

# Una Propuesta de Instalación basada en Habilidades de Visión por Computador

Y. Rodríguez Domínguez, M. Castrillón Santana, O. Déniz  
Suárez, C. Guerra Artal y M. Hernández Tejera

IUSIANI - Edif. Ctral. del Parque Científico Tecnológico

Univ. de Las Palmas de Gran Canaria

35017 - Gran Canaria

mcastrillon@iusiani.ulpgc.es

## Resumen

Se propone en este trabajo la utilización de habilidades de detección de personas basadas en Visión por Ordenador para su integración en instalaciones artísticas de vídeo, explorando someramente sus capacidades, y mostrando una propuesta preliminar.

## 1. Motivación

A lo largo de la historia del arte, diversas disciplinas han hecho uso de la representación bidimensional para mostrar sus propuestas creativas. El ordenador es en la actualidad una herramienta de enorme potencial para el arte visual [12], tanto en el marco de la imagen estática, como en el contexto donde el factor tiempo se introduce en el proceso expresivo.

Desde el punto de vista de la imagen estática, una imagen es una matriz de píxeles los cuales básicamente representan un punto en un espacio de muy alta dimensionalidad. De forma análoga a otras técnicas como por ejemplo la pintura, permite representar elementos muy dispares, ver Figura 1. La tecnología digital en este contexto, presenta la singularidad de la no existencia de un original único, el arte digital permite disponer del original en cualquier parte, éste es copiable hasta la saciedad sin pérdida.

Por otro lado, nuevos ámbitos tecnológicos han ido abriendo capacidades y posibilidades

expresivas. Las tecnologías del vídeo digital y la animación introducen el factor tiempo en el proceso expresivo, la mutabilidad, la fugacidad y la narrativa temporal. La introducción de la interactividad a través del uso de las tecnologías de visión por computador aporta un nuevo canal expresivo y unas posibilidades para la generación de sensaciones a través de los conceptos de obra viva e interactiva, tal y como ya describiera Krueger en su concepto de Realidad Artificial [8].

La obra se puede convertir así en única y cambiante, reactiva a la interacción con su entorno en cada momento. Recupera el concepto de exclusividad, siendo además posible registrar la *vida* de la instalación. Este enfoque se relaciona de forma clara con el concepto de instalación manejado en el mundo artístico *una obra es instalación si dialoga con el espacio que la circunda* [6].



Figura 1: Somero ejemplo de la variabilidad del espacio de imágenes posibles.

Nuestra motivación es la de hacer uso de capacidades actuales de Visión por Computador para su integración en instalaciones artísticas. No cabe duda de que nuestra experiencia se relaciona básicamente con el mundo tecnológico, no es nuestro objetivo por tanto presentar una obra de creación sino mostrar las posibilidades interactivas que la visión artificial puede introducir en el arte.

## 2. Descripción del prototipo

El prototipo diseñado y utilizado para las pruebas preliminares hace uso del bucle tradicional de percepción-acción, ver Figura 2. En base a este esquema, como primera aproximación hemos restringido por un lado las percepciones del ordenador, a las imágenes adquiridas por medio de una cámara web estándar localizada en la sala donde acceden los visitantes; y por otro, las acciones a la alteración de una imagen proyectada en una pantalla que los visitantes de la instalación podrán observar.

Los dos submódulos considerados de percepción y acción, respectivamente la detección de personas en la sala, y la producción visual alterada por dicha presencia humana, se describen brevemente en las siguientes secciones.



Figura 2: Esquema del bucle percepción acción tradicional.

### 2.1. Detección de individuos

La detección de individuos debe ser una habilidad básica a incluir en cualquier Interfaz Basada en Visión [13] con el objetivo de proporcio-

nar al ordenador un mecanismo para percibir al usuario dentro de un contexto de Interacción Hombre-Máquina.

Son numerosas las soluciones propuestas para detectar personas en la literatura, teniendo en cuenta distintos elementos de la anatomía humana: la cara [5, 16], la cabeza [2], el cuerpo [15] o simplemente las piernas [10], e incluso la piel humana [7]. En años recientes la cara humana ha recibido gran atención en este marco dada su enorme influencia en el proceso de comunicación humana. Tal es así, que el problema ha sido repetidamente tratado, existiendo interesantes propuestas [11, 14], que sin embargo restringen enormemente las dimensiones aceptables de la cara, y en su mayoría se limitan a poses frontales.

La detección de personas ha sido anteriormente integrada en el proceso de producción creativa. Sin embargo tanto desde la utilización de la silueta realizada por Krueger [8], hasta desarrollos más recientes que hacen uso de técnicas de detección de movimiento o color para el módulo de percepción [1], todas las aplicaciones asumen distintas restricciones para poder realizar una detección robusta en tiempo real. La aproximación que hemos utilizado para este fin combina varias pistas en el tiempo obteniendo un sistema resultante más robusto y veloz, que tiene en el peor de los casos, un rendimiento análogo al, en estos días muy referenciado, detector de Viola-Jones [14]. Y todo ello sin restringirse a contextos donde por ejemplo el movimiento sea necesario.

El esquema de detección descrito muy brevemente debido a limitaciones de espacio, para más detalles consultar [3], proporciona una colección de hilos de detección durante la sesión interactiva,  $SI = \{hdt_1, hdt_2, \dots, hdt_n\}$ . Cada uno de estos hilos contiene información sobre las distintas detecciones que tienen lugar de forma consecutiva, y que han sido relacionadas por el sistema de detección en términos de posición, tamaño y apariencia.

La primera detección de un individuo se realiza por medio de detectores basados en el esquema de Viola-Jones [14], de forma oportunista cuando la vista sea frontal o los hombros y la cabeza sean visibles. Cada cara se

describe a partir de ese momento en términos de su última posición, tamaño, color, apariencia tanto global como de sus ojos, y su persistencia. Estas características dirigen a continuación la búsqueda en zonas próximas a la detección previa, aplicando siempre en primer lugar las técnicas más económicas hasta que una de ellas detecte el rostro:

- Seguimiento ocular: Un esquema rápido de seguimiento [4] busca los ojos.
- Detector facial: El detector facial de Viola-Jones [14].
- Detector basado en el contexto: Sólo cuando fallan las anteriores, buscamos la cabeza basada en el contexto [9].
- Color piel: Localizamos zonas con el color de la cara.
- Seguimiento facial: Si todo falla aplicamos un esquema de seguimiento para toda la cara [4].

Aplicado el detector a 70 sesiones de interacción con un total aproximado de 27000 imágenes requirió un tiempo promedio de 45 milisegundos, detectando un 99% de las caras con un 6% de error.

## 2.2. Producción visual

Este módulo genera efectos visuales en función de los datos proporcionados por el detector de caras. Se compone de dos submódulos principales: 1) proceso de datos, encargado de elaborar a partir de los datos de detección un vector global que representa las caras visibles en cada instante, y 2) visualización, que produce un efecto visual distinto para cada cara visible en el fotograma.

La asociación entre la cara detectada y el efecto visualizado se realiza en función de la etiqueta única que el detector asignó a cada hilo de detección, *hdt*, en su primera detección. El aspecto visual de cada efecto se modifica en función de los parámetros de cada cara, siendo éstos interpretados de forma diferente para cada uno de los efectos considerados.

En esta primera propuesta, la configuración de cada efecto utiliza algunos de los parámetros característicos de la cara (posición y tamaño, posición de cada ojo, color promedio, persistencia, etc.), así como otros parámetros independientes (funciones de ruido, valores aleatorios, etc.).

El catálogo de efectos disponibles para la versión actual se compone de: 1) un sistema de partículas cuya posición se obtiene transformando la posición de la cara a la que se asocia, y cuya densidad depende de la persistencia de la detección de ésta; 2) una matriz de marcos concéntricos cuyo tamaño depende de la posición de los ojos, y su cardinalidad depende de la proximidad de la cara al objetivo de la cámara; y 3) una nube de destellos cuyo tamaño depende del ojo derecho y que se mueven según una función Browniana.

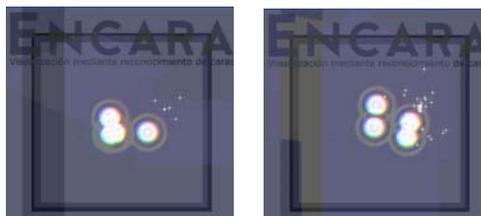


Figura 3: Dos instantes durante la visualización.

## 3. Conclusión

La integración de la Visión por Computador en el arte no es una novedad, si bien en este trabajo hemos querido mostrar las posibilidades que el uso de la percepción realizada por un ordenador puede aportar al desarrollo de instalaciones artísticas, haciendo uso de un sistema más robusto para la detección de personas que aquellos utilizados por desarrollos creativos previos. Los resultados obtenidos cumplen además la restricción del tiempo real para resoluciones estándar de cámaras web, sin asumir restricciones sobre el contexto de aplicación. El uso de esta herramienta permite partir de conceptos cognitivos de mayor

nivel que los proporcionados por técnicas de Visión de bajo nivel.

El prototipo inicial descrito se limita a detectar personas y almacenar sus parámetros, que son posteriormente utilizados para producir una imagen. Actualmente estamos analizando la ampliación de capacidades de visualización añadiendo capacidades de memoria, deliberativas, de representación tridimensional, y otras modalidades sensoriales como, en primer término, las auditivas.

## Referencias

- [1] Jeffrey E. Boyd, G. Hushlak, and C. J. Jacob. Swarmart: Interactive art swarm intelligence. In *Proceedings of the ACM Multimedia*, October 2004.
- [2] Marco La Cascia and Stan Sclaroff. Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3d models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(4):322–336, April 2000.
- [3] M. Castrillón Santana, O. Déniz Suárez, M. Hernández Tejera, and C. Guerra Artal. Real-time detection of faces in video streams. In *2nd Workshop on Face Processing in Video*, Victoria, Canada, May 2005.
- [4] Cayetano Guerra Artal. *Contribuciones al seguimiento visual precategórico*. PhD thesis, Universidad de Las Palmas de Gran Canaria, Octubre 2002.
- [5] Erik Hjelmas and Boon Kee Low. Face detection: A survey. *Computer Vision and Image Understanding*, 83(3):236–274, 2001.
- [6] José Iges. El espacio. el tiempo en la mirada del sonido. Catalogo de exposición. Kulturanea. España, 1999.
- [7] Michael J. Jones and James M. Rehg. Statistical color models with application to skin detection. Technical Report Series CRL 98/11, Cambridge Research Laboratory, December 1998.
- [8] Myron W. Krueger, Thomas Gionfriddo, and Katrin Hinrichsen. Videoplace – an artificial reality. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 35–40, May 1985.
- [9] Hannes Kruppa, Modesto Castrillón Santana, and Bernt Schiele. Fast and robust face finding via local context. In *Joint IEEE Internacional Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS)*, pages 157–164, October 2003.
- [10] C. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In *Proceedings of the International Conference on Computer Vision*, pages 555–562, 1998.
- [11] Henry Schneiderman and Takeo Kanade. A statistical method for 3d object detection applied to faces and cars. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1746–1759, 2000.
- [12] Anne Morgan Spalter. *The Computer in the Visual Arts*. Addison-Wesley Professional; 1st edition (1999), 1999.
- [13] M. Turk. Computer vision in the interface. *Communications of the ACM*, 47(1):61–67, January 2004.
- [14] Paul Viola and Michael J. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition*, pages 511–518, 2001.
- [15] Paul Viola, Michael J. Jones, and Daniel Snow. Detecting pedestrians using patterns of motion and appearance. In *Proc. of the International Conference on Computer Vision*, volume 2, pages 734–741, October 2003.
- [16] Ming-Hsuan Yang, David Kriegman, and Narendra Ahuja. Detecting faces in images: A survey. *Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.