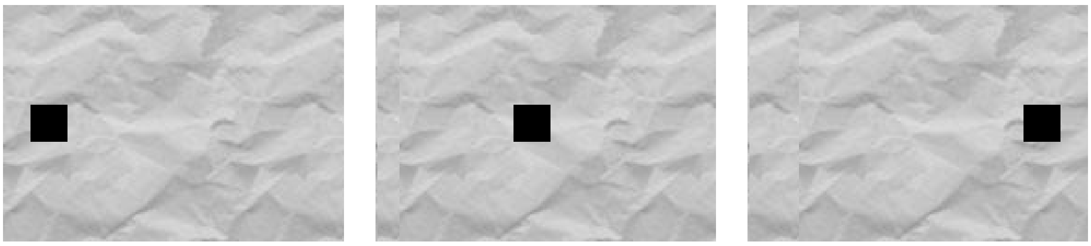


Figura 2.12: Frames 0, 4 y 9 de la *Secuencia0*.

En la tabla 2.4 y en las figuras 2.13, 2.15 se muestran los resultados cuantitativos de los tres métodos utilizando las métricas AEE y AAE. En los resultados cuantitativos se aprecia cierta estabilidad de los métodos temporales respecto al espacial. El AAE y AEE es muy parecido en todos los frames, hecho que no ocurre con el método espacial. Si analizamos el AAE observamos que los métodos *Bi–Temporal* y *Temporal* son un 7,56 % y 6,47 % más preciso que el *Spatial*. Si nos centramos en el AEE vemos que el resultado del *Temporal* es algo superior al *Spatial*, más concretamente un 3,18 %. Como se ha comentado en los apartados anteriores el último flujo del método *Temporal* es puramente espacial. El término temporal transfiere esta información a los frames anteriores; una mala estimación del último frame empeora a la de sus predecesores. El método *Bi–Temporal* empeora ligerísimamente respecto al *Spatial* en un 0,26 %. Sin embargo, la desviación estándar del AAE y AEE demuestra la estabilidad de las estimaciones obtenidas.

En la figura 2.14 se muestra varias imágenes con el desplazamiento real del cuadrado y los flujos ópticos obtenidos con los métodos *Spatial* y *Bi-Temporal* que hemos comparado en los experimentos. Pese a que visualmente no existe grandes diferencias entre las

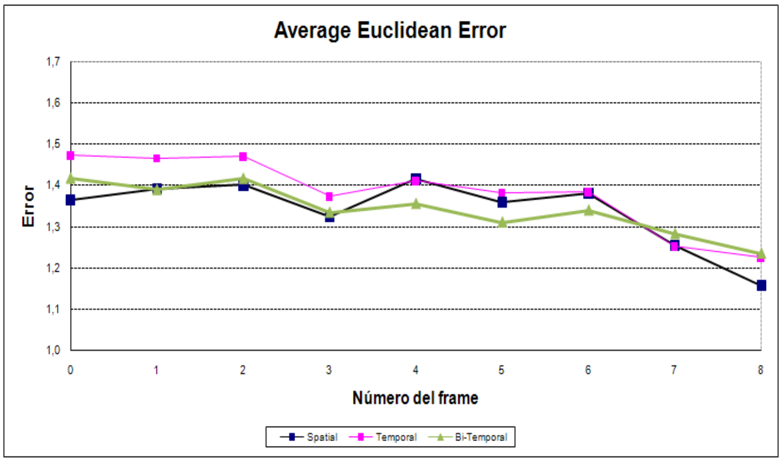


Figura 2.13: AEE de la secuencia del *Cuadrado*. Las diferencias entre las estimaciones del método espacial y espaciotemporal son mínimas aunque si se observa la estabilidad que aporta la información temporal. Los cambios en la magnitud del error entre los distintos frames es menos abrupta que en el caso espacial.

Cuadro 2.3: Lista de parámetros utilizados por los métodos *Spatial*, *Temporal* y *Bi–Temporal* en la secuencia del *cuadrado*. Escalas = número de escalas en el enfoque multipiramidal. C = parámetro de regularización espacial. β = parámetro de regularización temporal. γ = parámetro de control de la función Φ del término de regularización temporal. λ = factor de isotropía del operador de Nagel–Enkelmann.

Cuadrado

Método	Escalas	C	β	γ	λ
Spatial	4	0.6	-	-	0.1
Temporal	4	0.6	0.01	0.1	0.3
Bi–Temporal	4	0.6	0.01	0.5	0.1

soluciones si que se puede percibir la aportación del término de regularización temporal que ha permitido controlar el proceso de difusión respecto a la estimación espacial. El desplazamiento del fondo de la escena se aprecia más homogéneo en la solución temporal que en la espacial. En la tabla 2.3 mostramos los parámetros utilizados en los experimentos. Dado que el desplazamiento del objeto es de 15 píxeles es necesario emplear al menos cuatro escalas en el enfoque multipiramidal. A la hora de estimar correctamente el flujo óptico en la escena y evitar que el proceso de difusión actúe más allá de las discontinuidades del cuadrado debemos utilizar un valor de λ bajo, anisotrópico, mientras que C tendría un valor intermedio.

Como se ha visto, el término de regularización temporal permite mejorar la estimación hecha por el espacial. Cuando el movimiento en la secuencia es traslacional y constante se manifiesta la eficacia de este nuevo término de regularización temporal. En la siguiente secuencia veremos cómo se comporta el método ante las aceleraciones de los objetos en

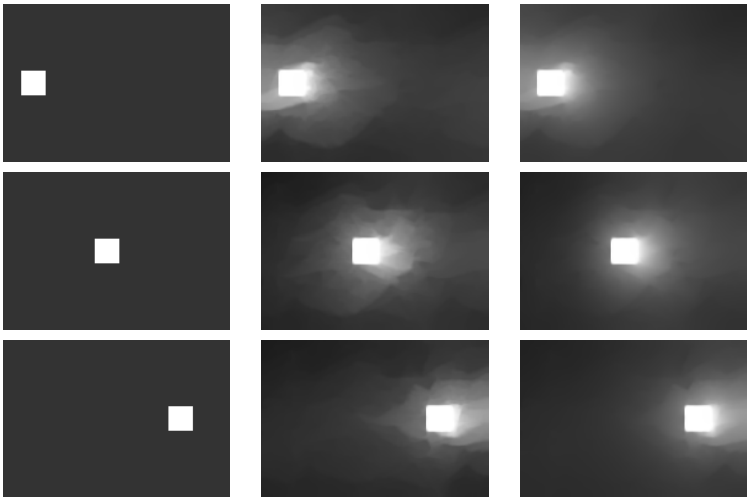


Figura 2.14: De arriba a abajo los campos de desplazamiento correspondiente a los frames 0, 4 y 8 de la secuencia del *Cuadrado*. En la primera columna los desplazamientos reales. En las siguientes columnas las estimaciones obtenidas con los métodos *Spatial* y *Bi–Temporal*. Las estimaciones obtenidas por los distintos métodos son muy parecidas. Si comparamos las soluciones temporales con la espacial observamos que la aportación del término de regularización temporal ha permitido controlar o acotar el proceso de difusión. En la estimación espacial se aprecia una regularización algo descontrolada.

determinados frames.

Marble Blocks

Marble Blocks se trata de una secuencia donde la cámara se mueve horizontalmente mientras una serie de torres de mármol permanecen estáticas (figura 2.16). La secuencia se compone de treinta frames. En nuestros resultados experimentales sólo se ha mostrado la estimación de quince de ellos (los situados en la parte central, del frame 5 al 20).

El movimiento aparente de la escena indica que todas la torres se mueven a distinta velocidad, en función de la distancia a la cámara. En principio, el movimiento registrado en esta secuencia se asemeja mucho a la anterior. Sin embargo, la velocidad de los objetos en cada uno de los frames no es constante como se pudiera pensar. En los frames 4, 8, 13, 18, 23 y 28 existen unas aceleraciones de los objetos respecto a los frames contiguos. Estos cambios en la velocidad se pueden percibir sutilmente a simple vista. Para verificar esta hipótesis hemos calculado la velocidad media entre frames, utilizando el flujo real suministrado por Nagel, y hemos observado que este valor se incrementa notablemente en los frames especificados anteriormente. La velocidad media en los distintos frames de la secuencia es de $1,33$ con una desviación típica de $0,03$. En los frames indicados previamente la velocidad media se incrementa hasta los $1,58$ con una desviación típica de $0,02$.

En la secuencia de *Marble Blocks* las superficies de los objetos están cubiertas por texturas. Los regularizadores del tipo *image-driven*, como es el caso del operador de Nagel–Enkelmann, no se comportan muy bien ya que podrían dar lugar a la creación de flujos

Secuencia del *Cuadrado*				
Método	AAE$_\mu$	AAE$_\sigma$	AEE$_\mu$	AEE$_\sigma$
Spatial	6,2116^o	0,3455^o	1,3393	0,0830
Temporal	5,8100^o	0,1347^o	1,3819	0,0902
Bi-Temporal	5,7423^o	0,1612^o	1,3428	0,0612

Cuadro 2.4: AAE y AEE en la secuencia del cuadrado. Los subíndices μ y σ denotan la media y la desviación típica, respectivamente. Según vemos en los resultados cuantitativos el método Bi–Temporal mejora las estimaciones respecto a los otros dos métodos. La mejora del temporal unidireccional respecto al espacial no es tan evidente debido a la dependencia que existe con la estimación del último frame; aunque como podemos ver en la tabla, la desviación típica es muy baja lo que refleja la estabilidad ofrecida por la información temporal.

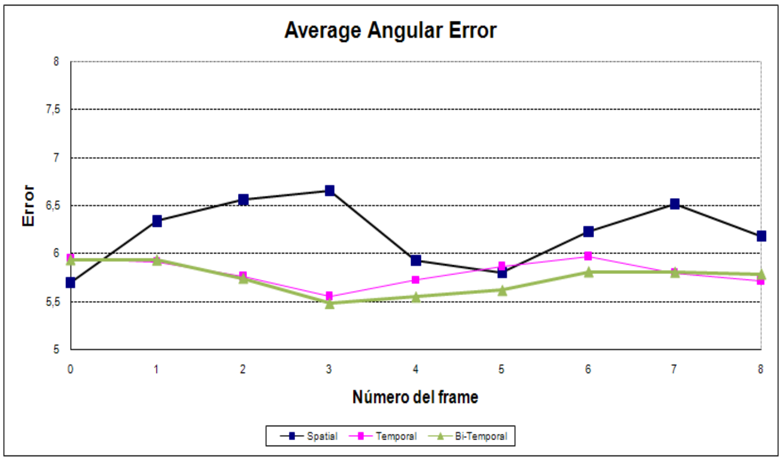


Figura 2.15: AAE de la secuencia del *Cuadrado*. Las estimaciones obtenidas por los métodos temporales son ligeramente más precisas. Nuevamente, la estimación del método espacial se vuelve inestable debido a la ausencia de la información temporal.

hipersegmentados al *confundir* las fuertes variaciones en la magnitud del gradiente en la textura con discontinuidades en la imagen. La presencia de las aceleraciones y las texturas en los objetos no supone el escenario ideal para nuestro método. A pesar de estos dos inconvenientes se verá que los resultados experimentales son relativamente buenos.

En la figura 2.17 podemos ver una comparación visual entre el *ground truth* y los flujos obtenidos por los métodos *Spatial* y *Bi–Temporal*. Si le echamos un vistazo con más detenimiento observamos que las estimaciones obtenidas por la versión *Spatial* y *Bi–Temporal* del método son bastante parecidas aunque existen pequeñas diferencias. En la estimación temporal el suelo se percibe más suavizado y el fondo de la escena está mucho más definido. Otras diferencias a destacar están en la torre situada en la parte derecha de la imagen que parece ser más continua, sin grandes artificios en el flujo. Por otro lado, el flujo de las dos torres situadas en la parte izquierda alejadas de la cámara parecen estar subestimadas en su parte central.

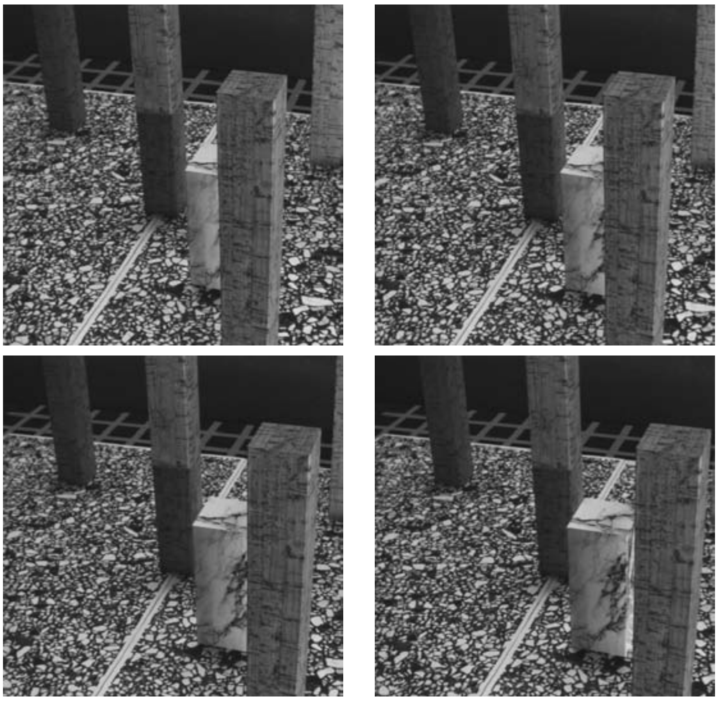


Figura 2.16: Frames 5, 10, 15, y 20 de la secuencia *Marble Blocks*.

Marble Blocks				
Método	AAE_μ	AAE_σ	AEE_μ	AEE_σ
Spatial	$6{,}695^o$	$2{,}698^o$	0,2480	0,0963
Temporal	$5{,}402^o$	$1{,}327^o$	0,2081	0,0638
Bi–Temporal	$4{,}731^o$	$1{,}330^o$	0,1848	0,0661

Cuadro 2.5: AAE y AEE para la secuencia *Marble Blocks*. En esta secuencia se aprecia la ventaja que ofrece la información espaciotemporal respecto a los métodos espaciales. La mejora del temporal unidireccional es notable y en una proporción similar al Bi–Temporal.

En las figuras 2.18 y 2.19 disponemos de dos gráficas que nos muestran los AEE y AAE de cada frame de la secuencia. Al igual que ocurría con la secuencia de los cuadrados la estimación hecha por el método puramente *Spatial* es más inestable y menos precisa. Las versiones temporales del método mejoran sustancialmente respecto al espacial. El comportamiento del *Bi–Temporal* es similar al *Temporal* pero con un error en magnitud relativamente inferior. Si observamos las figuras 2.18 y 2.19 existen unos picos en los valores de AEE y AAE. Estos frames coinciden con aquéllos en los que había aceleraciones de los objetos. Por este motivo, la estabilidad de las estimaciones se ve ligeramente comprometida, es decir, que el término de regularización temporal no identifica estas aceleraciones en la imagen y la estimación no es tan buena como en los frames vecinos donde si existe cierta continuidad.

En la tabla 2.5 se muestra los AAE y AEE globales junto con las desviaciones típicas

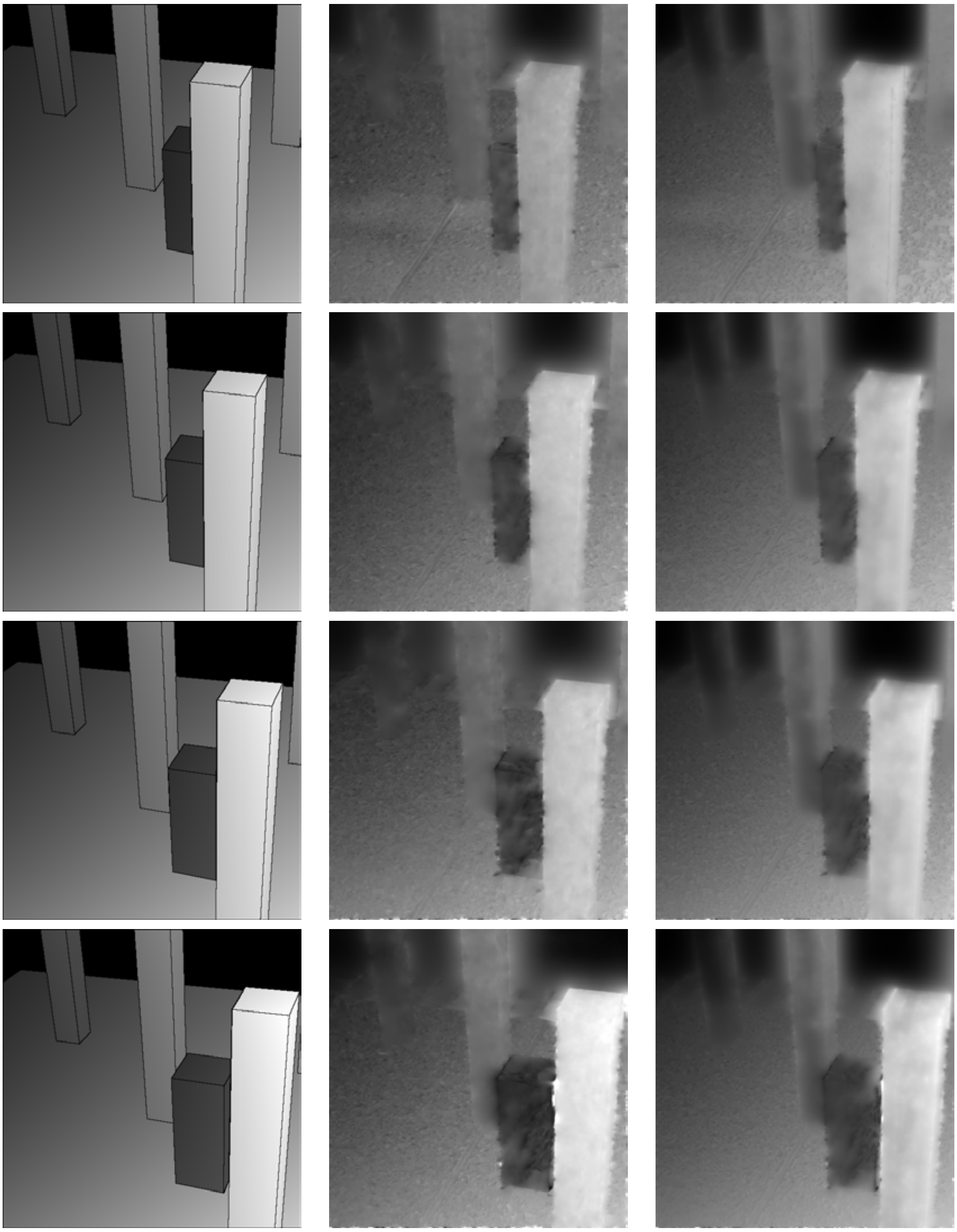


Figura 2.17: De arriba a abajo, el campo de desplazamiento correspondiente a los frames 5, 10, 15 y 20 de la secuencia del *Marble-Blocks*. En la primera columna los desplazamientos reales. En la columna central las estimaciones obtenidas con el método *Spatial* y en la columna de la derecha, las estimaciones del método *Bi–Temporal*. Existe bastante parecido entre las estimaciones. Sin embargo, si nos fijamos con más detenimiento la estimación del método Bi–Temporal es mucho más suave en las zonas homogéneas y no se aprecian grandes artificios. En el caso de la estimación espacial, la presencia de artificios es manifiesta y los bordes de los objetos están borrosos y menos definidos.

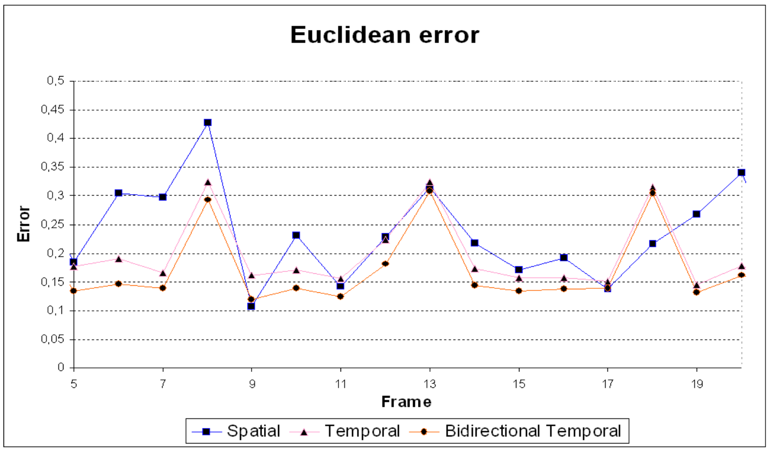


Figura 2.18: AEE para la secuencia de *Marble Blocks*. En esta gráfica sólo se muestra los errores desde el frame 5 hasta el 20. Se observa el efecto de las distintas aceleraciones detectadas en la secuencia. En los frames anteriormente señalados (8, 13 y 18) hay unos cambios notables de velocidad. El método temporal, en sus dos variantes, penaliza los cambios bruscos entre los flujos ópticos por lo que esos frames el error en la estimación será mayor. La última estimación del método temporal unidireccional debe coincidir con la del espacial. Sin embargo, los datos mostrados en la gráfica sólo recogen los errores del frame 5 al 20. Por último, cabe decir que se han seleccionado estos frames porque, por un lado, reflejan la estabilidad de los métodos temporales y, por otro lado, muestra el aumento del error en aquellos frames donde se producen las aceleraciones en los objetos.

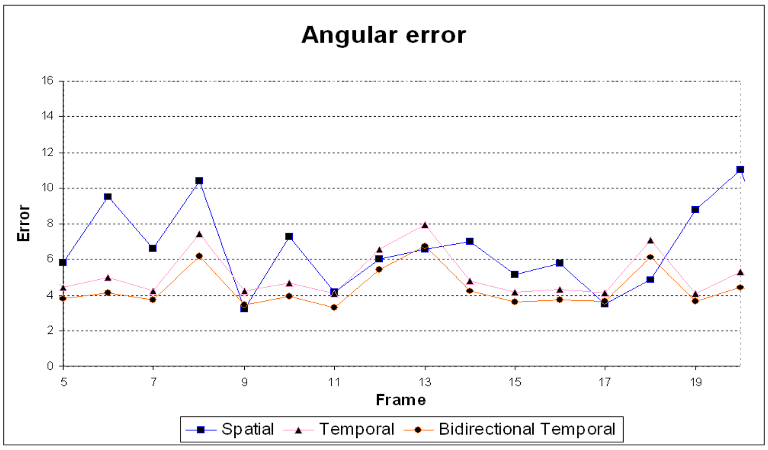


Figura 2.19: AAE para la secuencia de *Marble Blocks*. En esta gráfica sólo se muestra los errores desde el frame 5 hasta el 20. Al igual que se observaba en la gráfica 2.18 el AAE se incrementa en determinados frames que coinciden con los que se produce ciertas aceleraciones detectadas. El comportamiento de los métodos es similar al comentado en la gráfica 2.18.

Cuadro 2.6: Lista de parámetros utilizados por los métodos *Spatial*, *Temporal* y *Bi–Temporal* en la secuencia del *Marble Blocks*. Escalas = número de escalas en el enfoque multipiramidal. C = parámetro de regularización espacial. β = parámetro de regularización temporal. γ = parámetro de control de la función Φ del término de regularización temporal. λ = factor de isotropía del operador de Nagel–Enkelmann.

Marble Blocks

Método	Escalas	C	β	γ	λ
Spatial	2	0.6	-	-	0.5
Temporal	2	0.6	0.3	0.5	0.5
Bi–Temporal	2	0.6	0.3	0.5	0.5

asociadas. Según los datos de esta tabla el método *Temporal* obtiene respecto al *Spatial* en torno al 19,31 % de mejora del AAE y un 16,09 % en el caso de AEE. Si comparamos la mejora del *Bi–Temporal* respecto al *Spatial*, ésta es del 29,34 % y 25,48 % para el AAE y AEE, respectivamente. También cabe destacar la reducción de la desviación típica que se produce en las versiones temporales. En la tabla 2.6 mostramos los valores de los parámetros empleados en los experimentos.

En general, la inclusión del término temporal influye positivamente en la precisión de las estimaciones hechas por el método. Hay que tener en cuenta que esta secuencia no era, a priori, la más idónea debido a los inconvenientes comentados anteriormente.

Taxi de Hamburgo

La última secuencia utilizada en los experimentos es la conocida como *Taxi de Hamburgo* (fig. 2.20). En el apartado 1.4.1 del capítulo anterior podemos encontrar más información acerca de ella. Se trata de una secuencia real en la que se observan varios objetos en movimiento: tres vehículos y una persona; el resto de objetos de la secuencia permanecen estáticos. Al tratarse de una secuencia real no disponemos de *ground truth* por lo que solamente se mostrarán resultados visuales. En la figura 2.21, se muestran las soluciones de los métodos *Spatial* y *Bi–Temporal*.

Figura 2.20: Frames 0, 10 y 19 de la secuencia del *Taxi* de Hamburgo.

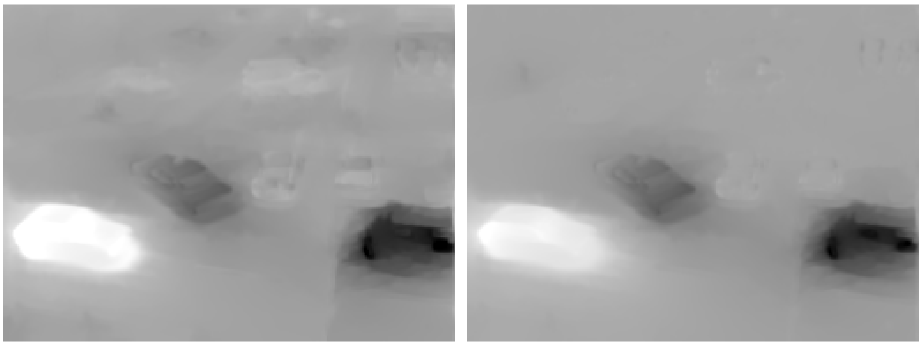


Figura 2.21: A la izquierda, la estimación obtenida por el método espacial. A la derecha, la solución ofrecida por el método temporal bidireccional. Los valores de los parámetros son: Número de escalas = 2, $C = 0{,}6$, $\beta = 0{,}1$, $\gamma = 0{,}5$, $\lambda = 0{,}1$.

Los dos vehículos situados en la parte inferior de la secuencia se mueven a velocidades constantes en la misma dirección pero en sentido opuesto. El tercer vehículo, un taxi, realiza un pequeño giro a baja velocidad y un transeúnte camina por la acera. Este fenómeno es casi imperceptible a simple vista. Nuestro modelo de energía no incorpora ninguna técnica de robustificación lo que le hace altamente sensible al ruido. En el caso espacial esta sensibilidad se hace más notable. Para una correcta identificación del movimiento de los tres vehículos se ha optado por un regularizador espacial con un comportamiento anisotrópico ($\lambda = 0{,}1$) y un peso para ambos términos de suavizado de $C = 0{,}6$ y $\beta = 0{,}1$ primando el espacial sobre el temporal.

Como vemos en la figura 2.21 el término de regularización temporal suaviza la solución espacial eliminando en gran medida los falsos movimientos detectados en esta solución. Teniendo en cuenta todo esto, existen importantes diferencias entre las estimaciones de ambos métodos y que comentamos a continuación:

1. Fondo de la escena.

   Salvo los cuatro objetos mencionados anteriormente en la escena no se registra ningún tipo de movimiento. El flujo óptico en esos píxeles debería ser nulo. La solución temporal es bastante suave aunque se observan algunos artificios producto de los efectos del ruido local. Por el contrario, la solución espacial dista mucho de la realidad, el número de movimientos fantasmas es elevado y el flujo no es homogéneo.

2. Peatón.

   El desplazamiento del peatón por la acera es un movimiento casi imperceptible a simple vista. El método temporal es capaz de detectarlo mientras que este movimiento al espacial se le solapa con los falsos positivos registrados en el fondo de la escena.

3. Movimiento de los vehículos.

   Los desplazamientos del taxi y del camión son similares en ambos métodos. El movimiento del coche situado a la izquierda parece estar mejor estimado en el caso espacial, pero en general, la estimación obtenida con el método temporal es más estable y precisa.

2.5. Métodos Variacionales basados en el Análisis Espectral del Tensor de Movimiento

En los últimos años los métodos variacionales presentados han incrementado aún más la precisión y robustez de sus estimaciones. Esta mejora se ha debido principalmente gracias a la incorporación de nuevos elementos al modelo de energía, a la utilización de nuevos esquemas numéricos y al uso de discretizaciones más exactas.

Los modelos de energía normalmente se componen de dos términos. El término de ligadura atrae a las estructuras en movimientos mientras que el de regularización actúa en aquellas zonas donde el término de ligadura no dispone de información suficiente para realizar una buena estimación. El funcionamiento idóneo de un método consistiría en la detección del desplazamiento en los bordes de los objetos (regiones donde se aprecia claramente el movimiento) y, a partir de esa información, rellenar el flujo para el resto de píxeles. El proceso de rellenado de los objetos sólo actuaría en el interior de los objetos y se detendría exactamente en las discontinuidades, no se propagaría más allá de los bordes de los objetos. A pesar de que en la literatura se han propuesto todo tipo de regularizadores que permiten guiar el proceso de difusión, sigue presente la sensación que los términos del modelo colaboran pero no se complementan del todo. El mayor problema radica en la identificación y elección de la dirección o direcciones a suavizar.

En esta sección se presenta un método para la estimación del flujo óptico usando secuencias de imágenes. El modelo de energía se apoya en técnicas recientes, como es el *tensor de movimiento*. A través del análisis espectral del tensor de movimiento es posible definir un modelo de energía que sea capaz de intercambiar información complementaria entre el término de ligadura y el de regularización. De esta forma, se establece un mecanismo directo que suaviza en las direcciones no dominantes del flujo. Para mejorar la precisión de las estimaciones se utiliza términos no-cuadráticos que han demostrado ser robustos frente a los *outliers*. Nuestro modelo es capaz de detectar los desplazamientos largos gracias al uso de técnicas de *warping*.

Esta sección se divide en cinco apartados. En primer lugar, se describe la notación y las técnicas utilizadas para la definición de nuestro modelo de energía genérico. Veremos como gracias a estas técnicas podemos establecer una relación de complementariedad entre ambos términos de la energía. En el apartado 2.5.2, comenzamos presentando el modelo de energía de Nagel–Enkelmann. La capacidad de este modelo para direccionar el proceso de difusión encierra la idea base de nuestro método. La combinación de las técnicas anteriores y del modelo de Nagel–Enkelmann dará como resultado la estructura general de nuestro *framework*. Una vez descrito las distintas variantes del modelo general, en los apartados 2.5.3 y 2.5.4, se minimizarán y se comentará el esquema numérico utilizado. En el apartado 2.5.5, se muestran los resultados experimentales obtenidos con secuencias reales y sintéticas. Comparamos nuestros resultados con los mejores métodos de la literatura demostrando que las ventajas de este nuevo framework no son sólo apreciables desde el punto de vista teórico.

2.5.1. Tensor de Movimiento

Para detectar las correspondencias de los píxeles entre dos imágenes se suele suponer que alguna propiedad de la imagen no varía a lo largo del tiempo. Esta suposición se puede representar tal que

$$f(x+u, y+v, t+1) - f(x, y, t) = 0, \tag{2.44}$$

donde f es algún tipo de propiedad en la imagen y, $t+1$ y t, representan dos imágenes de la escena en distintos instantes de tiempo. Esta no linealidad se suele solventar a través del desarrollo de Taylor de la ecuación (2.44) y eliminado los términos de orden superior con lo que obtendríamos

$$f_x u + f_y v + f_t = 0,$$

donde los subíndices indican derivadas parciales.

Para estimar el flujo óptico, $\mathbf{h}(\mathbf{x}) := (u(\mathbf{x}), v(\mathbf{x}), 1)^\top$, normalmente se recurre a la minimización por mínimos cuadrados de la expresión anterior

$$(f_x u + f_y v + f_t)^2 \;=\; \mathbf{h}^\top \nabla f \nabla f^\top \mathbf{h} \;=\; \mathbf{h}^\top J\, \mathbf{h}. \tag{2.45}$$

La expresión (2.45) se puede descomponer de tal forma que obtengamos una matriz, que por definición es, semidefinida positiva, [Bruhn06a]. Esta matriz se conoce como *tensor de movimiento*, $J := \nabla f \nabla f^\top$. Entre las numerosas ventajas que ofrece destaca: (i) la posibilidad de representar cualquier invarianza en una estructura compacta y, (ii) la inclusión múltiples suposiciones dentro de esa misma estructura. En tal situación obtendríamos un *tensor de movimiento* acumulado

$$J_{mc} = \sum_{i=1}^{N_c} \gamma_i \nabla f_i \nabla f_i^\top,$$

donde N_c es el número de invarianzas y γ_i un peso asociado a cada invarianza.

A continuación, se muestra un ejemplo de dos invarianzas, suposición lambertiana y gradiente constante, en el que se puede comprobar cómo quedan expresadas varias invarianzas mediante un *tensor de movimiento* acumulado.

$$\begin{aligned}
((f(x+u,y+v,t+1) - f(x,y,t))^2 \quad &+ \quad \gamma\, |\nabla f(x+u,y+v,t+1) - \nabla f(x,y,t)|^2 \;\approx \\
(f_x u + f_y v + f_t)^2 \quad &+ \quad \gamma\, \big((f_{xx} u + f_{xy} v + f_{xt})^2 \\
&\qquad + (f_{xy} u + f_{yy} v + f_{yt})^2\big) \;= \\
\mathbf{h}^\top \nabla f \nabla f^\top \mathbf{h} \quad &+ \quad \gamma\, \mathbf{h}^\top (\nabla f_x \nabla f_x^\top + \nabla f_y \nabla f_y^\top)\, \mathbf{h} \;= \\
\mathbf{h}^\top J_f\, \mathbf{h} \quad &+ \quad \gamma\, \mathbf{h}^\top J_{\nabla f}\, \mathbf{h} \;= \\
\mathbf{h}^\top J_{mc}\, \mathbf{h} &
\end{aligned} \tag{2.46}$$

donde $J_{mc} = \nabla f \nabla f^\top + \gamma\, (\nabla f_x \nabla f_x^\top + \nabla f_y \nabla f_y^\top)$.

Una característica de las matrices semidefinidas positivas es que todos sus autovalores son mayores o iguales a cero. Además, este tipo de matrices se puede expresar como una combinación lineal del producto de sus autovalores y autovectores:

$$J = \mu_1 r_1 r_1^\top + \mu_2 r_2 r_2^\top + \cdots + \mu_N r_N r_N^\top = \sum_{i=1}^{N} \mu_i\, r_i r_i^\top, \tag{2.47}$$

donde $\mu_1 > \mu_2 > \cdots > \mu_N \geq 0$ son los autovalores, $r_1, r_2, \cdots, r_N$ los autovectores y N la dimensión de la matriz cuadrada.

El análisis espectral es una herramienta utilizada en álgebra lineal para el estudio de datos. Dado un conjunto de elementos permite extraer la estructura de la distribución de los elementos. Dicho de otra manera, extrae la organización de los elementos del conjunto representándolos a través de direcciones (autovectores) y magnitud (autovalor). A primera vista, la descomposición en autovalores y autovectores del *tensor de movimiento* es simplemente otra forma de expresar la matriz. Sin embargo, detrás de esta elegante descomposición se esconden unas propiedades que encierran la idea principal del método que describimos en esta sección. Los autovectores indican las direcciones dominantes de la información almacenada en la matriz y los autovalores la importancia de cada dirección. Esta representación puede ser útil para el guiado del proceso de difusión. En principio, esta descomposición ofrece dos propiedades muy útiles: (i) la magnitud de un autovalor representa la importancia del autovector asociado y, (ii) los autovectores son ortogonales entre sí.

Los autovectores y autovalores en los que se descompone el *tensor de movimiento* tiene un significado diferente en función de la invarianza o invarianzas que incorpore. El caso más simple lo encontramos cuando representamos la suposición lambertiana. El *motion tensor* se convierte en el *tensor de estructuras.* En las zonas homogéneas los autovalores son de pequeña magnitud o cero. En las discontinuidades habrá un autovalor dominante de gran magnitud mientras que los otros dos, iguales o próximos a cero. El autovector asociado al autovalor de gran magnitud indica la dirección del gradiente del borde. En casos más complejos, como el expresado en la ecuación (2.46), la interpretación de los autovectores se puede ver como una dirección combinada entre las distintas invarianzas que componen el tensor. Aquí el peso asociado a cada invarianza juega un papel fundamental para contrarrestar el dominio de una invarianza respecto a la otra.

2.5.2. Modelo de Energía

El modelo de energía que se propone se asemeja al de Nagel-Enkelmann. Se compone de dos términos: el de ligadura y el de regularización. Usa información espaciotemporal y puede detectar los desplazamientos largos.

Para entender la idea que subyace a nuestro modelo de energía vamos a comenzar explicando la versión más simple e iremos incorporando una serie de técnicas hasta alcanzar la estructura general. En la ecuación (2.48) presentamos un modelo espaciotemporal con la suposición lambertiana en el término de ligadura que funciona con la misma filosofía que el de [Nagel90]. El flujo óptico se obtiene mediante la minimización de dicha expresión

$$\begin{aligned} E(\mathbf{h}) &= \int_\Omega \mathbf{h}^\top \nabla f \nabla f^\top \mathbf{h} \; dx\,dy\,dz \\ &+ \alpha \int_\Omega \left(\nabla u^\top \nabla f^\perp \nabla f^{\perp\top} \nabla u + \nabla v^\top \nabla f^\perp \nabla f^{\perp\top} \nabla v \right) \, dx\,dy\,dz, \end{aligned} \tag{2.48}$$

donde $\nabla := (\partial_x, \partial_y, \partial_z)^\top$ representa el gradiente espaciotemporal y la matriz $\nabla f^\perp \nabla f^{\perp\top}$ es la proyección perpendicular al gradiente de la imagen, ∇f. El primer término representa una constancia en la intensidad de los píxeles que atrae a los objetos en movimiento, mientras el segundo término expresa el suavizado aplicado a lo largo de contorno de los objetos. Si no fijamos existe un efecto complementario entre ambos términos de forma que el término de ligadura restringe el flujo óptico a lo largo del gradiente de la imagen mientras que el término de suavizado actúa perpendicular a él. Esta elegante descomposición distingue el método de Nagel–Enkelmann de otros métodos variacionales.

El método que se presenta en esta sección pretende funcionar de forma similar al modelo de Nagel–Enkelmann; la regularización sólo actúa en direcciones ortogonales al flujo dominante. El modelo de energía anterior se puede generalizar si la suposición lambertiana y el operador de Nagel–Enkelmann se expresan en forma de tensores, obteniendo así el primer prototipo que hemos desarrollado. Lo llamaremos *prototipo A*, ec. (2.49).

Prototipo A:

$$E_A(\mathbf{h}) = \int_\Omega \left(\mathbf{h}^\top J_1\,\mathbf{h} + \alpha\left(\nabla u^\top J_2 \nabla u + \nabla v^\top J_2 \nabla v\right)\right)\, dx\, dy\, dz. \tag{2.49}$$

Las matrices J_1 y J_2 son *tensores de movimiento* en los que existe una relación de complementariedad entre ellos. La matriz J_1 representa un *tensor* que puede incorporar cualquier invarianza, mientras que J_2 es un tensor inverso a J_1 que se construye a partir de los autovalores de éste.

$$J_1 = \sum_{i=1}^{3} \mu_i\, r_i r_i^\top, \quad J_2 = \sum_{i=1}^{3} g(\mu_i)\, r_i r_i^\top. \tag{2.50}$$

De esta forma, se crea un efecto complementario entre la información de ambos tensores: cuando los autovalores de J_1 (μ_1,...,μ_3) son grandes, los de J_2 son pequeños y viceversa. Para invertir los autovalores se utiliza la función: $g(s^2) := \frac{1}{s^2+\varepsilon^2}$ donde $\varepsilon > 0$ es una constante de estabilización que evita una posible división por cero. Cuando la estimación del término de ligadura es buena en una dirección el suavizado sólo actúa en las direcciones ortogonales. De igual modo, cuando la aportación del término de ligadura no es suficiente, el término de regularización suaviza en todas las direcciones.

Si sustituimos J_1 y J_2 en la ecuación (2.49) por las expresiones definidas en (2.50) obtenemos el *prototipo A* representado mediante autovalores y autovectores, ec. (2.51).

$$\begin{aligned} E_A(\mathbf{h}) &= \int_\Omega \mathbf{h}^\top \left(\sum_{i=1}^{3} \mu_i\, r_i r_i^\top\right) \mathbf{h}\, dx\, dy\, dz \\ &+ \alpha \int_\Omega \left(\nabla u^\top \left(\sum_{i=1}^{3} g(\mu_i)\, r_i r_i^\top\right) \nabla u + \nabla v^\top \left(\sum_{i=1}^{3} g(\mu_i)\, r_i r_i^\top\right) \nabla v\right)\, dx\, dy\, dz. \end{aligned} \tag{2.51}$$

Prototipo B:

El tensor de movimiento permite la inclusión de múltiples invarianzas dentro de la misma estructura sin hacer ningún cambio en la notación. Aprovechando esta característica podemos hacer que nuestro modelo sea flexible y fácilmente escalable. El *tensor de movimiento* acumulado también puede ser expresado mediante la descomposición en autovalores y autovectores de una forma parecida a como veíamos en la ecuación (2.47).

$$\sum_{c=1}^{N_c} \gamma_c \nabla f_c \nabla f_c^\top = \sum_{c=1}^{N_c} \gamma_c J_c = J = \sum_{i=1}^{3} \lambda_i \, r_i r_i^\top, \tag{2.52}$$

donde N_c es el número de invarianzas incluidas en el tensor, γ_c es un peso que indica la importancia de cada invarianza c. Además λ_{i_c} es un autovalor que representa μ_{i_c} o $g(\mu_{i_c})$ en el caso de J_1 y J_2, respectivamente.

El prototipo A se puede extender si empleamos tensores de movimiento acumulados. Pese a esta ampliación seguimos manteniendo una estructura sencilla y compacta. Esta extensión será etiquetada como *prototipo B*, ec. (2.53). La apariencia del modelo de energía del prototipo B es idéntica al A. La diferencia radica en la utilización de varias invarianzas en el proceso de contrucción del tensor de movimiento. En el modelo se define una serie de invarianzas con una cierta ponderación. Se contruye el tensor de movimiento a partir de esa información para posteriormente descomponerlo en autovalores y autovectores. El tensor de movimiento acumulado está representado en dicha descomposición.

$$\begin{aligned} E_B(\mathbf{h}) &= \int_\Omega \mathbf{h}^\top \left(\sum_{i=1}^{3} \mu_i \, r_i r_i^\top \right) \mathbf{h} \, dx \, dy \, dz \\ &+ \alpha \int_\Omega \left(\nabla u^\top \left(\sum_{i=1}^{3} g(\mu_i) \, r_i r_i^\top \right) \nabla u + \nabla v^\top \left(\sum_{i=1}^{3} g(\mu_i) \, r_i r_i^\top \right) \nabla v \right) \, dx \, dy \, dz. \end{aligned} \tag{2.53}$$

En los modelos variacionales tradicionales las invarianzas del término de ligadura se representan en forma cuadrática. Sin embargo, se ha demostrado que la utilización de términos no cuadráticos ofrecen resultados más robustos frente a outliers, además de preservar las discontinuidades en el flujo ([Black91]) lo que mejora notablemente la precisión de las estimaciones.

La inclusión de términos no cuadráticos como *funciones de robustificación* en el *prototipo B* aporta a nuestro modelo insensibilidad frente a los outliers. Dependiendo de cómo se combinen las funciones de robustificación y la descomposición del *tensor de movimiento* acumulado en autovalores y autovalores se abre un abanico de posibilidades pudiéndose obtener hasta cuatro variantes de nuestro modelo inicial. Si le echamos un vistazo a la ecuación (2.52) observamos que las funciones de robustificación se pueden aplicar en dos niveles: (i) a nivel de la *invarianza* o (ii) a nivel del *autovector*. Si intercambiamos el orden de la función de robustificación y del sumatorio podemos obtener dos tipos de tensores, que podrían ampliarse hasta cuatro si aplicamos la robustificación independientemente a los tensores que componen el acumulado. En la tabla 2.7 hemos representado la estructura de los tensores para cada uno de los casos.

Cada uno de estos tensores se comporta de una forma distinta dependiendo de la posición de la función de robustificación. Gracias a la experiencia aportada en otros trabajos de la literatura [Papenberg06, Bruhn05a, Weickert01a] podemos hacernos una idea de la influencia del término no cuadrático sobre cada tensor. En la ecuación (2.54) se muestra un esquema de las distintas posiciones en las que nos podemos colocar la función de robustificación.

$$\text{Tensor de movimiento} = (1) \sum_{i=1}^{3} (2) \ \ldots \tag{2.54}$$

- *A nivel de invarianza.* Cuando las invarianzas se agrupan bajo la misma función de robustificación hay que tener en cuenta que una mala estimación de alguna de ellas arrastra al resto obteniendo globalmente una estimación mala. En este caso, sólo es posible obtener una buena estimación cuando todas ellas son relativamente buenas. Para evitar este problema se suele aplicar las funciones de robustificación por separado. Aunque la aportación de una o varias invarianzas no sea lo suficientemente buena mientras haya alguna que sí será posible obtener una estimación precisa. Cuando las funciones se aplican a todas las invarianzas conjuntamente diremos que los tensores son del tipo **joint** y si se emplea a cada invarianza de forma independiente, **separate**.
- *A nivel de autovector.* La función de robustificación se puede aplicar conjuntamente sobre todos los autovectores o cada dirección por separado.

En los cuatro modelos de energía que se describen a continuación el tensor se puede expresar de varias formas. La notación que utilizaremos será la expresada en la ecuación (2.55). En ella se formula la ecuación del flujo óptico mediante una expresión compacta que incluye el flujo y la descomposición espectral del tensor de movimiento. Hemos llamado a estos cuatro modelos *prototipo C*, *D*, *E* y *F*. La función de robustificación escogida es la conocida como *Total Variation*, $\psi(s^2) := \sqrt{s^2 + \varepsilon^2}$. Esta función puede ser intercambiada por cualquier otra dependiendo de las características de la secuencia.

$$\begin{aligned}(f_x u + f_y v + f_z)^2 \ = \ (\mathbf{h}^\top \nabla f)^2 \ = \ (\mathbf{h}^\top \nabla f)(\mathbf{h}^\top \nabla f) \ = \ \mathbf{h}^\top \nabla f \nabla f^\top \mathbf{h} \ = \\ \mathbf{h}^\top J \, \mathbf{h} \ = \ \mathbf{h}^\top \left(\sum_{i=1}^{N} \mu_i \, r_i r_i^\top \right) \mathbf{h} \ = \ \sum_{i=1}^{N} \mu_i \, \mathbf{h}^\top \, r_i r_i^\top \, \mathbf{h} \ = \ \sum_{i=1}^{N} \mu_i \, (\mathbf{h}^\top r_i)^2.\end{aligned} \tag{2.55}$$

En caso de los prototipos C y D el tensor de movimiento acumulado se contruye a partir de las todas las invarianzas. Como resultado obtenemos un único tensor, el cual descomponemos en autovalores y autovectores. Las funciones de robustificación actuarán sobre todo el conjunto de autovectores o sobre cada uno independientemente. Por lo tanto, el sumatorio de las N invarianzas no se aparece en el modelo de energía porque está implícitamente definido en la descomposición espectral.

En el caso de los prototipos E y F las funciones de robustificación actúan de forma separada sobre cada invarianza. Por ello dispondremos de N tensores de movimiento distintos. En el modelo de energía si que aparece definido explicitamente el sumatorio de las invarianzas, para actuar sobre la información de cada tensor independientemente.

En la tabla 2.7 se ha representado todos los tensores definidos en los términos de ligadura de los distintos prototipos. En la primera columna indicamos el nombre del

***Tensores de movimiento* definidos en los términos de ligadura de los modelos de energía**

Prototipo	Término de ligadura	Descomposición del Tensor
A	$\mathbf{h}^\top J\,\mathbf{h}$	$\sum_{i=1}^{3} \mu_i\,(\mathbf{h}^\top r_i)^2$
B	$\mathbf{h}^\top \left(\sum_{c=1}^{N_c} \gamma_c\,J_c\right)\mathbf{h}$	$\sum_{i=1}^{3} \mu_i\,(\mathbf{h}^\top r_i)^2$
C	$\psi\left(\mathbf{h}^\top \left(\sum_{c=1}^{N_c} \gamma_c J_c\right)\mathbf{h}\right)$	$\psi\left(\sum_{i=1}^{3} \mu_i(\mathbf{h}^\top r_i)^2\right)$
D	$J^{\psi,\mathbf{h}}_{\sum_{c=1}^{N_c}}$	$\sum_{i=1}^{3} \psi\left(\mu_i(\mathbf{h}^\top r_i)^2\right)$
E	$\sum_{c=1}^{N_c} \gamma_c\,\psi\left(\mathbf{h}^\top J_c\,\mathbf{h}\right)$	$\sum_{c=1}^{N_c} \gamma_c\,\psi\left(\sum_{i=1}^{3} \mu_{i_c}(\mathbf{h}^\top r_{i_c})^2\right)$
F	$\sum_{c=1}^{N_c} \gamma_c\,J_c^{\psi,\mathbf{h}}$	$\sum_{c=1}^{N_c} \gamma_c\left(\sum_{i=1}^{3} \psi\left(\mu_{i_c}(\mathbf{h}^\top r_{i_c})^2\right)\right)$

Cuadro 2.7: Las seis variantes del *tensor de movimiento* producto de la descomposición del tensor y la combinación de las funciones de robustificación, $\psi(s^2)$. γ_c es un peso asociado a cada invarianza, μ_{i_c} y r_{i_c} son autovalores y autovectores i del tensor de movimiento de la invarianza c. El contenido de las columnas de izquierda a derecha: nombre del prototipo, definición del tensor utilizado en el término de ligadura y la descomposición espectral del tensor.

prototipo asociado al tensor. En la segunda columna se expresa la estructura del tensor: si se define una única invarianza o varias, o si las funciones de robustificación actuarán sobre todas las invarianzas juntas o de forma separada. En la tercera columna, la descomposción espectral de cada tensor.

Prototipo C:

El modelo de energía etiquetado como *prototipo C* incorpora los tensores del tipo **Joint**. La ubicación de la función de robustificación correspondería a la posición (1) de la ecuación (2.54). En el término de ligadura se permite la inclusión de múltiples invarianzas que son penalizadas conjuntamente. El término de regularización tiene un comportamiento inverso al término de ligadura y la robustificación engloba a todas las autodirecciones de todas las invarianzas aglutinas en el tensor.

$$
\begin{aligned}
E_C(\mathbf{h}) &= \int_\Omega \psi\left(\sum_{i=1}^{3} \mu_i(\mathbf{h}^\top r_i)^2\right) dx\,dy\,dz \\
&+ \alpha \int_\Omega \psi\left(\sum_{i=1}^{3} g(\mu_i)\left((\nabla u^\top r_i)^2 + (\nabla v^\top r_i)^2\right)\right) dx\,dy\,dz.
\end{aligned}
$$

Prototipo D:

El *prototipo D* incluye tensores que penaliza por separado las autodirecciones del tensor. La posición de la función de robustificación correspondería con la (2) en la ecuación (2.54). Todas las invarianzas se representan bajo un único *tensor de movimiento*. En el modelo de energía la función de robustificación actúa sobre cada autodirección de ese único tensor. La interpretación geométrica de cada autovector no es sencilla cuando el tensor incorpora más de una invarianza ya que se trata de una dirección surgida de la combinación de todas las invarianzas.

$$\begin{aligned} E_D(\mathbf{h}) &= \int_\Omega \sum_{i=1}^{3} \psi\left(\mu_i(\mathbf{h}^\top r_i)^2\right)\,dx\,dy\,dz \\ &+ \alpha \int_\Omega \sum_{i=1}^{3} \psi\left(g(\mu_i)\left((\nabla u^\top r_i)^2 + (\nabla v^\top r_i)^2\right)\right)\,dx\,dy\,dz. \end{aligned}$$

Prototipo E:

Los modelos de energía E y F disponen de mayor flexibilidad que los dos anteriores. Esto se debe a que cada invarianza es tratada por separado. El modelo de energía del *prototipo* C y D sólo tiene dos términos. En el caso de E y F existirán un término de ligadura y regularización por cada invarianza que se incluya. Esta situación no creará un cambio drástico en la apariencia del modelo gracias a la descripción compacta que nos ofrece el tensor de movimiento y su descomposición espectral, ec. (2.56).

$$\sum_{c=1}^{N_c} \gamma_c \nabla f_c \nabla f_c^\top = \sum_{c=1}^{N_c} \gamma_c J_c = \sum_{c=1}^{N_c} \gamma_c \left(\sum_{i=1}^{3} \lambda_{i_c}\, r_{i_c} r_{i_c}^\top\right). \tag{2.56}$$

El *prototipo E* incorpora los tensores del tipo **Separate**, las invarianzas son tratadas independientemente. La posición de la función de robustificación sería la (1) en la ecuación (2.54) pero requeriría la inclusión de un sumatorio que referenciara a las invarianzas. En esta ocasión, el sumatorio de las invarianzas queda fuera de la influencia de la función de robustificación creando ese tratamiento diferenciado de cada una de ellas. Sin embargo, la función engloba a todas las direcciones de cada invarianza.

$$\begin{aligned} E_E(\mathbf{h}) &= \int_\Omega \sum_{c=1}^{N_c} \gamma_c\, \psi\left(\sum_{i=1}^{3} \mu_{i_c}(\mathbf{h}^\top r_{i_c})^2\right)\,dx\,dy\,dz \\ &+ \int_\Omega \sum_{c=1}^{N_c} \alpha_c\, \psi\left(\sum_{i=1}^{3} g(\mu_{i_c})\left((\nabla u^\top r_{i_c})^2 + (\nabla v^\top r_{i_c})^2\right)\right)\,dx\,dy\,dz. \end{aligned}$$

Ahora por cada invarianza tenemos un parámetro de regularización asociado, α_c, pudiéndose controlar de una forma más precisa e independiente el proceso de difusión.

Prototipo F:

El modelo de energía de mayor grado de libertad de todos los presentados en este apartado corresponde al F. En este prototipo la función de robustificación se aplica por separado sobre cada autovector de cada invarianza. La ubicación coincide con el punto (2) de la ecuación (2.54). Está colocado a la derecha de ambos sumatorios sin ejercer ninguna influencia sobre ellos, o dicho de otra forma, actuando sólo en cada dirección de cada invarianza. De igual modo, el término de suavizado actúa inversamente proporcional sobre las direcciones penalizadas por el término de ligadura.

La estructura de cada término del modelo de energía se descompone en tres niveles: (i) en el nivel superior se sitúa el sumatorio de las invarianzas creando el tratamiento diferenciado de cada invarianza, (ii) en un nivel intermedio, el sumatorio de la descomposición del tensor, que determina las direcciones dominantes y la magnitud del tensor asociado y, por último, (iii) en el nivel inferior, la función de robustificación que se aplica únicamente a cada dirección de cada invarianza definida en el modelo de energía.

$$\begin{aligned} E_F(\mathbf{h}) &= \int_\Omega \sum_{c=1}^{N_c} \gamma_c \left(\sum_{i=1}^{3} \psi\left(\mu_{i_c} (\mathbf{h}^\top r_{i_c})^2 \right) \right) dx\, dy\, dz \\ &+ \int_\Omega \sum_{c=1}^{N_c} \alpha_c \left(\sum_{i=1}^{3} \psi\left(g(\mu_{i_c}) \left((\nabla u^\top r_{i_c})^2 + (\nabla v^\top r_{i_c})^2 \right) \right) \right) dx\, dy\, dz. \end{aligned}$$

Para finalizar la presentación de los distintos modelos de energía de este trabajo quisiera hacer mención a las principales aportaciones de este framework. (i) El uso de tensores en todo el modelo de energía y (ii) su descomposición en autovalores y autovectores. Gracias a la combinación del tensor de movimiento y las funciones de robustificación ha sido posible el desarrollo de un *framework* que establece una relación de *complementariedad* entre el término de ligadura y el de regularización. Además, los prototipos propuestos son fácilmente adaptables y escalables para cualquier tipo de invarianza.

2.5.3. Minimización de la Energía

En este apartado se describe la minimización de los seis *prototipos* presentados en el apartado anterior. Comenzaremos con la minimización del prototipo A dado su sencillez. Iremos derivando cada uno de los restantes modelos incrementando la complejidad de los sistemas de ecuaciones obtenidos. En la derivación de los prototipos C al F aparecen términos *no-lineales* que hacen que la minimización no sea trivial.

Vamos a seguir una estructura muy parecida a la del apartado anterior. Se presentan cada una de las ecuaciones de Euler–Lagrange asociadas a los modelos de energía comentando las peculiaridades de cada uno de ellos. Para facilitar la comprensión al lector comenzaremos con el prototipo más básico (A, lineal con una sola invarianza) y terminaremos con el modelo más complejo y flexible (F, múltiples invarianzas con términos no cuadráticos separados).

Prototipo A/B:

El modelo de energía es el más simple que nos podemos encontrar ya que no incluye ninguna técnica de robustificación. El término de ligadura puede incluir una o varias invarianzas y el término de regularización que incorpora el tensor inverso al definido en el de ligadura. Como vimos en el apartado anterior los modelos de energía de los prototipos A y B son idénticos. La única diferencia está en la interpretación de la descomposición espectral. Por este motivo hemos unificado la minimización de ambos prototipos.

Las ecuaciones de Euler-Lagrange asociadas a estos prototipos son

$$\begin{aligned}\sum_{i=1}^{3} \mu_i(\mathbf{h}^\top r_i)\, r_{i_1} - \alpha\, div\, (D(u,v,r_i)\nabla u) &= 0,\\ \sum_{i=1}^{3} \mu_i(\mathbf{h}^\top r_i)\, r_{i_2} - \alpha\, div\, (D(u,v,r_i)\nabla v) &= 0,\end{aligned}$$

donde

$$D(u,v,r_i) = \sum_{i=1}^{3} g(\mu_i)\left(r_i r_i^\top\right).$$

En las ecuaciones de Euler-Lagrange se aprecia el formalismo y elegancia del framework que definimos en este trabajo. Independientemente de la invarianza que se incluya en el término de ligadura ésta se expresará en función de los autovalores y autovectores del tensor de movimiento. Esta forma de representación de los tensores ofrece la máxima flexibilidad a nuestro modelo ya que una vez hecha la descomposición espectral de las invarianzas el resto del esquema permanece inalterado.

Prototipo C:

Las ecuaciones de Euler-Lagrange de los siguientes prototipos crecen en complejidad y su minimización se convierte en una tarea no trivial. Estos prototipos incluyen términos no cuadráticos. Además los tensores incluidos en estos términos permiten varias invarianzas lo que complica aún más la formulación de las ecuaciones de Euler-Lagrange.

Las ecuaciones de Euler-Lagrange asociadas a este modelo son

$$\begin{aligned}\psi'\left(\sum_{i=1}^{3} \mu_i(\mathbf{h}^\top r_i)^2\right)\sum_{i=1}^{3} \mu_i(\mathbf{h}^\top r_i)\, r_{i_1} - \alpha\, div\, (D(u,v,r_i)\nabla u) = 0,\\ \psi'\left(\sum_{i=1}^{3} \mu_i(\mathbf{h}^\top r_i)^2\right)\sum_{i=1}^{3} \mu_i(\mathbf{h}^\top r_i)\, r_{i_2} - \alpha\, div\, (D(u,v,r_i)\nabla v) = 0,\end{aligned}$$

donde

$$D(u,v,r_i) = \psi'\left(\sum_{i=1}^{3} g(\mu_i)\left((\nabla u^\top r_i)^2 + (\nabla v^\top r_i)^2\right)\right)\sum_{i=1}^{3} g(\mu_i)\left(r_i r_i^\top\right).$$

El prototipo C incorpora una función de robustificación que trata de forma conjunta a las invarianzas y a todas las direcciones de cada tensor. Cada ecuación de Euler-Lagrange se compone de la derivada de la función de robustificación que actúa como un peso sobre la estimación del término de ligadura. Ese peso se agrega a la estimación obtenida de fusionar todas las direcciones a través del sumatorio situado a la derecha de la función de robustificación. En el término de regularización ocurre exactamente lo mismo pero lo único que cambia es el valor de los autovalores, que son los inversos a los definidos en el término de ligadura.

Prototipo D:

Partiendo del modelo anterior e intercambiando la función de robustificación y el sumario que expresa las tres direcciones en la que se ha descompuesto el tensor de movimiento obtendríamos el prototipo D. Ahora cada autovector es penalizado independientemente.

Las ecuaciones de Euler-Lagrange asociadas a este modelo son

$$\begin{aligned} \sum_{i=1}^{3} \psi' \left(\mu_i (\mathbf{h}^\top r_i)^2 \right) \mu_i (\mathbf{h}^\top r_i)\, r_{i_1} - \alpha\, div\, (D(u,v,r_i)\nabla u) &= 0, \\ \sum_{i=1}^{3} \psi' \left(\mu_i (\mathbf{h}^\top r_i)^2 \right) \mu_i (\mathbf{h}^\top r_i)\, r_{i_2} - \alpha\, div\, (D(u,v,r_i)\nabla v) &= 0, \end{aligned}$$

donde

$$D(u,v,r_i) = \sum_{i=1}^{3} \psi' \left(g(\mu_i) \left((\nabla u^\top r_i)^2 + (\nabla v^\top r_i)^2 \right) \right) g(\mu_i) \left(r_i r_i^\top \right).$$

Al igual que ocurría en el modelo anterior todas las invarianzas se tratan conjuntamente. La derivada de la función de robustificación se convierte ahora en un peso asociado a cada dirección del autovector. Por un lado, nuestro modelo de energía se compone únicamente dos términos, independientemente del número de invarianzas que incorpore. Por otro lado, la complejidad de cada término no varía cuando se incrementa el número de invarianzas. Lo que aumenta es la complejidad de la estructura del tensor pero una vez hecha su descomposición espectral su manipulación es sencilla. Las direcciones expresadas por los autovectores representan una combinación de todas las invarianzas.

Prototipo E:

Los dos últimos modelos tratan de forma separada cada una de las invarianzas. Estos prototipos crean funcionales de energía que disponen por cada invarianza de un par de términos de ligadura y regularización. Ahora el sumatorio que engloba a todas las invarianzas queda situado fuera de la robustificación creando esa diferenciación entre invarianzas.

Las ecuaciones de Euler-Lagrange asociadas a este modelo son

$$\begin{aligned}
\sum_{c=1}^{N_c} \gamma_c \, \psi' \left(\sum_{i=1}^{3} \mu_{i_c} (\mathbf{h}^\top r_{i_c})^2 \right) \left(\sum_{i=1}^{3} \mu_{i_c} (\mathbf{h}^\top r_{i_c}) \, r_{i_{c1}} \right) - \sum_{c=1}^{N_c} \alpha_c \, div \left(D_c(u, v, r_i) \nabla u \right) &= 0, \\
\sum_{c=1}^{N_c} \gamma_c \, \psi' \left(\sum_{i=1}^{3} \mu_{i_c} (\mathbf{h}^\top r_{i_c})^2 \right) \left(\sum_{i=1}^{3} \mu_{i_c} (\mathbf{h}^\top r_{i_c}) \, r_{i_{c2}} \right) - \sum_{c=1}^{N_c} \alpha_c \, div \left(D_c(u, v, r_i) \nabla v \right) &= 0,
\end{aligned}$$

donde

$$D_c(u, v, r_i) = \psi' \left(\sum_{i=1}^{3} g(\mu_{i_c}) \left((\nabla u^\top r_{i_c})^2 + (\nabla v^\top r_{i_c})^2 \right) \right) \left(\sum_{i=1}^{3} g(\mu_{i_c}) \left(r_{i_c} r_{i_c}^\top \right) \right).$$

Los sumatorios de las invarianzas están situados a la izquierda de cada término generando un nuevo elemento por cada invarianza. Ahora por cada invarianza tenemos un parámetro de regularización asociado, α_c, pudiéndose controlar de una forma más precisa e independiente el proceso de difusión.

Prototipo F:

Como ya se ha comentado, el *prototipo F* es el más flexible de todos los presentados en esta sección. La función de robustificación está situada de tal forma que ofrece total grado de libertad tanto a nivel de invarianza como de autodirección. Este caso recoge dos de las ideas presentadas en los *prototipos D* y *E*. Del *D* se toma la idea de aplicar la función de robustificación sobre cada autodirección por separado. El prototipo E desacopla las estimaciones de cada invarianza en distintos términos.

Las ecuaciones de Euler-Lagrange asociadas a este prototipo son

$$\begin{aligned}
\sum_{c=1}^{N_c} \gamma_c \left(\sum_{i=1}^{3} \psi' \left(\mu_{i_c} (\mathbf{h}^\top r_{i_c})^2 \right) \mu_{i_c} (\mathbf{h}^\top r_{i_c}) \, r_{i_{c1}} \right) - \sum_{c=1}^{N_c} \alpha_c \, div \left(D_c(u, v, r_i) \nabla u \right) &= 0, \\
\sum_{c=1}^{N_c} \gamma_c \left(\sum_{i=1}^{3} \psi' \left(\mu_{i_c} (\mathbf{h}^\top r_{i_c})^2 \right) \mu_{i_c} (\mathbf{h}^\top r_{i_c}) \, r_{i_{c2}} \right) - \sum_{c=1}^{N_c} \alpha_c \, div \left(D_c(u, v, r_i) \nabla v \right) &= 0,
\end{aligned}$$

donde

$$D_c(u, v, r_i) = \sum_{i=1}^{3} \psi' \left(g(\mu_{i_c}) \left((\nabla u^\top r_{i_c})^2 + (\nabla v^\top r_{i_c})^2 \right) \right) g(\mu_{i_c}) \left(r_{i_c} r_{i_c}^\top \right).$$

Al igual que ocurría en el prototipo anterior, *E*, en el término de regularización se define un parámetro de regularización asociado cada invarianza pudiéndose controlar independientemente la cantidad de difusión aplicada por cada invarianza.

2.5.4. Esquema Numérico

En este apartado se describe el esquema numérico del prototipo F, que por su flexibilidad consideramos el caso más completo de todos los presentados. Sólo habrá pequeñas diferencias respecto al resto de prototipos.

Para hallar las componentes del flujo óptico (u, v) vamos a resolver el siguiente sistema elíptico

$$\begin{aligned}
0 &= \sum_{c=1}^{N_c} \alpha_c\, div\left(D_c(u,v,r_i)\nabla u\right) - \sum_{c=1}^{N_c} \gamma_c \left(\sum_{i=1}^{3} \psi' \left(\mu_{i_c} (\mathbf{h}^\top r_{i_c})^2 \right) \mu_{i_c} (\mathbf{h}^\top r_{i_c})\, r_{i_{c1}} \right) \\
0 &= \sum_{c=1}^{N_c} \alpha_c\, div\left(D_c(u,v,r_i)\nabla v\right) - \sum_{c=1}^{N_c} \gamma_c \left(\sum_{i=1}^{3} \psi' \left(\mu_{i_c} (\mathbf{h}^\top r_{i_c})^2 \right) \mu_{i_c} (\mathbf{h}^\top r_{i_c})\, r_{i_{c2}} \right),
\end{aligned} \quad (2.57)$$

donde

$$D_c(u,v,r_i) = \sum_{i=1}^{3} \psi' \left(g(\mu_{i_c}) \left((\nabla u^\top r_{i_c})^2 + (\nabla v^\top r_{i_c})^2 \right) \right) g(\mu_{i_c}) \left(r_{i_c} r_{i_c}^\top \right).$$

El tensor de movimiento se define en un dominio continuo donde se producen desplazamientos pequeños y la existencia de las derivadas está asegurada. Dado que nuestro modelo se basa en la descomposición espectral de dicho tensor, cuando nos encontramos ante desplazamientos largos de los objetos toda la teoría alrededor de esta técnica no se cumple. A continuación vamos a describir cómo se puede extender esta estructura para que pueda detectar los desplazamientos largos conservando la notación y propiedades comentadas hasta ahora. La descripción es similar al desarrollo hecho en [Bruhn06a].

$$(f_x u + f_y v + f_t)^2 \;=\; \mathbf{h}^\top \nabla f \nabla f^\top \mathbf{h} \;=\; \mathbf{h}^\top J\, \mathbf{h}. \quad (2.58)$$

El tensor de movimiento es una forma de representar cualquier invarianza dentro de una estructura compacta. En la ecuación (2.58) se obtiene su estructura a partir de la ecuación de restricción del flujo óptico. Para manejar los desplazamientos largos será necesario deducir el tensor a partir de la ecuación no lineal, ec. (2.44). Esto es posible hacerlo fácilmente en el momento de la discretización. Siguiendo la misma estrategia que [Bruhn06a] se ha combinado un enfoque multipiramidal con una discretización *fixed point iteration*.

Para eliminar los términos nolineales presentes en el sistema de ecuaciones se ha optado por la linealización del término de ligadura y por dividir el flujo óptico a estimar en la siguiente iteración, $\mathbf{h}^{k+1} = (u^{k+1},\ v^{k+1}, 1)^\top$, como la suma del flujo calculado en la iteración actual, $\mathbf{h}^k = (u^k,\ v^k, 1)^\top$, más un pequeño incremento, $d\mathbf{h}^k = (du^k, dv^k, 1)^\top$. Fijando como referencia el flujo actual queremos estimar su valor en la siguiente iteración hallando un flujo desconocido que necesitamos para ir desde $(u^k,\ v^k)$ a $(u^{k+1},\ v^{k+1})$. Dicho de otra manera

$$\begin{aligned}
u^{k+1} &= u^k + du^k, \\
v^{k+1} &= v^k + dv^k, \\
\mathbf{h}^{k+1} &= \mathbf{h}^k + d\mathbf{h}^k.
\end{aligned}$$

Si expresamos la ecuación del flujo óptico para largos desplazamientos, ec. (2.44), siguiendo la notación utilizada en nuestro esquema numérico tendríamos

$$f(\mathbf{x}+\mathbf{h}^k)-f(\mathbf{x})=0.$$

Para eliminar la no-linealidad de la expresión anterior y establecer una formulación en función de las nuevas incógnitas del flujo óptico (du^k, dv^k) linealizamos $f(\mathbf{x}+\mathbf{h}^{k+1})$ obteniendo

$$f(\mathbf{x}+\mathbf{h}^{k+1})=f(\mathbf{x}+\mathbf{h}^k)+f_x(\mathbf{x}+\mathbf{h}^k)\,du^k+f_y(\mathbf{x}+\mathbf{h}^k)\,dv^k. \tag{2.59}$$

Si sustituimos la ecuación (2.59) en la discretización de cualquier invarianza podremos obtener una expresión cuadrática en la que tengamos un *tensor de movimiento* que sea capaz de reflejar los largos desplazamientos de la escena.

$$\begin{aligned}(f(\mathbf{x}+\mathbf{h}^{k+1})-f(\mathbf{x}))^2 &\approx\\ (f(\mathbf{x}+\mathbf{h}^k)+f_x(\mathbf{x}+\mathbf{h}^k)\,du^k+f_y(\mathbf{x}+\mathbf{h}^k)\,dv^k-f(\mathbf{x}))^2 &=\\ (f_x(\mathbf{x}+\mathbf{h}^k)\,du^k+f_y(\mathbf{x}+\mathbf{h}^k)\,dv^k+(f(\mathbf{x}+\mathbf{h}^k)-f(\mathbf{x})))^2 &=\\ (d\mathbf{h}^k\widetilde{\nabla}f(\mathbf{x}+\mathbf{h}^k))^2 &=\\ (d\mathbf{h}^k)^\top\underbrace{\widetilde{\nabla}f(\mathbf{x}+\mathbf{h}^k)\widetilde{\nabla}f^\top(\mathbf{x}+\mathbf{h}^k)}_{J(\mathbf{x}+\mathbf{h}^k)}(d\mathbf{h}^k)&,\end{aligned} \tag{2.60}$$

$d\mathbf{h}^k=(du^k,dv^k,1)^\top$ representa el incremento espaciotemporal del movimiento y $\widetilde{\nabla}$ es una variante del operador gradiente donde la última componente es una aproximación de la derivada temporal mediante una diferencia.

De acuerdo a la ecuación (2.60) el tensor de movimiento se puede utilizar para representar cualquier invarianza que tenga en cuenta los desplazamientos largos. La descomposición espectral de este tensor seguirá conservando esta característica. Por lo tanto, nuestro modelo está preparado para detectar cualquier movimiento presente en la imagen. Sin embargo, el modelado de los desplazamientos largos no se hace explícitamente en el funcional de energía. Esto se pospone al momento de la discretización. Además, nuestro esquema numérico será embebido en un enfoque multipiramidal.

Si fusionamos las ideas descritas en las ecuaciones (2.55) y (2.60), el esquema numérico quedaría expresado en función del incremento del flujo $d\mathbf{h}^k$, ec. (2.61).

$$\begin{aligned}d\mathbf{h}^{k\top}\left(\widetilde{\nabla}f(\mathbf{x}+\mathbf{h}^k)\widetilde{\nabla}f^\top(\mathbf{x}+\mathbf{h}^k)\right)d\mathbf{h}^k \;=\; d\mathbf{h}^{k\top}\left(J(\mathbf{x}+\mathbf{h}^k)\right)d\mathbf{h}^k \;=\\ d\mathbf{h}^{k\top}\left(\sum_{i=1}^{3}\mu_i\,r_ir_i^\top\right)d\mathbf{h}^k \;=\; \sum_{i=1}^{3}\mu_i\,d\mathbf{h}^{k\top}\,r_ir_i^\top\,d\mathbf{h}^k \;=\; \sum_{i=1}^{3}\mu_i\,(d\mathbf{h}^{k\top}r_i)^2.\end{aligned} \tag{2.61}$$

Esta notación es fácilmente extensible para múltiples invarianzas.

Método de Resolución del Sistema de Ecuaciones

La solución del sistema de ecuaciones de Euler-Lagrange descrito en el apartado anterior, ec. (2.57), se obtiene mediante un método numérico iterativo. *Successive*

Overrelaxation (*SOR*, [Young71]) se trata de un método iterativo basado en una extrapolación de los resultados de Gauss-Seidel. Es un compromiso entre simplicidad y eficiencia. Aunque el coste por iteración sea algo superior que el método de Gauss-Seidel necesita menos iteraciones para converger lo que lo convierte en una buena opción.

La estructura general de método SOR se expresa en la siguiente ecuación

$$x_i^{k+1} = (1-\beta)x_i^k + \beta \underbrace{(A_{i,i})^{-1} \left(b_i - \sum_{j=1}^{i-1} A_{i,j}\, x_j^{k+1} - \sum_{j=i+1}^{2N} A_{i,j}\, x_j^k \right)}_{\text{Iteración de Gauss-Seidel}}, \tag{2.62}$$

donde x puede ser una de las componentes del flujo óptico (u, v) y $i = 1, ..., 2N$; A es una matriz que almacena los coeficientes del sistema de ecuaciones a resolver. Este método numérico será utilizado en la implementación del nuestro método.

Discretización del Sistema de Ecuaciones de Euler–Lagrange

Una vez definido el modelo de energía y minimizarlo el siguiente paso consiste en la resolución del sistema de ecuaciones de Euler–Lagrange, ec. (2.57). Para ello debemos resolver un sistema de ecuaciones no-lineal del tipo

$$A^m(x^m) = b^m, \tag{2.63}$$

donde $A^m(x^m)$ es un operador no-lineal y b^m es la parte derecha del sistema. $A^m(x^m)$ se puede descomponer en

$$A^m(x^m) = B^m(x^m)x^m + c^m(x^m), \tag{2.64}$$

donde $B^m(x^m)$ y $c^m(x^m)$ son operadores no-lineales pero para cada valor de x^m, $B^m(x^m)$ es una matriz simétrica y definida positiva; $c^m(x^m)$ es un vector.

Los métodos *Lagged-Diffusivity* ([Kacur68, Fucik73, Chan99]) se basan en la idea de la resolución de un sistema de ecuaciones nolineal, como es el caso de (2.63), a través de su descomposición en un conjunto de problemas lineales. La resolución de estos problemas lineales se puede llevar a cabo con técnicas estándar, como son el método de SOR o Gauss–Seidel.

La descomposición hecha en la ecuación (2.64) permite aprovechar las propiedades de la matriz $B^m(x^m)$ y el vector $c^m(x^m)$ para convertir el sistema inicial, ec. (2.63), en lineal, gracias a la evaluación de los operadores $B^m(x^m)$ y $c^m(x^m)$ en el instante anterior k

$$\begin{aligned} x^{m,k+1} &= (B^m(x^{m,k}))^{-1}(b^m - c^m(x^{m,k})), \\ B^m(x^{m,k})x^{m,k+1} &= (b^m - c^m(x^{m,k})). \end{aligned} \tag{2.65}$$

Para la resolución del sistema de ecuaciones descrito en la ecuación (2.57) se ha optado por la combinación de las técnicas de *fixed point iterations* y el método *Lagged-Diffusivity.* A continuación, vamos a desarrollar el esquema numérico resultante. Para ello, vamos a descomponer el proceso en tres niveles:

1. *Enfoque Multipiramidal.*

   Para detectar los desplazamientos largos nuestro método se apoya en un enfoque multipiramidal. En este enfoque se crean un cierto número de escalas $s_1, s_2, \ldots, s_n$, donde cada una representa imágenes de distinto tamaño. Durante el cambio de escala, los valores del flujo deben ser actualizados y adaptados a la nueva escala.

   Para hallar las componentes del flujo óptico $\{u^s, v^s\}$ tenemos que resolver en cada escala el sistema de ecuaciones descrito en la ecuación (2.57). El índice s refleja la escala actual. El tensor de movimiento se recalcula y descompone espectralmente en cada escala antes de volver a resolver el sistema de ecuaciones no lineal.

2. *Lagged-Diffusivity.*

   En cada escala se resuelve el sistema de ecuaciones de Euler-Lagrange. Este sistema se convierte en lineal gracias al *Lagged-Diffusivity*. El sistema se resuelve un cierto número de iteraciones. El índice l representa el punto de iteración en el que está fijado cada elemento del sistema, p.e. $(u^{s,l}, v^{s,l})$. En este nivel se resuelve un problema convexo nolineal a través de un conjunto de problemas convexos lineales manteniendo la nolinealidad en el término de ligadura y regularización. En cada iteración se recalculan las derivadas de las funciones de robustificación y el tensor de difusión. En nuestro caso, el sistema resultante es

$$\begin{aligned} 0 &= \sum_{c=1}^{N_c} \alpha_c \, div \left(D_c(u^{s,l} + du^{s,l}, v^{s,l} + dv^{s,l}, r_i^s) \nabla (u^{s,l} + du^{s,l}) \right) \\ &\quad - \sum_{c=1}^{N_c} \gamma_c \sum_{i=1}^{3} \psi' \left(\mu_{i_c}^s ((d\mathbf{h}^{s,l})^\top r_{i_c}^s)^2 \right) \mu_{i_c}^s ((d\mathbf{h}^{s,l})^\top r_{i_c}^s) \, r_{i_{c1}}^s, \\ 0 &= \sum_{c=1}^{N_c} \alpha_c \, div \left(D_c(u^{s,l} + du^{s,l}, v^{s,l} + dv^{s,l}, r_i^s) \nabla (v^{s,l} + dv^{s,l}) \right) \\ &\quad - \sum_{c=1}^{N_c} \gamma_c \sum_{i=1}^{3} \psi' \left(\mu_{i_c}^s ((d\mathbf{h}^{s,l})^\top r_{i_c}^s)^2 \right) \mu_{i_c}^s ((d\mathbf{h}^{s,l})^\top r_{i_c}^s) \, r_{i_{c2}}^s, \end{aligned} \tag{2.66}$$

   donde $D_c(u^{s,l} + du^{s,l}, v^{s,l} + dv^{s,l}, r_i^s)$ es

$$D_c = \sum_{i=1}^{3} \psi' \left(g(\mu_{i_c}^s) \left((\nabla (u^{s,l} + du^{s,l})^\top r_{i_c}^s)^2 + (\nabla (v^{s,l} + dv^{s,l})^\top r_{i_c}^s)^2 \right) \right) g(\mu_{i_c}^s) \left(r_{i_c}^s r_{i_c}^{s\top} \right).$$

3. *Resolución del sistema de ecuaciones.*

   En el nivel inferior se resuelve el sistema de ecuaciones lineales calculando los nuevos valores del incremento del flujo. Para ello, se utiliza el método numérico SOR. En este nivel es necesario la inclusión de un nuevo índice k que plasme la iteración del *SOR* en el que nos encontramos. Para simplificar la notación de la discretización vamos a introducir una serie de expresiones

$$\begin{aligned} (d\mathbf{h}^{s,l,k})^\top r_{i_c}^s &= (r_{i1_c}^s du^{s,l,k} + r_{i2_c}^s dv^{s,l,k} + r_{i3_c}^s) \\ \psi_D'^{s,l} &= \psi' \left(\mu_{i_c}^s ((d\mathbf{h}^{s,l})^\top r_i^s)^2 \right) \end{aligned}$$

La matriz D, se define en cada píxel i como:

$$D_i = \begin{pmatrix} a_i & b_i & c_i \\ b_i & d_i & e_i \\ c_i & e_i & f_i \end{pmatrix}.$$

La discretización del término de la divergencia en cada píxel i se realiza de la siguiente manera:

$$\begin{aligned} div(D_i \nabla u) = \begin{pmatrix} a_i\partial_x u + b_i\partial_y u + c_i\partial_t u \\ b_i\partial_x u + d_i\partial_y u + e_i\partial_t u \\ c_i\partial_x u + e_i\partial_y u + f_i\partial_t u \end{pmatrix} &= \partial_x (a_i\partial_x u) + \partial_x (b_i\partial_y u) + \partial_x (c_i\partial_t u) \\ &+\partial_y (b_i\partial_x u) + \partial_y (d_i\partial_y u) + \partial_y (e_i\partial_t u) \\ &+\partial_t (c_i\partial_x u) + \partial_t (e_i\partial_y u) + \partial_t (f_i\partial_t u), \end{aligned}$$

N_i^* se define como los vecinos alrededor del píxel i. Utilizando un esquema de diferencias estándar la divergencia se puede expresar como:

$$div(D_i \nabla u_i) = \sum_{j \in N_i^*} w_j\, u_j + w_i u_i,$$

para los coeficientes adecuados w_j. Para el caso de $div(D_i \nabla v)$ se procede de la misma manera.

En el sistema de ecuaciones (2.57) utilizamos un esquema semi-implícito en el término de ligadura e implícito en el de regularización. El esquema iterativo a implementar surge de la inclusión de las ecuaciones (2.66) en el esquema SOR, ec. (2.62).

$$\begin{aligned} du_i^{s,l,k+1} &= (1-\beta)\, du_i^{s,l,k} \\ &+\beta \frac{\sum_{c=1}^{N_c} \alpha_c \left(\sum_{j \in N_i^*} w_j \,(u_j^{s,l} + du_j^{s,l,k}) \right)}{\sum_{c=1}^{N_c} \alpha_c \left(\sum_{j \in N_i^*} w_j \right) + \sum_{c=1}^{N_c} \gamma_c \left(\sum_{i=1}^{3} \psi_D'^{s,l}\, \mu_i^s\, r_{i1}^{s2} \right)} \\ &-\beta \frac{\sum_{c=1}^{N_c} \alpha_c \left(\sum_{j \in N_i^*} w_j\, u_j^{s,l} \right) + \sum_{c=1}^{N_c} \gamma_c \left(\sum_{i=1}^{3} \psi_D'^{s,l}\, \mu_i^s\, (r_{i2}^s dv_i^{s,l,k} + r_{i3}^s)\, r_{i1}^s \right)}{\sum_{c=1}^{N_c} \alpha_c \left(\sum_{j \in N_i^*} w_j \right) + \sum_{c=1}^{N_c} \gamma_c \left(\sum_{i=1}^{3} \psi_D'^{s,l}\, \mu_i^s\, r_{i1}^{s2} \right)} \end{aligned}$$

en la dirección vertical

$$\begin{aligned} dv_i^{s,l,k+1} &= (1-\beta)\, dv_i^{s,l,k} \\ &+\beta\, \frac{\sum_{c=1}^{N_c} \alpha_c \left(\sum_{j \in N_i^*} w_j \,(v_j^{s,l} + dv_j^{s,l,k}) \right)}{\sum_{c=1}^{N_c} \alpha_c \left(\sum_{j \in N_i^*} w_j \right) + \sum_{c=1}^{N_c} \gamma_c \left(\sum_{i=1}^{3} \psi_D'^{s,l}\, \mu_i^s\, r_{i2}^{s2} \right)} \\ &-\beta\, \frac{\sum_{c=1}^{N_c} \alpha_c \left(\sum_{j \in N_i^*} w_j\, u_j^{s,l} \right) + \sum_{c=1}^{N_c} \gamma_c \left(\sum_{i=1}^{3} \psi_D'^{s,l}\, \mu_i^s\, (r_{i1}^s du_i^{s,l,k} + r_{i3}^s)\, r_{i2}^s \right)}{\sum_{c=1}^{N_c} \alpha_c \left(\sum_{j \in N_i^*} w_j \right) + \sum_{c=1}^{N_c} \gamma_c \left(\sum_{i=1}^{3} \psi_D'^{s,l}\, \mu_i^s\, r_{i2}^{s2} \right)} \end{aligned}$$

donde β es un *parámetro de relajación* cuyo valor está en el intervalo $(0, 2)$. El subíndice i representa un píxel en la imagen. Los valores más comunes para β suelen oscilar entre 1.5 y 1.99.

2.5.5. Resultados Experimentales

En este apartado se presentan los resultados obtenidos con los distintos prototipos descritos en este trabajo. Se ha establecido una comparación entre todos ellos y los mejores métodos propuestos en la literatura. Nuestros prototipos se basan en un modelo anisotrópico que utiliza información espaciotemporal y la descomposición espectral de los tensores del término de ligadura y de regularización. La inclusión de términos no cuadráticos da robustez a nuestro método frente al ruido.

En estos experimentos se quiere demostrar que la elegante descomposición de nuestro método no sólo se aprecia desde punto de vista teórico sino también práctico. En los tests se ha utilizado una secuencia sintética, *Yosemite con nubes*, y una real, *Rheinhafen.* Antes de comentar los resultados conviene recordar los distintos parámetros que intervienen en nuestro método.

- El parámetro de regularización, α, determina el peso del término de suavizado. Cuando el valor de este parámetro es alto genera campos de desplazamiento más suaves, el término de suavizado gana importancia respecto al de ligadura. Los prototipos del A al D tienen un sólo parámetro de regularización. Para los prototipos E y F existirá un parámetro por cada invarianza definida en término de ligadura, α_c para $c = 1, ..., N$ siendo N el número de invarianzas.

- El ratio de descenso $\eta \in (0, 1)$ para el cálculo de las escalas. En [Weickert04] se recomienda que el factor de escala η esté dentro del rango $[0{,}5, 0{,}95]$. En todos los experimentos se ha fijado su valor a $\eta = 0{,}9$.

- γ_c es un peso que se utiliza en el término de ligadura para ponderar la importancia de cada invarianza c.

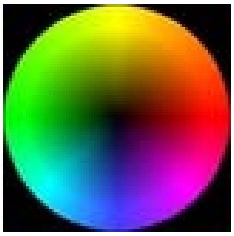


Figura 2.22: Círculo que representa en color la dirección y magnitud del desplazamiento de cada píxel.

- β es un *parámetro de relajación* cuyo valor está en el intervalo $(0, 2)$. En los experimentos el valor de este parámetro ha sido 1,99.
- $\varepsilon_d, \varepsilon_s, \varepsilon_g$. Se trata de unas constantes de estabilización que evitan una posible división por cero en las derivadas de las funciones de robustificación. A estos parámetros se le asignan un valor próximo a cero, p.e. 0,001.
- σ_s, σ_t. Antes de la estimación del flujo óptico se suele aplicar un proceso de *suavizado* a las imágenes de entrada. De esta forma se atenúa o elimina el ruido presente en las imágenes. Este proceso consiste en una convolución con una gaussiana de desviación estándar σ. El parámetro σ_s y σ_t expresan la desviación estándar aplicada en el dominio espacial y temporal, respectivamente. En las pruebas se realizadas los valores han sido $\sigma_s = 1{,}0$ y $\sigma_t = 0{,}001$.
- *Outer iterations*. Número de veces que se resuelve el sistema de ecuaciones en cada escala del enfoque multipiramidal. Su valor en los experimentos se ha fijado a 15 por escala.
- *Inner iterations*. Número de iteraciones utilizadas en el método numérico SOR. Su valor en los experimentos ha sido de 5.
- Dimensiones del píxel (h_x, h_y, h_z). Se trata de un parámetro utilizado en la discretización del método cuyo valor normalmente es de 1,0.

Para poder obtener resultados cuantitativos que nos permitan comparar la precisión de las estimaciones de nuestros prototipos se ha escogido como métricas el Average Angular Error (AAE, ec. 1.15), ampliamente utilizado en la literatura [Barron94]. En los resultados visuales los desplazamientos de los píxeles se han representado en color. La dirección se representa con un color y la magnitud del desplazamiento variando la intensidad de ese color. En la figura 2.22 se muestra un círculo de color utilizado para representar los desplazamientos de los píxeles.

Secuencia sintética. *Yosemite con Nubes.*

La secuencia de *Yosemite con nubes*, figura 2.23, ha sido ampliamente utilizada para evaluar la calidad de los métodos de flujo óptico. Combina varios tipos de movimientos, como puede ser traslacional y divergente. En el centro de la escena el movimiento es

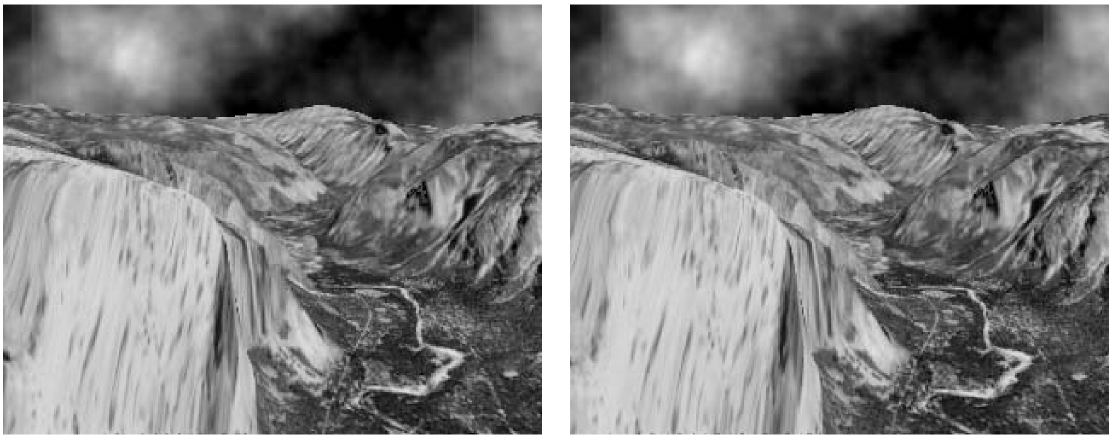


Figura 2.23: Frames 8 y 9 de la secuencia de *Yosemite con nubes.*

próximo a cero mientras que en el cielo es horizontal, constante y existe cambios en la iluminación de los píxeles a lo largo de la secuencia. En otras zonas de la imagen, cerca de los bordes de la imagen, se observan distintas velocidades y aceleraciones. En el apartado 1.4.1 de esta tesis podemos conocer más detalles acerca de esta secuencia.

Dado los cambios de iluminación presentes en el cielo de Yosemite es imprescindible la inclusión en el modelo de energía de una invarianza que sea insensible a estos cambios (p.e. gradiente constante). En el trabajo de Papenberg et al. [Papenberg06] se investigó la influencia de distintas invarianzas en el término de ligadura. De este trabajo se desprende que para la secuencia de Yosemite la combinación de las invarianzas suposición lambertiana y gradiente constante ofrece mejores resultados que la utilización de otras invarianzas de orden superior.

Los experimentos realizados ponen de manifiesto, a nivel cuantitativo, la ganancia ofrecida por las ligeras modificaciones realizadas a cada prototipo. Dado que el campo de desplazamiento en la secuencia de Yosemite es muy suave no se producirán diferencias significativas entre las distintas variantes del método. Pese a que los prototipos E y F guardan cierta similitud respecto a C y D, el tratamiento diferenciado de cada invarianza produce una notable reducción en el error de las estimaciones. En la figura 2.24 se puede observar el *ground truth* y las distintas estimaciones obtenidas con cada uno de los prototipos presentados. Las diferencias más importantes entre las distintas soluciones están en las discontinuidades entre el cielo y las montañas. En el caso de la solución del prototipo A se aprecia como la ausencia de la segunda invarianza y las funciones de robustificación impiden una correcta estimación del flujo óptico, sobre todo el desplazamiento de los píxeles del cielo. En la tabla 2.9 se muestran los parámetros utilizados por cada prototipo. En la tabla 2.8 disponemos de resultados cuantitativos en el que se puede comparar el AAE de cada prototipo respecto a los más importantes métodos de la literatura. En esta tabla también se indica las distintas técnicas incluidas en el modelo de energía de los métodos. Si analizamos los resultados de cada uno de los prototipos observamos que la descomposición espectral de los tensores y el guiado de la difusión a través de las direcciones dominantes del flujo mejora sensiblemente las estimaciones respecto a otros métodos que incluyen técnicas muy parecidas. El *prototipo A*, que sólo incorpora una invarianza y términos cuadráticos, es capaz reducir el error frente a otros métodos como pueden ser [Mémin98b, Alvarez00] y empeora ligeramente frente a [Bruhn05c], un método que incorpora técnicas de robustificación. La inclusión en el término de ligadura de una segunda invarianza que sea insensible a los cambios de iluminación del cielo permite mejorar ligeramente la solución del *prototipo A*.

Cuando se incorporan términos no cuadráticos al *prototipo B* se observa una reducción significativa en los errores de las estimaciones. En los resultados cuantitativos se observa cómo los prototipos que penalizan conjuntamente las autodirecciones (C,E) mejoran ligeramente respecto a los que los hacen por separado (D,F). La estructura del modelo de energía de Papenberg *et al.* [Papenberg06] es muy parecida a los prototipos C y D. Todos ellos son espaciotemporales, incluyen términos no cuadráticos, incorporan invarianzas de orden superior y utilizan un enfoque multipiramidal. La diferencia más notable está en el término de regularización. Papenberg incluye una variante de Horn–Schunck y nuestro método un regularizador que sigue la filosofía del operador de Nagel–Enkelmann. Otro trabajo de referencia en la literatura es el de Bruhn–Weickert [Bruhn05a]. Es una modificación del método de [Papenberg06] en el que se aplica robustificación separada a las distintas varianzas definidas en el término de ligadura. En este sentido se asemeja a los prototipos E y F. El prototipo D mejora un 3,93 % respecto a [Papenberg06] e incluso se equipara a [Bruhn05a], el mejor método puramente variacional existente en la literatura. En el caso del prototipo C la reducción respecto a los métodos anteriores es de un 5,62 % y 1,75 %, respectivamente. La separación de la robustificación de las invarianzas consigue reducir en más de un 8 % respecto a la robustificación conjunta. La separación del proceso de difusión por invarianzas en el término de regularización ha permitido tal mejora ya que en esencia se trata de la única diferencia de relevancia si lo comparamos con [Bruhn05a]. La reducción de los prototipos E y F respecto a [Bruhn05a] es de 14,04 % y 9,94 %, respectivamente. De esta forma, nuestro modelo E se sitúa en la segunda posición en la tabla comparativa de los mejores métodos de la literatura (tabla 2.8). Es capaz de mejorar la estimación hecha por Amiaz *et al.* [Amiaz07], un método que incluye técnicas de segmentación.

El guiado de la difusión mediante las autodirecciones se percibe como una de las formas más precisa de las vistas hasta ahora en la literatura.

Secuencia real. *Rheinhafen.*

La secuencia real utilizada en los experimentos es la conocida como *Rheinhafen* (fig. 2.25). Se trata de una secuencia de tráfico en la que se observan varios vehículos moviéndose a distinta velocidad y en distintas direcciones. Próxima a la cámara hay una furgoneta que circula a gran velocidad. Debido al efecto de entrelazado de la cámara los bordes del objeto aparece en forma de dientes de sierra. Al fondo de la secuencia, se observa unos coches cambiando de dirección en un cruce, el desplazamiento de ellos es pequeño. En el apartado 1.4.1 podemos encontrar más detalles acerca de esta secuencia.

Las imágenes reales se ven influenciadas por una serie de perturbaciones, como son cambios de iluminación y ruido, que dificultan la estimación del flujo óptico. Para atenuar el efecto de estas perturbaciones nuestros modelos de energía incluyen técnicas de

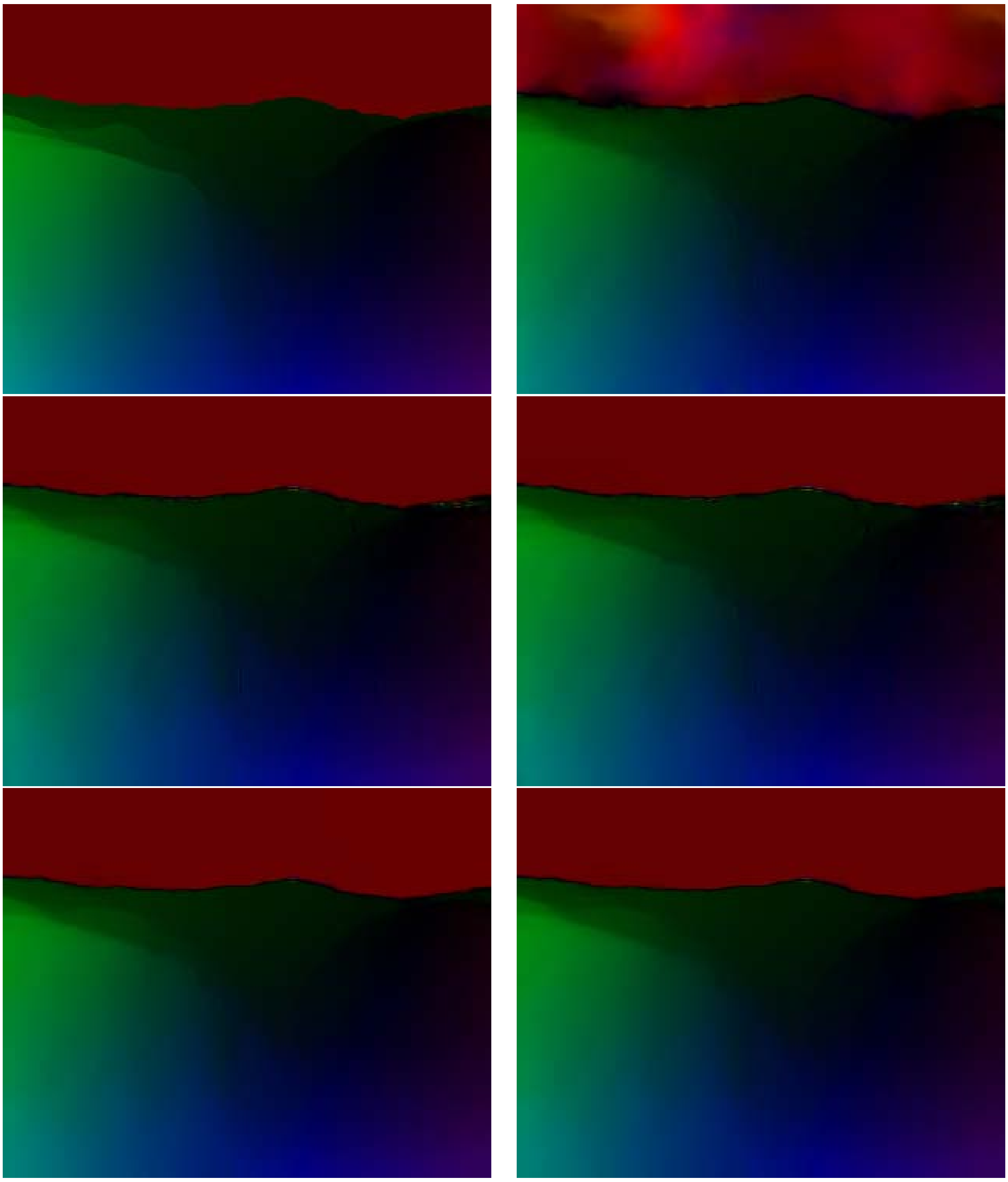


Figura 2.24: Los distintos campos de desplazamiento obtenidos con los prototipos A al F. En la parte superior, el *ground truth* y el campo de desplazamiento estimado por el prototipo A. En el medio, la estimación del prototipo C y D. En la parte inferior, la solución ofrecida por E y F.

Cuadro 2.8: Comparativa entre las estimaciones de los prototipos A al F y los métodos más relevantes de la literatura. HC= invarianzas con derivadas de orden superior. NQ-D= término de ligadura no cuadrático. NQ-S= término de suavizado no cuadrático o algún tipo de estrategia que preserve las discontinuidades. 3D= término de suavizado espaciotemporal. MS= enfoque multiescala o multipiramidal. S= Segmentación. AAE= Average Angular Error.

Yosemite con nubes

Método	HC	NQ-D	NQ-S	3D	MS	S	AAE
Anandan [Barron94]	-	-	-	-	-	-	13,36°
Nagel [Barron94]	-	-	✓	-	-	-	10,22°
Horn-Schunck, mod. [Barron94]	-	-	-	-	-	-	9,78°
Uras *et al.* [Barron94]	-	-	-	-	-	-	8,94°
Bruhn *et al.* linear [Bruhn05c]	-	-	-	✓	-	-	6,24°
Álvarez *et al.* [Alvarez00]	-	-	✓	-	✓	-	5,53°
Mémin–Pérez *et al.* [Mémin98b]	-	✓	✓	-	✓	-	4,69°
Prototipo A	-	-	-	✓	✓	-	**4,44°**
Prototipo B	✓	-	-	✓	✓	-	**4,20°**
Bruhn *et al.* nonlinear [Bruhn05c]	-	✓	✓	✓	✓	-	4,17°
Papenberg *et al.* [Papenberg06]	✓	✓	✓	✓	-	-	2,78°
Brox *et al.* [Weickert04]	✓	✓	✓	✓	✓	-	1,94°
Papenberg *et al.* [Papenberg06]	✓	✓	✓	✓	✓	-	1,78°
Bruhn–Weickert [Bruhn05a]	✓	✓	✓	✓	✓	-	1,71°
Prototipo D	✓	✓	✓	✓	✓	-	**1,71°**
Prototipo C	✓	✓	✓	✓	✓	-	**1,68°**
Amiaz-Kiryati (2D-LD)[Amiaz06]	✓	✓	✓	-	✓	✓	1,64°
Prototipo F	✓	✓	✓	✓	✓	-	**1,54°**
Amiaz *et al.* [Amiaz07]	✓	✓	✓	-	✓	✓	1,48°
Prototipo E	✓	✓	✓	✓	✓	-	**1,47°**
Brox *et al.* [Brox06]	✓	✓	✓	✓	✓	✓	1,22°

Cuadro 2.9: Lista de parámetros utilizados en los prototipos A al F para la secuencia de *Yosemite con nubes*. N_c = Número de invarianzas definidas en el modelo de energía. γ_1 = peso asignado a la primera invarianza en el término de ligadura. γ_2 = peso asignado a la segunda invarianza en el término de ligadura. α_1 = peso asignado al primer término de suavizado. α_2 = peso asignado al segundo término de suavizado.

Yosemite con Nubes

$\varepsilon_d = 10^{-3}$, $\varepsilon_s = 10^{-3}$, $\varepsilon_g = 10^{-3}$,
$\eta = 0{,}9$, $\beta = 1{,}99$, $\sigma_s = 1{,}0$, $\sigma_t = 10^{-3}$

Método	N_c	γ_1	γ_2	α_1	α_2
Prototipo A	1	1	-	0.001	-
Prototipo B	2	1	0.05	10	-
Prototipo C	2	1	200	200	-
Prototipo D	2	1	200	150	-
Prototipo E	2	1	10	0.1	0.05
Prototipo F	2	1	100	1	0.5

Figura 2.25: Dos frames no consecutivos de la secuencia de *Rheinhafen*.

robustificación. En las secuencias reales los cambios de iluminación es un fenómeno muy frecuente. Si estos cambios son muy grandes los métodos que asumen constancia en la intensidad de los píxeles no obtendrán buenos resultados. Si por el contrario, se tratan de ligeras variaciones las diferencias entre métodos que incluyan términos de orden superior y los que no serán tan significativas. El efecto de entrelazado es otro de los problemas al que nos enfrentamos cuando disponemos de secuencias reales. Este efecto crea bordes en los objetos en forma de dientes de sierra. La velocidad de los objetos determina la alteración de los bordes por el efecto de entrelazado: mayor velocidad, bordes más aserrados. Los regularizadores anisotrópicos deberían comportarse mejor que los isotrópicos en los bordes de los objetos.

En la figura 2.26 se muestran las distintas estimaciones obtenidas por cada uno de los prototipos presentados en esta sección. En la tabla 2.10 se muestra los parámetros utilizados por los distintos prototipos. El prototipo A incluye la suposición lambertiana

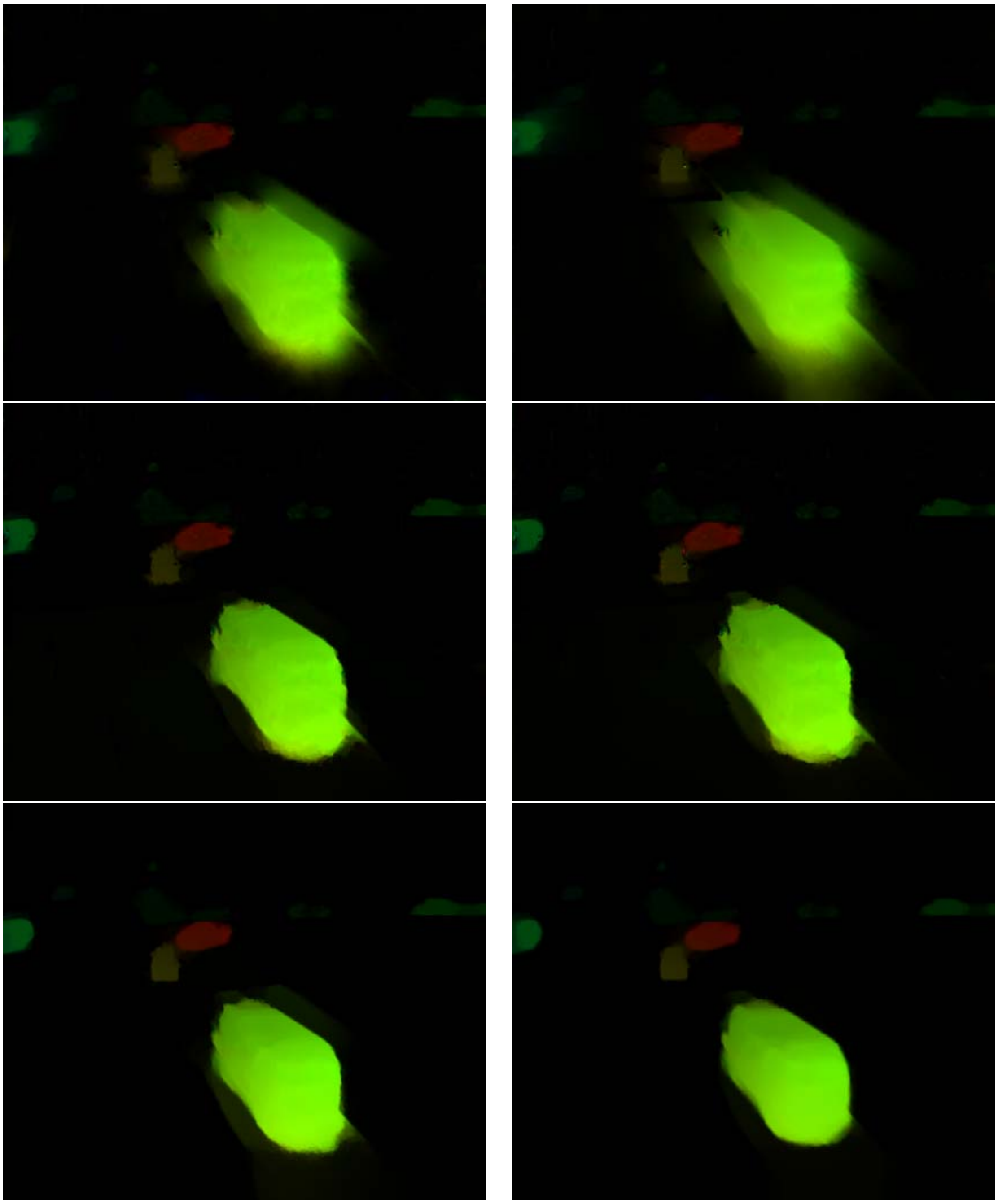


Figura 2.26: Los distintos campos de desplazamiento obtenidos con los prototipos A al F. En la parte superior, los campos de desplazamiento estimados por el prototipo A y B. En el medio, la estimación del prototipo C y D. En la parte inferior, la solución ofrecida por E y F.

Cuadro 2.10: Lista de parámetros utilizados en los prototipos A al F para la secuencia de *Rheinhafen*. N_c = Número de invarianzas definidas en el modelo de energía. γ_1 = peso asignado a la primera invarianza en el término de ligadura. γ_2 = peso asignado a la segunda invarianza en el término de ligadura. α_1 = peso asignado al primer término de suavizado. α_2 = peso asignado al segundo término de suavizado.

Rheinhafen

$\varepsilon_d = 10^{-3}$, $\varepsilon_s = 10^{-3}$, $\varepsilon_g = 10^{-3}$,
$\eta = 0{,}9$, $\beta = 1{,}99$, $\sigma_s = 1{,}0$, $\sigma_t = 10^{-3}$

Método	N_c	γ_1	γ_2	α_1	α_2
Prototipo A	1	1	-	0.00045	-
Prototipo B	2	1	0.001	50	-
Prototipo C	2	1	0.001	0.01	-
Prototipo D	2	1	0.001	0.01	-
Prototipo E	2	1	0.05	0.03	0.001
Prototipo F	2	1	1	0.04	0.01

en el término de ligadura. La solución obtenida por este prototipo se ve afectada por los cambios de iluminación y el ruido presente en la secuencia. Dado que en *Rheinhafen* el ruido influye más que los cambios de iluminación la inclusión de una segunda invarianza (prototipo B) no consigue mejorar la estimación del prototipo A. El resto de los prototipos incluyen funciones de robustificación lo que da al método una mayor insensibilidad frente al ruido. No existe visualmente grandes diferencias entre las versiones **joint** C y D o entre **separate** E y F. Si se percibe la influencia de la robustificación junta o separada sobre las invarianzas definidas en el modelo. Cabe destacar que el movimiento de los coches situados al fondo ha sido detectado con bastante nitidez.

2.6. Conclusiones

En este capítulo se han presentado tres trabajos relativos a la estimación del flujo óptico con información procedente de varias imágenes. En los métodos descritos se ha querido plasmar distintos enfoques a la hora de estimar el flujo óptico en una secuencia de imágenes: (1) un método espacial que combina información multicanal, (2) un método espacial al que se le ha añadido un término de regularización temporal y, (3) un método espaciotemporal que comparte información entre el término de ligadura y suavizado. A continuación, se describe las conclusiones más importantes extraídas de estos tres trabajos.

2.6.1. Método Variacional Multicanal

La incorporación en los satélites de múltiples sensores ha multiplicado la cantidad de información disponible haciendo necesario la automatización de las tareas de almacenamiento y procesado. La meteorología y la climatología son campos de estudio que más hacen uso de las imágenes satélites. Estas imágenes facilitan el análisis y seguimiento de los distintos fenómenos que ocurren en la atmósfera. Tradicionalmente se han utilizado técnicas de correlación para el análisis de los fenómenos atmosféricos debido a su simplicidad y buenos resultados. Sin embargo, son computacionalmente costosos y no ofrecen soluciones densas. Los métodos variacionales han demostrado un enorme éxito a la hora de resolver problemas de visión por ordenador, como es la estimación del flujo óptico.

En este trabajo se ha propuesto un método variacional para la detección del desplazamiento de las estructuras nubosas en las imágenes satélites multicanal. Se trata de una extensión multicanal de un conocido método para la estimación del flujo óptico ([Alvarez00]). El modelo de energía se compone de un término de ligadura incluye la suposición lambertiana, aplicada a cada canal por separado, mientras que en el término de suavizado se utiliza el operador de Nagel-Enkelmann ([Nagel86]) con algunas mejoras. Dado que el tensor de difusión de Nagel-Enkelmann usa el gradiente de la imagen para determinar la cantidad de difusión se han desarrollado dos estrategias para calcular dicho gradiente a partir de la información multicanal disponible. La combinación de la información de todos los canales permite mejorar la estimación del movimiento de las nubes.

Dado que las secuencias satélites son datos reales y no disponemos de los desplazamientos de los píxeles se han creado dos secuencias sintéticas usando un modelo de movimiento realista. Con este tipo de secuencias es posible realizar una comparación cuantitativa de las estimaciones ofrecidas por los métodos. Los resultados experimentales demuestran que el método multicanal mejora significativamente la precisión de las estimaciones si la comparamos con su versión monocanal. Los canales visibles e infrarrojo son los que más información aportan acerca del desplazamiento de las nubes.

2.6.2. Método Variacional con Regularización Temporal no Continua

En este trabajo se ha descrito un método variacional que incorpora un novedoso término de regularización exclusivamente temporal. Algunos autores han propuesto regularizadores que tratan de forma conjunta la información espacial y temporal. Estos regularizadores imponen una continuidad en el modelo obligando a que los desplazamientos de los objetos sean pequeños tanto en la dirección espacial como en la temporal. En situaciones reales un modelo discontinuo, como el nuestro, se ajusta mejor.

El funcional de energía propuesto se trata de una modificación del modelo de Nagel-Enkelmann [Nagel86] al que se le ha añadido un término de regularización temporal. La división del término de regularización en dos, uno espacial y otro temporal, pretende evitar los inconvenientes derivados del acoplamiento de la información espaciotemporal. Se han diseñado dos variantes del término temporal: (1) *Temporal* cuya estimación se basa en la información del flujo del frame siguiente y, (2) *Bi–Temporal* que utiliza la información del flujo en ambos sentidos. Un inconveniente que tiene la versión *Temporal* está en su fuerte dependencia respecto al último flujo de la secuencia, de forma que si la estimación del último flujo es mala ese error se propaga a los flujos anteriores. Por el contrario, en la versión *Bi–Temporal* los flujos se compensan entre sus vecinos sin verse influenciados por uno en concreto. Este hecho se aprecia en los experimentos donde el método bidireccional mejora sustancialmente las estimaciones respecto a su homólogo unidireccional.

En los resultados experimentales se observa que la incorporación del nuevo término de regularización temporal aporta estabilidad a las soluciones del método espacial. El método propuesto en esta tesis se trata de unos de los primeros trabajos que considera regularizadores temporales para largos desplazamientos.

2.6.3. Método Variacional basado en el Análisis Espectral

En el último trabajo de este tema se propone un *framework* para la estimación del flujo óptico usando secuencias de imágenes. Este *framework* permite la definición de un modelo de energía genérico que gracias a la combinación de algunas de las técnicas más innovadoras propuestas en la literatura es fácilmente adaptable y escalable para cualquier tipo de invarianzas.

El *tensor de movimiento* es una técnica que permite representar cualquier invarianza en una estructura compacta. La descomposición espectral de este tensor permite la fácil identificación de las direcciones dominantes del flujo. Esta información resulta muy práctica a la hora de guiar de una forma más precisa el proceso de difusión. Al mismo tiempo, se puede establecer una complementariedad entre el término de ligadura y el de regularización. La utilización de funciones de robustificación aportan insensibilidad al método frente a la presencia de *outliers*. La combinación de esta técnica junto con la descomposición del *tensor de movimiento* permite la creación de cuatro modelos de energía distintos, donde las autodirecciones son penalizadas conjuntamente o por separado en función de la colocación de la robustificación.

Los experimentos realizados con las distintas variantes de nuestro modelo nos han permitido comprobar que la elegante descomposición de nuestro método no sólo es apreciable desde punto de vista teórico sino también práctico. Para ello, se ha utilizado la secuencia sintética de *Yosemite con nubes* para comparar los modelos con los mejores métodos propuestos en la literatura. Los resultados son bastante buenos mejorando las estimaciones de otros métodos que incorporan técnicas parecidas. Incluso uno de los modelos obtiene el segundo mejor registro de la literatura. Recordemos que este método no incluye ninguna técnica de segmentación. En la secuencia real de *Rheinhafen* también se pone de manifiesto la calidad de las soluciones obtenidas con los distintos prototipos.

Las principales aportaciones realizadas en este trabajo son: (i) el uso de tensores en toda el modelo de energía y (ii) la descomposición espectral de los mismos. Gracias a la combinación del tensor de movimiento y las funciones de robustificación ha sido posible el desarrollo de un *framework* que establece una relación de *complementariedad* entre el término de ligadura y el de regularización.

Capítulo 3

Estimación del Mapa de Disparidad en Pares Estéreo

3.1. Introducción

En la visión estereoscópica disponemos de dos vistas de la misma escena en el mismo instante de tiempo. La estimación del mapa de disparidad consiste en el cálculo del desplazamiento de los píxeles de una vista a la otra. Este problema se reduce a una búsqueda de correspondencias entre dos imágenes por lo que en cierto modo tiene muchas similitudes con la estimación del flujo óptico. La componente temporal que tienen las secuencias de imágenes frente a los pares estéreos supone un impedimento para la adaptación de muchas de las ideas surgidas en el problema del flujo óptico. Sin embargo, la visión estereoscópica dispone de una herramienta muy útil como es la *geometría epipolar*. La geometría epipolar permite acotar el área de búsqueda de las correspondencias. Toda esta teoría se basa en la existencia de un sistema de cámaras perfectamente calibradas.

Como parte fundamental de esta tesis se ha dedicado un capítulo completo a la descripción de los distintos métodos desarrollados para la estimación del mapa de disparidad. En este capítulo se describen algunas de las técnicas más exitosas surgidas en los últimos años y cómo la combinación de ellas permite aumentar la precisión de los estimaciones. Antes de entrar en detalle en cada uno de ellos se comentarán las contribuciones hechas a la literatura.

3.1.1. Contribuciones de este Capítulo

Las contribuciones en el ámbito científico que se han hecho en este capítulo son las siguientes:

- **Método para la Estimación del Mapa de Disparidad utilizando una Secuencia de Pares Estéreo:**

  La primera aportación novedosa supone la creación de un método que combina la información estéreo Alvarez/Deriche/Sánchez/Weickert (2002) [Alvarez02b] con la del flujo óptico Alvarez/Weickert/Sánchez (2000) [Alvarez00] aumentando la robustez de las estimaciones al fusionar dentro del mismo modelo de energía las ideas de dos métodos de gran precisión.

La estimación del mapa de disparidad se apoya en la geometría epipolar para establecer las correspondencias. Esas mismas correspondencias se pueden obtener mediante el cálculo del flujo óptico a lo largo de la secuencia de ambas cámaras. Por lo tanto, en una secuencia de pares estéreo disponemos de hasta cuatro vistas de un punto 3D en dos instantes de tiempo distintos. Este método pretende aprovechar las ventajas de los métodos de flujo óptico y estéreo para mejorar la estimación de los mapas de disparidad. Al disponer de más información permite que la búsqueda de las correspondencias sea más robusta.

- **Combinación de un Método Variacional Espacial y uno de Graph–cuts para la Estimación del Mapa de Disparidad:**

  La segunda aportación novedosa consiste en la combinación de un método variacional Alvarez/Deriche/Sánchez/Weickert (2002) [Alvarez02b] y una técnica de graph–cuts [Kolmogorov01, Boykov04]. El método de graph–cuts se ha utilizado para obtener una aproximación inicial del mapa de disparidad que el método variacional se encargaría de refinar. Normalmente, la técnica más empleada para estimar la inicialización es una basada en la correlación a ventanas.

  La solución en los métodos variacionales se alcanza mediante la minimización de un funcional de energía. Para garantizar y acelerar la obtención de una solución próxima al mínimo global se suele *ayudar* al algoritmo mediante una inicialización. O sea, se le ofrece una solución relativamente próxima al mínimo global de forma que el algoritmo converja rápidamente. Es muy importante que la inicialización esté cerca del mínimo porque en el caso contrario la convergencia no está garantizada.

  En los últimos años se han propuesto métodos graph–cuts para la estimación del mapa de disparidad. Con esta técnica se obtienen muy buenos resultados pero en precisión entera. Una buena inicialización no tiene porqué tener una gran precisión sólo se requiere que esté próxima a la solución final. Por este motivo, la combinación de un método variacional [Alvarez02b] y otro de graph–cuts [Kolmogorov01, Boykov04] parece más ventajosa que un variacional con una técnica de correlación.

3.2. Método para la Estimación del Mapa de Disparidad utilizando una Secuencia de Pares Estéreo

En este apartado se describe un nuevo método para la recuperación de la geometría 3D a partir de una secuencia de pares estéreo. Disponemos de dos cámaras captando una escena y que están fijadas sobre un soporte rígido orientadas en una cierta dirección. En esa escena pueden haber objetos estáticos o dinámicos. Suponemos que el sistema de cámaras está débilmente calibrado y que las cámaras están situadas en posición frontoparalela (ver apartado 1.2.1). Debido a las perturbaciones presentes en las imágenes captadas por un sistema de visión estereoscópico no es posible hacer un perfecto seguimiento del desplazamiento de los píxeles de una vista a la otra. Una forma de minimizar el efecto que producen algunas perturbaciones, como puede ser las oclusiones o el ruido, consistiría en transferir información desde los flujos ópticos de cada cámara a los mapas de disparidad. De esta forma podríamos incluir información temporal en las estimaciones del flujo estéreo y mejorar su precisión.

El método presentado en este apartado combina la información estéreo con la del flujo óptico para mejorar la estimación de los mapas de disparidad. La estimación del flujo óptico en ambas cámaras se utiliza como restricción en el cálculo de los mapas de disparidad haciendo que la búsqueda de las correspondencias sea más robusta. El modelo de energía permite detectar los desplazamientos largos de los objetos. Para ello, se utiliza un enfoque multipiramidal en el que la solución de la escala inferior se utilizará como aproximación inicial de la superior.

Este método es una continuación de los trabajos presentados en Alvarez/Weickert/Sánchez (2000) [Alvarez00] y Alvarez/Deriche/Sánchez/Weickert (2002) [Alvarez02b]. Los métodos de ambos trabajos se basan en una técnica de minimización de energía que ofrecen resultados precisos y densos. La organización de este trabajo es la siguiente: en primer lugar, se presenta la notación utilizada para la descripción del modelo de energía. Una vez descrita la notación, se comenta con todo lujo de detalle el modelo de energía y el esquema numérico utilizado. Finalmente, se muestra los experimentos realizados con secuencias sintéticas y reales que pondrán de manifiesto la aportación de la información del flujo óptico en la estimación del mapa de disparidad.

3.2.1. Notación del Flujo Óptico y Mapas de Disparidad

En este apartado se describe la notación utilizada en el modelo de energía para combinar la información del flujo óptico y del mapa de disparidad.

El flujo óptico representa el desplazamiento de cada píxel de una imagen a la siguiente. Este desplazamiento se suele representar como la función $\mathbf{h}(\mathbf{x}) = (u(\mathbf{x}), v(\mathbf{x}))^t$. En nuestro caso, existen dos flujos ópticos a estimar: el de la imagen derecha e izquierda. Para poder identificar correctamente a cada flujo es necesario introducir índices que hagan referencia a cada cámara. Un índice i para especificar un determinado flujo de la secuencia; y otro l, r para indicar si ese flujo pertenece a la cámara izquierda o derecha, respectivamente. Dado que la secuencia de entrada esta compuesta por N pares estéreos, existirán $N-1$

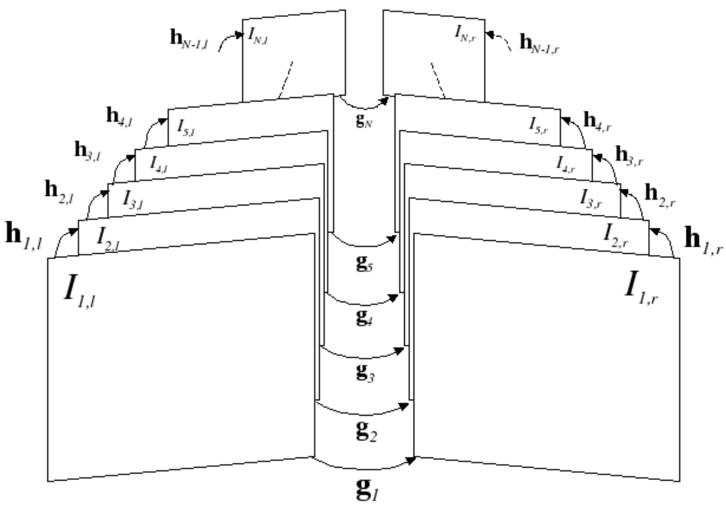


Figura 3.1: $I_{i,l}(\mathbf{x})$ representa a las imágenes tomadas por la cámara izquierda y $I_{i,r}(\mathbf{x})$ a la de la derecha. $\mathbf{h}_{i,l}(\mathbf{x})$ y $\mathbf{h}_{i,r}(\mathbf{x})$ son los flujos ópticos de ambas cámaras. $\mathbf{g}_i(\mathbf{x})$ es el mapa de disparidad desde la cámara izquierda a la derecha del par estéreo i.

flujos ópticos y N flujos estéreos. En la figura 3.1, se muestra un gráfico que resume los distintos campos de desplazamiento estimados: $\mathbf{h}_{i,l}(\mathbf{x})$ para la cámara izquierda, $\mathbf{h}_{i,r}(\mathbf{x})$ para la derecha y $\mathbf{g}_i(\mathbf{x})$ es el mapa de disparidad.

Siguiendo esta notación el flujo óptico se puede expresar como

$$\begin{aligned} I_{i,l}(\mathbf{x}) &\simeq I_{i+1,l}\left(\mathbf{x}+\mathbf{h}_{i,l}(\mathbf{x})\right) \\ I_{i,r}(\mathbf{x}) &\simeq I_{i+1,r}\left(\mathbf{x}+\mathbf{h}_{i,r}(\mathbf{x})\right) \end{aligned}$$

para la imagen izquierda y derecha respectivamente. $I_{i,l}(\mathbf{x})$ es la imagen tomada en el instante i por la cámara izquierda y $\mathbf{x} = (x, y)$ es la coordenada de un píxel en la imagen. $I_{i,r}(\mathbf{x})$ es la imagen de la cámara derecha.

Como se comentó en el apartado 1.2.1 la búsqueda de las correspondencias entre los píxeles de un par estéreo se puede simplificar si disponemos de un calibrado de las cámaras correcto. Apoyándonos en la geometría epipolar es posible reducir la zona de búsqueda a la *línea epipolar*. En la figura 3.1 el flujo estéreo se define como $\mathbf{g}_i(\mathbf{x}) = (u_{i,s}(\mathbf{x}), v_{i,s}(\mathbf{x}))^t$. Para diferenciar los campos de desplazamiento del flujo óptico de los del estéreo se ha incluido la etiqueta s. De forma que el flujo estéreo se expresará en función de dos índices (i, s), donde s indica que se trata del mapa de disparidad e i, la posición dentro de la secuencia. El flujo estéreo depende de un escalar $\lambda(\mathbf{x})$ que expresa el desplazamiento sobre la recta epipolar. En función de esto, el campo de desplazamiento se puede expresar como

$$\begin{aligned} u_{i,s}(\mathbf{x}) &= \frac{-\lambda_i(\mathbf{x})b(\mathbf{x})}{\sqrt{a^2(\mathbf{x})+b^2(\mathbf{x})}} - \frac{a(\mathbf{x})x+b(\mathbf{x})y+c(\mathbf{x})}{a^2(\mathbf{x})+b^2(\mathbf{x})}a(\mathbf{x}) \\ v_{i,s}(\mathbf{x}) &= \frac{\lambda_i(\mathbf{x})a(\mathbf{x})}{\sqrt{a^2(\mathbf{x})+b^2(\mathbf{x})}} - \frac{a(\mathbf{x})x+b(\mathbf{x})y+c(\mathbf{x})}{a^2(\mathbf{x})+b^2(\mathbf{x})}b(\mathbf{x}) \end{aligned} \tag{3.1}$$

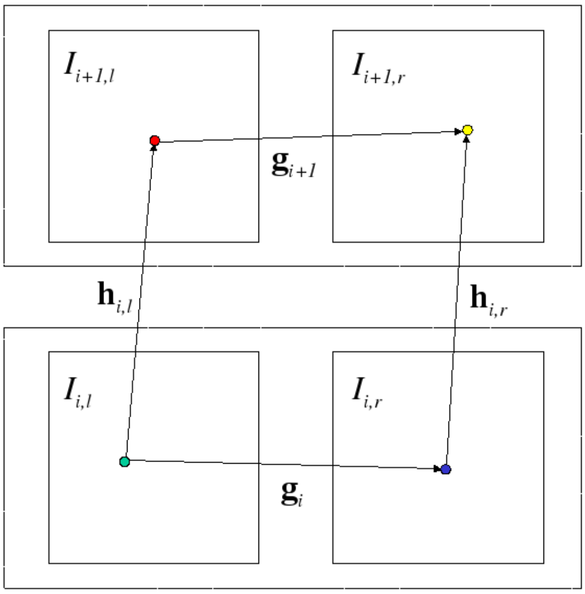


Figura 3.2: Las cuatro vistas del mismo punto 3D están interconectadas a través del flujo óptico y estéreo de ambas cámaras. En un caso ideal, si partimos desde el punto situado en la cámara izquierda en el instante i y seguimos cualquiera de los dos caminos: (1) a través del flujo óptico ($\mathbf{h}_{i,l}$) y luego con el flujo estéreo ($\mathbf{g}_{i+1}$) o (2) a través del flujo estéreo ($\mathbf{g}_i$) y posteriormente el flujo óptico ($\mathbf{h}_{i,r}$), debemos llegar a la correspondencia de la cámara derecha en el instante $i+1$. Esta restricción temporal se puede expresar como $\mathbf{h}_{i,l}(\mathbf{x}) + \mathbf{g}_{i+1}(\mathbf{x} + \mathbf{h}_{i,l}(\mathbf{x})) \equiv \mathbf{g}_i(\mathbf{x}) + \mathbf{h}_{i,r}(\mathbf{x} + \mathbf{g}_i(\mathbf{x}))$.

En el trabajo de [Alvarez02b] se pueden encontrar más detalles acerca de la parametrización descrita en la ecuación (3.1).

Para establecer las correspondencias entre los píxeles en un par estéreo es necesario escoger alguna propiedad invariante de las imágenes. La ecuación de restricción del flujo estéreo puede expresarse como

$$I_{i,l}(\mathbf{x}) \simeq I_{i,r}\left(\mathbf{x} + \mathbf{g}_i(\mathbf{x})\right) \tag{3.2}$$

Si nos fijamos en la figura 3.2 se puede ver fácilmente que para cada dos frames consecutivos de la secuencia es posible establecer una relación entre las proyecciones de los puntos que pertenecen al mismo punto 3D mediante el flujo óptico y el estéreo

$$\mathbf{h}_{i,l}(\mathbf{x}) + \mathbf{g}_{i+1}(\mathbf{x} + \mathbf{h}_{i,l}(\mathbf{x})) \equiv \mathbf{g}_i(\mathbf{x}) + \mathbf{h}_{i,r}(\mathbf{x} + \mathbf{g}_i(\mathbf{x})) \tag{3.3}$$

Esta relación se puede expresar matemáticamente mediante la ecuación (3.3). En una situación ideal si partimos de un punto y nos desplazamos de acuerdo al flujo estéreo y óptico, se alcanzaría el mismo destino independientemente si se ha tomado primero el flujo óptico o estéreo.

3.2.2. Modelo de Energía

El objetivo de este método es la estimación de los flujos ópticos y estéreos de una secuencia de pares de imágenes, introduciendo la restricción de coherencia entre ambos flujos. Esta restricción se compone de tres incógnitas: (i) el flujo estéreo $\mathbf{g}_i(\mathbf{x})$ donde $i = 1, ..., N$ indica un par estéreo de la secuencia, (ii) el flujo óptico izquierdo $\mathbf{h}_{j,l}(\mathbf{x}) = (u_{j,l}(\mathbf{x}), v_{j,l}(\mathbf{x}))^t$ y (iii) el flujo óptico derecho $\mathbf{h}_{j,r}(\mathbf{x}) = (u_{j,r}(\mathbf{x}), v_{j,r}(\mathbf{x}))^t$, donde $j = 1, ..., N-1$.

La estructura del modelo variacional descrito en esta sección viene definido por la ecuación:

$$E(\lambda_i, \mathbf{h}_{j,l}, \mathbf{h}_{j,r}) = E_s(\lambda_i) + E_{o_l}(\mathbf{h}_{i,l}) + E_{o_r}(\mathbf{h}_{i,r}) + E_c(\lambda_i, \mathbf{h}_{i,l}, \mathbf{h}_{i,r}) \tag{3.4}$$

donde $E_s(\lambda_i)$ es la energía correspondiente a la estimación del flujo estéreo, $E_{o_l}(\mathbf{h}_{i,l})$ y $E_{o_r}(\mathbf{h}_{i,r})$ las energías asociadas al flujo óptico izquierdo y derecho respectivamente, y $E_c(\lambda_i, \mathbf{h}_{i,l}, \mathbf{h}_{i,r})$ es la energía que relaciona las tres incógnitas según lo especificado en la ecuación (3.3).

A continuación vamos a describir con más detalles cada uno de los términos del modelo de la ecuación (3.4). El funcional de energía para la estimación del mapa de disparidad es

$$\begin{aligned} E_s(\lambda_i) \quad &= \sum_{i=1}^{N} \left(\int_\Omega (I_{i,l}(\mathbf{x}) - I_{i,r}(\mathbf{x} + \mathbf{g}(\lambda_i(\mathbf{x}))))^2 \, dx \; dy \right. \\ &\left. + \alpha \int_\Omega \nabla\lambda_i(\mathbf{x})^t \mathbf{D}(\nabla I_{i,l}) \nabla\lambda_i(\mathbf{x}) dx \; dy \right) \end{aligned} \tag{3.5}$$

El modelo de energía definido en la ecuación (3.5) es similar al propuesto en [Alvarez02b]. La diferencia más importante está en la utilización de un conjunto de pares estereoscópicos lo que implica tener en cuenta todos los frames de la secuencia. El primer término es el de ligadura y expresa la ecuación de restricción del flujo estéreo con la suposición lambertiana, ec. (3.2). El segundo término es el de regularización propuesto por Nagel–Enkelmann [Nagel86] y permite obtener una solución única y suave preservando los contornos de los objetos.

Las ecuaciones de las energías del flujo óptico son

$$\begin{aligned} E_{o_l}(\mathbf{h}_{i,l}) \quad &= \sum_{i=1}^{N-1} \left(\int_\Omega (I_{i,l}(\mathbf{x}) - I_{i+1,l}(\mathbf{x} + \mathbf{h}_{i,l}(\mathbf{x})))^2 \, dx \; dy \right. \\ &\left. + \alpha \int_\Omega \operatorname{trace}\left((\nabla \mathbf{h}_{i,l}(\mathbf{x}))^t \, \mathbf{D}(\nabla I_{i,l}) \nabla \mathbf{h}_{i,l}(\mathbf{x}) \right) dx \; dy \right) \end{aligned}$$

para la cámara izquierda y

$$\begin{aligned} E_{o_r}(\mathbf{h}_{i,r}) \quad &= \sum_{i=1}^{N-1} \left(\int_\Omega (I_{i,r}(\mathbf{x}) - I_{i+1,r}(\mathbf{x} + \mathbf{h}_{i,r}(\mathbf{x})))^2 \, dx \; dy \right. \\ &\left. + \alpha \int_\Omega \operatorname{trace}\left((\nabla \mathbf{h}_{i,r}(\mathbf{x}))^t \, \mathbf{D}(\nabla I_{i,r}) \nabla \mathbf{h}_{i,r}(\mathbf{x}) \right) dx \; dy \right) \end{aligned}$$

para la derecha. Estas dos energías son similares a la descrita en [Alvarez00] salvo el sumatorio que aglutina la información de todos los frames de la secuencia. Cabe destacar que el índice de ese sumatorio tiene un rango de valores de $[1, ..., N-1]$ ya que sólo existen $N-1$ flujos ópticos. α es una constante que define el peso del término de regularización y $\mathbf{D}(\nabla I)$ es el tensor de difusión utilizado en el operador de Nagel–Enkelmann [Nagel86].

$$\mathbf{D}(\nabla I) = \frac{1}{\|\nabla I\|^2 + 2\zeta^2} \left(\nabla I_\perp \nabla I_\perp^t + \zeta^2 \mathbf{Id}\right) \tag{3.6}$$

$\nabla I_\perp$ es el vector ortogonal al gradiente $\nabla I_\perp = (-I_y \quad I_x)^t$. Se trata de una matriz de proyección en la dirección perpendicular al gradiente y, por lo tanto, en la dirección del contorno de los objetos.

El último término del modelo definido en la ecuación (3.4) representa la relación existente entre los flujos ópticos y estéreos. Siguiendo la notación descrita en la figura 3.2, este término quedaría expresado como

$$\begin{aligned} E_c(\lambda_i, \mathbf{h}_{i,l}, \mathbf{h}_{i,r}) \quad &= \beta \sum_{i=1}^{N-1} \int_\Omega \Phi \left(\|\mathbf{h}_{i,l}(\mathbf{x}) + \mathbf{g}_{i+1}(\mathbf{x} + \mathbf{h}_{i,l}(\mathbf{x})) \right. \\ &\qquad \left. -\mathbf{g}_i(\mathbf{x}) - \mathbf{h}_{i,r}(\mathbf{x} + \mathbf{g}_i(\mathbf{x}))\|^2\right) dx\ dy \end{aligned} \tag{3.7}$$

La ecuación (3.7) impone una restricción para los frames consecutivos. En esta ecuación β es una constante que pondera la importancia de este término en el conjunto global de la energía y $\Phi\,(.)$ una función de robustificación, que podría ser

$$\Phi\,(s) = \rho \left(1 - e^{\frac{-s}{\rho}}\right)$$

En una situación ideal en la que los flujos ópticos y estéreos estén perfectamente estimados, la restricción expresada en la ecuación (3.3) es cierta para todos los píxeles. Sin embargo, en situaciones reales esto rara vez ocurre. La presencia de pequeñas desviaciones en las estimaciones hará que pasemos del idílico escenario mostrado en la figura 3.2 al de la figura 3.3. Dado que no concuerdan ambos caminos será necesario la inclusión de un elemento que imponga o atraiga los puntos en correspondencia independientemente de la ruta seguida. Por este motivo, se ha utilizado la función $\Phi\,(.)$ cuyo objetivo es la atracción de los puntos en correspondencia en la imagen $I_{i+1,r}$.

3.2.3. Minimización de la Energía

La solución de los métodos variacionales se obtiene a través de la minimización del modelo de energía. La solución debe satisfacer las ecuaciones de Euler–Lagrange y se obtiene tras la resolución de este sistema de ecuaciones en derivadas parciales (PDE) que se resuelve mediante la técnica del descenso por gradiente. En nuestro caso tenemos un sistema de ecuaciones se compone de tres incógnitas $\lambda_i, h_{i,l}$ y $h_{i,r}$.

El mapa de disparidad se puede expresar mediante una función escalar λ_i, ec. (3.1). Para simplificar esta ecuación vamos a introducir una serie de variables en la notación empleada en las ecuaciones de la minimización de la energía.

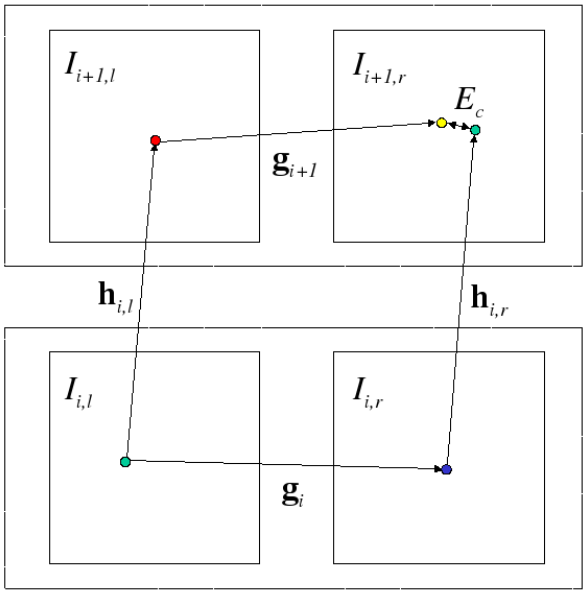


Figura 3.3: Efecto de la función $\Phi(.)$ en el término E_c de la energía, ec. (3.7). En un escenario ideal, las estimaciones del flujo óptico y estéreo serían exactas y, los caminos descritos en la figura 3.2 terminarían en el mismo punto en la imagen $I_{i+1,r}$. En un escenario real, se producirán pequeñas desviaciones. Para que se cumpla la ecuación (3.7) será necesario la inclusión de una función $\Phi(.)$ que atraiga los puntos en correspondencia.

$$A := \frac{a(\mathbf{x})}{\sqrt{a^2(\mathbf{x})+b^2(\mathbf{x})}} \qquad B := \frac{-b(\mathbf{x})}{\sqrt{a^2(\mathbf{x})+b^2(\mathbf{x})}}$$

$$Ca := -\frac{a(\mathbf{x})x+b(\mathbf{x})y+c(\mathbf{x})}{a^2(\mathbf{x})+b^2(\mathbf{x})}a(\mathbf{x}) \qquad Cb := -\frac{a(\mathbf{x})x+b(\mathbf{x})y+c(\mathbf{x})}{a^2(\mathbf{x})+b^2(\mathbf{x})}b(\mathbf{x})$$

Por lo tanto, el flujo estéreo queda simplificado como

$$\begin{aligned} u_{i,s} &= B\lambda_i + Ca \\ v_{i,s} &= A\lambda_i + Cb, \end{aligned}$$

y la ecuación del descenso del gradiente para el flujo estéreo queda como

$$\begin{aligned} \frac{\partial \lambda_i}{\partial t} &= \alpha \operatorname{div}\left(\mathbf{D}\left(\nabla I_{i,l}\right) \nabla \lambda_i\right) + \left(I_{i,l} - I_{i,r}^{\lambda_i}\right)\left(AI_{i,r,y}^{\lambda_i} + BI_{i,r,x}^{\lambda_i}\right) \\ &+\beta \Phi_i' \begin{pmatrix} u_{i,l} + u_{i+1,s}^{\mathbf{h}_{i,l}} - B\lambda_i - Ca - u_{i,r}^{\mathbf{g}_i} \\ v_{i,l} + v_{i+1,s}^{\mathbf{h}_{i,l}} - A\lambda_i - Cb - v_{i,r}^{\mathbf{g}_i} \end{pmatrix}^t \begin{pmatrix} 1 + u_{i,r,x}^{\mathbf{g}_i} & u_{i,r,y}^{\mathbf{g}_i} \\ v_{i,r,x}^{\mathbf{g}_i} & 1 + v_{i,r,y}^{\mathbf{g}_i} \end{pmatrix} \begin{pmatrix} B \\ A \end{pmatrix} \end{aligned}$$

Las ecuaciones del descenso del gradiente para el flujo óptico izquierdo serían

$$\begin{aligned}
\frac{\partial u_{i,l}}{\partial t} &= \alpha \operatorname{div}\left(\mathbf{D}\left(\nabla I_{i,l}\right) \nabla u_{i,l}\right) + \left(I_{i,l} - I_{i+1,l}^{\mathbf{h}_{i,l}}\right) I_{i+1,l,x}^{\mathbf{h}_{i,l}} \\
&\quad -\beta \Phi_i' \begin{pmatrix} u_{i,l} + u_{i+1,s}^{\mathbf{h}_{i,l}} - B\lambda_i - Ca - u_{i,r}^{\mathbf{g}_i} \\ v_{i,l} + v_{i+1,s}^{\mathbf{h}_{i,l}} - A\lambda_i - Cb - v_{i,r}^{\mathbf{g}_i} \end{pmatrix}^t \begin{pmatrix} 1 + u_{i+1,s,x}^{\mathbf{h}_{i,l}} \\ u_{i+1,s,y}^{\mathbf{h}_{i,l}} \end{pmatrix} \\
\frac{\partial v_{i,l}}{\partial t} &= \alpha \operatorname{div}\left(\mathbf{D}\left(\nabla I_{i,l}\right) \nabla v_{i,l}\right) + \left(I_{i,l} - I_{i+1,l}^{\mathbf{h}_{i,l}}\right) I_{i+1,l,y}^{\mathbf{h}_{i,l}} \\
&\quad -\beta \Phi_i' \begin{pmatrix} u_{i,l} + u_{i+1,s}^{\mathbf{h}_{i,l}} - B\lambda_i - Ca - u_{i,r}^{\mathbf{g}_i} \\ v_{i,l} + v_{i+1,s}^{\mathbf{h}_{i,l}} - A\lambda_i - Cb - v_{i,r}^{\mathbf{g}_i} \end{pmatrix}^t \begin{pmatrix} v_{i+1,s,x}^{\mathbf{h}_{i,l}} \\ 1 + v_{i+1,s,y}^{\mathbf{h}_{i,l}} \end{pmatrix}
\end{aligned}$$

Para el derecho sería

$$\begin{aligned}
\frac{\partial u_{i,r}}{\partial t} &= \alpha \operatorname{div}\left(\mathbf{D}\left(\nabla I_{i,r}\right) \nabla u_{i,r}\right) + \left(I_{i,r} - I_{i+1,r}^{\mathbf{h}_{i,r}})\right) I_{i+1,r,x}^{\mathbf{h}_{i,r}} \\
\frac{\partial v_{i,r}}{\partial t} &= \alpha \operatorname{div}\left(\mathbf{D}\left(\nabla I_{i,r}\right) \nabla v_{i,r}\right) + \left(I_{i,r} - I_{i+1,r}^{\mathbf{h}_{i,r}}\right) I_{i+1,r,y}^{\mathbf{h}_{i,r}}
\end{aligned}$$

Para poder superar el problema de la detección de los desplazamientos largos se ha utilizado un esquema multipiramidal.

3.2.4. Resultados Experimentales

En este apartado se presentan los resultados obtenidos con el método presentado en esta sección. Para evaluar la mejora que ofrece este nuevo método se ha realizado una comparación tanto cualitativa como cuantitativa con su versión puramente estéreo [Alvarez02b]. Los datos de entrada que manejamos en los experimentos son secuencias de pares de estéreo. No ha sido posible encontrar una secuencia de este tipo en la literatura. Por este motivo, sólo se ha podido realizar experimentos con tres secuencias: una real y dos sintéticas. La secuencia real ha sido cedida por la empresa *Mediapro* y en ella se capta un partido de fútbol celebrado *MiniEstadi*. Las otras dos secuencias han sido generadas de forma sintética y consisten en unos objetos texturados desplazándose horizontalmente delante de dos cámaras. A continuación, se presentan los resultados obtenidos con estas secuencias. Para poder analizar los resultados es necesario definir una serie de métricas que nos permitan cuantificar la precisión de las estimaciones. Al igual que en el problema del flujo óptico se ha utilizado el *Average Euclidean Error* (AEE, ec. 1.14) y el *Average Angular Error* (AAE, ec. 1.15). Para más detalles acerca de estas métricas ir al apartado 1.5 de la tesis.

Cilindro

La secuencia sintética del *Cilindro* ha sido creada por el grupo AMI para poder evaluar en condiciones ideales la precisión del método. En ella se ha simulado la presencia de dos cámaras situadas en posición frontoparalela (ver figura 3.4). Delante de las cámaras se mueve de derecha a izquierda un cilindro con una textura compuesta por unas hebras de

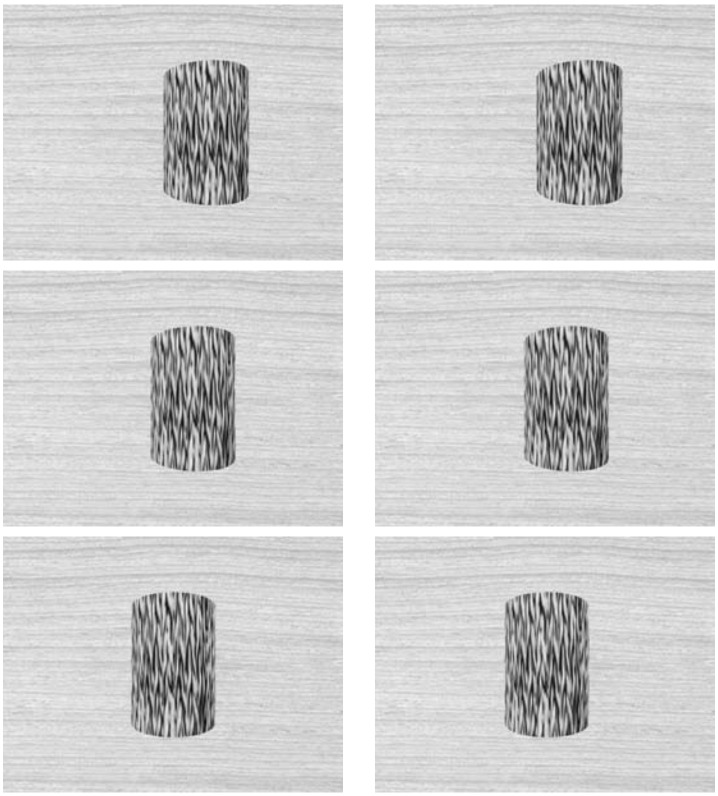


Figura 3.4: Distintos pares estéreos de la secuencia del *Cilindro*. En la primera columna, las imágenes de la cámara izquierda. En la segunda columna, las imágenes de la cámara derecha.

lana entrelazadas. En el fondo de la escena se ha planchado una textura de tono claro. La distancia respecto a la cámara no es la misma para todos los puntos del cilindro. El mapa de disparidad del cilindro consiste en un fondo con movimiento constante y, en el interior del cilindro, un degradado que será más claro en la zona central (máximo desplazamiento respecto a las cámaras) y se irá oscureciendo a medida que se aproxime a los laterales del cilindro. Los píxeles que están situados en los laterales están a mayor distancia de la cámara y, por ello, registrarán un desplazamiento menor. Para más detalles acerca de la secuencia ver el apartado 1.4.2 de la tesis.

En los resultados cuantitativos que se muestran en este apartado sólo se han tenido en cuenta los píxeles que pertenecen al cilindro. Para ello, se ha aplicado una máscara sobre la solución calculada por el método variacional.

En la figura 3.5 se muestran los mapas de disparidad densos asociados a los pares estéreo de la figura 3.4. Como se puede observar a simple vista en el interior del cilindro se produce el efecto de degradado presente en los mapas de disparidad de referencia. La solución es relativamente suave aunque se perciben varios artificios sobre la superficie del cilindro debido a la regularidad de la textura. Esto puede provocar una *confusión* en nuestro método al no establecer correctamente las correspondencias originando valores de disparidad erróneos. Otra deficiencia que observamos en la solución obtenida está en que

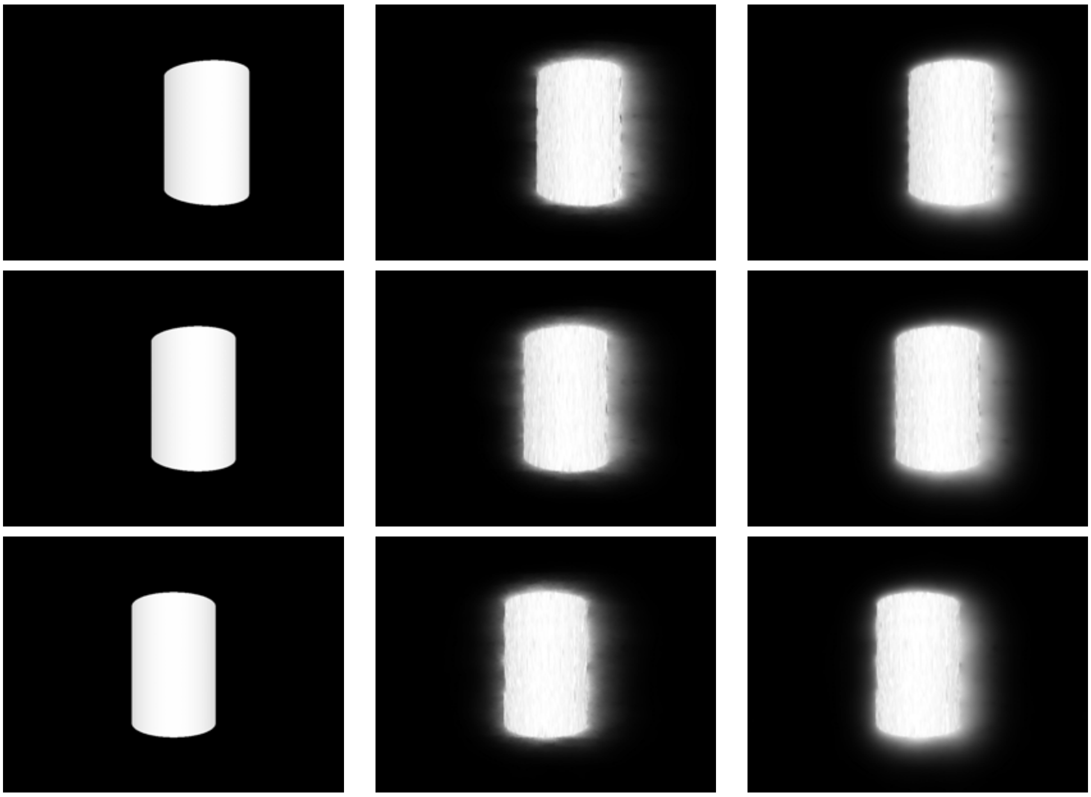


Figura 3.5: En la columna izquierda, los mapas de disparidad reales correspondiente a los frames 0, 4 y 10. Píxeles con tonos claros indican desplazamientos mayores y, los oscuros, los menores. En la columna central, los mapas de disparidad estimados con el método espacial. En la columna derecha, los mapas de disparidad obtenidos con nuestro método.

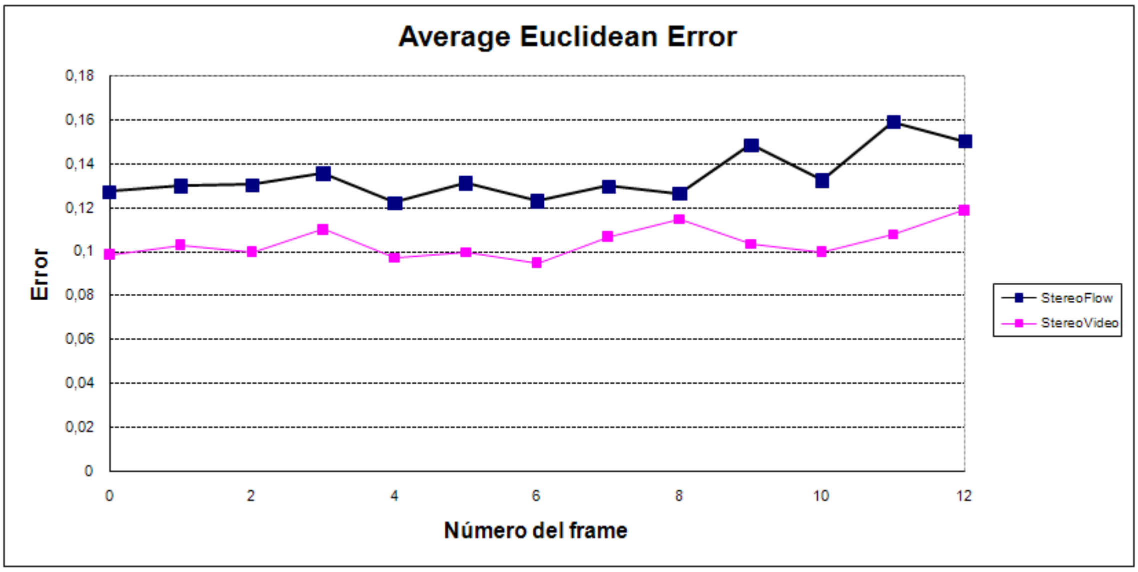


Figura 3.6: Gráfica del AEE obtenido en la secuencia del *Cilindro*.

Cilindro AEE				
Frame	**Método espacial**	**Nuestro método**	**Diferencia**	**% Mejora**
0	0.127043	0.098521	0.028521	22.45 %
1	0.129855	0.103104	0.026751	20.60 %
2	0.130148	0.099989	0.030159	23.17 %
3	0.135433	0.110151	0.025282	18.67 %
4	0.122247	0.097061	0.025186	20.60 %
5	0.131151	0.099486	0.031665	24.14 %
6	0.122922	0.094766	0.028156	22.91 %
7	0.129470	0.106723	0.022747	17.57 %
8	0.126313	0.114762	0.011551	9.14 %
9	0.148327	0.103394	0.044933	30.29 %
10	0.132380	0.099746	0.032634	24.65 %
11	0.158627	0.107786	0.050841	32.05 %
12	0.150153	0.118870	0.031283	20.83 %

Cuadro 3.1: Secuencia del *Cilindro*: AEE obtenido en los distintos frames con el método estéreo espacial [Alvarez02b] y el método propuesto en esta sección.

Cilindro AAE				
Frame	**Método espacial**	**Nuestro método**	**Diferencia**	**% Mejora**
0	0.387795	0.294634	0.093161	24.02 %
1	0.383048	0.305028	0.078020	20.37 %
2	0.382562	0.291168	0.091394	23.89 %
3	0.396488	0.389646	0.006842	1.73 %
4	0.356934	0.285340	0.071594	20.06 %
5	0.376084	0.289093	0.086991	23.13 %
6	0.346269	0.272438	0.073831	21.32 %
7	0.372812	0.319556	0.053256	14.28 %
8	0.372630	0.415654	-0.043024	-11.55 %
9	0.606374	0.308366	0.298008	49.15 %
10	0.390006	0.296871	0.093135	23.88 %
11	0.480648	0.313414	0.167234	34.79 %
12	0.442590	0.348788	0.093802	21.19 %

Cuadro 3.2: Secuencia del *Cilindro*: AAE obtenido en los distintos frames con el método estéreo espacial [Alvarez02b] y el método propuesto en esta sección.

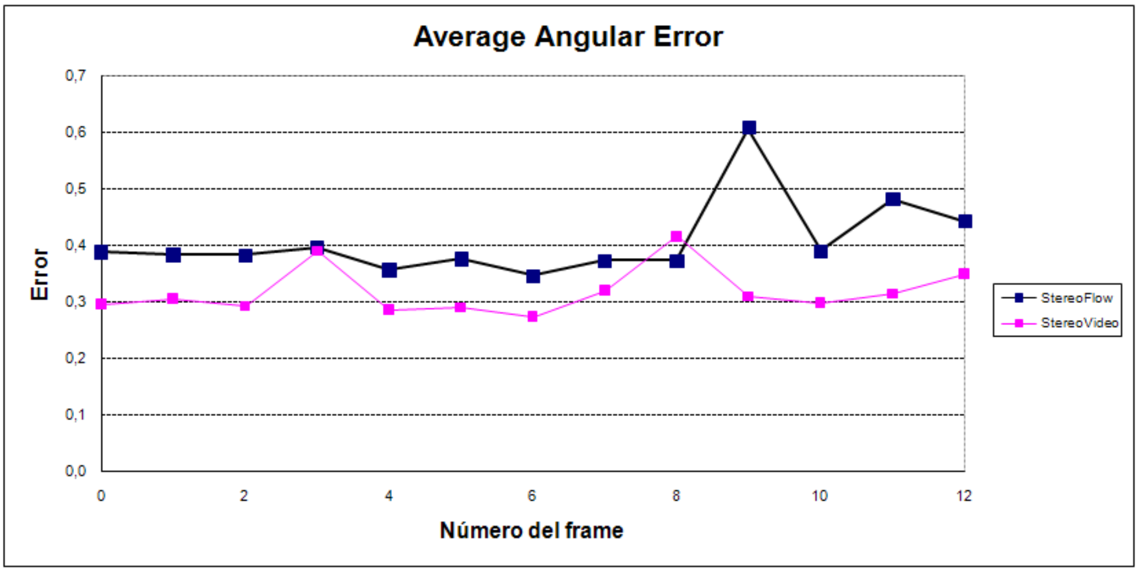


Figura 3.7: Gráfica del AAE obtenido en la secuencia del *Cilindro*.

el proceso de difusión no se ha detenido en los contornos de cilindro pese a su contraste con el fondo de la escena. Debido al uso del gradiente de la imagen en el operador de Nagel–Enkelmann, si en ciertas zonas del contorno del objeto el gradiente no es lo suficientemente alto puede provocar que la regularización actúe más allá de los límites del objeto creando una especie de áurea alrededor del mismo. En este caso, la utilización de funciones de robustificación permitiría mejorar las estimaciones.

Los modelos de energía de [Alvarez00] y [Alvarez02b] en los que se basa nuestro método tienen como característica la estabilidad de sus parámetros, pequeñas variaciones en sus valores ofrecen resultados similares. Esta característica facilita las pruebas y la optimización de los parámetros no se convierte en una tarea tediosa. Los parámetros que configuran nuestro método variacional son: α, β, ζ y el número de escalas. α es el peso que pondera la importancia de los términos de los flujos óptico y estéreo por separado y, β es el peso del término que combina la energía de ambos flujos. ζ establece el comportamiento isotrópico o anisotrópico del tensor de difusión de Nagel–Enkelmann. Con valores pequeños el tensor de difusión tiene un comportamiento anisotrópico. Los parámetros α y ζ se calculan a través de las ecuaciones descritas en el apartado 2.3.1. Los datos de entrada son λ y C_α. Los valores de estos parámetros son $C_\alpha = 1{,}5$ ($\alpha = 1{,}86 \times 10^{-4}$), $\beta = 1 \times 10^{-3}$ y $\lambda = 0{,}1$. Aunque los valores de C_α y β parecen muy dispares una vez que se ha estimado el peso del término de regularización en función de la normalización de las imágenes de entrada obtenemos valores de similar magnitud. Se ha tomado un valor de $\lambda = 0{,}1$ para evitar que el proceso de difusión actuara más allá de los contornos del cilindro. Nuestro método utiliza un enfoque multipiramidal para recuperar los desplazamientos largos y, al mismo tiempo, evitar converger a un mínimo local irrelevante. El número de escalas óptimo depende del desplazamiento máximo registrado en los datos. En nuestro caso, ese desplazamiento máximo es de alrededor de 8 píxeles por frame. Dado que el factor de reducción aplicado en cada escala se ha fijado a 0,5 el número de escalas utilizadas ha sido tres. En los experimentos quedó demostrado que un número superior de escalas no mejoraba los resultados obtenidos. Los valores de los parámetros en el método estéreo espacial han sido $C = 0{,}7(\alpha = 3{,}91 \times 10^{-4})$, $\lambda = 0{,}1$ y el número de escalas de tres. Dado que nuestro método se basa en éste es lógico que los valores de los parámetros sean similares.

La mejora aportada por el método presentado en esta sección no se puede cuantificar sino establecemos una comparación con otros métodos. En los experimentos esta comparación se ha hecho con el método estéreo [Alvarez02b]. De esta forma es posible determinar el grado de influencia del término de flujo óptico en el nuevo método. En las tablas 3.1 y 3.2 podemos encontrar los resultados cuantitativos obtenidos de la comparación de ambos métodos. Para facilitar el análisis de la información de ambas tablas hemos representado dicha información de forma gráfica (figuras 3.6 y 3.7). Se observa de manera general la reducción en el error que nuestro método ofrece. Sin embargo, existen ciertos frames, más concretamente el tercero y el octavo, en el que la estimación es ligeramente peor. Las diferencias apreciables en la figura 3.5 se reflejan en los errores detectados en todos los frames de la secuencia. Analizando los resultados de las tablas podemos sacar dos conclusiones: (1) la estabilidad de las estimaciones del método y, (2) la contribución de la información del flujo óptico en la mejora de la estimación del mapa de disparidad.

Cilindro y Esfera

La segunda secuencia sintética utilizada se trata de una versión algo más compleja que la secuencia del *Cilindro* (figura 3.8). Ahora disponemos de dos objetos, un cilindro y una esfera, que se desplazan a distinta velocidad por la escena. Se produce oclusiones entre ambos objetos. Cerca del fondo de la escena se aprecia un panel estático con una textura clara que ocupa casi la totalidad de la imagen. Para más detalles acerca de la secuencia ver el apartado 1.4.2 de la tesis.

A diferencia de los experimentos realizados con la secuencia anterior hemos tenido en cuenta todos los píxeles de la imagen a la hora de obtener los resultados cuantitativos. En la figura 3.9 se muestran los mapas de disparidad densos asociados a los pares estéreo de la figura 3.8. Como se puede observar se ha producido una correcta detección de los tres objetos presentes en la escena. El cilindro es el objeto más próximo a las cámaras. Por este motivo aparece en un tono claro. En el interior del cilindro se aprecia el efecto de degradado presente en los mapas de disparidad de referencia. Sin embargo, en ambas soluciones se observan algunos artificios y zonas de disparidad subestimadas. La esfera también es detectada correctamente por ambos métodos. La diferencia más significativa entre ellos está en el contorno de la esfera y sobre todo en la discontinuidad con el cilindro, zonas donde se producen oclusiones. El último objeto presente en la escena es el panel situado al fondo, justo detrás del cilindro y la esfera. En el interior de dicho objeto la disparidad ha sido identificada correctamente y apenas existe diferencia entre ambos métodos. Sin embargo, en su contorno se aprecia los errores más relevantes del mapa de disparidad. En nuestro método el proceso de difusión se detiene con mayor precisión en el contorno del panel.

Los parámetros que configuran nuestro método variacional son: α, β, ζ y el número de

Cilindro y Esfera AEE				
Frame	**Método espacial**	**Nuestro método**	**Diferencia**	**% Mejora**
0	0.079917	0.064413	0.015503	19.40 %
1	0.082677	0.069917	0.012761	15.43 %
2	0.078151	0.067839	0.010312	13.20 %
3	0.099163	0.066084	0.033079	33.36 %
4	0.075023	0.065726	0.009297	12.39 %
5	0.085858	0.071507	0.014351	16.71 %
6	0.071050	0.067640	0.003409	4.80 %
7	0.085979	0.068504	0.017475	20.32 %
8	0.082747	0.065617	0.017130	20.70 %
9	0.085170	0.063675	0.021496	25.24 %
10	0.067150	0.062035	0.005115	7.62 %
11	0.103858	0.073978	0.029880	28.77 %
12	0.097868	0.055609	0.042259	43.18 %

Cuadro 3.3: Secuencia del *Cilindro* y la *Esfera*: AEE obtenido en los distintos frames con el método estéreo espacial [Alvarez02b] y el método propuesto en esta sección.

escalas. Los valores de entrada son: $C_\alpha = 0,03$ ($\alpha = 9{,}48 \times 10^{-3}$) y $\beta = 1 \times 10^{-3}$. Dado los fuertes contrastes que se producen en los contornos de los objetos nos interesa que el término de regularización tenga un comportamiento anisotrópico. En nuestro método $\lambda = 0{,}1$. En esta secuencia el desplazamiento máximo es de alrededor de 12 píxeles por frame. Dado que el factor de reducción aplicado en cada escala se ha fijado a 0,5 el número de escalas utilizadas ha sido cuatro. En los experimentos quedó demostrado que un número superior de escalas no mejoraba los resultados obtenidos. Los valores de los parámetros en el método estéreo espacial han sido también $C = 0,03$ ($\alpha = 9{,}48 \times 10^{-3}$), $\lambda = 0{,}1$ y cuatro escalas.

El aumento en la precisión de las estimaciones ofrecida por nuestro método no se puede cuantificar sino establecemos una comparación con [Alvarez02b]. En las tablas 3.3 y 3.4 podemos encontrar los AEE y AAE obtenidos por ambos métodos. Para facilitar el análisis de la información de ambas tablas hemos representado dicha información de forma gráfica (figuras 3.10 y 3.11). En estas gráficas podemos observar la estabilidad del error en cada uno de los frames y cómo la combinación de información del flujo óptico y estéreo permite aumentar la precisión de las estimaciones. En general nuestro método ha permitido la obtención de soluciones más suaves en el interior de los objetos y una mejor identificación de la disparidad en las discontinuidades de los objetos, zonas éstas últimas donde los métodos fallan debido a las oclusiones.

Secuencia de Fútbol

La secuencia de fútbol corresponde a un partido celebrado en el MiniEstadi del F.C. Barcelona. En ella se observa un área del campo desde dos cámaras situadas en la grada (fig. 3.12). El juego se desarrolla en la parte del campo que no es visible; en el área se

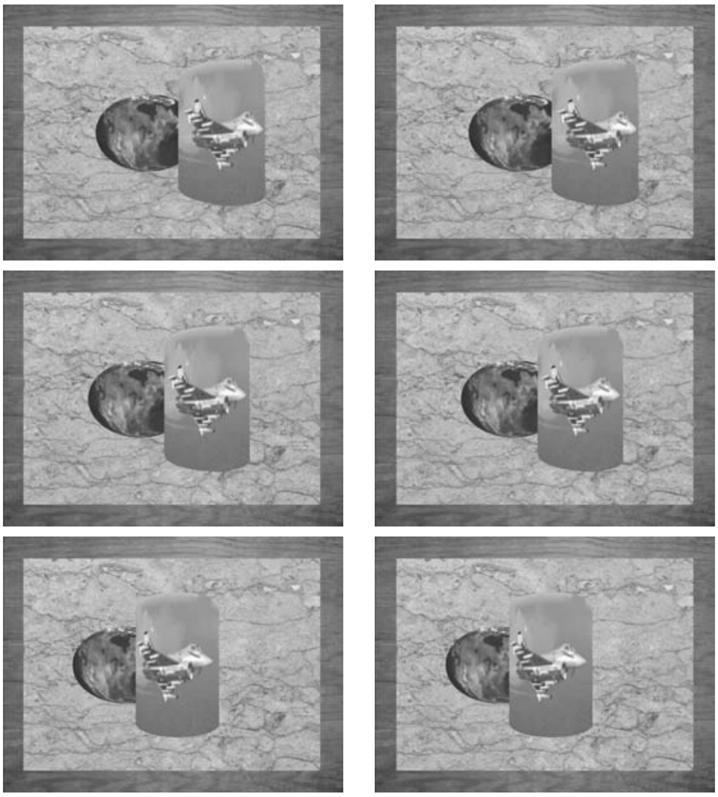


Figura 3.8: Distintos pares estéreos de la secuencia del *Cilindro* y la *Esfera*. En la primera columna, las imágenes de la cámara izquierda. En la segunda columna, las imágenes de la cámara derecha.

Cilindro y Esfera AAE				
Frame	**Método espacial**	**Nuestro método**	**Diferencia**	**% Mejora**
0	4.059050	2.946890	1.112160	27.40 %
1	4.056370	3.016360	1.040010	25.64 %
2	4.041220	3.051960	0.989260	24.48 %
3	4.666410	2.998850	1.667560	35.74 %
4	3.943700	2.991910	0.951790	24.13 %
5	4.201020	3.071630	1.129390	26.88 %
6	3.871540	3.182070	0.689470	17.81 %
7	3.921320	3.108290	0.813030	20.73 %
8	4.467930	3.071850	1.396080	31.25 %
9	4.372420	3.034130	1.338290	30.61 %
10	3.934800	3.116080	0.818720	20.81 %
11	4.750460	3.330940	1.419520	29.88 %
12	4.050140	3.281430	0.768710	18.98 %

Cuadro 3.4: Secuencia del *Cilindro* y la *Esfera*: AAE obtenido en los distintos frames con el método estéreo espacial [Alvarez02b] y el método propuesto en esta sección.

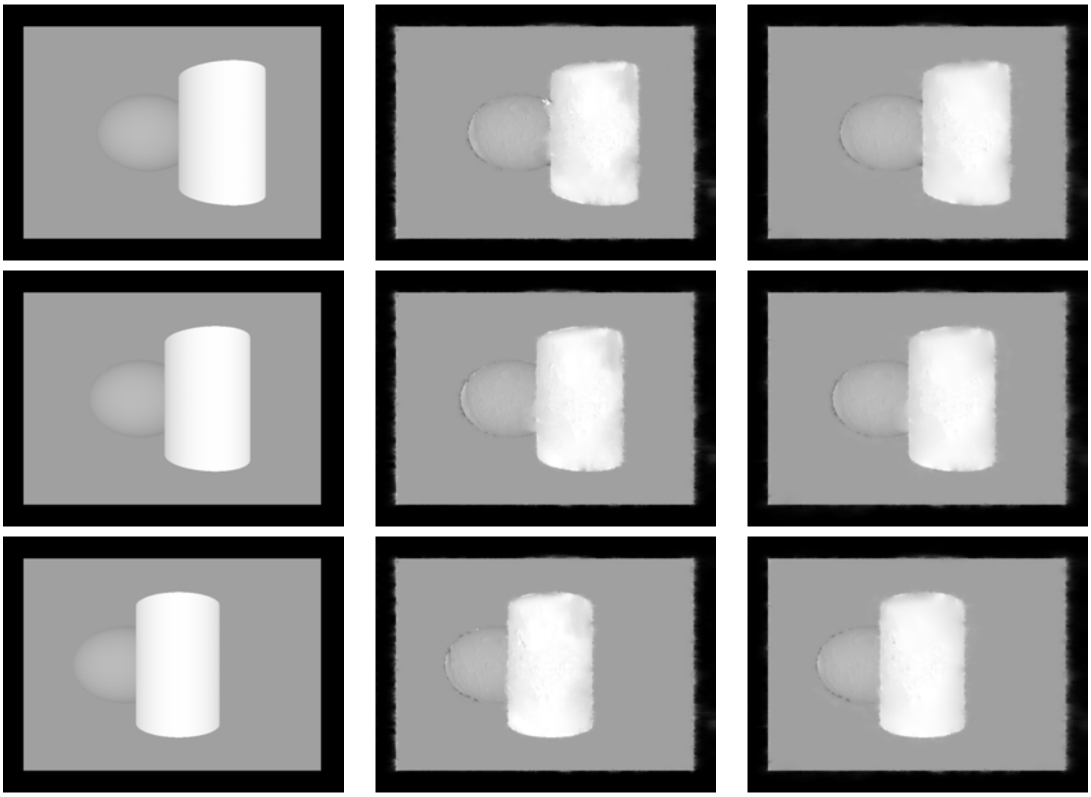


Figura 3.9: En la columna izquierda, los mapas de disparidad reales correspondiente a los frames 1, 4 y 10. Píxeles con tonos claros indican desplazamientos mayores y, los oscuros, los menores. En la columna central, los mapas de disparidad estimados con el método espacial. En la columna derecha, los mapas de disparidad obtenidos con nuestro método.

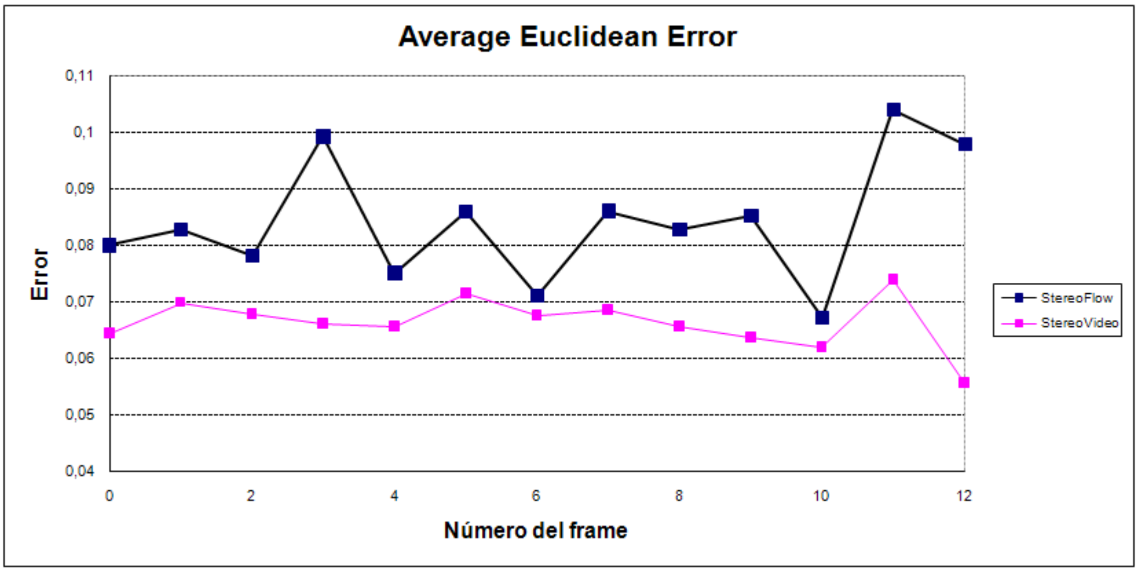


Figura 3.10: Gráfica del AEE obtenido en la secuencia del *Cilindro* y la *Esfera*.

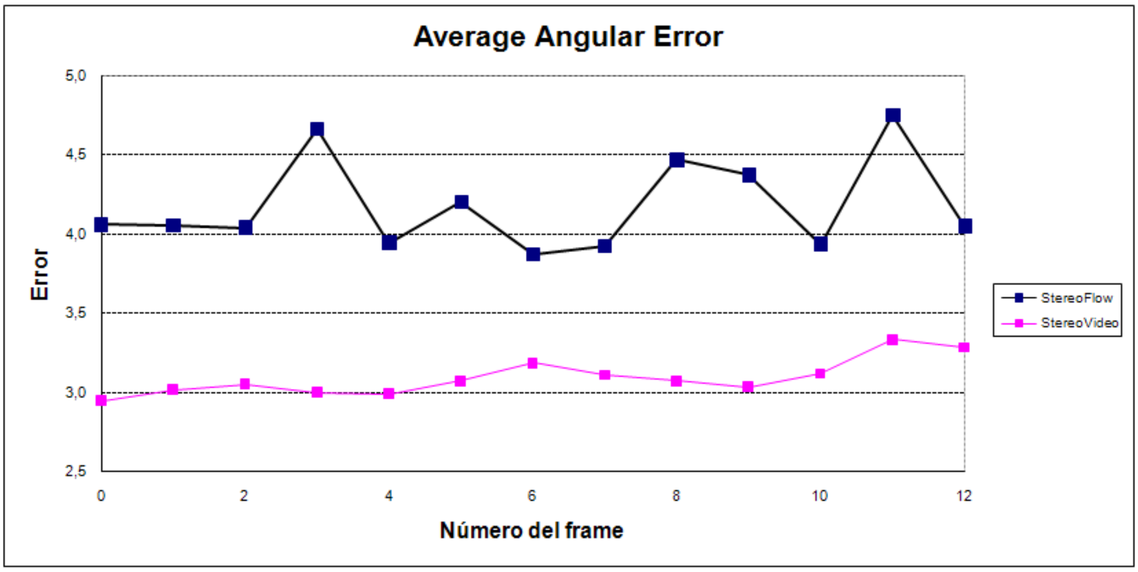


Figura 3.11: Gráfica del AAE obtenido en la secuencia del *Cilindro* y la *Esfera*.

puede ver al portero caminando desde el punto de penalti hacia el borde del área. Dada las dimensiones de las imágenes (1900×1080 píxeles) se ha seleccionado una región más pequeña, de tamaño 430×170 píxeles, donde se percibe el único movimiento de interés de la secuencia, el caminar del portero en el área. Desgraciadamente la posición y orientación de las cámaras no es frontoparalela. Por ello, se ha recurrido a un proceso de rectificación de las imágenes de ambas cámaras. La secuencia original está compuesta por miles de frames. En nuestros experimentos se han seleccionado once frames con el objetivo de verificar, en un tiempo razonable, el correcto funcionamiento del método en situaciones reales. En esta secuencia no disponemos de los mapas de disparidad reales, pero dada su simplicidad será muy fácil verificar visualmente la validez de las estimaciones obtenidas. En función de la distancia a las cámaras los objetos aparecerán en un tono claro los más próximos y oscuros los más lejanos.

En la figura 3.14 se muestran los mapas de disparidad obtenidos con el método estéreo [Alvarez02b] y el nuestro en los frames 0, 5 y 9 de la secuencia. Los dos objetos presentes en las imágenes son el portero y el punto de penalti. La estimación de la disparidad de ambos objetos se ha hecho correctamente gracias al contraste que se produce con el césped. La figura del portero se aprecia con distintas intensidades dependiendo de la distancia respecto a la cámara de cada parte del cuerpo. Las partes más cercanas a la cámara como son el guante izquierdo y el punto de penalti aparecen en tono cercano al blanco. Las partes del cuerpo más lejanas pierna y brazo derecho tonalidades oscuras. El césped tiene una tonalidad constante debido a la actuación del término de suavizado sobre el de ligadura ya que éste no tiene información suficiente para establecer las correspondencias entre las imágenes. Si comparamos los mapas de disparidad estimados por cada método podemos apreciar que la solución del nuestro es mucho más suave. Como era de esperar la utilización de información del flujo óptico aporta estabilidad a la solución.

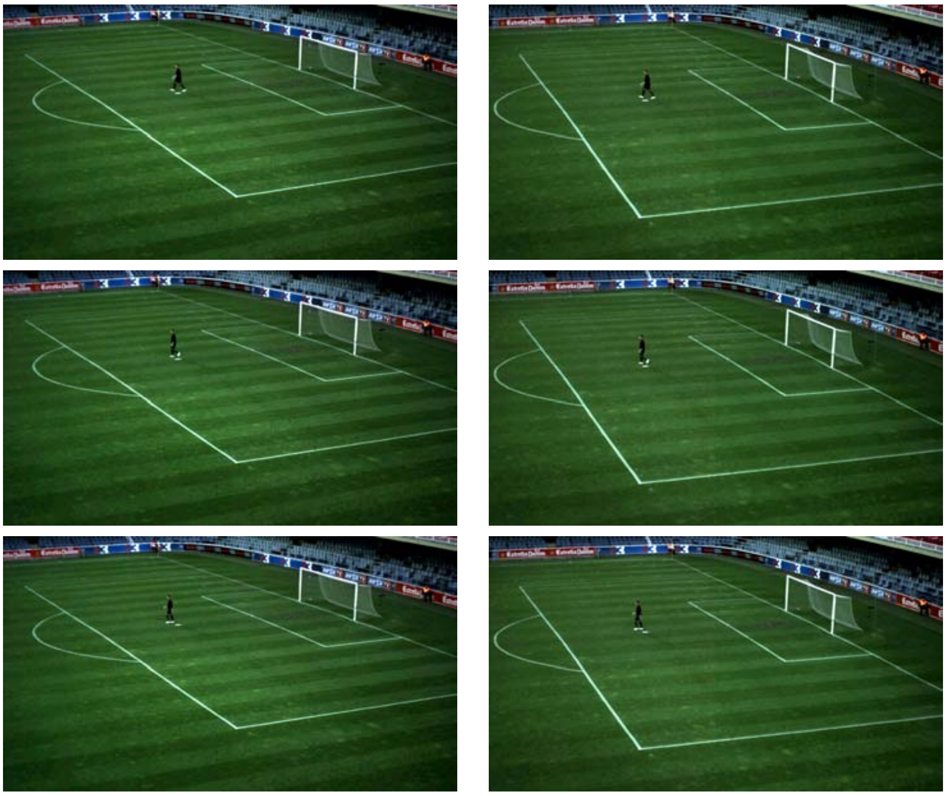


Figura 3.12: Secuencia de fútbol. En la primera columna, los frames 0, 5 y 9 tomados por la cámara izquierda. En la segunda columna, los mismos frames pero captados por la cámara derecha.

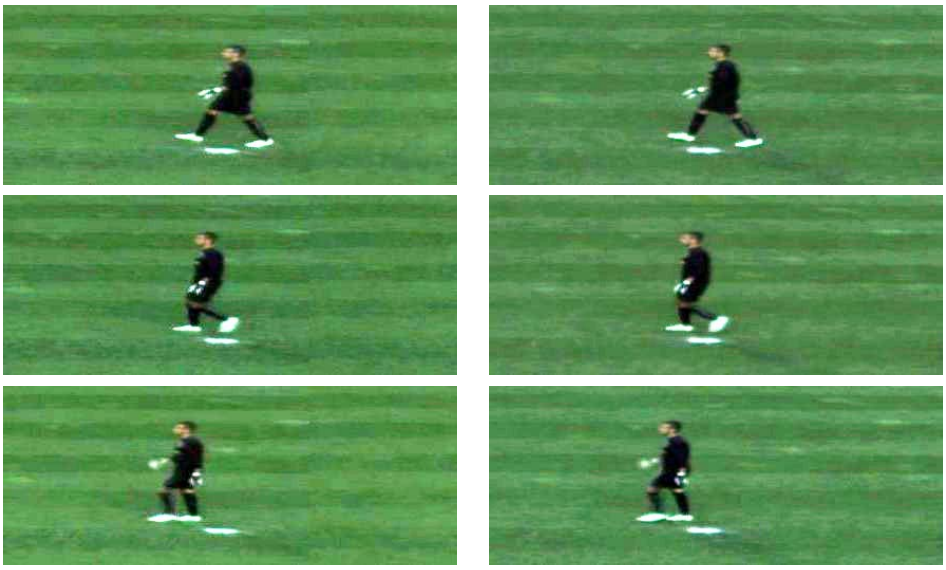


Figura 3.13: Imágenes rectificadas modificadas de las cámaras izquierda y derecha correspondientes a los frames mostrados en la figura 3.12.

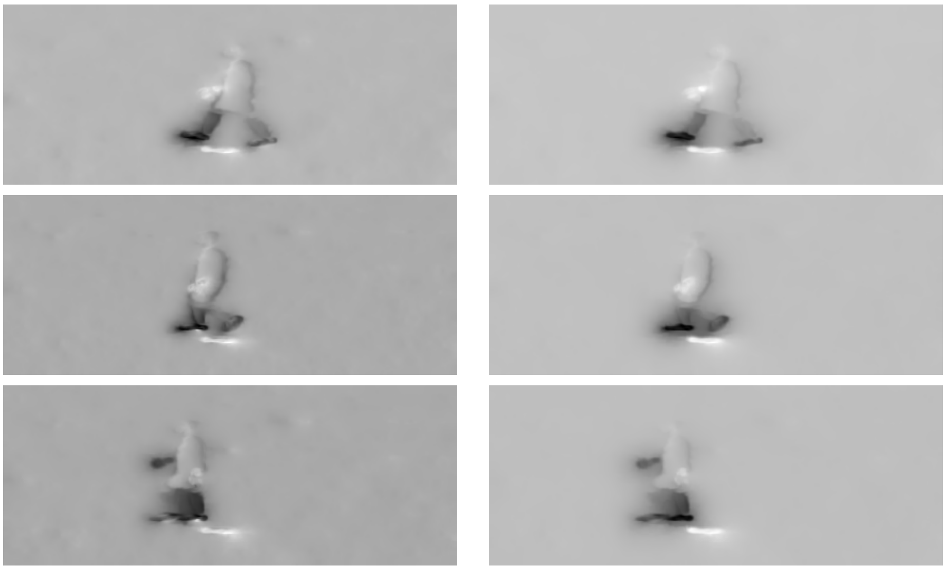


Figura 3.14: Mapas de disparidad obtenidos con las imágenes rectificadas presentadas en la figura 3.13. En la columna de la izquierda, los mapas de disparidad calculado con un método variacional [Alvarez02b]. En la columna derecha, el mapa de disparidad obtenido con nuestro método.

El modelo de energía de nuestro método no incorpora ninguna técnica que lo haga insensible a las perturbaciones típicas de cualquier secuencia real (p. e. ruido o cambios de iluminación). Esto provoca que las estimaciones hechas por los términos estéreo y del flujo óptico se vean influenciadas por estas perturbaciones. Durante las pruebas hemos observado que cuando la relación entre el peso del parámetro C_α y β era inferior a 100 la estabilidad del método se veía seriamente comprometida y los resultados obtenidos no eran correctos. Por este motivo, el valores de los parámetros son $C_\alpha = 0,3$ ($\alpha = 1{,}61 \times 10^{-4}$), y $\beta = 3x10^{-3}$. Para valores de C_α mayores que 0,3 el mapa de disparidad era demasiado suave. Dado en las imágenes el desplazamiento de los objetos es pequeño solamente se ha utilizado una única escala en el enfoque multipiramidal, con un factor de reducción fijado a 0,5. Para evitar crear una sobresegmentación en el flujo y favorecer una solución suave el parámetro del regularizador de Nagel–Enkelmann es de $\lambda = 0{,}5$.

3.3. Estimación del Mapa de Disparidad mediante la Combinación de un Método de Graph–cuts y uno Variacional

Los métodos variacionales han tenido bastante éxito en el problema de la estimación del mapa de disparidad y han demostrado poder obtener muy buenos resultados. Para evitar la convergencia de la solución a mínimos locales se recurre a un enfoque multipiramidal aunque otra alternativa sería la utilización de una inicialización. Para obtener esta inicialización se suele recurrir a una técnica basada en la correlación. Una alternativa que ha surgido en los últimos años a los métodos variacionales han sido los métodos de *graph–cuts*. En esencia comparten muchas características: basados en la minimización de energías, resultados densos, etc. Aunque proceden de enfoques matemáticos diferentes.

Dado las similitudes y los buenos resultados que ofrecen parece razonable que la unión del método de graph–cuts y el método variacional puede ser más ventajosa que la aportada por la técnica basada en correlación. Por este motivo, en esta sección se propone un método para la estimación de los mapas de disparidad que combina uno basado en graph–cuts y otro variacional. El método de graph–cuts ha sido propuesto por Kolmogorov–Zabih [Kolmogorov01, Boykov04]. Ofrece muy buenos resultados en muchas aplicaciones como pueden ser la segmentación o la estimación de los mapas de disparidad. Es capaz de detectar y manejar las oclusiones pero las soluciones tienen precisión entera. El método variacional es el descrito en el artículo Alvarez/Deriche/Sánchez/Weickert (2002) [Alvarez02b] y se trata de la versión estéreo del método de flujo óptico Alvarez/Weickert/Sánchez (2000) [Alvarez00]. El modelo de energía descrito en Alvarez/Deriche/Sánchez/Weickert (2002) [Alvarez02b] no tiene en cuenta las oclusiones.

A continuación se describe con detalle el trabajo desarrollado en esta sección. En primer lugar, se define la estructura de la escena y el proceso de rectificación de imágenes que se aplica a las secuencias cuando la disposición de las cámaras no es frontoparalela. En segundo lugar, se comenta brevemente las técnicas utilizadas como aproximación inicial del método variacional: en 3.3.1, una implementación básica de la correlación a ventanas utilizada en [Alvarez02b] y, en el apartado 3.3.2, el método de graph–cuts. Una vez comentadas estas dos técnicas se explica cómo se ha combinado el método variacional con el del graph-cuts y las similitudes que existen entre ellos, además del enfoque multiescala para la detección de los desplazamientos. En el apartado 3.3.5 se muestra los resultados experimentales obtenidos con el método propuesto y se establecerá una comparación entre el método variacional con las dos aproximaciones iniciales. Con esta comparación se pretende evaluar la influencia de cada una de las inicializaciones en la solución final. Para ello, se utilizarán pares estéreos reales y sintéticos ofrecidos por la base de datos Middlebury (versión 2001). Se concluirá con las observaciones más importantes de la combinación de estas dos técnicas.

Estructura de la escena

La disposición de las cámaras en la escena puede ser arbitraria. Muchos de los métodos actuales asumen que el sistema de cámaras está débilmente calibrado. La matriz fundamental encierra información acerca de la posición relativa de una cámara respecto a la otra. En un sistema de calibrado fuerte esta matriz se puede calcular a partir de las matrices de proyección.

Cuando nos encontramos ante una disposición arbitraria de las cámaras los elementos de la matriz fundamental suele ser distintos de cero. Con el objetivo de simplificar el problema y obtener expresiones más simples se supone que la disposición de las cámaras es frontoparalela lo que da lugar a una matriz fundamental en la que la mayoría de sus elementos son cero, ec. (3.8).

$$F = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \tag{3.8}$$

Dado que no es sencillo obtener una sistema de cámaras situadas horizontalmente ya sea por no disponer del material adecuado o por una incorrecta configuración de las cámaras, siempre es posible *simular* esta disposición utilizando un proceso denominado *rectificación*.

El proceso de rectificación consiste en reproyectar las imágenes, ver figura 3.15, de tal forma que la matriz fundamental asociada tenga la estructura de la matriz de la ecuación (3.8). Esta transformación representa un cambio en la disposición de los píxeles en ambas imágenes. Pese a que este proceso de rectificación es algo complejo obtenemos varios beneficios como son (i) una matriz fundamental muy simple y (ii) la posibilidad de expresar la disparidad como un escalar, un desplazamiento en el eje horizontal. Según la geometría epipolar, cuando la cámaras están situadas en disposición frontoparalelas las líneas epipolares serán horizontales lo que simplifica la búsqueda y la disparidad quedaría expresada como un escalar.

El proceso de rectificación es independiente al método. Se trata de un preprocesado que se hace a las imágenes antes de calcular la disparidad. Por lo tanto, la disparidad obtenida por el método corresponde con las imágenes rectificadas; es necesario deshacer la transformación inicial *desrectificando* el mapa de disparidad. Al realizar este proceso de *desrectificación* es posible que queden huecos entre los píxeles del mapa de disparidad final. Para obtener un mapa denso se suele aplicar en estos píxeles un proceso de relleno con la información de los píxeles vecinos.

Un sistema de cámaras situadas horizontalmente suele ponerse como precondición en muchos de los métodos actuales. Cuando la secuencia no cumple con este requisito es necesario un proceso de rectificación de las imágenes.

3.3.1. Método de Correlación

Los métodos basados en la correlación realizan la búsqueda de correspondencias utilizando ventanas o patrones alrededor de un píxel. Esta técnica suele ofrecer buenos resultados ya que la búsqueda se apoya no sólo en la información de un sólo píxel

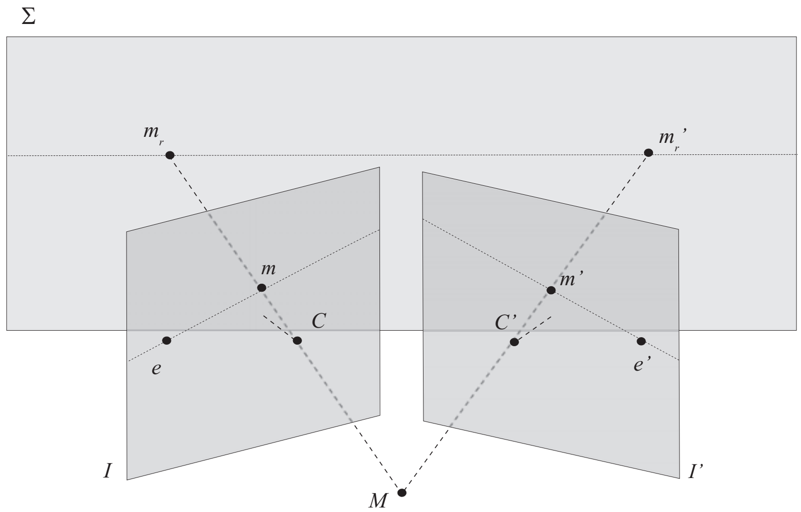


Figura 3.15: Proceso de rectificación de un sistema de cámaras estereoscópicas. Se aprecia el sistema de cámaras comentado en el apartado 1.2.1 y cómo la información se proyecta sobre el plano frontoparalelo $\sum$.

sino también en la de sus vecinos. El mayor inconveniente que tiene es su alto coste computacional.

La presencia en la imagen de ruido de baja intensidad no afecta demasiado a las estimaciones dado que éstas se calculan a partir de la información de muchos píxeles, un píxel incorrecto no tiene porqué alterar el resultado final. Debido al alto coste computacional que requiere, el desplazamiento sólo se calcula en determinados píxeles, aquéllos que pertenezcan a una rejilla predefinida previamente. Para obtener soluciones densas se suele aplicar un algoritmo de interpolación basándonos en la información de la rejilla. Una rejilla muy tupida nos aporta muy buenos resultados ya que el algoritmo de interpolación tendrá más información real para rellenar los huecos. Esto implica la aplicación del método un mayor número de veces, lo que requerirá mayor tiempo de cómputo. El tamaño de la ventana suele ser un parámetro crítico. Una ventana reducida puede verse más afectada por el ruido dado que el peso de los píxeles defectuosos será mayor y, habrá menos información para contrastar. En las zonas homogéneas podrá ofrecer resultados erróneos. Una ventana amplia aumenta la cantidad de cálculos a realizar aunque reduce la influencia del ruido. A la hora de aplicar la técnica de correlación debemos establecer un compromiso entre los valores del (i) tamaño de la rejilla y (ii) el de la ventana, para obtener buenos resultados en un tiempo razonable.

La técnica de correlación utilizada como aproximación inicial de nuestro método variacional es muy básica. Dada dos imágenes, la correlación a ventanas cuantifica la similaridad entre dos ventanas dentro de esas imágenes. Consideramos dos ventanas rectangulares de tamaño $(2P+1) \times (2N+1)$ centradas en los puntos (u_0, v_0) y (u_1, v_1), respectivamente. El valor de correlación entre las dos ventanas, $C_{12}(\tau)$, se calcula como:

$$C_{12}(\tau) = \frac{1}{k} \cdot \sum_{x=-N}^{+N} \sum_{y=-P}^{+P} (I_1(x+u_0, y+v_0) - \overline{I_1(u_0, v_0)})(I_2(x+u_1, y+v_1) - \overline{I_2(u_1, v_1)})$$

donde

$$\begin{aligned}(2P+1) &\text{ es el ancho de las ventanas}\\ (2N+1) &\text{ es el alto de las ventanas}\\ k = (2P+1) \times (2N+1) &\times \sigma_1(u_0, v_0) \times \sigma_2(u_1, v_1)\end{aligned}$$

$$\tau = (u_1 - u_0, v_1 - v_0)$$

$\overline{I(u,v)}$ y $\sigma(u,v)$ es la media de las intensidades y la desviación estándar de la imagen I en el punto (u,v).

$$\overline{I(u,v)} = \frac{1}{(2P+1) * (2N+1)} \cdot \sum_{x=-N}^{+N} \sum_{y=-P}^{+P} I(x+u, y+v)$$

$$\sigma^2(u_0, v_0) = \frac{1}{(2P+1) * (2N+1)} \cdot \sum_{x=-N}^{+N} \sum_{y=-P}^{+P} (I(x+u, y+v) - \overline{I(u,v)})^2$$

Debido a la normalización de σ_1 y σ_2, los valores devueltos por la correlación $C_{12}(\tau)$ están acotados entre $[-1, 1]$. Si las ventanas son exactamente idénticas en valores de intensidad y colocación de los puntos el valor de la correlación será máximo (+1). Este valor decrece cuanto más diferencia haya entre las ventanas.

Dada dos ventanas, una tomada de la imagen izquierda y otra de la imagen derecha, el valor máximo de la correlación determina el píxel de la imagen derecha que guarda mayor similaridad con el píxel seleccionado en la imagen izquierda. La disparidad para ese píxel será la diferencia entre las coordenadas de esos píxeles.

3.3.2. Método de Graph-cuts

Los métodos de graph–cuts se han popularizado en los últimos años. Su éxito se debe a los buenos resultados que han obtenido en un amplio número de problemas, como pueden ser la segmentación o la estimación del mapa de disparidad. Este tipo de métodos asigna una etiqueta a cada píxel basándose en la información ofrecida por un algoritmo de *minimum cut/maximum flow*. Se trata de un método de minimización de energía en la que ésta es mínima en aquellas zonas donde el flujo sea máximo. En esos puntos se establece un corte que divide los píxeles que están a ambos lados de dicho corte, creando agrupaciones de píxeles que serán etiquetados con un valor.

El nombre de *graph*, *grafo* en inglés, se debe a que la información se representa en forma de grafo. El grafo está compuesto por tres tipos de elementos: *nodos*, *conectores* y *terminales*. Tenemos un conjunto de *nodos* interrelacionados a través de los *conectores*. Los *terminales* son las etiquetas a las que se le asociarán a los píxeles, en el caso de imágenes. En la figura 3.16 vemos un gráfico que muestra la estructura de un grafo de nueve nodos.

El método de *graph–cuts* utilizado en este trabajo es el propuesto por Kolmogorov–Zabih en [Kolmogorov01]. Este método ha sido apodado como *kz2* y permite la detección de las oclusiones producidas en el par estéreo. El modelo de energía, ec (3.9), incluye a

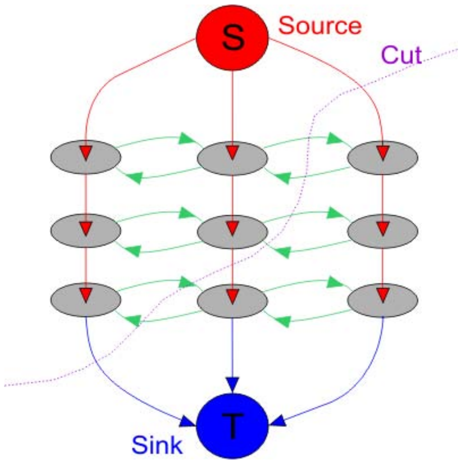


Figura 3.16: Ejemplo de la estructura de un grafo utilizado en los métodos de *graph–cuts*. Las elipses grisáceas son *nodos* (píxeles de la imagen). El círculo rojo y azul son terminales y corresponden con los puntos donde nace y muere la energía, respectivamente. La idea que subyace a este tipo de método es que la energía es mínima entre aquellos puntos donde es más sencillo de llegar desde *source* a *sink*. Para el conjunto de nodos que cumplan esta condición se agrupan bajo una misma etiqueta. En aquellos nodos situados en los límites de una agrupación y que tienen distinta etiqueta que su vecino se produce una discontinuidad.

los ya tradicionales términos de ligadura y suavizado un término de oclusión que permite identificar esos píxeles que no tienen correspondencia en la otra imagen.

$$E(h) = E_{data}(h) + E_{smooth}(h) + E_{occ}(h) \tag{3.9}$$

El funcional de energía está compuesto por tres términos que comentaremos con más detalle a continuación:

$$\begin{aligned} E(h) &= \sum (I_l(p) - I_r(q))^2 \\ &+ \sum_{\{\langle p,q\rangle,\langle p',q'\rangle\}\in\rho} K_{\{\langle p,q\rangle,\langle p',q'\rangle\}} \cdot T(h_{\langle p,q\rangle} \neq h_{\langle p',q'\rangle}) \\ &+ \sum_{p\in P} C_p \cdot T(|N_p(h)| = 0) \end{aligned}$$

donde P es un conjunto de píxeles de ambas imágenes y h, el mapa de disparidad. ρ es un conjunto de tuplas con las parejas de vecinos $\{\langle p,q\rangle, \langle p',q'\rangle\}$ que tienen la misma disparidad. C_p es la penalización del píxel p por ser una oclusión. $T(x)$ es una función que devuelve 1 si el argumento es *verdadero* y 0 si es *falso*. $N_p(h)$ es el conjunto de asignaciones activas en h, e indica las correspondencias candidatas. Cuando $|N_p(h)| = 0$, no hay asignaciones activas y se considera a ese píxel como una oclusión.

- El *término de ligadura*, E_{data}, impone una restricción de constancia sobre alguna característica de la imagen. Normalmente se utiliza la suposición lambertiana. En el caso del método *kz2* este término se define como:

$$E_{data} = \sum (I_l(p) - I_r(q))^2$$

donde $I_l(p)$ y $I_r(q)$ son los valores de intensidad de los píxeles p y q, respectivamente.

- El *término de suavizado*, E_{smooth}, penaliza cualquier desviación respecto al flujo suave por trozos. Este término penaliza a los píxeles vecinos que no tengan la misma disparidad.

$$E_{smooth} = \sum_{\{\langle p,q\rangle,\langle p',q'\rangle\}\in\rho} K_{\{\langle p,q\rangle,\langle p',q'\rangle\}} \cdot T(h_{\langle p,q\rangle} \neq h_{\langle p',q'\rangle})$$

- El *término de oclusión*, E_{occ}, penaliza a aquellos píxeles que sean considerados oclusiones. La penalización C_p se impondrá sobre el píxel "p" cuando nos encontremos ante una oclusión, es decir, $|N_p(h)| = 0$.

$$E_{occ}(h) = \sum_{p\in P} C_p \cdot T(|N_p(h)| = 0)$$

Para más detalles acerca del funcionamiento de este método podremos encontrarlo en el artículo [Kolmogorov01].

3.3.3. Método Variacional

El funcional de energía utilizado en este trabajo es exactamente el mismo que el propuesto en Alvarez/Deriche/Sánchez/Weickert (2002), [Alvarez02b]. Esta energía permite la estimación de los mapas de disparidad a partir de un par estéreo, preservando las discontinuidades utilizando la información del gradiente de la imagen. El método presupone que el sistema de cámaras está débilmente calibrado.

El funcional de energía utiliza el regularizador de Nagel–Enkelmann [Nagel86]. Este regularizador asume que los bordes de los objetos serán zonas con valores de gradiente altos. En las imágenes altamente texturadas esta suposición no se cumple siempre dado que una textura puede formar parte de una misma superficie de un objeto. En este caso, se produce una sobresegmentación del mapa de disparidad. Para la detección de los desplazamientos largos se utiliza un esquema multirresolución para evitar la convergencia hacia mínimos locales irrelevantes.

La energía propuesta para la reconstrucción de la geometría 3D de la escena se define en la ecuación (3.10).

$$E(\lambda) = \int_\Omega (I_l(\mathbf{x}) \; - \; I_r(\mathbf{x} + \mathbf{h}(\lambda(\mathbf{x}))))^2 \, d\mathbf{x} + \alpha \int_\Omega \Phi\left(\nabla I_l, \nabla\lambda\right) \, d\mathbf{x}. \tag{3.10}$$

El desplazamiento de los píxeles en la imagen $\mathbf{h}$ se expresa en base a un escalar, λ. Este escalar representa el desplazamiento de los píxeles sobre la línea epipolar. En este caso $\Phi(\nabla I_l, \nabla\lambda) = \nabla\lambda^t \cdot D(\nabla I_l) \cdot \nabla\lambda$, $D(I_l\nabla)$ se trata de una matriz de proyección perpendicular a ∇I_l,

$$D(\nabla I_l) = \frac{1}{|\nabla I_l|^2 + 2\zeta^2} \cdot \left\{ \begin{bmatrix} \frac{\partial I_l}{\partial y} \\ \frac{-\partial I_l}{\partial x} \end{bmatrix} \begin{bmatrix} \frac{\partial I_l}{\partial y} \\ \frac{-\partial I_l}{\partial x} \end{bmatrix}^t + \zeta^2 Id \right\},$$

donde Id es la matriz identidad y ζ es un parámetro que controla el comportamiento anisotrópico del tensor de difusión. El operador de Nagel-Enkelmann se ha utilizado por su simplicidad y por el rendimiento demostrado en el contexto del flujo óptico.

Para minimizar el funcional de energía se resuelve las ecuaciones de Euler-Lagrange. Después de la minimización se aplica el método de descenso del gradiente obteniendo las siguientes ecuaciones de reacción/difusión:

$$\frac{\partial\lambda}{\partial t} = \alpha \operatorname{div}\left(D(\nabla I_l)\ \nabla\lambda\right) + \left(I_l(\mathbf{x}) - I_r^{\lambda}(\mathbf{x})\right) \cdot \left(\frac{a\left(\frac{\partial I_r}{\partial y}\right)^{\lambda}(\mathbf{x})}{\sqrt{a^2+b^2}} + \frac{-b\left(\frac{\partial I_r}{\partial x}\right)^{\lambda}(\mathbf{x})}{\sqrt{a^2+b^2}} \right)$$

$D(I_l\nabla)$ juega el rol de un tensor de difusión y α es un parámetro de regularización. Más detalles acerca de este método lo podemos encontrar en [Alvarez02b].

3.3.4. Combinación del Método de Graph-cuts y Variacional

En los apartados anteriores se ha comentado de forma general cada una de las técnicas que se utilizan en este trabajo: la *correlación*, el *graph–cuts* y el método *variacional*.

A continuación se describe cómo se combinan el método *kz2* (graph–cuts) y el *StereoFlow* (método variacional). Teóricamente ambas técnicas comparten algunas similitudes. *StereoFlow* y *kz2* son métodos basados en una minimización de energía. Los modelos de energía tienen estructuras similares; se componen de varios términos como son: el de *ligadura* o el de *suavizado*. En ambos casos el término de ligadura es exactamente el mismo (salvando las distancias del modelo discreto y continuo). En el término de suavizado existen otras diferencias propias debido a la naturaleza de cada energía. Pese a estas diferencias podemos considerar a los métodos como complementarios. *kz2* detecta correctamente los bordes de los objetos y además incluye información de las oclusiones. Dada una buena inicialización, el método *StereoFlow* es capaz de alcanzar soluciones muy precisas por lo que *kz2* se convierte en el complemento perfecto.

El método variacional se ha introducido en un enfoque multipiramidal con el objetivo de mejorar la robustez del mismo frente a los mínimos locales. Las dos imágenes del par estéreo se reducen en tamaño por un factor de escala, σ, que se aplica en cada escala. El método de graph-cuts sólo se utiliza en la escala inferior para calcular la aproximación inicial. A partir de ahí la solución de la escala inferior es utilizada como inicialización en la superior donde se vuelve a aplicar el método variacional. El método variacional se emplea $N+1$ veces, siendo N el número de escalas.

3.3.5. Resultados Experimentales

En este apartado se establece una comparación entre los distintos métodos presentados anteriormente. Se quiere poner de manifiesto que la combinación del método variacional (*SF, StereoFlow*) y el método de graph–cuts (*kz2*) ofrece mayor precisión y robustez respecto a otro tipo de técnicas basadas en correlación (*Corr*). Llamaremos *kz2+SF* al método surgido de la combinación del método variacional y el de graph–cuts. La combinación del método variacional y la correlación se definirá como *Corr+SF*. A continuación se comparan estos cuatro métodos: *Corr*, *kz2*, *kz2+SF* y *Corr+SF*.

En los experimentos también se demuestra la robustez del método variacional frente al ruido presente en los pares, independientemente del tipo de inicialización utilizada. El método de graph–cuts identifica a los píxeles ruidosos como oclusiones, hecho que reduce notablemente la precisión de la estimación. El método variacional es capaz de suavizar el flujo y obtener soluciones precisas aun cuando el mapa de disparidad ofrecido por *kz2* está *contaminado* por píxeles ruidosos.

Las secuencias de pares estéreo reales y sintéticos que se han empleado forman parte de la base de datos de Middlebury. Estos pares han sido ampliamente utilizados en la literatura para evaluar los métodos dado que incorporan distinto tipo de texturas, objetos, movimientos y el mapa de disparidad es conocido. Los pares estéreo son: *Venus* (Fig. 3.17), *Map* (Fig. 3.19), *Sawtooth* (Fig. 3.21) y *Tsukuba* (Fig. 3.23). Se pueden descargar en `http://cat.middlebury.edu/stereo/data.html`). Estos pares son a color. Dado que nuestro método utiliza un sólo canal se ha transformado estas imágenes del espacio a color RGB al de HSV. De la descomposición de la imagen a color en los canales *Tonalidad*, *Saturación*, *Valor* se ha tomado este último canal por ser el que más información con contraste aporta. Por último, otra secuencia empleada ha sido la archiconocida como par de pasillo (*Corridor*, en inglés) (Fig. 3.25, de la Universidad de Bonn, `http://www-student.informatik.uni-bonn.de/ ~gerdes/MRTStereo/`). En el par real de *Tsukuba* no es posible conocer la disparidad exacta pero se ha creado *artificialmente* un mapa que representa aproximadamente el desplazamiento de los objetos. El desplazamiento de las cámaras en todas las secuencias es horizontal.

Para poder comparar los resultados cuantitativos es necesario definir una serie de métricas. Al igual que en el problema del flujo óptico se ha utilizado el *Average Euclidean Error* (AEE, ec. (1.14)) y el *Average Angular Error* (AAE, ec. (1.15)). Para más detalles acerca de estas métricas ir al apartado 1.5 de la tesis. Para evitar que la influencia de los bordes de la imagen afecte a los resultados cuantitativos no se ha tenido en cuenta los píxeles que están a una distancia menor a diez píxeles respecto al borde. En los experimentos con el método de correlación se ha tomado como tamaño de la ventana 21×21 píxeles.

El método *StereoFlow* utiliza un enfoque multipiramidal para recuperar los desplazamientos largos y evitar converger a un mínimo local irrelevante. Normalmente, este algoritmo converge a un mínimo situado en la vecindad de la aproximación inicial. Por este motivo, es tan importante la precisión de la aproximación inicial. El número de escalas depende del desplazamiento máximo detectado en los datos. En cada uno de los experimentos se comentará el número de escalas utilizadas. Dado que el mapa de disparidad ofrecido por el método de graph–cuts es bueno no será necesario definir muchas escalas en el enfoque multipiramidal. En el caso de la correlación, tampoco se observó grandes diferencias en los resultados cuando se variaba este parámetro.

Base de datos de Middlebury

Middlebury es una base de datos que agrupa una serie de secuencias utilizadas por los métodos para la estimación del flujo óptico y del mapa de disparidad. Esta base de datos ha ido evolucionando a lo largo de los años incluyendo nuevas secuencias. Muchas de ellas han sido aceptadas como estándar *de facto* por parte de la comunidad científica. El objetivo de este tipo de base de datos es la distribución de secuencias que sean utilizadas en la evaluación de los métodos. Existe una amplia variedad de secuencias, reales o sintéticas que cubren todo tipo de movimientos (traslacional, rotacional o divergente). En este trabajo se han utilizado cuatro pares estéreo presentes en la versión del 2001.

Los parámetros utilizados por el método variacional son: α (peso del término de regularización), ζ (parámetro del operador de Nagel–Enkelmann) y *el número de escalas*. α y ζ son muy estables por lo que pequeñas variaciones en sus valores no ofrecerán grandes diferencias en el resultado. Tanto α como ζ son parámetros que se calculan siguiendo las ecuaciones descritas en el apartado 2.3.1. Por lo tanto los parámetros de entrada serán υ y C. En las distintas tablas que se presentan en este apartado se indican los valores de estos parámetros.

Los mapas de disparidad disponibles en Middlebury están en precisión entera y, escalados entre $[0, 255]$. Un incremento de ocho unidades en la intensidad equivale al desplazamiento de un píxel. El mapa de disparidad se podría expresar en precisión *pseudo–real* con un error máximo de 0,125 píxel. Este error puede resultar significativo en algunas aplicaciones y deberá tenerse en cuenta cuando se compare los resultados cuantitativos. Se podría pensar que la naturaleza discreta del método *kz2* le favorecería en la comparación. Lo cierto es que el escalonamiento en el flujo le perjudicará en zonas donde las superficies no sean frontoparalelas. En el caso del método variacional este pequeño error de cuantificación se traducirá en una pequeña penalización en los resultados cuantitativos.

Par de *Venus*

Venus es un par estéreo sintético en el que se observa dos texturas, una página de un periódico de deportes y otra de un póster, sobre un lienzo (figura 3.17). Para más información acerca de esta secuencia podemos ir al apartado 1.4.2 de la tesis.

En la tabla 3.5 se presentan los resultados cuantitativos obtenidos. Estos resultados indican una sustancial mejora del método *kz2+SF* respecto al *Corr+SF*. Esta mejora es de un 49,74 % y de un 48,25 % para el AEE y AAE, respectivamente. Si analizamos las estimaciones obtenidas por cada método (fig. 3.18) se percibe que los errores de la solución de *kz2+SF* se concentran principalmente en los bordes de los objetos y, alguno de ellos tienen su origen en la aproximación inicial. En la solución obtenida por *Corr+SF* se observa que la mala aproximación inicial dada por la técnica de correlación ha impedido alcanzar un resultado preciso. En la figura 3.18, disponemos los mapas de disparidad calculados por las técnicas de graph–cuts y correlación. La estimación del método de graph–cuts (*kz2*) está muy próxima a la solución real y ha sido calculada a partir de las imágenes de

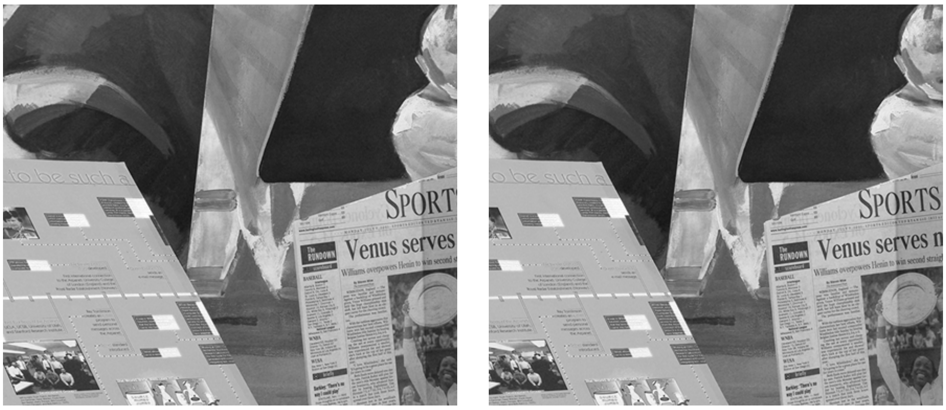


Figura 3.17: Par estéreo de Venus.

entrada originales (sin aplicar ninguna reducción en su tamaño). La inicialización ofrecida por la técnica de correlación fue obtenida con una escala por lo que su tamaño no se corresponde con el de las imágenes de entrada. En la figura 3.18 hemos obtenido el mapa de disparidad con la correlación en la escala cero para que pueda ser comparada con el resto de estimaciones. Como se puede ver a simple vista la precisión de esta aproximación es claramente inferior a la de graph–cuts. A la hora de establecer los parámetros de configuración del método de correlación nos encontramos ante un dilema: qué tamaño de ventana escoger y el radio de búsqueda. Ventanas pequeñas son más sensibles al ruido y en zonas homogéneas dan lugar a estimaciones incorrectas. Ventanas excesivamente grandes no permiten discriminar con claridad qué desplazamiento entre píxeles candidatos correspondería con el real. Cuando las zonas de búsqueda son pequeñas pueden dar lugar a la no detección de ciertos desplazamientos. Zonas de búsqueda grandes podría provocar que, en ciertas circunstancias, se pudiera comparar regiones similares situadas a gran distancia aunque por coherencia su desplazamiento fuera superior al de sus vecinos. En el mapa de disparidad obtenido con la técnica de correlación se aprecia como ha habido una mala estimación. La utilización de una ventana de correlación grande ha evitado malas estimaciones en zonas homogéneas pero ha originado una incorrecta estimación de la disparidad sobre todo en las dicontinuidades de los objetos.

La técnica de correlación utilizada es muy básica. Con una versión más sofisticada la precisión de los resultados aumentaría. Sin embargo, a diferencia del método de graph–cuts las técnicas de correlación más avanzadas carecen de una formulación matemática en la que se plasman todas las restricciones. Por este motivo, la combinación del método variacional y del graph–cuts está más justificada desde el punto de vista teórico ya que ambas técnicas reflejan sus restricciones en un modelo de energía. Desde el punto de vista práctico el esfuerzo necesario para implementar un método de graph–cuts hace que en muchas aplicaciones convenga el empleo de una técnica de correlación básica.

Los parámetros a configurar en el método variacional: C, v y el *número de escalas*. *Venus* es un par compuesto de imágenes texturadas donde se aprecia con bastante nitidez las discontinuidades de los objetos. Por un lado, un término de regularización isotrópico actuaría correctamente en el interior de los objetos pero el proceso de difusión no se detendría con precisión en las discontinuidades. Por otro lado, un

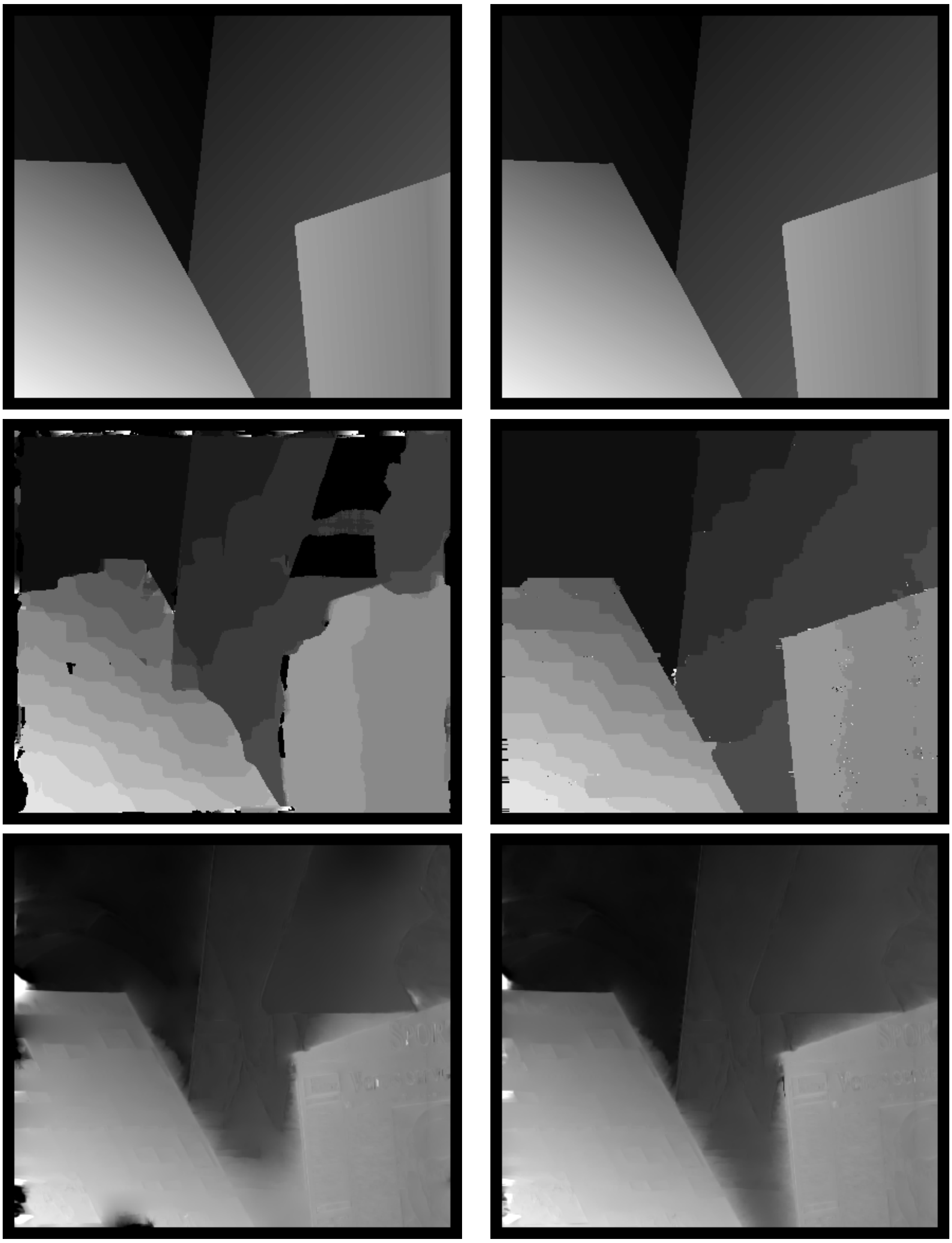


Figura 3.18: En la primera fila, tenemos por duplicado el mapa de disparidad real. En la segunda fila, las aproximaciones iniciales calculadas con los métodos de correlación y graph–cuts, respectivamente. En la última fila, los mapas de disparidad obtenidos por los métodos *Corr+SF* y *kz2+SF*.

Venus				
Método	**AEE**	**% mejora**	**AAE**	**% mejora**
Corr ($sc = 1$)	1.02682	167.19 %	7.50134	1009.50 %
kz2	0.30470	-20.71 %	0.47434	-29.84 %
Corr+SF ($C = 0{,}1, \upsilon = 0{,}1,\ sc = 1$)	0.38430	0.00 %	0.67610	0.00 %
***kz2*+SF** ($C = 0{,}1, \upsilon = 0{,}1,\ sc = 0$)	0.19315	-49.74 %	0.34989	-48.25 %

Cuadro 3.5: Par estéreo de *Venus*: AEE y AAE obtenidos con los métodos *Corr*, *kz2*, *Corr+SF* y *kz2+SF*.

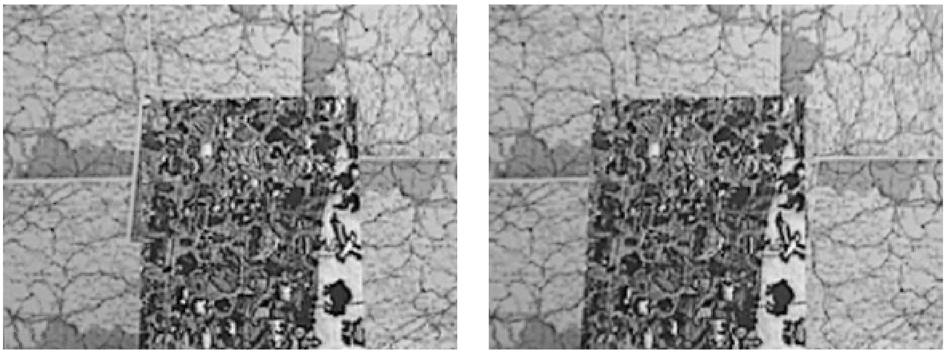


Figura 3.19: Par estéreo de *Map*.

regularizador anisotrópico si que funcionaría bien en las discontinuidades pero crearía una sobresegmentación del flujo en zonas interiores. En estos experimentos creemos conveniente primar el comportamiento anisotrópico pese a las *deficiencias* que genera en zonas homogéneas. Por este motivo, los valores utilizados en los parámetros del método variacional son: $C = 0{,}1$ (la influencia del término de suavizado es baja), $\upsilon = 0{,}1$ (el comportamiento del regularizador es totalmente anisotrópico) y cero escalas. De esta forma, cuando nos encontremos en los contornos de los objetos sólo se suavizará a lo largo de ellos. Como se puede apreciar en los mapas de disparidad obtenidos la difusión aplicada en las discontinuidades sólo se propaga ligeramente más allá de los bordes de los objetos. Cabe destacar la capacidad del método variacional de calcular buenos flujos partiendo de inicialización poco precisas, como es el caso de la correlación. El efecto de la sobresegmentación en zonas interiores es apreciable en los mapas de disparidad obtenidos con el método variacional.

Par de *Map*

El par estéreo conocido como *Map* consta de un objeto oscuro rectangular situado sobre un fondo claro (Fig. 3.19). En el apartado 1.4.2 de esta tesis encontramos más información acerca de este par.

En los resultados de la tabla 3.6 se puede extraer como conclusión la mejora que ofrece el método variacional independientemente de la aproximación inicial utilizada. Pese a la correcta identificación de las discontinuidades del objeto en movimiento por parte del

Map				
Método	**AEE**	**% mejora**	**AAE**	**% mejora**
Corr ($sc = 0$)	0.33147	79.50 %	0.78094	119.33 %
kz2	0.34991	89.49 %	0.63947	79.6 %
Corr+SF ($C = 0{,}1, \upsilon = 0{,}5,\ sc = 0$)	0.18466	0.00 %	0.35606	0.00 %
***kz2*+SF** ($C = 0{,}3, \upsilon = 1{,}0,\ sc = 0$)	0.16566	-10.29 %	0.31975	-10.2 %

Cuadro 3.6: Par estéreo de *Map*: AEE y AAE obtenidos con los métodos *Corr*, *kz2*, *Corr+SF* y *kz2+SF*.

método *kz2*, se ha visto penalizado por la discretización de su flujo en el interior del objeto. La técnica de correlación ha sido capaz de obtener una aproximación inicial más precisa. Si establecemos *Corr+SF* como referencia, el AEE es un 10,29 % peor respecto a *kz2+SF* y 89,49 % mejor en comparación con *kz2*. El método variacional es capaz de suavizar la inicialización dada por el método de graph–cuts y no se ve excesivamente penalizado por la difusión aplicada en el borde del objeto rectangular. Cuando analizamos el error angular (AAE) observamos que el método *kz2+SF* es un 10,2 % mejor que *Corr+SF* y éste a su vez un 79,6 % más preciso que *kz2*.

En la figura 3.20 es posible comparar visualmente los mapas de disparidad obtenidos con los distintos métodos respecto al mapa real. En esta imagen se observa que las soluciones ofrecidas por el método variacional son muchos más suaves en el interior del objeto. Con el método de graph–cuts se produce un efecto de suavizado escalonado. La gran diferencia entre todos los mapas se percibe en el borde del objeto rectangular. *kz2* detecta perfectamente el borde del objeto y establece una clara disparidad respecto al fondo. Sin embargo, el método variacional aplica el proceso de difusión más allá del borde real del objeto creando un áurea alrededor de él. Esto se debe a que el operador de Nagel–Enkelmann no es capaz, con la información de la imagen, de delimitar los píxeles donde se produce las discontinuidades del objeto. A pesar de la incorrecta identificación de ciertos desplazamientos localizados en el exterior del objeto rectangular la inicialización del método de correlación es ligeramente mejor que la dada por *kz2*.

El método variacional dispone de tres parámetros a determinar: C, υ y el *número de escalas*. El par de *Map* se trata de una secuencia altamente texturada. Esto provoca la existencia de gradientes altos en regiones interiores. El operador de Nagel–Enkelmann identificará estos gradientes como discontinuidades en la imagen generando una sobresegmentación en el flujo. Para evitar que esta sobresegmentación afecte a la precisión del mapa de disparidad se ha aumentado el valor de υ para que nuestro término de suavizado tenga un comportamiento isotrópico ($\upsilon = 1{,}0$) y, así no penalizar en las zonas homogéneas. Como es de esperar, la difusión aplicada se propagará más allá del contorno del objeto rectangular aumentando el error en estas zonas. Para contrarrestar el efecto de la difusión se ha reducido el valor del parámetro de regularización, $C = \{0{,}1, 0{,}3\}$. Al igual que ocurría en el par anterior no se ha observado diferencias significativas entre los resultados cuando se han utilizado varias escalas. En las estimaciones que se muestran en las tablas el número de escalas es cero.

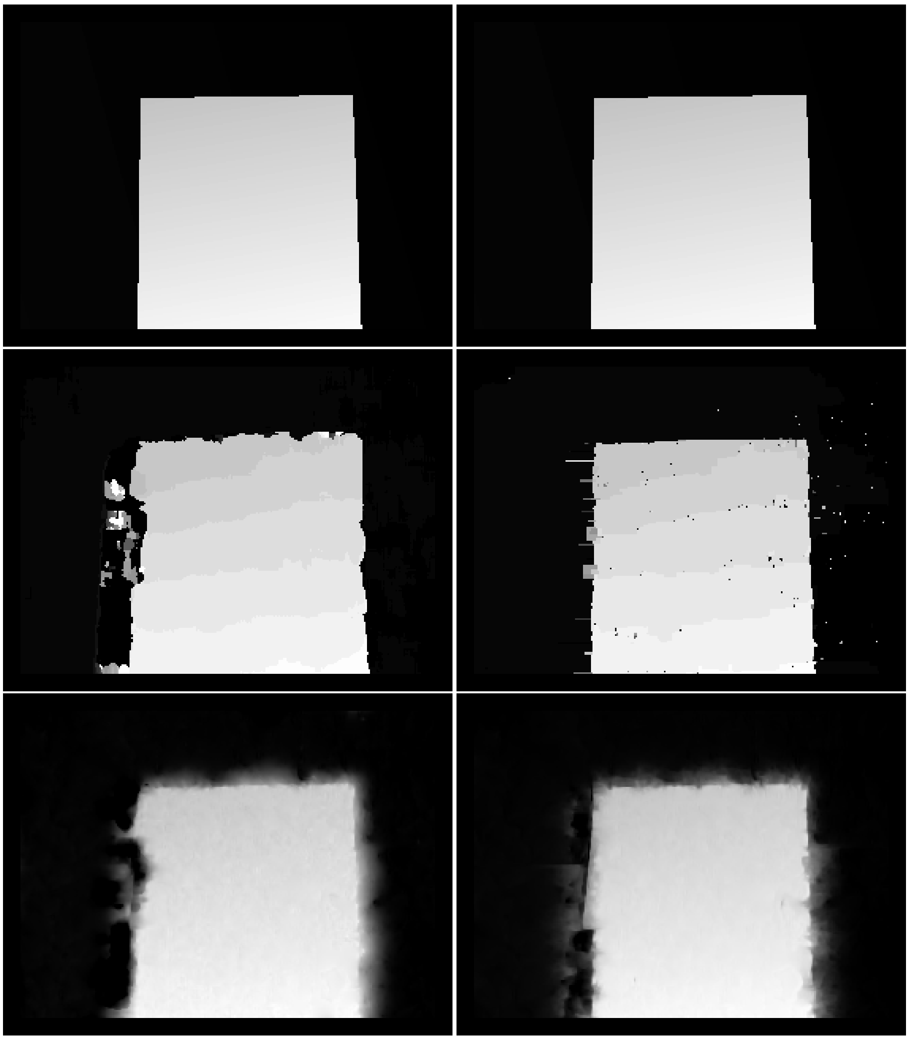


Figura 3.20: En la primera fila, tenemos por duplicado el mapa de disparidad real. En la segunda fila, las aproximaciones iniciales calculadas con los métodos de correlación y graph–cuts, respectivamente. En la última fila, los mapas de disparidad obtenidos por los métodos *Corr+SF* y *kz2+SF*.

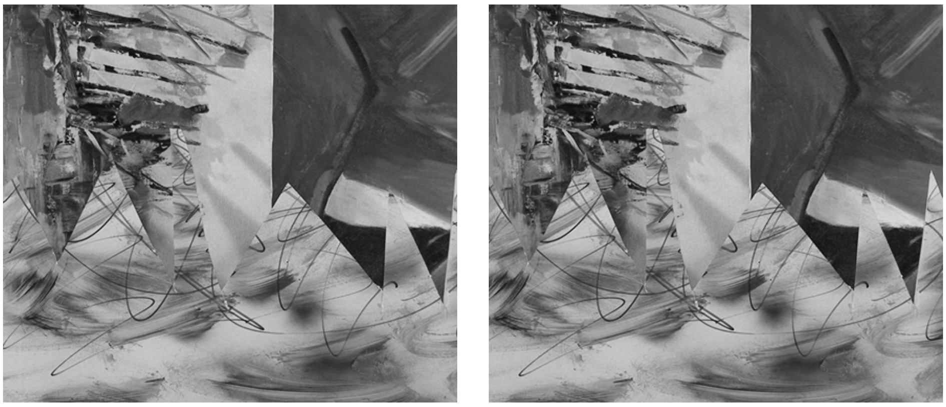


Figura 3.21: Par estéreo de *Sawtooth*.

Par de *Sawtooth*

El par estéreo *Sawtooth* debe su nombre a una textura grisácea situada en la parte central de la imagen que se asemeja a los dientes de una sierra (fig. 3.21). En el apartado 1.4.2 podemos encontrar información detallada acerca de este par.

En la figura 3.22 se observa los mapas de disparidad real y los calculados por los distintos métodos presentados en esta sección. Al igual que ocurría con las secuencias anteriores en el interior de los objetos el mapa de disparidad es muchísimo más suave con el método variacional y se sigue produciendo el efecto escalera con el de graph–cuts. El método *kz2* identifica con bastante precisión las discontinuidades de los objetos y, en el caso del método variacional, los bordes aserrados se ven ligeramente difuminados. La solución obtenida con la inicialización con la técnica de correlación tiene algunos artificios tanto en las discontinuidades como en el interior de los objetos y la identificación de los bordes del objeto aserrado no se ajusta al desplazamiento real. La utilización de distintos valores en los parámetros de configuración de la técnica de correlación no consiguieron la eliminación de estos artificios. La aproximación inicial obtenida por la correlación fue calcula en el primer nivel de escalado en el enfoque multipiramidal. Para poder comparar las soluciones de todos los métodos se ha estimado el mapa de disparidad de la correlación en el nivel superior, con las imágenes de entrada. Los problemas manifestados en el par de Venus se vuelven a repetir en este caso.

En la tabla 3.7 se presentan los resultados cuantitativos que nuevamente ponen de manifiesto que el método *kz2+SF* funciona mejor que el graph–cuts (*kz2*) y que *Corr+SF*. El método variacional compensa la precisión del cálculo del flujo en zonas homogéneas interiores con los pequeños errores en la difusión producida en las discontinuidades. Si establecemos *Corr+SF* como referencia, el AEE es 55,29 % peor que *kz2+SF* y un 41,35 % para el *kz2*. Nuevamente se observa que la solución dada por *kz2+SF* es muchísimo mejor salvo en las discontinuidades. Si analizamos el AAE, *Corr+SF* es un 47,11 % y un 49,71 % peor comparado con *kz2* y *kz2+SF*, respectivamente. Nuevamente el método variacional es capaz de mejorar la estimación dada por su inicialización. La mala aproximación inicial dada por la técnica de correlación condiciona a la solución obtenida por el método variacional.

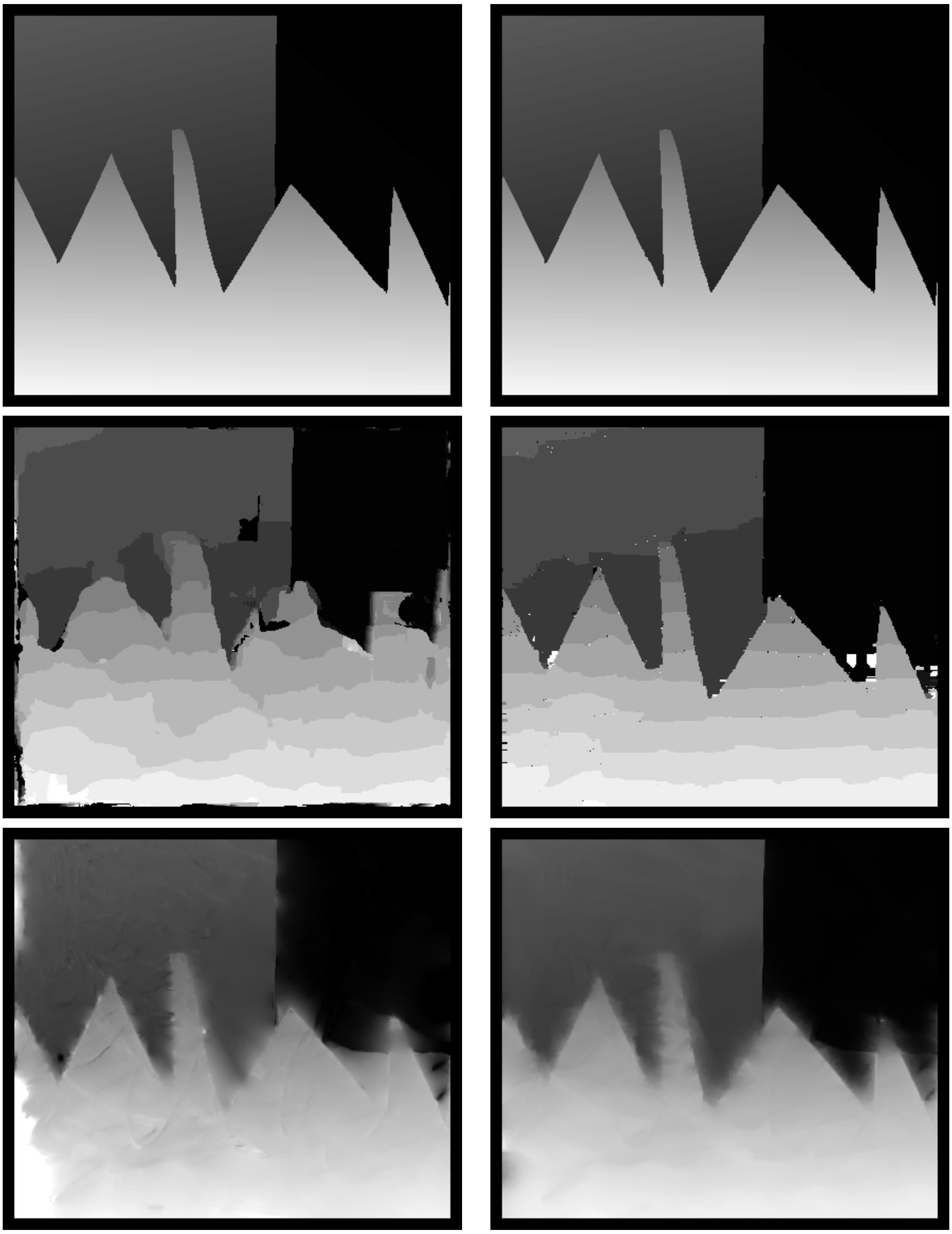


Figura 3.22: En la primera fila, tenemos por duplicado el mapa de disparidad real. En la segunda fila, las aproximaciones iniciales calculadas con los métodos de correlación y graph–cuts, respectivamente. En la última fila, los mapas de disparidad obtenidos por los métodos *Corr+SF* y *kz2+SF*.

Sawtooth				
Método	**AEE**	**% mejora**	**AAE**	**% mejora**
Corr ($sc = 1$)	0.543393	38.20 %	1.73947	332.05 %
kz2	0.23059	-41.35 %	0.21293	-47.11 %
Corr+SF ($C = 0{,}1, \upsilon = 0{,}5,\ sc = 1$)	0.393187	0.00 %	0.40261	0.00 %
***kz2*+SF** ($C = 0{,}9, \upsilon = 1{,}0$)	0.17578	-55.29 %	0.20246	-49.71 %

Cuadro 3.7: Par estéreo de *Sawtooth*: AEE y AAE obtenidos con los métodos *Corr*, *kz2*, *Corr+SF* y *kz2+SF*.

El par de *Sawtooth* se trata de una secuencia muy parecida a *Venus*. *Sawtooth* es un par compuesto por imágenes texturadas donde se aprecia con bastante nitidez las discontinuidades de los objetos. Por lo tanto, debemos suavizar bastante en las zonas interiores a los objetos y detener ese suavizado en los contornos de los objetos. Por este motivo, los valores utilizados en los parámetros del método variacional son: $C = 0{,}9$ (peso alto del término de suavizado) y $\upsilon = 1{,}0$ (el comportamiento isotrópico). Para identificar los contornos de los objetos correctamente deberíamos haber aplicado un suavizado anisotrópico. Sin embargo, en los experimentos se observó que la diferencia en las discontinuidades no eran notables y en las zonas interiores a los objetos se producía una sobresegmentación del flujo cuando el valor de υ era próximo a 0.1. Por ello, se prefirió penalizar en los bordes y no en las regiones interiores.

Par de *Tsukuba*

El par estéreo de *Tsukuba* capta la escena de un despacho (ver figura 3.23). En ella, hay varios objetos: una lámpara naranja, un busto blanco sobre una caja de cartón, una videocámara sobre un trípode y, al fondo, unas estanterías con libros. Pese a tratarse de una secuencia real se ha creado un mapa de disparidad en precisión entera que expresa los desplazamientos de los objetos anteriores. Como es lógico, esta disparidad es ficticia y no expresa con detalle el desplazamiento real de los píxeles de los objetos. Por ejemplo, la disparidad de los píxeles de la lampara es constante. Sin embargo, al no tratarse de una superficie plana, en un escenario real, este desplazamiento no puede ser constante y varía ligeramente en función de la profundidad de los píxeles respecto a las cámaras. En el apartado 1.4.2 se dispone de más información acerca de esta secuencia.

En la figura 3.24 se observa los mapas de disparidad *pseudo-real* y los obtenidos por los distintos métodos. Como se puede ver a simple vista, las diferencias entre el mapa *pseudo-real* y el calculado por *kz2* son mínimos. Nuevamente, con el método variacional, el interior de los objetos es más suave y realista pero los bordes están algo difuminados, restando precisión al compararlo con el mapa *real*. La solución obtenida por *Corr+SF* no es muy satisfactoria y, otra vez se percibe algunos artificios en el fondo de la escena. La estimación de *Corr+SF* pone de manifiesto, una vez más, que la técnica de correlación empleada no ofrece una buena aproximación inicial. Esta inicialización fue estimada en la escala uno del enfoque multipiramidal. Esta solución sufre las mismas deficiencias percibidas en los experimentos anteriores. A la hora de configurar los parámetros de la técnica de correlación

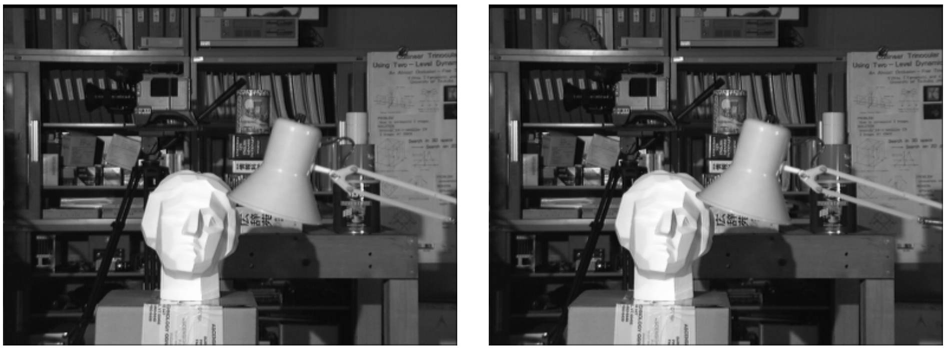


Figura 3.23: Par estéreo de *Tsukuba*.

se ha considerado que para el resultado global era más conveniente la penalización por un tamaño de ventana grande que por las malas estimaciones de una ventana de comparación pequeña. La aproximación inicial de la correlación mostrada en la figura 3.24 ha sido calculada en la escala cero y así poder establecer una comparación visual en la misma escala entre los distintos mapas de disparidad.

Pese a que la solución de *kz2* es la más parecida al mapa de disparidad *pseudo-real*, no sería útil para reconstruir la escena 3D. El modelo 3D obtenido estaría compuesto por objetos planos, carentes de profundidad; no ajustándose a la realidad. Sin embargo, el mapa ofrecido por el método *kz2+SF* permitiría calcular un modelo 3D de la escena mucho más realista, resaltando la necesidad de combinar ambas técnicas si queremos obtener resultados coherentes en ciertas aplicaciones.

Secuencia del pasillo

El par del pasillo, *Corridor* en inglés, es uno de los más famosos que se han utilizado para la evaluación de los métodos estéreos. Se trata de una secuencia sintética generada a través de un programa de ray tracing. En la escena nos encontramos un pasillo en donde hay una esfera, un cono y un objeto al fondo; todos situados a distintas profundidades respecto a las cámaras. Las cámaras están situadas en posición frontoparalela. En el apartado 1.4.2 se da más detalle acerca de esta secuencia. En la figura 3.25 se puede observar tanto la secuencia original del *Corridor* como los mapas de disparidad de ambas cámaras (de izquierda a derecha y viceversa).

El enfoque de los experimentos realizados con esta secuencia es distinto a los anteriores. Se pretende comparar la precisión y robustez de cuatro métodos (*Corr*, *kz2*, *kz2+SF* y *Corr+SF*) ante la presencia de ruido. Se utilizarán tres versiones del *Corridor*: una sin ruido y, otras con dos con un ruido gaussiano de media cero y de varianza de 10 y 100, respectivamente. Otra diferencia de estos experimentos está en la precisión del mapa de disparidad de referencia (*ground truth*). Ahora disponemos de un *ground truth* con valores reales; ya no se ve perjudicado por el efecto de escalado presente en las secuencias de Middlebury.

El método *kz2* funciona muy bien cuando las superficies de los objetos son paralelas al plano de proyección de las cámaras ya que todos los píxeles de dicha superficie están a

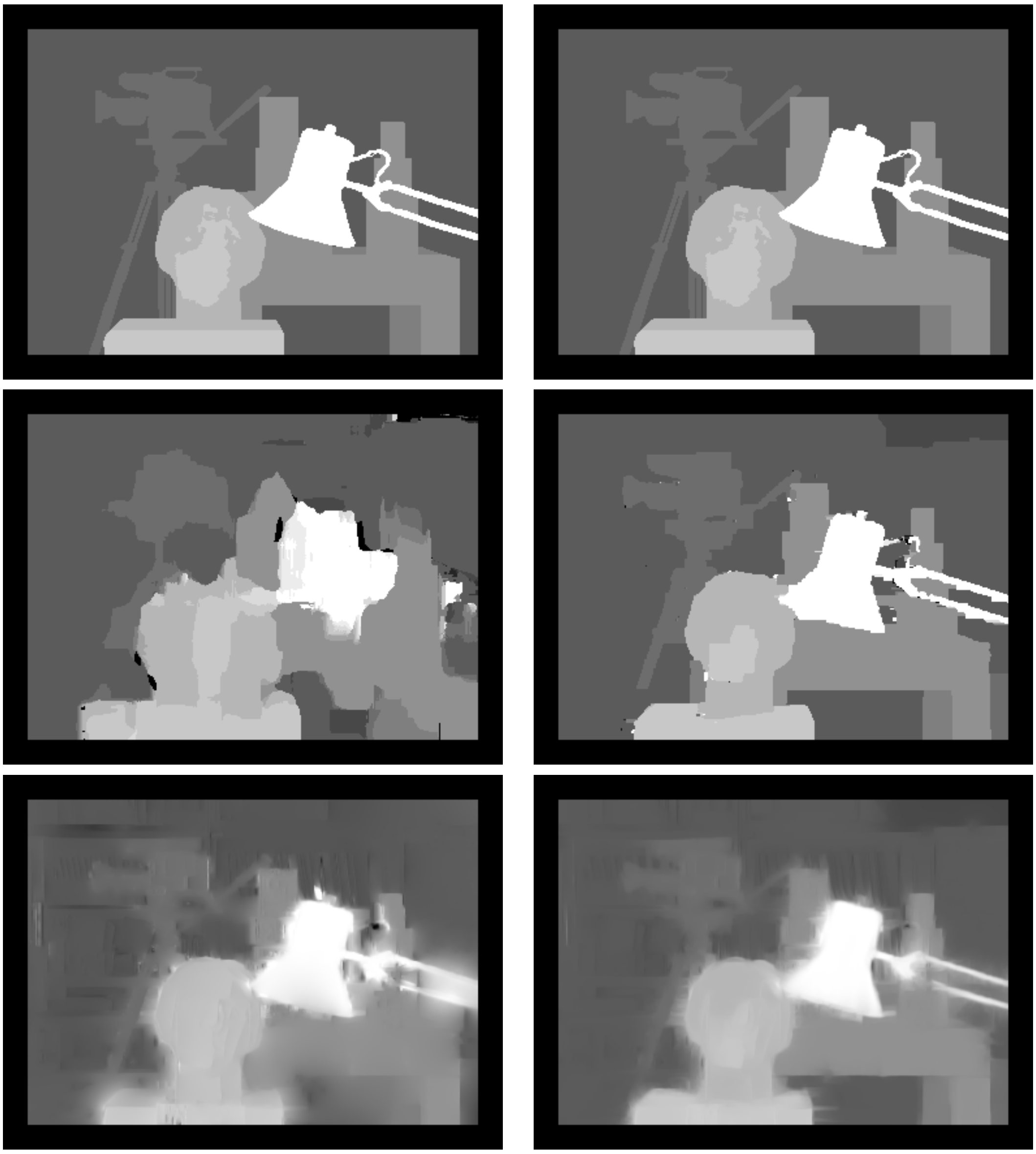


Figura 3.24: En la primera fila, tenemos por duplicado el mapa de disparidad real. En la segunda fila, las aproximaciones iniciales calculadas con los métodos de correlación y graph–cuts, respectivamente. En la última fila, los mapas de disparidad obtenidos por los métodos *Corr+SF* y *kz2+SF*.

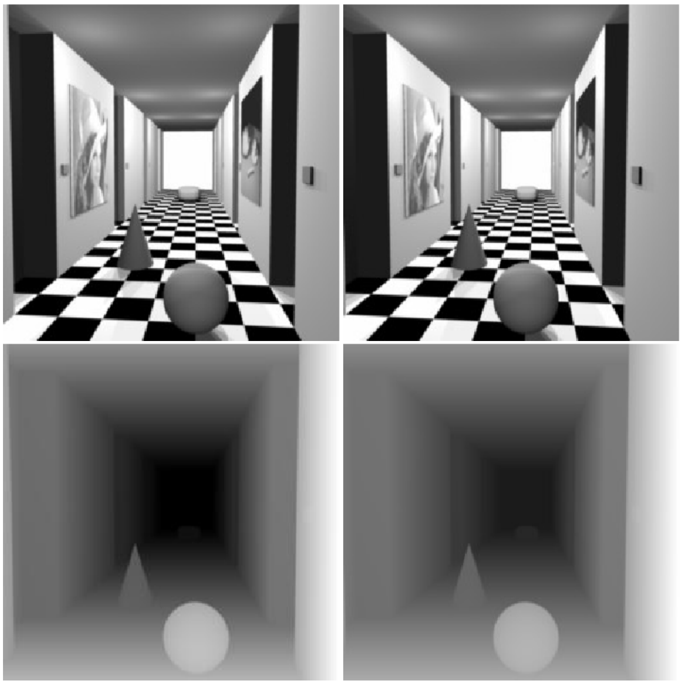


Figura 3.25: En la parte superior, la imagen izquierda y derecha del par del *Corridor*. En la parte inferior, el mapa de disparidad real en ambos sentidos.

la misma profundidad. En la secuencia del *Corridor* este fenómeno no ocurre en todas las superficies; hecho que hará que las soluciones *kz2* no sean óptimas. La disparidad calculada con *kz2* en las superficies no paralelas provocará un efecto de escalera. El desplazamiento suave de los píxeles se convertirá en una sobresegmentación del flujo. En estas superficies, la precisión del método variacional será mucho mayor. Además, el método variacional es capaz de atenuar la influencia de los píxeles ruidosos mejorando aún más la precisión del mapa de disparidad. Sin embargo, el método de graph–cuts identificará estos píxeles como oclusiones generando una estimación muy segmentada.

En estos experimentos no sólo se compara la precisión de los métodos sino que se quiere poner de manifiesto la robustez frente al ruido del método variacional. Vamos a presentar resultados tanto cuantitativos como cualitativos de los mapas de disparidad obtenidos. En la figura 3.26 se muestran las distintas aproximaciones iniciales utilizadas en el método variacional para las tres versiones del par del *Corridor*. En la parte superior, la inicialización suministrada por el método de graph–cuts, mientras que en la inferior, la dada por la técnica de correlación. En las tablas 3.8 y 3.9 se muestra un resumen con los AEE y AAE obtenidos con los tres métodos en las distintas versiones del *Corridor* (sin ruido, con ruido de 10 y 100). En cada casilla también se indican los valores de los parámetros utilizados para configurar el método variacional. Los parámetros C y υ son muy estables por lo que pequeñas variaciones en sus valores no ofrecerán grandes diferencias en el resultado. En las figuras 3.28 y 3.29 se muestran dos gráficas que representan los AEE y AAE de los distintos métodos para las distintas variantes del par del pasillo. En ellas, se observa la estabilidad ofrecida por el método variacional frente a la solución obtenida por el método de graph–cuts, altamente sensible al ruido. A medida que el ruido aumenta en el pasillo el aumento del error euclídeo y angular de los métodos *Corr+SF* y *kz2+SF* es insignificante. Pese a que las inicializaciones ofrecidas por *kz2* y *Corr* no son buenas, *StereoFlow* permite alcanzar mapas de disparidad de bastante precisión. Como es de esperar las aproximaciones iniciales calculadas con *kz2* y *Corr* empeoran con la presencia de ruido en las imágenes. Cuando la varianza del ruido es alta se forman pequeñas regiones en los mapas de disparidad debido a esos píxeles ruidosos. Una forma de eliminar o atenuar la influencia de esos píxeles es la aplicación de un filtro mediana a la aproximación inicial antes de que sea utilizada por el método variacional. En el mapa de disparidad obtenido con la correlación se aprecia cómo esta técnica es menos sensible al ruido si lo comparamos con la solución ofrecida por el método *kz2*.

El método StereoFlow permite la recuperación de los desplazamientos largos en la escena. Para ello, se ha utilizado un enfoque multipiramidal donde un parámetro a configurar es el número de escalas; el factor de reducción a aplicar en cada escala se ha fijado a 0,5. El método variacional, actuando en solitario, puede ofrecer buenas soluciones cuando el número de escalas es suficientemente grande y los desplazamientos en el nivel más inferior son de menos de un píxel. Sin embargo, si se dispone de una *ayuda* en forma de aproximación inicial podemos reducir el número de escalas aumentando el rendimiento global del método y la precisión de las soluciones. En los experimentos hemos utilizado sólo dos escalas para *Corr+SF* ya que un número superior de escalas no ofrecía mejores resultados.

La inicialización ofrecida por *kz2* se ve claramente afectada por el escalonamiento que ocurre en los métodos de graph–cuts cuando la disparidad de una superficie no es constante. Como era de esperar, esta inicialización no es demasiado precisa y se ve empeorada ante la presencia de ruido. Para corregir las malas estimaciones de la aproximación inicial se optó por aumentar el número de escalas del esquema multipiramidal cuando el ruido era alto. En los casos de ruido de baja intensidad con una o ninguna escala el método variacional se comportaba bien y era capaz de obtener buenas soluciones.

En la figura 3.27 se muestran los resultados visuales obtenidos con *Corr+SF* y *kz2+SF* utilizando los pares del *Corridor* con ruido de distinta intensidad. Si miramos con detenimiento la disparidad ofrecida por *kz2+SF* observamos que es mucho más suave. En todos los experimentos se percibe que la combinación *kz2+SF* es claramente más ventajosa que las otras alternativas llegando a ofrecer en los pares con ruido mejoras superiores a 15 % y 14 % para el AEE y AAE, respectivamente.

Los valores de los parámetros del método variacional determinan su comportamiento. El método buscará una solución que satisfaga las ecuaciones de Euler-Lagrange y que se encuentre próxima a la inicialización. Dicho de otra forma, el método se apoyará en la aproximación inicial para buscar una solución que cumpla una serie de restricciones. Por lo tanto, los valores de los parámetros del método permitirán sacar provecho a la inicialización y/o corregirá ciertas deficiencias que se encuentre en ella. El término de regularización que utiliza nuestro método se basa en la información del gradiente de la imagen. En imágenes muy texturadas, un valor de υ próximo a cero (anisotrópico) provocaría una sobresegmentación en el flujo. En estos casos conviene darle a υ valores no cercanos a cero para que la estimación no penalice demasiado en las zonas homogéneas. Sin embargo, en imágenes con las discontinuidades bien definidas sería deseable disponer de un regularizador anisotrópico. En la mayoría de las situaciones se optará por un valor

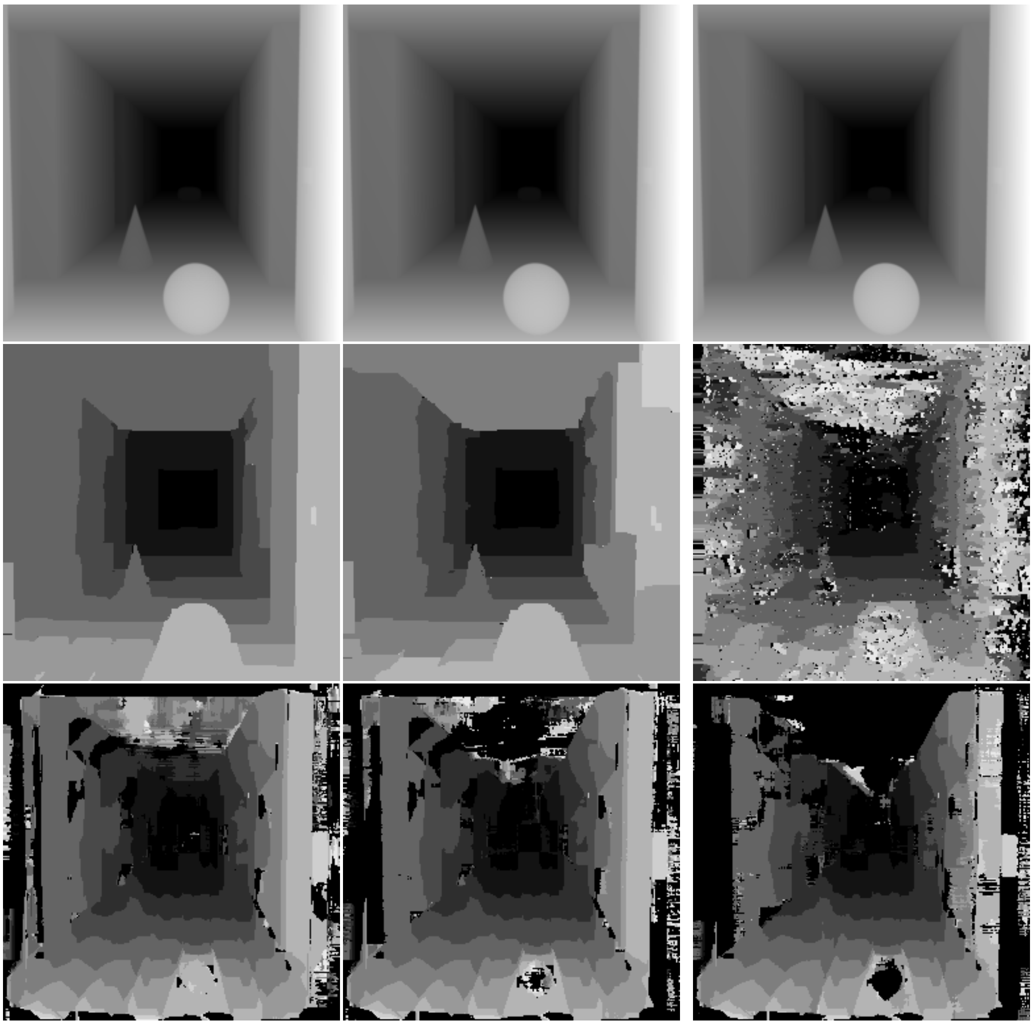


Figura 3.26: Aproximaciones iniciales obtenidas con las tres variantes del *corridor* con el método de graph–cuts (en la parte superior) y con la técnica de correlación (en la parte inferior).

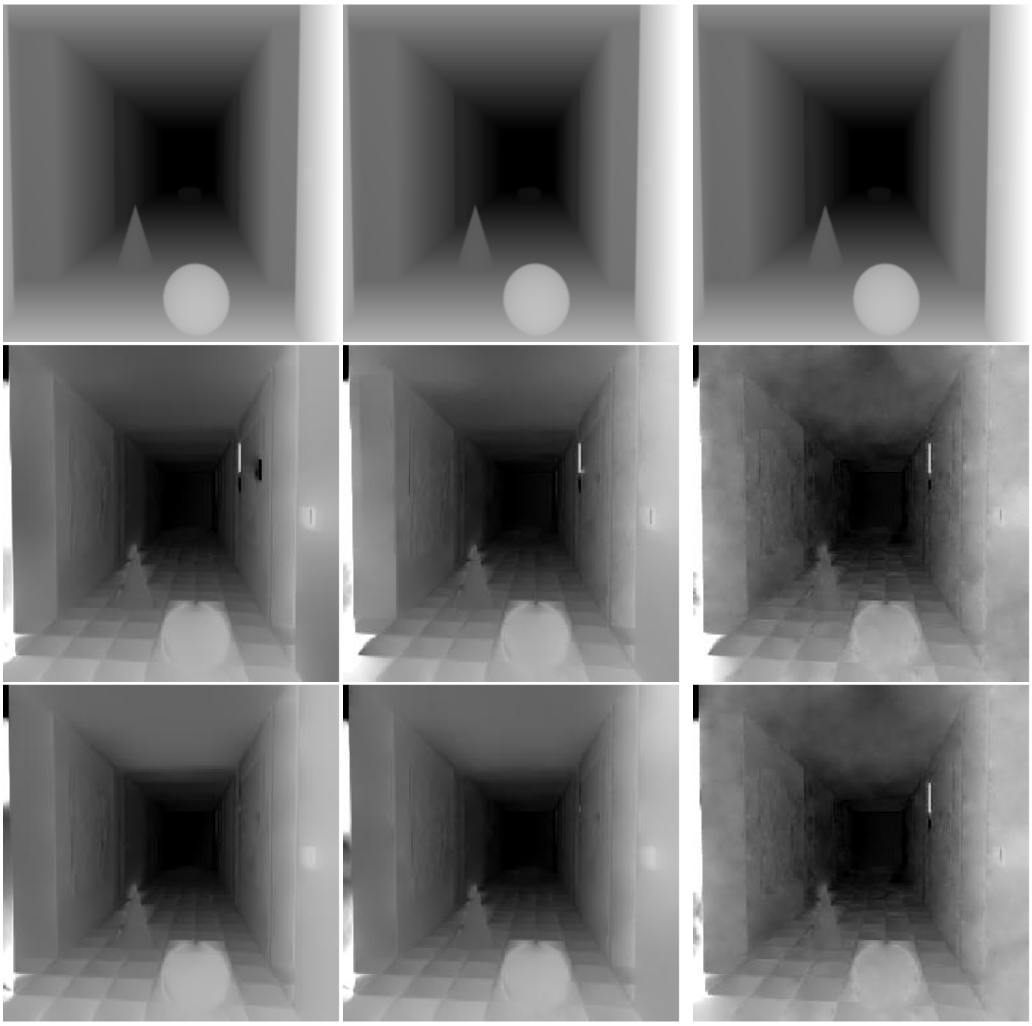


Figura 3.27: De izquierda a derecha los mapas de disparidad obtenidos utilizando las tres variantes del par del pasillo (sin ruido, con ruido=10 y 100). En la parte superior, el mapa de disparidad real. En la parte central, los malos mapas obtenidos con el método *kz2+SF*. En la parte inferior, los mapas de disparidad ofrecido por el método $Corr + SF$.

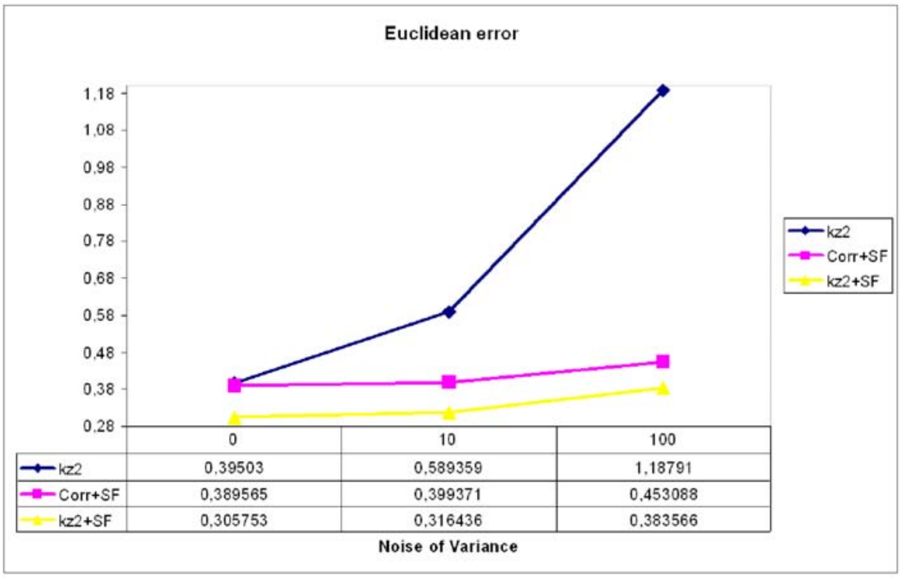


Figura 3.28: AEE obtenido en el par *Corridor* en las tres variantes presentadas (sin ruido y con ruido de varianza 10 y 100).

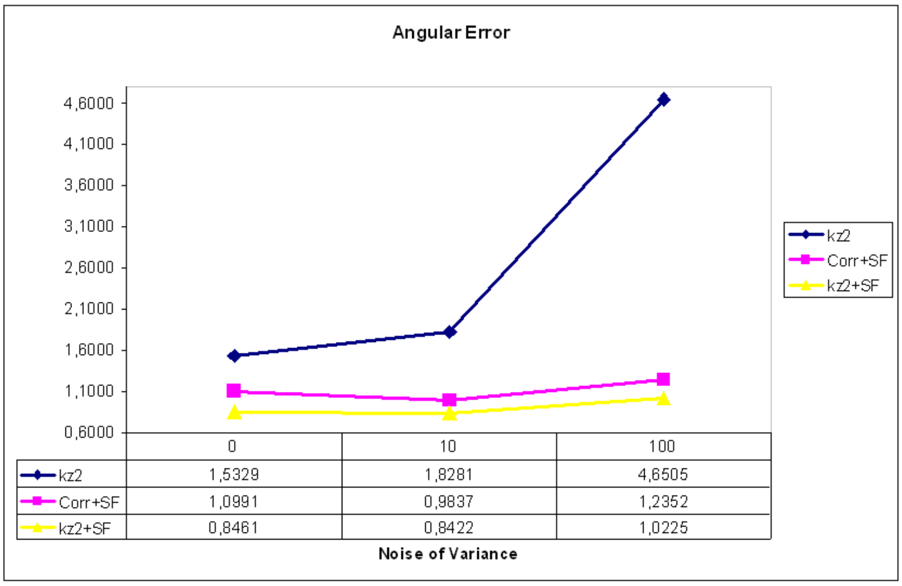


Figura 3.29: AAE obtenido en el par *Corridor* en las tres variantes presentadas (sin ruido y con ruido de varianza 10 y 100).

Par del Corridor	**AEE**		
Método	**Sin ruido**	$\sigma = 10$	$\sigma = 100$
Corr	0.8878	1.3653	1.7859
kz2	0.3950	0.5893	1.1879
Corr+SF ($C = 0{,}1$)	0.3895 (1) ($\upsilon = 0{,}5$)	0.3993 (2) ($\upsilon = 1{,}0$)	0.4530 (2) ($\upsilon = 1{,}0$)
***kz2*+SF** ($C = 0{,}1$)	0.3057 (0) ($\upsilon = 0{,}5$)	0.3164 (0) ($\upsilon = 0{,}5$)	0.3835 (2) ($\upsilon = 0{,}1$)

Cuadro 3.8: Par estéreo de *Corridor*: AEE obtenido con los métodos *Corr*, *kz2*, *Corr+SF* y *kz2+SF*.

Par del Corridor	**AAE**		
Método	**Sin ruido**	$\sigma = 10$	$\sigma = 100$
Corr	7.5422	15.4	22.4235
kz2	1.5329	1.8281	4.6505
Corr+SF ($C = 0{,}1$)	1.0991 (1) ($\upsilon = 0{,}5$)	0.9837 (2) ($\upsilon = 1{,}0$)	1.2352 (2) ($\upsilon = 1{,}0$)
***kz2*+SF** ($C = 0{,}1$)	0.8461 (0) ($\upsilon = 0{,}5$)	0.8422 (0) ($\upsilon = 0{,}5$)	1.0225 (2) ($\upsilon = 0{,}1$)

Cuadro 3.9: Par estéreo de *Corridor*: AAE obtenido con los métodos *kz2*, *Corr+SF* y *kz2+SF*.

de υ intermedio (por ejemplo, 0,5); se trata de un compromiso entre el comportamiento isotrópico y anisotrópico.

Una inicialización que identifique con precisión las discontinuidades facilitaría a nuestro modelo de energía la identificación de las correspondencias en los contornos de los objetos. Dada mala aproximación de la técnica de correlación ha obligado a incrementar el valor de υ (de 0.5 a 1.0) para que el regularizador de Nagel–Enkelmann tenga un comportamiento isotrópico y compensar los errores que pueda tener la aproximación inicial. Cuando la aproximación inicial fue aportada por *kz2* se mejoró sensiblemente la identificación de las discontinuidades del cono y esfera. Sin embargo, se produjo una sobresegmentación de las paredes del pasillo lo que obligó suavizar la inicialización. Por ello, υ tiene un valor intermedio 0,5.

El parámetro C especifica la importancia del término de regularización respecto al del ligadura. Cuando las estimaciones del término de ligadura son buenas no es necesario que el de regularización actúe. En caso contrario, el término de suavizado debería actuar con firmeza. En la secuencia del *Corridor* variaciones de C no mejoraban sensiblemente las estimaciones obtenidas. Por ello, se optó por asignarlo un valor bajo ($C = 0{,}1$) ya que dado que υ tiene valores superiores a 0,5 (isotrópico) no convenía que el proceso de difusión se propagará más allá de las discontinuidades.

3.4. Conclusiones

3.4.1. Estimación de la Disparidad a partir de una Secuencia Estereoscópica

En este trabajo se ha desarrollado un método variacional para la estimación de los mapas de disparidad de una secuencia de vídeo estereoscópica. Para ello, se ha combinado la información aportada por el flujo óptico de cada secuencia y el flujo estéreo calculado usando cada par estéreo. Con el objetivo de obtener una solución consistente se ha introducido una restricción temporal entre el flujo óptico y estéreo, de forma que se ha establecido una relación entre ambos flujos. El método propuesto se trata de una extensión de un método estéreo Alvarez/Deriche/Sánchez/Weickert (2002) [Alvarez02b] y otro de flujo óptico Alvarez/Weickert/Sánchez (2000), [Alvarez00]. El modelo de energía resultante ha combinado las ideas reflejadas en ambos métodos.

En la evaluación de este método se han utilizado tres secuencias de pares estéreo, dos sintéticas y una real. Los métodos que combinan información estéreo y flujo no abundan en la literatura por lo que la disponibilidad de secuencias de pares estéreo es bastante reducida. Normalmente existen bases de datos con secuencias de imágenes o pares estéreos sueltos. La secuencias sintéticas utilizadas consisten en el desplazamiento horizontalmente de una serie de objetos delante de las cámaras. Con estas secuencias se ha querido comprobar la precisión de nuestro método. Nos encontrábamos ante una situación ideal, objetos texturados con un movimiento simple, por lo que su detección sería sencilla. El resultado obtenido fue muy satisfactorio ya que se alcanzaron errores de magnitud subpíxel. Para evaluar la calidad del método en condiciones más complejas se decidió utilizar una secuencia real de un partido de fútbol celebrado en el MiniEstadi. Con dos cámaras situadas en la grada se captó el caminar de un portero por el área. Esta secuencia tiene mayor complejidad que la anterior ya que (1) las condiciones de iluminación no están bajo control y, además, (2) las cámaras no están en posición frontoparalela. Desgraciadamente nuestro método no incorpora ninguna técnica que lo haga insensible frente a cambios en la iluminación por lo que esta perturbación produjo errores en la estimación. Para corregir la orientación de las cámaras y simplificar el cálculo de la disparidad se optó por la rectificación de las imágenes reales. De esta forma el desplazamiento de los objetos, en este caso del portero, sólo ocurriría en una dirección. Los resultados obtenidos con esta secuencia reflejan el buen funcionamiento del método, pese a la ausencia de técnicas de robustificación e invarianzas insensibles a los cambios de iluminación en el modelo de energía.

3.4.2. Estimación del Mapa de Disparidad mediante la Combinación de un Método de Graph–cuts y uno Variacional

En este trabajo se han combinado dos técnicas para la estimación del mapa de disparidad usando pares estéreo con el objetivo de aumentar la precisión y la robustez de los resultados. Se ha utilizado el método variacional propuesto en Alvarez/Deriche/Sánchez/Weickert (2002) [Alvarez02b] y un método de graph–cuts, [Kolmogorov01, Boykov04]. Como inicialización de los métodos variacionales se suele recurrir a una técnica basada en correlación. Una posible alternativa a este tipo de técnica podría ser los métodos de *graph–cuts*. En los últimos años, han demostrado su *precisión* en el problema estéreo. El modelo de energía de [Kolmogorov01, Boykov04] incorpora un término que detecta las oclusiones, hecho que lo convierte en el complemento idóneo para nuestro método variacional.

El método variacional de Alvarez/Deriche/Sánchez/Weickert (2002) [Alvarez02b] permite la detección de cualquier desplazamiento presente en la secuencia. Para ello, se apoya en un enfoque multipiramidal. En la escala más pequeña es donde el método de *graph–cuts* interviene calculando el mapa de disparidad que se utilizará como la aproximación inicial. A partir de ahí entra en funcionamiento el método variacional refinando en cada nivel la estimación hecha por la escala inferior.

En los experimentos se han perseguido dos objetivos: (1) por un lado, se ha querido analizar la influencia de las distintas aproximaciones iniciales, *graph–cuts* y técnica de correlación; y (2) por otro lado, se ha querido comprobar la robustez de los distintos métodos ante la presencia en la secuencia de ruido de distinta intensidad. Para la consecución del primer objetivo se ha utilizado varias secuencias disponibles en la base de dados de Middlebury (versión 2001). Las soluciones ofrecidas por el método de *graph–cuts* ponen de manifiesto su buen comportamiento. Este método detecta con gran exactitud las discontinuidades en la imagen pero su naturaleza discreta le resta precisión en los resultados. Sin embargo, al método variacional le ocurre lo contrario, las estimaciones en las regiones interiores de los objetos son muy suaves pero en los contornos el proceso de difusión se propaga más allá, reduciendo su precisión. Al analizar los resultados experimentales debemos hacerlo teniendo en cuenta que los mapas de disparidad de referencia ofrecidos por Middlebury están escalados por lo que implícitamente introducen un pequeño error. Para comprobar la robustez de los distintos métodos ante la presencia de ruido se ha utilizado la secuencia del *Corridor*. En este par se percibe una de las limitaciones del método de graph–cuts: no detectan correctamente la disparidad en superficies que no están a una distancia constante respecto a la vista de las cámaras. Pese a ello, la inicialización ofrecida por este método ofrece mejores resultados que la técnica de correlación. Los resultados demuestran que el método variacional es bastante robusto frente al ruido y que pese a que las inicializaciones no son del todo precisas es posible obtener muy buenos resultados.

Para finalizar podemos afirmar que desde el punto de vista teórico la combinación de los métodos variacionales y graph–cuts parece bastante acertada ya que cada modelo de energía incorporan un mecanismo, como son la detección de oclusiones y precisión subpíxel, que cubre una carencia o limitación de la otra técnica. Desde el punto de vista práctico el esfuerzo necesario para implementar un método de graph–cuts hace que en muchas aplicaciones convenga el empleo de una técnica de correlación básica.

Capítulo 4

Conclusiones

4.1. Resumen

El objetivo de esta tesis era el desarrollo de una serie de *métodos variacionales para la estimación del flujo óptico y mapas de disparidad*. En los trabajos expuestos en este documento se ha intentado contribuir a la literatura siguiendo tres estrategias: (i) la incorporación pequeñas modificaciones a métodos ya existentes, (ii) la combinación de métodos con el fin de obtener lo mejor de cada uno de ellos y, (iii) el diseño nuevos modelos de energía que fueran fácilmente adaptables y extensibles.

Los métodos de flujo óptico desarrollados en esta tesis se han centrado en el estudio de la influencia de la información temporal frente a métodos espaciales tradicionales. Por otro lado, los métodos para la estimación del mapa de disparidad se han enfocado en el análisis de la mejora que se produce a través de la combinación de métodos existentes. Es decir, a partir de métodos de cierto éxito y gran precisión se ha considerado que su fusión podría mejorar las estimaciones que ofrecen por separado. Un objetivo que se ha dejado en un segundo plano ha sido la mejora del rendimiento de las implementaciones de los métodos propuestos en esta tesis.

En ciertos campos, como la meteorología y la climatología, el uso de imágenes satélites ha supuesto toda una revolución. La estimación del desplazamiento de las estructuras nubosas en la atmósfera se trata de un problema prioritario en este campo. Para solucionar este problema se sigue recurriendo a métodos basados en técnicas de correlación. Los métodos variacionales no gozan del éxito y la expansión cosechada en problemas similares. El método variacional propuesto por [Alvarez00] se trata de una variante del modelo de energía de Nagel–Enkelmann en el que se ha introducido una serie de mejoras. Este método ha sido embebido en una estrategia multipiramidal para evitar que el algoritmo converja hacia mínimos locales irrelevantes. Este método permite recuperar desplazamientos de más de diez píxeles con una buena precisión, supone una mejora significativa respecto al método original de Nagel y es bastante robusto. Aprovechando las ventajas ofrecidas por este método se ha decidido extenderlo para que utilice información multicanal y aplicarlo en la detección del movimiento de las nubes en imágenes de satélites multisensoriales. A la hora de incorporar información multicanal al método de [Alvarez00] se deben aplicar una serie de modificaciones. El término de ligadura se ha definido mediante una suma ponderada de la suposición lambertiana de cada uno de los canales. Esta solución es similar a la ofrecida por otros métodos multicanales de la literatura. En el término de suavizado radica la aportación más novedosa de este método. El operador de Nagel-Enkelmann se caracteriza por usar el gradiente de la imagen como *métrica* para designar la dirección y la cantidad de difusión a aplicar en cada píxel de la imagen. La utilización de varias imágenes supone la creación de una estrategia diferente ya que ahora ese gradiente debe aglutinar información de todos los canales disponibles. Esto abre un abanico de posibilidades a la hora de establecer el criterio de construcción de dicho gradiente. En nuestro modelo se ha definido dos estrategias que consideramos las más apropiadas. En la primera estrategia, el gradiente en cada píxel se calcula como el mayor en magnitud de los gradientes de todos los canales. En la segunda, se calcula como la dirección dominante de los gradientes de todos los canales. A partir del modelo de energía se obtuvo el sistema de ecuaciones de Euler-Lagrange y en la discretización se ha utilizado un esquema implícito de diferencias finitas. Para poder evaluar con más objetividad los resultados ofrecidos por este método se crearon dos secuencias sintéticas a partir de las imágenes multisensoriales del satélite usando un modelo de movimiento realista calculado con un método de correlación. Los resultados experimentales demostraron la mejora que ofrecían las modificaciones realizadas al modelo original y que la utilización de métodos variacionales en imágenes de satélites podían tener éxito.

Uno de los primeros trabajos de la tesis fue el desarrollo de un método variacional que incorporaba un novedoso término de regularización exclusivamente temporal. Este nuevo término desacoplaba la regularización del dominio espacial respecto al temporal permitiendo que el proceso de difusión actuase de forma independiente en cada dominio. Algunos autores han propuesto regularizadores que tratan de forma conjunta la información espacial y temporal. Nuestro método se trata de una modificación del modelo de energía propuesto por Nagel-Enkelmann [Nagel86]. Su diseño flexible da la posibilidad de cambiar el modelo de energía con distintas estrategias temporales favoreciendo la detección de ciertos desplazamientos. Nosotros supusimos la existencia de cierta constancia en el flujo óptico entre frames vecinos premiando cualquier movimiento constante y uniforme registrado en la secuencia de imágenes. Esta suposición dio lugar a dos posibilidades: (i) constancia del flujo en una sola dirección o (ii) en ambas direcciones. Dada que la primera estrategia es un caso particular de la segunda, se minimizó y discretizó el modelo de energía de esta segunda estrategia. La solución se obtuvo al aplicar la técnica de descenso del gradiente en el sistema de ecuaciones de Euler–Lagrange. Para ello discretizamos el sistema de ecuaciones parabólico por diferencias finitas. Las derivadas espaciales se han aproximado por diferencias centradas y para la discretización en la dirección temporal mediante un esquema lineal implícito. Al igual que se ha hecho en todos los métodos desarrollados en esta tesis la detección de los desplazamientos largos se ha podido llevar a cabo gracias al uso de un enfoque multipiramidal. La restricción temporal impuesta sobre el flujo requiere el cálculo del *flujo inverso*. A diferencia del flujo directo el inverso tiene valores únicamente en algunas posiciones discretas. Para calcular el valor del flujo en cada píxel es necesario hacer un promediado de las porciones del flujo que caen sobre cada píxel. En los experimentos realizados con este método se estableció una comparación entre la versión espacial y las dos estrategias comentadas extrayendo como conclusión que el método temporal bidireccional mejoraba sustancialmente las estimaciones respecto a su homólogo unidireccional y espacial. La incorporación del nuevo término de regularización temporal aporta estabilidad a las soluciones del método espacial.

A pesar de todos los métodos propuestos en la literatura la coordinación entre el término de ligadura y el de suavizado seguía sin estar claramente definida; no se había descrito un vía de comunicación explícita entre ambos términos. Por ello, se ha desarrollado de un método espaciotemporal basado en el *análisis espectral del tensor de movimiento* capaz de intercambiar información complementaria entre el término de ligadura y el de regularización. La descomposición espectral del tensor de movimiento establece un mecanismo que indica las direcciones en las que suavizar, las no dominantes del flujo. Para ello, se han utilizado algunas de las técnicas más innovadoras propuestas en la literatura, como es el tensor de movimiento [Bruhn06a] o las funciones de robustificación [Black91]. El resultado de la combinación de estas técnicas da lugar a un *framework* capaz de representar cualquier modelo de una forma compacta, sencilla, fácilmente extensible y adaptable. La inclusión de términos no-cuadráticos aporta robustez y junto a la descomposición espectral del tensor de movimiento permite la construcción de seis modelos de energía distintos: cuatro con funciones de robustificación y dos sin ellas. Los cuatro modelos de energía resultantes se diferencian entre ellos en la disposición de las función de robustificación y los autovalores/autovectores del tensor de movimiento. A partir de cada prototipo se han derivado las ecuaciones de Euler-Lagrange asociadas. Dado que todos los prototipos son similares se ha optado por la descripción del esquema numérico del modelo de energía más completo. Tanto modelo de energía como el tensor de movimiento están definidos en el caso continuo. Sin embargo, nuestro modelo es capaz de detectar los desplazamientos largos gracias al uso de técnicas de *warping*. El uso de estas técnicas no se hace explícito en el modelo de energía sino que se ha pospuesto a la etapa de discretización. Para la resolución del sistema de ecuaciones se ha optado por la combinación de las técnicas de *fixed point iterations* y el método *Lagged-Diffusivity*. Los métodos *Lagged-Diffusivity* ([Kacur68, Fucik73, Chan99]) se basan en la idea de la resolución de un sistema de ecuaciones nolineal a través de su descomposición en un conjunto de problemas lineales. La solución del sistema de ecuaciones de Euler-Lagrange se obtiene mediante el método numérico iterativo *Successive Overrelaxation* (*SOR*). Los experimentos realizados con las distintas variantes de nuestro modelo nos han permitido comprobar que las ventajas que ofrece la elegante descomposición de nuestro método no sólo es apreciable desde punto de vista teórico sino también práctico, obteniendo uno de las mejores estimaciones en la secuencia de *Yosemite con nubes*. Las principales aportaciones realizadas en este método son: (i) el uso de tensores en toda el modelo de energía, (ii) la descomposición espectral de los mismos y (iii) el desacoplamiento de las invarianzas en el término de suavizado. Los nuevos modelos desarrollados tienen la característica de ser fácilmente adaptable y escalables para cualquier tipo de invarianzas.

El otro gran problema tratado en esta tesis corresponde con la estimación de la carta de disparidad. Nuestra mayor contribución ha sido el desarrollo de un método variacional para la estimación de los mapas de disparidad de una secuencia de vídeo estereoscópica. En una secuencia de vídeo estereoscópica las proyecciones de un mismo punto 3D se perciben entre los pares estéreos o a lo largo de las imágenes de una misma cámara. Con la combinación del flujo óptico y estéreo se podría obtener unas estimaciones más robustas. El modelo de energía propuesto combina las ideas reflejadas en el método de flujo óptico [Alvarez00] y flujo estéreo [Alvarez02b]. Para obtener una solución consistente se ha añadido una restricción temporal entre el flujo óptico y estéreo. Gracias a que la geometría epipolar es conocida, nos permite simplificar el modelo de energía así como la ecuación de Euler-Lagrange asociada. Para capturar los desplazamientos grandes y con el fin de evitar que el algoritmo converja a mínimos locales irrelevantes el método se ha introducido en una técnica de análisis multipiramidal. Los resultados experimentales obtenidos con nuestro método son muy prometedores. En el caso de un par estéreo sintético donde se conoce el desplazamiento exacto de los objetos de la escena demostramos que las estimaciones mejoran sensiblemente a las ofrecidas por el método estéreo espacial [Alvarez02b]. La evaluación de este método en pares estéreos de imágenes reales también han ofrecido resultados muy positivos. Este método pretende aprovechar las ventajas de los métodos de flujo óptico y visión estereoscópica para mejorar la estimación de los mapas de disparidad.

La solución en los métodos variacionales se alcanza mediante la minimización de una energía. Una forma de *ayudar* al algoritmo a alcanzar la solución óptima consiste en la inicialización con una estimación relativamente próxima al mínimo global. Normalmente se emplea una técnica basada en la correlación a ventanas. Sin embargo, las buenas soluciones ofrecidas por los métodos de graph–cuts en el cálculo del mapa de disparidad lo erigen como una alternativa. Por este motivo se ha creado un método que consiste en la combinación de uno variacional [Alvarez02b] y una técnica de graph–cuts [Kolmogorov01, Boykov04]. El método variacional [Alvarez02b] se apoya en un enfoque multipiramidal por lo que es capaz de identificar cualquier desplazamiento presente en los pares estéreo. Los puntos fuertes de esta combinación frente a la ofrecida por la correlación está en los buenos mapas de disparidad obtenidos por el método de graph–cuts. Supone una inicialización fiable e incorpora un término que detecta las oclusiones, hecho que lo convierte en el complemento idóneo para nuestro método variacional. Los puntos débiles de esta unión está en la complejidad de la implementación de los métodos de graph–cuts frente a una técnica de correlación convencional. Consideramos que si el término de regularización del método variacional fuera *flow-driven* en vez de *image-driven* el proceso de difusión en los contornos de las imágenes se detendría con mayor precisión y los resultados serían muchísimo mejores. En los experimentos se ha analizado por un lado (i) la influencia de la inicialización dada por la técnica de graph–cuts o la de correlación sobre el método global y, por otro lado, (ii) la estabilidad de la combinación del método variacional y la aproximación inicial ante la presencia de ruido en el par estéreo. Los resultados ponen de manifiesto un ligera mejora en la precisión de las estimaciones dadas por la inicialización del graph–cuts y la gran estabilidad del método variacional.

4.2. Trabajo futuro

Durante esta tesis nos hemos planteado otros trabajos que se podrían desarrollar para mejorar los métodos propuestos, pero por cuestiones de tiempo y esfuerzo no hemos podido realizar. Entre estos podemos destacar:

Términos de regularización temporales. En uno de los trabajos presentados en este documento se ha descrito un término de regularización temporal que se acoplaba a un método espacial. Este término de suavizado asumía cierta constancia en el flujo de los frames vecinos, es decir, favorecía los desplazamientos traslacionales constantes. Una posible mejora supondría la definición de otros términos de regularización temporales que no favorezcan ni presupongan un determinado movimiento de los objetos por la escena.

Nuevas invarianzas. En los términos de ligadura de los modelos de energía descritos en esta tesis se utiliza predominantemente la invarianza de la intensidad de los píxeles en la imagen. En la mayoría de las situaciones esta suposición se viola y el establecimiento de las correspondencias no se realiza correctamente. Se debería emplear un término más complejo que tuviera en cuenta las propiedades de reflectancia de las superficies de la escena y de la posición de las fuentes de luz que inciden sobre los objetos. De esta forma, las soluciones obtenidas en secuencias reales serían mucho más precisas y coherentes. Las distintas invarianzas fotométricas analizadas en [Mileva07] podrían ser utilizadas en los nuevos métodos.

Imágenes a color. Solamente en uno de los trabajos desarrollados se ha utilizado modelos de energía multicanal. La información suministrada por las imágenes a color podría facilitar la identificación de las correspondencias tanto en el cálculo del flujo óptico como en el del mapa de disparidad. La adaptación de los métodos a versiones multicanal (RGB o HSV) sin duda mejorarían las estimaciones.

Aumentar el rendimiento de los métodos. Los métodos descritos en esta tesis se basan en técnicas de optimización que emplean esquemas numéricos iterativos básicos. El rendimiento de los métodos ha sido un objetivo dejado en un segundo plano. La optimización de los métodos se podría realizar a varios niveles:

- *Software*: Un primer paso consistiría en la mejora del código desarrollado. Se ha utilizado de forma casi generalizada el método de Gauss–Seidel para la resolución iterativa de los sistemas de ecuaciones. Otros métodos como SOR o multigrid mejorarían sustancialmente el rendimiento global de los métodos, como se demuestra en las pruebas realizas en [Bruhn06a]. La paralelización del método a través de hilos también podría ser una alternativa.

- *Hardware*: La derivación a la GPU de la resolución de los sistemas de ecuaciones de los modelos de energía, sin duda, mejorarían el rendimiento global del método. Incluso sería posible calcular las estimaciones del flujo en tiempo real. El gran problema se centra en la paralelización de los esquemas numéricos iterativos que tienen fuerte dependencia de datos respecto a los píxeles vecinos.

Técnicas de robustificación. Las técnicas de robustificación han demostrado ser una buena solución para atenuar los efectos provocados por los *outliers* y el ruido presente en las imágenes. Nuestros métodos no incorporan ninguna técnica de este tipo. La inclusión de estas técnicas supondría la aparición de términos no lineales en el sistema de ecuaciones a resolver. Dado los excelentes resultados obtenidos en otros métodos este aumento en la complejidad de los modelos se vería recompensado por una notable mejora en la precisión de las estimaciones.

Método variacional con un regularizador guiado por el flujo. En uno de los trabajos se proponía la combinación de un método variacional y otro de graph–cuts para mejorar la estimación del mapa de disparidad. La inicialización dada por el método de graph–cuts, pese a tener precisión entera, detectaba perfectamente las discontinuidades de los objetos. Esta inicialización sería mejor aprovechada por un método variacional cuyo término de regularización fuera guiado por el flujo (*flow-driven*) ya que le permitiría suavizar con mayor precisión en los contornos. Los términos de regularización guiados por la imagen (*image-driven*), como el utilizado en nuestro método, no son los idóneos cuando nos enfrentamos a imágenes altamente texturadas.

Técnicas de detección de oclusiones. El problema de las oclusiones influye negativamente sobre la estimación del flujo óptico y estéreo ya que generan errores que se propagan a otras partes de la imagen. La inclusión en el modelo de energía de técnicas que identificaran las oclusiones evitaría los efectos que éstas producen en las soluciones.

Evaluación más exhaustiva del método variacional basado en el análisis espectral del tensor de movimiento. Este método fue probado con la secuencia de Yosemite con nubes y la de Rheinhafen. Allí se puso de manifiesto la calidad de las estimaciones que se pueden obtener con este método. Hemos dejado como trabajo futuro una evaluación más exhaustiva de los modelos presentados con las secuencias propuestas en el trabajo [Baker07].

Conclusions

The aim of this dissertation is the development of a set of variational methods to deal with the optical flow and disparity maps estimation from image sequences. The main contributions done in this thesis have focused on three aspects: (i) modifications to existing methods, (ii) combination of methods in order to get the best of them and (iii) design of new energy functionals easily adaptable and extensible.

The optical flow methods developed in this thesis have been focused on studying the influence of temporal information compared to traditional spatial variational methods. Multichannel meteorological satellite image analysis is a challenging problem. In order to estimate the cloud structure motion across the satellite image sequence, we have extended a variational motion estimation technique, proposed by [Alvarez00], to deal with multichannel satellite image sequences. The main idea have been the combination of the information of all channels in the energy model. Our method has been embedded in a multi-piramidal strategy to prevent the algorithm from converging to irrelevant local minima.

We have also proposed a novel variational method that includes a temporal smoothness term. We have overcome the drawbacks of other methods that use a coupled temporal–spatial regularizer term which is more suitable for small displacements due to the inclusion of temporal derivatives. We have introduced a new term that minimizes the difference of the optical flows and permits large displacements of the objects. The design of this new term promotes the constant velocity of particles all through the sequence, so it is well suited for sequences that are translating in a fixed direction. There will not be any improvement for non translating motions, but due to the $\Phi(.)$ function in the temporal smoothness term, there will be a small penalisation if other kind of displacements are present. All the comparisons show that the bidirectional temporal method clearly outperforms the simpler temporal method. To our knowledge, this is the first approach that considers temporal regularisations with large displacements.

The third main contribution in the optical flow problem has been the definition of a new framework that establishes a strong complementary relationship between the data and smoothness term. The design of this framework was made possible by using topics recently introduced in the optical flow field and the principal component analysis of the motion tensor. From the anisotropic energy model of Nagel-Enkelmann and the motion tensor, six spatio-temporal *prototypes* have been proposed. In order to make the energy model more robust with respect to outliers, robustification functions have been included. In the traditional variational methods, a collaboration exists between all the terms of the energy functional. It is based on the compensation of the estimations done for each term. There is no *complementarity* between the terms. In our novel energy model, the use of inverse motion tensors and their decomposition in eigenvalues and eigenvectors allows us to share information within the energy. Moreover, this model can control the diffusion process in a precise way. The smoothness can be applied only in the directions where the eigenvalues of the motion tensor are small. Another advantage of this framework is that it is easily extensible because the motion tensor may encode many assumptions keeping the same structure.

In addition, methods for estimating the disparity map have focused on the analysis of the improvement that occurs through a combination of different techniques. In the stereoscopic problem, we have proposed two methods called *Stereo Video* and kz2 & StereoFlow. The first method for the estimation of dense disparity maps from a stereoscopic video sequence establishes a unified framework to deal with optical flow estimation and stereo flow computation in continuous stereo video. We have extended some well studied stereoscopic and optical flow techniques for pair of images to a sequence of images. In order to have a consistent solution, we have introduced a temporal constraint between the stereo and the optical flows. We obtain dense and accurate solutions owing to a variational formulation. We also deal with large displacements by means of a pyramidal approach. The second method combines two different techniques on disparity maps estimation in order to obtain more accurate and reliable solutions. We have used a pixel accurate method based on graph–cuts as initialization for another method based on PDE's. The latter depends on an initial approximation which is supported by the former one. The solution we obtain is in floating-point precision and the accuracy is considerably improved. We have compared the combination of the PDE and graph–cuts with the combination of the PDE and a correlation–based method. We may conclude that the use of the kz2 at the first stage provides better results than the correlation method.

Notación

En esta sección se presenta un resumen de la notación utilizada en este documento. Se ha dividido en cinco apartados con tal de facilitar al lector la comprensión de la misma. En el primer apartado, se comentan los elementos utilizados en la definición de los modelos de energía de los distintos métodos. En los tres apartados siguientes, se expone la notación utilizada en los apartados de minimización y discretización de los funcionales de energía. Por último, se enumera los distintos símbolos utilizados como parámetros de los métodos.

- Definiciones generales:

f	Secuencia de imágenes o pares estéreo
I	Secuencia de imágenes o pares estéreo
I_1, I_2	Primera y segunda imagen perteneciente a una secuencia o un par estéreo
I_l, I_r	Imagen izquierda y derecha de un par estéreo, respectivamente
x, y	Coordenadas espaciales en una imagen
t	Coordenadas temporales en una imagen
$\mathbf{x}$	Vector de coordenadas (x, y, t)
$\tilde{\mathbf{x}}$	Vector de coordenadas (x, y)
Ω	Dominio de la imagen, $\Omega \subset R^2$ en el caso de imágenes y, $\Omega \subset R^2 \times R_0^+$ para secuencia de imágenes
$\|\Omega\|$	Tamaño del dominio de la imagen Ω
u, v	Componentes del flujo óptico/estéreo en la dirección x e y respectivamente
$\mathbf{h}(x, y, t)$	Vector de componentes del flujo óptico, $(u(x, y, t), v(x, y, t), 1)^\top$
$\mathbf{h}_t(\tilde{\mathbf{x}})$	$\mathbf{h}(x, y, t)$
$\mathbf{h}(x, y)$, $\mathbf{g}(x, y)$	Vector de componentes del flujo estéreo, $(u(x, y), v(x, y))^\top$
$D(f, \mathbf{h})$	Término de ligadura
$R(\nabla f, \nabla \mathbf{u})$	Término de regularización o suavizado
u^k, v^k	Componentes del flujo en la iteración k

- Derivadas:

f_x	Derivada en la dirección x de la característica f de la imagen. En caso $f := I$ la derivada de la imagen (niveles de grises)
f_y	Derivada en la dirección y de la característica f de la imagen. En caso $f := I$ la derivada de la imagen (niveles de grises)
f_t	Derivada en la dirección t de la característica f de la imagen. En caso $f := I$ la derivada de la imagen (niveles de grises)
∂_a	Abreviación de $\frac{\partial}{\partial_a}$
$\nabla_2 a$	Gradiente espacial de a, p.e., $(\partial_x a, \partial_y a)^\top$
$\nabla_3 a$	Gradiente espaciotemporal de a, p.e., $(\partial_x a, \partial_y a, \partial_t a)^\top$
$div(a)$	$\partial_{x1} a_1 + \partial_{x2} a_2$ en el caso espacial y $\partial_{x1} a_1 + \partial_{x2} a_2 + \partial_{x3} a_3$ en el caso espaciotemporal

- Matrices y vectores:

Id	Matriz identidad
$tr(A)$, $trace(A)$	Traza de la matriz A
A_{ij}	Elemento (i, j) de una matriz
$\mu_1, ..., \mu_n$	Autovalores de una matriz $n \times n$
$r_1, ..., r_n$	Autovectores de una matriz $n \times n$
$\lvert\mathbf{a}\rvert$, $\lVert\mathbf{a}\rVert$	Módulo del vector **a**
A	Operador lineal (matriz del sistema)
$A(\mathbf{x^h})$	Operador no lineal
$x^{\mathbf{h}}$	Vector de incógnitas
$\tilde{\mathrm{x}}^{\mathbf{h}}$	Aproximación de $\mathbf{x^h}$
$\mathbf{b^h}$	Parte derecha del sistema a resolver

- Vecindad de un píxel:

$\mathcal{N}(i, j)$, $\mathcal{N}^*(i, j)$	Vecindad local alrededor del píxel (i, j)
$\lvert\mathcal{N}(i, j)\rvert$	Número de vecinos del píxel (i, j)
$\mathcal{N}(i, j)$	Vecindad local alrededor del píxel (i, j)
$\mathcal{N}^-(i, j)$	Vecindad local alrededor del píxel (i, j) que todavía no han sido evaluados
$\mathcal{N}^+(i, j)$	Vecindad local alrededor del píxel (i, j) que ya han sido evaluados

- Funciones:

J	Tensor de movimiento: matriz 3×3
D	Tensor de difusión
$\Psi(.), \Phi(.)$	Funciones de robustificación no cuadráticas utilizadas
$\Psi_D(s^2), \Psi_S(s^2)$	en los términos de ligadura y regularización
$\Psi'_D(s^2), \Psi'_S(s^2), \Phi'(s^2)$	Derivadas de $\Psi_D(s^2)$, $\Psi_S(s^2)$ y $\Phi'(s^2)$ respecto a s^2

- Parámetros de los métodos:

α, β	pesos asociado al término de regularización espacial y/o temporal
ζ	Parámetro de regularización del operador de Nagel–Enkelmann
C, λ	Parámetros de configuración de los pesos asociado al término de regularización y peso del operador de Nagel–Enkelmann
γ_i	peso asociado a la invarianza i
σ	Desviación estándar de una gaussiana
ϵ_D, ϵ_S	Parámetro de regularización utilizado en las funciones de robustificación de los términos de ligadura y suavizado
β	Parámetro de relajación utilizado en el método SOR
η, s	Factor de escalado en el enfoque multipiramidal. Valor dentro del intervalo (0,1)
τ, dt	Tamaño de paso utilizado en el método de descenso del gradiente
T	Número de iteraciones a realizar por el algoritmo
n	Número de escalas utilizadas en la estrategia multipiramidal

Índice de figuras

Índice de cuadros

Bibliografía

[Adelson85] E. Adelson, J. Bergen. Spatiotemporal Energy Models for the Perception of Motion. *Journal of the Optical Society of America A*, vol. 2, num. 2, pags. 284–299, 1985.

[Alvarez99] L. Alvarez, J. Esclarín, M. Lefébure, J. Sánchez. A PDE model for computing the optical flow. *XVI Congreso de Ecuaciones Diferenciales y Aplicaciones*, pags. 1349–1356, 1999.

[Alvarez00] L. Alvarez, J. Weickert, J. Sánchez. Reliable Estimation of Dense Optical Flow Fields with Large Displacements. *International Journal of Computer Vision*, vol. 39, num. 1, pags. 41–56, 2000.

[Alvarez02a] L. Alvarez, R. Deriche, T. Papadopoulo, J. Sánchez. Symmetrical Dense Optical Flow Estimation with Occlusions Detection. *European Conference on Computer Vision (ECCV)*, volumen 1, pags. 721–735, 2002.

[Alvarez02b] L. Alvarez, R. Deriche, J. Sánchez, J. Weickert. Dense Disparity Map Estimation Respecting Image Derivatives: A PDE and Scale-Space based Approach. *Journal of Visual Communication and Image Representation*, vol. 13, pags. 3–21, 2002.

[Alvarez07a] L. Alvarez, R. Deriche, T. Papadopoulo, J. Sánchez. Symmetrical Dense Optical Flow Estimation with Oclussions Detection. *International Journal of Computer Vision*, vol. 75, num. 3, pags. 371–386, December 2007.

[Alvarez07b] L. Alvarez, C. C. no, M. García, K. Krissian, L. Mazorra, A. Salgado, J. Sánchez. Symmetric Optical Flow. *EUROCAST 2007. Lecture Notes on Computer Science*, vol. 4739, , 2007.

[Alvarez08] L. Alvarez, C. C. no, M. García, K. Krissian, L. Mazorra, A. Salgado, J. Sánchez. Multi-Channel Satellite Image Analysis Using a Variational Approach. *Pure and Applied Geophysics*, vol. , , 2008.

[Amiaz06] T. Amiaz, N. Kiryati. Piecewise-Smooth Dense Optical Flow via Level Sets. *International Journal of Computer Vision*, vol. 68, num. 2, pags. 111–124, 2006.

[Amiaz07] T. Amiaz, E. Lubetzky, N. Kiryati. Coarse to over–fine optical flow estimation. *Pattern Recognition*, vol. 40, num. 9, pags. 2496–2503, 2007.

[Anandan89] P. Anandan. A Computational Framework and an Algorithm for the Measurement of Visual Motion. *International Journal of Computer Vision*, vol. 2, pags. 283–310, 1989.

[Ari07] R. Ari, N. Sochen. Variational stereo vision with sharp discontinuities and occlusion handling. *International Conference on Computer Vision*, pags. 1–7. IEEE Computer Society Press, 2007.

[Aubert99] G. Aubert, D. Deriche, P. Kornprobst. Computing Optical Flow via Variational Techniques. *SIAM Journal on Applied Mathematics*, vol. 60, num. 1, pags. 152–182, 1999.

[Ayache87] N. Ayache, F. Lustman. Fast and reliable passive trinocular stereovision. *In Proceedings of the 1st International Conference on Computer Vision*, vol. , pags. 422–427, June 1987.

[Baker07] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, R. Szeliski. A Database and Evaluation Methodology for Optical Flow. *International Conference on Computer Vision (ICCV 2007)*, 2007.

[Barron94] J. Barron, D. Fleet, S. Beauchemin. Performance of Optical Flow Techniques. *International Journal of Computer Vision*, vol. 12, num. 1, pags. 43–77, 1994.

[Battiti91] R. Battiti, E. Amaldi, C.Koch. Computing optical flow across multiple scales: an adaptive coarse-to-fine strategy. *International Journal of Computer Vision*, vol. 6, num. 2, pags. 133–145, 1991.

[Beauchemin95] S. Beauchemin, J. Barron. The Computation of Optical Flow. *ACM Computing Surveys*, vol. 27, num. 3, pags. 433–467, 1995.

[Bertero88] M. Bertero, T. Poggio, V. Torre. Ill-posed problems in early vision. volumen 76, pags. 869–890, 1988.

[Black91] M. Black, P. Anandan. Robust Dynamic Motion Estimation Over Time. pags. 292–302, June 1991.

[Black92] M. Black, P. Anandan. Robust Incremental Optical Flow. pags. 296–302, 1992.

[Black96a] M. Black, A. Rangarajan. On the Unification of Line Processes, Outlier Rejection and Robust Statistics with Applications in Early Vision. *International Journal of Computer Vision*, vol. 19, pags. 75–104, 1996.

[Black96b] M. Black, P. Anandan. The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Fields. *Computer Vision and Image Understanding*, vol. 63, num. 1, pags. 75–104, January 1996.

[Bobick99] A. F. Bobick, S. S. Intille. Large occlusion stereo. *International Journal of Computer Vision*, vol. 33, num. 3, pags. 181–200, 1999.

[Bornemann96] F. Bornemann, P. Deuflhard. The cascadic multigrid method for elliptic problems. *Numerische Mathematik*, vol. 75, pags. 135–152, 1996.

[Boykov04] Y. Boykov, V. Kolmogorov. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. , , 2004.

[Brad02] R. Brad, I. Letia. Cloud motion detection from infrared satellite images. *Second International Conference on Image and Graphics, SPIE*, volumen 4875, pags. 408–412, 2002.

[Brint90] A. Brint, M. Brady. Stereo Matching of Curves. *Image and Vision Computing*, vol. 8, num. 1, pags. 50–56, 1990.

[Brown03] M. Brown, D. Burschka, G. Hager. Advances in Computational Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, num. 3, pags. 993–1008, 2003.

[Brox02] T. Brox, J. Weickert. Nonlinear matrix diffusion for optic flow estimation. volumen 2449, pags. 446–453, 2002.

[Brox06] T. Brox, A. Bruhn, J. Weickert. Variational Motion Segmentation with Level Sets. volumen 3951, pags. 471–483, 2006.

[Bruhn05a] A. Bruhn, J. Weickert. Towards ultimate motion estimation: Combining highest accuracy with real-time performance. pags. 749–755, 2005.

[Bruhn05b] A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger, C. Schnörr. Variational optical flow computation in real–time. *IEEE Transactions on Image Processing*, vol. 14, num. 5, pags. 608–615, 2005.

[Bruhn05c] A. Bruhn, J. Weickert, C. Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, vol. 61, num. 3, pags. 211–231, 2005.

[Bruhn06a] A. Bruhn. *Variational Optic Flow Computation: Accurate Modelling and Efficient Numerics*. Tesis Doctoral, Department of Mathematics and Computer Science, Saarland University, Germany, 2006.

[Bruhn06b] A. Bruhn, J. Weickert, T. Kohlberger, C. Schnörr. A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. *International Journal of Computer Vision*, vol. 70, num. 3, pags. 257–277, 2006.

[Buxton84] B. Buxton, H. Buxton. Computation of optical flow from the motion of edge features in image sequences. *Image and Vision Computing*, vol. 2, pags. 59–75, 1984.

[Cachier00] P. Cachier, D. Rey. Symmetrization of the non-rigid registration problem using inversion-invariant energies: Application to multiple sclerosis. *LNCS (MICCAI 2000)*, volumen 1935, pags. 472–481, 2000.

[Campani90] M. Campani, A. Verri. Computing optical flow from an overconstrained system of linear algebraic equations. *3rd International Conference on Computer Vision*, pags. 22–26, 1990.

[Castro87] E. D. Castro, C. Morandi. Registration of translated and rotated images using finite fourier transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, pags. 700–703, 1987.

[Chan99] T. Chan, P. Mulet. On the convergence of the lagged diffusivity fixed point method in total variation image restoration. *SIAM Journal on Numerical Analysis*, vol. 36, (2), pags. 354–367, 1999.

[Christensen01] G. Christensen, H. Johnson. Consistent image registration. *IEEE Transactions on Medical Imaging*, vol. 20, num. 7, pags. 568–582, 2001.

[Cohen93] I. Cohen. Nonlinear Variational Method for Optical Flow Computation. 1993.

[Corpetti02] T. Corpetti, E. Memin, P. Perez. Dense Estimation of Fluid Flows. *IEEE Transactions on Pattern Analysis and Machine Inteligence*, vol. 24, num. 3, pags. 365–380, March 2002.

[Côté95] S. Côté, A. Tatnall. A neural network-based method for tracking features from satellite sensor images. *International Journal Remote Sensing*, vol. 16, num. 16, pags. 3695–3701, 1995.

[Deriche95] R. Deriche, P. Kornprobst, G. Aubert. Optical-Flow Estimation while Preserving Its Discontinuities: A Variational Approach. *Asian Conference on Computer Vision*, volumen 2, pags. 290–295, 1995.

[Devernay94] F. Devernay, O. Faugeras. Computing differential properties of 3-D shapes from stereoscopic images without 3-D models. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, vol. , pags. 208–213, June 21–23 1994.

[Elad98] M. Elad, A. Feuer. Recursive optical flow estimation – adaptive filtering approach. *Journal of Visual Communication and Image Representation*, vol. 9, num. 2, pags. 119–138, 1998.

[Enkelmann88] W. Enkelmann. Investigation of Multigrid Algorithms for the Estimation of Optical Flow Fields in Image Sequences. *Comput. Vis. Graph. Image Process*, vol. 43, pags. 150–177, 1988.

[Evans06] A. Evans. Cloud Motion Analysis Using Multichannel Correlation-Relaxation Labeling. *IEEE Geoscience and Remote Sensing Letters*, vol. 3, num. 3, pags. 392–396, July 2006.

[Farnebäck01] G. Farnebäck. Very high accuracy velocity estimation using orientation tensors, parametric motion, and simultaneous segmentation of the motion field. *IEEE Computer Society Press*, vol. 1, pags. 171–177, 2001.

[Faugeras93a] O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, 1993.

[Faugeras93b] O. Faugeras, B. Hotz, H. Mathieu, T. Viéville, Z. Zhang, P. Fua, E. Théron, L. Moll, G. Berry, J. Vuillemin, P. Bertin, C. Proy. Real time correlation based stereo: algorithm implementations and applications. *International Journal of Computer Vision*, vol. , , 1993.

[Faugeras01] O. Faugeras, Q. Luong, T. Papadopoulo. *The Geometry of Multiple Images*. MIT Press, 2001.

[Fleet90] D. Fleet, A. Jepson. Computation of Component Image Velocity from Local Phase Information. *International Journal of Computer Vision*, vol. 5, num. 1, pags. 77–104, 1990.

[Fleet92] D. Fleet. *Measurement of Image Velocity*. Engineering and Computer Science. Kluwer Academic Publishers, 101 Philip Drive, Assinippi Park, Norwell, Massachusetts 02061 USA, 1992.

[Fleet93] D. Fleet, A. Jepson. Stability of phase information. *IEEE Trans. Patt. Anal. Mach. Intel.*, vol. 15, pags. 1253–1268, 1993.

[Franklin06] J. Franklin. Tropical Cyclone Report: Hurricane Vince. Draft, National Hurricane Center, Feb. 2006.

[Froehlinghaus96] T. Froehlinghaus, J. Buhmann. Regularizing phase based stereo. *International Conference on Pattern Recognition*, volumen I, pags. 451–455, 1996.

[Fua93] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. *Machine Vision and Applications*, vol. 6, num. 1, pags. 35–49, 1993. Available as INRIA research report 1369.

[Fucik73] S. Fucik, A. Kratochvil, J.Ñecas. Kacanov-Galerkin method. *Commentationes Mathematicae Universitatis Carolinae*, vol. 14, (4), pags. 651–659, 1973.

[Galvin98] B. Galvin, B. McCane, K.Ñovins, D. Mason, S. Mills. Recovering Motion Fields: An Evaluation of Eight Optical Flow Algorithms. *British Machine Vision Conference*, 1998.

[Greig89] D. Greig, B. Porteous, A. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society, Series B*, vol. 51, num. 2, pags. 271–279, 1989.

[Grimson85] W. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, num. 1, pags. 17–34, 1985.

[Hampel86] R. Hampel, E. Ronchetti, P. Rousseeuw, W. Stahel. Robust Stitistics: The Approach Based on Influence Functions. *MIT Press*, vol. , , 1986.

[Hartley03] R. Hartley, A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, 2003.

[Hasler90] A. Hasler. Stereoscopic Measurements. P. Rao, S. Holms, R. Anderson, J. Winston, P. Lehr, editors, *Weather Satellites:Systems, Data and Environmental Applications*, Boston, MA, 1990. Ameri. Meteor. Soc.

[Haussecker01] H. W. Haussecker, D. J. Fleet. Computing Optical Flow with Physical Models of Brightness Variations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, num. 6, , June 2001.

[Héas06] P. Héas, E. Mémin, N. Papadakis, A. Szantai. Layered estimation of atmospheric mesoscale dynamics from satellite imagery. *IEEE Transactions on Geosciences and Remote Sensing*, vol. , , 2006. (under revision).

[Heeger88] D. Heeger. Optical Flow Using Spatio-Temporal Filters. *International Journal of Computer Vision*, vol. 1, num. 4, pags. 279–302, 1988.

[Heitz93] F. Heitz, P. Bouthemy. Multimodal estimation of discontinuous optical flow using Markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, num. 12, pags. 1217–1232, 1993.

[Hermosillo02] G. Hermosillo, C. Chefd'Hotel, O. Faugeras. Variational Methods for Multimodal Image Matching. *International Journal of Computer Vision*, vol. 50, num. 3, pags. 329–343, 2002.

[Hildreth84] E. Hildreth. The computation of the velocity field. *Royal Society London*, volumen B 221, pags. 189–220, 1984.

[Horn81] B. Horn, B. Schunck. Determining Optical Flow. *MIT Artificial Intelligence Laboratory*, vol. 17, pags. 185–203, 1981.

[Huber81] P. Huber. Robust Stitistics. *Wiley*, vol. , , 1981.

[Inoue87] T. Inoue. A cloud type classification with NOAA-7 split-window measurements. *J. Geophys. Res.*, vol. 92, num. D4, pags. 3991–4000, 1987.

[Ishikawa98] H. Ishikawa, D. Geiger. Occlusions, discontinuities, and epipolar lines in stereo. *5th European Conference on Computer Vision*, pags. 232–248, 1998.

[Jenkin94] M. Jenkin, A. Jepson. Recovering local surface structure through local phase difference methods. *CVGIP: Image Understanding*, vol. 59, pags. 72–93, 1994.

[Kacur68] J. Kacur, J.Ñecas, J. Polák, J. Soucek. Convergence of a method for solving the magnetostatic field in nonlinear media. *Aplikace Matematiky*, vol. 13, pags. 456–465, 1968.

[Kalivas91] D. Kalivas, A. Sawchuk. A region matching motion estimation algorithm. *CVGIP*, vol. 54, num. 2, pags. 275–288, 1991.

[Kambhamettu95] C. Kambhamettu, K. Palaiappan, A. Hasler. Coupled, Multi-Resolution Stereo and Motion Analysis. *Proc. IEEE Int'l Symp. Computer Vision*, pags. 43–48, November 1995.

[Kanade94] T. Kanade, M. Okutomi. A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, num. 9, pags. 920–932, 1994.

[Kim03] H. Kim, K. Sohn. Hierarchical Disparity estimation with Energy Based Regularization. *IEEE International Conference on Image Processing*, volumen 1, pags. 373–376, 2003.

[Klaus06] A. Klaus, M. Sormann, K. Karner. Segment-Based Stereo Matching Using Belief Propogation and Self-Adapting Dissimilarity Measure. *ICPR*, volumen 3, pags. 20–24, 2006.

[Kolmogorov01] V. Kolmogorov, R. Zabih. Computing Visual Correspondence with Occlusions using Graph Cuts. *In International Conference on Computer Vision (ICCV)*, volumen 2, pags. 508–515, 2001.

[Kolmogorov02] V. Kolmogorov, R. Zabih. Multi–camera scene reconstruction via graph cuts. *7th European Conference on Computer Vision, LNCS 2352(3)*, pags. 82–96, 2002.

[Kolmogorov04] V. Kolmogorov, R. Zabih. What energy functions can be minimized via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 26, num. 2, pags. 147–159, 2004.

[Kories86] R. Kories, G. Zimmerman. A versatile method for the estimation of displacement vector fields from image sequences. *In IEEE Proc. of Workshop on Motion: Representation and Analysis*, vol. , pags. 101–106, 1986.

[Kuglin75] C. Kuglin, D. Hines. The phase correlation image alignment method. *In Conference on Cybernetics and Society*, vol. , pags. 163–165, Septiembre 1975.

[Kumar97] A. Kumar, S. Haker, C. Vogel, S. Zucker, A. Tannenbaum. Stereo Disparity and L1 Minimization. *IEEE Conference on Decision and Control*, volumen 2, pags. 1125–1129, 1997.

[Leese71] J. Leese, C.Ñovak, B. Clark. An automated technique for obtaining cloud motion from geosynchronous satellite data using cross correlation. *Journal of Applied Metheorology*, vol. 10, pags. 118–132, 1971.

[Lei06] C. Lei, J. Selzer, Y. Yang. Region Tree Based Stereo Using Dynamic Programming Optimization. *Computer Vision and Pattern Recognition*, volumen 2, pags. 2378–2385, 2006.

[LH80] H. Longuet-Higgins, K. Prazdny. The interpretation of a moving retinal image. *Royal Society London*, volumen B 208, pags. 385–397, 1980.

[Li06] G. Li, S. Zucker. Differential geometric consistency extends stereo to curved surfaces. *European Conference on Computer Vision (ECCV)*, volumen 3953, pags. 44–57. Lecture Notes in Computer Science, 2006.

[Li08] Y. Li, D. Huttenlocher. Learning for Optical Flow Using Stochastic Optimization. *European Conference on Computer Vision (ECCV) '2008*, 2008.

[Little88] J. Little, H. Butlhoff, T. Poggio. Parallel optical flow using local voting. *International Conference on Computer Vision*, pags. 454–459, 1988.

[Lucas81] B. Lucas, T. Kanade. An iterative image-registration technique with an application to stereo vision. *Image Understanding Workshop*, pags. 121–130, 1981.

[Luettgen94] M. Luettgen, W. Karl, A. Willsky. Efficient multiscale regularization with applications to the computation of optical flow. *IEEE Transactions on Image Processing*, vol. 3, num. 1, pags. 41–64, 1994.

[Luong96] Q.-T. Luong, O. D. Faugeras. Fundamental matrix: Theory, algorithms, and stability analysis. *International Journal of Computer Vision*, vol. 17, num. 1, pags. 43–75, 1996.

[LZ01] G. D. L. Zhou, C. Kambhamettu, K. Palaniappan, A. Hasler. Tracking Nonrigid Motion and Structure from 2D Satellite Cloud Images without Correspondences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, num. 11, pags. 1330–1336, November 2001.

[Mansouri98] A. Mansouri, A. Mitiche, J. Konrad. Selective Image Diffusion: Application to Disparity Estimation. *Proc. Int'l Conf. Image Processing*, volumen 3, pags. 284–288. IEEE Computer Society Press, 1998.

[McIntosh88] J. McIntosh, K. Mutch. Matching Straight Lines. *Computer Vision, Graphics and Image Processing*, vol. 43, num. 3, pags. 386–408, sep 1988.

[Medioni85] G. Medioni, R.Ñevatia. Segment Based Stereo Matching. *Computer Vision, Graphics and Image Processing*, vol. 31, num. 1, pags. 2–18, 1985.

[Mémin98a] E. Mémin, P. Pérez. Dense Estimation and Object-Based Segmentation of the Optical Flow with Robust Techniques. *IEEE Transactions on Image Processing*, vol. 7, num. 5, pags. 703–719, May 1998.

[Mémin98b] E. Mémin, P. Pérez. A multigrid approach for hierarchical motion estimation. *6th International Conference on Computer Vision*, pags. 933–938, 1998.

[Mémin02] E. Mémin, P. Pérez. Hierarchical Estimation and Segmentation of Dense Motion Fields. *International Journal of Computer Vision*, vol. 46, num. 2, pags. 129–155, 2002.

[Mileva07] Y. Mileva, A. Bruhn, J.Weickert. Illumination-Invariant Variational Optical Flow with Photometric Invariants. *29th DAGM Symposium DAGM 2007, Lecture Notes in Computer Science*, volumen 4713, pags. 152–162, 2007.

[Mitiche87] A. Mitiche, Y. F. Wang, J. K. Aggarwal. Experiments in computing optical flow with the gradient-based, multiconstraint method. *Pattern Recognition*, vol. 20, num. 2, pags. 173–179, 1987.

[Mitiche96] A. Mitiche, P. Bouthemy. Computation and analysis of image motion: a synopsis of current problems and methods. *International Journal of Computer Vision*, vol. 19, num. 1, pags. 29–55, 1996.

[Mumford89] D. Mumford, J. Shah. Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems. *Communications on Pure and Applied Mathematics*, vol. 42, pags. 577–684, 1989.

[Nagel83] H.-H. Nagel. Constraints for the estimation of displacement vector fields from image sequences. *Eighth International Joint Conference on Artificial Intelligence*, volumen 2, pags. 945–951, 1983.

[Nagel86] H.-H. Nagel, W. Enkelmann. An Investigation of Smoothness Constraints for the Estimation of Displacements Vector Fields from Image Sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pags. 565–593, 1986.

[Nagel90] H. H. Nagel. Extending the 'oriented smoothness constraint' into the temporal domain and the estimation of derivatives of optical flow. *European Conference on Computer Vision (ECCV) '90, Lecture Notes in Computer Science*, volumen 427, pags. 139–148, 1990.

[Nasrabadi92] N.Ñasrabadi. A Stereo Vision Technique Using Curve-Segments and Relaxation Matching. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, num. 5, pags. 566–572, May 1992.

[Nesi93] P.Ñesi. Variational approach to optical flow estimation managing discontinuities. *Image and Vision Computing*, vol. 11, num. 7, pags. 419–439, 1993.

[Nishihara84] K.Ñishihara. Practical real-time imaging stereo matcher. *Optical Engineering*, vol. 23, num. 5, pags. 536–545, Agosto 1984.

[Ohta85] Y. Ohta, T. Kanade. Stereo by intra- and inter-scanline search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7, num. 2, pags. 139–154, 1985.

[Ottenbacher97] A. Ottenbacher, M. Tomassini, K. Holmlund, J. Schmetz. Low-Level Cloud Motion Winds from Meteosat High-Resolution Visible Imagery. *Weather and Forecasting*, vol. 12, num. 1, pags. 175–184, March 1997.

[Papenberg06] N. Papenberg, A. Bruhn, T. Brox, S. Didas, J. Weickert. Highly Accurate Optic Flow Computation with Theoretically Justified Warping. *International Journal of Computer Vision*, vol. 67, num. 2, pags. 141–158, 2006.

[Perona90] P. Perona, J. Malik. Scale-space and Edge Detection Using Anisotropic Diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, num. 7, pags. 629–639, 1990.

[Phillips72] D. Phillips, E. Smith, V. Suomi. Comment on 'An automatied Technique for Obtaining Cloud Motion from Geosynchronous Satellite Data Using Cross-Correlation'. *Journal of Applied Meteorology*, vol. 11, pags. 752–754, 1972.

[Pollard85] S. Pollard, J. Mayhew, J. Frisby. PMF: A stereo correspondence algorithm using a disparity gradient constraint. *Perception*, vol. 14, pags. 449–470, 1985.

[Proesmans94] M. Proesmans, L. J. Van Gool, E. J. Pauwels, A. Oosterlinck. Determination of Optical Flow and its Discontinuities using Non-Linear Diffusion. *ECCV '94: Proceedings of the Third European Conference-Volume II on Computer Vision*, pags. 295–304, London, UK, 1994. Springer-Verlag.

[Robert91] L. Robert, O. Faugeras. *Curve-based stereo: Figural continuity and curvature.* In *Proceedings of the International Conference on Computer Vision and Pattern Recognition.* IEEE Computer Society Press, Lahaina, Hawai, June 3–6 1991.

[Robert96] L. Robert, R. Deriche. Dense Depth Map Reconstruction: A Minimization and Regularization Approach which Preserves Discontinuities. *Lecture Notes in Computer Science*, vol. 1064, pags. 439–451, 1996.

[Roth07] S. Roth, M. Black. On the Spatial Statistics of Optical Flow. *International Journal of Computer Vision*, vol. 74, (1), pags. 33–50, Agosto 2007.

[Roy98] S. Roy, I. Cox. A maximum–flow formulation of the n–camera stereo correspondence problem. *In International Conference on Computer Vision (ICCV)*, pags. 492–499, 1998.

[Roy99] S. Roy. Stereo without epipolar lines: A maximum–flow formulation. *International Journal of Computer Vision*, vol. 34, num. 2/3, pags. 147–162, 1999.

[Salgado05a] A. Salgado, J. Sánchez. 3D Geometry Reconstruction from a Stereoscopic Video Sequence. *ICIAR 2005, Lecture Notes on Computer Science*, vol. 3656, pags. 609–616, 2005.

[Salgado05b] A. Salgado, J. Sánchez. Accurate Estimation of Dense Disparity Maps Through a PDE. *Workshop on PDE Methods in Computer Graphics*, 2005.

[Salgado05c] A. Salgado, J. Sánchez. Combining two methods to accurately estimate dense disparity maps. *ICINCO'05: International Conference on Informatics in Control, Automation and Robotics*, pags. 210–216, 2005.

[Salgado05d] A. Salgado, J. Sánchez. A Unified Framework for Optical Flow and Disparity Map Estimation from a Sequence of Stereoscopic Images. *CEDI, I Congreso Español de Informática*, pags. 145–152, 2005.

[Salgado06a] A. Salgado, J. Sánchez. Optical Flow Estimation with Large Displacements: A Temporal Regularizer. Paper 33, Cuadernos del Instituto Universitario de Ciencias y Tecnologías Cibernéticas, Abril 2006.

[Salgado06b] A. Salgado, J. Sánchez. A Temporal Regularizer for Large Optical Flow Estimation. *ICIP 2006: IEEE International Conference on Image Processing*, pags. 1233–1236, 2006.

[Salgado07a] A. Salgado, J. Sánchez. Combining two methods to accurately estimate dense disparity maps. *Informatics in Control, Automation and Robotics II*, vol. , pags. 137–144, 2007.

[Salgado07b] A. Salgado, J. Sánchez. Temporal Constraints in Large Optical Flow Estimation. *EUROCAST 2007. Lecture Notes on Computer Science*, vol. 4739, pags. 709–716, 2007.

[Scharstein02] D. Scharstein, R. Szeliski. A Taxonomy and Evaluation of Dense two Frame Stereo Correspondence Algorithms. *International Journal on Computer Vision*, vol. 47, num. 1, pags. 7–42, 2002.

[Schmetz93] J. Schmetz, K. Holmlund, J. Joffman, B. Strauss, B. Mason, V. Gaertner, A. Koch, L. van de Berg. Operational cloud motion winds from Meteosat infrared images. *Journal of Applied Meteorology*, vol. 32, pags. 1206–1225, 1993.

[Schmetz02] J. Schmetz, P. Pili, S. Tjemkes, D. Just, J. Kerkmann, S. Rota, A. Ratier. An Introduction to Meteosat Second Generation (MSG). *American Meteorological Society*, vol. , pags. 977–992, July 2002.

[Schnörr94a] C. Schnörr. Bewegungssegmentation von Bildfolgen durch die Minimierung konvexer nicht-quadratischer Funktionale. *Mustererkennung*, vol. , pags. 178–185, 1994.

[Schnörr94b] C. Schnörr. Segmentation of visual motion by minimising convex non-quadratic functionals. *IEEE Computer Society Press*, vol. A, pags. 661–663, 1994.

[Shah91] J. Shah. Segmentation by Nonliniar Diffusion. *Computer Vision and Pattern Recognition (CVPR)*, pags. 202–207, 1991.

[Shah93] J. Shah. A Non-linear Diffusion Model for Discontinious Disparity and Half Occlusions in Stereo. *Computer Vision and Pattern Recognition (CVPR)*, pags. 34–40, 1993.

[Shulman89] D. Shulman, J. Hervé. Regularization of discontinuous flow fields. *Workshop on Visual Motion, IEEE Computer Society Press*, volumen A, pags. 81–90, 1989.

[Slesareva05] N. Slesareva, A. Bruhn, J. Weickert. Optical Flow Goes Stereo: A Variational Method for Estimating Discontinuity–Preserving Dense Disparity Maps. *27th DAGM Symposium, Lecture Notes in Computer Science*, volumen 3663, pags. 33–40, 2005.

[Sun08] D. Sun, S. Roth, J. Lewis, M. Black. Learning optical flow (SRF-LFC). *European Conference on Computer Vision (ECCV) '2008*, volumen 5304, pags. 83–97, 2008.

[Sutton83] M. Sutton, W. Walters, W. Peters, W. Ranson, S. McNeil. Determination of displacement using an improved digital correlation method. *Image and Vision Computing*, vol. 1, num. 3, pags. 133–139, 1983.

[Tretiak84] O. Tretiak, L. Pastor. Velocity estimation from image sequences with second order differential operators. *Proc. 7th International Conference on Pattern Recognition*, vol. , pags. 20–22, 1984.

[Ullman79] S. Ullman. The interpretation of visual motion. Cambridge, London, 1979.

[Uras88] S. Uras, F. Girosi, A. Verri, V. Torre. A computational approach to motion perception. *Biological Cybernetics*, vol. 60, pags. 79–97, 1988.

[Viola97] P. Viola, W. M. Wells. Alignment by Maximization of Mutual Information. *International Journal of Computer Vision*, vol. 24, pags. 137–154, 1997.

[Watson85] A. Watson, A. A. Jr. Model of human visual motion sensing. *J. Opt. Soc. Am.*, vol. 2, num. 2, pags. 322–341, 1985.

[Weickert98] J. Weickert. On discontinuity-preserving optic flow. *Computer Vision and Mobile Robotics Workshop, CVMR '98*, pags. 115–122, 1998.

[Weickert01a] J. Weickert, C. Schnörr. A Theoretical Framework for Convex Regularizers in PDE-based Computation of Image Motion. *International Journal of Computer Vision*, vol. 45, num. 3, pags. 245–264, 2001.

[Weickert01b] J. Weickert, C. Schnörr. Variational Optic Flow Computation with a Spatio-Temporal Smoothness Constraint. *Journal of Mathematical Imaging and Vision*, vol. 14, num. 3, pags. 245–255, 2001.

[Weickert04] J. Weickert, N. Papenberg, A. Bruhn, T. Brox. High Accuracy Optical Flow Estimation Based on a Theory for Warping. volumen 4, pags. 25–36, May 2004.

[Wells96] W. M. Wells, P. Viola, H. Atsumi, S.Ñakajima, R. Kinikis. Multimodal Volume Registration by Maximization of Mutual Information. *Medical Image Analysis*, vol. 1, pags. 35–51, 1996.

[Wiklund92] J. Wiklund, C. Westelius, H. Knutsson. Hierarchical phase based disparity estimation. *Proc. Second Int. Singapore Conf. on Image Proc.*, vol. , pags. 128–131, Sept. 7–11 1992.

[Wohn83] K. Wohn, L. Davis, P. Thrift. Motion estimation based on multiple local constraints and nonlinear smoothing. *Pattern Recognition*, vol. 16, num. 6, pags. 563–570, 1983.

[Wu93] Z. Wu, R. Leahy. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 15, num. 11, pags. 1101–1113, 1993.

[Wu95] Q. Wu. A correlation-relaxation-labeling framework for computing optical flow: Template matching from a new perspective. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, num. 9, pags. 843–853, September 1995.

[Yoon05] K. Yoon, I. Kweon. Locally Adaptive Support-Weight Approach for Visual Correspondence Search. *Computer Vision and Pattern Recognition (CVPR)*, volumen 2, pags. 924–931, 2005.

[Young71] D. Young. Iterative Solution of Large Linear Systems. *Academic Press*, vol. , , 1971.

[Young90] G. Young, R. Chellappa. 3D Motion Estimation Using a Sequence of Noisy Stereo Images: Models, Estimation, and Uniqueness Results. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, num. 8, pags. 735–759, August 1990.

[Zimmer08] H. Zimmer, A. Bruhn, L. Valgaerts, M. Breuss, J. Weickert, B. Rosenhahn, H.-P. Seidel. PDE–Based Anisotropic Disparity-Driven Stereo Vision. *Vision, Modeling, and Visualization*, pags. 263–272. AKA Heidelberg, 2008.

[Zimmer09] H. Zimmer, A. Bruhn, J. Weickert, L. Valgaerts, A. Salgado, B. Rosenhahn, , H.-P. Seidel. Complementary Optic Flow. *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR). Lecture Notes on Computer Science*, vol. 5681, pags. 207–220, 2009.