Published in XXXII Jornadas de Automática, Sevilla. 2011

CÁMARAS BASADAS EN TIEMPO DE VUELO. USO EN LA MEJORA DE MÉTODOS DE DETECCIÓN DE CARAS

J.R. Ruiz-Sarmiento, C. Galindo, J. González-Jiménez
Dpto. de Ingeniería de Sistemas y Automática, Universidad de Málaga, Campus de Teatinos, 29071 Málaga (España)
jotaraul@isa.uma.es, {cipriano, jgonzalez}@ctima.uma.es

Resumen

Este artículo presenta la utilización de cámaras basadas en Tiempo de Vuelo (TOF) para detectar la presencia de personas mediante búsqueda de caras. La ventaja principal de este tipo de cámaras es que proporcionan información de la escena tanto de intensidad como de rango (distancia). El algoritmo propuesto utiliza el método de Viola-Jones para detectar caras en la imagen de intensidad, y posteriormente reduce el alto número de falsos positivos que se obtienen mediante una batería de tests morfológicos basados en la información de rango. Las pruebas realizadas ponen de manifiesto la utilidad del algoritmo desarrollado en la detección de caras en escenarios no controlados. Se presentan también experiencias en las que se integra la cámara TOF en un robot móvil, ilustrando su utilidad para aplicaciones de robótica de servicios.

Palabras Clave: Cámara de Tiempo de Vuelo, Detección Facial, Visión por computador, Robótica Móvil.

1 INTRODUCCIÓN

La detección de rostros resulta de gran importancia en multitud de aplicaciones que implican interacción con personas, como por ejemplo en robótica de servicios o inteligencia ambiental. Un robot guía para un museo, por ejemplo, debería detectar eficazmente la presencia de visitantes para activar correctamente la reproducción de la presentación correspondiente. En el campo de inteligencia ambiental, la detección de personas también resulta relevante, como por ejemplo, en aplicaciones que se encargan de regular la temperatura de una habitación, operando de forma distinta en función de la presencia o no de personas. Para este tipo de situaciones, se podrían considerar dispositivos simples como detectores volumétricos. Sin embargo, estos dispositivos presentan ciertas limitaciones ya que se disparan al detectar cambios en la escena, sin discriminar si se ha producido por el movimiento de una persona o no.

En el campo de la Visión por Computador se han desarrollado, a lo largo de su historia, una gran diversidad de algoritmos de detección de caras ([2],[11],[12]). Sin embargo la eficacia obtenida en entornos no controlados, con iluminación variable, es aún pobre. Recientemente ha aparecido en el mercado un nuevo tipo de cámara que proporciona tanto imágenes de intensidad como imágenes de rango o distancia. Estas cámaras, denominadas cámaras de Tiempo de Vuelo (TOF), proporcionan información valiosa para la detección fiable de caras, pudiéndose utilizar no solo la información visual característica de una cara humana, como su color, forma 2D, etc., sino también la información tridimensional que representa la morfología típica de las caras, como por ejemplo el hecho de que el área de la nariz sobresalga del área de las mejillas.

En la literatura reciente, algunos trabajos han explorado el uso de cámaras TOF. La mayoría se han centrado en su caracterización [1], [13], y algunos estudian su aplicación a diferentes campos [4], [14]. En este artículo se presenta una técnica basada en una batería de tests morfológicos que mejora los resultados obtenidos en otros trabajos relacionados que también proponen el uso de estas cámaras en tareas de detección facial. Ejemplos de estos trabajos son [3], donde se utiliza una técnica basada en triángulos semejantes para la eliminación de falsos positivos, y [7] donde se emplean imágenes de rango para la detección facial. Se presentan experimentos realizados en escenarios no controlados en los que se obtienen una alta eficacia, reduciendo el número de falsos positivos que constituyen el punto débil de la mayoría de detectores. Se han realizado también experimentos en los que la cámara TOF se ha montado sobre un robot móvil [6] demostrando su aplicabilidad a la robótica de servicios.

2 MÉTODO PROPUESTO

El objetivo de este trabajo es desarrollar un método robusto de detección facial, empleando para ello una cámara de Tiempo de Vuelo. En este contexto, robusto se refiere a que el método produzca el menor número de falsos positivos posibles, manteniendo a la vez altos ratios de detecciones.

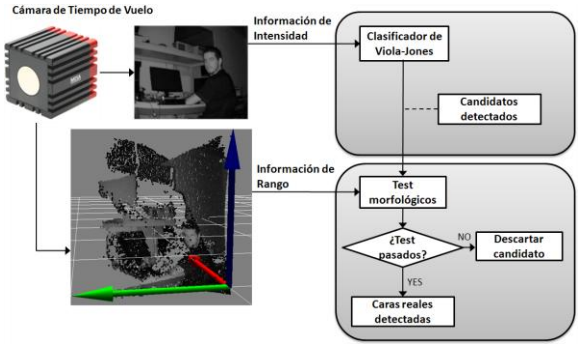

Figura 1. Esquema del proceso de dos etapas propuesto para detección facial.

El método propuesto está dividido en dos fases (ver figura 1). En la primera se aplica el clasificador de Viola-Jones [11] para la detección facial sobre la imagen de intensidad, obteniéndose un conjunto de regiones candidatas que presumiblemente contienen una cara. En la segunda fase, las regiones candidatas son sometidas a una batería de tests que estudian su morfología 3D operando sobre la imagen de rango. Esta segunda fase trata de descartar la mayor cantidad posible de los falsos positivos obtenidos en la primera.

2.1 Cámaras de Tiempo de Vuelo

En términos generales, una cámara de Tiempo de Vuelo es un dispositivo que provee datos de intensidad y de rango, es decir, información de distancia a los objetos percibidos. Más concretamente, la cámara de Tiempo de Vuelo considerada en este trabajo, fabricada por *Mesa Imaging* [8] y distribuida en España por Infaimon [5], trabaja emitiendo una onda modulada de forma continua empleando un *array* de leds infrarrojos (ver figura 2). Cuando la onda rebotada desde el objeto es captada por la cámara, sus celdas CCD/CMOS se excitan con una señal que exhibe un cierto desfase, el cual permite al dispositivo calcular la distancia y reflectividad del objetivo empleando correlación cruzada, hasta una distancia de 5 metros y con una precisión de ±10 cm.

A pesar de las grandes posibilidades que ofrecen las cámaras de Tiempo de Vuelo, éstas tienen aún una serie de limitaciones a tener en cuenta (más información en [13]):

- Baja resolución. La cámara empleada tiene una resolución de 176 *x* 144 píxeles.
- Inadecuadas para escenarios con un alto grado de dinamismo. Cuando una escena presenta objetos desplazándose rápidamente, las mediciones de distancia son propensas a presentar errores elevados.
- Baja precisión de las mediciones. El color y el tipo del material de los objetos afectan a las mediciones de distancia.

2.2 Clasificador de Viola-Jones

El clasificador de Viola-Jones aplica una cascada de test de complejidad creciente sobre la imagen de intensidad (ver figura 3). Las primeras etapas de la cascada son simples y descartan rápidamente regiones que no presentan las características generales de las caras humanas, por ejemplo, el área de los ojos ha de ser más oscura que el de la nariz. Las etapas siguientes son incrementalmente más selectivas y complejas, minimizando el ratio de falsos positivos a expensas de un mayor coste computacional.

Los test considerados en cada etapa son entrenados usando AdaBoost [15]. AdaBoost es un algoritmo de aprendizaje que escoge clasificadores débiles entre una familia de características simples o *Haar features*. Estos clasificadores son combinados en clasificadores fuertes, los cuales forman las etapas del clasificador de Viola-Jones. Más información sobre cómo se construyen estas etapas en [11].

Una vez entrenado, el clasificador es aplicado a una secuencia de vídeo. Sobre cada fotograma se desplaza una ventana de 20*x*20 píxeles, comprobando si estas subregiones superan todas las etapas de la cascada, lo que permite detectar caras en distintas localizaciones de la imagen. Además, esta ventana se escala hasta que alcanza el mismo tamaño que el fotograma, lográndose de esta manera detectar caras con diferentes tamaños o escalas. El tamaño mínimo de ventana usado por la implementación del

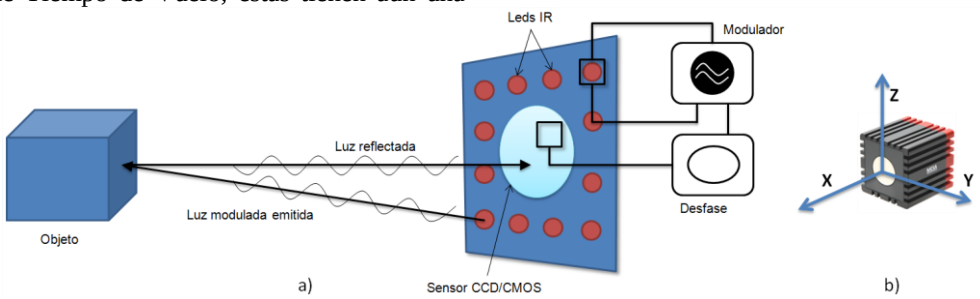

Figura 2. a) Esquema del funcionamiento de una cámara de Tiempo de Vuelo. b) Sistema de referencia establecido por la cámara usada.

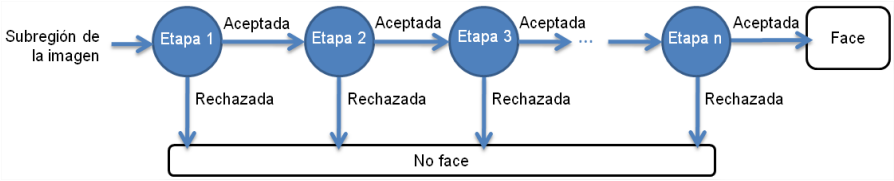

Figura 3. Estructura en cascada para la detección facial. Las etapas que requieren menos computación se sitúan al comienzo.

clasificador de Viola-Jones y la baja resolución de la cámara de Tiempo de Vuelo limitan la distancia a la que son detectables las caras a 2,5 metros.

Aunque este detector facial proporciona buenos resultados en comparación con otros métodos [10], el número de falsos positivos que produce es un serio inconveniente para muchas aplicaciones reales.

2.3 Tests morfológicos

Para afrontar el problema que suponen estas falsas detecciones se propone una segunda etapa donde se aplican tres tests morfológicos sobre la imagen de rango para confirmar o rechazar las regiones candidatas. Cada test comprueba si una característica morfológica está presente en la región candidata o no, identificando y eliminando distintos tipos de falsos positivos. La figura 4 muestra varios ejemplos de falsos positivos detectados como tales por dos de los test propuestos y descritos en las siguientes secciones.

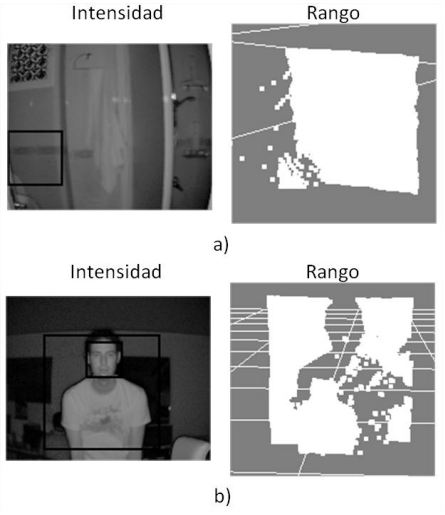

Figura 4. a) Falso positivo detectado por ser una región plana. b) Falso positivo por tener un tamaño no común para ser una cara a esa distancia.

Nótese que, aunque el método presentado en este trabajo se ha particularizado para el caso de la detección facial, el concepto es genérico para cualquier objeto detectable siempre y cuando los objetos de la misma clase compartan una morfología común. Así, bastaría con elegir el detector adecuado en la primera fase y adaptar los filtros morfológicos de la segunda a la morfología particular de dicha clase de objeto.

2.3.1 Test #1: Región plana

El primer test está basado en el hecho de que las caras humanas no son planas, por lo que presentan un cierto relieve que se ha de percibir en la imagen de rango. Un ejemplo de falso positivo de este tipo se muestra en la figura 4-a). En la implementación de este test se calcula la matriz de dispersión, C, de la posición espacial (x,y,z) de los píxeles que forman la región candidata. Los autovalores de la matriz C dan información acerca de la distribución espacial de los píxeles que forman la región. Concretamente, cuanto menor sea el menor autovalor, más plana será la región.

Un razonamiento similar podría aplicarse estudiando la desviación típica de las posiciones espaciales a lo largo del eje x (profundidad con respecto a la cámara), lo cual requiere menos cálculos. No obstante, se ha comprobado que este enfoque da problemas con regiones que presentan un cierto escorzo.

2.3.2 Test #2: Ratio de tamaño-distancia

Este test descarta candidatos que no cumplen con el tamaño esperado para una cara humana a una cierta distancia. La figura 4-b) muestra un ejemplo de falso positivo rechazado por este test. En la implementación se ha considerado que el tamaño normal de una cara humana, de media, es 209 cm^2, con una desviación típica de aproximadamente 60 cm^2. Estos datos han sido obtenidos analizando alrededor de 10.000 imágenes de caras.

2.3.3 Test #3: Estructuras faciales

El último test explota el hecho de que las caras humanas presentan una morfología común. Su

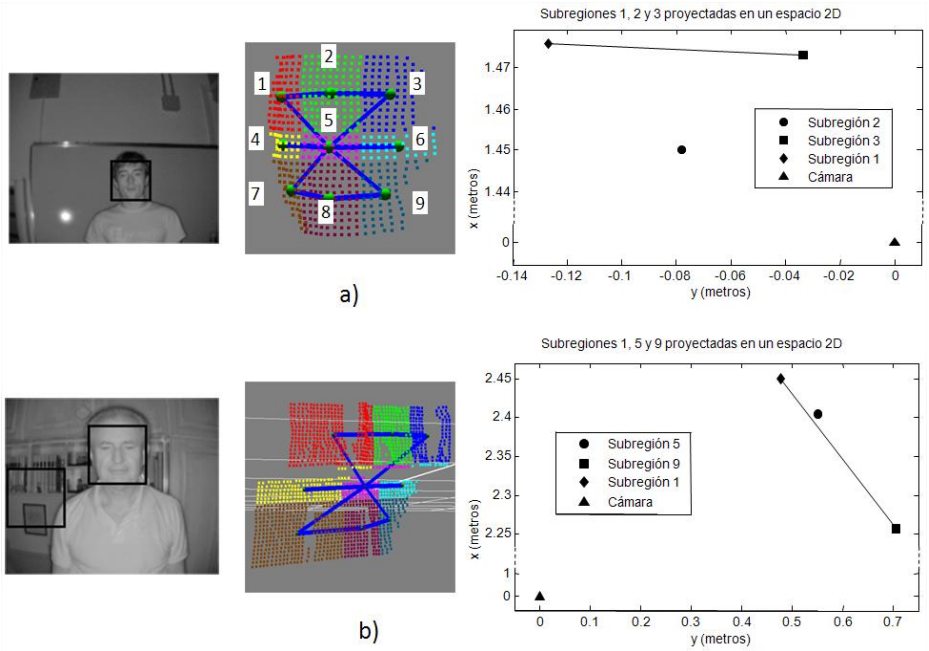

Figura 5. a) Izquierda: una región candidata. Centro: los centroides son representados con esferas, mientras las líneas representan las restricciones a comprobar. Derecha: proyección en el plano *y*-*x*. b) Izquierda: un falso positivo (detrás de la persona). Centro: Subdivisión de la región y las restricciones a verificar. Derecha: proyección de tres subregiones en el plano *y*-*x*. Se puede ver como el test no se satisface en este caso

funcionamiento es el siguiente: primero, se segmentan los candidatos usando un método de crecimiento de regiones para extraer la cara candidata del fondo, y el resultado es dividido en 9 subregiones, tal y como se muestra en la figura 5-a). La profundidad de estas subregiones con respecto a la cámara ha de satisfacer ciertas restricciones características de las caras humanas. Por ejemplo, la subregión que presumiblemente contiene la nariz, etiquetada como 5 en la figura 5-a), debe resaltar con respecto a las regiones laterales, etiquetadas como 4 y 6, correspondiente con los pómulos. Un ejemplo de falso positivo detectado por este test se muestra en la figura 5-b).

En la implementación de este test, la posición de cada subregión es considerada como el centroide de los píxeles que forman dicha subregión. Se realizan un total de cinco comprobaciones para verificar si las subregiones centrales de las tres filas y las dos diagonales resaltan con respecto a las subregiones laterales. Concretamente, se realizan las comprobaciones en el plano *y*-*x* de la siguiente manera:

Definamos $x = a \cdot y + b$ como la línea recta que une los centroides de las subregiones laterales. Siendo $P_c = (y_c, x_c)$ el centroide de la subregión central, se comprueba si la subregión central resalta con respecto a las laterales, o lo que es igual, la distancia $d = a \cdot y_c + b - x_c$ debe ser positiva.

3 EVALUACIÓN

Para probar la efectividad del método desarrollado se han realizado un total de 16 experimentos divididos en dos tipos de escenarios: *cámara fija*, donde la cámara permanece fija mientras hay personas moviéndose alrededor, y *cámara en movimiento*, donde se situó la cámara en un robot móvil [6], estando, de este modo, afectada por importantes cambios de iluminación. Un vídeo de ejemplo del método desarrollado funcionando en un escenario del segundo tipo puede verse en la siguiente dirección: *http://www.youtube.com/watch?v=GJE4A7R6LNs*

La implementación del clasificador en cascada de Viola-Jones escogida para detectar las regiones candidatas es la presente en la librería OpenCV, concretamente el clasificador ya entrenado *haarcascade_frontalface_alt2*. Aunque éste provee buenos resultados [9], se produce un considerable porcentaje de falsos positivos. En nuestros experimentos, de un total de 11.184 candidatos, se obtuvieron 685 (5.77%) falsas detecciones. Estos falsos positivos se identificaron por inspección visual.

Los test morfológicos propuestos han sido implementados en una aplicación C++ *multi-thread*, usando un computador con un procesador Intel® Core™ 2 Quad CPU Q6600 2.4GHz y 4Gb de memoria RAM, consiguiendo procesar hasta 14 fps. La arquitectura *multi-núcleo* que presentan los

Tabla 1. Falsos positivos rechazados por cada test de forma individual y la combinación de todos ellos con respecto al número total de falsos positivos producidos por el clasificador de Viola-Jones.

Escenarios	Fotogramas	Falsos positivos de la primera etapa	Test #1	Test #2	Test #3	Todos
Cámara fija	11230	122	60,66%	68,85%	53,27%	100%
Cámara en movimiento	27649	563	22,74%	53,46%	93,25%	98,03%

procesadores modernos nos permite, gracias a la implementación *multi-thread*, ejecutar simultáneamente todos los test, aunque también se podría haber adoptado una solución secuencial.

El uso de los test morfológicos reduce drásticamente el número de falsos positivos, tal y como puede verse en la tabla 1. Nótese que cada test por si sólo tiene un modesto porcentaje de eliminación de falsos positivos, pero cuando son combinados proporcionan excelentes resultados, eliminando la totalidad de falsos positivos en el caso de escenarios con la cámara fija, y sobre el 98% en escenarios con la cámara montada sobre el robot. En lo referente a la tasa de falsos negativos, esto es, caras que son erróneamente eliminadas por la segunda fase, la batería de test descarta sobre el 3% de caras reales en ambos escenarios, lo que no representa un inconveniente ya que trabajamos con una secuencia de vídeo.

4 CONCLUSIONES

En este trabajo se ha presentado un método de detección facial robusto que emplea cámaras de Tiempo de Vuelo para reducir radicalmente el número de falsas detecciones, manteniendo a la vez una tasa baja de falsos negativos. Los resultados obtenidos muestran el interés de usar no sólo información de intensidad, sino también información de rango para alcanzar el nivel de robustez demandado por las aplicaciones reales. Aunque actualmente las cámaras de Tiempo de Vuelo tienen un precio elevado, recientemente están emergiendo sensores similares (como por ejemplo *Kinect*) empleados como interfaces para la interactuación con videojuegos que presentan buenas prestaciones y un precio ajustado, lo que hace prever que su coste en el mercado bajará. En el futuro se planea integrar el método desarrollado en aplicaciones en las que la detección facial es una fase necesaria.

Referencias

[1] Andreas Kolb, Erhardt Barth, and Reinhard Koch. ToF Sensors: New Dimensions for Realism and Interactivity. In *CVPR 2008 Workshop on Time-of-Flight-based Computer Vision*, 2008.

[2] Cha Zhang and Zhengyou Zhang. A Survey of Recent Advances in Face Detection. Technical Report, MSR-TR-2010-66. June 2010.

[3] Dan Witzner Hansen, Rasmus Larsen, and Francois Lauze. Improving Face Detection with TOF Cameras. In *Proc. of the IEEE Int. Symposium on Signals, Circuits & Systems* (ISSCS), vol. 1, pages 225-228, Iasi, Romania, 2007.

[4] David Droeschel, Stefan May, Dirk Holz, Paul Ploeger, and Sven Behnke. Robust Ego-Motion Estimation with ToF Cameras. In *Procc of the European Conf. on Mobile Robots (ECMR'09)*, Croatia 2009.

[5] Infaimon website: http://www.infaimon.com/.

[6] J. Gonzalez, C. Galindo, J.L. Blanco, J.A. Fernandez-Madrigal, V. Arevalo, F.A. Moreno. SANCHO, a Fair Host Robot. A Description. *IEEE Int. Conf. on Mechatronics (ICM'09)*, 2009.

[7] Martin Bhme, Martin Haker, Kolja Riemer, Thomas Martinetz, and Erhardt Barth. Face Detection Using a Time-of-Flight Camera. In *Lecture Notes in Computer Science*, vol. 5742, pages 167-176, 2009.

[8] Mesa Imagin website: http://www.mesa-imaging.ch/.

[9] M. Castrillon-Santana, O. Deniz-Suarez, L. Anton-Canalis and J. Lorenzo-Navarro. Face and Facial Feature Detection Evaluation. *III Int. Conf. on Computer Vision Theory and Applications*, 2008.

[10] Nikolay Degtyarev, and Oleg Seredin. Comparative testing of face detection algorithms. *Proc. of the 4th Int. Conf. on Image and signal processing, ICISP'10*, Canada, 2010.

[11] Paul Viola and Michael Jones. Robust real-time face detection. In Porc. *Int. Conf. on Computer Vision*, 57(2):137-150, 2004.

[12] Rainer Lienhart and Jochen Maydt. An extended set of haar-like features for rapid object detection. In *IEEE ICIP* 2002, vol. 1, pp. 900-903, 2002.

[13] Robert Lange and Peter Seitz. Solid-State Time-of-Flight Range Camera. In *IEEE Journal of Quantum Electronics*, Vol 37, No. 3, March 2001.

[14] Sergi Foix, Guillem Alenya, Juan Andrade-Cetto and Carme Torras. Object Modeling using a ToF Camera under an Uncertainty Reduction Approach. *IEEE Int. Conf on Robotics An Automation (ICRA'10)*, Anchorage, Alaska, 2010.

[15] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Computational Learning Theory: Eurocolt' 95*,pages 2337. Springer-Verlag, 1995.