

Detección facial basada en una distancia de Hausdorff normalizada

Pablo Suau¹, Francisco Pujol², Ramón Rizo¹, Mar Pujol¹

¹ Departamento de Ciencia de la Computación e Inteligencia Artificial

² Departamento de Tecnología Informática y Computación

Universidad de Alicante, Ap. de correos 99, 03080, Alicante

pablo@dccia.ua.es, fpujol@dtic.ua.es, rizo@dccia.ua.es, mar@dccia.ua.es

Abstract. Los sistemas de detección de caras basados en reconocimiento de patrones son utilizados a menudo como paso previo en diversas aplicaciones biométricas. En este tipo de sistemas, la distancia de Hausdorff es una métrica muy adecuada para encontrar la parte de una imagen más similar a una cara, ya que es muy precisa y hace que el porcentaje de error sea bajo. Sin embargo, la robustez global puede aumentar si se adapta al problema concreto. En el presente artículo mostramos un sistema de detección de caras basado en una distancia de Hausdorff normalizada, atendiendo a las características peculiares del proceso. Los experimentos indican que nuestro sistema realiza una detección de caras muy precisa incluso en presencia de fondos complejos o condiciones de iluminación variables.

1 Introducción

Son muchas las aplicaciones del tipo interfaz humano-máquina que deben ser construidas sobre un sistema fiable de detección de caras. El comportamiento de estas aplicaciones no será lo suficientemente robusto si no se realiza una correcta localización facial; un sistema automático de reconocimiento de expresiones faciales no será muy útil si, por ejemplo, este paso fundamental no devuelve unos resultados correctos y ciertas características esenciales como las arrugas de la parte superior de la cara o las cejas quedan fuera del área detectada. A pesar de este hecho, son muchos los sistemas complejos desarrollados en los que se evita este paso previo, ya que sus desarrolladores opinan que es un problema resuelto, y prefieren utilizar una técnica manual para determinar de una manera muy sencilla donde está localizada una cara en la imagen [10].

Recientes estudios sobre el estado del arte en esta disciplina [1] revelan la existencia de diversos métodos simples de detección de caras que podrían ser útiles como parte de estos sistemas más completos y que se basan en el término de *localización facial* [2]. El objetivo de la localización facial es determinar la posición de una única cara en la imagen, suponiendo que dicha imagen incluya tan solo una.

En el caso concreto de la localización facial, estos mismos estudios [1] indican que el ajuste de modelos (o *template matching*) basado en la distancia

de Hausdorff es uno de los métodos más robustos. En un sistema de detección basado en ajuste de modelos, se predefine un patrón correspondiente a una cara estándar (normalmente frontal), y se busca la zona de la imagen más similar a este patrón, mediante correlación. En [3], Jesorsky et al. describen un sistema de localización facial de este tipo basado en la distancia de Hausdorff. Su método permite encontrar caras de distintos tamaños en una imagen, y obtiene mejores resultados que otros sistemas de detección de caras basados en métricas diferentes. Sin embargo, los resultados de este trabajo se ven muy afectados en el caso de que existan fondos complejos.

El siguiente texto presenta un sistema de localización facial mejorado basado en ajuste de modelos y en una distancia de Hausdorff normalizada. Se ha escogido un método de localización debido a que nuestro interés es que sea utilizado como primer paso en un sistema más complejo de reconocimiento de expresiones faciales. En la sección 2, se define la distancia de Hausdorff y se explica cómo puede ser utilizada para encontrar objetos en una imagen. A continuación, en la sección 3, se describe el funcionamiento de nuestro sistema. Finalmente, en la sección 4, se muestran resultados experimentales, incluyendo aquellos que nos permitieron extraer un factor de normalización adecuado para la distancia de Hausdorff.

2 Distancia de Hausdorff en el ajuste de modelos

La distancia de Hausdorff es una métrica que permite comparar conjuntos de puntos. Es usado en reconocimiento de patrones para determinar el grado de similitud entre dos objetos ([4],[5]). La distancia de Hausdorff puede ser definida como la máxima distancia desde un conjunto de puntos al punto más cercano de otro conjunto de puntos. Si $A = \{a_1, \dots, a_m\}$ y $B = \{b_1, \dots, b_n\}$ son dos conjuntos de puntos, la distancia de Hausdorff entre A y B se define como

$$YH(A, B) = \max(h(A, B), h(B, A)) \quad (1)$$

donde

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (2)$$

recibe el nombre de *distancia de Hausdorff directa* del conjunto A al conjunto B , usando una norma subyacente $\|\cdot\|$ en los puntos de A y B , y

$$h(B, A) = \max_{b \in B} \min_{a \in A} \|b - a\| \quad (3)$$

recibe el nombre de *distancia de Hausdorff inversa*. Como se puede comprobar, la distancia de Hausdorff no es simétrica. Esto nos proporciona una medida interesante de la proximidad.

A la hora de utilizar la distancia de Hausdorff para el procesamiento de imágenes, aparecen dos problemas principales: la escala (tema que será discutido posteriormente) y los puntos atípicos. Los experimentos en [7] probaron que el

efecto de los puntos atípicos puede ser atenuado utilizando una modificación de la distancia de Hausdorff denominada *distancia media de Hausdorff*. Esta variante obtiene mejores resultados que otras medidas basadas en Hausdorff ([7]). La distancia media de Hausdorff se define como:

$$h_{avg}(A, B) = \frac{1}{|A|} \sum_{a \in A} \min_{b \in B} \|a - b\| \quad (4)$$

$$h_{avg}(B, A) = \frac{1}{|B|} \sum_{b \in B} \min_{a \in A} \|b - a\| \quad (5)$$

Como se ha comentado anteriormente, la distancia de Hausdorff es muy útil para detectar un objeto en el interior de una imagen. Sea A una representación de la imagen en la que deseamos buscar un determinado objeto, y sea B un patrón representando al propio objeto. Sea también $T = (t_x, t_y, s_x, s_y)$ nuestro espacio de transformaciones, donde t_x y t_y representan translación en x e y , y donde s_x y s_y representan la escala. Sea por último P un conjunto que contenga todos los posibles valores de T (por lo tanto, si $b = (b_x, b_y)$ es un punto perteneciente a B y $p \in P$, entonces $T_p(b) = (s_x \cdot b_x + t_x, s_y \cdot b_y + t_y)$). El rango de transformaciones permitidas y su espacio paramétrico P dependen de la aplicación). La Figura 1 muestra el proceso que permite encontrar un modelo B en el interior de una imagen A mediante el uso de este espacio de transformaciones. Para poder encontrar B dentro de A , es necesario buscar los valores

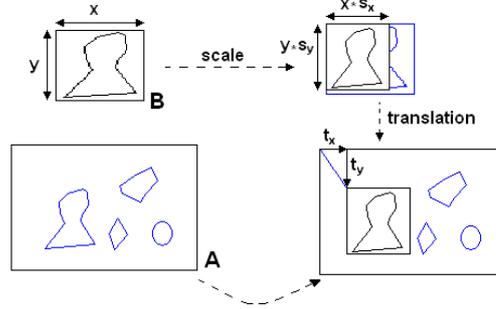


Fig. 1. Aplicando transformaciones para encontrar un patrón que represente a un determinado objeto en el interior de una imagen.

$p \in P$ pertenecientes al espacio de transformaciones que hagan que la distancia de Hausdorff entre el modelo transformado $T_p(B)$ y A se minimice. La distancia de Hausdorff a ser minimizada es la siguiente:

$$d_{\hat{p}} = \min_{p \in P} H(A, T_p(B)) \quad (6)$$

donde \hat{p} es el conjunto de valores $p \in P$ para los parámetros del espacio de transformaciones T que minimizan la distancia de Hausdorff entre la imagen A y el modelo transformado B .

De esta forma, llamaremos a $f_B(T_p) = h(T_p(B), A)$ *distancia directa* y será usada para encontrar un objeto en una imagen, y llamaremos a $f_A(T_p) = h(A, T_p(B))$ *distancia inversa* y permitirá a nuestro sistema detectar falsos positivos. Debemos tener en cuenta que tan solo se debe considerar, a efectos de cálculo de distancias, la parte de la imagen que esta cubierta por el modelo.

3 Descripción del sistema

Nuestro objetivo es utilizar nuestro sistema de detección facial como parte de un sistema completo de reconocimiento de expresiones faciales. Por lo tanto, como en [3], nuestro método se basa en localización facial: solo habrá una cara en las imágenes, y se busca la parte de la imagen más similar a una cara. Nuestro sistema solo devuelve el conjunto de valores del espacio de transformaciones T que minimizan la distancia de la plantilla a una región de la imagen, por lo que en caso de haber presentes varias caras, tan solo se detectaría una de ellas.

La Figura 2 muestra el proceso completo de localización, que se encuentra dividido en una fase de segmentación y una fase de localización. La fase de segmentación transforma la imagen de entrada en una imagen binaria compuesta por aristas, de tal forma que se pueda realizar un ajuste de modelos en dicha imagen. Una vez la segmentación ha finalizado, la localización utiliza Eq.(6) como medida de distancia para efectuar un ajuste de modelos utilizando una determinada plantilla que representa a una cara, con el fin de detectar la posición y tamaño exactos de una cara en la imagen. Estos dos pasos se describen con mayor detalle a continuación.

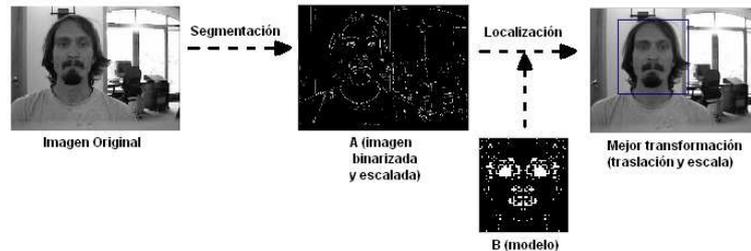


Fig. 2. Esquema del funcionamiento del sistema de localización facial.

3.1 Fase de segmentación

De forma previa a la localización, La imagen de entrada debe ser transformada en una imagen binaria de aristas. Esta imagen binaria debería contener sufi-

ciente información como para que el proceso de localización sea lo suficientemente robusto. Diversos operadores extractores de aristas, como Canny, Sobel, etc., podrían ser utilizados para conseguirlo. De todos ellos, Sobel es el más usado en sistemas de ajuste de modelos basados en distancia de Hausdorff ([3], [4], [5]). Sin embargo, como podemos comprobar en la Figura 3a y en la Figura 3b, usar un operador simple de Sobel en una imagen con un fondo complejo puede ser inapropiado para nuestro problema concreto. Si se aplica un umbral bajo (Figura 3a) obtendremos como resultado una imagen conteniendo demasiadas aristas, incluyendo una gran cantidad de puntos correspondientes al fondo que afectarían en gran medida al paso de localización. En el caso de utilizar un umbral mayor (Figura 3b), el número detectado de aristas disminuirá, pero también desaparecerán puntos muy importantes, como los relacionados con elementos faciales (ojos, nariz, boca, etc.). Nuestro sistema resuelve este problema combinando dos detecciones de aristas con el operador de Sobel empleando diferentes umbrales (Figura 3c). Para conseguir una segmentación adecuada, en la que no haya tantos segmentos correspondientes al fondo y no se pierdan los puntos correspondientes a los elementos de la cara, se calcula la diferencia entre la imagen obtenida tras aplicar un algoritmo de Sobel con un umbral bajo y la imagen obtenida tras aplicar el mismo algoritmo con un umbral alto.

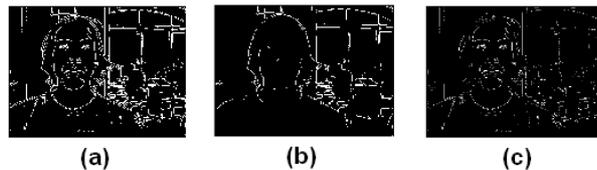


Fig. 3. a) Detección de segmentos mediante el algoritmo de Sobel con un umbral de 0.04. b) Detección de segmentos mediante el algoritmo de Sobel con un umbral de 0.09. c) Detección de segmentos tras combinar los dos resultados anteriores.

Con el objetivo de aumentar la velocidad del sistema, y conociendo el hecho de que el coste de ejecución de la distancia de Hausdorff depende del número de puntos presente en los conjuntos A y B , se realiza un escalado de la imagen de entrada *antes* de que comience la fase de segmentación. Nuestros experimentos nos indicaron que la imagen de entrada podía ser escalada hasta un 40% de su tamaño original sin afectar el resultado final de la detección y disminuyendo considerablemente el coste de ejecución.

3.2 Fase de localización

Como resultado de la fase de segmentación, el sistema dispone de una imagen binaria escalada donde se pueden encontrar diferentes patrones de puntos que podrían representar a una cara. La plantilla que se utiliza es la misma que en

[3] (se puede observar en la Figura 2). Tal como se indica en [6], esta plantilla de puntos se creó como la media de un conjunto de caras a partir de algoritmos genéticos. Casi todos los puntos de la misma representan elementos faciales como ojos, boca y nariz, correspondiéndose algunos otros con los límites de la cara.

A partir de la plantilla B que representa una cara media, y la imagen binaria A obtenida tras la segmentación, se busca la cara que está presente en la imagen utilizando Eq.(6). Para ello, se aplican diferentes valores para los parámetros de escala y traslación de B . Los valores del conjunto de posibles transformaciones \hat{p} que minimizan $H(A, T_p(B))$ nos proporcionan la posición y el tamaño de la cara.

Tal como se comenta anteriormente, la escala de la plantilla es uno de los principales problemas prácticos a la hora de aplicar la distancia de Hausdorff en el procesamiento de imágenes. Debido a que el tamaño de las caras que podemos encontrar en nuestras imágenes de entrada puede ser muy variable, durante la búsqueda se deberá probar un gran rango de valores de escala (s_x, s_y) para el modelo B . Al hacer esto, se observa que los valores de la distancia son menores cuando la plantilla es pequeña; esto significa que aunque las transformaciones correctas de traslación y escala resulten en una distancia de Hausdorff baja, los valores más pequeños de entre todos los posibles valores de escala tendrán como resultado distancias todavía menores. Las consecuencias se pueden observar en la Figura 4a. Por lo tanto, parece necesario aplicar un factor de normalización para adaptar la distancia de Hausdorff a diferentes tamaños de plantilla. Este factor de normalización debe ser escogido con cuidado, pues modificar demasiado la distancia puede producir el efecto contrario: será muy complicado detectar las caras más pequeñas.

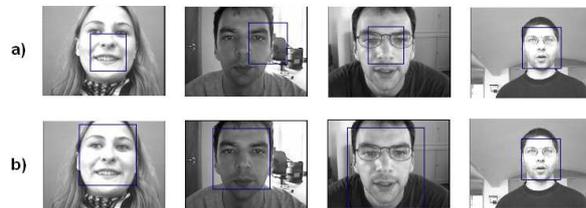


Fig. 4. Problema de escala. a) Detección facial sin factor de normalización. b) Detección facial usando factor de normalización.

Tras un proceso de experimentación (explicado en la siguiente sección), se determinó que al dividir $h(T_p(B), A)$ y $h(A, T_p(B))$ entre $\sqrt{\frac{s_x + s_y}{2}}$ el sistema detecta de una forma más correcta y exacta la posición y el tamaño de las caras, sin este efecto que se ha comentado. La Figura 4b muestra como este factor de normalización resuelve el problema.

Una combinación correcta de las distancias de Hausdorff directa e inversa puede ser muy útil a la hora de evitar falsos positivos. En [3], los autores usan el

producto de las dos distancias para determinar si lo que se detecta es realmente una cara o no. La Figura 5a muestra varios ejemplos de localización incorrecta cuando tan solo se utiliza la distancia directa. Para obtener mejores resultados, se debe encontrar el conjunto de valores para los parámetros de transformación \hat{p} que minimicen el producto de la distancia directa $h(A, T_{\hat{p}}(B))$ y la distancia inversa $h(T_{\hat{p}}(B), A)$. La Figura 5b muestra el resultado de la localización para las mismas imágenes que en el caso anterior, pero usando la distancia inversa para descartar falsos positivos.



Fig. 5. Usando la distancia inversa para descartar falsos positivos. a) Localización en el caso de no utilizar distancia inversa. b) Localización en el caso de utilizar distancia inversa.

4 Experimentación

Durante el desarrollo de los experimentos se utilizó la base de datos BioID³, que se puede obtener de forma gratuita en Internet y que incluye 1521 imágenes en tonos de gris, todas ellas de un tamaño de 384x288 píxeles. Esta completa base de datos contiene imágenes de 23 sujetos distintos, con una gran variedad de condiciones de iluminación, tamaños de cara, y fondos complejos. La plantilla utilizada se muestra en la Figura 2 y tiene un tamaño de 81x86 píxeles.

La región de búsqueda se restringe, al igual que en [3], a una zona cuadrada centrada en la imagen, que cubre el total de la altura de la misma. Previamente a todo el proceso, la imagen de entrada es escalada hasta un 40% de su tamaño original. En la fase de segmentación se realiza la diferencia de dos imágenes obtenidas tras aplicar Sobel con umbrales 0.04 y 0.09, respectivamente. Finalmente, el rango de escalas para la plantilla durante el proceso de localización se encontró entre el 55% y el 120%. Si se dibuja una caja en la posición en la que la cara ha sido detectada, consideraremos que dicha cara ha sido correctamente localizada si la caja dibujada engloba todos los elementos faciales más importantes (ojos, nariz, boca, etc.) y además la cara ocupa al menos el 80% de la misma.

Para demostrar que el valor de la distancia y la escala de la plantilla están relacionados, se calculó la mínima distancia para cada valor de escala aplicado a

³ <http://www.humanscan.de/support/downloads/facedb.php>

la plantilla, para todas las imágenes de la base de datos BioID. Como se puede observar en la Figura 6a, la distancia aumenta con la escala de la plantilla, por lo que las caras más grandes, como se explicó anteriormente, no son correctamente detectadas (Figura 4a). Una demostración más formal puede ser realizada utilizando el factor de correlación de Pearson, una medida de la asociación entre dos variables ampliamente utilizada en estudios estadísticos. El valor del coeficiente de Pearson se encuentra entre -1 y +1, siendo este valor más alto cuando existe una estrecha relación entre las variables. El factor de correlación de Pearson de nuestras dos variables (distancia de Hausdorff y escala aplicada a la plantilla) es $p = 0.795$ con un nivel de significación del 99%, lo cual indica que existe una fuerte relación lineal.

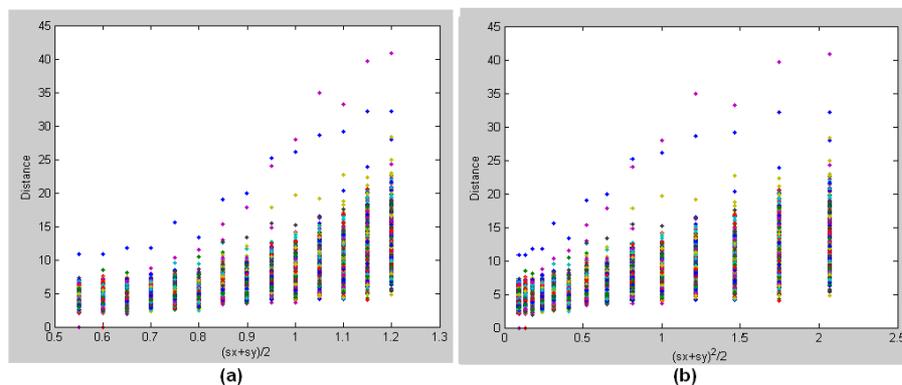


Fig. 6. Correlación entre la distancia de Hausdorff y la escala de la plantilla. a) Distancia frente a escala. b) Distancia frente a $escala^2$.

Sin embargo, la Figura 6a nos muestra que la relación entre estas dos variables parece más cuadrática que lineal. En estos casos suele ser común transformar una o ambas variables con el objetivo de eliminar la curvatura y entonces recalculamos la correlación. En nuestro caso, representamos en la Figura 6b una gráfica en la que se muestra el valor de la distancia de Hausdorff frente a $[\frac{s_x+s_y}{2}]^2$. Tras esta transformación, el valor del coeficiente de correlación fue de $p = 0.8278$ con un nivel de significación del 99%. Por lo tanto, existe una relación cuadrática entre estas dos variables que es más fuerte que una mera relación lineal. Debemos tener en cuenta esta dependencia cuadrática a la hora de definir un factor de normalización. Si se utilizara $\frac{s_x+s_y}{2}$ con este fin, la distancia se modificaría en gran medida, y obtendríamos como resultado el efecto contrario al indicado anteriormente: sería más complicado detectar las caras más pequeñas. Si embargo, al utilizar $\sqrt{\frac{s_x+s_y}{2}}$ como factor de normalización, la distancia se reduce de forma significativa para los valores grandes de escala, pero más ligeramente para las plantillas más pequeñas, con lo que el error en la localización será menor (Figura 4b).

La tasa de error de nuestro sistema fue 7.3%, usando la base de datos BioID y $\sqrt{\frac{s_x+s_y}{2}}$ como factor de normalización. En el caso de que usemos $\frac{s_x+s_y}{2}$ como factor de normalización, la tasa de error se incrementará hasta un 23.73%, debido a que la distancia de Hausdorff es modificada en gran medida.

Otros sistemas que emplean la misma base de datos para comprobar sus resultados tienen una tasa de error mayor [8], o se basan en unas condiciones menos restrictivas para aceptar una localización como correcta [9]. En la Figura 7a se pueden observar algunas detecciones correctas, incluyendo distintas escalas, poses y condiciones de iluminación, así como una variedad de fondos complejos. Estas imágenes demuestran que nuestro sistema es lo suficientemente flexible para localizar de forma correcta algunas caras rotadas fuera de plano, incluso utilizando una plantilla que representa una media de caras frontales, como es nuestro caso.



Fig. 7. a) Algunos ejemplos de detección correcta, incluyendo caras no frontales y fondos complejos, usando un factor de normalización cuadrático. b) Resultados para las mismas imágenes al utilizar un factor de normalización lineal. c) Resultados para las mismas imágenes en el caso de no emplear ninguna normalización.

La Figura 7 también muestra como el factor de normalización afecta a la localización de la cara. En la Figura 7a se pueden observar los resultados correctos cuando se utiliza $\sqrt{\frac{s_x+s_y}{2}}$ como factor de normalización; en la Figura 7b, los resultados en el caso de que se utilice $\frac{s_x+s_y}{2}$ como factor de normalización; y finalmente, en la Figura 7c, los resultados en el caso de no utilizar ningún factor de normalización. Por lo tanto, gracias a dicho factor de normalización, las caras grandes son correctamente detectadas sin afectar a la detección de las caras más pequeñas.

5 Conclusiones y líneas futuras

Se ha presentado un sistema de localización facial basado en ajuste de modelos en una imagen binaria obtenida tras un algoritmo de detección de aristas utilizando

una distancia de Hausdorff, normalizada según la escala de la plantilla, como medida de similitud. La robustez se ve incrementada debido a la introducción de dicho factor de normalización. Nuestro sistema es capaz de detectar caras en imágenes con fondos complejos (debido a nuestro método de segmentación basado en la combinación de diferentes operadores de Sobel) bajo condiciones variables de iluminación, e incluso caras ligeramente rotadas fuera de plano.

Siguiendo esta misma línea sería interesante comprobar en un futuro los beneficios de utilizar multirresolución o el uso de diferentes plantillas para detectar caras en diferentes orientaciones. Finalmente, nuestro sistema de localización facial formará parte de un sistema más completo de reconocimiento de expresiones faciales, que esta siendo desarrollado actualmente.

References

- [1] Yang, M. H., Kriegman, D., Ahuja, N., Detecting Faces in Images: A Survey, *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 24, no. 1, pp. 34-58, 2002.
- [2] Lam, K., Yan, H., Fast Algorithm for Locating Head Boundaries, *J. Electronic Imaging*, vol.3, no.4, pp. 351-359, 1994.
- [3] Jesorsky, O., Kirchberg, K. J., Frischholz, R. W., Robust Face Detection Using the Hausdorff Distance, *Third International Conference on Audio and Video-based Biometric Person Authentication*, Springer Lecture Notes in Computer Science, LNCS-2091, pp. 90-95, Halmstad, Sweden, 2001.
- [4] Huttenlocher, D. P., Klanderman, G. A., Rucklidge, W. J., Comparing Images Using the Hausdorff Distance, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 9, pp. 850-853, 1993.
- [5] Huttenlocher, D. P., Rucklidge, W. J., A multi-resolution technique for comparing images using the Hausdorff distance, *Technical Report 1321, Cornell University, Department of Computer Science*, 1992.
- [6] Kirchberg, K. J., Jesorsky, O., Frischholz, R. W., Genetic Model Optimization for Hausdorff Distance-Based Face Localization, *International ECCV 2002 Workshop on Biometric Authentication*, Springer Lecture Notes in Computer Science, LNCS-2359, pp. 103-111, Copenhagen, Denmark, 2002.
- [7] Shapiro, M. D., Blaschko, M. B., On Hausdorff Distance Measures, *Technical Report UM-CS-2004-071, Department of Computer Science, University of Massachusetts Amherst*, 2004.
- [8] V. Manian, A. Ross, A Texture-based Approach to Face Detection, *Biometric Consortium Conference (BCC)*, (Crystal City, VA), September 2004.
- [9] B. Fröba, C. Küblbeck, Robust Face Detection at Video Frame Rate Based on Edge Orientation Features, *Fifth IEEE International Conference on Automatic Face and Gesture Recognition (FGR 2002)*, pp. 342-347, Washington, USA, 2002.
- [10] M. Rosenblum, Y. Yacoob, L. Davis, Human Expression Recognition from Motion using a Radial Basis Function Network Architecture, *IEEE Transactions on Neural Networks*, Vol. 7, no. 5, pp. 1121-1138, 1996.