

DEPARTAMENTO DE INFORMATICA Y SISTEMAS

TESIS DOCTORAL

**Hacia una Teoría de las Transformaciones en
Campos Receptivos y Campos de Datos.**
Implicaciones en Teoría Retinal y Proceso de Imágenes.

MEMORIA

AUTORA: **OLGA BOLIVAR TOLEDO**

DIRECTORES: **Dr. ROBERTO MORENO DIAZ**
Dr. SANTIAGO CANDELA SOLA

OCTUBRE, 1989



Universidad Politécnica de Canarias

DEPARTAMENTO DE INFORMATICA Y SISTEMAS

DON ROBERTO MORENO DIAZ, CATEDRATICO DE UNIVERSIDAD Y DON SANTIAGO CANDELA SOLA PROFESOR TITULAR DE UNIVERSIDAD, PERTENECIENTES AL AREA DE CIENCIA DE LA COMPUTACION E INTELIGENCIA ARTIFICIAL DEL DEPARTAMENTO DE INFORMATICA Y SISTEMAS DE LA UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA.

CERTIFICAN:

Que la memoria titulada HACIA UNA TEORIA DE LAS TRANSFORMACIONES EN CAMPOS RECEPTIVOS Y CAMPOS DE DATOS: IMPLICACIONES EN TEORIA RETINAL Y PROCESO DE IMAGENES", ha sido realizada por D. Olga-Bolivar Toledo bajo mi dirección en el Departamento de Informática y Sistemas de la Universidad de Las Palmas de Gran Canaria.

Esta memoria constituye la Tesis que presenta para optar al grado de Doctora en Informática.

Las Palmas , a 27 de Octubre de 1.989.

Vº Bº

EL DIRECTOR DE TESIS

EL DIRECTOR DEL DPTO DE
INFORMATICA Y SISTEMAS.

Roberto Moreno Díaz.

FDo.- Santiago Candela Solá.

A mis Padres

AGRADECIMIENTOS

Quiero expresar profunda y sinceramente mi agradecimiento al Profesor D. Roberto Moreno Díaz por haberme permitido colaborar con él durante estos años. Sus trabajos, sus ideas y sus amenas reuniones científicas han constituido tanto la base de mi formación científica como las directrices fundamentales de este trabajo. Su categoría científica, su apoyo constante y su inmensa calidad humana me han aportado el estímulo y la seguridad que he necesitado en todo momento.

Al Profesor D. Santiago Candela Solá, mi agradecimiento especial por haberme animado siempre. Gracias a sus reflexiones, consejos y dirección he podido realizar esta Tesis.

También quiero agradecer a mis compañeros del Departamento de Informática y Sistemas de la Universidad de Las Palmas la colaboración, ayuda y disponibilidad que siempre me han prestado.

Particularmente quiero agradecer a Vicente Umpiérrez su gran predisposición y esmero en todas las tareas de edición.

Por último y por ser el más entrañable, quiero agradecer a Jose Antonio Muñoz Blanco, su ayuda incondicional a lo largo de todo el trabajo. A él le debo mucho.

RESUMEN

RESUMEN

La presente memoria está estructurada en 4 capítulos. El primero de ellos constituye una exposición de conceptos básicos sobre Visión Artificial y Visión Natural, en los que se fundamenta gran parte del trabajo realizado. En la primera parte de este capítulo, presentamos los planteamientos básicos de la visión artificial. La idea central consiste en utilizar las imágenes para obtener representaciones visuales y considerar que la caracterización de formas en un espacio de representación es anterior a la interacción con una estructura de conocimiento.

Presentamos una de las mas fuertes tendencias actuales en visión artificial, cuyos puntos básicos se refieren a una doble existencia de niveles en el aspecto funcional y/o estructural y en el aspecto metodológico.

A continuación se presenta un esquema para la visión artificial, donde el objetivo central es de obtener información sobre formas a partir de las imágenes, imponiendo restricciones que vienen determinadas por la propia naturaleza del mundo exterior. Se plantea el concepto de visión multimodal, en el que el sistema realiza representaciones usando distintos modos visuales, donde se ha de llegar a un diagnóstico final a partir de la conversación cruzada entre las modalidades.

En el segundo apartado de este capítulo dentro de los métodos mas utilizados en el reconocimiento de formas, hacemos un énfasis especial en la Transformada de Fourier y en la teoría de momentos, por constituir éstos las bases funcionales de las transformaciones empleadas fundamentalmente en el trabajo realizado.

En el tercer apartado de este capítulo presentamos las bases e hipótesis fundamentales de los procesos retinales. La característica funcional mas relevante de la retina de los vertebrados, y posiblemente de la corteza, es su estructura operativa por capas. Las capas externas proporcionan a las ganglionares el mínimo conjunto de señales que les permita construir sus respuestas especializadas. Estos niveles de señales son entregados a la capa plexiforme interna. Señalamos la importancia y relevancia de la CPI como lugar definitivo donde ocurre la especialización, a través de una interacción lateral no lineal, en la que las células amacrinas son importantes.

A continuación presentamos un esquema conceptual de la operación retinal, que incluye las ideas fundamentales de los trabajos de Moreno-Díaz y colaboradores.

En el segundo capítulo titulado Transformaciones no Lineales introducimos los conceptos de Linealidad, no Linealidad y Complitud, con relación al campo receptivo y a la función. Posteriormente presentamos un marco general para las transformaciones globales, basado en la aplicación de un conjunto de reglas de distinta naturaleza a un campo de datos.

El aspecto mas interesante ha sido el incluir las convoluciones generalizadas como caso particular de una única regla aplicada a todos o parte de los "lugares de memoria" contenidos en un campo de datos. A partir de esta nueva formulación, se deduce de forma elegante la convolución convencional, en el caso de que la regla pueda expresarse de forma algebraica y lineal.

De la misma manera se incluyen en la formulación típicas transformaciones no lineales correspondientes ahora a convoluciones generalizadas de naturaleza analítica y algorítmica.

Dentro de las convoluciones generalizadas por reglas analíticas nos centramos en los sistemas no lineales polinómicos y teniendo en cuenta que dichos sistemas son equivalentes a un expansor del número de funciones de entrada seguido de un sistema lineal, aplicamos este esquema a la generación de transformaciones no lineales en proceso de imágenes.

Finalmente desarrollamos de forma práctica un teorema propuesto por J.A. Muñoz relativo a transformaciones no lineales en preproceso de imágenes y estudiamos su incidencia en teoría cortical.

El tercer capítulo de esta tesis, titulado Campos de Datos y Particiones, constituye el núcleo fundamental del trabajo realizado.

Como continuación del estudio realizado por Candela en 1987 sobre las Transformaciones de resolución progresiva, realizamos una implementación de las Transformaciones Foveales, denominadas así porque para una retina (campo de datos) proporcionan la máxima resolución en la parte central ó fovea.

Desde el punto de vista teórico, nos replantemos el concepto de complitud, de forma que al considerar un campo de datos hemos de tener en cuenta los lugares de memoria y la resolución de los mismos.

Después de un análisis de las T.R.P, proponemos un enfoque alternativo de las mismas. Para ello, en lugar de fijarnos solamente en las particiones independientes de los campos receptivos, nos centramos en el concepto de descripción completa de un campo de datos, tanto en el aspecto funcional como en el del campo receptivo.

Analizamos primero, por razones de sencillez las transformaciones homogéneas de campo receptivo, que corresponderían a las particiones foveales de una capa, y establecemos y demostramos un Teorema, relativo al conjunto no redundante de funcionales que actuando sobre las direcciones de cada partición, genera una representación completa.

A partir de este teorema, que constituye en núcleo fundamental del trabajo práctico realizado, presentamos una aplicación del mismo al estudio de la complitud en proceso de imágenes. Los funcionales elegidos en este caso fueron los descriptores de Fourier.

Posteriormente analizamos las transformaciones no homogéneas de campo receptivo, que en definitiva son aquellas en las que el número de grados de libertad de cada subpartición no es constante.

Este análisis nos lleva a formular y demostrar un nuevo teorema que facilita la elección de un conjunto mínimo de funcionales, que garantice la complitud en estos casos. Realizamos una aplicación al proceso de imágenes utilizando los núcleos de la transformada de Haar.

En el último apartado de este capítulo introducimos un planteamiento general de las Redes Neuronales, a partir del cual aclaramos la limitación y alcance de los algoritmos rápidos de computación.

Por último en el capítulo 4, presentamos la estructura computacional concreta que se ha implementado en ordenador y detallamos las pruebas experimentales llevadas a cabo en el reconocimiento y clasificación de formas.

INDICE

CAPITULO 1: ANTECEDENTES	Páginas
1.1 Planteamientos en Visión Artificial.....	1.1
1.2 Métodos en Reconocimiento de Formas.....	1.12
1.2.1 Características Locales.....	1.10
1.2.1 Características Globales.....	1.14
1.3 Proceso de Datos en Retina y Proyección al Cortex Visual	
1.3.1 Estructura por capas. Relevancia de la C.P.I.....	1.28
1.3.2 El concepto de Preproceso.....	1.32
1.3.3 Interacción lateral no lineal en la C.P.I.....	1.33
1.3.4 Esquema conceptual de la Operación Retinal.....	1.35
CAPITULO 2: TRANSFORMACIONES NO LINEALES	
2.1 Linealidad, no linealidad y Complejidad.....	2.1
2.2 Producción y aplicación de Transformaciones no lineales.....	2.4
2.2.1 Marco general para Transformaciones Globales.....	2.5
2.2.2 Convoluciones Generalizadas.....	2.6
2.2.3 Producción de Convoluciones Generalizadas por reglas analíticas.....	2.12
2.3 Sistemas no lineales polinómicos.....	2.13
2.4 Transformaciones no lineales en Preproceso de Imágenes.....	2.23
2.5 Incidencia en Teoría de Redes Neuronales.....	2.27

CAPITULO 3: CAMPOS DE DATOS Y PARTICIONES

3.1 Sobre las Transformaciones Campo Receptivo Funcional.....	3.1
3.1.1 Transformaciones de Campo receptivo variable.....	3.1
3.1.2 Transformaciones Foveales.....	3.5
3.2 Campos de Datos. Transformaciones y Representaciones Completas.	3.10
3.2.1 Una reconsideración de la capacidad de los ficheros.....	3.14
3.2.2 Los casos usuales en visión artificial.....	3.15
3.3 Expresión de las Transformaciones Completas. Campo Receptivo Funcional.....	3.17
3.3.1 Transformaciones homogéneas de Campo Receptivo.....	3.19
3.3.2 Transformaciones no homogéneas de Campo Receptivo.....	3.25
3.4 Redes Neuronales y Redes por Capas.....	3.35
3.5 Algoritmos rápidos de computación.....	3.41

CAPITULO 4: APLICACION AL PROCESO Y RECONOCIMINETO VISUAL

4.1 Un sistema práctico para Visión Artificial.....	4.1
4.2 Preproceso y Normalización.....	4.7
4.3 Generación de la Base de datos referencial y de la máquina de inferencia.....	4.16
4.4 Resultados Experimentales.	4.25

CAPITULO 5: CONCLUSIONES Y PRINCIPALES APORTACIONES...5.1

REFERENCIAS.....1-3

CAPITULO 1

ANTECEDENTES

1.1 PLANTEAMIENTOS EN VISION ARTIFICIAL

El problema de la visión artificial trata, en general, de la definición de "máquinas" posibles (existentes o no) que, partiendo de imágenes proporcionen una descripción del mundo exterior. Así, un concepto que parece ser útil tanto en los sistemas naturales como en los artificiales, consiste en considerar la visión como la "operación de un sistema que proporciona, a partir de imágenes, una representación que es útil al que ve y que no está cargada de información irrelevante" [1].

El proceso visual presenta dos características iniciales de importancia. Primero, el requisito de que la información extraída sea útil para los fines de un sistema más amplio que incluye al visual como parte, y segundo, la reducción, en general drástica, de la información contenida en las imágenes, con la esperanza de que se mantenga sólo aquella información que es relevante a los fines del sistema total.

Partiendo de que el sistema visual, artificial o natural, está integrado en un sistema total cuyos fines sirve, la investigación lógica en los mecanismos de visión debe partir de los requisitos y/o restricciones que vienen impuestas por esos fines. Ahora bien, aunque en el caso de la mayor parte de los sistemas visuales artificiales, el establecimiento de esas restricciones puede ser realizado a priori con una cierta precisión, en el caso de los sistemas naturales éstos no son conocidos, y son, por consiguiente, tema también de la investigación en visión.

La tarea de modelización de la visión biológica podría consistir en el desarrollo de una teoría efectiva de como se construye la representación visual invariante del mundo exterior, es decir, cómo se llega a la percepción visual invariante a partir de las imágenes proyectadas en las retinas. Como es bien

sabido, la percepción visual está influida por lo que se necesita reconocer. Asimismo, la solución de los problemas asociados al reconocimiento visual exige, como ha sido reiteradamente demostrado [2], el uso de la información multisensorial para la elaboración de un modelo del medio, que en definitiva es algo así como una estructura de conocimiento, de naturaleza relacional, donde las pistas iniciales proporcionadas por las modalidades sensoriales son usadas, a través de mecanismos cooperativos, para actualizar su estructura e indagar en el mundo exterior, requiriendo más información de las modalidades sensoriales.

Como consecuencia, una de las cuestiones que de forma más o menos explícita subyace en la investigación de la visión artificial actual es la de la medida en la que el proceso visual está determinado por los datos visuales o por la estructura de conocimiento. Una creencia bastante antigua en visión artificial es que los fenómenos de percepción visual son análogos a los de resolución de problemas en inteligencia artificial, y que, por consiguiente, están determinados por la estructura del conocimiento. Ello lleva a considerar que la función fundamental del proceso visual es el establecimiento, la prueba y la modificación de hipótesis acerca de lo que hay en la escena.

Esta idea sigue siendo común en la visión por computador y tiene algunas contrapartidas en la psicología de la visión. El concepto básico es invocar el conocimiento especializado que se dispone de la escena que se ve para ayudar en la segmentación de la imagen en regiones que se deben corresponder con los objetos que se esperan en la escena. Este tipo de enfoque tuvo buenos representantes en el laboratorio de Inteligencia Artificial del MIT y en el Stanford Research Institute y su expresión clara en los conceptos de "entender imágenes".

Marr [3] propuso un enfoque alternativo al anterior, en el sentido de considerar que el problema de la visión es un "proceso determinado por los

datos". Su tesis mantiene que los conocimientos adquiridos apenas desempeñan papel alguno en las primeras etapas del procesamiento visual, de manera que la visión puede ser considerada como una tarea de procesamiento de información, es decir de extracción a partir de las imágenes, de los diferentes aspectos de la escena observada que son útiles para el observador, y de una representación interna de esta información con toda su profusión de detalle, disponiéndose así de una base para decidir, en interacción con la estructura de conocimiento, sobre pensamientos y acciones.

La idea central de Marr es que las imágenes han de ser usadas para obtener representaciones visuales y que la caracterización de formas en un espacio de representación es anterior a la interacción con la estructura de conocimiento.

Los puntos básicos de las ideas de Marr se refieren a una doble existencia de niveles en el aspecto funcional y/o estructural y en el aspecto metodológico.

En el aspecto funcional/estructural, la existencia de niveles corresponde, en el caso particular de la visión, a la distribución de niveles más general que caracteriza a la Biocibernética y en el reconocimiento de que cada nivel requiere sus propias herramientas formales que impone limitaciones a lo que se puede obtener [4]. Los niveles funcionales van desde el bioquímico - o físico -, al neuronal o de componentes, al de operaciones lógicas, hasta el nivel simbólico, con el establecimiento de fines.

En el aspecto metodológico, la estructura de tres niveles de Marr se refiere al nivel computacional, al nivel de los algoritmos y al nivel de la realización con hardware. Marr y sus seguidores, argumentan que al nivel computacional, las teorías de visión natural y artificial interactúan fuertemente, de tal forma que la visión por computador puede:

- a) Proporcionar un lenguaje para expresar las teorías del proceso de datos visuales.
- b) Tener capacidad de establecer experimentos claramente, sin las confusiones que se suelen derivar de la consideración de la visión humana.
- c) Permitir que las evidencias de la percepción visual humana proporcionen pistas para la visión artificial.
- d) Establecer que las teorías a distintos niveles son apropiadas para imponer restricciones en el aspecto teórico y experimental, tanto en visión artificial como en la percepción humana.

Esquema de Visión Artificial

En el modelo computacional de visión, el objetivo central consiste en conseguir aquellos métodos computacionales efectivos para pasar de imágenes a representaciones, a través de la imposición de las restricciones suficientes. Se ilustra con la búsqueda de métodos para pasar desde imágenes a superficies y desde superficies a formas, con el problema final de la representación de formas para el reconocimiento.

Este modelo, mas que una teoría completa es una propuesta general de un procedimiento para investigar en visión artificial. El objetivo central es obtener información sobre formas a partir de las imágenes, imponiendo restricciones que vienen determinadas por la propia naturaleza del mundo exterior, y no por la estructura del conocimiento impuesta a priori.

Partiendo de las imágenes, la teoría computacional propone obtener un "boceto primario", de aquí un boceto en 2 1/2 dimensiones y finalmente, una representación de las formas en 3-D. El conocimiento final de que éste es un enfoque apropiado para analizar el proceso visual les vino a Marr y Poggio del hecho de que es posible encontrar un procedimiento computacional efectivo para interpretar estereogramas formados por puntos al azar y que por consiguiente, tal representación debe ser anterior al reconocimiento. El programa de trabajo de la teoría computacional de la visión se presenta a continuación.

Según este esquema, el marco representacional presenta al menos cuatro niveles, en el que cada representación se ha de expresar en términos de unas primitivas propias. Las primitivas correspondientes a la imagen son el valor de la intensidad en cada punto. El boceto primario tiene como finalidad explicitar información relevante acerca de la imagen bidimensional, esencialmente los cambios de intensidad, y su distribución geométrica y organización. Sus primitivas incluyen cruces por cero, terminaciones y discontinuidades, segmentos de bordes, líneas virtuales, grupos, organización curvilínea y límites. El boceto 2 1/2 D tiene por objeto explicitar la orientación y profundidad aproximada de las superficies visuales, y los contornos y discontinuidades en estas magnitudes, en un sistema de coordenadas centrado en el observador. Sus primitivas incluyen la orientación local de superficies, distancia hasta el observador, discontinuidades en profundidad y en orientación de superficies.

Finalmente, está la representación en forma de modelo 3D, cuyo objetivo es describir formas y su organización espacial referidas a un sistema de coordenadas centrado en el objeto, usando una representación modular jerarquizada que incluye descriptores volumétricos tanto como descriptores superficiales. Las primitivas son un conjunto de modelos 3-D dispuestos jerárquicamente, cada uno

de los cuales posee un propio sistema de referencias intrínseco, al que se refieren los descriptores superficiales y de volumen.

El esquema anterior se completa con un procedimiento de reconocimiento visual, comandado fundamentalmente por conocimiento visual pre-estructurado y que echa mano de bases de datos conteniendo los modelos standard. El esquema total se ilustra en la fig. 1.

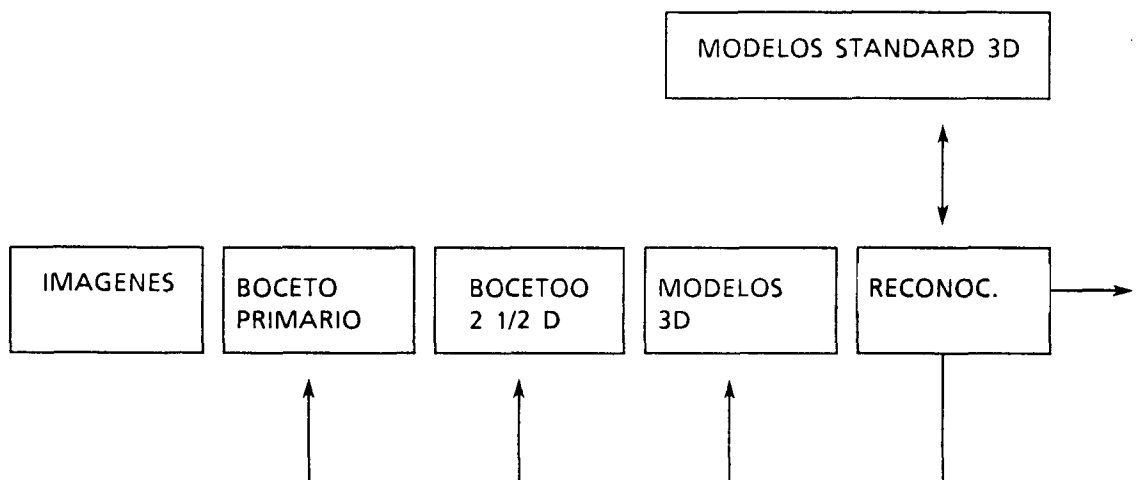


Fig. 1. Esquema de visión artificial (adaptado de Marr).

La correspondencia entre las primitivas y los descriptores y las formas que representan, es directa según lo anterior. En realidad, esto no es un requisito esencial aunque ayuda a un metaobservador a entender el proceso subyacente de una forma más próxima a la intuitiva. Las propiedades de las superficies y de las formas a las que refiere el esquema de la fig. 1 son propiedades deducibles y expresables en un lenguaje formal muy próximo al de la geometría diferencial e integral. En esencia, las restricciones impuestas son debidas a las propiedades

físicas del mundo exterior, sin llegar a implicar niveles simbólicos o semánticos hasta que se llega al reconocimiento.

La citada correspondencia no tiene en principio por que ser directa y podría llevarse a cabo en un espacio transformado, donde ciertas primitivas de interés pueden ser determinadas y manejadas con mayor flexibilidad. De hecho, esto no pasa de ser una suposición y una esperanza. En muchos casos, el trabajar con descriptores procedentes de espacios transformados en sistemas de visión artificial descansa solamente en la adecuación posterior comprobada experimentalmente, tanto en tiempo de cómputo como en aciertos. Esto ha de tomarse con cierto cuidado, porque, normalmente, los investigadores no dan cuenta de los errores, sino de los aciertos.

En todo caso, es muy difícil, hoy por hoy, proporcionar un teorema negativo acerca de la inutilidad de pasar a espacios transformados. Una transformada integral biunívoca que ha tenido gran aceptación en la representación visual ha sido la transformada de Fourier espacio-temporal. Incluso se la intentó justificar fisiológicamente por Campbell y otros [5], llegándose a afirmar que el objetivo de la retina es realizar una descomposición en frecuencias espacio-temporales, que es transmitida al cortex. Posteriormente, se ha visto que esa afirmación no tiene más fundamento que la de Hubel y Wiesel acerca de las células corticales especializadas en detectar líneas y bordes orientados, es decir, ninguno.

El que un sistema tenga una cierta respuesta espectral, no implica nunca que el que usa la información resultante esté interesado en descomposiciones espectrales.

Sin proseguir en la cuestión de que si algunos sistemas naturales de visión usan o no descriptores procedentes de transformaciones integrales, es cierto que

estos descriptores tienen propiedades que los hacen atractivos, al menos para un sistema de visión artificial. Uno de ellos es la desaparición de los escotomas que aparecen en las representaciones punto a punto. Otra es la posibilidad de lograr invarianzas en la representación de una forma rápida y elegante [6],[7].

Como consecuencia de admitir la potencial utilidad de una representación múltiple se llega a los conceptos correspondientes a la representación visual multimodal. En este caso, el sistema visualiza representaciones usando modos visuales y de la conversación cruzada entre las modalidades se ha de llegar a un diagnóstico final. Convergemos así en una propuesta de esquema de visión artificial, que se ilustra en la fig 1.2.

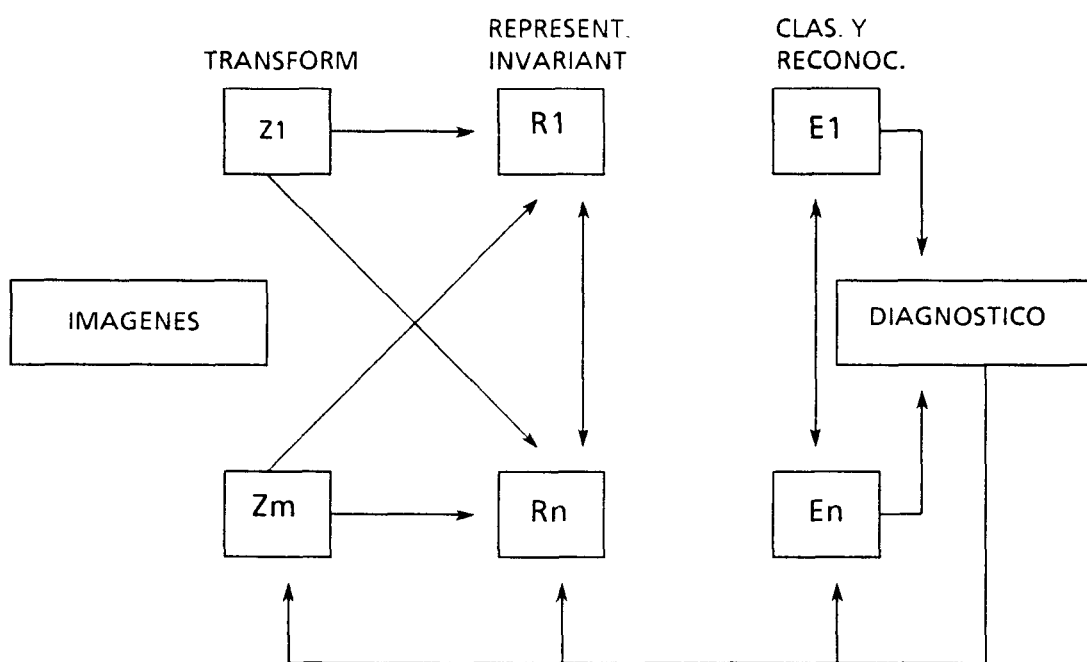


Fig. 1.2. Esquema de visión artificial

Según éste, las transformaciones básicas iniciales actúan directamente sobre las imágenes para proporcionar descriptores relevantes de los mismos, con las

correspondientes etiquetas. De acuerdo con la idea básica del reconocimiento multimodal, las transformaciones deben utilizar descriptores invariantes, si bien los parámetros para determinar las invarianzas, deben, en general, proceder de una modalidad diferente. En todo caso, ha de tenerse en cuenta que las transformaciones no corresponden a una única expresión, sino a una cascada de ellas, como se ilustró en el esquema de la fig. 1.1. El significado concreto de alguna de las transformaciones posibles será considerado en capítulos posteriores.

Por último, conviene destacar que la ausencia de una teoría completa a cerca de la percepción visual, ha dado lugar a que el área de investigación en Visión Artificial (que utilizando la definición de R.Nevatia [8] es "el área dedicada a permitir a las máquinas la percepción de su entorno visual tal y como lo hacen los humanos y los animales"), sea, como ocurre en la mayoría de las áreas de Inteligencia Artificial, una ciencia fundamentalmente experimental. Evidentemente que se hace uso en ella de principios de Óptica, Electrónica e Informática en la adquisición y procesamiento de imágenes, pero la interpretación de las imágenes para hacer un uso pragmático de las mismas, depende en gran medida de un elevado y creciente número de algoritmos y métodos, que son concebidos heurísticamente, utilizados racionalmente y verificados experimentalmente [9]. En este entorno es precisamente en el que hay que entender las aportaciones del presente trabajo.

1.2 METODOS EN RECONOCIMIENTO DE FORMAS

En este apartado presentamos una revisión de las técnicas y métodos más utilizados en el reconocimiento automático de formas. Hacemos un énfasis especial en la Transformada de Fourier y en la Teoría de Momentos ya que estas

han sido las transformaciones mas frecuentemente utilizadas en la realización de este trabajo.

El estado del arte en reconocimiento de formas ha avanzado, desde el uso de esquemas primitivos para el reconocimiento de un conjunto limitado de formas estandar, generadas por software, a la de sofisticadas aplicaciones técnicas para el reconocimiento de una gran variedad de formas reales adquiridas mediante cámaras de video.

Las técnicas de reconocimiento varían mucho dependiendo de los descriptores elegidos para caracterizar las formas, el modo en que se obtienen estos descriptores y los métodos de clasificación empleados.

Las técnicas utilizadas en reconocimiento de formas pueden agruparse en tres grandes categorías:

- a) Características Locales.
- b) Características Globales.
- c) Características Geométricas y Topológicas.

1.2.1 Características Locales

Una forma de caracterizar una imagen es obtener características de la distribución de los puntos de la misma [10-14]. Los descriptores locales elementales coinciden con los valores de los pixels de la imagen. El más empleado, principalmente por su economía y por existir comercialmente los

sensores apropiados, es el nivel de luminancia (o nivel de gris) de los puntos que forman la imagen formada en una cámara de vídeo. La señal analógica generada se digitaliza en forma matricial.

La mayoría de los objetos de la escena tendrán unas características distintas de luminosidad, y se encontrarán respecto de la fuente de luz en situaciones distintas, lo que dará lugar a cambios relativamente bruscos de nivel, que serán una de las principales pistas para la segmentación.

Por ello, el borde es un tipo de descriptor local muy interesante. Un método para calcular bordes suele consistir en calcular la convolución con algún núcleo apropiado, que de valores máximos para determinado tipo de transiciones. En general, al ser las transiciones de distinto valor y de distinta pendiente, suele ser conveniente utilizar una variedad de núcleos de distintas características.

En el caso particular de imágenes muy contrastadas, un simple análisis del histograma frecuencia/tono, nos permitirá calcular el valor de umbral adecuado que separe piezas de fondo. Existen diversos métodos: ajuste a dos gaussianas, ajuste por el método de la cuartica, suavizado progresivo hasta encontrar un solo mínimo entre dos únicos máximos, etc. Hay que tener en cuenta que, si se pretende obtener las fronteras definitivas entre regiones en un solo paso, es muy conveniente reducir el ruido que acompaña toda imagen mediante algún filtrado pasa baja, por ejemplo; e incluso, posteriormente al umbralizado, desechar regiones pequeñas o "poco conexas" (poco densas en número de puntos de pieza/número puntos fondo) etc.

Se ha de tener en cuenta que en la determinación del umbral se ha despreciado la información espacial. Esta es evidentemente importante, pudiéndose atacar la umbralización como clusterización en espacio tridimensional (las dos dimensiones

tridimensional (las dos dimensiones espaciales, más la de tono de gris). Este procedimiento es de aplicación mucho más general (suponiendo que las distintas partes de la imagen poseen tono uniformes o de variación lenta, y localmente distintos). Cualquier método de clusterización que no precise información adicional es en teoría aplicable, aunque de gran lentitud de proceso.

Calculando histogramas locales cada cierta distancia e interpolando posibles umbrales, se obtendrían umbrales dependientes de la posición que podrían servir para la clusterización tridimensional antes apuntada.

En el caso general de que las distintas partes no tengan tonos homogéneos, ni aún ligeramente variables, sino que estén formados por texturas distintas u orientadas diferentemente, los métodos vistos anteriormente pueden aplicarse aisladamente (aunque su validez es reducida) o incluso en paralelo con otros. En este tipo de imágenes se pueden extraer descriptores locales en forma de valores de transformadas aplicadas localmente o de medidas locales de textura.

Cualquier resultado de cualquier transformada bidimensional, aplicada en una zona pequeña alrededor de un punto puede considerarse como descriptor local asociado a ese punto de forma análoga al proceso de convolución, con la diferencia de que para hacer tratables los resultados, es preciso disminuir el volumen de datos resultante, tomando los más significativos. En el caso de los coeficientes de una transformada, se pueden tomar aquéllos de mayor módulo, o ajustar la superficie a una predeterminada cuyos parámetros serían los descriptores locales. Como ejemplo de estos procesos lineales, se pueden citar las conocidas transformadas de Fourier, Coseno, Seno, Hadamard, Haar y Karhunen-Loeve.

Por otro lado, las convoluciones, en principio lineales, pueden generalizarse incorporando ciertos procesos no lineales aplicados localmente, que hagan los resultados más discriminantes para el tipo de patrón en cuestión. En estos casos el descriptor sería normalmente un escalar.

Otros descriptores podrían ser los obtenidos de cualquier conjunto de parámetros de un ajuste local a un determinado tipo de función.

Por último, medidas estadísticas aplicadas localmente, como mediana, desviación, etc. y, en general, valores de momentos de distinto orden, invariantes o no.

Se pueden considerar dos tipos de descriptores semilocales, aquellos que se refieren a un área de la imagen y que ayudan a una segmentación de ésta, y aquéllos que se calculan sobre regiones de imágenes previamente segmentadas (quizá primera y provisionalmente).

Podemos entender como de los primeros, todos aquellos descriptores locales vistos anteriormente en los que, para su cálculo, haya intervenido una pequeña zona de la imagen. Cuando esta zona no se desplaza a todas las posiciones posibles, sino a incrementos de más de un pixel, podemos entenderlo como semilocal. Pueden ayudar a una segmentación inicial de baja resolución.

En imágenes complejas, donde se hayan obtenido por algún procedimiento puntos candidatos a borde, pero inconexos, se pueden extraer descriptores locales candidatos a líneas de contorno por el siguiente procedimiento: los puntos candidatos a borde se expanden, de forma que muchos de ellos (próximos) quedarán conectados. A continuación se realiza un proceso de adelgazamiento con lo que se obtendrán líneas formadas por puntos de borde

próximos. El proceso puede repetirse varias veces. Las líneas inferiores a un cierto tamaño pueden despreciarse.

Cuando ya se ha procedido a una segmentación de la imagen, se pueden extraer descriptores semilocales que reflejarán la presencia de patrones que probablemente formen parte de una entidad mayor. El caso más claro lo tenemos en imágenes binarias: en las formas podrán detectarse agujeros, esquinas, lados rectos, etc.; su posición relativa respecto del centro de la forma, su tipo y tamaño serán los valores que los definan.

A la hora de describir el contorno de una forma (para su reconocimiento), la curvatura de esta, o en concreto, patrones presentes en esta, pueden considerarse descriptores locales (más generales que esquinas, lados rectos, etc.).

1.2.2 Características Globales

Esta categoría comprende técnicas que extraen características que dependen de todos los puntos que están dentro de un rectángulo, (denominado marco) que circunscribe a la forma. Tales características no reflejan ninguna propiedad local, geométrica o topológica de la imagen.

Medidas de la Coincidencia y Correlación.-

El objetivo de esta técnica (Matching) [14-15], es medir la similitud entre una forma de entrada y las almacenadas como referencia, denominadas patrón o máscara, mediante coincidencia y correlación de puntos o grupos de puntos dentro del marco. Este proceso se ha aplicado a un conjunto de aplicaciones, como pueden ser:

a) Aplicaciones de segmentación y detección de bordes, utilizando primitivas sencillas, tales como desniveles o flancos de nivel de gris.

b) Aplicaciones de lectura automática.

c) Visión artificial para Robots.

Para realizar este proceso se han utilizado distintas funciones que definen la semejanza entre la imagen incógnita y el patrón o imagen de referencia [13]. Una medida de la diferencia entre patrón e imagen de entrada podría ser el criterio del error cuadrático medio, calculado para diferentes desplazamientos entre patrón e imagen de entrada. Si denominamos $f(x,y)$ a la imagen de entrada y $g(x,y)$ al patrón, se puede expresar la función del error cuadrático medio como:

$$D^2(u,v) = \iint [f(x,y) - g(x + u, y + v)]^2 dx dy$$

Una de las medidas de semejanza más utilizada en los procesos de medida de coincidencia es la función de correlación, que se define como:

$$C(u,v) = \iint f(x,y) g(x + u, y + v) dx dy$$

El valor máximo de la correlación no asegura la semejanza absoluta o igualdad, pero realizando una transformación en los niveles de gris de una de las imágenes puede obtenerse esa igualdad.

Transformadas Globales

Un problema que surge de tomar los puntos de la imagen como característica de la misma, como en el caso anterior, es la gran dimensión del vector resultante de características. Se han realizado intentos para reducir la dimensionalidad y extraer características invariantes a algunas transformaciones afines (traslaciones, ampliaciones y giros) mediante la utilización de transformadas y desarrollos en serie.

Consideremos la imagen digitalizada como una matriz de N^2 puntos colocados en N filas y N columnas, donde el valor de cada punto representa el brillo de cada elemento correspondiente. Las transformaciones de esta imagen conducirán también a tablas de N^2 puntos.

Una transformación bidimensional se representa por una expresión:

$$F(u, v) = \sum_x \sum_y f(x, y) a(x, y, u, v)$$

donde:

$f(x, y)$ corresponde a la imagen original;

$F(u, v)$ es la imagen transformada;

$a(x, y, u, v)$ es el núcleo de la transformación.

La transformación inversa, si existe, es:

$$f(x, y) = \sum_u \sum_v F(u, v) a(x, y, u, v)$$

El núcleo de la transformación se dice que es separable si

$$a(x, y, u, v) = a_x(x, u) a_y(y, v)$$

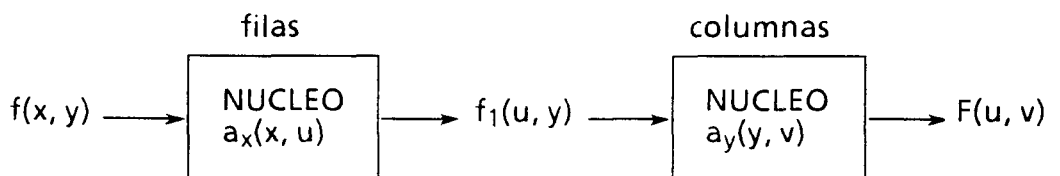
En este caso, la transformación puede efectuarse en dos tiempos, una vez sobre las líneas y otra sobre las columnas.

$$f_1(u, y) = \sum_x f(x, y) a_x(x, u)$$

y después

$$F(u, v) = \sum_y f_1(u, y) a_y(y, v)$$

Podemos esquematizarlo de la forma:



La transformación es simétrica si a_x es idéntica a a_y . Esta propiedad se cumple en general. Se pueden encontrar, sin embargo, ciertos casos en donde las transformaciones son diferentes. En el tratamiento de ciertas imágenes de radar, la transformada de "x" puede ser una transformada de Fourier y la de "y" una transformada de Fresnel. Para una transformación, supuesta separable y simétrica, se puede escribir, utilizando notación matricial:

$$[F]^T = [[f] [A]^T]^T [A]^T$$

Para la transformada directa, $[A]^T$ es la matriz traspuesta de $[A]$. La transformada inversa se efectúa con la matriz $[B]$ tal que:

$$[B] = [A]^{-1}$$

Se dice que la matriz $[A]$ es unitaria si: $[A]^{-1} = [A^*]^T$
 $[A^*]$ es la matriz conjugada de la matriz $[A]$

Para una matriz unitaria real: $[A]^{-1} = [A]^T$
 en una matriz ortogonal.

La transformación puede ser interpretada también como una descomposición espectral generalizada. En efecto, para unos valores dados u_0 y v_0 de las variables del espacio transformado, se tiene:

$$F(u_0, v_0) = \sum_x \sum_y f(x, y) a(x, y, u_0, v_0)$$

que es el producto escalar de la función a transformar $f(x, y)$ por la función $a(x, y, u_0, v_0)$ definido sobre el mismo espacio. $F(u_0, v_0)$ representa la proporción de la función $a(x, y, u_0, v_0)$ que está presente en la función $f(x, y)$.

$$\sum_i \sum_j a(x, y, u_i, v_i) a(x, y, u_j, v_j) = d_{ij} D_{ij}$$

Donde:

d_{ij} es un número y D_{ij} representa la delta de Kronecker.

En este caso, hay conservación de la energía a partir de la transformación. Existe un gran número de transformaciones que presentan propiedades de ortogonalidad y de separabilidad [16,17,18].

Transformada De Fourier

La transformada de Fourier es la más conocida de las transformaciones ortogonales, por el hecho del significado físico de sus componentes en el caso de tratamiento de señales que son función del tiempo. También pueden aplicarse a las imágenes, lo que conduce a la noción de frecuencias espaciales.

La definición de la transformada de Fourier directa e inversa viene dada por la expresión:

$$F(u, v) = \frac{1}{N} \sum_x \sum_y f(x, y) \exp \left[-\frac{2\pi i}{N} (u_x + v_y) \right]$$

$$F(x, y) = \frac{1}{N} \sum_u \sum_v F(u, v) \exp \left[\frac{2\pi i}{N} (u_x + v_y) \right]$$

La transformación es de núcleo separable y simétrico, por lo que se puede reemplazar por una secuencia de transformaciones unidimensionales, definida por la matriz de transformación.

$$[A] = \frac{1}{\sqrt{N}} \exp \left[-\frac{2\pi i u x}{N} \right]$$

Si hacemos

$$w = \exp \left[-\frac{2\pi i}{N} \right]$$

$$[A] = \frac{1}{\sqrt{N}} w^k$$

con $k = ux$, siendo $0 \leq u, x \leq N - 1$.

La siguiente figura da la expresión de esta matriz para $N = 8$.

$$\frac{1}{\sqrt{8}} \begin{bmatrix} w_0 & w_0 & w_0 & w_0 & w_0 & w_0 & w_0 & w_0 \\ w_0 & w_1 & w_2 & w_3 & w_4 & w_5 & w_6 & w_7 \\ w_0 & w_2 & w_4 & w_6 & w_8 & w_{10} & w_{12} & w_{14} \\ w_0 & w_3 & w_6 & w_9 & w_{12} & w_{15} & w_{18} & w_{21} \\ w_0 & w_4 & w_8 & w_{12} & w_{16} & w_{20} & w_{24} & w_{28} \\ w_0 & w_5 & w_{10} & w_{15} & w_{20} & w_{25} & w_{30} & w_{35} \\ w_0 & w_6 & w_{12} & w_{18} & w_{24} & w_{30} & w_{36} & w_{42} \\ w_0 & w_7 & w_{14} & w_{21} & w_{28} & w_{35} & w_{42} & w_{49} \end{bmatrix}$$

$$w = \exp \left[-\frac{2\pi i}{8} \right]$$

Propiedades de la transformación.-

De todas las propiedades de la transformada de Fourier, hay que señalar dos características que es preciso tener en cuenta en su empleo:

a) La imagen inicial comporta N^2 puntos y después de la transformación se encuentran N^2 coeficientes complejos, por lo tanto N^2 coeficientes reales. Este

efecto proviene de la hipótesis de la anulación de la parte imaginaria del brillo para cada punto de la imagen. Existe una relación entre los coeficientes de la transformada.

$$F(-u, -v) = F^*(u, v).$$

Basta con guardar N^2 coeficientes reales (correspondientes a $u > 0$, por ejemplo), puesto que los N^2 restantes se pueden deducir de estos últimos.

b) La transformada de Fourier discreta bidimensional es esencialmente una representación de Fourier de un campo bidimensional. La representación sólo es válida si el campo es periódico. La imagen original no es únicamente el conjunto de los N^2 puntos dados, sino que comporta también las imágenes trasladadas vertical y horizontalmente de estos N^2 puntos. Esta propiedad puede conducir a efectos parásitos de recubrimiento en ciertas operaciones sobre las transformadas de Fourier, con aparición de estructuras no reales.

Algoritmo de cálculo rápido.-

El desarrollo del cálculo numérico de las transformadas de Fourier, que hacen intervenir un número elevado de puntos, como es el caso de las imágenes, está ligado a la descripción por Cooley y Turkey de un algoritmo de cálculo rápido. En lugar de tener N^2 operaciones aritméticas para la transformación, el método de la transformada rápida sólo necesita $2 N \log_2 N$ operaciones, para $N = 2^n$.

La existencia de una transformada rápida de Fourier es debida a la simetría de la matriz de transformación $[A]$, que no comporta, finalmente más que N términos diferentes para transformar N puntos. Los términos son las diferentes potencias.

$$w^k = \left\{ \exp \left[-\frac{2\pi i}{N} \right] \right\}^k$$

con $k = 0, 1, \dots, N-1$.

$$\text{pues: } \exp \left[-\frac{2\pi i}{N} k \right] = \exp \left[-\frac{2\pi i}{N} (k + N) \right]$$

La siguiente figura 1.3 representa la matriz de transformación reducida para $N = 8$. Esta matriz puede ser factorizada en un producto de matrices más simples, teniendo numerosos coeficientes nulos.

$$\frac{1}{\sqrt{8}} \begin{bmatrix} w^0 & w^0 & w^0 & w^0 & w^0 & w^0 & w^0 & w^0 \\ w^0 & w^1 & w^2 & w^3 & w^4 & w^5 & w^6 & w^7 \\ w^0 & w^2 & w^4 & w^6 & w^0 & w^2 & w^4 & w^6 \\ w^0 & w^3 & w^6 & w^1 & w^4 & w^7 & w^2 & w^5 \\ w^0 & w^4 & w^0 & w^4 & w^0 & w^4 & w^0 & w^4 \\ w^0 & w^5 & w^2 & w^7 & w^4 & w^1 & w^6 & w^3 \\ w^0 & w^6 & w^4 & w^2 & w^0 & w^6 & w^4 & w^2 \\ w^0 & w^7 & w^6 & w^5 & w^4 & w^3 & w^2 & w^1 \end{bmatrix}$$

Fig 1.3: Matriz de la transformada de Fourier con $N = 8$.

Ya que la DFT es una transformación lineal reversible, no hay ni ganancia ni pérdida de información en este proceso. Sin embargo, ciertas manipulaciones simples de esta representación de la forma en el dominio de la frecuencia pueden eliminar la dependencia con la posición, tamaño y orientación.

Momentos

La Teoría de los Momentos fue aplicada por primera vez al reconocimiento de formas bidimensionales por Hu en el año 1962, [19], el cual obtuvo un sistema completo de invarianzas frente a traslaciones, homotecias y giros. Posteriormente se han realizado distintos trabajos en este tema por miembros de nuestro Departamento [20-21].

Según dicha teoría, cualquier imagen bidimensional puede representarse mediante una función $f(x,y)$, respecto de un par de ejes fijos en el campo visual. Una imagen puede representarse por los momentos espaciales de su función de distribución, definiéndose el momento generalizado jk , de orden $j+k$, de la siguiente forma

$$m_{jk} = \iint_{-\infty}^{\infty} N_{jk}(x, y) f(x, y) \, dx dy$$

donde $N_{jk}(x,y)$ es una función general de x e y , por ejemplo $x^j y^k$, en cuyo caso la ecuación anterior constituye la definición usual de los momentos, que es la que se utiliza en la mayoría de los casos, quedando entonces la ecuación anterior de la forma:

$$m_{jk} = \iint_{-\infty}^{\infty} x^j y^k f(x, y) dx dy; \quad j, k = 0, 1, 2, \dots$$

Si $f(x,y)$ es continua, acotada y puede tomar valores distintos de cero sólo en la parte finita del plano xy , entonces existen momentos de cualquier orden y se verifica el teorema de unicidad: La secuencia de momentos $\{m_{jk}\}$ e, inversamente, $f(x,y)$ está unívocamente determinada por $\{m_{jk}\}$.

Es decir, podemos reemplazar una descripción total $f(x,y)$, correspondiente a un conjunto infinito de intensidades, por una descripción equivalente $\{m_{jk}\}$, un conjunto infinito de momentos, y viceversa.

Los momentos anteriormente citados no resultan, por definición, ser invariantes frente a transformaciones afines (traslaciones, homotecias y giros) del objeto a caracterizar. No obstante es posible definir a partir de los mismos un espacio de representación que sí lo sea.

Para la normalización frente a traslaciones se calculan los momentos centrales μ_{jk} definidos como:

$$\mu_{jk} = \iint (x-x_c)^j (y-y_c)^k f(x,y) dx dy$$

que son invariantes frente a traslaciones.

Respecto a la normalización frente a homotecias, existen diferentes formas de llevarlo a cabo, la más extendida de las cuales consiste en trabajar con los nuevos momentos:

$$\eta_{jk} = \frac{\mu_{jk}}{\mu^{(j+k+2)/2}}$$

que son invariantes frente a traslaciones y homotecias.

Por último para la obtención de momentos normalizados frente a giros son dos los métodos más utilizados, a saber:

a) Método de los ejes principales, que son aquellos solidarios a la imagen y que forman un ángulo respecto de los ejes (x,y). Los momentos determinados respecto de dichos ejes son independientes de la orientación de la imagen.

b) Método de los invariantes absolutos, desarrollado por Hu a partir de la teoría de invariantes algebraicos. En la construcción de estos momentos se emplean los momentos centrales normalizados frente a homotecias, y para los momentos de orden dos y tres, éstos son:

$$I_1 = n_{20} + n_{02}$$

$$I_2 = (n_{20} - n_{02})^2 + 4n_{11}$$

$$I_3 = (n_{30} - 3n_{12})^2 + (3n_{21} - n_{03})^2$$

$$I_4 = (n_{31} + n_{12})^2 + (n_{21} + n_{03})^2$$

$$I_5 = (n_{30} - 3n_{12})(n_{30} + n_{12}) [(n_{30} + n_{12})^2 - 3(n_{21} + n_{03})^2] \\ + (3n_{21} - n_{03})(n_{21} + n_{03}) [3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2]$$

$$I_6 = (n_{20} - n_{02}) [(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2] \\ + 4n_{11}(n_{30} + n_{12})(n_{21} + n_{03})$$

$$I_7 = (3n_{12} - n_{30})(n_{30} + n_{12}) [(n_{30} + n_{12})^2 - 3(n_{21} + n_{03})^2] \\ + (3n_{12} - n_{03})(n_{21} + n_{03}) [3(n_{30} + n_{12})^2 - (n_{21} + n_{03})^2]$$

1.3 PROCESO DE DATOS EN RETINA Y PROYECCION AL CORTEX VISUAL

Distintos trabajos de Moreno-Díaz y sus colaboradores [22,23,24] han puesto de manifiesto que el análisis apropiado de los procesos retinales ha de basarse en los conceptos de computación por capas, contenidos en germen en los escritos de McCulloch y Pitts [25], en relación a la corteza visual. En este sentido, la retina de los vertebrados puede considerarse como un sistema que procesa por capas la información visual procedente del exterior; ésta información progresa tanto longitudinal como transversalmente, implicando en este recorrido a todas las neuronas retinales, desde las células fotorreceptoras a las ganglionares. El objetivo final de una teoría de proceso de datos en la retina consiste en explicar los mecanismos que ocurren en las diferentes capas y la codificación de las correspondientes células retinales. Desde el punto de vista neurofisiológico, la codificación está soportada por los registros de señales eléctricas (trenes de pulsos de las ganglionares y potenciales lentos de las células intrarretinales). Debido a la dificultad de obtener registros de la actividad eléctrica en casi todas las neuronas retinales con excepción de las ganglionares y parcialmente de las horizontales, los estudios retinales se centran fundamentalmente en dichas células.

1.3.1 Estructura por capas: relevancia de la C.P.I.

La retina de los vertebrados en un amplio espectro de especies desde los peces, anfibios, reptiles y aves, hasta los mamíferos como el gato, el mono y el hombre, presenta una estructura anatómica por capas de células, que tiene un claro sentido funcional. Hay abundante evidencia sobre las peculiaridades funcionales de las capas nucleares externa e interna, plexiforme y ganglionar. Esta estructura operativa por capas ha sido también encontrada en vías más altas

del sistema visual, como, por ejemplo, en la bien definida estructura de 4 capas de proceso en el tectum de la rana.

En las especies que carecen de cortex -o es muy rudimentario (especies inferiores)-, el tectum es el centro visual dominante. Las descargas que conduce el nervio óptico están mas fuertemente relacionadas con determinadas formas del estímulo visual y no proporcionan un mapa detallado de la iluminación sobre la retina, como ocurre en las especies superiores. Así, en la rana las ganglionares son complejas en su mayor parte y realizan una operación mas elaborada y especializada que en el gato o en el mono, donde son fundamentalmente simples.

Los campos receptivos retinales, estudiados en un número considerable de especies, pueden clasificarse en dos grandes grupos: aquellos que son selectivos para alguna característica espacial o temporal del estímulo, y aquellos que son no selectivos. Sorprende, en principio, que esta mayor especificidad y eficiencia en la forma de codificar la información esté correlacionada inversamente con una buena discriminación de formas, la cual está mucho mas desarrollada en los primates (casi carentes de campos selectivos), y menos en las aves y los anfibios (cuyos campos son selectivos).

Se observa que en los vertebrados superiores la selectividad aparece retrasada al nivel cortical. Una posible explicación es que la selectividad en el estímulo implica una pérdida de generalidad, y además se requiere una gran densidad celular para cubrir cada una de las variantes de una característica de disparo que pueden suceder en cualquier punto de la retina. Esta proliferación de células está presente en la corteza visual y no antes. Por otra parte, el procesamiento de dicha información por la corteza puede coincidir con otros procesamientos, tales como la visión binocular.

Puede decirse entonces que a medida que se desciende en la escala biológica las funciones operativas del cortex son realizadas por capas de proceso inferiores, a la vez que pierden versatilidad y calidad, e incluso la retina participa en mayor grado en estas tareas. Las distintas ganglionares de los anfibios y las aves muestran una arborización dendrítica que recoge información de estratos bien definidos de la capa plexiforme interna. En casi todas las especies junto a algunas células de procesos difusos la mayor parte termina sus arborizaciones en capas determinadas de la retina donde encuentran las señales que a su vez van a procesar.

Además, los estímulos existentes a nivel de fotorreceptores se diferencian de los que hay a nivel de bipolares, y otro tanto de los que constituyen las señales aferentes a las ganglionares. Sin embargo, estas señales -aún de varios tipos- son similares entre sí en cada capa.

Por otra parte, existe una relativa semejanza estructural de las distintas retinas, a las que debe acompañar un esquema funcional general también semejante.

Así pues, una primera base para la explicación del proceso visual en la retina está dada por su estructura operativa por capas. Ahora bien, la correspondencia entre las capas de proceso que nos sirven para distinguir los pasos sucesivos en los procesos de cómputo, y las capas reales anatómicas y funcionales, no es necesariamente biunívoca.

En segundo lugar la capa plexiforme interna (CPI) es muy compleja y su papel parece ser bastante relevante. Existe en la retina una gran variedad de células ganglionares según su morfología, y más significativamente, según su fisiología.

Sus campos receptivos pueden ser concéntricos o no, antagonistas, con anillo inhibitor, con respuestas ON, OFF, ON-OFF, sostenidas o transitorias, con direccionalidad mas o menos acusada, y mayor o menor sensibilidad al movimiento, con respuesta estacionaria o no, con respuesta espontánea, etc. Sin embargo, cada tipo específico de ganglionar no tiene sus propias y exclusivas bipolares. Así pues, las ganglionares de distinto tipo deben compartir la información procedentes de las mismas células bipolares, lo que conduce a que las peculiaridades diferenciales son debidas a operaciones que se realizan en la CPI y/o en el cuerpo de la ganglionar.

Las capas externas de la retina deben entonces proporcionar a las ganglionares el mínimo conjunto de señales que les permita construir sus respuestas especializadas. Este mínimo conjunto de operaciones puede consistir, como veremos, en un cierto grado de extracción de propiedades locales (mayor en las especies inferiores), tales como detección de abrillantamiento local, oscurecimiento y contraste. Para esta extracción las capas exteriores ofrecen dispersión local de la señal puntual, retrasos en las cadenas que implican varias células e interacciones laterales entre sus procesos. Estos niveles de señales son entregados a la CPI.

La CPI es lógicamente un lugar adecuado para la extracción de propiedades de los estímulos en la retina, puesto que la información espacial está todavía allí presente, antes de su posterior integración por las ganglionares, y porque esta capa está interconectada en horizontal profusamente. Además, las amacrinas y sus sinapsis con las ganglionares parecen estar relacionadas con operaciones complejas. Como la anatomía de la CPI evidencia, la densidad de sinapsis en retinas complejas es mucho mayor que en las simples. Las células ganglionares simples son excitadas principalmente por contacto directo con las bipolares, mientras que las complejas muestran un gran número de sinapsis con amacrinas.

Por todo ello, la CPI parece ser el lugar definitivo donde ocurre la especialización, a través de una interacción lateral, en la que las células amacrinas son importantes.

1.3.2 El concepto de Preproceso

Consideremos ahora la capa plexiforme externa y las capas nucleares externa e interna. Ellas constituyen las capas de preproceso. De especies inferiores a superiores se observa una variación desde muy drásticas no linealidades en las ganglionares, a pocas no linealidades. Así, en la rana, las respuestas de las ganglionares son poco dependientes de contrastes y de otros parámetros del estímulo, estando relacionadas mas bien con las operaciones previas realizadas por el preproceso.

Tambien se ha observado que los estímulos óptimos para la rana no son del tamaño del campo receptivo, sino del orden de la mitad de éste, lo que se explica facilmente cuando la ganglionar actúa sobre una detección de contraste previa [24].

Ya a nivel de bipolares se registran campos receptivos organizados antagónicamente centro-periferia, con respuestas de potenciales lentos, diferenciándose de las descargas impulsivas ON u OFF de las ganglionares simples. Diversos autores han hecho notar el esparcimiento horizontal de estimulación que se da ya a nivel de contactos entre los pedículos de los fotorreceptores, así como su mútua interacción; asimismo, el papel fundamental en este traslado de la información -que incluirá probablemente un cierto retraso- es juzgado por las células horizontales. En principio este esparcimiento debe ser isótropo, ya que en los vertebrados superiores las células ganglionares anisótropas son la excepción y no la regla.

La mayor o menor lejanía de las sinapsis dendríticas del cuerpo celular puede ser responsable, a causa del amortiguamiento en la transmisión del potencial, del mecanismo de ponderación que realizan las células: desde excitación prácticamente uniforme hasta grandes variaciones en la misma, según su proximidad al centro.

Por último es conocida la existencia de umbrales en las capas externas, de cuyo mecanismo resultan no linealidades locales.

El fenómeno de adaptación al estímulo luminoso es debido a los umbrales de estas capas, umbrales que deben ser función de las señales sobre grandes zonas de la retina, lo que introduce de nuevo fenómenos no lineales. Este resultado es sugerido también por las ganglionares desprovistas de actividad espontánea, pues en ellas no están presentes los resultados negativos de las operaciones de preproceso, indicando que las señales al menos son recortadas por el umbral adaptivo. Otras ganglionares con respuestas poco dependientes de la intensidad de la luz y del contraste sugieren incluso una acción no lineal drástica.

De esta manera las capas exteriores de la retina poseen la estructura anatómico-fisiológica necesaria para la realización de un conjunto mínimo de operaciones de preproceso. Al menos pueden estar presentes señales rápidas y retardadas, de abrillantamiento, contraste y oscurecimiento local, lo que depende de la estructura retinal concreta en cada especie.

1.3.3 Interacción lateral no lineal en la C.P.I.

Se ha indicado antes la capacidad de ponderación de las excitaciones recibidas en una arborización dendrítica, sobre todo en las ganglionares. Además, estos

potenciales lentos excitatorios e inhibitorios postsinápticos son integrados en el cuerpo ganglionar dando origen a la excitación generadora del potencial de acción. La frecuencia instantánea de repetición de impulsos aparece en muchos registros fuertemente correlacionada con la amplitud de la excitación.

La dispersión en estratos de algunos procesos dendríticos de ganglionares, y en todo caso la existencia de distintos niveles de señales de las capas externas, da pie a pensar que las operaciones de ponderación e integración son realizadas de forma separada para cada nivel. Posteriormente estas señales pueden actuar como excitadoras o simplemente como facilitadoras o desfacilitadoras en la generación de potencial, lo que conduce a una interacción integrada de tipo multiplicativo, y por lo tanto a la no linealidad en la función que relaciona la frecuencia instantánea de disparo con las informaciones que manipula el cuerpo ganglionar.

Todo elemento neuronal se caracteriza también por sus constantes de tiempo de respuesta. Así, en las células ganglionares existen límites superiores de respuesta que implican no linealidades por saturación.

Las células ganglionares podrían ofrecer capacidad de memoria corta en algún grado, tal como ocurre en otras unidades, por ejemplo las bipolares. Sin embargo, no parece existir evidencia de actuación de mecanismos de aprendizaje, a nivel ganglionar (o bien es despreciable), por lo que hemos de considerarlas aproximadamente como células de memoria de largo término nula.

1.3.4 Esquema conceptual de la Operación retinal.

Acción de las capas exteriores

Con los conceptos anteriores y a modo de resumen, el funcionamiento hipotético de la retina puede sintetizarse como sigue:

Los datos espacio-temporales de entrada a la retina son procesados por los fotorreceptores, bipolares y horizontales de las capas exteriores, ofreciendo canales o niveles de propiedades espacio-temporales, según la especie en cuestión. La información de cada canal es el resultado de operaciones que pueden ser cualitativamente descompuestas en una transformación lineal espacio-temporal seguida de umbrales adaptivos.

En la CPI cada uno de los canales interactúa lateralmente de un modo peculiar para cada tipo de ganglionar. Esta interacción lateral no lineal estaría gobernada por las sinapsis bipolar-ganglionar y amacrina-ganglionar.

Tal como se ha concluido en trabajos anteriores [26,27], las capas exteriores generan versiones con diversos retardos de las señales que al interactuar en la capa plexiforme interna explican las propiedades tanto de las ganglionares simples como de las complejas y codificadoras de color.

Las retinas simples y no especializadas como por ejemplo la del gato, indican que las señales procedentes de las capas exteriores son dos versiones espacio temporales de datos de entrada transformados cuasi linealmente: una señal rápida y una señal retardada. Aunque el papel de las células horizontales no está del todo clarificado, es posible que estas desempeñen un papel en la generación de las versiones retardadas, las cuales son trasladadas lateralmente.

Para una retina continua, bajo simples suposiciones espacio temporales y considerando solo términos de primer orden, la expresión de lo anterior según la notación convencional del análisis de sistemas es:

$$f_F(r, s) = H_F(s) \int_{r'} W_F(r, r') I(r', s) dr \quad (1)$$

$$f_R(r, s) = f_F(r, s) / (1 + \tau s) \quad (2)$$

donde r y r' son las correspondientes coordenadas espaciales, f_F y f_R son las transformadas de Laplace de las señales rápidas y retardadas, y $I(r',s)$ es la transformada de alguna función no lineal de la intensidad de la luz incidente en el punto r' (e.g una función logarítmica). W_F y H_F corresponden a una transformación espacio temporales de las capas externas, que corresponde esencialmente a un filtro paso baja espacio temporal.

Posteriormente, las señales f_F y f_R interactúan lateralmente en la capa plexiforme interna. Esta interacción puede descomponerse en una inhibición lateral lineal seguida de una no linealidad local de rectificación. Las dos posibilidades resultantes son inhibición lateral lineal de f_F por f_R y viceversa.

La inhibición resulta de pequeñas superficies S_R o S_F , de áreas A_R y A_F . $U(r',r)$ es una función de valor 1 sobre dichas superficies y 0 en el exterior; K_R y K_F son los pesos constantes de la inhibición es decir:

$$x_1(r, s) = f_F(r, s) - (k_R/A_R) \int_{r'} U(r, r') f_R(r', s) dr' \quad (3)$$

$$x_2(r, s) = f_R(r, s) - (k_F/A_F) \int_{r'} U(r, r') f_F(r', s) dr' \quad (4)$$

La no linealidad local de rectificación, genera las señales x_1' y x_2' dadas por:

$$x_1'(r, t) = \text{Pos} (L^{-1} (x_1(r, s))) \quad (5)$$

$$x_2'(r, t) = \text{Pos} (L^{-1} (x_2(r, s))) \quad (6)$$

L^{-1} es la transformada de Laplace inversa y $\text{Pos } x$ se define por: $\text{Pos } x = x$ para $x > 0$ y cero en otro caso.

Las señales x_1' y x_2' son posteriormente pesadas y sumadas, tal que para una célula ganglionar en el origen, su frecuencia instantánea de disparo $G(t)$ está dada por:

$$G(t) = \text{Pos} \left\{ \int_r \int_{t'} k_1(t-t', r) x_1(r, t') dt' dr + \int_r \int_{t'} k_2(t-t', r) x_2'(r, t') dt' dr + G_0 \right\} \quad (7)$$

siendo G_0 la respuesta espontánea, cuando existe.

Estas bases, unificadas por la formalización matemática [28] son apropiadas a nivel conceptual para elucidar los mecanismos que determinan la totalidad de los procesos retinales en los vertebrados, tal como han sido registrados en neurofisiología.

La completa sistematización y generalización de estos conceptos no ha sido realizada por anatomistas y neurofisiólogos posiblemente por no poseer - o quizá no estar interesados en ellas - las herramientas de la moderna teoría de sistemas (lineales y no lineales), que son precisas para la formulación de una verdadera teoría del sistema visual de la retina.

En efecto, algunos aspectos (los aspectos puramente temporales) de la formulación matemática caen de lleno dentro de la ortodoxa o clásica teoría general de análisis de sistemas de la ingeniería eléctrica y de comunicaciones. Pero la dificultad más importante ocurre con el carácter altamente no lineal de los procesos espacio-temporales de las neuronas. La teoría de sistemas (análisis e identificación) ofrece gran ayuda, y también los conceptos de codificación y transmisión familiares en las técnicas de telecomunicaciones. Sin embargo, el sistema vivo de proceso en la retina tiene una complejidad poco habitual en los sistemas físicos, añadiendo aspectos nuevos, espaciales, con los que el ingeniero eléctrico no se había enfrentado. De esta complejidad se beneficia, asimismo, la teoría de sistemas, que encuentra así una fuerte motivación adicional en su desarrollo para cubrir las necesidades surgidas en la descripción de la operación retinal.

Generalización de la operación retinal a nivel de las ganglionares.-

Los retardos de la señal en las capas exteriores son suficientes para explicar con éxito las operaciones retinales a nivel de las ganglionares simples y complejas en el gato, las aves e incluso en la rana esculenta. Sin embargo, para la rana *pipiens* ha sido necesario introducir altas no linealidades, como son la facilitación y desfacilitación multiplicativas. ¿Cómo pasar de esas no linealidades drásticas a las expresiones casi lineales de las ganglionares del gato, el mono y quizá el hombre?

Un primer enfoque será proveer el sistema con desarrollo funcional en serie de potencias, pudiéndose restringir, para el caso lineal, a los primeros términos. Esto equivale a efectuar una descripción del sistema no lineal del tipo de las propuestas por Volterra y Wiener o Marmarellis. Sin embargo, es muy difícil de justificar y verificar experimental el que por el hecho de que menos amacrinas conecten con una ganglionar, o porque haya pocas fibras centrífugas, ésta se decida por los primeros términos de un desarrollo de potencias que en definitiva afecta a productos múltiples de las señales. En cambio, por otro lado, está el probado efecto de las fibras centrífugas aferentes a la retina sobre la extensión de los campos receptivos ganglionares, si bien no sobre su cualidad. Es claro que este efecto no se refiere a las señales, sino a la naturaleza del sistema.

Las formulaciones clásicas de Volterra o Wiener de la teoría de sistemas no lineales mantienen una descripción externa del sistema, como si fuera una "caja negra". Este modo de operación apenas resulta útil en la teoría que, pese a su generalidad, puede calificarse desde nuestro enfoque como "ciega", en cuanto ignora la naturaleza del sistema. En el mejor de los casos sabremos (incluso en detalle) qué hace el sistema, pero muy poco cómo lo hace.

Surge también la limitación de trabajar con señales monocanales o en todo caso multicanales (multientrada), pero pocas e independientes entre sí, sin interacción espacio temporal de importancia. Más en realidad muchos sistemas biológicos presentan aspectos espaciales no abordables desde esa perspectiva.

Por otra parte, partiendo de una representación geométrica cualquiera de un sistema, ¿cómo pensar que para procesar la señal requiera la realización de convoluciones sobre los productos dobles, triples, etc. de las señales de entrada? ¿No será que esa representación encubre un proceso adaptativo no lineal? En ese caso tales productos deben corresponder a núcleos de interacción y no a las señales. Supongamos un proceso discreto en el tiempo. Entonces en el instante t el núcleo de interacción es función de lo procesado en instantes anteriores, o bien está controlado por los resultados de un proceso superior. (Esto es precisamente lo que los resultados experimentales en la retina de las aves sugieren.)

En la representación funcional de Volterra o la de Wiener lo más característico del procesamiento de un sistema no lineal es el núcleo o peso con que contribuye el valor que tiene la señal en cada uno de los puntos de su espacio de variabilidad expandido, en la construcción de la salida. Esta construcción es una suma (lineal) en cualquier caso. Esta expansión del espacio de entrada en $1, 2, \dots, n$ espacios superpuestos se precisa para tratar adecuadamente la influencia (peso) que tiene no sólo la señal puntualmente considerada, sino como un todo, que incluye todas sus tendencias propias (derivadas espaciales y temporales) y forma definitiva y peculiar de sucederse a sí misma en espacio y tiempo. Ahora bien, esto no es debido a la señal, sino que viene impuesto por el sistema siempre que éste tenga sensibilidad a la señal, es decir, tenga alguna capacidad de discriminar sus valores (respondiendo a ellos selectivamente), y tenga quizá también capacidad de compararlos o relacionarlos.

Se precisa, pues, en primer lugar, sensibilidad a la señal, lo que requiere núcleos adaptivos. En segundo lugar, puede existir capacidad de relacionar los valores de la señal que han penetrado en el sistema, lo que ocurre si tiene "memoria" temporal o espacial. La ausencia de memoria temporal indica que los núcleos se adaptan instantáneamente al valor de la señal. La falta de memoria espacial indica que los núcleos se adaptan puntualmente al valor de la señal: son locales. El grado de adaptación o sensibilidad es, precisamente, el grado de no linealidad del sistema. En particular, si no se adaptan a la señal, el sistema es lineal.

La transparencia y la consistencia ha conducido a los núcleos de interacción. Pero una vez allí el socorrido desarrollo en funciones ortonormales no es utilizable.

Así pues, se ha de abandonar el enfoque de desarrollo funcional en serie de potencias de Volterra. Es preciso considerar que los núcleos de operación son núcleos adaptivos, de acuerdo con la señal de entrada, aunque invariantes (no evolutivos). Para ello una reciente formulación de Moreno-Díaz [29], nos pone en el camino adecuado de lograr la generalización de la operación retinal de las ganglionares.

A nivel cualitativo, podemos llegar a algunas consecuencias relevantes. En los vertebrados superiores la mayoría de los núcleos de interacción - las propias ganglionares - toman escasa cuenta de lo que se hace en centros superiores o en ellas mismas, al contrario que en los vertebrados inferiores. Se plantea entonces una cuestión experimental: ¿a medida que avanzamos en la escala biológica visual disminuye la influencia de las fibras centrifugas sobre las ganglionares de la retina?

Según ello, una ganglionar es "casi lineal", cuando su propia realimentación sobre el núcleo o la realimentación desde la corteza es casi nula. En otro caso sus propias funciones de peso sinápticas dependen o de la historia global tal como sea interpretada por los centros visuales superiores o de su propia historia si estas ganglionares muy especializadas - como en la rana *pipiens* - son sus propios centros superiores. La mayor realimentación procederá en estos casos del tectum ópticum.

Entonces, los aparentes efectos de facilitación y desfacilitación que dan lugar a los resultados experimentales de "gating" en la rana *pipiens* pueden proceder del tectum vía las fibras centrifugas. Si se seccionan éstas aparecerán los efectos más localistas y cuantitativos.

Y otra consecuencia. Los resultados en la rana *pipiens* se refieren a registros en el tectum ópticum. Así, esta retina y la del gato parecen muy lejanas. Pero un bloqueo de las fibras centrifugas y el registro en los propios axones ganglionares han de llevar a que la función retinal de animales tan lejanos en la escala como el gato y la rana presentan, en esencia, el mismo tipo de operaciones, aunque con muy distinta distribución entre fibras especializadas o no en cada especie.

CAPITULO 2

TRANSFORMACIONES NO LINEALES

2.1 LINEALIDAD, NO LINEALIDAD Y COMPLITUD

Un concepto básico en proceso de imágenes y en visión es el de "descripción completa". Una descripción se admite que es "completa" en su entorno visual, si la descripción contiene todos los datos, y propiedades de los datos, necesarios para cubrir unos objetivos.

A nivel analítico, los objetivos que se establecen con frecuencia consisten en admitir que una descripción es completa cuando de ella puede recuperarse el entorno visual.

Dentro de la teoría de sistemas, una transformación corresponde a la acción de un sistema sobre el espacio de las entradas y quizás el de las salidas, para generar un nuevo "slide" del espacio de salidas.

Los conceptos de complitud de una transformación con relación al campo receptivo y a la función fueron introducidos por Candela en 1987 [30]. Según dicho autor, y desde el punto de vista analítico, la descripción completa (que, según la heurística y el experimento, podrían ser "cortados" para un sistema de V.A. práctico), requiere "a priori", la conservación del número de grados de libertad, o del número de propiedades independientes, N , del entorno visual. Dicha constancia permitió establecer una especie de principio de conservación que viene determinado por los funcionales computados por F y por los campos receptivos, que, además reflejan una situación de dualidad, en el sentido de que la complitud requiere aumentar el número de ellos si el otro disminuye. Seguiremos profundizando en estos conceptos en el capítulo 3 al tratar las transformaciones de campos de datos, donde además generalizaremos el concepto de complitud.

Como ya indicamos en la sección 1.2, suele ser frecuente en visión artificial trabajar con descriptores procedentes de espacios transformados. La forma típica de plantear el problema consiste en expandir $f(x,y)$ en términos de coeficientes $F(u, v)$. La justificación de esta expansión para el caso de proceso de imágenes y reconocimiento visual, se basa en una conjetura que aún no ha sido reducida a teoremas formales, y por lo tanto, no tiene ningún sustento teórico.

Desde el punto de vista de la matemática finita, a la que se está avocando a priori en proceso de imágenes, el planteamiento es sencillo y transparente.

De acuerdo con esto, una imagen es un conjunto de valores f_i ($i=1\dots N^2$). Admitimos, a priori, que no hay posiciones privilegiadas ni minusvalorables (si las hubiese, adjudicaríamos más pixels -números i - aumentando la resolución de las zonas privilegiadas, y viceversa para las de minus-valor).

Desde el punto de vista de la teoría de sistemas, la representación lineal completa equivale a una transformación lineal

$$F_i = \sum_j \alpha_{ij} f_j$$

donde el núcleo de la transformación, α_{ij} , es tal que la transformada inversa

$$f_j = \sum_i \beta_{ji} F_i$$

exista.

Es decir, la generación de los coeficientes F_i es la salida de un sistema lineal, con función de transferencia α_{ij} , tal que el sistema lineal inverso existe, tal como se ilustra en la figura 2.1.

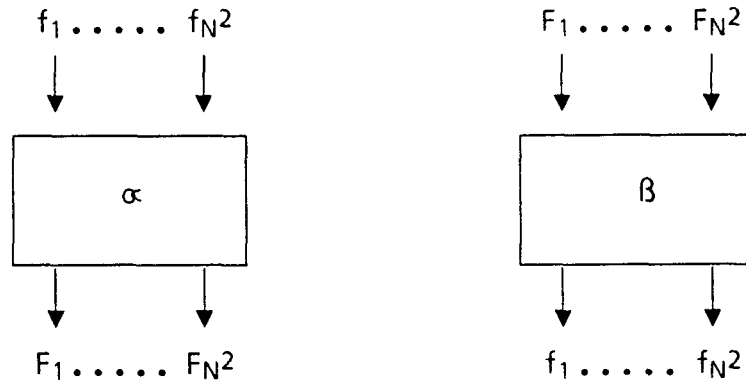


Figura 2.1: Representación de la generación de coeficientes de representación como la acción de un sistema lineal con Inverso.

Lo primero que salta a la vista en este planteamiento es que los α_{ij} pueden ser cualesquiera, con la única restricción de independencia lineal, o sea $|\alpha_{ij}| \neq 0$. Si son ortogonales, es decir, si

$$\sum \alpha_{il} \alpha_{lj} = \delta_{ij}$$

entonces, la analítica se simplifica considerablemente, aunque no necesariamente la algorítmica.

Esta representación permite generalizar fácilmente el concepto de coeficientes de representación, de tal modo que los coeficientes F_i ($1, 2, \dots, N^2$) forman un conjunto completo, no redundante, si éstos son obtenidos por cualquier transformación arbitraria ϕ_i

$$F_i = \varphi_j (f_1, \dots, f_N^2)$$

tal que la transformación inversa

$$f_j = \beta_j (F_1, \dots, F_N^2)$$

existe.

Es decir, los F_i son obtenidos por un sistema arbitrario con la condición de que el sistema inverso exista.

Estos conceptos de linealidad, no linealidad y complitud, serán tratados mas profundamente en posteriores apartados del presente trabajo.

2.2 PRODUCCION Y APLICACION DE TRANSFORMACIONES NO LINEALES.

Este apartado está dedicado esencialmente a transformaciones sobre toda la retina, transformaciones que caen típicamente dentro de lo que es considerado como proceso de imágenes -que no tiene el reconocimiento o la visión artificial como objetivo esencial- y que corresponde a lo que en un sistema visual se realiza como preproceso y proceso global de las imágenes. Ello sin menoscabo de que algunas de las transformaciones que consideramos puedan usarse para obtener descriptores de aplicación en la generación de espacios para el reconocimiento global visual.

2.2.1 Marco General para Transformaciones Globales.-

El marco general para la producción de transformaciones no lineales está basado en una propuesta de Moreno-Diaz y Rubio [31] y modificaciones posteriores [32], actualizada y adaptada a los objetivos del proceso visual.

Los conceptos básicos coinciden con los expuestos en los citados trabajos; pero al considerar aquí la regla de decisión como una transformación, en lugar de considerarla como definitoria o determinante de un "sistema", se introducen conceptos nuevos que son de crucial importancia en estos nuevos casos y que están relacionados con la complejidad de la transformación.

Consideremos un espacio de entrada, caracterizado por n subíndices (o de dimensión n), uno de los cuales es el tiempo y tal que el "lugar" $ij\dots n$ está ocupado por un número, $F_{ij\dots n}$, típicamente, real. En el caso de imágenes que nos ocupa, $F_{ij\dots n}$ indica, por ejemplo, la intensidad luminosa de una componente cromática, C , en un punto (x,y) en un instante t , es decir sería del tipo F_{cxyt} .

Puesto que, en definitiva, nos movemos en el dominio de los funcionales, las variables a considerar son precisamente esas F_{cxyt} para todos los c,x,y,t y **no** los subíndices.

Una transformación de $F_{ij\dots n}$ es un conjunto de reglas $R_{i'j'\dots n'}$ (una regla para cada posible n -tupla de los subíndices $i'j'\dots n'$), tal que, aplicadas a todos los números $F_{ij\dots n}$, generan un nuevo espacio $F_{i'j'\dots n'}$ (que es el espacio de salida). Es decir:

$$F'_{i'j'...n'} = R_{i'j'...n'} \{ \forall_{ijn} F_{ijn} \} \quad \text{Ec. 2.1}$$

para el caso de imágenes, se tiene:

$$F'_{c'x'y't'} = R_{c'x'y't'} \{ \forall_{cxyt} F_{cxyt} \} \quad \text{Ec. 2.2}$$

Para un espacio unidimensional:

$$F'_{i'} = R_{i'} \{ \forall_i F_i \} \quad \text{Ec.2.3}$$

Haremos el desarrollo teórico para una dimensión, ya que los resultados son fácilmente extensibles a mas dimensiones, teniendo cuidado de que cuando una de las dimensiones es el tiempo, se ha de estar de acuerdo con el principio de causalidad.

Se tiene que i e i' son subíndices, que toman los valores ordenados de los números naturales. Así, tenemos de inmediato la siguiente generalización.

2.2.2 Convoluciones Generalizadas

Una **Convolución Generalizada** es una regla de transformación tal que:

$$F'_{i'+k} = R_{i'} \{ \forall_i F_{i+k} \} \quad \forall_k \quad \text{Ec.2.4}$$

La regla $R_{i'}$ puede entonces ser la del elemento de índice $i' = 0$. Es decir, que, para una convolución generalizada, se tiene:

$$F'_k = R_0 \{ \forall_i F_{i+k} \} \quad \text{Ec. 2.5}$$

Por ejemplo, supongamos que en la ec. 2.3, la regla R_2 ($i' = 2$) sea la siguiente:

$$R_2 = \{ \text{Calcular } F_8 * F_{16}: \quad \text{Si } F_8 * F_{16} > F_{24} \text{ entonces } F_{2'} = F_{24} \\ \text{en otro caso } F_{2'} = 0 \}$$

Si la regla para cualquier $i' \neq 2$ es la R_2 , cambiando los subíndices de modo que:

$$R_{i'} = \{ \text{Calcular } a = F_{6+i'} * F_{14+i'}; \\ \text{Si } a > F_{22+i'} \quad \text{entonces} \quad F'_{i'} = F_{22+i'} \\ \text{si no} \quad \quad \quad F'_{i'} = 0 \}$$

se tendrá entonces una convolución generalizada. Como se ve la R_0 es tal que:

$$F'_0 = R_0 (F_6, F_{14}, F_{22}) \text{ , siendo:}$$

$$R_0: \left\{ \begin{array}{l} \text{Calcular } a = F_6 * F_{14}; \\ \text{si } a > F_{22} \text{ entonces } F_{0'} = F_{22} \\ \\ \text{si } a \leq F_{22} \text{ entonces } F_{0'} = 0 \end{array} \right\}$$

con lo que: $F'_k = R_0 \{F_{6+k}, F_{14+k}, F_{22+k}\}$

Una transformación multidimensional en la que existe una convolución generalizada para una (o mas) dimensiones, es invariante en esas dimensiones. Así, si la dimensión en la que existe la convolución generalizada es el tiempo, se tiene una transformación "invariante en el tiempo", es decir, en la que no existe aprendizaje.

La regla de decisión en las ecuaciones anteriores puede considerarse como una relación sobre los F_i ; y de hecho, en general, adopta la forma de un "programa" que, para el caso unidimensional, trabaja sobre los datos contenidos en los N ($i = 1 \dots N$) "lugares" de memoria, F_i .

Consideremos algunas otras ilustraciones de convoluciones generalizadas que servirán para establecer con claridad el alcance de la formulación.

Primero, es inmediato ver que en el caso de que la regla R_i' sea una regla algebraica y lineal, se tiene la convolución convencional. En efecto, la regla:

$$F'_{i'} = \left\{ \sum_j \alpha_{i',j} F_j \right\} \tag{Ec. 2.6}$$

— corresponde a una transformación lineal arbitraria. Poniendo la condición de la ecuación 2.5, resulta la regla R_o

$$R_o \longrightarrow \sum_i \alpha_{oi} F_i$$

que con la condición 2.4:

$$\begin{aligned} F'_k &= \sum_i \alpha_{oi} F_{i+k} = \\ &= \sum_i \alpha_i F_{i+k} = (\alpha_{oi} = \alpha_i) \end{aligned} \quad 2.7$$

El proceso se ilustra en la figura 2.2. Los α_i son los "factores de peso", o el núcleo de la convolución, referidos al origen.

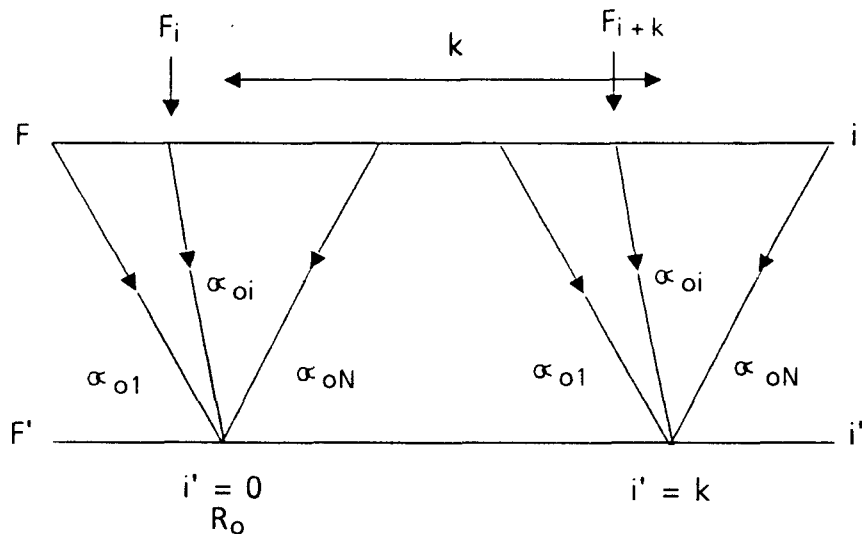


Figura 2.2: Ilustración de una convolución standar como un caso particular (lineal) de una convolución generalizada.

El cambio de subíndices:

$$i + k = j; \quad i = j - k$$

permite escribir la ecuación 2.7 en la forma mas usual:

$$F'_k = \sum_j \alpha_{j-k} F_j$$

que en el caso continuo, queda:

$$F'(x) = \int \alpha(x'-x) F(x') dx'$$

ecuación familiar de la convolución, con el núcleo de diferencias $\alpha(x'-x)$.

Casos interesantes -que asimismo se corresponden con convoluciones lineales que actúan localmente- son los operadores diferenciales. Por ejemplo, el operador laplaciana, detector de contrastes, en una dimensión es:

$$-\frac{1}{2} \frac{d^2F}{dx^2} = F_i - \frac{1}{2} [F_{i-1} + F_{i+1}]$$

es decir, en términos de convolución:

$$F'_k = F_k - \frac{1}{2} [F_{-1+k} + F_{1+k}]$$

Ilustraciones de convoluciones generalizadas no algebraicas (algorítmicas) han sido desarrolladas por Moreno-Diaz Y Mira [33] y por Santana et al, [34]. A título de ejemplo presentamos un proceso correspondiente a una inhibición lateral simbólica:

Sean C y P las áreas central y periférica correspondientes a un punto (x,y) de una imagen. La inhibición lateral simbólica requiere la definición de un proceso A en C y un proceso B en P. De esta forma P inhibe a C de acuerdo con una regla de decisión que tiene en cuenta los resultados de los procesos A y B.

Este algoritmo simple pero potente para la inhibición lateral simbólica es tal que los procesos A y B son procesos de selección del máximo. Es decir:

$$\text{Proceso A Select Max (C) = } C_m$$

$$\text{Proceso B Select Max (Abs(P)) = } P_m$$

$C_m(x,y)$ es el valor máximo de las señales presentes en el área central del punto (x,y). $P_m(x,y)$ es el correspondiente máximo de los valores absolutos en la periferia. La regla de decisión (D,R) es tal que si $F(x,y)$ es la salida en el punto (x,y):

Regla de decisión ---- Si $C_m \geq P_m$ entonces

$$F(x,y) = C(x,y)$$

en otro caso

$$F(x,y) = -P_m$$

Este algoritmo es apropiado para la generación de ondas, como ha sido apuntado por los citados autores.

2.2.3 Producción de Convoluciones Generalizadas por reglas analíticas

Como se ha visto en la sección anterior, una convolución generalizada se basa en una única regla que se aplica a todos o parte de los valores F_i contenidos en los lugares i . A partir de la regla del elemento en posición cero (F'_0), R_0 , la regla para el elemento en la posición k se obtiene sustituyendo cada F_i en la regla por F_{i+k} .

Una familia importante de convoluciones generalizadas es la que caracteriza los sistemas no lineales polinómicos, invariantes en el tiempo-espacio, que corresponde a la caracterización Volterra-Wiener. En esta situación, se admite una regla, R_0 , expresable por una función analítica de los F_i (que aparecen como variables). Nótese que, estrictamente, la regla R_0 es un funcional, ya que, desde el punto de vista del análisis clásico, las variables no son las F_i , sino los subíndices i , ésto es las direcciones de "memoria". Las incidencias análisis-estructura computacional serán reconsideradas en el capítulo 3.

Si la regla R_o es una expresión analítica, puede desarrollarse en serie de potencias. Para una dimensión es un desarrollo en N variables (subíndices), de tal forma que R_o es:

$$R_o = R_o(0, \dots, 0) + \sum_i (\delta R_o / \delta F_i)_{F_i=0} F_i + \sum_i \sum_j (\delta^2 R_o / \delta F_i \delta F_j)_{F_i=F_j=0} F_i F_j + \dots$$

donde W_i, W_{ij} , etc, son:

$$W_i = (\delta R_o / \delta F_i)_{F_i=0} \quad ; \quad W_{ij} = (\delta^2 R_o / \delta F_i \delta F_j)_{F_i=F_j=0}$$

correspondientes a los Kernels de 1° orden, 2° orden,... introducidos en la teoría de sistemas no lineales.

2.3 SISTEMAS NO LINEALES POLINOMICOS

Como ya indicamos, un caso particular interesante de sistema no lineales son los sistemas polinómicos, descriptibles por medio de los núcleos de orden 1,2,...,n de Wiener. Moreno-Diaz y Rubio [28], demuestran facilmente gracias a la notación sencilla -la misma aquí usada- que un sistema polinómico es equivalente a un **expansor** del número de funciones de entrada seguido de un sistema lineal, tal como se ilustra en la figura 2.3

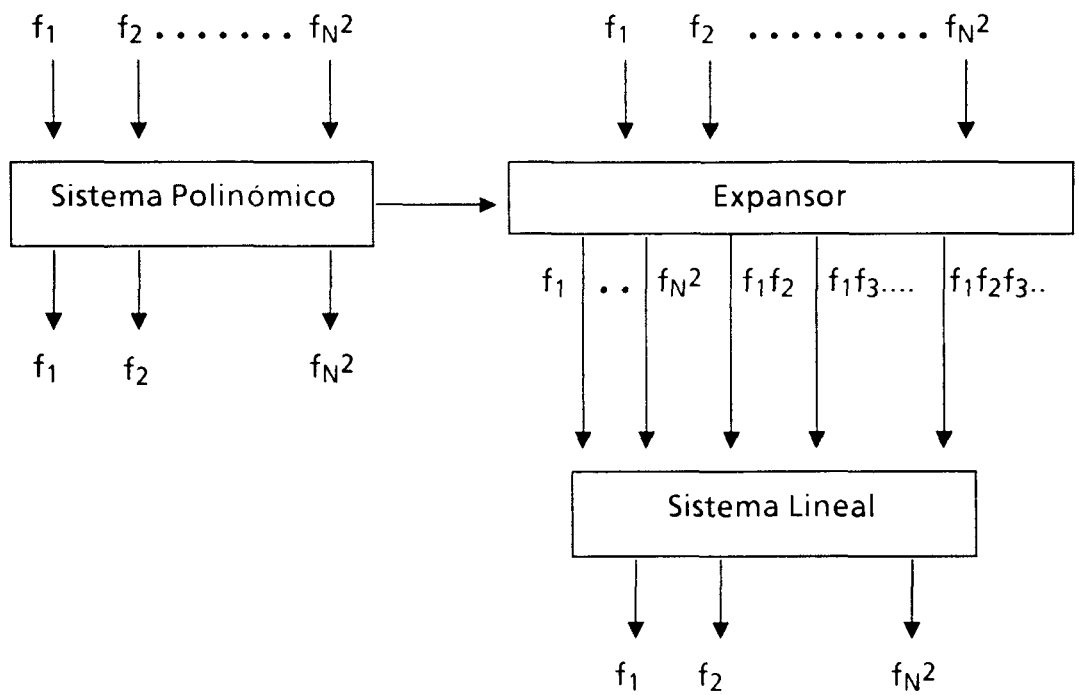


Figura 2.3

Investigar, en general, las condiciones necesarias y suficientes para la existencia del sistema inverso a uno polinómico dado cae fuera de los objetivos de este trabajo, aunque es una cuestión a estudiar y a investigar experimentalmente sobre imágenes, en el futuro. Consideraremos, como ilustración, el caso de un sistema de segundo orden, es decir, un sistema que vendría caracterizado hasta los núcleos de Wiener de orden dos. Desde el punto de vista de la Teoría de Sistemas, preocuparía determinar si la clase de sistemas representables por este polinomio, en la vida real justifica su estudio. Desde nuestro punto de vista, esta preocupación no existe, por la misma razón que la representación por Fourier o Momentos en proceso de imágenes no descansa en la existencia de sistemas naturales o artificiales que a priori realizan tal transformación, sino en su utilidad posterior.

Un sistema no lineal polinómico de orden dos comporta los coeficientes F_i :

$$F_i = \sum_j \alpha_{ij} f_j + \sum_j \sum_k \alpha_{ijk} f_j f_k$$

Introduciendo las nuevas variables funcionales $\Gamma_{jk} = f_j f_k$ y redenominando los índices tales que, a cada pareja jk le hacemos corresponder un nuevo índice l ($l=1, \dots, N^4$), resulta

$$\Gamma_l = \Gamma_{jk} = f_j f_k$$

Es decir

$$F_i = \sum_{j=1}^{N^2} \alpha_{ij} f_j + \sum_{l=1}^{N^4} \alpha_{il} \Gamma_l$$

o sea, un sistema lineal **expandido** y redundante, tal como preveé el teorema sobre representación lineal de sistemas polinómicos ya citado.

Renombrando las variables, resulta:

$$F_i = \sum_1^{N^2 + N^4} \alpha_{ij} \Gamma_j \quad (i = 1, \dots, N^2)$$

donde α_{ij} es una matriz de $N^2 \times (N^2 + N^4)$.

En principio, la solución inversa de la ecuación anterior, con más incógnitas funcionales $\{f_j\}$ que variables funcionales $\{F_i\}$ lleva a un problema diofantino en funcionales. Aunque ello pueda ser estrictamente el caso, el número de variables funcionales independientes no es $N^2 + N^4$ ya que, obviamente, por encima del índice N^2 , las nuevas variables funcionales no son independientes de las anteriores.

Pueden hallarse con cierta facilidad condiciones suficientes para la existencia del sistema inverso.

1°) Las variables funcionales presentes en la expansión han de ser independientes, para evitar redundancia. Por ejemplo, si f_i y f_j están presentes, no lo puede estar $f_i f_j$. Igualmente, si f_i está presente, no lo podrá estar f_i^2 y viceversa.

2°) Lo anterior reduce a N^2 , número original de grados de libertad, el número de los $\{f_{ij}\}$ que están presentes en la expansión. Además, si $\{f'_{ij}\}$ es la matriz reducida, debe cumplirse $|\{f'_{ij}\}| \neq 0$.

3°) En general, ningún f_i debe ser cero. Este es un requisito fácil de conseguir en el preproceso de imágenes, por medio de una traslación del origen de intensidades.

Según esto, si se cumplen los requisitos anteriores, la transformación no lineal inversa existe tal como ha indicado J.A. Muñoz en su Tesis Doctoral [35].

TEOREMA: Si una representación polinómica de segundo grado cumple 1°, 2° y 3°, existe la transformación inversa, que viene dada por las expansiones y reglas de solución expresadas a continuación:

En efecto, partimos de la representación

$$F_i = \sum_1^{N^2} \alpha_{ij} \Gamma_j \quad \text{con } \alpha'_{ij} \neq 0$$

computamos

$$\Gamma_i = \sum_1^{N^2} \beta_{ij} F_j \quad \text{donde } \beta = \alpha^{-1}$$

Puesto que Γ_i contiene los f_k, f_l como f_k o bien $f_k f_l$ y según las condiciones (1,2,3) resulta:

Si f_k está presente, así como $f_k f_l$, entonces

$$\Gamma_k = f_k \text{ y } \Gamma_{k'} = f_k f_l, \text{ de donde } f_l = \Gamma_{k'} / \Gamma_k$$

De forma semejante se procede para otras combinaciones de productos de las funciones de origen.

Como ilustración, consideremos una retina de cinco fotoreceptores. Una forma sistemática de conseguir un conjunto Γ independiente de orden dos es el siguiente:

$$\Gamma_1 = l_1; \Gamma_2 = l_1 l_2; \Gamma_3 = l_2 l_3; \Gamma_4 = l_3 l_4; \Gamma_5 = l_4 l_5$$

luego, la transformación no lineal sería

$$F_i = \alpha_{i1} l_1 + \alpha_{i2} l_1 l_2 + \alpha_{i3} l_2 l_3 + \alpha_{i4} l_3 l_4 + \alpha_{i5} l_4 l_5$$

$$(i = 1, \dots, 5)$$

los α_{ij} pueden ser los coeficientes de cualquier transformada lineal con inverso ($|\alpha_{ij}| \neq 0$), por ejemplo, la de Fourier, Fourier-Bessel o los Momentos.

La transformación inversa es:

$$\Gamma_i = \sum_1^{N^2} \beta_{ij} F_j \quad \text{con } \beta = \alpha^{-1}$$

y

$$f_1 = \Gamma_1; f_2 = \Gamma_2 / \Gamma_1; \dots; f_5 = \Gamma_2 / \Gamma_4$$

La extensión a dos dimensiones no es única, dependiendo de la regla de definición. Podríamos, por ejemplo, realizar una expansión por etapas, primero por filas y posteriormente por columnas tal que dada una imagen de $N \times N$ la expansión sería:

Por Filas: Si $j = 1$ entonces $I'_{ij} = I_{ij}$

sino entonces $I'_{ij} = I_{ij} * I'_{i,j-1}$

Por columnas: Si $i = 1$ entonces $I''_{ij} = I'_{ij}$

sino $I''_{ij} = I'_{ij} * I'_{i-1,j}$

de esta forma nos aparecerían los productos cuádruples

$$I_{ij} * I_{i,j-1} * I_{i-1,j-1} * I_{i-1,j}$$

Esta transformación al ser multiplicativa tiene la ventaja de ser simétrica direccionalmente.

Otra posible definición sería tomar como referencia la diagonal y generalizar el ejemplo unidimensional del siguiente modo:

$$\begin{array}{ll}
\text{Si } i=j \text{ entonces} & I'_{i,j} = I_{i,j} \\
\text{Si } i < j & " & I'_{i,j} = I_{i,j-1} * I_{i,j} \\
\text{si } i > j & " & I'_{i,j} = I_{i-1,j} * I_{i,j}
\end{array}$$

Observamos que en este caso solo aparecen los dobles productos.

Otros ejemplos de convoluciones generalizadas analíticas serían las siguientes transformaciones no lineales

$$I'_{i,j} = \frac{I_{i,j}}{1 + I_p}$$

$$I'_{i,j} = I_{i,j} * e^{-I_p}$$

donde I_p sería la contribución periférica de cada punto de la imagen.

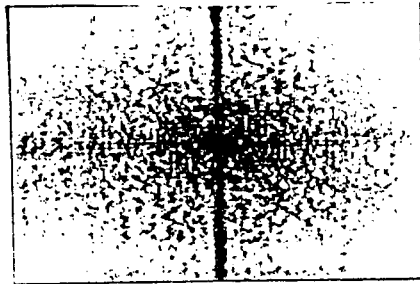
En la figura 2.4 se presenta un ejemplo de una transformación no lineal donde el sistema lineal utilizado ha sido una transformación de Fourier. En la figura 2.4 a) observamos la imagen original; en b) imagen la expandida y en c) la transformada de Fourier de la imagen expandida. Por otra parte, comprobamos experimentalmente la completitud de la transformación.



a)



b)



c)

Figura 2.4

2.4 TRANSFORMACIONES NO LINEALES EN PREPROCESO DE IMAGENES.-

Un objetivo central en el preproceso de imágenes es la normalización de las intensidades (iluminación), esto es, conseguir cuasi-invariancias frente a modificaciones generales de la condiciones de iluminación. En general, una transformación que es adecuada puede considerarse como una traslación más una especie de homotecia de las intensidades, con un factor de homotecia que puede no ser constante para toda la imagen y que, desde luego, es función de la imagen, y por consiguiente, resulta ser una operación no lineal sobre cada pixel.

Un procedimiento de conseguir estas cuasi-invarianzas, según Muñoz [35], es a través de un filtro lineal pasa alta, no necesariamente invariante en el espacio, seguido de un procedimiento de escalado para estandarizar el rango de variación de las intensidades. El proceso de filtro pasa alta puede expresarse como un caso particular de la transformación :

$$I'(x,y) = \int_C I(x,y) W(x',y',x,y) dx'dy'$$

donde $I(x,y)$ es la imagen incidente, $W(x',y',x,y)$ son los factores de peso del filtro espacial y C es la zona de actuación sobre el mosaico (campo receptivo), que dependerá a su vez de las coordenadas x, y .

La invarianza conseguida es relativa a desplazamientos en el rango de las intensidades por adición o sustracción de iluminación adicional dentro del dominio C .

Este filtrado, como todos los pasa alta, es muy sensible al ruido, y existen varias alternativas para disminuir tal sensibilidad, que consisten, en esencia, en utilizar filtros pasa banda desplazados, más o menos, hacia abajo en el dominio de frecuencias.

De todas formas, al ser la anterior una transformación lineal, no presenta invarianzas frente a cambios homotéticos en el dominio de las intensidades.

El conseguir este tipo de invarianza es, como indicamos antes, una operación no-lineal, que obliga a la selección de valores de referencia en la imagen de entrada. Esta selección de referencia puede ser realizada de forma muy diversa.

Una transformación sencilla que proporcione las invarianzas traslacionales y homotéticas se consigue por medio de la expresión local :

$$I''(x,y) = K \frac{I(x,y) - \text{Min}[I(x,y)]}{\text{Max}[I(x,y)] - \text{Min}[I(x,y)]}$$

De hecho, el requisito de las invarianzas no impone la condición de carácter local de la transformación. Por otro lado, el utilizar valores máximos y mínimos en toda la imagen va en contra de una estructura computacional paralela, ya que los procesos paralelos han de **pararse** hasta que un parámetro global (el histograma) se calcule.

Como hemos indicado anteriormente, las invarianzas requeridas suponen la acción de un filtro no lineal. La naturaleza de la no linealidad viene dada por la homotecia en el dominio de las intensidades. En general, ésta es una propiedad

que tiene como expresión el cociente de dos transformaciones lineales, cuyos núcleos cumplen la propiedad

$$\int_C W(x,y,x',y') dx'dy' = 0$$

donde C es, como antes, el dominio para cada (x,y), ésto es, el campo receptivo. En el lenguaje de las estructuras computacionales paralelas, la expresión anterior requiere que el campo receptivo C_j de cada unidad computacional tenga partes excitadoras e inhibidoras que se cancelen.

Lo anterior permitió establecer y demostrar con gran facilidad a Muñoz [35], el siguiente teorema :

Teorema. Si T₁ y T₂ son dos transformaciones lineales, caracterizadas por los núcleos W₁ y W₂, tales que

$$\int_C W_1(x,y;x',y') dx'dy' = 0 \quad y \quad \int_C W_2(x,y;x',y') dx'dy' = 0$$

la transformación no lineal T, definida por

$$T(f(x,y)) = \frac{T_1(f(x,y))}{T_2(f(x,y))} = f'(x,y)$$

es tal que $f'(x,y)$ es invariante frente a traslaciones y homotecias en intensidades.

Este teorema tiene especial interés y repercusión en el proceso de imágenes para su estandarización y abre un campo experimental apreciable, porque amplía considerablemente el abanico de los métodos para esos fines, con la inclusión de las propiedades que son deseables de ciertas transformaciones lineales.

La transformación no lineal anterior tiene la dificultad de presentar singularidades cuando se han realizado transformaciones previas según los núcleos con las propiedades de cancelación expresadas. Ello se solventa imponiendo sólo la condición de invarianza homotética y posteriormente la de desplazamientos.

Este teorema, aunque demostrado teóricamente por Muñoz (1987), no se desarrolló de forma práctica, tarea que hemos llevado a cabo en el presente trabajo.

Un ejemplo de aplicación, de entre los muchos posibles, se muestra en la figura 2.5.

En este caso hemos utilizado un operador diferencial de segundo orden, que como ya dijimos corresponde a un tipo particular de convoluciones generalizadas.

En la figura 2.5 a) podemos observar una imagen directamente adquirida a través de una cámara de video B/N, cuyos niveles de grises están comprendidos entre $a = 0$ y $b = 80$. En la figura b) puede observarse la imagen resultante de procesar la anterior, proceso consistente en llevar a cabo una homotecia de valor 2 mas una traslación de valor 50 sobre la imagen anterior.

Por último en las figuras c) y d) puede verse el resultado de preprocesar las imágenes anteriores según la transformación no lineal T expuesta anteriormente, que también corresponde a una convolución generalizada.



a)



b)



c)



d)

Figura 2.5

2.5 INCIDENCIA EN TEORIA DE REDES NEURONALES.

El efecto no lineal clave en el teorema anterior es un proceso divisivo. En términos de la computación en redes de elementos neuronales, el proceso divisivo es un proceso de inhibición presináptica y fué reconocido como tal hace más de veinte años [36]. Ello sugiere de inmediato que tal mecanismo de inhibición puede ser usado en una estructura computacional en paralelo, para conseguir representaciones que sean invariantes ante las transformaciones citadas. El logro de representaciones invariantes sigue siendo, a distintos niveles, una cuestión abierta en la teoría de la función cerebral, y por ello resulta de interés el desarrollar el modelo neuronal correspondiente a la transformación T antes citada.

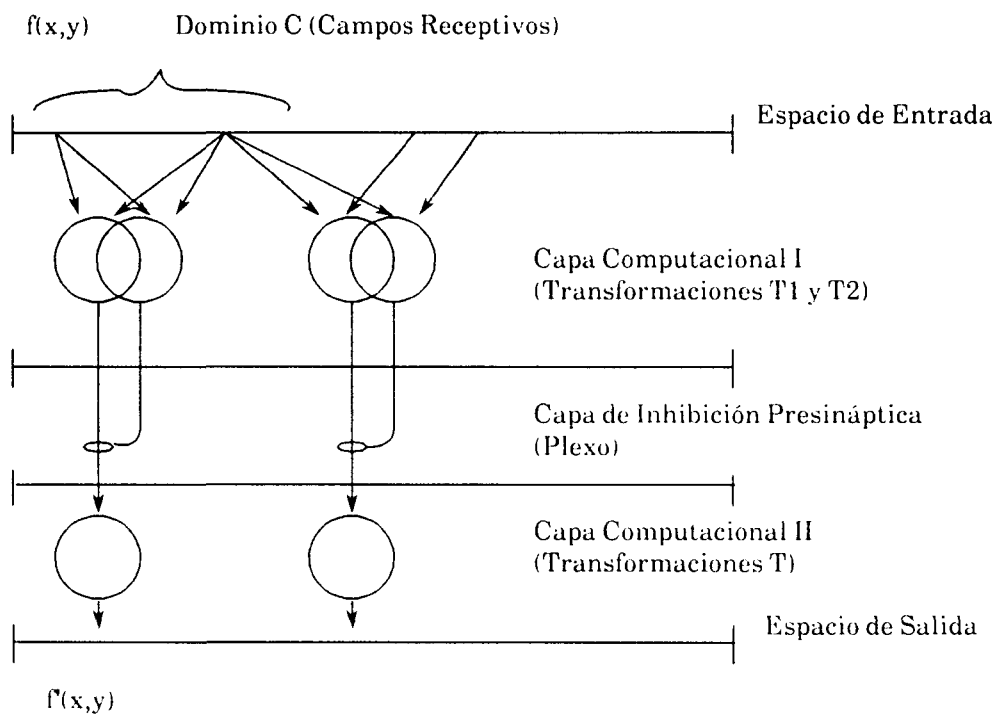


Figura 2.6

Como tal estructura computacional en paralelo, el proceso es representado en la figura 2.6. Los datos del espacio de entrada son tomados, de los campos receptivos, por dos tipos de unidades de computación en la capa computacional I, de tal forma que cada pareja de unidades de computación tienen un campo receptivo aproximadamente igual. La capa de inhibición presináptica tiene estructura de **plexo**, en el sentido neuronal. Las representaciones invariantes aparecen en la salida de la capa computacional II, que, a los efectos que nos ocupan, es una simple transmisión o a lo más, codificación, de los datos de la capa anterior.

Este mecanismo proporciona un esquema sencillo para explicar la adaptación neuronal, por ejemplo en retina. En este caso, el espacio de entrada es la imagen incidente sobre los receptores, o bien las salidas de éstos. La capa computacional I es la correspondiente a dos tipos de bipolares, que, en teoría, debía doblar en número a los fotorreceptores (tal como ocurre en la fovea). La capa de inhibición presináptica se localiza en la capa plexiforme interna. La capa computacional II es, a los efectos de adaptación neuronal indicados, la formada por las células ganglionares.

Esta estructura de computación retinal se representa en el diagrama de la figura 2.7.

Tal como acabamos de indicar, la fovea es la única región que en la retina de los vertebrados superiores cumple el hecho (requerido por este planteamiento teórico) de que el número de bipolares doble (al menos) al de fotorreceptores. Como consecuencia, se deduce de inmediato la siguiente predicción teórica, que habrá de ser contrastada experimentalmente : **la adaptación neuronal (no pupilar) a cambios a referencias y escalas de la intensidad incidente sólo tiene lugar**

de forma notoria en las regiones foveales, o en las parafoveales con estructura cuasi-foveal.

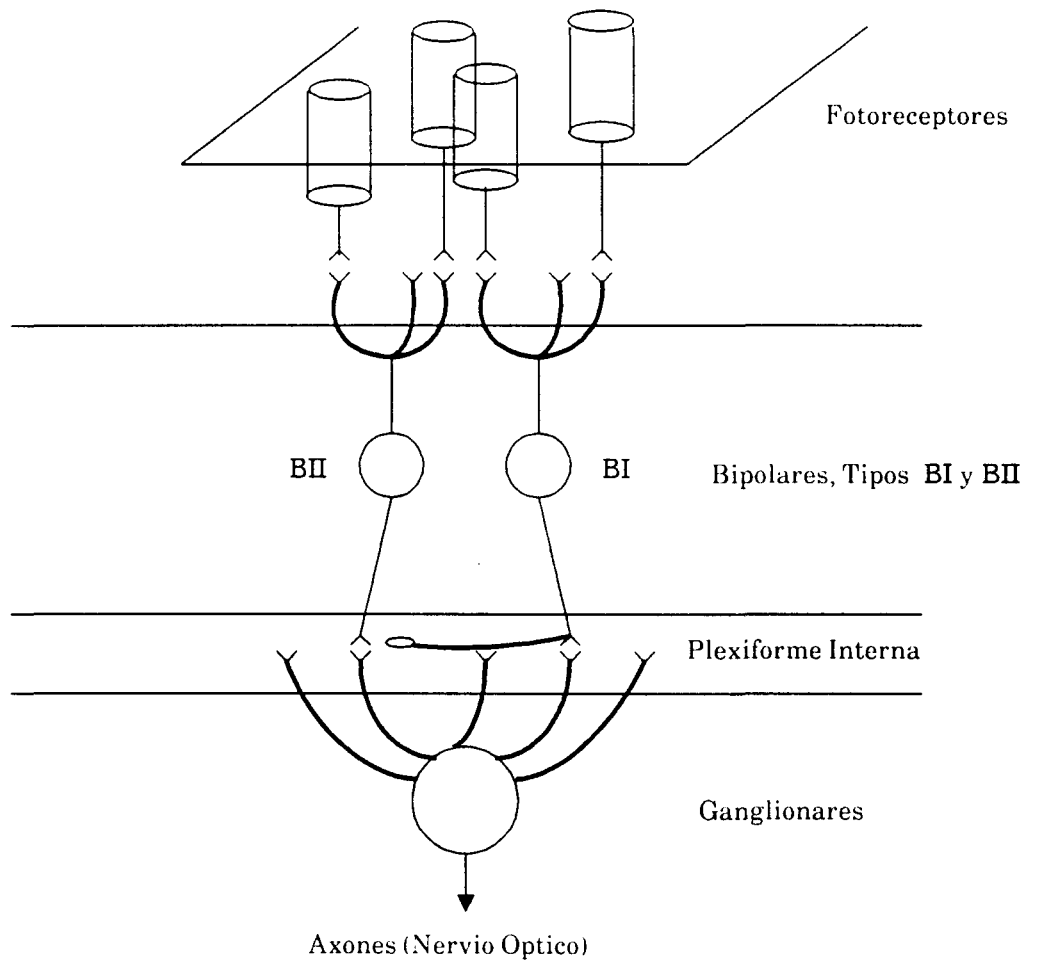


Figura 2.7

CAPITULO 3

CAMPOS DE DATOS Y PARTICIONES

3.1 SOBRE LAS TRANSFORMACIONES CAMPO RECEPTIVO-FUNCIONAL.

3.1.1 Transformaciones de Campo Receptivo variable.

El concepto de TCRV fué introducido por Candela, [30] en su Tesis Doctoral y desarrollado allí y en trabajos posteriores [37,38], partiendo de la definición de transformaciones de campo receptivo constante y en su negación posterior. Los argumentos proceden de la siguiente forma:

Dentro de la Teoría de Sistemas, una transformación corresponde a la acción de un sistema sobre el espacio de las entradas y quizá el de las salidas, para generar un nuevo "slide" del espacio de salidas.

En tal estructura, no es difícil definir y visualizar lo que se entiende por "campo receptivo".

Trabajaremos en una dimensión, que es extensible a varias dimensiones sin esfuerzo conceptual adicional importante.

El sistema bajo consideración transforma $f(x)$ en $F(x')$, donde x y x' son dominios que pueden coincidir, a través de una transformación arbitraria:

$$F(x') = T[f(x)]$$



Ahora bien, admitimos que existe un "dominio máximo" de definición de $f(x)$, en lo que respecta a la variable independiente x .

Por ejemplo, si $x = t$ (el tiempo), x va de $-\infty$ (ó 0) a t (instante actual), igual que x' , "toma nota" de todos los valores posibles de $f(t)$ antes de t para "decidir" su valor $F(t)$. Si la variable es espacial, el anterior "principio" de causalidad puede ser roto es decir x' puede ser tal que $x' > x$, tal como se ilustra en la figura 3.1.

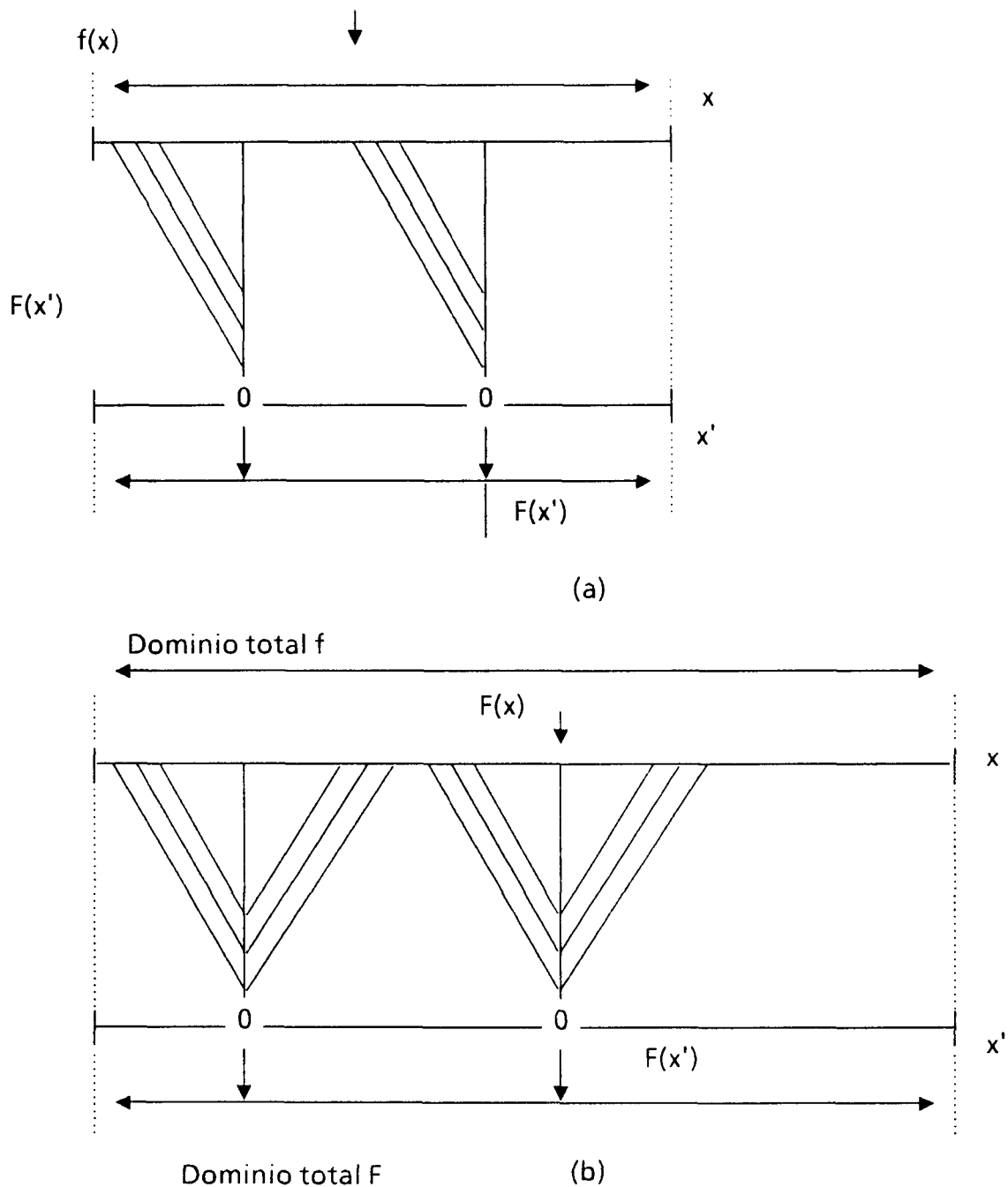


Figura 3.1 Ilustración de sistema causal (a) y de sistema incausal (b).

Primeramente, hablaremos de "dominios totales" de las variables x y x' (que pueden ir de $-\infty$ a $+\infty$, o de $-\infty$ a x_0 , o viceversa o de x_1 a x_2). Este se corresponde al dominio de x donde $f(x)$ (como familia) puede existir. Lo que esto quiere decir es claro en el caso de una retina con una imagen estática, pero también queda claro si x es el tiempo.

Para una "operación" $F(x')$, en la posición x' , su dominio es la "zona" de x donde se adquieren valores de $f(x)$ para tomar la decisión $F(x')$. Así, tenemos la siguiente condición (C_1), obvia como condición suficiente:

C_1 . Establecido el dominio total de $f(x)$, una transformación

$F(x')$ es TCRC si, para todo x' , el dominio de $F(x')$ es el dominio total de x . (Condición suficiente).

Por ejemplo, todo sistema causal lineal en el tiempo realiza una transformación TCRC, cosa que es muy fácil de demostrar.

Siguiendo con lo anterior, es fácil llegar a una condición más restrictiva para las TCRC. Para ello, se define el concepto de "tamaño" del campo receptivo de una transformación $F(x')$. Admitimos que x soporta una métrica de distancias tal que la distancia entre dos puntos x_1 y x_2 es el valor numérico $d(x_i, x_j)$. Así, pasamos a la condición más restrictiva de TCRC siguiente:

C_2 . Dada una distancia $d(x_i, x_j)$ medida en x , una transformación $F(x')$ es TCRC si $F(x')$ decide sobre el dominio comprendido por $(x_i < x < x_j)$ tal que $d(x_i, x_j) = cte$ para todo x' .

Así, la generalización de lo anterior resulta en cierto modo trivial. En efecto:

- Sea el dominio x y la definición de distancia $d(x_1, x_2)$ entre dos "puntos" de x . Sea D el dominio definido por los intervalos determinados por las parejas $(x_1 x_2, x_3 x_4, x_5 x_6, \dots x_l x_m)$, es decir $D = (d_{12}, d_{34}, \dots d_{lm})$ donde $D(d_{lm})$ es una ristra que contiene las distancias d_{lm} , para los valores predefinidos de l y m . D es constante si cada d_{lm} lo es, uno a uno.

Entonces, en general, llegamos a la condición C_3

C_3 . Una TCRC es una transformación $F(x')$ tal que, para todo x' , su dominio cumple la condición $D = \text{cte}$. D es el campo receptivo.

No es difícil ver que las condiciones anteriores son tales que $C_3 > C_2 > C_1$. Hemos procedido en ese orden por razones de claridad.

Luego una TCRV es una transformación donde D no es constante.

Siguiendo a Candela y como ya indicamos en el capítulo 2, al separar el campo receptivo de la función en sus raíces, se muestra que ambos tienen propiedades genuinas respecto de la complitud, que dependen de la propia naturaleza de ambos.

Sea un espacio de entrada $f(x)$ de N grados de libertad. Admitamos la existencia de un conjunto de funcionales $[F]$ tales que de ellos se selecciona un conjunto, $[A]$, "independiente", de coeficientes descriptores de las $f(x)$, tal que estos actúan en cada dominio o campo receptivo. Sea un conjunto, de cardinal B , de campos receptivos independientes.

Una descripción es completa si: $A * B = N$

Para $B=1$, $A=N$ tenemos las transformaciones globales puras clásicas, y para $B=N$, $A=1$ tenemos las transformaciones puras de campo receptivo (o de resolución progresiva) introducidas por Moreno-Diaz y Candela en 1987. [30]

3.1.2 Transformaciones Foveales

Una clase particular de TRP's, que están inspiradas en la estructura del "moving average", son las transformaciones denominadas "foveales", ya que para una retina, proporcionan la máxima resolución "directa" en la parte central, o "fovea".

La idea de TRP foveales parte del hecho de que en una partición completa global de una retina podemos jugar con los grados de libertad iniciales, para deducir, o imponer, campos receptivos o dominios independientes, a nuestra elección. La idea se ilustra imponiendo la existencia de todos los campos receptivos de una dimensión dada, d , sobre la retina, de forma que sean independientes (como en el caso del "moving average"). Para una dimensión dada, d , existen

$$D = N - d + 1$$

dominios linealmente independientes. Estos D , aumentados con un número $d-1$ de dominios linealmente independientes de dimensión 1 arbitrarios (es decir, dominios correspondientes a la transformación identidad), forman una TRP.

Puesto que los dominios unitarios pueden ser cualesquiera los podemos elegir de forma que estén "centrados", es decir que su unión proporcione un dominio de

dimensión única situado en el centro de la retina. Este "dominio unión" puede ser considerado como una subretina, a la que es aplicable de nuevo el procedimiento anterior, o bien cualquier otra TRP.

Existen procedimientos sistemáticos para conseguir las TRP foveales como el siguiente:

Definimos el número de subconjuntos de dominios de dimensiones diferentes que se van a admitir. A cada subconjunto se le denomina "capa de proyección". Sea K el número de tales capas, tales que el cardinal de cada capa es constante.

Se tiene que dicho cardinal es

Parte entera de $[N / K] = [N / K]_E$

Para la primera capa se tiene una dimensión

$$d_1 = N - [N / K]_E + 1$$

e iterativamente, para la capa i

$$d_i = d_{i-1} - [N / K]_E$$

En general si N/K no es entero, la última capa proporciona una subretina de dimensión:

$$N - (k - 1) [N / K]_E$$

Como ilustración, consideremos una retina de $N=16$, con 4 capas de proyección. Se tiene que el cardinal de cada capa es:

$$[16/4]_E = 4$$

Es decir, cada subconjunto de dominios está formado por 4 elementos.

La dimension de los elementos de las capas son:

$$d_1 = 16 - 4 + 1 = 13$$

$$d_2 = 13 - 4 = 9$$

$$d_3 = 9 - 4 = 5$$

$$d_4 = 5 - 4 = 1$$

y las de las subretinas:

$$SUB_1 \quad 13 - 1 = 12$$

$$SUB_2 \quad 9 - 1 = 8$$

En la figura 3.2 se representa un esquema de este ejemplo.

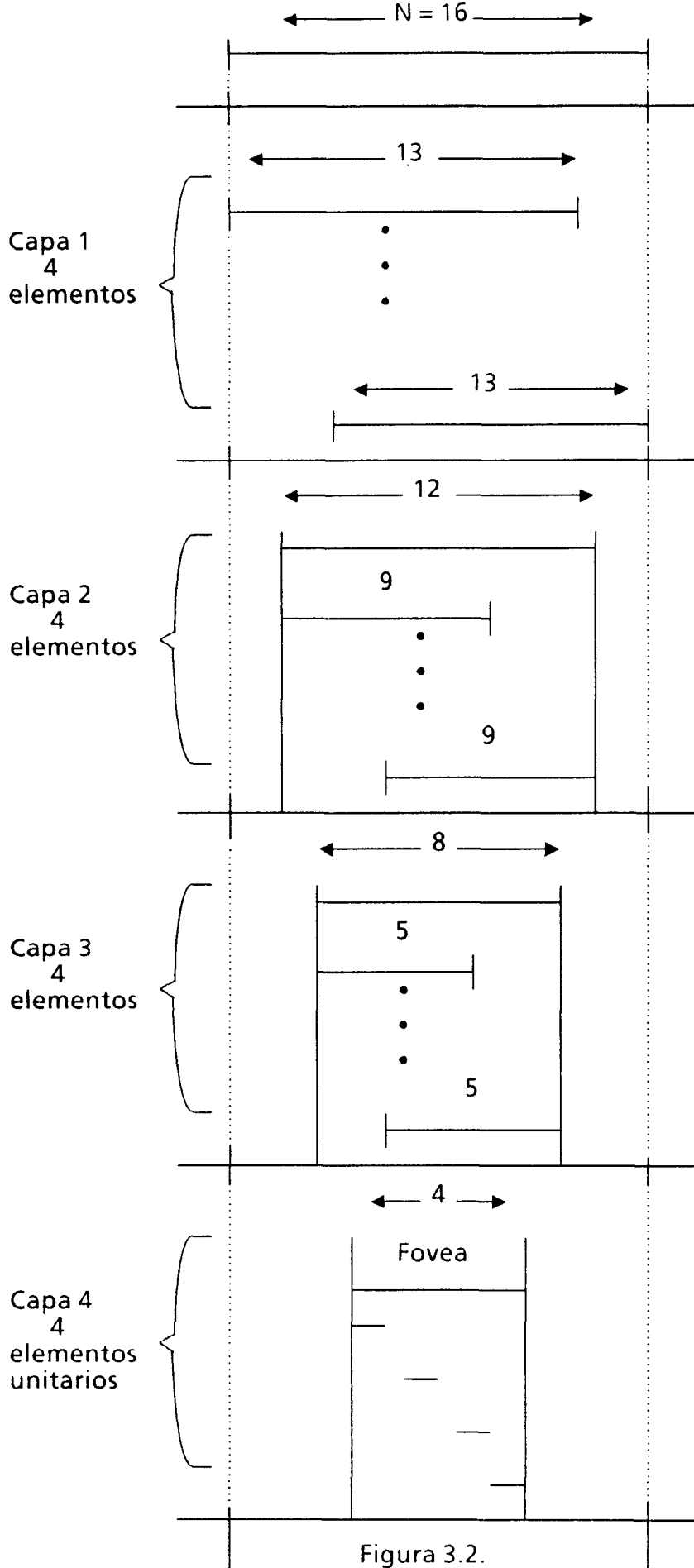


Figura 3.2.

Y en las siguientes figuras 3.3 se muestran los resultados de aplicar una transformacion foveal a una imagen de 128*128 variando el número de capas.

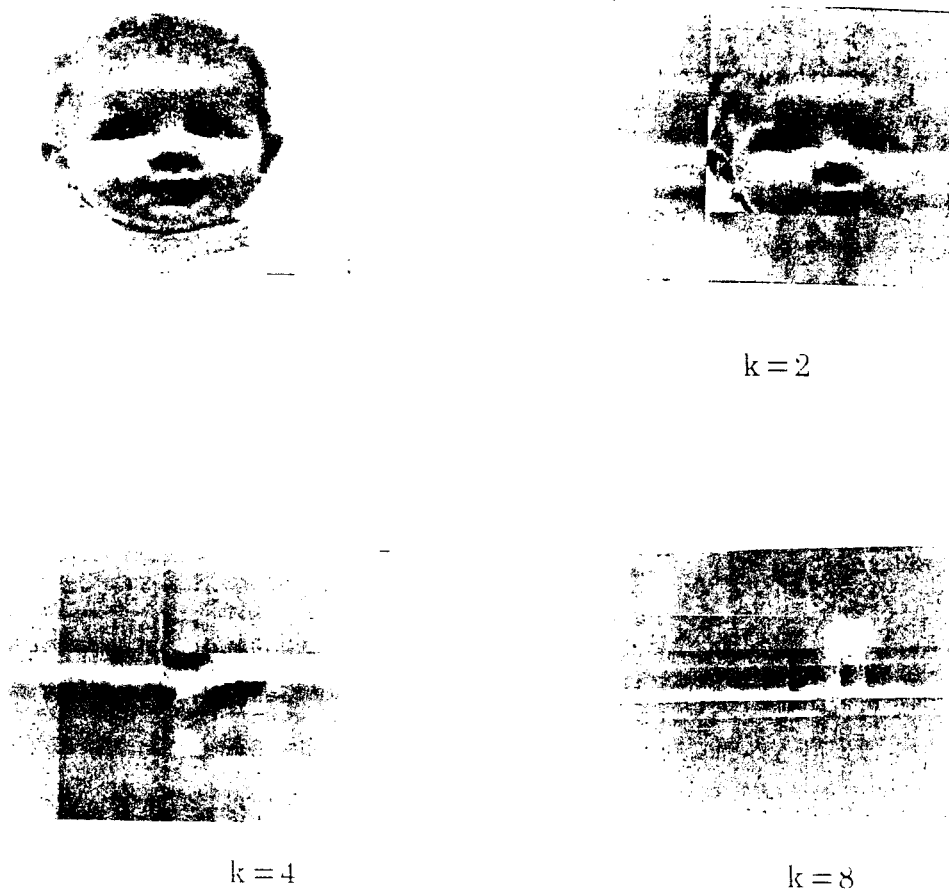


Figura 3.3

3.2. CAMPOS DE DATOS. TRANSFORMACIONES Y REPRESENTACIONES COMPLETAS.

El procedimiento definitorio de las TCRV, introducidos anteriormente, aunque formalmente correcto, adolece del inconveniente de dejar con cotas demasiado extensas el conjunto de transformaciones que son considerables como TCRV. Es algo similar a lo que ocurriría si, para definir una transformación no lineal, definiéramos "ad principium", las transformaciones lineales y se procediese a la negación posterior. Aquí se ve que resulta apropiado acotar el campo de formulación, por ejemplo, restringiéndonos a las transformaciones formulables en analítica. En este caso, se procedería definiendo primero como tales, las transformaciones analíticas y, con posterioridad, la subclase de transformaciones lineales.

Si bien estamos de acuerdo en la importancia de separar el campo receptivo y la función en proceso visual y en general en computación paralela, coincidiendo plenamente en ello con los argumentos de los citados autores, creemos que un enfoque alternativo permite converger en una amplia, pero manejable, clase de transformaciones de campos receptivos, de las que las TCRC son un caso particular. Para ello, en lugar de fijarnos solamente en las particiones independientes de los campos receptivos, centraremos nuestra atención en el concepto de descripción completa de un "campo de datos", tanto en el aspecto funcional como en el del campo receptivo.

El concepto de Campo de Datos se obtiene como generalización inmediata de la representación de imágenes en una, dos, o mas dimensiones. Trabajaremos en una dimensión, ya que los resultados son generalizables a mas dimensiones sin dificultad conceptual adicional.

Un Campo de Datos Unidimensional, de longitud N y resolución R es un conjunto ordenado de N lugares, i , ($i = 1, \dots, N$), tal que a cada lugar es asignable un número I_i (real, complejo o de otra índole) con resolución R . Los lugares, i , deben entenderse como "hojas en archivo" mientras que R debe entenderse como el máximo número de huecos binarios necesarios para alojar los posibles datos I_i .

Según ello, podemos hablar del número de grados de libertad "extendido" de un campo de datos, que, es en general $N \cdot R$. Este enfoque permite replantear el concepto de "complitud", que en formulaciones anteriores se centraban en N , número de lugares.

Así, dado un campo de datos:

$$D = (N, R)$$

Cuando los valores N están llenos con valores concretos para una situación, decimos que existe una realización del campo, Dr .

Una transformación de D es una regla, r , que permite pasar de D a D' .

$$D(N, R) \xrightarrow{r} D'(N', R') \quad \text{para toda } Dr.$$

Así, podemos centrarnos en el concepto de complitud.

DEFINICION 3.1

Una transformación r proporciona una representación completa si y solo si la transformación inversa r' , tal que:

$$D'(N',R') \xrightarrow{r'} D(N,R) \quad \text{existe.}$$

El planteamiento anterior, aunque aparentemente "naiv", tiene sin embargo consecuencias inmediatas en el uso óptimo de las memorias, cuando debe existir un compromiso de uso en direcciones y contenidos, aparte de las implicaciones que son esperables en el problema inicial de proceso de imágenes, como es la representación de transformaciones de histogramas.

Sin desarrollar las cuestiones de inter-relación entre dirección-contenido (o de memorias relacionales), nos centraremos en una subclase de transformaciones de Campos de Datos para los cuales:

$$R = R'$$

No es conveniente olvidar que estamos tratando la Visión Artificial dentro de la Visión por computadora, como herramienta. De hecho, en la práctica totalidad de los trabajos actuales se supone $R = R' =$ suficientemente grande para albergarlo que sea según la resolución buscada, lo cual puede no ser el caso- y posiblemente, debe no ser el caso, en Visión Natural.

En tal situación, sin embargo, si la complitud requiere como condición necesaria que

$$N * R = N' * R'$$

resultaría $N = N'$, aunque la suficiencia no se concluye.

Sea r_r la subclase de transformaciones tales que $R = R'$. Entonces:

DEFINICION 3.2

Una transformación r_r tal que:

$$D(N,R) \xrightarrow{r_r} D'(N',R)$$

es completa si existe r'_r tal que:

$$D'(N',R) \xrightarrow{r'_r} D(N,R) \quad \text{existe.}$$

Por lo tanto, y siguiendo con la analogía inspiradora de la relación entre transformaciones analíticas y analíticas-lineales, introduciremos el concepto de transformaciones de lugares de resolución constante, es decir, transformaciones de la forma de r_r , según la definición 3.2.

3.2.1 Una Reconsideración de la Capacidad de los Lugares de Memoria

Volvamos por el momento al caso:

$$D(N,R) \xrightarrow{\mathbf{r}} D'(N',R')$$

con la condición de complitud, $N \cdot R = N' \cdot R'$.

Como ilustración del método inductivo que se usa aquí, proponemos el siguiente problema elemental: Dada una imagen D de N lugares y una resolución $1 (0,1)$, ¿existe una representación completa D' , de $N = 1$ [lugar(0,1)] y resolución N ? La respuesta es también elemental: No, al menos que exista un proceso de codificación, con reglas codificadoras bien establecidas. (Por regla codificadora bien establecida se entiende una regla común, al dato inicial y a su transformado.

Por consiguiente, una posible tentación, que es el asignar a los lugares i , una capacidad resolutive mayor tal que si $R' \gg R$ entonces $N \ll N'$, solo sería posible a través de una estructura adicional de asignación de significados, lo que supone un cierto salto de nivel. Y viceversa, si en un proceso (natural o artificial) encontramos una situación en la que $R' \gg R$ y, como consecuencia $N \ll N'$, debe existir una estructura adicional que se encarga del establecimiento del código necesario, a un nivel superior.

Estas consideraciones tienen una incidencia conceptual interesante en las líneas actuales de investigación de codificación retinal, en los llamados "múltiples

significados" de las salidas de proceso visual en la retina de los vertebrados. Así, parece que las células retinales "comprimen" el número de "lugares" necesario a expensas de un aumento de resolución en tectum (rana) o cortical (gato), con la ayuda de un mecanismo codificador - descifrador de mayor nivel existente en los dichos tectum o corteza.

3.2.2 Los Casos Usuales en Visión Artificial ($R = R'$)

Si se plantea la cuestión de complitud de una representación, tal que de:

$$D(N,R) \xrightarrow{r_r} D'(N',R)$$

se pueda recuperar $D(N,R)$ por r'_r , se llega, obviamente, a que una transformación completa es un proceso de rellenado de $N' = N$ lugares (ordenados, cada uno a su manera), partiendo de una realización D_r a otra D'_r y tal que desde D'_r se puede recuperar D_r , para todo r .

A nuestro entender, éste es un problema de transformaciones reversibles de lugares, según la nomenclatura de la Visión Artificial). Según esto, el problema de complitud debe plantearse, en general, de la siguiente forma conceptual:

Dado un Campo de Datos D , accesible según el n° de orden $i (i = 1, \dots, N)$, ¿Se puede construir otro campo de datos D' , accesible por el n° de orden $j (j = 1, \dots, N)$, tal que de D' se pueda recuperar D ? ¿Cuales son las condiciones necesarias y suficientes?.

La visión artificial impone otras condiciones más poderosas que, usualmente, vienen dictadas por razones de aplicabilidad o de reducción, en definitiva, de la dimensión del campo inicial.

Así, se tiene el factor adicional proporcionado por el conocimiento condicionante exterior, que en su forma más sencilla supone el siguiente planteamiento:

Dado un campo de datos D , según el n° de orden i ($i=1\dots N$), ¿es posible construir otro campo D' de j ($j=1\dots M$) (donde $M \ll N$) tal que una tarea visual se complete con sólo consultas al campo D' ?

La típica solución de la visión artificial consiste en ordenar adecuadamente los D'_j , tal que, siguiendo los condicionantes exteriores, el número de consultas a D'_j sea nítidamente inferior a los que se requerirían si se consultase a D_i , para los casos en cuestión, con una prueba de validez, que suele ser, en definitiva, heurística.

Según lo anteriormente expuesto, se replantean dos problemas de interés teórico-práctico:

a) El relativo a las transformaciones reversibles de campos de datos, accediendo por uno o varios números de índice.

b) El de selección del conjunto de índices óptimo (mínimo en coste computacional) a través del cual acceder al campo de datos correspondiente para resolver una tarea visual.

3.3 EXPRESION DE LAS TRANSFORMACIONES COMPLETAS. CAMPO RECEPTIVO FUNCIONAL.-

Consideremos un campo de datos $D(N)$ de resolución R y consideremos un conjunto de L particiones independientes de las direcciones i ($i=1,\dots,N$). Este conjunto se puede representar por una matriz binaria, P_{ik} , de N filas y L columnas tal que el elemento ik ($k=1\dots L$) es uno si la dirección i está en la partición K y cero en otro caso. Los vectores verticales P_{ik} (para cada k) son independientes en el sentido de que ninguno de ellos puede obtenerse por unión de los otros. Por ejemplo, la matriz P_{ik} correspondiente a una transformación de resolución progresiva, para $N=L=8$, es:

$$\begin{array}{c}
 1 \\
 2 \\
 \cdot \\
 \cdot \\
 \cdot \\
 \cdot \\
 \cdot \\
 N
 \end{array}
 \left[\begin{array}{cccccccc}
 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\
 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0
 \end{array} \right]$$

Ec. 3.3.1

Una partición para $L=4$, es por ejemplo, representable por la matriz.

$$\begin{array}{c}
 1 \\
 . \\
 . \\
 . \\
 . \\
 . \\
 . \\
 N
 \end{array}
 \left[\begin{array}{ccccc}
 1 & 1 & 1 & 0 & \\
 1 & 1 & 1 & 0 & \\
 1 & 1 & 0 & 0 & \\
 1 & 1 & 0 & 0 & \\
 1 & 0 & 0 & 1 & \\
 1 & 0 & 0 & 1 & \\
 1 & 0 & 0 & 0 & \\
 1 & 0 & 0 & 0 &
 \end{array} \right]$$

1, . . . , L

Ec. 3.3.2

Otra partición - que usaremos profusamente más adelante - es la correspondiente a una "capa" de una transformación foveal. Por ejemplo, para campos receptivos de cuatro lugares, se obtendría.

$$\begin{array}{c}
 1 \\
 . \\
 . \\
 . \\
 . \\
 . \\
 . \\
 N = 8 \\
 . \\
 . \\
 N
 \end{array}
 \left[\begin{array}{ccccc}
 1 & 0 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 0 \\
 1 & 1 & 1 & 0 & 0 \\
 1 & 1 & 1 & 1 & 0 \\
 0 & 1 & 1 & 1 & 1 \\
 0 & 0 & 1 & 1 & 1 \\
 0 & 0 & 0 & 1 & 1 \\
 0 & 0 & 0 & 0 & 1
 \end{array} \right]$$

L = 5

Ec. 3.3.3

Nótese que estas particiones no son, a priori, completas por ser en general $L < N$, aunque sean independientes.

Consideremos una subclase de particiones de las direcciones i , donde cada dirección se toma, al menos, una vez. La partición cubre entonces, todo el campo de datos - es decir, no hay ninguna dirección que sea "desconsiderada". Tal ocurre en las dos particiones anteriores. En general, consideraremos solamente particiones de este tipo, que llamaremos particiones P .

Se plantea, entonces, la siguiente cuestión. Dada una partición P de un campo de datos de dimensión N , ¿cuál es un conjunto de funcionales que actuando sobre las direcciones de cada partición genera una representación completa?

3.3.1 Transformaciones Homogéneas de Campo Receptivo

Por razones de sencillez, consideramos primeramente las particiones foveales de 1 capa, que llamaremos P_1 . Para una de estas particiones de campo d , existen d grados de libertad, mientras que se genera un número de ellas, $L = N - d + 1$. Es decir la matriz correspondiente será de $N \times (N - d + 1)$. Como cada partición tiene d grados de libertad, se precisarían, para cada partición un funcional que generase d coeficientes independientes, si los contenidos de las direcciones de cada partición fuesen a ser recuperados independientemente. Puesto que ese no es el caso, se requieren, en media $M = N/L$ coeficientes diferentes por cada partición. Así, en el ejemplo anterior, se requieren, en media ($N = 8$; $L = 5$), $M = 8/5$ coeficientes independientes en media, valor no entero que, lógicamente, ha de ser obviado, calculando, por ejemplo, dos coeficientes (los mismos) para tres de las particiones y 1 coeficiente para cada una de las dos particiones restantes. El cómputo por exceso nos da una redundancia, que en general, puede no ser fácil de eliminar.

Supongamos, por ejemplo, que se calculan (por exceso) dos coeficientes para cada una de las particiones. El total será de 10 coeficientes para describir un campo de datos de 8 direcciones. Para la representación completa no redundante habrá que eliminar dos coeficientes, que no son cualesquiera, sino que han de pertenecer a particiones diferentes.

Con estas indicaciones, estamos en condiciones de formular y demostrar el siguiente teorema.

TEOREMA 3.1

Sea un campo de datos de N direcciones. Sea una partición P_1 de L columnas tal que $M=N/L$ es entero. Entonces, el cómputo de M coeficientes de funcional, linealmente independientes y distintos de cero, en cada partición, proporciona una descripción completa del campo de datos.

En efecto, sean los M vectores de "factores de peso" del funcional representados en una matriz de M filas y d columnas F_{ij} ($i = 1, \dots, M; j = 1, \dots, d$).

$$F = \begin{bmatrix} F_{11} & F_{12} & \dots & F_{1d} \\ F_{21} & F_{22} & \dots & F_{2d} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ F_{m1} & F_{m2} & \dots & F_{md} \end{bmatrix}$$

Los vectores F_{ij} (para cada i) son independientes por construcción.

Consideremos la matriz P_{ki} de particiones:

$$\begin{bmatrix} P_{11} & P_{12} & \dots & P_{1l} \\ P_{21} & P_{22} & \dots & P_{2l} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ P_{n1} & P_{n2} & \dots & P_{nl} \end{bmatrix}$$

Construyamos la matriz, M , de transformaciones de $N \times N$, por la "aplicación" de cada vector del funcional a cada partición:

$$\begin{bmatrix} F_{11} & \dots & F_{1d} \\ F_{21} & \dots & F_{2d} \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ F_{m1} & \dots & F_{md} \end{bmatrix} * \begin{bmatrix} P_{11} & \dots & P_{1l} \\ P_{21} & \dots & P_{2l} \\ \dots & \dots & \dots \\ \dots & \dots & \dots \\ P_{n1} & \dots & P_{nl} \end{bmatrix} = \begin{matrix} \left[\begin{array}{cccc} \text{---} & \dots & \dots & \dots \\ \text{---} & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{array} \right] \left. \vphantom{\begin{matrix} M_1 \\ M_2 \\ \dots \\ M_k \end{matrix}} \right\} M_1 \\ \left[\begin{array}{cccc} \circ \circ \circ & \dots & \dots & \dots \\ \circ \circ \circ & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{array} \right] \left. \vphantom{\begin{matrix} M_2 \\ \dots \\ M_k \end{matrix}} \right\} M_2 \\ \dots \\ \left[\begin{array}{cccc} \circ \circ \circ & \dots & \dots & \dots \\ \circ \circ \circ & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \end{array} \right] \left. \vphantom{\begin{matrix} M_k \end{matrix}} \right\} M_k \end{matrix}$$

\xleftarrow{N}

Como se ve, se han generado "grupos" M_1, M_2, \dots, M_k de M vectores, que tienen todos la misma extensión sobre el campo de datos. Dentro de cada grupo, los vectores son independientes entre sí, por construcción. Asimismo, los grupos entre sí son independientes, por corresponder a particiones independientes. Luego los N

vectores horizontales son independientes, y por consiguiente, la correspondiente transformación es completa. Es decir, la matriz M tiene inversa.

EJEMPLO 1

Como ilustración práctica, y para propósitos posteriores de reconocimiento, vamos a considerar el caso de una transformación de campo de datos homogénea de dimensión $N = 128$. Para la elección de los valores de L (número de particiones) y d (grados de libertad de cada una de ellas), la única restricción existente viene dada por la ecuación:

$$L = (N-d) / \delta + 1$$

siendo δ el desplazamiento interparticiones.

Esto quiere decir, que fijado el número de grados de libertad y el desplazamiento, obtenemos el número de particiones necesarias.

En este ejemplo, hemos elegido los valores $d = 32$ y $\delta = 16$, ambos potencias de 2, lo cual simplifica apreciablemente el costo computacional para realizar la transformación. Las 7 particiones que se generan son suficientes para cubrir todo el campo de datos y por construcción son linealmente independientes. Según el teorema anterior, la complitud requiere un número de funcionales por partición de $M = N/L$ en media. En este caso al no ser entero, lo obviamos de la siguiente manera: elegimos 18 funcionales para las particiones 1,2,3,5,6,7 y 20 funcionales para la partición 4. En la figura se presenta un esquema de estas particiones:

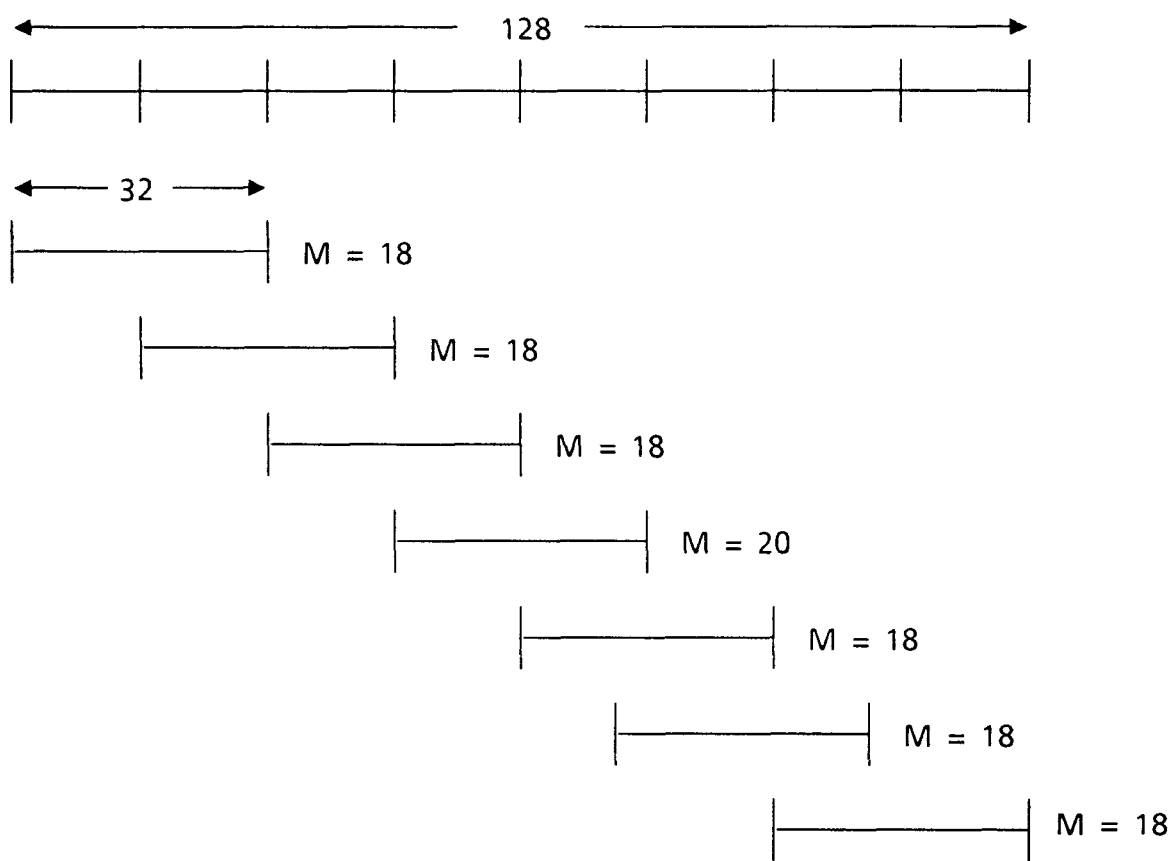


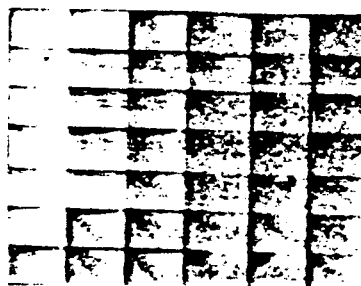
Figura 3.4

los funcionales elegidos en esta ilustración han sido los primeros coeficientes de la Transformada de Fourier.

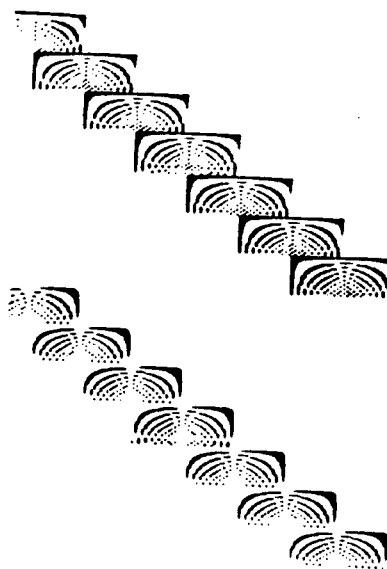
En la figura 3.5 a) se muestra la imagen original, en b) su transformada y c) el núcleo de la transformación.



a)



b)



c)

Figura 3.5

3.3.2 Transformaciones no Homogéneas de Campo Receptivo

Estamos ahora en condiciones de considerar particiones no homogéneas del campo de datos, es decir, aquellas en que las dimensiones de cada subpartición (columnas) no es la misma. Tal es el caso de la partición de la ecuación 3.2.2. Puesto que los grados de libertad son 8, resulta que, en media, se han de computar $8/4 = 2$ funcionales por columna. En este caso concreto se ve que ello llevaría a una representación completa, si la elección de los funcionales es apropiada. Es decir, la independencia de los funcionales ha de garantizarse cuando se aplican a cada columna, no bastando que lo sean para una columna.

EJEMPLO 2.

Dos funcionales independientes para la primera columna serían unos de tipo Haar:

$$F1 \longrightarrow 1/8 (1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1)$$

$$F2 \longrightarrow 1/8 (1 \ 1 \ 1 \ 1 \ -1 \ -1 \ -1 \ -1)$$

Se ve de inmediato que al aplicarlos a la segunda columna, los números que se obtienen no solo no son independientes, sino que son los mismos para los dos funcionales.

Sin embargo, pasando al otro caso límite, si elegimos dos funcionales que son independientes para la columna de menor dimensión, por ejemplo los de pesos:

$$(1,1); (+1,-1)$$

y se generan dos funcionales nuevos de dimensión el de la columna mas grande por yuxtaposición, es decir los,

$$(1, 1, 1, 1, 1, 1, 1, 1,)$$

$$(1, -1, +1, -1, +1, -1, +1, -1,)$$

es fácil ver que los números que se generan ya son independientes. La correspondiente matriz de transformación es:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \end{bmatrix}$$

y tal que todos los vectores fila son independientes. La transformada inversa es trivial: de las filas 5 y 6 se recuperarían los datos de las direcciones 1 y 2. De ellos y de las filas 3 y 4, los datos de las direcciones 3 y 4. De las filas 7 y 8 se recuperan las direcciones 5 y 6. Es decir, se han recuperado los datos de las direcciones 1,2,3,4,5 y 6. Entonces, de las dos primeras filas y de éstos se recuperan los datos

de las direcciones 7 y 8. Esta transformación es semejante a la de resolución progresiva detectora de contrastes introducida por Candela [30].

En las figuras 3.6 se muestra una aplicación de este ejemplo a un caso de una imagen bidimensional. La figura 3.6 a) corresponde a la imagen original y la b) corresponde a la imagen transformada.

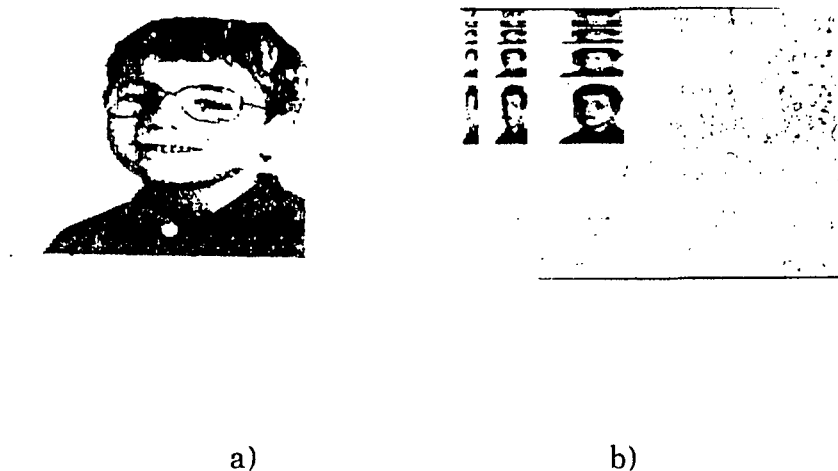


Figura 3.6

La solución encontrada en este ejemplo ha sido posible porque el número medio de funcionales a calcular por columna de la partición es igual (o menor) a la resolución de la columna de menor dimensión. Porque, evidentemente, no tiene sentido calcular tres descriptores, por ejemplo, de las columnas 3 y 4 de la partición, que contienen solo dos direcciones.

Lo anterior nos lleva, con claridad y sencillez, a formular y demostrar el siguiente teorema:

TEOREMA 3.2.

Sea un campo de datos de dimensión N . Sea una partición arbitraria P , de L columnas y tal que d_m es la dimensión de la columna de menor dimensión ($d_m > 1$) y además,

$$N/L \leq d_m$$

Entonces, un conjunto de N/L funcionales independientes de longitud d_m , yuxtapuestos, hasta la longitud de la columna de mayor dimensión proporciona una transformación completa.

La demostración del teorema se sigue con facilidad, según lo indicado en el ejemplo anterior. En efecto, formemos la matriz de transformación r :

$$r = \left[\begin{array}{c} \text{---} \\ \vdots \\ \text{---} \\ \vdots \\ \text{---} \\ \vdots \\ \text{---} \\ \vdots \\ \text{---} \\ \vdots \\ \text{---} \end{array} \right] \left. \begin{array}{l} \} C_1 \\ \vdots \\ \} C_2 \\ \vdots \\ \} C_1 \end{array} \right\}$$

C_1 corresponde al conjunto de columnas de menor dimensión, afectadas por los factores de peso correspondientes, que son independientes por construcción;

mientras que $C_2 \dots C_1$ corresponden a las restantes columnas, que, serán independientes por la condición adicional de independencia de las particiones y/o del funcional "extendido" que se ha usado.

Este teorema puede demostrarse de una forma quizás mas elegante (o purista), formulando y demostrando el siguiente teorema auxiliar:

Teorema Auxiliar 3.2

Sean dos dimensiones a y b tales que $a < b$. Sea $\{a\}$ un conjunto de n vectores linealmente independientes de dimensión a . Sea el conjunto $\{b\}$ de vectores de dimensión b construido por el siguiente proceso:

1) Se yuxtaponen las componentes de cada vector de $\{a\}$ hasta lograr vectores de dimensión b , en total n .

2) Se completan los vectores de $\{a\}$ hasta la dimensión b ($b-a$ lugares) con ceros.

Entonces, el conjunto $\{b\}$ forma un conjunto de $2n$ vectores (de dimensión b) que son linealmente independientes (obviamente, $2n < b$). La demostración es obvia. No se requiere que b/a sea entero.

Del teorema auxiliar 3.2 sigue de inmediato T.2 ya, que el proceso seguido en T.2 es el mismo.

[Ejemplo trivial: Tomemos el vector unitario en la dirección x ($n = 1, a = 1$) = 1. Consideremos el espacio tridimensional $(x, y, z), b = 3$. El conjunto de $2n = 2$ vectores formado por yuxtaposición $(1, 1, 1)$ y por el original completado con ceros $(1, 0, 0)$ es independiente.]

Algunos Casos Límites.

En el desarrollo de las transformaciones de campos de datos-funcional queda, por razones de complitud, considerar las situaciones en las que la dimensión de las columnas de dimensión mínima (d_m) sea tal que $d_m < N/L$. Pero en este caso se pierde la simetría que es una de las razones básicas para la separación del campo de datos y la función, en el sentido de que, dependiendo de la longitud de la partición, habría que disponer de más o menos vectores de factor de peso. En este caso, resulta cuestionable la utilidad de la separación realizada, frente a considerar la matriz de transformación directamente, con la condición de independencia entre sus vectores. Es decir, al necesitarse número $N' > N/L$ vectores de peso (que forman el funcional), el producto $N'*L$ supera el número inicial de grados de libertad, y para evitar redundancias, la aplicación de esos vectores a las columnas debe hacerse selectivamente.

Sin embargo, se pueden realizar algunas consideraciones que proporcionan métodos de generación de transformaciones completas. Según se ha indicado esta situación corresponde a particiones donde, en media, es preciso, para la complitud, extraer N/L descriptores independientes, pero ciertas de las columnas correspondientes no pueden proporcionarlos. Ello puede ser debido obviamente, a que:

- 1) Existen columnas donde es posible extraer un número de descriptores superior a la media.
- 2) Existen direcciones perdidas del campo de datos, con datos no recuperables por ninguna función del conjunto de procesadores.

El caso 1) corresponde a una partición tipo P. El caso 2) corresponde siempre a una situación de incompitud.

Nótese que el caso 2) tiene interés teórico en confiabilidad o seguridad de funcionamiento en cuanto corresponde al problema de los "escotomas" (desaparición local de datos), frente al problema de la "disfunción" (desaparición local de procesadores). Este segundo problema corresponde, en situación límite, al típico método usado en visión artificial de "cortar" el número de descriptores que, en una aplicación práctica, es suficiente, aunque no exista complitud.

Volviendo a la situación de particiones tipo P, donde cada dirección es considerada al menos una vez (caso 1 anterior), se pueden encontrar formulaciones para generar transformaciones completas. Consideremos un ejemplo, deducido de la partición de la ecuación 3.2.1. Sea la partición de $L=4$ representada por la matriz:

$$\begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

obtenida con las columnas 1,2,5,8 de la matriz de la ecuación 3.2.1. El número medio de descriptores por columna es $N/L = 8/4 = 2$. Pero como las columnas 3 y 4 no pueden proporcionarlos, las columnas 1 y 2 deberán proporcionar el número adicional de descriptores para mantener la media. Concretamente las columnas 3

y 4 pueden generar, a lo mas, un descriptor cada una, restando 6 a repartir entre las columnas 1 y 2. Puesto que las dimensiones en cada una de ellas es igual o superior a la media restante ($6/2=3$), para dichas columnas es aplicable el teorema T.3.2.

Desde el punto de vista de la teoría retinal actual, la solución anterior no parece, a priori, sostenible, puesto que equivaldría a admitir que las células ganglionares con campos receptivos amplios deben realizar una función de proceso y codificación muy superior a las que realizan las células con campo receptivo restringido. Pero, en definitiva, eso sería así, si y solo si se requiere la complitud para la totalidad de la retina, asunto cuestionable, e incluso rebatible a través del hecho de que es en la fovea (donde la precisión es alta) donde están las células de campo receptivo restringido.

Consideremos como ilustración, la construcción de una posible solución a la partición anterior. En efecto, podemos usar un único vector para obtener los descriptores correspondientes a las dos últimas columnas, por ejemplo el mas sencillo:

$$(1, 1, 1, 1, 1, 1, 1, 1)$$

Mientras que los seis descriptores de las dos primeras columnas los podemos calcular por el teorema 3.2, es decir, generamos tres vectores de dimensión 4, independientes, por ejemplo los:

$$\begin{pmatrix} 1 & -1 & 1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & -1 & -1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & -1 & 1 & -1 \end{pmatrix}$$

los que, por yuxtaposición, proporcionan:

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix}$$

Como se ve, el primero coincide, por lo que estos tres vectores, junto con las particiones, L, generan la transformación completa:

$$\mathbf{r} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Notemos que para obtener la transformación se han aplicado los tres vectores F a las dos primeras columnas de P, mientras que a las dos últimas se le aplica solo

el primero (en realidad se le podría haber aplicado uno cualquiera de los tres). Esto ilustra claramente la rotura de la simetría a la que antes nos referíamos.

En todas las transformaciones anteriores, se ha ignorado la ortogonalidad, y de hecho, la práctica totalidad no lo son, estando nuestro interés centrado precisamente en levantar la restricción de ortogonalidad. De hecho, el origen de los conceptos relativos a las transformaciones de campo receptivo variable, [30], fué precisamente utilizar una matriz ortogonal para generar otras no ortogonales.

Los métodos de generación de coeficientes de peso para los funcionales, utilizados en la sección anterior, sugieren otros procedimientos sistemáticos para generar transformadas no ortogonales. Nos referiremos, en lo que sigue, a un campo de datos donde no se realiza ninguna partición.

Consideremos el espacio de tres dimensiones y el cubo de lado 2 centrado en el origen:

Sea el vector $(1,1,1)$ el vértice del cubo. Una forma rápida de generar vectores no ortogonales pero independientes, usado en las secciones anteriores, es movernos en la dirección de las tres aristas confluyentes en $(1,1,1)$, para obtener los vectores:

$$(-1,1,1), (1,-1,1), (1,1,-1)$$

que forman una nueva base, pero no ortogonal. Su representación matricial es :

$$\begin{bmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix}$$

Una matriz de todos +1, excepto en la diagonal. Obviamente, los 1's pueden ser números cualesquiera (distintos de cero).

3.4 REDES NEURONALES Y REDES POR CAPAS

3.4.1 Planteamiento General.-

Una red neuronal en el sentido generalizado de McCulloch, Pitts, [39,40], es un conjunto N de módulos computacionales, donde, aparte de las M entradas exteriores a la red, cada módulo recibe entradas de si misma y de todas las demás en la red, computando una operación lógica, funcional o algorítmica, o combinación de ellas, arbitraria. La red se puede suponer síncrona sin pérdida notoria de generalidad. La estructura de tal red se muestra en la figura 3.7.

Como se ve, cada módulo computacional proporciona la salida (en el dominio adecuado), después de un retardo único (sincronizado), resultado de su operación sobre las N + M entradas en el instante anterior.

Como ya ha sido demostrado, [39], si la operación de cada módulo consiste en una función lógica arbitraria, la red duplica cualquier autómata arbitrario, determinista o probabilista según sea la naturaleza de la operación de unos o todas los módulos. La practicabilidad de este tipo de redes queda limitada por su generalidad. Al imponer restricciones se converge en redes que permiten duplicar

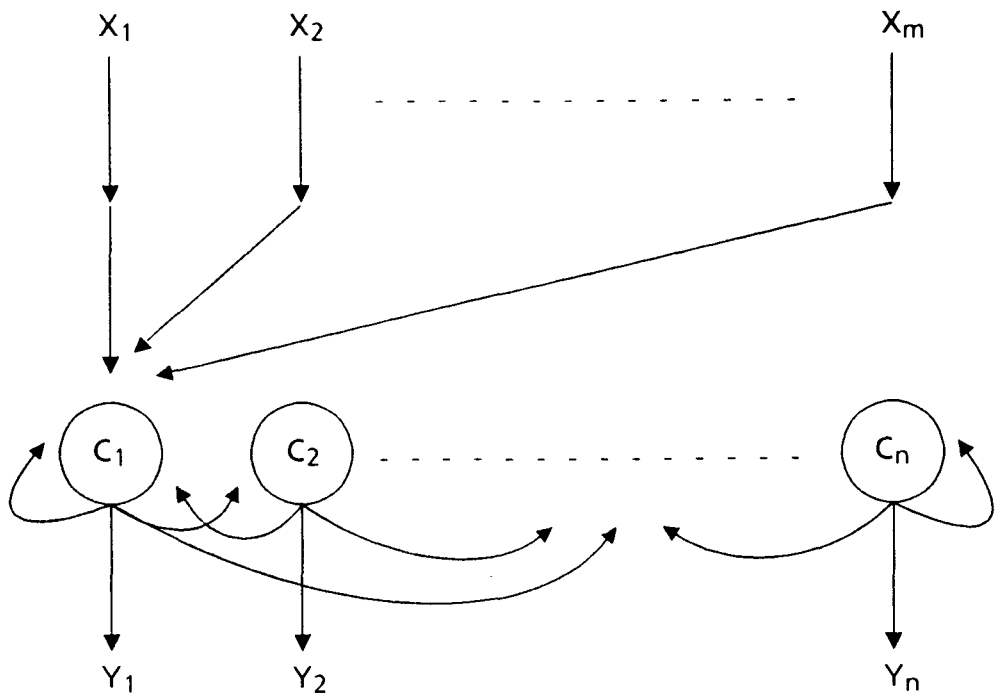


Figura 3.7

y aclarar la estructura y función de sistemas mas apropiados para teorizar sobre lo natural y lo artificial. Tal es el caso citado de limitar las posibles operaciones de cada módulo a las funciones lógicas arbitrarias, que permite una contrapartida modular estructural de todos los autómatas.

Como se infiere de inmediato, es posible imponer restricciones sobre la función, sobre la estructura o sobre ambas, en grado mas o menos fuerte.

3.4.2. Restricción Estructural Fuerte.-

Obviamente, al imponer restricciones estructurales sin hablar de la función, resultan consecuencias aplicables tanto a los dominios lógicos, funcionales, algorítmicos o cualquier otro, tal como hizo Candela [30] en lo que respecta a los campos receptivos o campos de datos.

Definición 1. Restricción Estructural Fuerte. (REF)

La restricción estructural fuerte (REF) consiste en que la red estructural generalizada NO tiene realimentación.

Consecuencia 1.

Una REF, cuando las operaciones de los módulos son funciones lógicas equivale a un conjunto de N circuitos combinacionales mas retardos. Por extensión, podemos hablar de REF's que son circuitos funcionales o circuitos algorítmicos. Así, se ve que una REF no tiene, por estructura, caracter secuencial, sino un estricto caracter paralelo, es decir, de los N elementos computacionales se obtienen, en el instante siguiente al de la aplicación de los datos del Campo de datos de entrada, el campo de datos de salida. Es el computador paralelo puro, tal como se ilustra en la Figura 3.8.

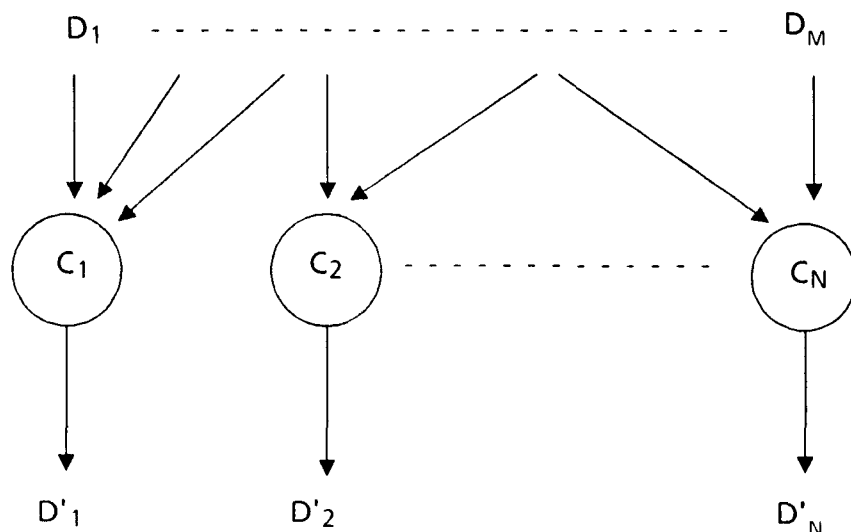


Figura 3.8

D_1, D_2, \dots, D_M es el campo de Datos de entrada; C_1, C_2, \dots, C_N son computadores actuando sobre todos o parte de los datos de entrada y D'_1, D'_2, \dots, D'_N es el campo de datos de salida.

3.4.3. Intercambio de espacio por tiempo.-

Es clásico [39] que, en redes, el espacio y el tiempo son intercambiables. Esta intercambiabilidad resulta que es desafortunadamente dependiente de la operación de los módulos y de la estructura, y de ello se pueden presentar ilustraciones desde las triviales a las muy sofisticadas. Trivial es, por ejemplo, pasar de una red de módulos computacionales que calculan la misma operación, sea convolución en un campo de datos con campos receptivos restringidos, a un único módulo computacional que proporciona las salidas de todos los módulos secuencialmente.

Menos trivial y cuando la operación no es del tipo de convolución -generalizada o no- porque es preciso, en momentos apropiados, cambiar el programa del elemento computacional, o añadir nuevos elementos computacionales capaces de nuevos programas.

Un caso práctico obvio de interés de intercambiabilidad de espacio y tiempo para REF's se presenta, cuando, en las operaciones de los módulos, en algunos de ellos existen suboperandos, de manera que algunos o todos de sus operandos están siendo computados por otros módulos. Bastaría pues, para esos módulos, "esperar" un instante posterior, y operar, no sobre el campo de datos, sino sobre los resultados de esos otros módulos. Como se ve, es, en esencia, una regla heurística que ha de ser buscada para cada REF particular. Como ejemplo trivial, consideremos la REF de tres módulos de la figura 3.9, donde C_1 computa R_1 sobre D_1 y D_2 , C_2 computa R_2 sobre D_3 y D_4 y C_3 computa una regla arbitraria R_3 sobre

$R_3(R_1(D_1 D_2), R_2(D_3 D_4))$. La REF estricta de tipo paralelo es la indicada en la figura 3.9.

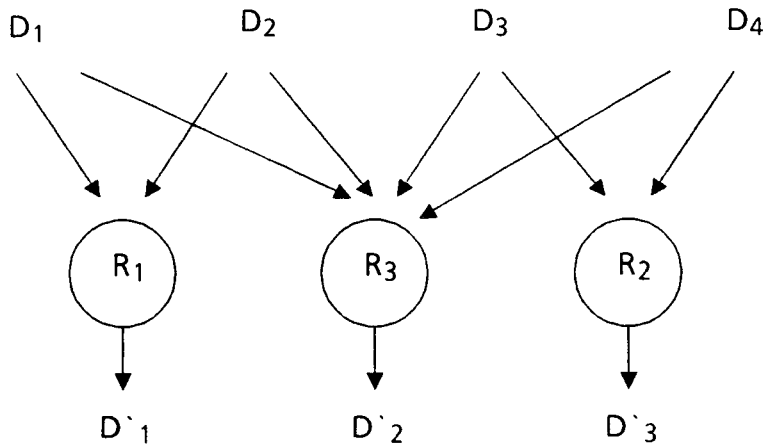


Figura 3.9

La intercambiabilidad obvia se ilustra en la figura 3.10

Si se usa una máquina secuencial para implementar la REF, evidentemente, la solución de la figura 3.10 es mas "barata", sobre todo si R_1 , R_2 y R_3' coinciden.

En la práctica, es preciso, dado un complejo REF-operaciones, analizar estructura y función para detectar posibles "trucos" de intercambiabilidad.

Ello nos lleva a las siguientes consecuencias, obvias, según lo anterior.

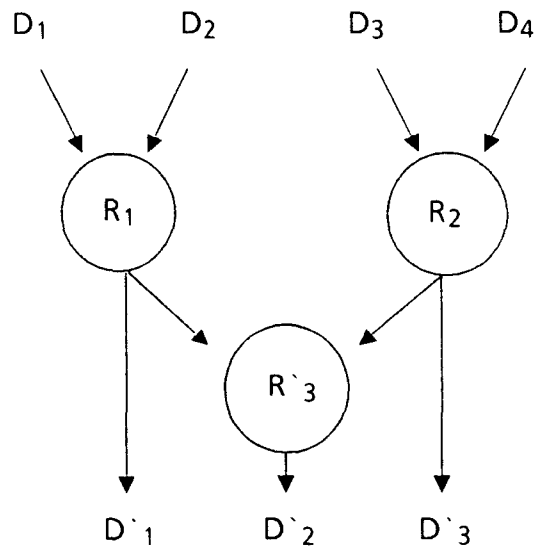


Figura 3.10

Consecuencia 2.-

2.1. Para la intercambiabilidad estricta entre puro espacio y espacio-tiempo, es necesario, aunque no suficiente, que la red tenga una estructura REF. (Restricción Estructural)

2.2. La conversión espacio-tiempo, en el problema analítico que estamos considerando, depende ultimamente de la asignación de operaciones a los módulos computacionales, y ha de ser investigada en cada caso. (Restricción Funcional). Si se trata de operaciones tipo convolución -"convolution like"-, la conversión es inmediata. Como se verá mas adelante, esto es la base de la existencia de "algoritmos rápidos" en proceso paralelo, que en definitiva, tratan de conseguir, con el menor costo computacional, operaciones equivalentes a las realizadas por redes tipo REF.

3.5 ALGORITMOS RAPIDOS DE COMPUTACION

Tal como ya hemos indicado, las transformaciones de campo de datos-funcional pueden representarse en el caso de una retina lineal de dimensión N por una matriz binaria, que en el caso de que la partición sea completa será de $N \times N$.

En el caso de que la matriz de transformación $[A]$ pueda ser factorizada en un producto de varias matrices:

$$[A] = [A_1][A_2] \dots [A_n]$$

estas diferentes matrices están caracterizadas por un número importante de elementos nulos, lo que reduce considerablemente el número de operaciones aritméticas. Decimos que este tipo de transformaciones es susceptible de algoritmo rápido.

Veamos un ejemplo ilustrativo: La figura representa la matriz de transformación del ejemplo 2 citado anteriormente, correspondeinete a las transformaciones no homogéneas de campo receptivo:

	11	12	13	14	15	16	17	18
1/8	1	1	1	1	1	1	1	1
1/4	1	1	1	1	0	0	0	0
1/2	1	1	0	0	0	0	0	0
1/2	0	0	0	0	1	1	0	0
H =	1	-1	1	-1	1	-1	1	-1
1/4	1	-1	1	-1	0	0	0	0
1/2	1	-1	0	0	0	0	0	0
1/2	0	0	0	0	1	-1	0	0

Esta matriz puede expresarse como producto de matrices mas sencillas con gran número de ceros:

$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$
A	B	C

$$H = A * B * C$$

Una característica atractiva de este tipo de transformaciones es que efectúan operaciones muy sencillas de tipo aritmético; podemos analizar los pasos del proceso de cálculo para un caso sencillo de 8 elementos:

Sean $I_1, I_2, I_3, I_4, I_5, I_6, I_7, I_8$ los valores de la función original. Se procederá a calcular la suma y la resta promediadas por parejas consecutivas, de forma que obtendremos unos nuevos valores:

$$\begin{aligned} A_1 &= (I_1 + I_2)/2 & A_5 &= (I_1 - I_2)/2 \\ A_2 &= (I_3 + I_4)/2 & A_6 &= (I_3 - I_4)/2 \\ A_3 &= (I_5 + I_6)/2 & A_7 &= (I_5 - I_6)/2 \\ A_4 &= (I_7 + I_8)/2 & A_8 &= (I_7 - I_8)/2 \end{aligned}$$

Con estos valores A_i se procederá a realizar el valor medio por parejas consecutivas, generando:

$$\begin{aligned} B_1 &= (A_1 + A_2)/2 & B_2 &= (A_3 + A_4)/2 \\ B_3 &= (A_5 + A_6)/2 & B_4 &= (A_7 + A_8)/2 \end{aligned}$$

Y así sucesivamente hasta llegar a un solo valor, en el que se finaliza. En este ejemplo obtendríamos los siguientes valores transformados:

$$\begin{aligned} F_1 &= (B_1 + B_2)/2 = (A_1 + A_2 + A_3 + A_4)/4 = (I_1 + I_2 + I_3 + I_4 + I_5 + I_6 + I_7 + I_8)/8 \\ F_2 &= B_1 = (A_1 + A_2)/2 = (I_1 + I_2 + I_3 + I_4)/4 \\ F_3 &= A_1 = (I_1 + I_2)/2 \\ F_4 &= A_3 = (I_5 + I_6)/2 \\ F_5 &= (B_3 + B_4)/2 = (A_5 + A_6 + A_7 + A_8)/4 = (I_1 - I_2 + I_3 - I_4 + I_5 - I_6 + I_7 - I_8)/8 \\ F_6 &= B_3 = (A_5 + A_6)/2 = (I_1 - I_2 + I_3 - I_4)/4 \\ F_7 &= A_5 = (I_1 - I_2)/2 \\ F_8 &= A_7 = (I_5 - I_6)/2 \end{aligned}$$

En forma algorítmica el cálculo de esta transformada se podría escribir de la siguiente forma:

Generar un vector A calculando sumas y restas ponderadas.

Mientras n° de elementos del vector ≥ 2 hacer

tomar los elementos impares del vector como coeficientes.

generar un nuevo vector promediando parejas de elementos

Fin mientras

Tomar los dos elementos del vector como coeficientes.

En la figura 3.11 se presenta un esquema de computación distribuida correspondiente a la transformación anterior.

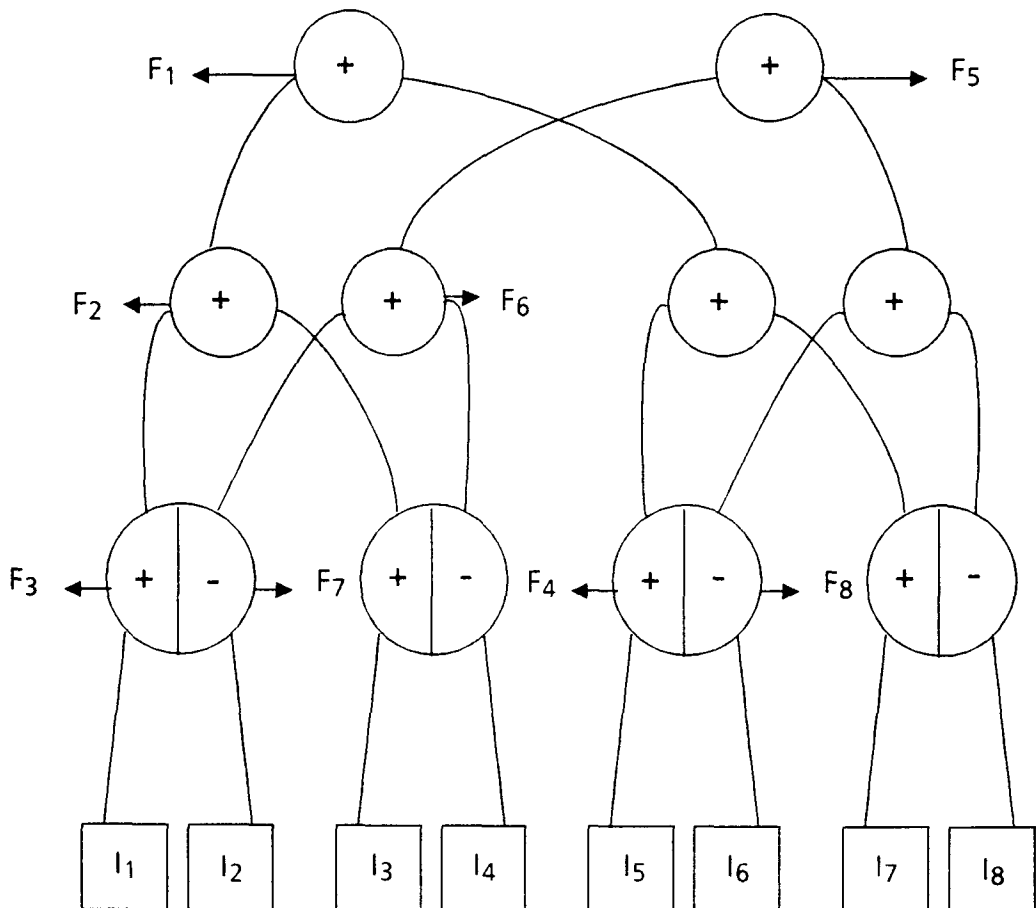


Figura 3.11

El núcleo de la transformada inversa puede expresarse mediante la siguiente matriz:

F ₁	F ₂	F ₃	F ₄	F ₅	F ₆	F ₇	F ₈
0	0	1	0	0	0	1	0
0	0	1	0	0	0	-1	0
0	2	-1	0	0	2	-1	0
0	2	-1	0	0	-2	1	0
0	0	0	0	0	0	0	1
0	0	0	1	0	0	0	-1
4	-2	0	-1	4	-2	0	-1
4	-2	0	-1	-4	2	0	1

matriz para N = 8

Primero se procederá a calcular sumas y diferencias de parejas:

$$A_1 = F_1 + F_5$$

$$A_5 = F_1 - F_5$$

$$A_2 = F_2 + F_6$$

$$A_6 = F_2 - F_6$$

$$A_3 = F_3 + F_7$$

$$A_7 = F_3 - F_7$$

$$A_4 = F_4 + F_8$$

$$A_8 = F_4 - F_8$$

Con estos valores A_i se generan los nuevos B_i mediante asignación directa y restando ponderadamente parejas alternadas, resultando:

$$B_1 = A_3$$

$$B_5 = A_4$$

$$B_2 = A_7$$

$$B_6 = A_8$$

$$B_3 = 2 \cdot A_2$$

$$B_7 = 4 \cdot A_1 - 2 \cdot A_2$$

$$B_4 = 2 \cdot A_6$$

$$B_8 = 4 \cdot A_5 - 2 \cdot A_6$$

A partir de los B_i obtenemos de forma análoga los C_i :

$$C_1 = B_1$$

$$C_5 = B_5$$

$$C_2 = B_2$$

$$C_6 = B_6$$

$$C_3 = B_3 - B_1$$

$$C_7 = B_7 - B_5$$

$$C_4 = B_4 - B_2$$

$$C_8 = B_8 - B_6$$

Luego la matriz núcleo de la transformada inversa se puede descomponer en el producto de tres matrices escasas:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 4 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & -2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \end{bmatrix}$$

$$H = A * B * C$$

Lo que hace que esta transformación sea susceptible, igualmente, de algoritmo rápido:

Generar un vector calculando sumas y diferencias de parejas

Colocar I° coeficiente multiplicado por $N/2$

Para $I = 1, \log_2 N$

colocar coeficientes

restar parejas

Fin para

En la figura 3.12 se muestra el esquema correspondiente a la computación distribuida.

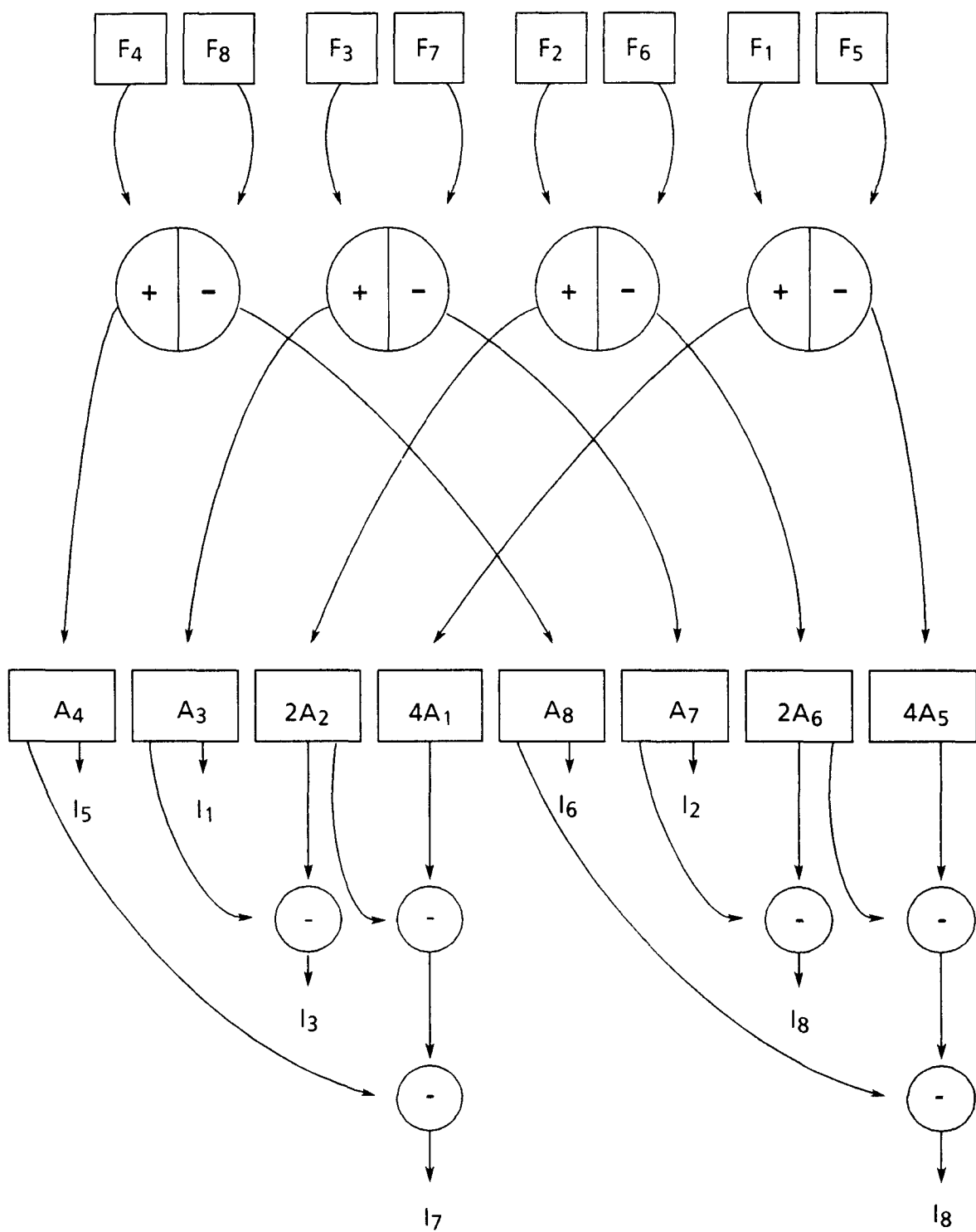


Figura 3.12

CAPITULO 4

APLICACION AL PROCESO Y RECONOCIMIENTO VISUAL

4.1 UN SISTEMA PRACTICO PARA VISION ARTIFICIAL

En este capítulo describimos las aplicaciones realizadas en base un sistema reducido de inferencia realizado por J.A Muñoz y S. Candela bajo la orientación de R.Moreno-Díaz [35].

Todas las aplicaciones desarrolladas, lo han sido en los sistemas existentes en el Departamento de Informática y Sistemas de la Universidad de Las Palmas de Gran Canaria, cuyo esquema se presenta en la figura 4.1.

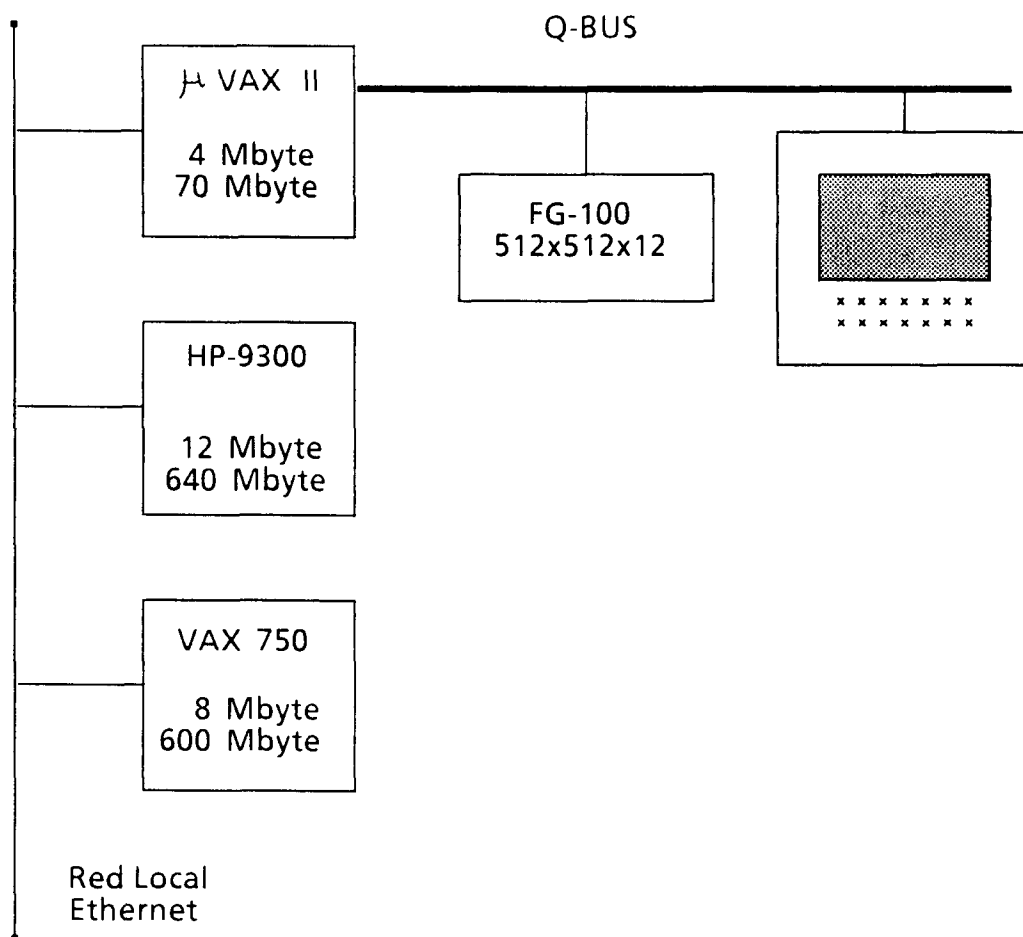


Figura 4.1

Como puede apreciarse en dicha figura existe una amplia proliferación de equipos en los cuales se ha ido desarrollando el software utilizado para llevar a cabo estas aplicaciones. Cronológicamente el proceso ha sido el siguiente:

1°) En el sistema VAX II y con las placas de adquisición FG-100 de Imaging Tec., así como cámaras de estado sólido, se realizó la adquisición de un banco de imágenes, que fueron preprocesadas para obtener imágenes invariantes frente a intensidades

2°) Trás la adquisición del VAX 750, las imágenes adquiridas fueron trasvasadas a dicho sistemas, donde se continuó con el proceso de las mismas.

Sistemas Reducidos de Inferencia.

En esta sección expondremos una estructura general de sistemas de inferencia que toman decisiones acerca del mundo visual directamente a partir de los datos visuales.

Un sistema reducido de inferencia (SRI) tiene la naturaleza de un Sistema Experto con los siguientes componentes:

a) Una base de datos referencial, que contiene un conjunto de frases **descriptoras** de referencia, a cada una de las cuales hay asignada una etiqueta. Esta base de datos está construida a priori, en base de conocimiento inyectado, y/o se construye a partir de un proceso de aprendizaje, donde se inyecta parcialmente el conocimiento que permite etiquetar las frases descriptoras.

b) Un dispositivo o máquina de inferencia, que a partir de ciertas reglas (normalmente, de naturaleza if..then..else) acepta una **frase descriptora incógnita** y, tras la interacción con la base de datos referencial, produce una decisión o diagnóstico.

La estructura relacional de estos dos componentes se muestra en la figura 4.2.

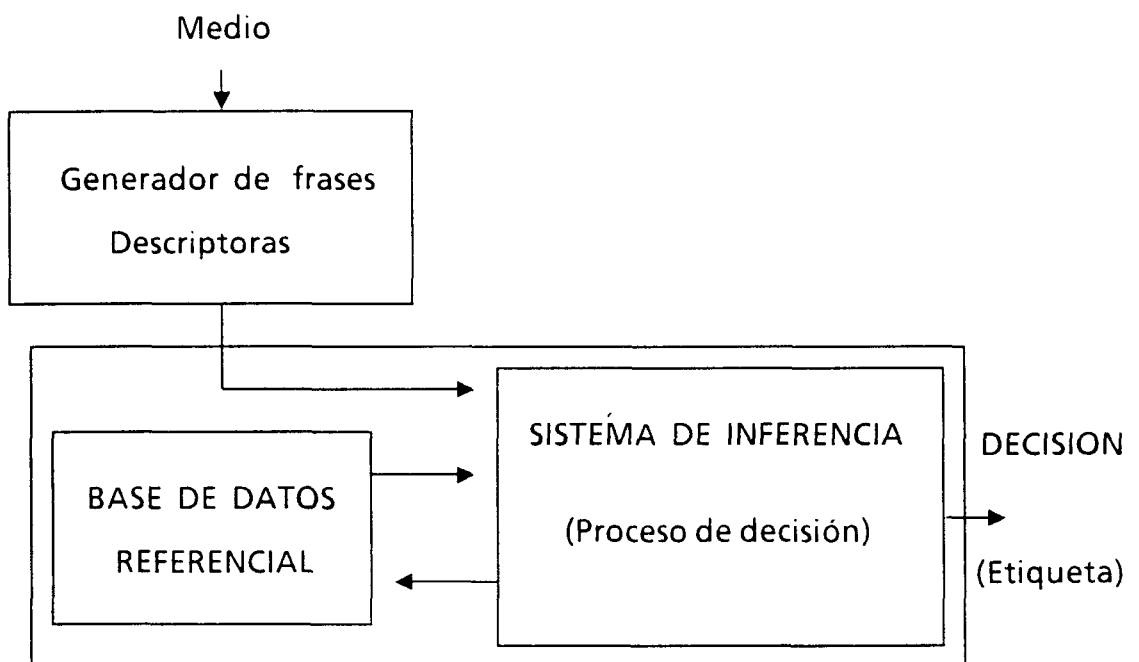


Figura 4.2

Integración de un Sistema Reducido de Inferencias en un Sistema Visual.

Consideraremos el caso de que el Sistema Visual no sea interactivo con el medio. Tal situación corresponde a la estructura básica de la figura 4.2, ampliada con las especificaciones sobre la naturaleza del generador de frases descriptoras y tres mecanismos de inyección de conocimiento (por aprendizaje o por conocimiento de procedencia puramente exterior). Hemos de notar que, en cualquier caso, la inyección de conocimiento requiere la acción del sistema exterior: en el

aprendizaje, éste conocimiento se utiliza como **asesor** acerca de la corrección de la decisión para modificar los distintos bloques funcionales, mientras que en el caso del **conocimiento exterior impreso** éste se implanta directamente en dichos bloques. La situación típica y general es híbrida.

En el caso de que el Sistema Visual no sea interactivo con el medio, la estructura computacional correspondiente se muestra en la figura 4.3

Los módulos operacionales del Sistema Visual propio son los indicados anteriormente y ahora desglosamos el proceso de generación de **frases descriptoras** en:

- a) Adquisición de la imagen.
- b) Preproceso y Normalización.
- c) Generación de las frases descriptoras al nivel elegido (analítico, algorítmico, simbólico), o a varios niveles.

De esta forma, es posible esquematizar la función de un sistema visual propio sin aprendizaje por experiencia, es decir, que funcione con conocimiento inyectado puro. Este esquema se muestra en la figura 4.4, y corresponde a la función realizada por la práctica totalidad de los sistemas actuales de reconocimiento visual (Sistema Visual Predeterminado).

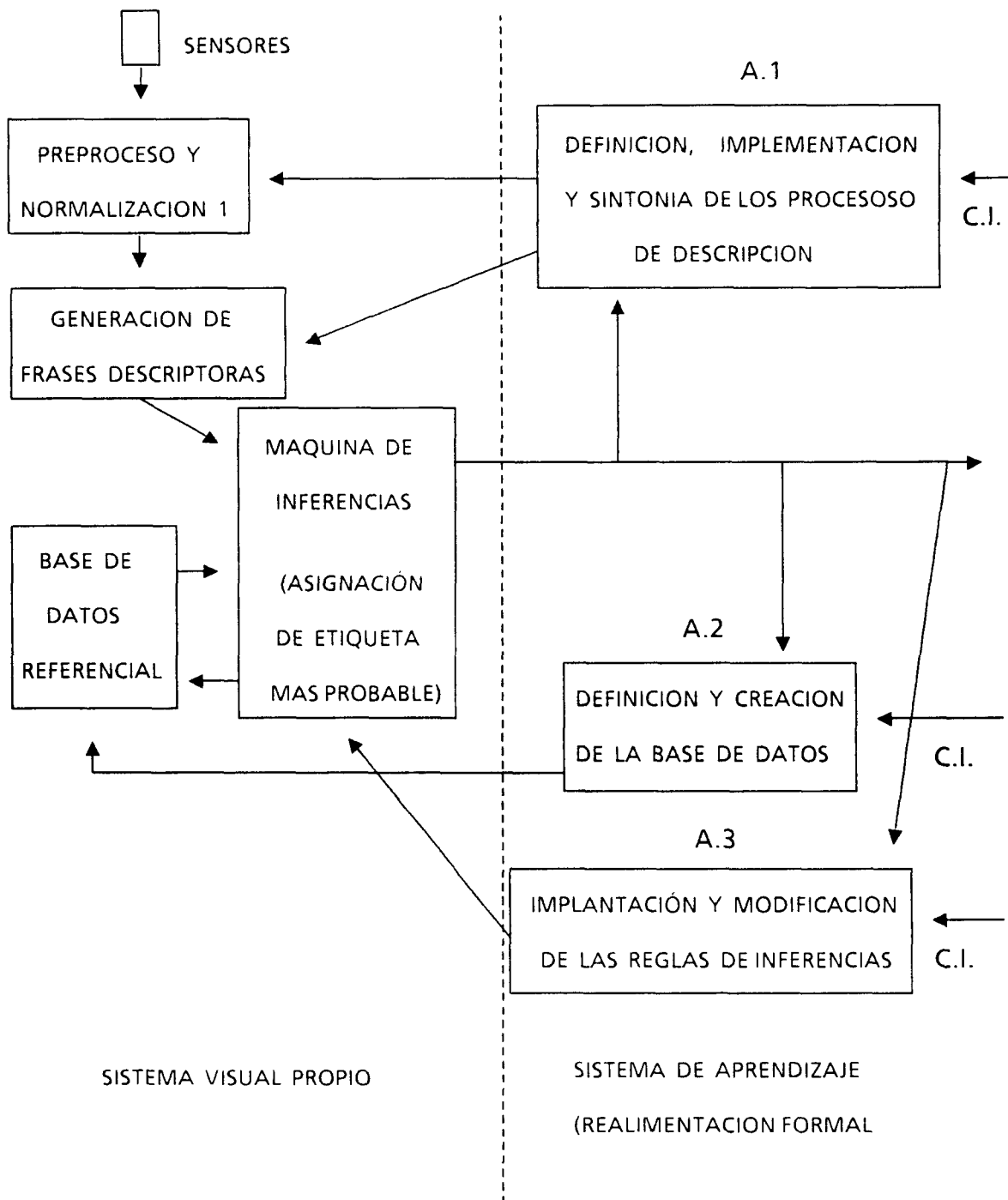


Figura 4.3. Estructura del Sistema Visual no Interactivo con el Medio

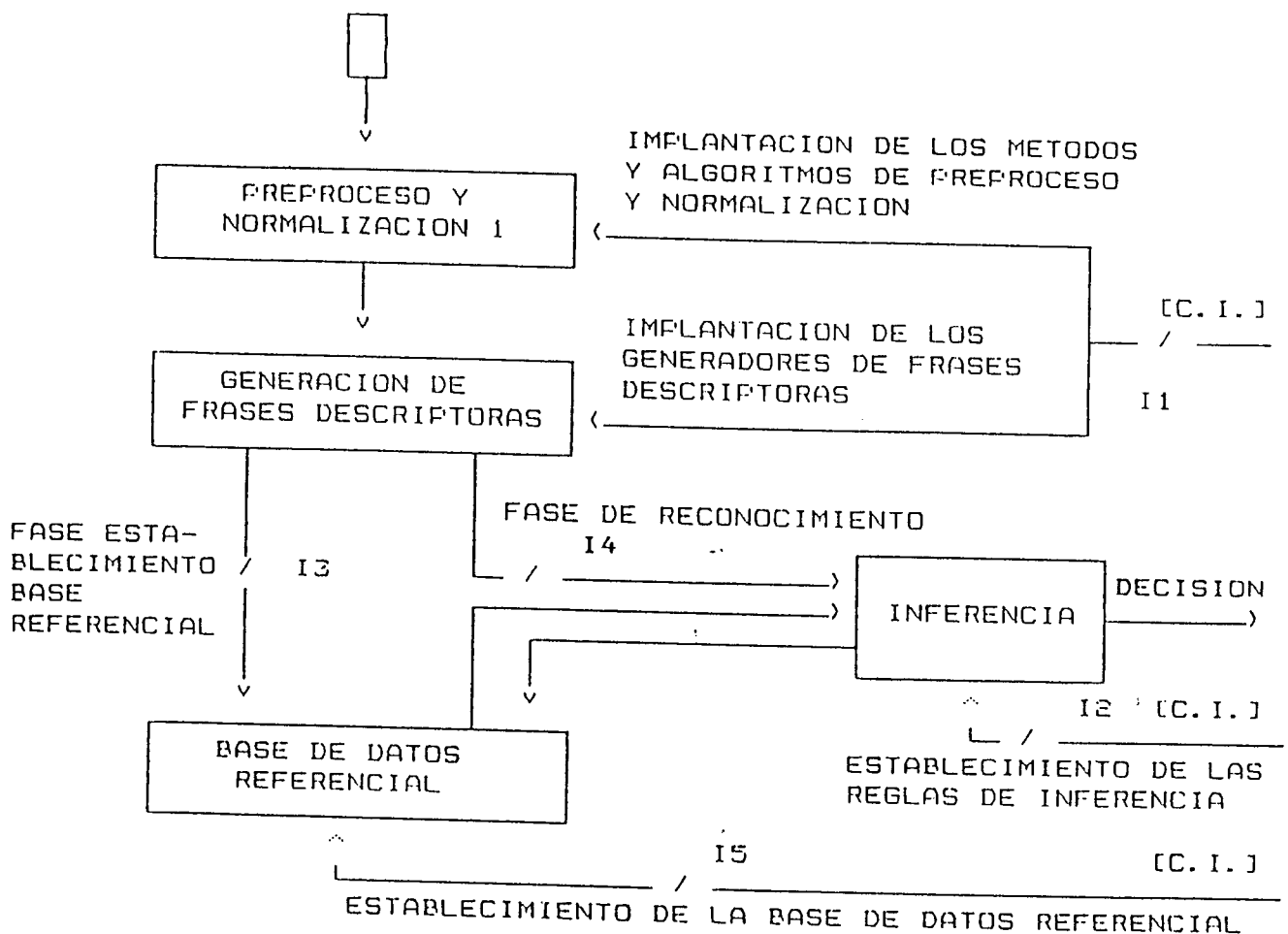


Figura 4.4. Estructura de un Sistema Visual Predeterminado

En estos sistemas se distingue la fase de establecimiento e implantación y la fase decisión. La primera fase corresponde a mantener cerrados los interruptores simbólicos I1, I2, I3 e I5 y abierto el interruptor simbólico I4. En la segunda fase de decisión (o de operación) el sistema queda desenganchado del medio, por la apertura de los interruptores simbólicos I1, I2, I3 e I5, quedando enganchado al mismo solamente a través del cierre del interruptor simbólico I4.

A continuación, en los siguientes apartados, pasamos a describir la implementación práctica de cada uno de los módulos del sistema propuesto según la figura 4.4.

4.2 PREPROCESO Y NORMALIZACION.-

En este apartado se describen los procesos llevados a cabo sobre las imágenes adquiridas, tanto las incógnitas a reconocer como las utilizadas para formar la base de datos referencial, para hacerlas invariantes tanto a diferencias de iluminación como a transformaciones afines.

4.2.1 Preproceso.

Uno de los aspectos que más influye en cualquier sistema de reconocimiento visual, aunque tal vez sea en el que menos hincapié se ha hecho hasta la fecha, es la iluminación tanto ambiental como directa, sobre las imágenes a adquirir para su posterior reconocimiento.

En nuestro caso particular en el que las frases descriptoras extraídas de las imágenes adquiridas van a depender solo de la forma de las mismas, es decir, del número de pixels que las conforman y su posición dentro del sistema de referencia utilizado, el método de preproceso utilizado, tendente a hacer a las imágenes adquiridas invariantes frente a las condiciones de iluminación, consiste simple y exclusivamente en un umbralizado binario de la imagen cuyo fin es separar fondo de forma.

De los distintos programas de umbralizado desarrollados por el grupo, se eligió el método de la cuartica [42], que consiste en el cálculo del histograma de la

imagen, frecuencia de todas las tonalidades de grises, versus tonos de gris (fig. 4.5) y aproximar la envolvente del histograma a una cuartica, por presentar el histograma dos lomas, una mayor con forma gaussiana correspondiente al fondo, otra más pequeña correspondiente a la pieza, y un valle entre estas.

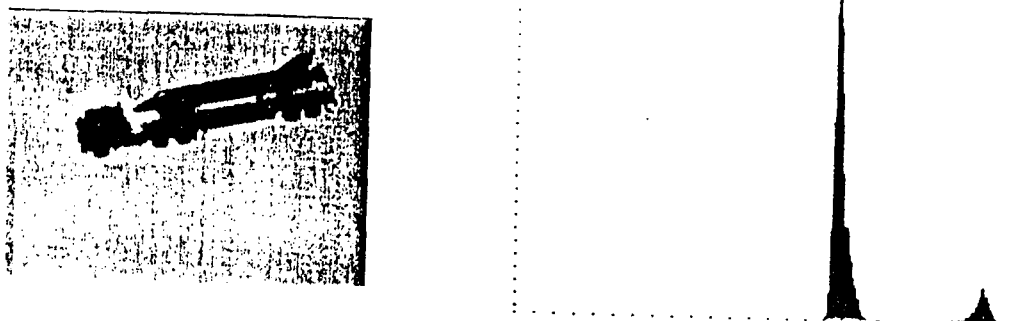


Figura 4.5



Figura 4.6

Calculando el mínimo de la cuartica (el valle) y cortando por este tono de gris, podemos separar el fondo de la imagen de la pieza en cuestión y umbralizar esta última a un valor. Este método tiene la desventaja del costo computacional, como el cálculo del histograma, la aproximación a una cuartica, el cálculo del mínimo,

corte y umbralizado, pero tiene la ventaja de ser automático, esto es, no interviene el operador, y los resultados son aceptables como lo ilustra la fig. 4.6.

4.2.2 Normalización.-

Una transformación afín es una transformación lineal cuya expresión general tiene la forma.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} c & d \\ e & f \end{bmatrix} \begin{bmatrix} x + a \\ y + b \end{bmatrix}$$

donde a, b, c, d, e y f son constantes.

El estudio de este tipo de transformaciones y sus invariantes algebraicos fueron iniciados por Boole, Cayley y Sylvester hace más de cien años. Aquí consideraremos las transformaciones afines con representación directa [30]. Si pretendemos la conservación del número de grados de libertad, el procedimiento es sencillo, mientras que si no coinciden, es preciso introducir reglas de transformación de los funcionales. Como veremos, las traslaciones y giros son del primer tipo, mientras que las homotecias son del segundo.

Traslaciones.

En el caso particular $c = f = 1, e = d = 0$

la expresión general de la transformada afín toma la forma:

$$\text{a) Transformación de dominios} \quad \left| \begin{array}{l} x' = x + a \\ y' = y + b \end{array} \right.$$

$$\text{b) Transformación funcional} \quad \left| f(x' y') = f(x y) \right.$$

Cuando se trabaja con imágenes, que representan piezas aisladas, una transformación que tiene especial interés es la traslación de la pieza al origen de coordenadas, esto es el caso que a y b coinciden con el centro de gravedad de la pieza X_g, Y_g .

Las coordenadas del centro de gravedad de una pieza, son de fácil obtención como lo apunta la teoría general de momentos, ya que precisamente una de las propiedades de los momentos de bajo orden es que nos dan las coordenadas del centro de gravedad, así que estas pueden ser calculadas mediante las expresiones.

$$X_G = M_{10}/M_{00} \quad Y_G = M_{01}/M_{00}$$

Aplicando este tipo de transformación, conseguimos que las imágenes sean invariantes frente a cualquier tipo de traslación de coordenadas.

Rotaciones.

La expresión de una transformación que gire la imagen un ángulo ϕ se consigue dando a las constantes los siguientes valores:

$$a = b = 0$$

$$c = d = \cos \phi$$

$$e = \text{sen } \phi$$

$$f = -\text{Sen } \phi$$

que nos dan la expresión:

a) Transformación de dominio

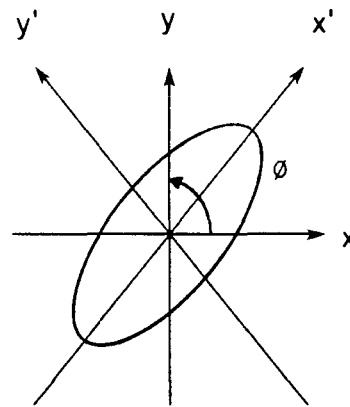
$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \text{Cos } \phi & \text{Sen } \phi \\ -\text{Sen } \phi & \text{Cos } \phi \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

b) Transformación funcional $f(x' y') = f(x y)$, donde

$$J = \begin{bmatrix} \text{Cos } \phi & \text{Sen } \phi \\ -\text{Sen } \phi & \text{Cos } \phi \end{bmatrix} = + 1$$

Para conseguir invarianza algebraica respecto a rotaciones debemos calcular el parametro ϕ . Una forma fácil de cálculo es la que aporta la teoría de momentos. Cuando una imagen se caracteriza sólo con momentos de hasta segundo orden,

esta es equivalente a una elipse homogénea, con un tamaño, orientación y excentricidad definidos. Los ejes (x' y') solidarios a esta elipse (figura) y por tanto, a la imagen, se denominan ejes principales. Es posible calcular el ángulo que forman los ejes (x' y') y (x , y) en función de los momentos determinados respecto a (x , y). Si a continuación se rota la imagen un ángulo ϕ en el sentido adecuado, habremos llevado la imagen a una posición estandar en la que coinciden los ejes (x' , y') y (x , y); y de este modo habremos normalizado frente a rotaciones.



El ángulo ϕ se determina mediante los momentos de bajo orden normalizados respecto al centro de gravedad, utilizando la expresión:

$$\phi = \frac{1}{2} \arctg \frac{2 M_{11}}{M_{20} - M_{02}}$$

Existe una ambigüedad en el cálculo del ángulo ϕ , a partir de la expresión anterior, la cual puede evitarse de diversas formas, entre otras, la siguiente consiste en elegir ϕ de tal forma que este sea el ángulo entre el eje x y el semieje mayor de la elipse a ($a \geq b$) y eligiendo el valor principal del arcotangente, es decir:

$$-\pi/2 < \arctg \phi < \pi/2$$

con estas consideraciones, obtenemos para ϕ los valores de la tabla:

M20-M02	M11	ϕ	
0	0	0°	$a = 2M11 / (M20 - M02)$
0	+	45°	
0	-	- 45°	
+	0	0°	
-	0	- 90°	
+	+	(½ arctg a)	(0° < ϕ < 45°)
+	-	(½ arctg a)	(-45° < ϕ < 0°)
-	+	(½ arctg a) + 90°	(45° < ϕ < 90°)
-	-	(½ arctg a) - 90°	(-90° < ϕ < -45°)

Realizadas varias pruebas con distintas piezas, se observa que el ángulo calculado mediante momentos tiene un cierto error, que se acentúa en aquellas piezas que tienen formas alargadas.

Un nuevo método, que hemos desarrollado, consiste en calcular dicho ángulo como el formado por el vector que une el punto más alejado del contorno de la forma y el centro de gravedad de la misma con respecto a un eje horizontal. Este método, como es fácil de comprobar, solo es válido en aquellos casos en que la forma no tenga ningún tipo de simetría respecto al centro de gravedad ni respecto a los ejes cartesianos que pasan por él, como es nuestro caso.

Para el cálculo de dicho vector nos hemos ayudado del código cadena del contorno de la forma, como puede verse en la figura 4.7.



Figura 4.7

La ventaja encontrada en este método frente al de los momentos es doble : Por un lado su mayor precisión y por otro su mayor velocidad de cómputo.

Homotecias.

Haciendo que las constantes a, b, d, e, sean cero la expresión general de la transformada afín se reduce a la ecuación:

a) Transformación de dominio

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} c & o \\ o & f \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

que representa la transformación de dominios para una homotecia. En el caso particular que las constantes c y f sean iguales, la relación de homotecia es la misma tanto para el eje x , como para el eje y .

b) Transformación funcional.

Los criterios de transformación del funcional, se pueden conseguir imponiendo invarianza respecto a ampliaciones, calculando el factor de homotecia c , imponiendo la condición de que la imagen transformada tenga un área predefinida constante A_k , esto es

$$c = \text{SQRT}(A_k/A)$$

donde A es el área de la pieza calculada mediante el momento M_{00} .

Mencionemos los casos.

- 1) Cuando $A_k > A$, existe una expansión apareciendo nuevos puntos, tendremos que definir un criterio para "rellenar" funcionalmente estos puntos que antes no existían. Un criterio utilizado, consiste en barrer la imagen transformada y asociar a cada punto el valor del pixel de la imagen original calculado por las expresiones $x = x'/c$; $y = y'/c$, esto es, se esta efectuando una aplicación inyectiva en la que a varios pixels de la imagen transformada le corresponde un unico punto en la imagen original.

- b) Caso en que $A_k < A$, existe una contracción, por lo que hay puntos que desaparecen, tendremos que definir un criterio para ver como estos puntos

que desaparecen, los hacemos intervenir, en la imagen transformada. Un criterio consiste en hacer un enrejado en la imagen original, y calcular para un punto transformado la media de los valores de una cuadrícula del enrejado.

4.3 GENERACION DE LA BASE DE DATOS REFERENCIAL Y DE LA MAQUINA DE INFERENCIA

Generación de Frases Descriptoras

Para la generación de las frases descriptoras que deben etiquetar a cada una de las clases que conforman las formas, tanto patrones como incógnitas, para su posterior reconocimiento, de entre todas las transformadas posibles expuestas en el capítulo anterior, hemos elegido una transformación de campo de datos homogénea caracterizada por los siguientes parámetros:

Dimensión : $N = 128$

Nº Particiones : $L = 7$

Grados de libertad de cada partición : $d = 32$

Desplazamiento interparticiones: $= 16$

Nº Funcionales por partición : Particiones 1,2,3,5,6,7 ----- 18

Partición 4-----20

Funcionales : Momentos

En la figura 4.8 se muestra el núcleo de la transformación, así como el esquema de las particiones.



Figura 4.8

Una vez elegida la transformación, para la generación de las frases descriptoras, solo falta definir que coeficientes se eligen para su construcción. Para ello se han llevado a cabo diferentes elecciones con el fin de buscar cual es la que da mejores resultados a la hora de clasificar.

Antes de indicar cuales han sido cada una de las elecciones, y con vistas a una mejor comprensión de las mismas, vamos a indicar en la figura 4.9 cual es la ordenación que se ha llevado a cabo para cada uno de los funcionales (momentos) de cada una de la particiones.

Como puede observarse, en la figura 4.9, se han agrupado los momentos de orden ij de cada una de las particiones. La colocación del momento ij correspondiente a las distintas particiones es la siguiente

$P_{0,0}$	$P_{0,1}$	$P_{0,2}$	-----	$P_{0,6}$
$P_{1,0}$	$P_{1,1} \dots P_{1,2}$		-----	$P_{1,6}$
.	.	.		.
.	.	.		.
.	.	.		.
.	.	.		.
$P_{6,0}$	$P_{6,1}$	$P_{6,2}$	-----	$P_{6,6}$

	7	7		7	1	1
7	M ₀₀	M ₀₁		M _{0,17}	M _{0,18}	M _{0,19}
7	M ₁₀	M ₁₁		M _{1,17}	M _{1,18}	M _{1,19}
7	M _{17,0}	M _{17,1}		M _{17,17}	M _{17,18}	M _{17,19}
1	M _{18,0}	M _{18,1}		M _{18,17}	M _{18,18}	M _{18,19}
1	M _{19,0}	M _{19,1}		M _{19,17}	M _{19,18}	M _{19,19}

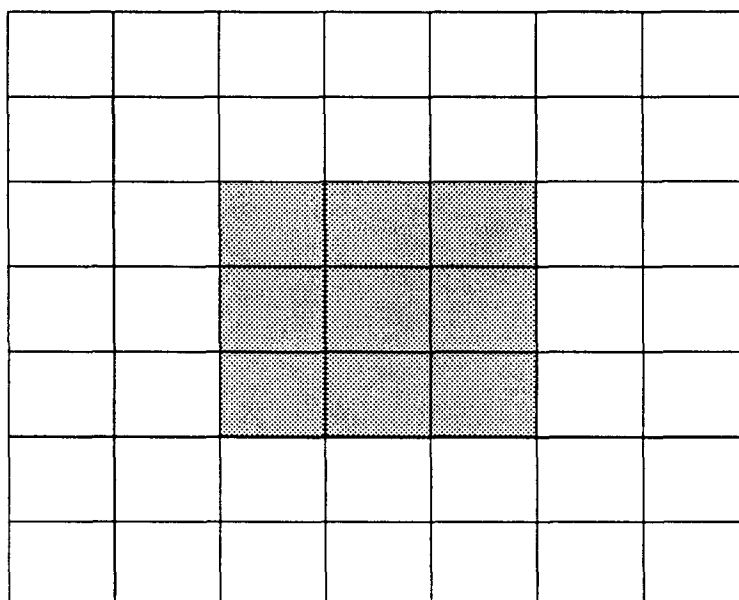
Figura 4.9

Los momentos utilizados para la generación de las frases descriptoras han sido los momentos M₀₃, M₀₄, M₁₂ y M₁₃ propuestos por Alt [7] como aquellos que dan mejores resultados a la hora de clasificar.

A continuación pasamos a detallar cada una de las elecciones realizadas para la generación de las frases descriptoras.

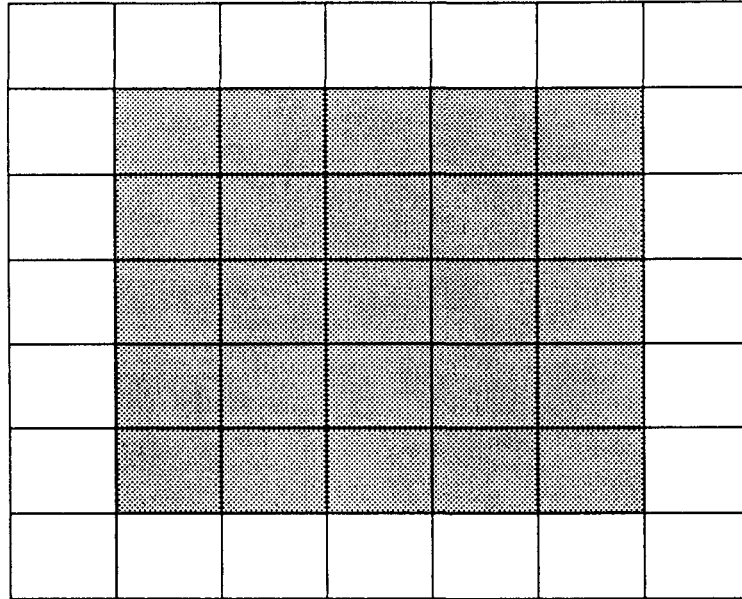
CENTRO CRECIENTE : Hemos denominado así al conjunto de los 12 tipos diferentes de frases descriptoras correspondientes a tres variaciones distintas de cada uno de los 4 momentos anteriormente indicados, a saber

a) Apartir de la particion central $P_{3,3}$ y en forma creciente se escogen un momento de cada una de las nueve particiones centrales, de la siguiente forma



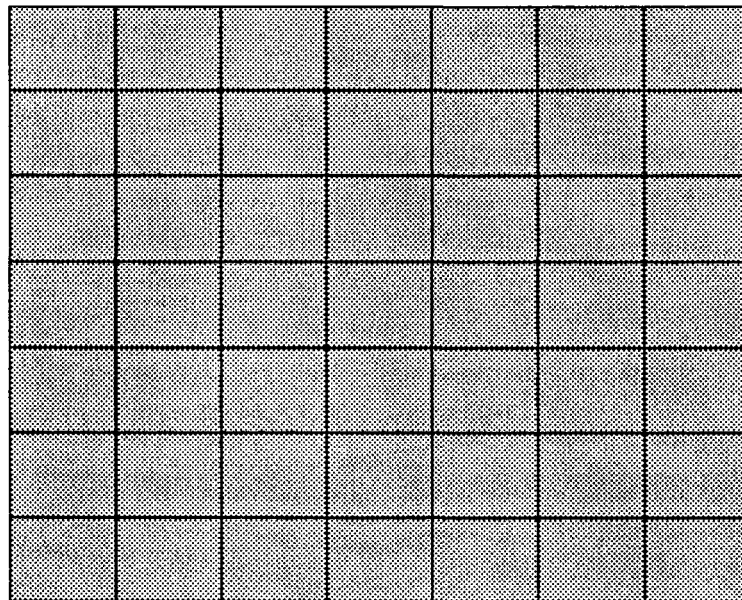
9

b) Apartir de la particion central $P_{3,3}$ y en forma creciente se escogen un momento de cada una de las veinte y cinco particiones centrales, de la siguiente forma



25

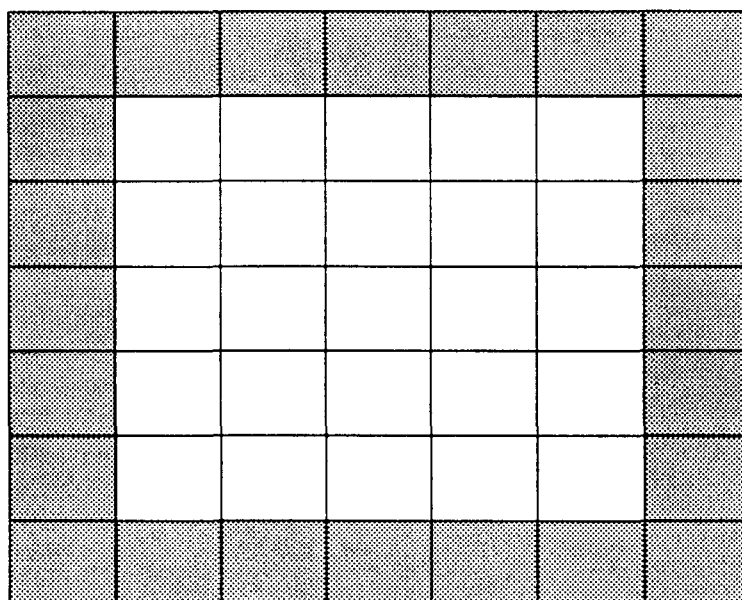
c) A partir de la particion central $P_{3,3}$ y en forma creciente se escogen un momento de cada una de las cuarenta y nueve particiones centrales, de la siguiente forma



49

PERIFERIA CRECIENTE : Hemos denominado así al conjunto de los 8 tipos diferentes de frases descriptoras correspondientes a dos variaciones distintas de cada uno de los 4 momentos anteriormente indicados, a saber

a) Apartir del contorno de la distribución de los momentos correspondientes a cada partición y en forma creciente se escogen un momento de cada una de las veinte y cuatro particiones periféricas, de la siguiente forma

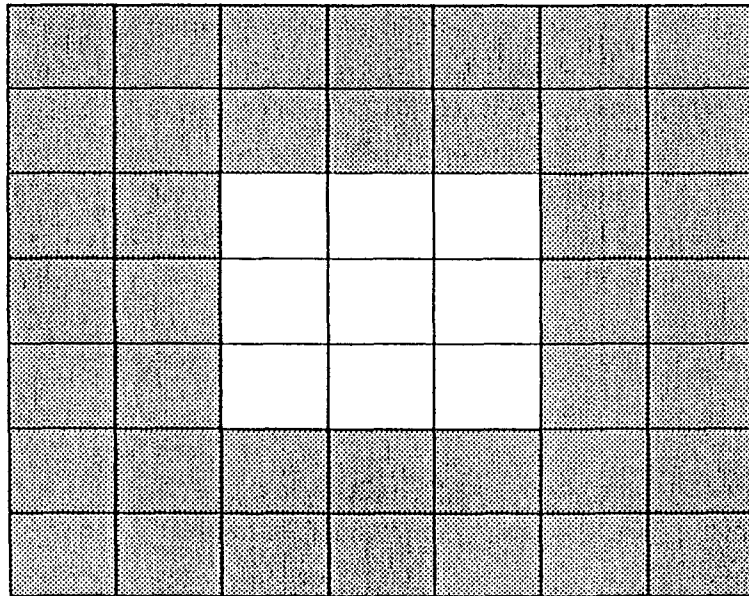


24

b) Apartir del contorno de la distribución de los momentos correspondientes a cada partición y en forma creciente se escogen un momento de cada una de las cuarenta particiones periféricas, de la siguiente forma

GENERACION DE LA BASE DE DATOS REFERENCIAL.-

Para llevar a cabo el proceso de reconocimiento se generó una base de datos referencial a partir de imágenes de prototipos de vehiculos, tanto militares



40

(camiones y tanques) como civiles (automóviles diversos), como se muestra en la figura 4.8

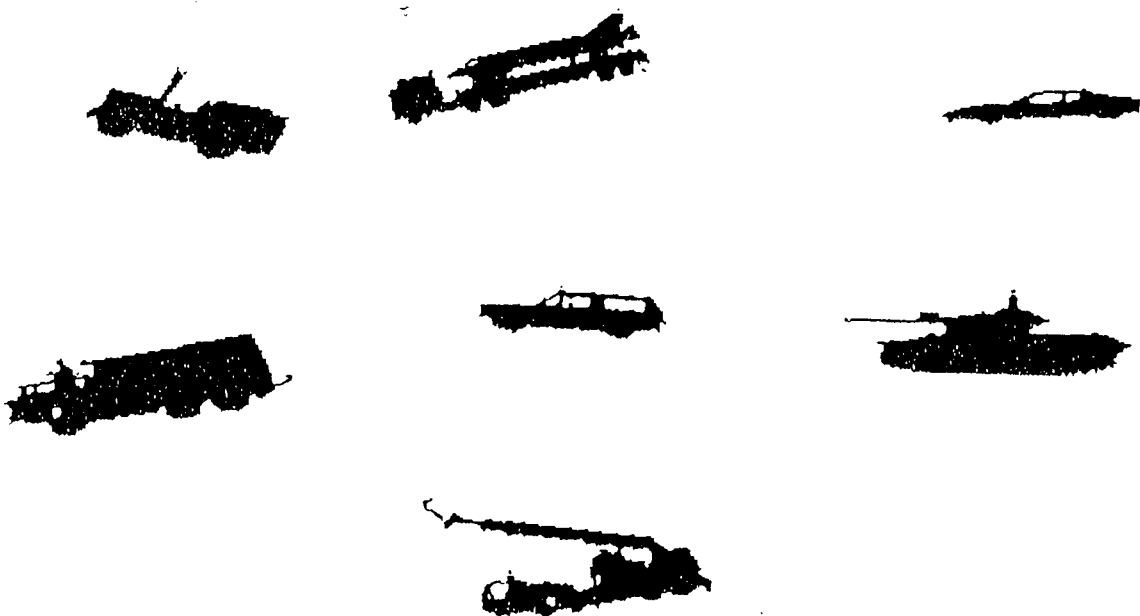


Figura 4.10

Mediante una cámara de estado sólido (CCD), y unas placas digitalizadoras, se toman imágenes digitalizadas de 512x512 pixels y cada pixel con 2^8 niveles de gris.

De cada pieza se realizaron diez vistas, entre toma y toma se variaba el zoom de la cámara para conseguir variaciones en amplitud, se movía la pieza para conseguir traslaciones y se giraba la pieza, dando como resultado que las vistas sean todas distintas.

Sólo se mantiene constante la luz ambiental, ya que en este experimento no se tiene en cuenta los cambios en intensidad, puesto que, como hemos visto anteriormente las imágenes se someten a un preproceso de umbralizado.

Tras comprobar que la frecuencia de muestreo de la cámara varía en horizontal y vertical con una relación aproximada de dos o tres, se corrige esta deformación de la imagen tomando 128x128 puntos, con la misma proporción anterior, de los originales 512x512.

Al finalizar este preproceso obtuvimos un banco de imágenes de siete piezas con diez vistas de cada una, (cada imagen consiste de 128x128 enteros) y están umbralizadas en intensidad a dos niveles, cero el fondo, y uno la pieza. Estas imágenes mediante los correspondientes programas de almacenamiento se guardan en ficheros sobre discos.

El siguiente paso consiste en elegir una imagen normalizada cualquiera, de las diez vistas que componen el set de una pieza, y tomarla como "clase patron" o referencia y asignarle una etiqueta que denota la pieza que representa. Repitiendo esta elección para las siete piezas que componen nuestro banco de

imagenes, tendremos un conjunto de frases descriptoras patrones con sus respectivas etiquetas, que seran los datos a almacenar en nuestra base de datos referencial.

En nuestra implementación, los datos almacenados en la base de datos referencial tienen una estructura sencilla de array. Se elige esta estructura por tener pocas clases patrones, lo cual obliga a utilizar algoritmos de búsqueda secuenciales de fácil implementación pero de tiempo de respuesta no optimizado.

MAQUINA DE INFERENCIA .-

Una vez descrito este bloque, vamos a pasar a describir ahora como hemos construido el sistema de inferencia utilizado para la toma de decisiones de asignación de formas incógnitas a una clase determinada, en otras palabras, el método de clasificación utilizado.

En nuestro caso hemos utilizado un método de clasificación no paramétrico, que no utiliza ningún conocimiento de las densidades de probabilidad de las clases, ni funciones discriminantes. Concretamente se utiliza una versión modificada de la distancia Euclídea [35] para asociar una forma incógnita a una clase. La distancia utilizada se define de la siguiente manera.

Dada una frase descriptora incógnita $D'(i)$ y un conjunto de frases descriptoras, correspondientes a cada una de las clases de la base de datos referencial, $D(i,j)$, se define la distancia Euclídea modificada como

$$[\text{Distancia}(j)]^2 = \sum [(D'(i) - D(i,j)) \cdot 10^{E_{ij}}]^2$$

donde

$$E(j) = \min (\text{exponente } \{D'(i)\}, \text{exponente } \{D(i,j)\}) \\ (i = 1, \dots, n)$$

siendo n el número de elementos de las frases descriptoras.

Una vez calculadas todas las distancias entre la frase descriptora incógnita y cada una de las frases descriptoras de las clases que componen la base de datos referencial, el sistema de inferencia asocia el patrón incógnita a aquella clase cuya distancia al patrón incógnita sea mínima.

Por último sólo queda hacer notar que la base de datos referencial está compuesta por las frases descriptoras de cada una de las tomas de las formas que representan a dicha clase.

4.5 RESULTADOS EXPERIMENTALES.-

En este apartado vamos a exponer los resultados experimentales alcanzados al llevar a cabo el proceso de clasificación de las formas incógnitas según los distintos modelos de frases descriptoras utilizados según se expuso en el apartado 4 del presente capítulo.

El conjunto de las imágenes incógnitas a reconocer está formado por 10 tomas diferentes de cada una de las 7 formas distintas.

Los resultados experimentales obtenidos han sido los siguientes, clasificados según momentos.

Momento 03 : Centro Creciente : a) N° Aciertos = 21

N° Fallos = 0

Eficiencia = 100 %

b) N° Aciertos = 17

N° Fallos = 4

Eficiencia = 80 %

c) N° Aciertos = 20

N° Fallos = 1

Eficiencia = 95 %

Periferia Creciente : a) N° Aciertos = 21

N° Fallos = 0

Eficiencia = 100 %

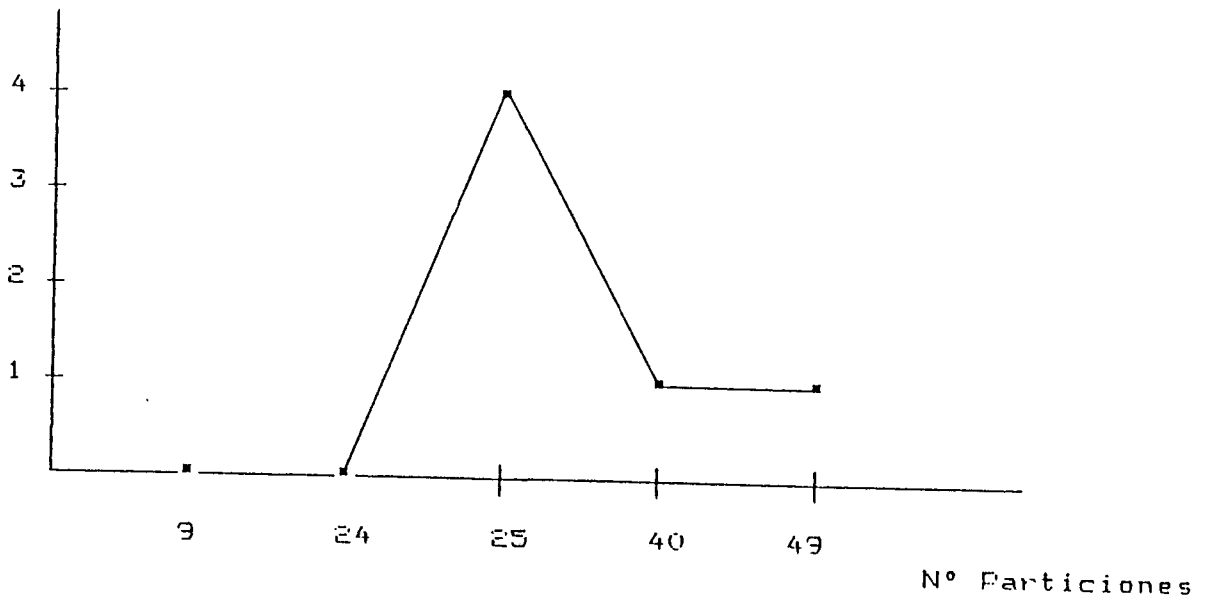
b) N° Aciertos = 20

N° Fallos = 1

Eficiencia = 95 %

Lo cual se puede mostrar de forma gráfica

N° Fallos



Momento 04: Centro Creciente : a) N° Aciertos = 21

N° Fallos = 0

Eficiencia = 100 %

b) N° Aciertos = 17

N° Fallos = 4

Eficiencia = 80 %

c) N° Aciertos = 20

N° Fallos = 1

Eficiencia = 95 %

Periferia Creciente : a) N° Aciertos = 21

N° Fallos = 0

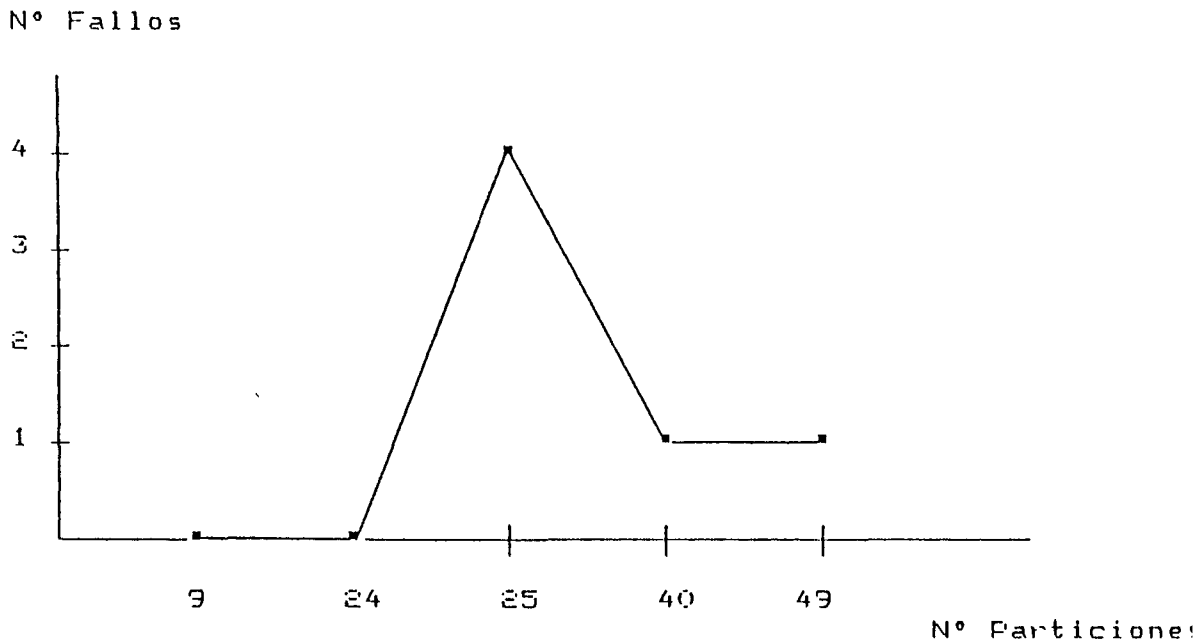
Eficiencia = 100 %

b) N° Aciertos = 20

N° Fallos = 1

Eficiencia = 95 %

Lo cual se puede mostrar de forma gráfica según



Momento 12 : Centro Creciente : a) N° Aciertos = 21

N° Fallos = 0

Eficiencia = 100 %

b) N° Aciertos = 19

N° Fallos = 2

Eficiencia = 90 %

c) N° Aciertos = 20

N° Fallos = 1

Eficiencia = 95 %

Periferia Creciente : a) N° Aciertos = 21

N° Fallos = 0

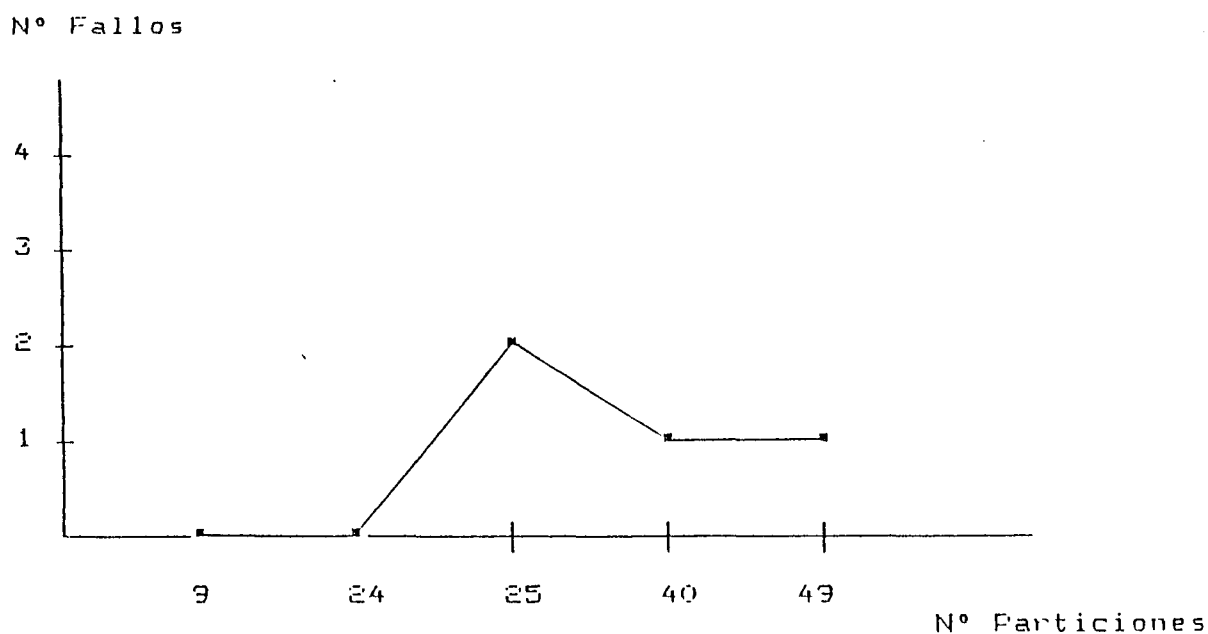
Eficiencia = 100 %

b) N° Aciertos = 20

N° Fallos = 1

Eficiencia = 95 %

Lo cual se puede mostrar de forma gráfica según



Momento 13 : Centro Creciente : a) N° Aciertos = 21

N° Fallos = 0

Eficiencia = 100 %

b) N° Aciertos = 18

N° Fallos = 3

Eficiencia = 86 %

c) N° Aciertos = 20

N° Fallos = 1

Eficiencia = 95 %

Periferia Creciente : a) N° Aciertos = 21

N° Fallos = 0

Eficiencia = 100 %

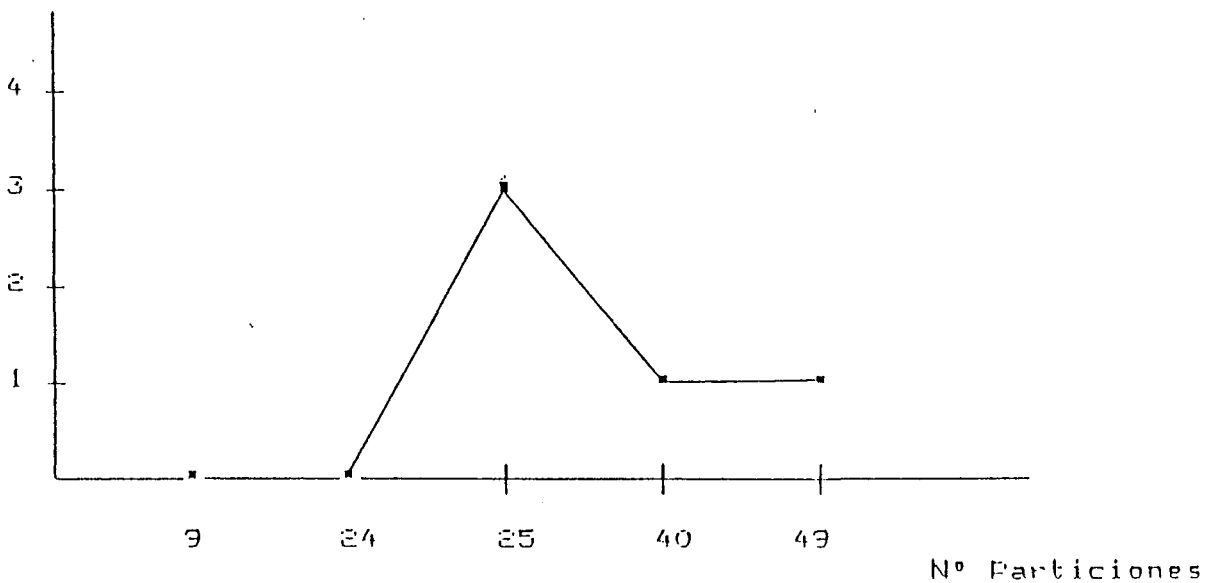
b) N° Aciertos = 20

N° Fallos = 1

Eficiencia = 95 %

Lo cual se puede mostrar de forma gráfica según

N° Fallos



La eficacia del sistema clasificador depende de varios factores como son la complejidad del campo de datos, la resolución, la mayor o menor extensión de los campos receptivos y el número de descriptores elegido en cada uno de éstos.

Podría parecer de forma intuitiva que ésta eficacia sería tanto mayor cuanto menor sea la complejidad del campo de datos y la resolución, y cuanto mayor sea el número de campos receptivos y los descriptores en cada uno de ellos. Sin embargo, cuando la resolución es relativamente baja y la complejidad del campo de datos es pequeña, un número alto de campos receptivos y de descriptores provoca una disminución de la eficacia en la clasificación, debido al ruido inherente y al aumento, no controlado por ser no detectable, de las redundancias en el campo de datos original.

CAPITULO 5

CONCLUSIONES Y PRINCIPALES APORTACIONES

5. CONCLUSIONES Y PRINCIPALES APORTACIONES

Como ha sido indicado, el objetivo central de ésta tesis ha sido el integrar dos constructos generales que clásicamente se han planteado en el análisis de imágenes, el constructo analítico y el algebraico, a los que recientemente han habido contribuciones importantes, sobre todo respecto del carácter diferenciador entre ambos, por miembros de nuestro grupo. Y desde ambos, proseguir hacia teoremas, lemas y demostraciones prácticas que permitan establecer un marco teórico unificado del proceso de imágenes. Dicho objetivo se estima que ha sido alcanzado en esencia, proporcionándose, como subproducto formalmente deseable, una conexión natural con la computación distribuida, la computación rápida y las funciones de la retina y el cortex visual.

Aparte de una disección descriptiva de los sistemas visuales artificiales y naturales, se considera que las principales aportaciones y conclusiones de este trabajo son las siguientes:

1.- Dentro del constructo analítico, hemos introducido el concepto de **convoluciones generalizadas**, que nos ha llevado a sustituir el espacio de entrada como campo receptivo, por un campo de datos, mas general y mas claro. Como tal, una imagen en el sistema artificial de adquisición o en la retina, es considerada, en definitiva, como lugares direccionables, con lectura anastomótica, no necesariamente serial o paralela.

2.- Este concepto es estrictamente formalizable hasta niveles no previamente predictibles por el análisis puro, pero que lo incluye como consecuencia. Como resultado se han desarrollado reglas de producción de convoluciones generalizadas, analíticas o no.

3.- Un teorema previo respecto de transformaciones no lineales divisivas ha sido integrado en nuestra formulación, y aplicado experimentalmente a imágenes reales. Las consecuencias teóricas respecto a las propiedades adaptivas de las células retinales han sido asimismo estudiadas, concluyéndose que dicho mecanismo explica procesos adaptivos ante variaciones de intensidad - invarianzas- de orden neuronal, que se suman a los mecánico-pupilares y a los bioquímicos de los fotorreceptores.

4.- Las contribuciones fundamentales de nuestro trabajo se refieren a los conceptos de Campos de Datos y Particiones de los mismos. Aparte de la reformulación de las Transformaciones de Campo Receptivo Variable y de las transformaciones foveales, que se ilustran experimentalmente, se plantea de raíz el manejo de los campos de datos, de sus grados de libertad y del significado de las transformaciones sobre los mismos.

5.- De la formulación se concluyen las definiciones formales precisas de representaciones completas, y las condiciones necesarias y suficientes para las mismas. Asimismo, se deducen como casos particulares, los casos típicos en Visión Artificial, al considerar la capacidad de los lugares de memoria y el papel del intercambio entre resolución y definición del campo de datos.

6.- Se estima que la contribución teórica central de nuestro trabajo gira alrededor del Teorema 3.1, sobre descripciones completas que pueden ser truncadas por motivos prácticos de reconocimientos. Este teorema integra el álgebra y el análisis de los campos de datos, al demostrar que, dada una partición algebraica con ciertos requisitos, entonces el cómputo de cierto número de coeficientes analíticos independientes en cada partición, proporcionan una descripción completa y no redundante del campo de datos.

7.- Otra contribución de cierto peso teórico ha sido conseguida por el teorema 3.2 respecto a la complitud en particiones arbitrarias de los campos de datos, donde un cierto conjunto de funcionales independientes lleva a la descripción no redundante y completa.

8.- Lo anterior inspira el concepto de transformaciones diagonales, que sirve de puente entre el proceso sobre campos de datos y los formalismos de los campos vectoriales.

9.- El parámetro tiempo "olvidado" en la formulación analítico-algebraica es reintroducido al considerar las redes neuronales y por capas. Lleva dicha consideración a aclarar la limitación y alcance de los algoritmos rápidos de computación.

10.- En el ensayo convencional acerca de los dichos algoritmos rápidos y del diseño de un sistema de reconocimiento, concluimos algo que era de esperar. Las descripciones completas y truncadas algebraico analíticas no son peores ni mejores que otras convencionales, puesto que no afectan a la estructura perceptual, sino a la sensorial. Sin embargo lo incluimos como requisito experimental en el planteamiento de nuestro trabajo.

11.- El futuro de nuestro trabajo, desde el punto de vista práctico, está en el replanteamiento formal de la computación distribuida, sin tener en cuenta las llamadas sin sentido a las redes neuronales, que según la actual moda, dedican un gran esfuerzo a buscar herramientas complicadas para resolver inexistentes problemas. Se estima así, como conclusión filosófica (en el sentido tradicional) que es primero una buena definición de la clase de problemas y luego la búsqueda de herramientas formales para copar con ella.

REFERENCIAS

- [1] Moreno Diaz, R.: "Visión Artificial". *Proc. Jornadas de Inteligencia Artificial, Texas Inst.*
- [2] Moreno-Díaz, R., Mira, J., Fórtes, J. y Delgado, A. "Proceso de Datos Visuales e Integración Multisensorial" *Proc. Jornadas y Fabricación Flexible. - C.S.I.C. - (1983)*
- [3] Marr, D. "Visión", W.H. Freeman, New York, 1980.
- [4] Moreno-Díaz, R., Mira Mira, J., "La Teoría de Automátas y la Teoría de Sistemas en Biocibernética". *Biocibernética. Ed. Siglo XXI. pp 85-90. 1984.*
- [5] Campbell, F.W., y Robson, J.G., "Application of Fourier Analysis to the visibility of gratings". *J. Physiol.* 197, pp 551-566. 1968.
- [6] Mendez, J.A., Falcon, A.: "On the Orthogonal Expansion of Images for Low Level Recognition. Fourier-Bessel Representation". *Proc. First Conf. Art. Int. App.*, pp. 167-169. *IEEE Press, Denver, Colorado. 198.*
- [7] Canton, M.: "La teoría de momentos en el análisis de formas: Formulación tensorial, invarianzas y aplicaciones". *Tesis Doctoral. Universidad de La Laguna. 1982.*
- [8] Nevatia, R., "Machine Perception", *Prentice Hall, New York, 1982.*
- [9] Rosen, C.A., "Machine Vision and Robotics", *Plenum Press, New York, 1979.*
- [10] Ballard, D.H., Brown, C.M.: "Computer Vision". *Prentice Hall, New Jersey, 1982.*
- [11] Faugeras, O.D.: "Fundamentals in Computer Vision: An Advanced Course". *Cambridge University Press, 1983.*
- [12] Pavlidis, T.: "Structural Pattern Recognition". *Springer-Verlag, Berlin, 1977.*
- [13] Rosenfeld, A., Kak, C.: "Digital Picture Processing". 2nd ed. *Academic Press, New York, 1982.*
- [14] Gonzalez, R.C., Wintz, P.: "Digital Image Processing". *Addison-Wesley. 1977.*
- [15] Rosenfeld, A., Thurston, M.: "Edge and curve detection for visual scene analysis". *IEEE Trans. Comput.*, C-21, 5, 1971, 562-
- [16] Andrews, H.C. "Orthogonal Transforms and Feature Selection in Pattern Recognition". *Univ. South. California. Proc. 4th Hawaii International Conf. on System Sciences, pp 675-7, 1971.*
- [17] Rao, K.R., Ahmed, N., "Orthogonal Transforms for Digital Signal Processing". *IEEE Intern. Conf. on Acoustics, Speech and Signal Processing. pp 136-40.*

[18] Cooper, J.A., "Orthogonal Transformations of Digital data". Dandia Labs., Albuquerque. N. Mex. GRAI7812.

[19] Hu, M.K.: "Visual Pattern Recognition by Moment Invariants". IRE Trans. Inform. Theory IT-8, pp. 179-187. 1962.

[20] Canton Garbín, M., Moreno-Díaz, R., "Normalized Moments for 3D characterization of Visual Patterns."Orwellian Symposium and International Conf. on Systems Research Informatics. Baden-Baden. 1984.

[21] Santana, O., Mendez J.A., Moreno Díaz, R., "Momentos normalizados para el proceso de datos visuales". Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Tomo LXXV. Cuaderno 1º. pp.287-289. 1981.

[22] Moreno-Díaz,R., Santana,O., Rubio, E., Nuñez, A. "Bases de una teoría de Proceso de Datos en la retina". Biocibernética. Ed. Siglo XXI de España Editores. pp.121-131. 1984.

[23] Núñez, A., "Una teoría unificada del proceso y transmisión de datos en la retina". Tesis Doctoral. Madrid, ETSII. 1981.

[24] Rubio, E., "Modelos retinales generalizados". Tesis Doctoral. Las Palmas. 1979.

[25] McCulloch, W. S., y Pitts, W., "How we know Universals". 1974.

[26] Moreno-Díaz, R.; Rubio, E., y Núñez, A., "A layered model for visual processing in avian retina."Biol. Cybern.,38, pp. 85-89.

[27] Moreno-Díaz, R., Rubio Royo, F., "Towards a theory of Visual Processing in Vertebrate Retina."Proc. of the Intern. Congress on Applied Systems Research and Cybernetics. 1981.

[28] Moreno-Díaz, R., Rubio,E., "A theoretical model for layered visual processing".Int.I.Biomed.Comp.,10,pp.231-243.1979.

[29] Moreno-Díaz, R., y Núñez, A., "Un enfoque intuitivo de los procesos no lineales, con aplicación a modelos neurofisiológicos". Act. I. Simp. Nac. Bioing., pp. 123-127. Madrid. ETSIT, 1981.

[30] Candela, S.: "Transformaciones de Campo Receptivo Variable en Proceso de Imágenes y Visión Artificial". Tesis Doctoral. Universidad Politécnica de Canarias. 1987.

[31] Moreno Diaz, R., Rubio Royo, E. : " A generalized model for non-linear retinal processing". Current topics in Cybernetics and systems. Rose J. (ed.). Berlin-Heidelberg-New York: Springer (1978).

[32] Moreno Diaz, R., Mira Mira, J.: Un marco teórico para interpretar la función Nueronal a altos niveles. "Biocibernética". Ed.Siglo XXI. Madrid 1984.

Moreno-Díaz, R., Mira Mira, J., y Delgado, A.: "Hacia ina teoria social de Redes Neuronales". Proc. II Simposium de Bioingenieria. (1983).

[34] Santana Suarez, O., Candela Solá, S. y Moreno Diaz,R.: "Computer non-linear and algorithmic simulation of static retinal processes". Proc. 6th Int. Congress of Cybernetics and Systems. AFCET. Paris 1984.

- [35] Muñoz Blanco, J.A. "Jerarquización de Estructuras de nivel bajo y medio para Reconocimiento Visual, Aplicaciones a texturas y formas". Tesis Doctoral. Univ de Las Palmas. (1987)
- [36] Lettvin, J.Y.: *Research Laboratory of Electronics Quarterly Progress Report No. 73, Massachussets Institute of Technology, Cambridge, Massachussets, 15 April 1965, pp. 199-208.*
- [37] Candela, S. Muñoz, J.A.: "Transformations of Variable Receptive Fields". *Ninth European Meeting on Cybernetics and Systems Research. Viena. 1988.*
- [38] Muñoz, J.A., Candela, S.: "On Computing Structures for Visual Processing". *7th Intern. Congress of Cibernetics and Systems, AFCET. Londres. 1987.*
- [39] Mc Culloch, W.S., Pitts, W. "A Logical Calculus of the ideas Immanent in Nervous Activity". *Bull. Math. Biophysics, 5, pp 115-133; 1943.*
- [40] Blum, M. "Properties of a Neuron with many Inputs" in *Principles of Self Organization. (H. von Foester and G.Gopf, eds). McMillan. (Pergamon). New York. 1961.*
- [41] Santana, O., De la Nuez, I. y Fortes, J.: *A camera Computer system from object isolation and recognition. Congreso de Cibernética y Sistemas Aplicados. Acapulco. 1980.*
- [42] Santana, O., De la Nuez, I. y Fortes, J.: *A camera computer system from object isolation and recognition. Congreso de Cibernética y Sistemas Aplicados. Acapulco. 1980.*