

Seguimiento de planos en tiempo real: aplicación al seguimiento del rostro^{*}

José Miguel Buenaposada, Cristian Del Pascua, Luis Baumela
Departamento de Inteligencia Artificial
Universidad Politécnica de Madrid
Campus de Montegancedo s/n, 28660 Boadilla del Monte (Madrid)
e-mail: {jmbuena, cris.one}@dia.fi.upm.es, lbaumela@fi.upm.es

Resumen: *Presentamos un método para el seguimiento 3D de regiones planas en tiempo real empleando una única cámara calibrada. Planteamos un modelo proyectivo para el seguimiento de planos. El seguimiento proyectivo, junto con la calibración del sistema de visión, permite la estimación de la orientación y posición del plano con respecto a la cámara. En los experimentos realizados el método se muestra adecuado incluso para el seguimiento de objetos que no son estrictamente planos, como el rostro humano.*

Palabras clave: visión artificial, seguimiento 3D, seguimiento de la cara

1. Introducción

El seguimiento del rostro humano mediante el análisis de una secuencia de imágenes es el paso previo para numerosas aplicaciones de visión computacional: lectura de labios como apoyo al reconocimiento de voz, codificación de gestos, animación de personajes mediante copia de movimientos, estimación de la dirección de la mirada, etc. Dependiendo de la precisión podemos clasificarlo como:

- *Seguimiento 2D*, que consiste en la estimación de la posición en el plano imagen.
- *Seguimiento 21/2D*, o seguimiento de la posición en el plano imagen más la estimación del ángulo de rotación en el plano y/o la escala.

^{*}Trabajo financiado por la CICYT, proyecto TIC1999-1021

- *Seguimiento 3D*, con el que se trata de recuperar la posición y rotación del rostro en el espacio con respecto a la cámara.

El seguimiento 3D se puede abordar desde la ausencia de modelo o sabiendo lo que se quiere seguir a priori. Con una única cámara es más fiable realizar el seguimiento 3D cuando tenemos un modelo de lo que queremos seguir. Por tanto, muchas técnicas de estimación de la posición y orientación 3D del rostro que emplean una única cámara, se basan en un modelo previo [SB96, MLC00, AG96, TH96].

En el seguimiento 3D existen aproximaciones al problema de lo más variadas: basadas en el ajuste de un modelo complejo de la cabeza a la imagen actual [SB96, MLC00], en el seguimiento de puntos característicos de la cabeza [AG96, TH96], en el seguimiento de contornos o en el seguimiento de planos [MJB97].

En el caso del rostro, interesa un seguimiento 3D preciso y robusto. Nuestro objetivo es realizar el seguimiento más preciso posible, de acuerdo con las condiciones de trabajo con las que nos encontremos [BM99] (cambios de iluminación, oclusiones, fondo complicado, etc) de forma que, empezando con un seguimiento 2D rápido pero poco preciso, se aumente paulatinamente la precisión hasta conseguir un seguimiento 3D a medida que las condiciones del entorno se hagan más y más favorables. Dentro de este marco de trabajo podemos ver a una técnica de recuperación de la posición y orientación de la cara basada en el seguimiento de planos como el paso previo al ajuste de un modelo 3D más complejo.

En este trabajo presentamos una técnica de seguimiento 3D que, partiendo de una vista calibrada de un objeto plano, permite estimar la orientación y posición del objeto en una secuencia de imágenes. El sistema es capaz de seguir el rostro humano incluso en posiciones desviadas de la frontal a la cámara a pesar de que en estas condiciones dista mucho de aparecer como plano.

2. Seguimiento de planos con SSD

El seguimiento visual de un objeto es más simple si partimos de un modelo, esto es, de un conocimiento previo de lo que queremos seguir. En el caso que nos ocupa vamos a tratar del seguimiento de regiones planas de las que hemos tomado una imagen previa al seguimiento (plantilla de referencia).

El sustrato del método de seguimiento que vamos a emplear [GDH96] lo constituye una idea muy sencilla. Si observamos un píxel p_x de la imagen tomada en el instante t_n , y que pertenece al objeto plano que queremos seguir, cuando el objeto se mueva, en el instante t_{n+1} , se modificará el valor de gris de p_x ya que sobre él se proyectará una zona diferente del objeto. Utilizando el cambio producido en todos los píxeles que pertenecían al objeto, y debido a que conocemos su estructura a través de la plantilla de referencia, seremos capaces de estimar los parámetros del movimiento que mejor explican todos los cambios en los niveles de gris en el sentido de mínimos cuadrados (SSD).

Ignorando los cambios en los niveles de gris debidos a la iluminación, el valor de gris de un píxel de coordenada \bar{x} en la imagen tomada en el instante t_0 (plantilla de referencia),

y que pertenece a la proyección del plano P , debe de ser igual al valor de gris del píxel sobre el que se proyecta el mismo punto del plano P después de un movimiento $f(\bar{\mu}, \bar{x})$:

$$I(\bar{x}, t_0) = I(f(\bar{x}, \bar{\mu}), t_n) \forall x \in P, \quad (1)$$

donde $I(f(\bar{x}, \bar{\mu}), t_n)$ es la imagen rectificadada en el instante t_n , que se obtiene mediante “warping” empleando los parámetros del instante anterior, t_{n-1} .

A partir de la ecuación (1) podemos plantear el cálculo de los parámetros de movimiento $\bar{\mu}$ mediante la minimización de las diferencias entre los niveles de gris de la plantilla de referencia y los valores de gris de la proyección del plano P sobre la imagen en el instante t_n :

$$\min_{\bar{\mu}} \left(\sum_{\bar{x} \in P} (I(f(\bar{x}, \bar{\mu}), t_n) - I(\bar{x}, t_0))^2 \right) \quad (2)$$

La minimización planteada en (2) puede resolverse de una manera lineal si calculamos $\bar{\mu}$ incrementalmente al tiempo que hacemos el seguimiento. Para ello basta realizar un desarrollo en serie de Taylor de (2) en el punto $(\bar{\mu}, t_n)$ y despejar el incremento de parámetros entre dos instantes de seguimiento $\delta\mu$, quedando [GDH96]:

$$\delta\bar{\mu} = -(M^t M)^{-1} M^t [I(f(\bar{x}, \bar{\mu}), t_n) - I(\bar{x}, t_0)] \quad (3)$$

donde M es la matriz jacobiana cuyas filas son las derivadas parciales del valor de gris de cada píxel con respecto a cada uno de los parámetros de movimiento. La matriz jacobiana representa el conocimiento a priori sobre la estructura del objeto a seguir ya que nos informa sobre cómo afecta al valor de gris de cada píxel un pequeño cambio en cada uno de los parámetros. Visto de otra forma, representa qué información proporciona cada píxel para realizar el seguimiento. La realidad es que no se puede seguir cualquier objeto con este método ya que pudiera ocurrir que una plantilla de seguimiento arrojarase una matriz $M^t M$ singular.

Si observamos la ecuación (3) veremos que para cada cuadro de la secuencia se incurre en un gran coste computacional con el cálculo de la matriz M de orden $N \times n$ (siendo N el número de píxeles de la plantilla de seguimiento y n el número de parámetros del modelo). La matriz jacobiana viene dada por:

$$M(\bar{\mu}) = \begin{pmatrix} \nabla_x I(\bar{x}_1, \bar{\mu}_0)^t \cdot f_x(\bar{x}_1, \bar{\mu})^{-1} \cdot f_\mu(\bar{x}_1, \bar{\mu}) \\ \nabla_x I(\bar{x}_2, \bar{\mu}_0)^t \cdot f_x(\bar{x}_2, \bar{\mu})^{-1} \cdot f_\mu(\bar{x}_2, \bar{\mu}) \\ \vdots \\ \nabla_x I(\bar{x}_N, \bar{\mu}_0)^t \cdot f_x(\bar{x}_N, \bar{\mu})^{-1} \cdot f_\mu(\bar{x}_N, \bar{\mu}) \end{pmatrix}, \quad (4)$$

siendo respectivamente $\nabla_x I$ y f_x las derivadas de la plantilla de referencia y f con respecto a las coordenadas del píxel, y f_μ la derivada de f con respecto a los parámetros de movimiento.

Si el modelo de movimiento lo permite, se puede factorizar la matriz jacobiana M en un producto de dos matrices [GDH96],

$$M(\bar{\mu}) = \begin{pmatrix} \nabla_x I(\bar{x}_1, \bar{\mu}_0)^t \cdot \Gamma(\bar{x}_1) \\ \nabla_x I(\bar{x}_2, \bar{\mu}_0)^t \cdot \Gamma(\bar{x}_2) \\ \vdots \\ \nabla_x I(\bar{x}_N, \bar{\mu}_0)^t \cdot \Gamma(\bar{x}_N) \end{pmatrix} \cdot \Sigma(\bar{\mu}) = M_0 \cdot \Sigma(\bar{\mu}) \quad (5)$$

una matriz constante M_0 de orden $N \times k$, que representa la estructura del plano a seguir, y una matriz Σ de orden $k \times n$, que depende de la posición del plano P , siendo N el número de píxeles de la plantilla de referencia y n el número de parámetros del modelo. Entonces los cálculos ha realizar en tiempo real se reducen a la inversión de Σ , una matriz del orden de los parámetros [GDH96]:

$$\delta \bar{\mu} = -\Sigma^{-1} (M_0^t M_0)^{-1} M_0^t [I(f(\bar{x}, \bar{\mu}), t_n) - I(\bar{x}, t_0)] \quad (6)$$

3. Seguimiento de planos con un modelo perspectivo

En [GDH96] se plantea el seguimiento de planos empleando un modelo afín. Éste es un modelo limitado, que falla cuando el objeto rota fuera del plano paralelo al de la cámara. En esta sección vamos a plantear un modelo proyectivo, que permite representar la distorsión que se produce al tomar la imagen de un objeto en cualquier orientación.

Introducir una nueva ecuación de movimiento en el seguimiento SSD consiste en encontrar la descomposición de la matriz jacobiana que induce este nuevo modelo de movimiento. Para el modelo proyectivo emplearemos coordenadas homogéneas en 2D para representar la posición de los píxeles en la imagen I en lugar de coordenadas cartesianas:

$$\bar{x}_h = \begin{pmatrix} r \\ s \\ t \end{pmatrix} \rightarrow \bar{x} = \begin{pmatrix} r/t \\ s/t \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix} \forall \bar{x} \in I(\bar{x}) \quad (7)$$

La ecuación f que describe el movimiento de la región plana será ahora una transformación lineal proyectiva 2D,

$$f(\bar{x}_h, \bar{\mu}) = H \cdot \bar{x}_h = \begin{pmatrix} a & d & g \\ b & e & h \\ c & f & 1 \end{pmatrix} \cdot \begin{pmatrix} r \\ s \\ t \end{pmatrix} = \begin{pmatrix} r' \\ s' \\ t' \end{pmatrix} \quad (8)$$

y por tanto el vector de parámetros, $\bar{\mu}$, del movimiento será:

$$\bar{\mu} = \begin{pmatrix} a & d & g & b & e & h & c & f \end{pmatrix}^T \quad (9)$$

Veremos que es posible la factorización de la matriz jacobiana mediante el cálculo los elementos que aparecen en la ecuación (4) para el modelo proyectivo:

$$\nabla_{x_h} I(\bar{x}_h, \bar{\mu}_0)^t = \begin{pmatrix} I_u & I_v & -(uI_u + vI_v) \end{pmatrix}^t \quad (10)$$

$$f_x(\bar{x}, \bar{\mu})^{-1} = H^{-1} \quad (11)$$

$$f_\mu(\bar{x}, \bar{\mu}) = \begin{pmatrix} r & 0 & 0 & s & 0 & 0 & t & 0 \\ 0 & r & 0 & 0 & s & 0 & 0 & t \\ 0 & 0 & r & 0 & 0 & s & 0 & 0 \end{pmatrix} \quad (12)$$

Con lo que podemos plantear la factorización de la matriz jacobiana según la ecuación (5),

$$\begin{aligned} f_x(\bar{x}, \bar{\mu})^{-1} \cdot f_\mu(\bar{x}, \bar{\mu}) &= H^{-1} \cdot \left(rI_{3 \times 3} \mid sI_{3 \times 3} \mid \frac{tI_{2 \times 2}}{0} \right) \\ &= \left(rH^{-1} \mid sH^{-1} \mid tH_{12}^{-1} \right) \\ &= \underbrace{\left(rI_{3 \times 3} \mid sI_{3 \times 3} \mid tI_{3 \times 3} \right)}_{\Gamma(\bar{x}_h)} \cdot \underbrace{\begin{pmatrix} H^{-1} & 0 & 0 \\ 0 & H^{-1} & 0 \\ 0 & 0 & H_{12}^{-1} \end{pmatrix}}_{\Sigma(\bar{\mu})} \end{aligned} \quad (13)$$

siendo H_{12}^{-1} la matriz formada por las dos primeras columnas de H^{-1} e $I_{q \times q}$ la matriz identidad con q filas y columnas. Con esta factorización de la matriz jacobiana es posible el seguimiento proyectivo de una región plana realizando durante el seguimiento únicamente la inversión de una matriz 8×8 .

4. Estimación de la posición y orientación

Esta técnica de seguimiento de un plano nos permite estimar la transformación entre una imagen previamente tomada del objeto (plantilla de referencia) y la proyección del objeto sobre la imagen actual en la secuencia. En esta sección veremos que es posible la estimación de la posición y orientación del objeto plano con respecto a la cámara empleada en el seguimiento.

Hasta ahora tenemos la relación entre dos proyecciones del plano en seguimiento pero no poseemos información sobre el objeto plano real. En realidad nos faltan dos transformaciones, la que nos lleva del plano P real al plano imagen en el instante t_0 , H_W^0 , y la que nos lleva del plano P al plano imagen en el instante t_n , H_W^n , (Fig. 1).

Para calcular la posición y orientación del plano P con respecto a la cámara, se coloca el sistema de referencia de la escena con los ejes X e Y coincidentes con P y el Z perpendicular al plano (de tal forma que la ecuación del plano sea $Z = 0$) (Fig. 1). Este sistema de referencia permitirá una simplificación en los desarrollos posteriores.

La transformación H_W^0 entre el plano P y la plantilla de referencia P_1 es fácil de calcular en la construcción del modelo, dados cuatro puntos sobre el plano imagen I_1 y sus correspondientes coordenadas del mundo en el sistema de referencia colocado en el plano.

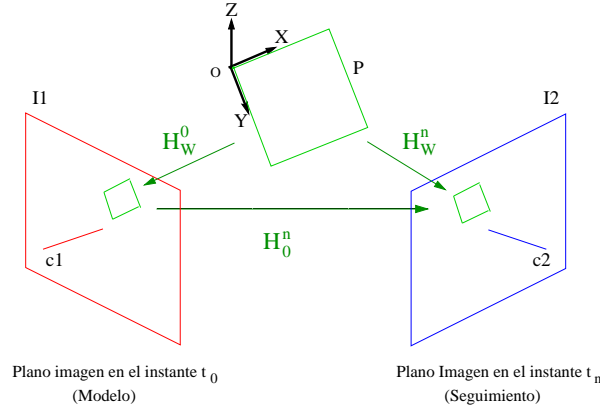


Figura 1. Las tres transformaciones proyectivas implicadas en el seguimiento 3D de un plano mediante SSD.

Con esta nueva transformación tenemos:

$$\begin{pmatrix} x_0 \\ y_0 \\ 1 \end{pmatrix} = H_W^0 \cdot \begin{pmatrix} X_P \\ Y_P \\ 1 \end{pmatrix} \quad (14)$$

Por otro lado H_W^n nos viene dada por la calibración de la cámara empleada en el seguimiento:

$$\begin{pmatrix} x_n \\ y_n \\ 1 \end{pmatrix} = \lambda \cdot K \cdot [R|t] \cdot \begin{pmatrix} X_P \\ Y_P \\ 0 \\ 1 \end{pmatrix} \quad (15)$$

donde λ es una constante de proporcionalidad, K es la matriz de intrínsecos, R es la orientación del plano P en el sistema de referencia de la cámara, y t es la posición del plano P en el sistema de referencia de la cámara (ya que el sistema de referencia de la escena ahora coincide con el plano P).

Teniendo en cuenta como hemos elegido el sistema de referencia de la escena, en el que los puntos de P tienen $Z = 0$, podemos plantear H_W^n :

$$\begin{pmatrix} x_n \\ y_n \\ 1 \end{pmatrix} = \lambda \cdot K \cdot [r_1 r_2 t] \cdot \begin{pmatrix} X_P \\ Y_P \\ 1 \end{pmatrix} = H_W^n \cdot \begin{pmatrix} X_P \\ Y_P \\ 1 \end{pmatrix} \quad (16)$$

donde r_i es la columna i de la matriz R .

Con H_W^0 y H_W^n tenemos:

$$\begin{pmatrix} x_n \\ y_n \\ 1 \end{pmatrix} = \underbrace{H_W^n \cdot (H_W^0)^{-1}}_{H_0^n} \cdot \begin{pmatrix} x_0 \\ y_0 \\ 1 \end{pmatrix} = \quad (17)$$

$$\underbrace{\lambda \cdot K \cdot [r_1 r_2 t]}_{H_0^n} \cdot (H_W^0)^{-1} \cdot \begin{pmatrix} x_0 \\ y_0 \\ 1 \end{pmatrix} \quad (18)$$

Con lo que tenemos otra expresión de H_0^n . Por tanto, suponiendo conocidas:

- La matriz de intrínsecos K de la cámara con la que se realiza el seguimiento.
- La matriz de transformación H_W^0 entre el plano real P y la plantilla de referencia tomada con la cámara con la que se construye el modelo.
- La matriz de transformación H_0^n que nos viene dada por el seguimiento SSD.

podemos encontrar la rotación y translación del plano seguido con respecto a la cámara (con la que se realiza el seguimiento) a partir de la matriz reducida H^* de H_0^n [SFZ00],

$$H^* = (K)^{-1} \cdot H_0^n \cdot H_W^0 = \lambda \cdot [r_1 r_2 t] \quad (19)$$

La translación se obtiene directamente de la tercera columna de H^* mientras que para la rotación hay que imponer ciertas condiciones:

- $\| r_1 \| = \| r_2 \| = 1$, ya que R es una matriz de rotación (= ortonormal). De esta forma obtenemos \hat{r}_1 y \hat{r}_2 .
- $r_3 \perp r_1$ y $r_3 \perp r_2$, con lo que obtenemos \hat{r}_3

5. Experimentos

La experimentación con el método de seguimiento de regiones planas propuesto se ha llevado a cabo en un sistema GNU/Linux sobre procesador AMD K7 a 750MHz y empleando en la captura de imágenes una cámara digital IEEE 1394 Sony VL500. El sistema de seguimiento funciona a 18 cuadros por segundo en línea (aunque el cuello de botella actual no se encuentra en el procesamiento sino en la forma en que se hace la captura).

Hasta ahora hemos obviado el efecto de los cambios de iluminación sobre el valor de gris de los píxeles pertenecientes a la región en seguimiento. Para aliviar parte del problema realizamos la ecualización de la plantilla de seguimiento y de la imagen rectificada [BM99], lo que permite eliminar el efecto de los cambios de intensidad de la iluminación pero no los del cambio de la dirección de la misma, cuyo resultado es la aparición de

sombras en partes de los objetos que cambian su estructura con respecto a la plantilla de seguimiento inicial.

Vamos a validar el algoritmo de seguimiento de planos en 3D mediante 3 experimentos¹: en el primero de ellos tomaremos una secuencia de un objeto plano diseñado por nosotros (Fig. 2), en la segunda secuencia seguiremos la cubierta de un libro (Fig. 3) y en la tercera secuencia veremos que el algoritmo se puede emplear en el seguimiento del rostro humano (Fig. 4).

Aunque no hemos validado directamente la estimación de la orientación que nos entrega el algoritmo, mediante la superposición del sistema de referencia del plano 3D sobre la imagen podemos realizar una estimación indirecta de la misma. En todos los casos los ejes sobre la imagen son coherentes con el movimiento del plano, excepto cuando la estimación SSD es menos precisa, como en el caso del movimiento del rostro hacia arriba y hacia abajo.

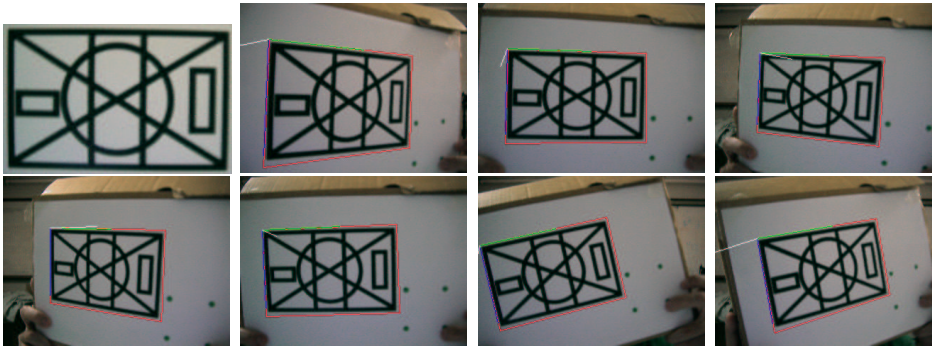


Figura 2. De izq. a drch. y de arriba a abajo, plantilla de seguimiento y cuadros 15, 85, 160, 200, 235, 280, 330 y 390 de una secuencia de 400. El recuadro rojo es la estimación de la posición de la plantilla de seguimiento del seguidor SSD. La estimación de la orientación nos la muestra el sistema de referencia del plano superpuesto sobre la imagen: en blanco el eje Z se dirige hacia dentro del plano imagen, en azul el eje Y y en verde el eje X

6. Conclusiones

Hemos presentado un seguidor proyectivo de planos con el que, partiendo de una plantilla calibrada, es posible estimar la orientación y posición del plano real con respecto a la cámara con la que se realiza el seguimiento. Éste seguidor 3D, cuando se emplea con el rostro, puede utilizarse como forma de encajar un modelo 3D completo del rostro, esto es, de un seguidor 3D más preciso. El sistema planteado funciona a 18 cuadros por segundo en un AMD K7 750 MHz, si bien, en la implementación actual la velocidad sólo depende de la forma en que se realiza la captura.

¹Ver los vídeos en <http://www.dia.fi.upm.es/~jmbuena>

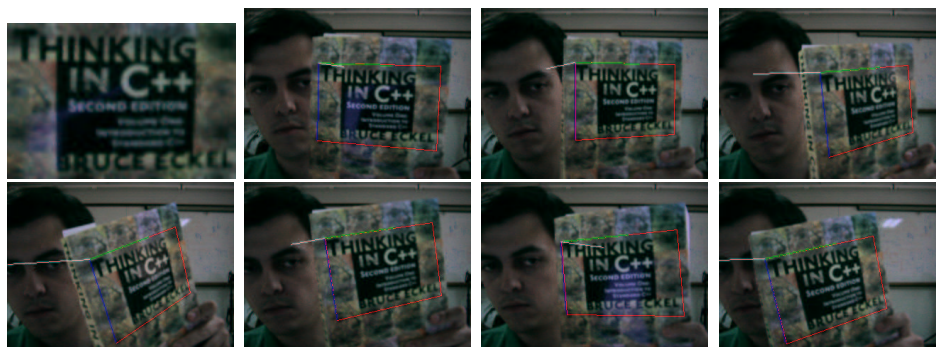


Figura 3. De izq. a drch. y de arriba a abajo, plantilla de seguimiento y cuadros 5, 35, 84, 115, 170, 200, 250 y 350 de una secuencia de 350. El recuadro rojo es la estimación de la posición de la plantilla de seguimiento del seguidor SSD. La estimación de la orientación nos la muestra el sistema de referencia del plano superpuesto sobre la imagen: en blanco el eje Z se dirige hacia dentro del plano imagen, en azul el eje Y y en verde el eje X



Figura 4. De izq. a drch. y de arriba a abajo, plantilla de seguimiento y cuadros 37, 99, 130, 193, 233, 289 y 349 de una secuencia de 350. El recuadro rojo es la estimación de la posición de la plantilla de seguimiento del seguidor SSD. La estimación de la orientación nos la muestra el sistema de referencia del plano superpuesto sobre la imagen: en blanco el eje Z se dirige hacia dentro del plano imagen, en azul el eje Y y en verde el eje X

El sistema actual no es independiente de los cambios de la dirección de la iluminación aunque actualmente se está estudiando una solución para incorporarla en el algoritmo de seguimiento.

Referencias

- [AG96] Roberto Cipolla Andrew Gee. Fast visual tracking by temporal consensus. *Image and Vision Computing*, 14(2):105–114, 1996.
- [BM99] José M. Buenaposada Biencinto and Luis Baumela Molina. Seguimiento robusto del rostro humano mediante visión computacional. In *Libro de actas de la VIII Conferencia de la Asociación Española para la Inteligencia Artificial, CAEPIA'99*, volume I, pages 48–53. AEPIA, 1999.
- [GDH96] Peter N. Belhumeur Gregory D. Hager. Efficient region tracking with parametric models of geometry and illumination. In *Proc. of International Conference on Computer Vision and Pattern Recognition, CVPR'96*, pages 403–410. IEEE, 1996.
- [MJB97] Y. Yacoob M. J. Black. Recognizing facial expressions in image sequences using local parameterized models of image motion. *Int. Journal of Computer Vision*, 25(1):23–48, 1997.
- [MLC00] V. Athitsos M. La Cascia, S. Sclaroff. Fast, reliable head tracking under varying illumination: An approach based on robust registration of texture-mapped 3d models. *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 22(4), April 2000.
- [SB96] A. Pentland S. Basu, I. Essa. Motion regularization for model-based head tracking. In *Proc. International Conference on Pattern Recognition, Viena, Austria*, 1996.
- [SFZ00] G. Simon, A. Fitzgibbon, and A. Zisserman. Markerless tracking using planar structures in the scene. In *Proc. International Symposium on Augmented Reality*, October 2000.
- [TH96] L.S. Davis T. Horprasert, Y. Yacoob. Computing 3-d head orientation from a monocular image sequence. In *Proc. Int'l Conf. Automatic Face and Gesture Recognition*, pages 242–247, 1996.