

# Detección y seguimiento de personas usando visión estéreo

Rafael Muñoz-Salinas, Eugenio Aguirre  
Miguel García-Silvente, Antonio González

Dpto. de Ciencias de la Computación e Inteligencia Artificial

E.T.S. Ingeniería Informática

Universidad de Granada - 18071 Granada

{salinas, eaguirre, mgs, A.Gonzalez}@decsai.ugr.es

## Resumen

En este trabajo presentamos un agente para la detección y seguimiento de personas usando visión estéreo. El agente emplea visión activa para realizar el seguimiento utilizando una cabeza robótica sobre la que está montado el sistema de visión. Inicialmente, se crea un mapa de altura del entorno que recoge las características inmóviles del mismo. Este mapa será utilizado para detectar los objetos móviles del entorno y buscar entre ellos a posibles personas utilizando un detector de caras. Una vez que se ha detectado a una persona, el agente es capaz de seguirla utilizando la cabeza robótica que permite girar el sistema estéreo. Para hacer más robusto el seguimiento hemos utilizado el filtro de Kalman. El agente mantiene localizada a la persona en todo momento enmarcando su cabeza y brazos en la imagen. Esta tarea podrá ser utilizada por otros agentes que necesiten analizar los gestos y expresiones de posibles usuarios de aplicaciones.

## 1. Introducción

Un aspecto fundamental en la creación de ciertos sistemas inteligentes es detectar la presencia humana. El tópico *interacción hombre-máquina* ha comenzado a atraer una gran atención en la última década. El objetivo es poder crear interfaces más inteligentes capaces de extraer información sobre el contexto o las acciones a realizar mediante una interacción natural con el usuario, por ejemplo por medio del uso de gestos o la voz.

Un aspecto fundamental en este sentido es la

detección y seguimiento de personas, existiendo una gran cantidad de literatura al respecto [3, 9, 10]. Las técnicas para realizar la localización suelen basarse en la integración de múltiples fuentes de información: color de la piel, detectores de caras, análisis del movimiento, etc.

Si bien la detección y seguimiento de personas con una única cámara es un tema ampliamente tratado, el uso de la tecnología estéreo para este propósito concentra ahora un importante interés. La aparición de hardware comercial para el cálculo estereoscópico, así como el abaratamiento de este tipo de sistemas, los convierte en un atractivo sensor con el que desarrollar sistemas inteligentes. El uso de visión estéreo dota de un grado más de información que es de gran utilidad en la tarea de detección e interacción hombre-máquina. Por un lado, la información de disparidad es más invariable a cambios de iluminación que la de una sola cámara, siendo un elemento muy ventajoso para la estimación del entorno (*background*). Además, la posibilidad de conocer la distancia al sujeto puede ser de gran ayuda para su seguimiento así como para un mejor análisis de sus gestos.

En este trabajo presentamos un agente para la detección y seguimiento de personas usando visión estéreo. El agente emplea visión activa para realizar el seguimiento utilizando una cabeza robótica sobre la que está montado el sistema de visión. Este agente servirá de base para el trabajo de otros agentes encargados de tareas como analizar los gestos y expresiones de posibles usuarios de aplicaciones. El método de detección se basa en la creación inicial

de un mapa de altura del entorno. Este mapa contiene información sobre la estructura del entorno y puede ser creado incluso en presencia de personas moviéndose por él. El mapa será utilizado para detectar los objetos móviles. Éstos son posibles candidatos a personas que serán encontradas utilizando un detector de caras. A diferencia de la mayoría de los trabajos que sólo modelan el entorno para una imagen fija, nuestro mapa del entorno modela toda la región visible por el sistema estéreo. Una vez que se ha detectado a una persona, la cabeza robótica permite girar el sistema estéreo para seguirla por el entorno. Para hacer más robusto el seguimiento hemos utilizado el filtro de Kalman. El método de seguimiento está diseñado para mantener visibles, siempre que sea posible, la cabeza y brazos de la persona y facilitar así el estudio de sus gestos.

### 1.1. Trabajos relacionados

Entre los trabajos más destacados sobre detección y seguimiento de personas usando visión estéreo encontramos el trabajo de Darrel et al [1]. Este trabajo presenta un sistema de display interactivo capaz de detectar y seguir múltiples personas. La detección de las personas se basa en integrar la información de un detector de piel, un detector de caras y el mapa de disparidad de la escena. Por un lado se detectan objetos (*blobs*) independientes en la imagen de disparidad que serán candidatos a personas. Por otro lado, se analiza el color de la imagen para detectar aquellas zonas que pudieran pertenecer a piel y se aplica un detector de caras. Estas tres informaciones son fusionadas para detectar a las personas que se encuentran visibles. Para realizar el seguimiento se utiliza información sobre el pelo, la ropa y la historia pasada de las personas localizadas. De esta manera los sujetos pueden ser reconocidos aunque desaparezcan de la imagen por un tiempo.

En [4] se muestra un sistema para la detección y seguimiento de personas para la interacción en entornos virtuales. El sistema permite a un usuario navegar por un entorno virtual simplemente andando por una habitación colocándose unas gafas de realidad virtual. En este trabajo la detección de la cara es un punto crucial que se resuelve utilizando el detector

de caras propuesto por Viola y Jones [12]. Una vez localizado el sujeto se emplea un histograma del color de la cara y del pecho del sujeto utilizando un filtro de partículas para estimar la posición de la cabeza. La información estéreo es utilizada para conocer la posición del sujeto en la habitación y poder así determinar su posición en un entorno virtual. En este trabajo, el procesamiento estéreo se realiza utilizando la información de varias cámaras situadas en distintos puntos del entorno.

En [6] se presenta un método para la localización y seguimiento de personas en imágenes estéreo utilizando mapas de ocupación. Antes de realizar la detección de personas se crea una imagen del entorno mediante un sofisticado método de análisis de imágenes. Una vez que se tiene la imagen del entorno se separan fácilmente los objetos que no pertenecen a él y se crea tanto un mapa de ocupación como un mapa de altura de la zona visible. La información de ambos mapas es fusionada para detectar personas utilizando heurísticas simples. El seguimiento de las mismas se realiza utilizando el filtro de Kalman combinado con plantillas deformables. En este trabajo se utiliza sistema estereoscópico que es situado a 3 metros del suelo y en una posición fija.

En la mayoría de los trabajos se usan posiciones elevadas de la cámara [5, 6, 7]. Sin embargo, en algunos otros trabajos en que se busca la interacción con el usuario, la cámara suele situarse en posiciones inferiores a la altura del mismo [1, 4, 11]. Además de mejorar la visibilidad de la cara y brazos del sujeto, estos enfoques son más adecuados para su implantación en sistemas robóticos que requieran interactuar con los humanos. Los estudios realizados muestran que para una mejor aceptación de los robots por parte de los humanos es importante que los primeros sean de menor tamaño que los segundos [2]. En caso contrario el humano podría sentirse intimidado.

En este trabajo proponemos un método para la detección y seguimiento de una personas utilizando un sistema estereoscópico móvil colocado en alturas inferiores a las de las personas. A diferencia de la mayoría de los trabajos estudiados, que sólo modelan el entorno para cámaras inmóviles [1, 4, 5, 6, 7], nosotros proponemos un método para crear un mapa que modele todo el espacio visible por el sis-

tema estereoscópico al mover la cabeza robótica. Una característica destacable del método es que el mapa puede ser creado incluso en presencia de objetos móviles. La utilización de este mapa nos permitirá detectar fácilmente los objetos que no forman parte del entorno y reducir la búsqueda de personas sólo a esos objetos. Esta reducción en la información a analizar nos permitirá, además de reducir el coste computacional, eliminar falsos positivos producidos por el detector de caras. El agente que hemos creado emplea visión activa para seguir al sujeto en su movimiento por el entorno. Esto nos permite realizar un seguimiento de la persona en un entorno más amplio que si hubiésemos usado una cámara inmóvil y de esta manera permitir una interacción más natural y cómoda.

## 2. Sistema hardware

El sistema hardware empleado se compone de un ordenador portátil para realizar el procesamiento, un sistema estereoscópico con dos cámaras y una cabeza robótica. La cabeza robótica (Pan-Tilt Unit o PTU) tiene dos grados de libertad, uno en el eje horizontal (pan) de  $\phi = [-139, 139]$  grados y otro en el vertical (tilt) de  $\psi = [-47, 31]$  grados.

El uso de nuestro sistema estereoscópico nos permite, capturar dos imágenes desde posiciones ligeramente distintas (par estéreo) y crear a partir de ellas una imagen de disparidad  $I_d$ . Conociendo los parámetros internos del sistema estereoscópico es posible estimar la posición tridimensional  $p_{cam}$  de los puntos de  $I_d$ . Debido a que la cámara está sometida a movimiento, estos puntos son trasladados a un sistema de referencia estático que tiene como centro la cabeza robótica a la altura del suelo mediante la Ecuación 1. La matriz de transformación  $T$  se crea utilizando los parámetros intrínsecos del sistema (proporcionados por el fabricante) y extrínsecos que son estimados previamente.

$$p_w = Tp_{cam} \quad (1)$$

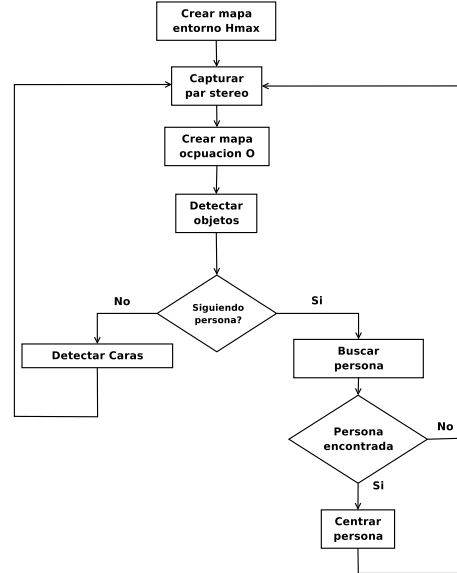


Figura 1: Máquina de estados del proceso.

## 3. Proceso de detección y seguimiento de personas

El método que se propone en este trabajo para la detección y seguimiento de personas es un proceso iterativo que puede verse esquematizado en la Figura 1.

Inicialmente, se crea un mapa del entorno (al que denominaremos  $\mathcal{H}_{max}$ ) que tiene por objetivo el registrar la posición de los objetos inmóviles presentes. Este mapa divide el entorno analizable en celdas de tamaño fijo e indica en cada una de ellas la altura máxima de los objetos detectados.

Una vez que el entorno ha sido aprendido, el sistema comienza un proceso continuo de captura de imágenes para crear un mapa de ocupación instantáneo  $\mathcal{O}$ . En este mapa aparecen aquellos objetos que están en la escena pero que no han sido registrados como objetos inmóviles en  $\mathcal{H}_{max}$ , es decir, aquellos objetos que están en movimiento. Los objetos presentes en  $\mathcal{O}$  son detectados y analizados para determinar cuáles de ellos son persona. Para ello hemos aplicado un detector de caras [12] sobre la imagen a color de la escena. Los falsos positivos generados por el detector de caras serán

rechazados gracias a la integración de la información de la imagen de disparidad y de  $\mathcal{O}$ . Si finalmente, se detectase que hay objetos en  $\mathcal{O}$  que son personas, el agente comenzará a seguir a aquella que se encontrase más cerca.

El seguimiento es también un proceso iterativo que en cada instante crea un mapa de ocupación  $\mathcal{O}$  para buscar en él a la persona deseada. Para realizar el seguimiento hemos usado el filtro del Kalman. Si se encuentra la persona deseada, se determina si se debe mover la PTU para centrarla en la imagen y de esta manera tenerla siempre localizada. El objetivo del proceso de centrado es mantener dentro de la imagen la cabeza y extremidades de la persona, siempre que esto sea posible. En las siguientes secciones se explican en detalle los procesos más relevantes anteriormente comentados.

### 3.1. Creación del mapa del entorno $\mathcal{H}_{max}$

En una fase previa al proceso de detección, se realiza una estimación del entorno de trabajo (*background*). Este proceso tiene por objeto aprender la estructura y objetos inmóviles del mismo. El modelo del entorno servirá para separar los objetos que no forman parte de él (objetos móviles). Nuestro enfoque se basa en la creación de un mapa geométrico de altura del entorno  $\mathcal{H}_{max}$ , que divide al plano del suelo en un conjunto de celdas de tamaño fijo. Los puntos detectados por el sistema estéreo  $p_w$  son proyectados sobre  $\mathcal{H}_{max}$ , que almacena en cada celda el valor máximo de las alturas de los puntos que se proyectan sobre ella. Para evitar agregar los puntos del techo en  $\mathcal{H}_{max}$ , aquellos que superen un umbral de altura  $h_{max}$  son excluidos del procesamiento. Por razones de eficiencia de cálculo, también se excluyen del procesamiento aquellos puntos que estén por debajo de un umbral de altura  $h_{min}$ . El intervalo de altura  $[h_{min}, h_{max}]$  debe ser tal, que una gran parte del cuerpo de las personas a detectar se encuentre en él. En aquellas celdas de  $\mathcal{H}_{max}(x, y)$  en las que no se localiza ningún punto, se asume que no hay objeto y que la altura es  $h_{min}$ .

Debido a que el sistema estereoscópico está sujeto a error, en lugar de proyectar solamente la altura del punto detectado sobre una celda, se hace sobre todo el área de incertidumbre

de dicho punto. Para esto se ha utilizado el modelo de error de nuestro sistema estereoscópico cuyos parámetros son proporcionados por el fabricante.

La creación de  $\mathcal{H}_{max}$  utilizando una única imagen de disparidad está sujeto a problemas. Por un lado, posibles objetos que no pertenezcan al entorno (como por ejemplo personas que estén de paso) pueden ser incorrectamente incluidas como parte del entorno. Además, los algoritmos de correlación para la detección estéreo están sujetos a errores que provocan que no todos los puntos de la escena se detecten. Por estas razones, en lugar de crear  $\mathcal{H}_{max}$  a partir de una única imagen de disparidad, se hará utilizando varias imágenes tomadas en distintos instantes de tiempo  $t$ . Por cada una de estas imágenes, se crea un mapa de altura instantáneo  $\mathcal{H}_{max}^t$ . Finalmente, los distintos  $\mathcal{H}_{max}^t$  creados son utilizados para calcular  $\mathcal{H}_{max}$  empleando un estimador robusto como es la mediana. Para ello, cada celda de  $\mathcal{H}_{max}$  tendrá como valor máximo de altura la mediana de los valores observados en los distintos  $\mathcal{H}_{max}^t$  para dicha celda utilizando la Ecuación 2.

$$\mathcal{H}_{max}(x, y) = \text{mediana}(\mathcal{H}_{max}^{t=1}(x, y), \dots, \mathcal{H}_{max}^{t=n}(x, y)) \quad (2)$$

En la Figura 2 se puede observar la creación del mapa de altura de una región del entorno. El mapa ha sido creado utilizando una secuencia de 13 imágenes, de las cuales en la Figura 2 sólo aparecen las correspondientes a los instantes  $t = \{1, 4, 10, 13\}$ . En la fila superior (Figuras 2(a-d)) se muestran, de izquierda a derecha, las imágenes correspondientes a los instantes anteriormente indicados. En la fila central (Figuras 2(e-h)), aparecen los mapas de altura instantáneos  $\mathcal{H}_{max}^t$  de las imágenes de la fila superior. Las zonas oscuras representan zonas de mayor altura y las zonas blancas indican la altura mínima  $h_{min}$ . En la fila inferior (Figuras 2(i-l)), se muestra la evolución del mapa de altura  $\mathcal{H}_{max}$  hasta el instante  $t$ . Se puede observar como para  $t = 1$ ,  $\mathcal{H}_{max}^{t=1} = \mathcal{H}_{max}$ . Pero a medida que se van utilizando más imágenes, éste último tiende a representar la verdadera estructura del entorno. Para crear estos mapas se ha usado un tamaño de celda  $\delta = 1$  cm y el intervalo de altura  $h_{min} = 0,5$  m y  $h_{max} = 2,5$  m.

Para crear un mapa completo del entorno,

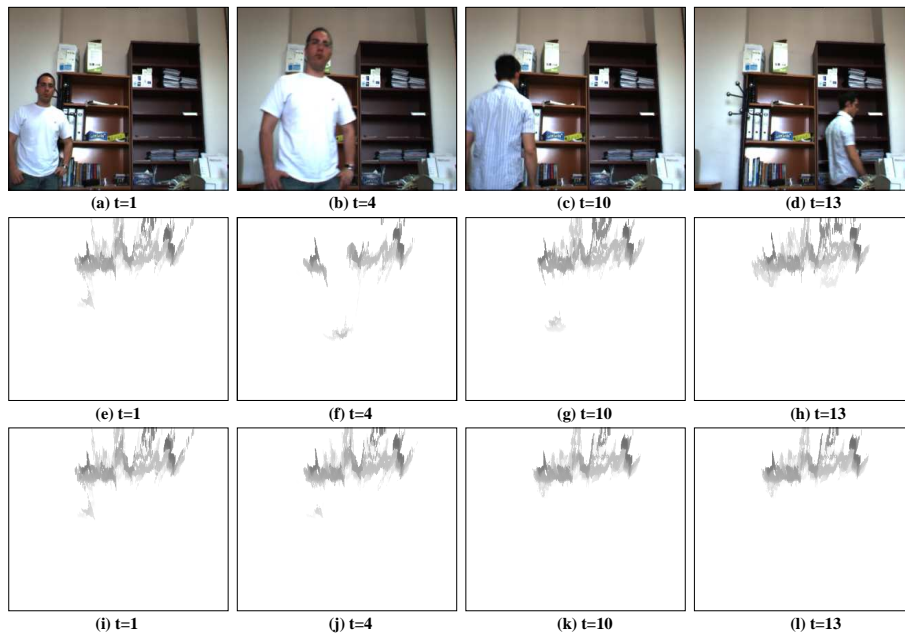


Figura 2: Creación del mapa de altura. En la fila superior (a,b,c,d) se muestra las imágenes capturadas para los instantes 1, 4, 10 y 13. En la fila central (e,f,g,h) se pueden observar los mapas de altura instantáneos  $\mathcal{H}_{max}^t$  para las imágenes de la fila superior. En la fila inferior (i,j,k,l) se presenta la evolución del mapa de altura  $\mathcal{H}_{max}^t$  creado como la mediana de los mapas de altura  $\mathcal{H}_{max}^t$  creados hasta ese instante

habrá que hacer girar la cámara para que capte información desde diferentes direcciones. Para ello, el proceso anteriormente explicado se repetirá para diferentes valores del ángulo  $\phi$  hasta cubrir todo el espacio observable por el sistema visual.

### 3.2. Creación del mapa de ocupación $\mathcal{O}$

Una vez que el mapa geométrico de altura  $\mathcal{H}_{max}$  ha sido creado, se puede comenzar la detección de personas. El primer paso consiste en crear un mapa geométrico de ocupación  $\mathcal{O}$ , que indica en cada celda la superficie ocupada por los objetos que no pertenece al entorno ( $\mathcal{H}_{max}$ ). Para esto, tras capturar un par estéreo de la zona del entorno que se desea examinar, se calcula la posición de los puntos detectados  $p_w$ . Para cada punto  $p_w$  se examina si su altura está dentro de los límites  $[h_{min}, h_{max}]$  y si supera el valor de la celda correspondiente en  $\mathcal{H}_{max}$ . Si es así, la celda correspondiente de  $\mathcal{O}$  es incrementada con un valor proporcional

a la superficie que ocupa el punto encontrado. Puntos más cercanos a la cámara corresponderán a superficies menores y viceversa. Por ello, los puntos más lejanos incrementarán el valor de la celda correspondiente mediante una cantidad mayor que los puntos más cercanos, tal y como se hace en [6]. Si el incremento fuese el mismo para cada celda, el mismo objeto tendría una suma de áreas menor cuanto más alejado de la cámara estuviese. Este escalado en el valor a incrementar compensa la diferencia en el tamaño observado de los objetos según su distancia a la cámara.

Una vez creado  $\mathcal{O}$ , lo analizamos para detectar los objetos que aparecen en él. En un primer paso, se realiza un proceso dilatación seguido de una erosión con el propósito de conectar posibles discontinuidades en los objetos. Tras esto, los objetos son detectados agrupando un conjunto de celdas conectadas cuya suma de áreas supera un umbral  $\theta_{min}$ . De esta manera, eliminamos el posible ruido que apa-

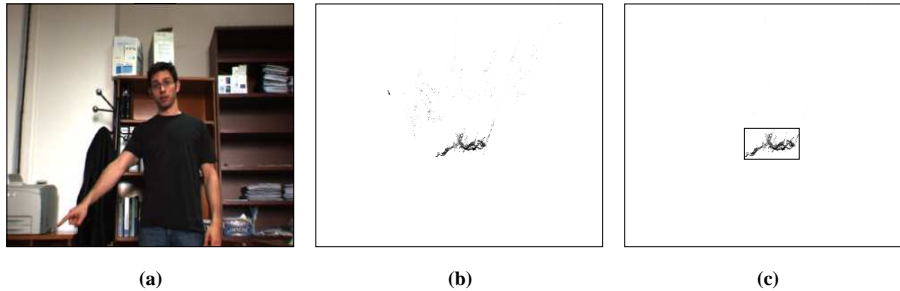


Figura 3: (a) Imagen del entorno (b) Mapa de ocupación  $\mathcal{O}$  (c) Objetos detectados en  $\mathcal{O}$

rece como consecuencia del procesamiento estereoscópico.

La Figura 3(b) muestra el mapa de ocupación  $\mathcal{O}$  de la escena de la Figura 3(a) utilizando como mapa de altura el  $\mathcal{H}_{max}$  de la Figura 2(1). Los valores más oscuros corresponden a zonas con mayor densidad de ocupación. La imagen ha sido manualmente retocada para hacer visibles las zonas ocupadas. Como se puede observar, aparecen en la zona superior de la Figura 3(b) pequeños puntos oscurecidos que son ruido que aparece al realizar el procesamiento estéreo. En la Figura 3(c) se muestra enmarcado el único objeto detectado tras el proceso de dilatación, erosión agrupamiento y umbralización.

### 3.3. Detección de caras

Si tras crear y analizar  $\mathcal{O}$  se detectase que ha entrado algún objeto en el entorno, se procedería aplicar un detector de caras para determinar cuáles corresponden a la cara de una persona. Como detector de caras se ha usado el método inicialmente propuesto por Viola y Jones [12]. Éste consiste en un detector general de objetos basado en la utilización de múltiples clasificadores simples dispuestos en cascada. La librería OpenCV [8] implementa este tipo de clasificador y proporciona uno apropiadamente entrenado para detectar caras. El método toma como entrada una imagen en niveles de gris y retorna las regiones de la imagen donde se detectan caras (ver Figura 4(a)).

Dado que el detector de caras es proclive a encontrar falsos positivos, es importante verificar que realmente la detección corresponde a

la cara de una persona. Como mecanismo de verificación se impondrá que los puntos detectados como pertenecientes a una posible cara no pueden estar repartidos entre objetos diferentes de  $\mathcal{O}$ . Para ello, analizaremos la región de cara en la imagen de disparidad y vemos si los puntos pertenecen a un único objeto de  $\mathcal{O}$ . Sin embargo, esta región puede contener puntos pertenecientes al fondo u otro objeto incluso cuando el detector acierte en encontrar una cara. En la Figura 4(a) se observa como se detecta satisfactoriamente una cara y como en la región indicada hay puntos que pertenecen a la pared del fondo. Por esta razón será necesario determinar de forma exacta qué puntos de la región indicada por el detector son verdaderamente de cara y cuáles no. Para ello realizamos un proceso que consiste en determinar en primer lugar la mediana de los valores de disparidad en el cuadrado indicado por el detector de caras. Los puntos con dicho valor de disparidad son utilizados como semilla para realizar un crecimiento de regiones. En la Figura 4(c) puede verse la región seleccionada con este método para la imagen de disparidad 4(b). Si tras este análisis se detectara más de una persona en el entorno, el sistema comenzaría el seguimiento de la más cercana.

### 3.4. Seguimiento

El proceso de seguimiento es iterativo y comienza al tomar un par estéreo y crear el correspondiente mapa de ocupación  $\mathcal{O}$ . Tras detectar los objetos presentes en  $\mathcal{O}$  (como se explicó anteriormente), se debe encontrar cuál

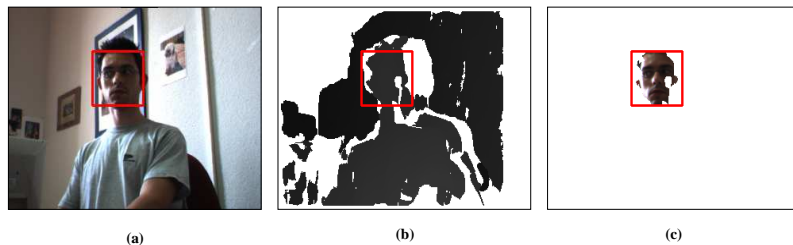


Figura 4: Detección de caras. (a) Imagen capturada y cara detectada (b) Imagen de disparidad y cara detectada (c) Región de la cara perteneciente a la persona

de ellos es el sujeto en seguimiento. Para fusionar la información disponible tomada en instantes anteriores así como la información procesada en el instante actual se ha utilizado el filtro de Kalman. Esta herramienta nos va a permitir fusionar de manera adecuada la posición que predice el modelo del sistema con la observación obtenida en nuestro proceso de búsqueda. Si el sujeto es encontrado, se emplea visión activa dirigiendo adecuadamente el sistema visual para mantenerlo siempre centrado en la imagen. El centrado tiene por objeto mantener visible al sujeto en la imagen situándolo en la mejor posición posible para poder analizar los gestos que realiza. Si el sujeto se encuentra de pie en posición normal, se pretende captar la cabeza y el tronco. Si eleva los brazos para señalar cualquier objeto o se agacha para recoger algo del suelo, es deseable poder captar la acción. En nuestro método, hemos contemplado la posibilidad tanto del desplazamiento de las personas así como movimientos que impliquen cambios de altura (agacharse o sentarse). Hemos comprobado experimentalmente, que la mejor opción para conseguir esto es mantener la zona visible más alta del sujeto en la zona superior de la imagen. Para determinar el movimiento que ha de realizar la PTU para realizar el centrado, hemos empleado un sistema basado en reglas difusas que se han diseñado usando conocimiento experto y que ha sido ajustado según la experimentación realizada.

#### 4. Experimentación

Durante la explicación del método hemos ido mostrando ejemplos de su funcionamiento.

Se ha realizado una experimentación más amplia que no podemos mostrar con imágenes por cuestiones de espacio pero que comentaremos brevemente. Esta experimentación ha consistido en la detección y seguimiento de diferentes personas bajo distintas condiciones de iluminación y a distintas distancias del sistema de visión. Para realizar el procesamiento estéreo hemos utilizado imágenes de tamaño  $320 \times 240$  utilizando interpolación de subpixel para ganar en precisión.

El uso del mapa de altura propuesto nos permite modelar todo el entorno haciendo girar el sistema estereoscópico en todas direcciones. El método de creación propuesto (usar la mediana de las alturas) permite crear el mapa incluso con personas en movimiento en la habitación tal y como se mostró en la Sección 3.1.

Hemos comprobado que el método de detección de personas funciona adecuadamente eliminando satisfactoriamente los falsos positivos producidos por el detector de caras. Las distancias más apropiadas para detectar personas varían entre 0,5 y 2,5 metros. Sin embargo, una vez se ha determinado la persona a seguir de entre las detectadas, el seguimiento puede realizarse a mayores distancias.

Los tiempos de cómputo son distintos para el proceso de detección que para el de seguimiento, si bien el procesamiento estéreo consume una gran parte del tiempo (120 ms). En el proceso de seguimiento, el detector de caras es la tarea más costosa (81 ms), alcanzándose ratios de 2,5 fps para el proceso de detección completo (incluyendo procesamiento estéreo). Sin embargo, en el proceso de seguimiento conseguimos ratios de hasta 5 fps lo que

hemos comprobado que es suficiente en nuestro caso. Estos valores podrían incrementarse sustancialmente si fuese posible optimizar el código del procesado estéreo o con el uso de hardware específico.

## 5. Conclusiones y trabajo futuro

En este trabajo hemos presentado un agente capaz de detectar personas y que emplea visión activa para seguir las. Para ello hemos utilizado un sistema estereoscópico montado sobre una cabeza robótica. El agente crea inicialmente de un mapa de altura del entorno que recoge las características inmóviles del mismo. Este mapa es posteriormente utilizado para detectar los objetos móviles del entorno y buscar entre ellos a posibles personas utilizando un detector de caras. Una vez que se ha detectado a una persona, el agente es capaz de seguirla utilizando la cabeza robótica que permite girar el sistema estéreo. Para hacer más robusto el seguimiento, hemos utilizado el filtro de Kalman. El agente mantiene localizada a la persona en todo momento enmarcando su cabeza y brazos en la imagen. Esta tarea podrá ser utilizada por otros agentes que necesiten analizar los gestos y expresiones de posibles usuarios para aplicaciones en el ámbito de la interacción hombre-máquina.

Como trabajos futuros, contemplamos la actualización del mapa del entorno. Para ello, el agente deberá ser capaz de agregar al mapa del entorno aquellos objetos, que no siendo personas, permanecen estáticos por largo tiempo y no se encontraban en el momento de creación del mapa. Otro aspecto a considerar es la utilización de un filtro de partículas para la tarea de seguimiento.

## Referencias

- [1] T. Darrell, G. Gordon, M. Harville, and J. Woodfill. Integrated Person Tracking Using Stereo, Color, and Pattern Detection. *Int. Journ. Computer Vision*, 37:175–185, 2000.
- [2] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots, 2002.
- [3] D. M. Gavrila. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding: CVIU*, 73(1):82–98, 1999.
- [4] D. Grest and R. Koch. Realtime multi-camera person tracking for immersive environments. In *IEEE 6th Workshop on Multimedia Signal Processing*, pages 387–390, 2004.
- [5] I. Haritaoglu, D. Beymer, and M. Flickner. Ghost 3d: detecting body posture and parts using stereo. In *Workshop on Motion and Video Computing*, pages 175 – 180, 2002.
- [6] Michael Harville. Stereo person tracking with adaptive plan-view templates of height and occupancy statistics. *Image and Vision Computing*, 2:127–142, 2004.
- [7] K. Hayashi, M. Hashimoto, K. Sumi, and K. Sasakawa. Multiple-person tracker with a fixed slanting stereo camera. In *6th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 681–686, 2004.
- [8] Intel. OpenCV: Open source Computer Vision library. <http://www.intel.com/research/mrl/opencv/>.
- [9] W. Liang, H. Weiming, and L. Tieniu. Recent developments in human motion analysis. *Pattern Recognition*, 36:585–601, 2003.
- [10] L. Snidaro, C. Micheloni, and C. Chiavedale. Video security for ambient intelligence. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 35:133 – 144, 2005.
- [11] R. Tanawongsuwan. Robust Tracking of People by a Mobile Robotic Agent. Technical Report GIT-GVU-99-19, Georgia Tech University, 1999.
- [12] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 511–518, 2001.