



Imago animi vultus est [1]

Modesto Castrillón Santana

e-mail: mcastrillon@iusiani.ulpgc.es

página web: <http://mozart.dis.ulpgc.es/Gias/modesto.html>

Oscar Déniz Suárez

e-mail: odeniz@dis.ulpgc.es

página web: <http://mozart.dis.ulpgc.es/Gias/oscar/oscar.html>

Javier Lorenzo Navarro

e-mail: jlorenzo@iusiani.ulpgc.es

página web: <http://mozart.dis.ulpgc.es/~javier>

Luis Antón Canalís

e-mail: lanton@iusiani.ulpgc.es

F. Mario Hernández Tejera

e-mail: mhernandez@dis.ulpgc.es

Departamento de Informática y Sistemas
Universidad de Las Palmas de Gran Canaria

Pinche sobre una fórmula para ampliarla. Vuelva a pinchar sobre ella para reducirla, o pinche manteniendo pulsada la tecla [shift] para reducir todas las que permanezcan ampliadas.

Introducción

Podemos discutir el título, un extracto de una cita de **Cicerón**, pero resulta evidente el valor que los humanos damos al rostro. Es difícil no ver un rostro en la portada de un periódico o de una revista, y son numerosos los artistas que han jugado y juegan con él en sus obras. Un ejemplo de esto último lo tenemos en la **Figura 1**. No es ésta una característica exclusiva de nuestra sociedad occidental sino, más bien, del género humano en general. Todas las culturas humanas cuentan con múltiples evidencias del interés que este conjunto de rasgos identificativos tienen entre nosotros. La importancia universal del rostro queda patente en pruebas variadas, como por ejemplo las caras de la Isla de Pascua, símbolos de poder que asustaron al viajero del siglo XIX Pierre Loti [**McN00**], que estaría hipotéticamente justificada por una razón evolutiva: un bebé que mira a sus padres facilitaría que estos lo hicieran algo suyo, y en definitiva le prestarán más atenciones, lo cual iría en beneficio de su supervivencia.

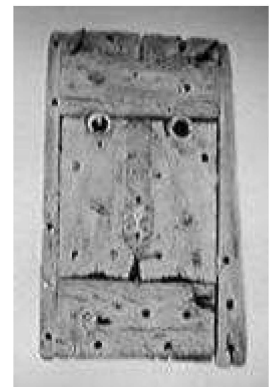


Figura 1. Autorretrato. César Manrique, 1975.



Figura 2. Bill Clinton y Al Gore
[Fuente: SiP96].

Sin embargo, no es el rostro la única fuente de información para identificar y describir a un individuo. Por mostrar un ejemplo, estudios sobre el sistema visual humano han sugerido una escasa habilidad para reconocer los rostros de las personas pertenecientes a una etnia no habitualmente presente en nuestro entorno cotidiano. Es el denominado *efecto de otra raza* [**BY98**]. Según este efecto, los habitantes de países occidentales haríamos uso de elementos del contexto, como por ejemplo la silueta de la cabeza, mientras que en países orientales, tradicionalmente con mayor homogeneidad, serían los elementos internos al rostro los más empleados para reconocer a una persona. Podemos observar este curioso efecto en la **Figura 2**, extraída de [**SiP96**], donde una primera observación nos muestra dos identidades, pero si la analizamos con más detenimiento descubrimos que ambas caras son idénticas. Podemos evidenciar esta habilidad de extracción de información del contexto, por ejemplo, en ciertos cuadros de **August Macke** (ver **Figura 3**) donde, a pesar de estar ausentes los elementos faciales, interpretamos fácilmente el género de quienes en ellos aparecen.

Si bien estas habilidades evidentes en los humanos siguen siendo tema de estudio en Psicología, la Inteligencia Artificial, y en particular la Visión por Computador, han

venido prestando atención a las posibilidades que el análisis automático del rostro puede tener en el campo de la interacción hombre máquina. En este contexto, en la literatura especializada se han descrito, durante los últimos años, aplicaciones dentro del mundo de la interacción, la seguridad, el entretenimiento, etc.

Desde un punto de vista de automatización del proceso humano de reconocimiento, identificación o descripción de caras, con el fin de construir sistemas capaces de replicar esta habilidad, un paso previo al análisis de cualquier rostro en un escenario concreto es el de la detección de la presencia del mismo.



Figura 3. Encuentros. August Macke, 1913.

Detección del rostro

El precio de las cámaras como dispositivo de adquisición de imágenes ha experimentado un marcado descenso en los últimos años, lo que ha posibilitado una enorme expansión de un amplio arsenal tecnológico de bajo coste en multitud de áreas de aplicación, incluida la informática de consumo en los hogares. Es por ello que hoy la disponibilidad de imágenes y vídeos es muy amplia. Esta circunstancia se ha dado a pesar de que los ordenadores actuales se distinguen por ser mayoritariamente *sordos, mudos y ciegos*, es decir, muy lejanos de los canales habituales de comunicación humana: la comunicación oral y la visual. Concretamente, la disponibilidad de estas tecnologías permite potencialmente utilizar las cámaras como entrada de datos para el *sentido de la vista* de un ordenador/máquina.



Figura 4 ^{a b}. Conjuntos *Cara* y *No cara*.

Una cámara proporciona imágenes que representan de forma numérica una proyección de la realidad. Estas imágenes pueden ser *estáticas*, es decir, estar aisladas en el tiempo, como una fotografía fija; o formar parte de una secuencia de imágenes que habitualmente conocemos como *vídeo*. Una imagen se compone de un número de puntos denominados *píxeles*, que se corresponden con las unidades sensibles a la luz en las cámaras. Cada uno de los píxeles posee uno o varios valores asociados al mismo, denominados *canales* (por ejemplo, una imagen de grises dispone de un único canal, que se correspondería con el nivel de iluminación de cada píxel, mientras que una imagen RGB, siglas en inglés para *rojo, verde y azul*, de tres, que se corresponderían con el nivel respectivo de cada uno de los tres colores anteriores para cada píxel de la imagen). Formalmente, una imagen estará definida por un ancho, n , y un alto, m , es decir, por $n \times m$ píxeles, y representa un punto en un espacio $n \times m$ -dimensional, que denominaremos *espacio imagen*:

$$I \in I_{n \times m}$$

El problema que abordamos es la detección del rostro humano en una imagen. La cuestión que se plantea es la siguiente: ¿cómo podemos saber si la imagen o parte de ella contienen un rostro? El problema de detección de rostros se puede definir como *determinar cualquier rostro en la imagen, si lo hubiera, indicando su localización y tamaño* [HL01], [YKA02].

Considerando esta definición, para resolver el problema basta con dar una aplicación funcional que nos permita establecer una correspondencia entre el espacio imagen y el de los conceptos. En este caso pretendemos quedarnos con aquella partición del espacio que lo segmenta en dos categorías: la de las caras, que denominaremos conjunto *Cara*, y la de las no caras o conjunto *No cara*, representados de forma ilustrativa en la Figura 4. Así, siendo ρ una imagen:

$$f: \rho \rightarrow (0,1),$$

$$f(\rho) = \begin{cases} 1, & \rho \in \text{Cara} \\ 0, & \rho \in \text{No cara} \end{cases} \quad \forall \rho \in I_{s \times t}, \quad 1 \leq s \leq n, \quad 1 \leq t \leq m,$$

donde f es la aplicación funcional denominada *clasificador*, ya que permite clasificar cada zona de la imagen candidata como *Cara* o *No cara*. Este clasificador deberá indicar, para una imagen dada, si contiene un rostro, varios o ninguno. Es este un problema difícil para el que sin embargo se han aportado interesantes soluciones durante los últimos años, tal y como se puede concluir de la consulta de las revisiones sobre el problema, [HL01], [YKA02]. Sin embargo, existen numerosas situaciones en las que el sistema de visión humano es mejor que cualquier sistema automático de detección. Es por ello que este problema sigue contando con la atención de los investigadores, con el objetivo de continuar mejorando las prestaciones de los sistemas desarrollados.

Las soluciones más exitosas abordan el problema desde el punto de vista de las técnicas de reconocimiento de formas. Se asume de esta manera que se dispone de conocimiento a priori del objeto o patrón a buscar, en nuestro caso la cara humana, lo que permite inferir la función del clasificador. Por simplicidad en cuanto al problema planteado, la literatura aborda principalmente el problema de detección de rostros frontales que no presenten una marcada inclinación lateral.

Desde el trabajo presentado en [SuP98], el conocimiento a priori de la cara no se basa únicamente en la observación de ejemplos de caras, sino que además incluye ejemplos de imágenes que podrían confundirse con ellas, aspecto que permite delimitar mejor la hipotética frontera entre ambas clases ilustrada en la Figura 4.

Debemos observar que para determinar si una imagen contiene una cara, varias o ninguna, el clasificador se aplica de forma exhaustiva por toda la imagen, analizando trozos rectangulares de distinto tamaño extraídos de la misma. La duración total del proceso dependerá del tamaño de la imagen, de la duración de cada clasificación, y del rango de tamaños diferentes contemplados durante el proceso. Éste, por tanto, puede ser, lógicamente, un proceso lento.

Teniendo en cuenta esta situación, al diseñar un clasificador se plantea un compromiso entre la precisión o bondad del mismo y su velocidad de procesamiento. Por ejemplo, si el detector de caras se va a utilizar integrado en un sistema que permita ejecutar interactivamente acciones con un ordenador, el proceso de detección no puede durar más de unas decenas de milisegundos, o sería impracticable desarrollar dicha aplicación. Es por ello que las aproximaciones más recientes han apostado por soluciones basadas en un esquema en cascada que intenta eliminar de forma rápida zonas muy obvias de ausencia de caras, y realizar procesos más elaborados y costosos sólo en aquellas zonas que presenten una mayor evidencia de contener un rostro. Esto se justifica por el hecho de que, típicamente, en la mayor parte de cada imagen a analizar no se contiene el objeto de interés.

Una analogía con este esquema en cascada podría ser una central de depuración de agua, donde en una serie de etapas se van eliminando las impurezas. Por ejemplo, el primer filtro eliminaría grandes objetos sólidos, un segundo filtro más delicado se encargaría de eliminar otros de menor tamaño, etc. Un clasificador en cascada aplica una cadena de *coladores* o *cedazos* que eliminan subimágenes de nuestra imagen original, por no superar el test de clasificación del clasificador o cedazo actual al no cumplir las características del patrón de un rostro humano, permitiendo así que un próximo cedazo refine aún más el resultado. En un proceso real de depuración, puede perderse parte del agua, bien porque se adhiere a los objetos, bien porque se evapora... En nuestro clasificador, dado que los filtros no son perfectos, se podrían rechazar, de forma incorrecta, algunas imágenes que contengan a nuestro objeto, la cara, y permitirán que pasen algunas impurezas, es decir, imágenes que no contengan una cara. Ahora bien, un diseño adecuado puede minimizar tanto la fracción de falsos negativos o imágenes que siendo rechazadas contienen caras, como de falsos positivos o fracción de imágenes que, no conteniendo caras, son identificadas como tales. Un clasificador de este tipo, tendrá unas tasas de detección y de rechazo que dependerán de las tasas individuales de cada filtro de la cascada, d_i y f_i , respectivamente, y del número de elementos clasificadores, K , en la cascada. La tasa resultante de detección, D , y la de falsas detecciones, F , para el detector serían:

$$D = \prod_{i=1}^K d_i, \quad F = \prod_{i=1}^K f_i.$$

A modo ilustrativo, un clasificador de 20 etapas con una tasa de detección para cada filtro del 99.9%, es decir, que sólo rechaza un 0.1% de las imágenes que contienen una cara, y una tasa de rechazo de falsos patrones del 50%, es decir, que elimina a la mitad de las imágenes que no contienen patrones, presentará una tasa conjunta esperada para las muestras correctas de $0.999^{20} \approx 0.98$, y para las incorrectas de $0.5^{20} \approx 0.9 \times 10^{-6}$. Esto muestra cómo, partiendo de unos clasificadores individuales con unas tasas de detección mediocres, pero computacionalmente ligeros y por tanto rápidos de respuesta, y de un esquema de conexión en cascada, se puede obtener un clasificador combinado mucho más robusto, con unas tasas mucho mejores.

Un clasificador basado en este planteamiento es el propuesto por Viola y Jones [VJ04], que utiliza un esquema de filtros de similar naturaleza para obtener resultados muy notables con rostros frontales. Sin embargo, no existe ninguna razón que obligue a limitar la naturaleza de los filtros. Por ejemplo, y una vez que dispongamos de la cara detectada, podríamos aplicar detección de elementos faciales como ojos, nariz o boca, para confirmar aún más que tenemos un rostro. Ciertamente, puede que perdamos alguna detección correcta pero, como veremos más adelante, si lo que nos interesa es un procesamiento rápido en vivo, el vídeo nos dará nuevas posibilidades para reducir la influencia de dichas pérdidas en el proceso.

Como hemos comentado, el detector de Viola-Jones proporciona resultados muy interesantes localizando rostros frontales, sin oclusiones ni inclinaciones laterales, ya que el proceso se basa en las características internas del rostro. ENCARA2 [CDHG07], el detector que hemos desarrollado, se concibe para mejorar el proceso de detección. Una primera idea ha sido la integración del contexto local en el proceso, ya que, si bien en ocasiones el rostro no se ve o está en una pose diferente a la frontal, nuestro cerebro sabe que está presente en la escena (y además extrae el género). Pinturas como *Hombre con bombín* de René Magritte (Figura 5) pueden darnos un ejemplo de lo que queremos expresar.





La detección del rostro basada en el contexto se ha llevado a la práctica utilizando clasificadores en cascada para obtener buena velocidad, y aunque los resultados no sean tan buenos como con rostros frontales, posibilitan la detección en situaciones de menor resolución, con inclinación lateral, y poses no limitadas a la frontal.

Figura 5. *Hombre con bombín*. René Magritte, 1964.

La combinación de varios detectores aumenta la tasa de detecciones pero tiene como desventaja un incremento de las falsas detecciones y del tiempo de procesamiento. La adición de nuevos filtros para detectar los elementos faciales permite obtener una solución que resulta más restrictiva, ya que eliminará falsas detecciones pero, a la vez, detectará menos buenos patrones. Sin embargo, esta pérdida de buenos patrones se minimiza a la hora de procesar vídeo, dado que este formato es adecuado para situaciones de interacción al proporcionar algunas ventajas:

- 1) se dispone de diversas oportunidades en cada segundo para detectar el rostro,
- 2) al existir *coherencia temporal* se pueden incorporar otras técnicas alternativas que permitan no realizar una búsqueda exhaustiva en la nueva imagen de la secuencia, sino que se pueda abordar de forma fiable y computacionalmente más ligera la relocalización del rostro en esa nueva imagen, y
- 3) el formato de vídeo permite además incluir la *persistencia* en el proceso de detección, de manera que una primera detección dudosa no se aceptaría, si bien varias detecciones algo dudosas, pero consecutivas en una misma localización, tendrían una mayor consideración.



Figura 6. Un modelo extraído en la imagen n , se busca en una zona coherente en la imagen $n+1$.

La coherencia temporal presente en una secuencia de vídeo nos permite acelerar el proceso de detección. La coherencia es consecuencia de la combinación de variabilidad de la imagen debida al movimiento de objetos físicos ante una cámara con las limitaciones de la física del movimiento de los cuerpos. Por ejemplo, el movimiento físicamente posible de una persona como la de la Figura 6, entre dos cuadros generados por una cámara típica que adquiere su imagen (con una frecuencia de 25 cuadros por segundo), está limitado a lo que la persona es capaz de moverse en los 40 milisegundos que transcurren entre ambas imágenes consecutivas. Esto introduce una restricción al posible desplazamiento de las partes de la imagen entre cuadros consecutivos. Debido a ello se puede limitar la búsqueda a una cierta zona de la imagen en las cercanías de la zona donde estaba presente en la imagen anterior. Esta sencilla idea permite afrontar con mayores garantías de velocidad el contexto de análisis de vídeo en tiempo real.

Existen múltiples aproximaciones para realizar dicha búsqueda en la nueva imagen. En nuestro caso, se ha aplicado una técnica sencilla y rápida que resuelve muchas situaciones de forma satisfactoria. La solución aplicada es la siguiente. Una vez que hemos detectado el rostro, extraemos un modelo de diversas zonas rectangulares que ofrecen contenido, como pueden ser la zona de los elementos faciales (ojos, nariz y boca). De forma ilustrativa en la Figura 6 se extrae uno de los ojos, conformando nuestro modelo, $M \in I_{w \times h}$, a buscar en la siguiente imagen pero solamente en una zona de interés, remarcada en la imagen derecha de la figura. La localización de esta zona de interés depende de la detección previa y de ella se extraen patrones candidatos de igual tamaño que el modelo a buscar. Una posible medida define una similitud basada en la suma de las diferencias absolutas de los píxeles entre el modelo y el candidato:

$$\arg \min D(u, v) = \sum_{i=1}^w \sum_{j=1}^h |M(i, j) - C(i + u, j + v)|$$

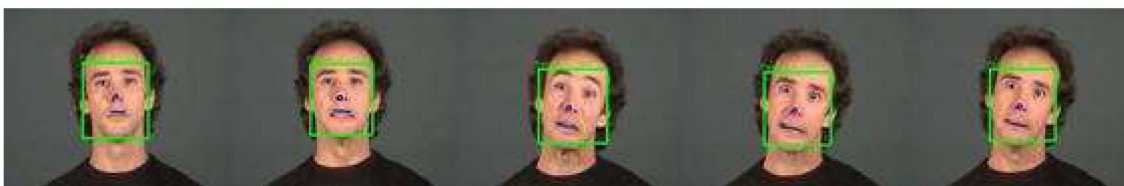


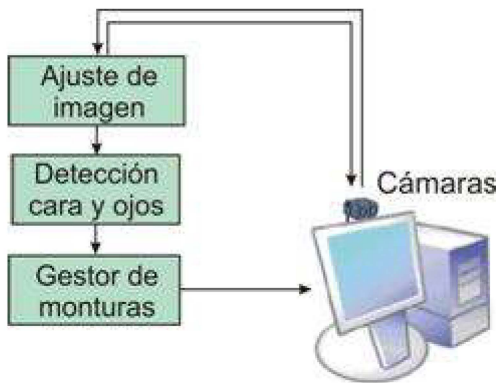
Figura 7. Ejemplos de detección en una secuencia de vídeo extraída de Dafex [BP04].

De forma ilustrativa se presentan resultados de detección para una secuencia en la Figura 7, donde, a pesar de aparecer el rostro inclinado de forma lateral, el detector aquí descrito es capaz de continuar su tarea de localización en situaciones donde el detector canónico de Viola-Jones no resuelve el problema.

Primera aplicación: selector de gafas

Son numerosos los contextos de aplicación de una solución como la descrita en la sección anterior, y nos gustaría remarcar que no todas ellas se restringen al ámbito de la seguridad y vigilancia. Un ejemplo de posible contexto de aplicación, dentro del sector comercial, es la selección de gafas en una óptica, facilitando al cliente la posibilidad de elegir, de forma virtual, entre una ingente gama de lentes, monturas, tintados, tratamientos polarizantes y fotocromáticos, etc.

Si bien en este ámbito existen algunos sistemas que ayudan en dicho proceso de selección, hay que valorar en ellos su coste, además de su atractivo y sencillez de uso para el cliente. Describiremos en esta sección un sistema de bajo coste, diseñado por los autores, que permite al cliente de una hipotética óptica virtual verse en una pantalla de vídeo, que hace el papel de un espejo para el cliente, mientras se prueba diferentes modelos de gafas. Éstas no son más que imágenes realistas de las gafas superpuestas sobre la cara del cliente, quien puede ajustar su colocación con la mano para comprobar cómo le sentarían.



El sistema, presentado de forma gráfica en la **Figura 8**, emplea dos cámaras. Incorpora, además del detector de caras y el gestor del modelo de gafas, un módulo de ajuste de la imagen que le permite mantenerla relativamente estable ante cambios en el entorno que afectan a las imágenes tomadas por las cámaras.

ENCARA2, el detector de rostros, no sólo permite la detección de caras, sino que adicionalmente, cuando el rostro está mirando hacia la cámara, provee la localización de los ojos.

Figura 8. Descripción del funcionamiento del selector de gafas.

El proceso con dos configuraciones de **zoom** diferentes, una de ellas con ángulo de visión mayor para localizar la cara, mientras que la otra se cierra sobre el rostro para localizar los ojos con más precisión y estabilidad.

La imagen de la montura elegida se superpone en la imagen de mayor ángulo de visión y se proyecta en el monitor que actúa de espejo para permitir al cliente observarse de forma completa. Sin embargo, como hemos comentado, los ojos se van a buscar también en la cámara de ángulo más cerrado dado que al ofrecer ésta mayor resolución del rostro la detección puede hacerse más precisa. Puesto que ambas cámaras proporcionan imágenes en escalas diferentes, se arbitra un mecanismo de reescalado entre ambas imágenes que las pone en correspondencia.

Como hemos comentado anteriormente, una vez determinada la posición, orientación y tamaño de las gafas a visualizar, se superponen sobre la imagen para obtener el efecto de realidad aumentada que permita al cliente tener la sensación visual de verse en el espejo con unas gafas puestas. La técnica empleada para combinar la imagen real y el modelo virtual de gafas es conocida como *alpha blending* en el ámbito de los gráficos por ordenador. Esta técnica consiste básicamente en combinar dos imágenes de forma ponderada. La ponderación permite incluir efectos de transparencia para conseguir el efecto de realismo en la imagen final, tal y como ocurriría con unas gafas reales, donde en función de la opacidad de los cristales podremos ver con mayor claridad los ojos o no.

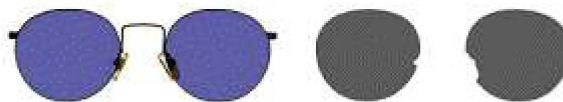
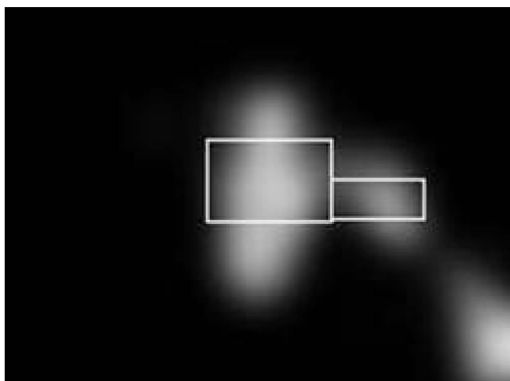


Figura 9. Modelo de gafas con sus dos imágenes.

De esta forma cada modelo de gafas está compuesto por dos imágenes: las gafas y el canal alfa, tal y como muestra el ejemplo de la **Figura 9**, siendo el canal alfa el encargado de indicar la ponderación para cada píxel de la imagen virtual, las gafas, en relación a la imagen real.

Todas estas acciones son responsabilidad del módulo de gestión de monturas, que se encargará de ajustar las dos imágenes correspondientes al modelo de montura activo en cada momento, tanto en lo referente a su tamaño como a su orientación sobre la base de las posiciones de los ojos suministradas por el módulo de detección de ojos.



Una vez llegados a este punto, disponemos de un sistema que al detectar una cara y sus ojos, es capaz de combinar un modelo de gafas virtual con la imagen de entrada. Para dar mayores posibilidades al sistema, hemos querido además integrar el reconocimiento de unos gestos realizados con alguna de las manos, con el objetivo de, por un lado, permitir la selección del modelo de montura activo, y por otro facilitar el ajuste de las gafas en su posicionamiento vertical.

La interacción con el sistema se flexibiliza y agiliza al incluir la detección de la mano del usuario. Esta capacidad se obtiene detectando las zonas de color similar a la del rostro detectado, es decir, las zonas de color piel. La **Figura 10** muestra el resultado de la detección de color piel para

Figura 10. Detección de zonas de color piel.

una escena en la que el individuo ha acercado su mano a la cara para ajustarse las gafas. Tras eliminar la zona de color piel coincidente con el rostro detectado, el resto de zonas con un tamaño coherente con el del

rostro serán consideradas candidatas a ser manos. Analizando la evolución de estas zonas dentro de la imagen a lo largo del tiempo, el sistema es capaz de interpretar una serie de gestos como órdenes de interacción.



Figura 11. Ajuste vertical de la posición de las gafas.

Por un lado, ciertas zonas de la imagen se activan como sensibles al movimiento, marcándose como tal en la imagen reproducida y permitiendo al usuario del sistema ordenar el cambio de modelo de gafas cuando éste mueva la mano en dichas zonas. Por otro lado, al acercar una mano a uno de los laterales de la cara se observa su movimiento en relación al rostro, y a partir del momento de contacto entre las regiones de la mano y la cara, las gafas se moverán de forma relativa hacia arriba o hacia abajo, acompañando al gesto realizado con la mano. Un ejemplo de uso de este gesto puede verse en la Figura 11. El lector puede acceder a mayor información sobre el sistema completo -incluyendo vídeos- en [Sel].

Segunda aplicación: detección de la sonrisa

El rostro humano nos proporciona información que va más allá de la identidad de la persona que tenemos ante nosotros: además podemos obtener conocimiento sobre su estado de ánimo y sensaciones, siendo ésta una importante característica social. Sin embargo, los ordenadores actuales evitan hacer uso de esta información.

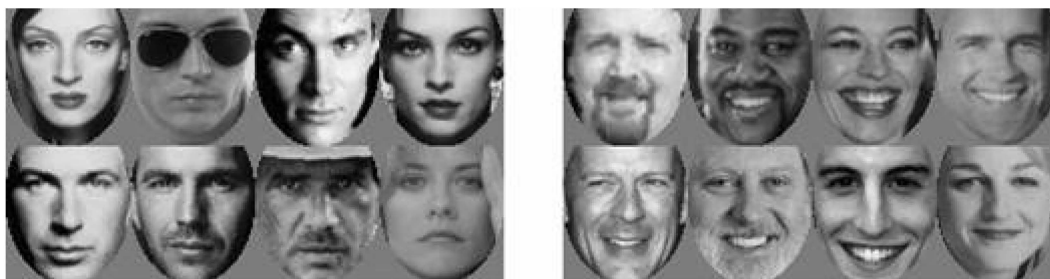


Figura 12. Muestras de los conjuntos de imágenes negativas (no sonrientes) y positivas (sonrientes).

Es probable que alguna vez hayamos utilizado los *emoticonos* (es decir, las combinaciones de caracteres ASCII que sintetizan expresiones faciales vistas con la cabeza inclinada, como por ejemplo “:-)” , que representa una sonrisa) para matizar un comentario en un correo electrónico o en una aplicación de mensajería instantánea, complementando así a nuestro interlocutor el sentido de las frases textuales. Ello se debe a que la carga semántica de la comunicación no sólo reside en la expresión textual, sino también en la gestual o la inflexiva. Para facilitar la utilización de este complemento gestual a la comunicación se ha desarrollado una pequeña aplicación de fácil integración en un sistema de mensajería instantánea, que tiene por función detectar la sonrisa del usuario e introducir de forma automática el emoticono correspondiente.

El proceso seguido para determinar la presencia de una sonrisa es el siguiente. En primer lugar hemos recopilado un conjunto de 7.000 imágenes faciales, las cuales, sobre la base de a la posición de los ojos, han sido rotadas y escaladas hasta alcanzar un tamaño normalizado. A continuación se han clasificado en dos conjuntos, según que el rostro presente una sonrisa (conjunto positivo) o no (conjunto negativo). Algunos ejemplos se muestran en la Figura 12.

Una vez en disposición de ambos conjuntos, hemos creado nuestro módulo de detección de sonrisa, basado en la solución de Viola-Jones, alimentado con ambos conjuntos. En ejecución, una vez que nuestro detector de caras, ENCARA2, detecta un rostro y sus ojos, procede a normalizar su tamaño teniendo en cuenta las posiciones de éstos, y en ese momento se aplica el detector de sonrisa. Una respuesta positiva del detector durante varias imágenes consecutivas permite asumir con alta fiabilidad que el usuario sonríe en ese preciso momento (Figura 13).

La información del detector permite, adicionalmente, detectar la presencia o ausencia del usuario ante el ordenador para, por ejemplo, activar el estado *Ausente* en un sistema de mensajería instantánea. El lector interesado puede obtener un demostrador de esta aplicación, para ser integrado en el sistema de mensajería instantánea MSN Messenger®, en este [enlace](#).

Referencias



Figura 13. Detector de sonrisa integrado en MSN Messenger®.

- [BP04] ^ A. Battocchi, F. Pianesi: Dafex: Un database di espressioni facciali dinamiche. *SLI-GSCP Workshop Comunicazione Parlata e Manifestazione delle Emozioni*, 2004.
- [BY98] ^ V. Bruce, A. Young: *The eye of the beholder*. Oxford University Press, 1998.
- [CDHG07] ^ M. Castrillón, O. Déniz, M. Hernández, C. Guerra: ENCARA2: Real-time detection of multiple faces at different resolutions in video streams. *Journal of Visual Communication and Image Representation* 18(2) (2007), 130-140.
- [HL01] a b E. Hjelmas, B.K. Low: Face detection: A survey. *Computer Vision and Image Understanding* 83(3) (2001), 236-274.
- [McN00] a b D. McNeill: *The Face: A Natural History*. Back Bay Books, 2000.
- [Sel] ^ *Selector de gafas*,
http://mozart.dis.ulpgc.es/Gias/oscar/SelectorGafas/index_sp.html.
- [SiP96] a b P. Sinha, T. Poggio: I think I know that face... *Nature* 384(6608) (1996), 384-404.
- [SuP98] ^ K.-K. Sung, T. Poggio: Example-based learning for view-based human face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 20(1) (1998), 39-51.
- [VJ04] ^ P. Viola, M.J. Jones: Robust real-time face detection. *International Journal of Computer Vision* 57(2) (2004), 151-173.
- [YKA02] a b M.-H. Yang, D. Kriegman, N. Ahuja: Detecting faces in images: A survey. *Transactions on Pattern Analysis and Machine Intelligence* 24(1) (2002), 34-58.

[1] *Imago animi vultus, indices oculi* [la cara es el reflejo del alma, y los ojos lo delatan]. Cicerón, *De oratore*, III, 59.



Sobre los autores

Los autores son miembros de la División de Inteligencia Artificial y Sistemas del Instituto Universitario de Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniería (IUSIANI) de la Universidad de Las Palmas de Gran Canaria. Sus líneas de trabajo están enmarcadas en la Investigación y el desarrollo en Interacción e Inteligencia Perceptual, en aspectos relacionados con la automatización de la interacción utilizando modalidades naturales de comunicación como la visual y la sonora y a su uso, tanto en la relación hombre máquina como en el desarrollo de soluciones autónomas para aplicaciones. Se trabaja en problemas de visión artificial, reconocimiento de formas y aprendizaje relacionados con la automatización de la detección y el reconocimiento visual de caras y gestos, la detección, localización y reconocimiento sonoro, la integración de las modalidades de interacción visual y sonora y la toma de decisión inteligente basada en esta información aferente. Se investiga en el desarrollo de agentes inteligentes que introduzcan estos elementos y su aplicación en los dominios de la robótica social y las interfaces de usuario avanzadas, que integren estos recursos o las interfaces adaptadas para usuarios con necesidades especiales: con disminución visual o auditiva o con movilidad o control reducidos de sus extremidades. Asimismo, se realiza investigación y desarrollo en soluciones para aplicaciones industriales y de seguridad. Un conjunto de aspectos importantes en los que se trabaja es en los relacionados con la integración de mecanismos de adaptación, ya sea mediante procesos de control como de aprendizaje basado en minería sobre flujos de datos perceptuales, tanto en procesos de conceptualización en modo lote (*batch*) como en aprendizaje incremental. Se dispone de un amplio bagaje de soluciones modulares desarrolladas y probadas



intensamente en seguimiento (*tracking*), detección, reconocimiento y aprendizaje, así como de una metodología para su integración en aplicaciones. Los miembros del grupo han publicado más de una treintena de artículos y más de cincuenta presentaciones en congresos y conferencias, han realizado más de una decena de proyectos de investigación financiados y son titulares de dos patentes, además de una solicitud más en trámite.