Reconocimiento de expresiones faciales mediante flujo óptico en secuencias de video

Filomen Incahuanaco Quispe

[1] Universidad Nacional de San Agustín, Escuela profesional de Ingenieria de Sistemas
Av. Independencia S/N,Arequipa-Perú

jincahuanaco@episunsa.edu.pe

Resumen

La expresión facial es una actividad cotidiana dentro del proceso de comunicación entre seres humanos. Por tanto la cuantificación de las expresiones faciales es un tema abierto en el campo de la visión por computador por su utilidad en la construcción de interfaces de comunicación con el ordenador. También es un tema de interés en otros campos, como el de la animación por computador.

En este trabajo se presentan los resultados de la aplicación de técnicas basadas en teoría de movimiento en secuencias de imágenes orientadas al análisis de expresiones faciales para detectar el estado de ánimo de la persona. El primer paso consiste en elegir los puntos característicos del rostro lo que conlleva a formar una malla de puntos; luego se mide el grado del movimiento facial de la expresión. Previo a este paso fue determinante detectar la posición de los ojos, con el objeto de calcular la inclinación o rotación del rostro. En seguida se establece una clasificación de estos grados de movimiento, usando el clasificador K-Support Vector Machine(K-SVM) con el objeto de determinar si la expresión facial representa una de las seis expresiones faciales más relevantes: normal o neutral, miedo, tristeza, enojo, sorpresa, asco, felicidad.

El aporte del presente trabajo consiste en que nuestra técnica está centrado en las características morfológicas y topológicas de los elementos del rostro relacionadas con las Unidades de Acción(UA), lo que permite reducir la cantidad de puntos y hacer un mejor seguimiento del movimiento, esto a su vez permite disminuir sustancialmente el costo computacional al momento de realizar el cálculo de flujo óptico. Los resultados son muy auspiciosos puesto que se consiguió detectar las expresiones de alegria, sorpresa, enojo y asco con un alto grado de eficiencia, alrededor del 75 %, en el mejor de los casos, por otro lado, las expresiones de tristeza y miedo son poco perceptible, ello porque la tristeza sufre una transición lenta y el miedo por ser una mezcla de expresiones[1].

1. Introducción

El rostro humano provee una fuente de información comunicativa acerca del comportamiento. Las expresiones faciales proporcionan señales que nos permite percibir la respuesta emocional y desempeña un papel importante en la interacción humana y la comunicación no verbal (Arias, 2005).
Así mismo las expresiones faciales muestran la emoción (Ekman, 1993), regulan la conducta social, las señales de intención comunicativa, está computacionalmente relacionado a la producción del habla, y puede revelar la función cerebral y patología (Arias, 2005).

[1]Material suplementario:
http://refaciales.awardspace.com

Por lo tanto, para hacer uso de la información ofrecida por las expresiones faciales es necesario métodos eficaces, confiables y validos para la medición.
En psicología se han propuesto "seis emociones básicas" (Alegria, sorpresa, enojo, tristeza, miedo y asco); cada una incluye cambios de los rasgos faciales en varias regiones del rostro. Estas expresiones básicas, sin embargo, se producen con relativa poca frecuencia en la vida cotidiana. Con mayor frecuencia son los cambios sutiles en uno o varios elementos discretos, como el endurecimiento de los labios que pueden comunicar la ira. Los seres humanos son capaces de producir muchas expresiones que varían en complejidad, intensidad y sentido (Ekman, 1993).

Posteriormente se definió un estándar basado en la anatomía del rostro, llamado FACS(Facial Action Coding System)(Ekman and Friesen, 1978; Cohn et al., 2005), para el uso en la codificación de las expresiones faciales. Con FACS, los observadores pueden codificar manualmente todos los posibles movimientos discretos del rostro; se denominan unidades de acción (AUs). Las AUs individualmente o en combinación pueden representar todas las expresiones visiblemente discriminables.

En cuanto a la detección automática de expresiones faciales, podemos destacar el de Ira Cohen en (Cohen et al., 2003), quien trabajó en la clasificación de expresiones faciales en secuencias de video, empleando el clasificador Naive–Bayes y el cambio de la distribución gaussiana a cauchy. En este trabajo proponen una nueva arquitectura de HMMs(Hidden Markov Models) para dividir y reconocer la expresión automáticamente en segmentos faciales humanos en secuencias de video, usando una arquitectura multi-nivel compuesta de una capa HMM y una del modelo de Markov.

También se trataron de clasificar las expresiones faciales mediante técnicas de Inteligencia Artificial(IA) (Chang and Chen, 2001), para ello fue empleando una red Radial Basis Functions Neural(RBFN).El proceso inicia con la detección de los contornos mas relevantes del rostro, así como los puntos que rodean las zonas de cejas, ojos y labios de la imagen del rostro. A partir de ello, Chang define 30 puntos para describir tres características faciales (Chang and Chen, 2001), teniendo como fundamento las UAs propuestas por Ekman y Friesen (Ekman and Friesen, 1978). Los resultado obtenidos en ese trabajo llegaron a un ratio de alrededor de 92.10 %, sobre las expresiones neutral, enojado y feliz.

Una de las grandes dificultades de los trabajos anteriores está relacionado al hecho que trabajan sobre una de las formas de reconocimiento de movimiento (segmentación) lo cual es un proceso pesado según la literatura existente. Por otro lado, una gran parte de los trabajos relacionados al tema (Black et al., 1997; Essa, 1995; Rosenblum et al., 1994; Yacoob and Davis, 1994) sólo discriminan un pequeño conjunto de emociones, los cuales están basados en el trabajo realizado por Charles Darwin. Adicionalmente, las anteriores técnicas no soportan eficazmente las traslaciones del rostro dentro de la secuencia de video.

1.1. Puntos faciales característicos

El estudio de la morfología facial y la relación entre sus elementos (Kobayashi and Hara, 1992), nos permite suponer que una vez hallada la relación entre los componentes del rostro podemos seleccionar zonas características más relevantes con el fin de enmallar(marcarlas) para así estimar el grado de movimiento. Kobayashi y Hara en (Kobayashi and Hara, 1992), propone un plano del rostro para la selección de zonas catacerísticas; usando una RBFN logran reconocer expresiones con un ratio del 70 %. Posteriormente

Chang en (Chang and Chen, 2001), emplea una métrica similar para el reconcimiento de expresiones faciales, dichas métricas son las de la figura 1.

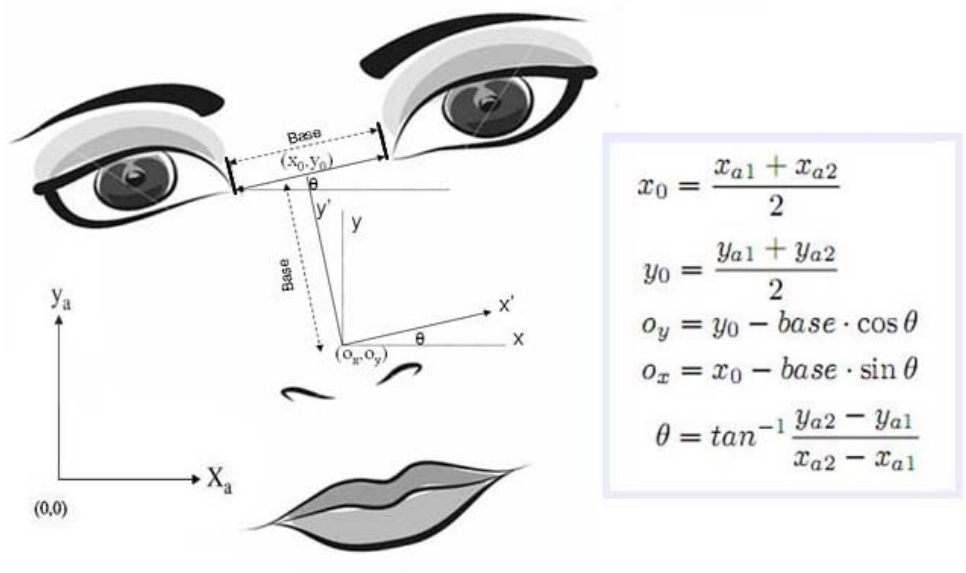

Figura 1: Sistema de coordenadas para el sistema propuesto

Esta referencia nos permite posicionar la malla sobre el rostro adecuadamente para lograr una correcta estimación del movimiento, sin importar la inclinación.

1.2. Arquitectura de la propuesta

La idea principal de este trabajo se centra en clasificar los patrones de movimiento en las secuencias de imágenes del rostro situado frente al ordenador, para ello primero enmallamos el rostro y apartir de ese instante inicia el proceso de clasificación.

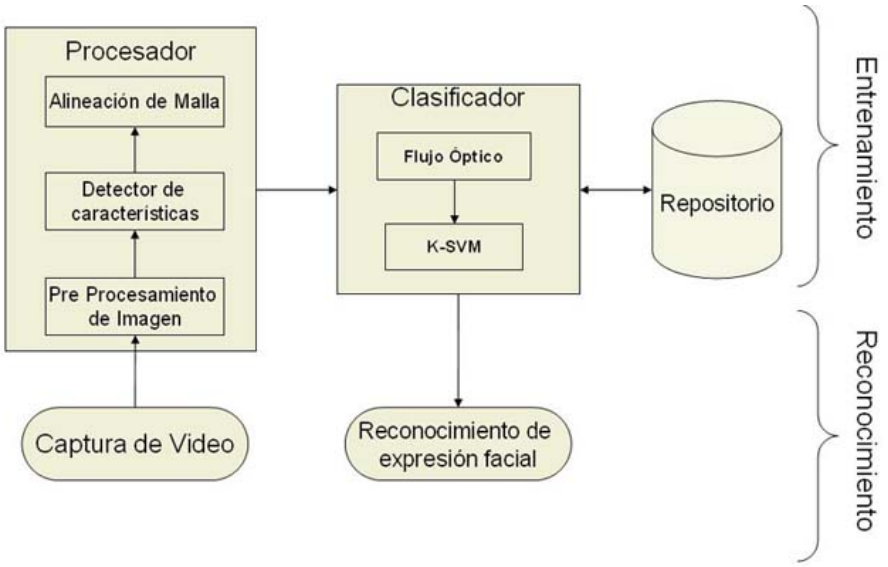

Figura 2: Esquema de trabajo del reconocedor de expresiones

2. Reconocimiento de expresiones faciales con flujo óptico

Dentro de nuestra propuesta están implicadas varias tecnologías, dentro de las cuales podemos nombrar: Detección de movimiento en secuencias de imágenes, seguimiento de objetos, clasificación de patrones entre los más importantes.

2.1. Detección de movimiento en imágenes

Considerando (Izaguirre et al., 2001; Romero, 2002), para detectar objetos en movimiento en secuencias de imágenes existen diferentes tipos de aproximaciones, las cuales se pueden agrupar de la siguiente forma:

- Diferencia de imágenes
- Métodos basados en características
- Flujo óptico
- Métodos basados en regiones

Por medio de la diferencia de imágenes se pueden encontrar los cambios que existen entre una imagen y otra. El costo computacional que se requiere invertir con este tipo de aproximación es reducido, en comparación con métodos donde se ejecuta correlación. Sin embargo, presenta el problema de no poder eliminar el ruido. Así mismo, la operación de diferenciación no proporciona información de cómo se llevó a cabo el movimiento, además detecta cambios de luz y sombra.

Los métodos basados en características se apoyan en la elección de un conjunto de primitivas que corresponden a un rasgo distintivo de la escena. Con este método se debe tener conocimiento de la geometría de proyección, la posición tridimensional relativa y el movimiento, los tipos de primitivas usadas son por ejemplo: Líneas, bordes o esquinas. El problema con los métodos basados en características es que dependen de la forma de los objetos, es decir, pequeños cambios en la forma de los objetos implican una modificación del patrón de búsqueda.

Las aproximaciones que se realizan por medio de flujo óptico, están basadas en el movimiento aparente de la intensidad de los patrones en la secuencia de imágenes. Este esquema está basado en dos suposiciones importantes: la primera de ellas está fundamentada en la observación, esto significa que aún cuando los objetos cambian de posición en la escena, su brillo debe permanecer constante, es decir, que la iluminación sea constante en el tiempo. Esta suposición es conocida como restricción de constancia de datos. La segunda suposición es conocida como la coherencia espacial, es decir, que los puntos de una superficie tridimensional correspondan a los mismos puntos de un vecindario pequeño de la imagen. Las aproximaciones realizadas por medio del flujo óptico caen dentro de dos grandes categorías: basados en correlación y en gradientes.

Los métodos basados en correlación tienen un costo computacional elevado. Por otra parte los métodos que están basados en gradiente pueden ser computacionalmente eficientes, sin embargo, el uso de una ventana de análisis pequeña origina el problema de apertura. El problema se presenta cuando existe una insuficiente variación de los tonos de gris de las regiones consideradas en la imagen, por lo tanto más de un candidato se adapta a los datos de la imagen observada de la misma forma.

En imágenes de rostros si cumplen las suposiciones anteriores debido a que la forma del objeto es constante y no presenta diferentes intensidades a lo largo del tiempo.

Finalmente, existen los métodos basados en regiones donde el problema central es segmentar los objetos en movimiento con respecto al fondo de la imagen. Una forma de estimar el movimiento de objetos por medio de esta técnica es mediante el apareamiento de regiones. Otra forma de realizar la operación de detección es a través de búsquedas de las regiones que definen a los objetos de interés entre un conjunto de cuadros, donde cada una de ellas experimenta un movimiento sencillo entre un par de imágenes consecutivas.

2.2. Clasificador de patrones

El Modelo de clasificación elegido es el Support Vector Machine Multi Class(K-SVM), el cual es una implementación libre llamada LIBSVM (Hsu et al., 2008), basada en (Wu et al., 2003). Cabe resaltar que los autores por su experiencia indican que si hay miles de atributos puede ser necesario seleccionar un subconjunto de ellos antes de proporcionar los datos al clasificador SVM. Sin embargo en nuestro trabajo reducimos las muestras a vectores de 216 atributos reales, correspondientes a las zonas que contienen las UAs.

3. Flujo óptico

El cálculo del flujo óptico que se usa en este proyecto se basa en el método de los gradientes y la "ecuación de restricción del Flujo Óptico", formulada por Horn y Schunck (Horn, 1986):

$$E_x \cdot u + E_y \cdot v + E_t = 0$$

Esta ecuación es válida bajo condiciones de patrón de brillo de la imagen constante, y gradiente local de intensidad lineal. Desarrollando a partir de ella llegamos a una solución de tipo iterativa para cada uno de los componentes del flujo óptico:

$$u = \bar{u} - \frac{E_x(E_x\bar{u} + E_y\bar{v} + E_t)}{\lambda^2 + E_x^2 + E_y^2} \quad (1)$$

$$v = \bar{v} - \frac{E_y(E_x\bar{u} + E_y\bar{v} + E_t)}{\lambda^2 + E_x^2 + E_y^2} \quad (2)$$

Para la obtención de resultados válidos este algoritmo requiere un mínimo de textura en los objetos de la imagen, desplazamientos pequeños entre imágenes y el movimiento continuo de los objetos (Ballard and Brown, 1982), el algoritmo implementado y usado es el de Lucas & Kanade (B.D.Lucas and Kanade, 1981).

Otro problema es el cálculo computacional como lo menciona G. Asensio (Asensio et al., 2007), donde indica que los métodos variacionales son especialmente útiles para resolver la ecuación de flujo óptico $u(x) \cdot \Delta f(t,x) + \partial_t f(t,x) = 0$, por que preservan las discontinuidades y funcionan incluso cuando hay variaciones de iluminación. Sin embargo, tienen alto costo computacional y demuestran que el método multimalla referente al cálculo es mucho mas eficientes. Tal método no es objeto de este trabajo, pero vemos que las tendencias ayudan a presumir una mayor efectividad del modelo propuesto.

4. Resultados

4.1. Datos empleados

Para el entrenamiento del localizador de ojos se empleo la base de datos(Fagertun and Stegmann, 2005), el cual cuenta con 240 imágenes de 40 personas sin lentes, de los cuales 33 son hombres y 7 son mujeres, a una resolución de 640x480 en formato JPEG , de los cuales tomamos 480 muestras positivas de ojos, transformadas a una resolución de 20x10 y 5000 negativas de resolución variable.

Para la obtención de imágenes se empleo una cámara CIF Single chip, con velocidad de 30 cuadros/seg a una resolución de 352x288pixels. El equipo dónde se realizaron las pruebas es una AMD ATHLON 2000XP+, con 304MB de memoria.

Para la etapa de entrenamiento y prueba del clasificador se tomaron muestras de 33 individuos de los cuales 10 son del sexo femenino entre edades de 19 a 23 años, todos ellos estudiantes de pre-grado. las muestras se tomaron durante el día bajo condiciones normales de iluminación, las pruebas también se realizaron bajo las mismas condiciones, pero con individuos que no participaron en el entrenamiento.

4.2. Detección y seguimiento de ojos

Antes de estimar el movimiento aparente en la escena, localizamos los ojos del rostro, para ello empleamos las técnicas de detección propuestas por Viola y Jones en (Viola and Jones, 2002). Obteniendo resultados como se muestra en las figuras 3 y 4.

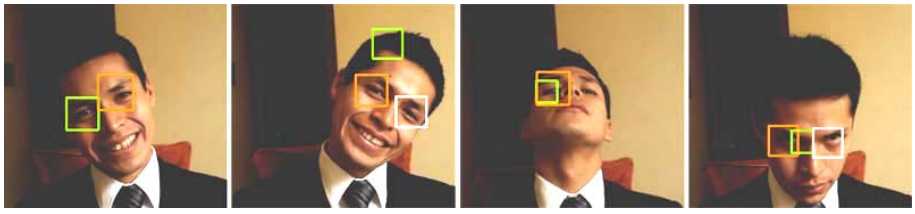

Figura 3: Localización de candidatos de ojos

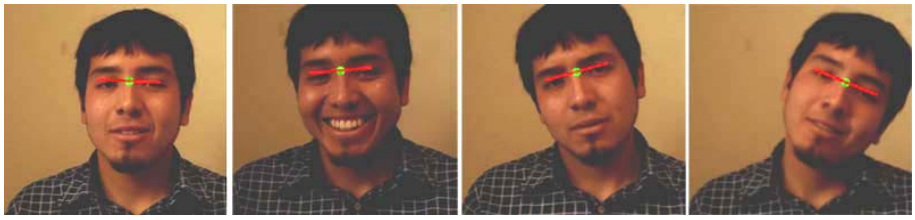

Figura 4: Seguimiento del par de ojos con punto medio marcado

En la figura 3 se observa escenas con recuadros de color para cada candidato a ojo, empleando umbrales numéricos sobre este resultado logramos reducir solo a dos candidatos, los necesarios para posterimente poder calcular el centro (x_0, y_0) del par de ojos. La figura 4 muestra el centro calculado(x_0, y_0), basado en el sistema de coordenadas mostrado en la

figura 1. Cada extremo de la línea representa la posicion de cada ojo detectado. Una de las ventajas de nuestra propuesta frente a los métodos tradicionales de flujo óptico es dar libertad de inclinación hacia ambos extremos sin perturbar los resultados, así se puede observar en la figura 4.

4.3. Enmallado

Denominamos enmallado al proceso de marcar pixeles en las zonas del rostro que contienen las llamadas UAs, estos pixeles marcados se visualizan en forma de puntos en las imágenes a) y b) de la figura 5. Mediante estos pixeles marcados podremos calcular el flujo óptico, es decir el movimiento aparente. Se puede observar el proceso de enmallado de nuestra propuesta en ejecución en la figura 6.

En la figura 5 item a) muestra el enmallado tradicional; en forma de cuadrícula y aparentemente estática. En el item b) el enmallado cubre la zona del rostro y existe mayor flexibilidad puesto que nuestra propuesta hace un seguimiento del rostro permitiendo mayor libertad de movimiento, así se observa en la figura 6.

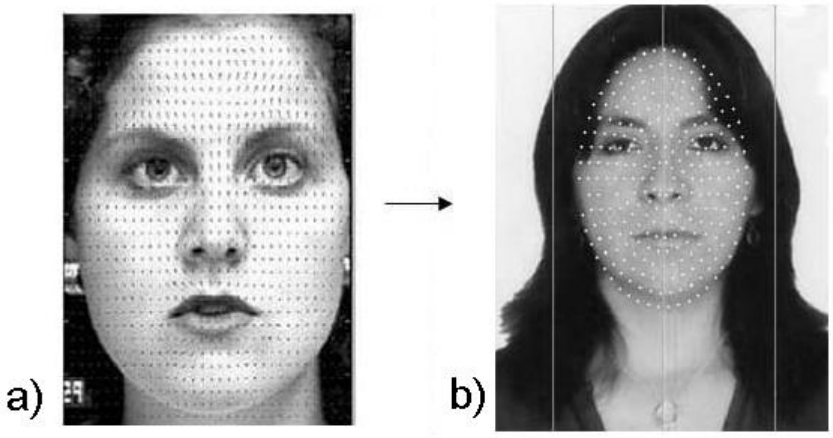

Figura 5: Enmallado tradiconal frente a nuestra propuesta.

Figura 6: Enmallado de rostro de nuestra propuesta.

4.4. Cálculo del flujo óptico

Marcadas ya las zonas de interés, se inicia el proceso de cálculo de flujo óptico, para lo cual usamos el algoritmo de Lucas & Kande(B.D.Lucas and Kanade, 1981). Mediante el cual, sabremos las nuevas posiciones de nuestros pixels marcados(estimación de movimiento aparente), obteniendo así vectores de 216 atributos, los cuales contienen información en módulo y dirección de nuestros pixels marcados.
Estos vectores intercalados y en grupos de 6 por segundo(nuestra cámara adquiere a 30 frames/seg.), son etiquetados y almacenados en un repositorio, para posteriormente entrenar el clasificador (Hsu et al., 2008). Las muestras de cada expresión son las transiciones de la estado facial normal o neutro hacia los demás estados como son alegría, sorpresa, ira, tristeza miedo y asco.

4.5. Clasificación de la expresión facial

Entrenado el clasificador (Hsu et al., 2008), con las muestras que constan de grupos de 6 vectores, esta cantidad fue estimada empíricamente, pudiendo observar lo siguiente: La transición de una expresión facial a otra varia en tiempo, se tiene por ejemplo el cambio de neutro a feliz demora mas que de neutro a sorpresa. Se muestra un resumen de las evaluaciones por expresión en la figura 7.

Expresion	Alegria	Sorpresa	Enojo	Tristeza	Miedo	Asco
						
Neutro	8	10	11	40	21	12
Alegria	74	10	9	10	10	10
Sorpresa	8	65	8	10	20	13
Enojo	3	2	58	5	8	10
Tristeza	1	1	5	20	6	8
Miedo	2	5	2	5	23	5
Asco	2	5	5	8	10	40
Frames	98	98	98	98	98	98
Ratio	75.51	66.33	59.18	20.41	23.47	40.82

Figura 7: Resumen de resultados obtenidos, con nuestra propuesta.

Los items Frames y Ratio nos explican el proceso de medición de nuestra propuesta, por ejemplo en el caso de Alegria se tomaron un promedio de 98 frames, de los cuales nuestra propuesta detectó 8 frames en estado Neutro, 74 Alegre, 8 Sorpresa, 3 Enojo, 1 Tristeza, 2 Miedo y 2 Asco, logrando un ratio de 75.51 %.

5. Conclusiones

La consideración de la geometría facial para enmallar el rostro que contienen la UAs, marcando los pixel candidatos para el cálculo del flujo óptico, permite reducir la dimensión de los vectores característicos de las expresiones faciales.

La estimación de expresiones faciales no solo debe considerar el cambio de las zonas que incluyen las UAs, sino también los tiempos de transición de una expresión a otra.

Las pruebas realizadas del modelo nos permitieron observar que las expresiones de alegría, sorpresa, enojo y asco, son las más detectables siendo las otras: Tristeza y miedo poco perceptibles, la tristeza por que su transición es lenta y el miedo por ser una mezcla de expresiones.

Referencias

Arias, M. G. (2005). *Valoración del efecto de diferentes fuentes de información sobre el reconocimiento de emociones en un contexto conversacional*. PhD thesis, Universidad de Chile, Facultad de Ciencias Sociales.

Asensio, G., Gonzalez, P., Platero, C., J.M.Poncela, J.Sanguino, and M.C.Tobar (2007). Métodos multimalla en problemas lineales de flujo óptico.

Ballard, D. and Brown, C. (1982). *Computer Vision*. Prentice-Hall.

B.D.Lucas and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *in Proceedings of Imaging Understanding Workshop*, pages 121–130.

Black, M. J., Yacoob, Y., Jepson, A. D., and Fleet, D. J. (1997). Learning parameterized models of image motion. In *IEEE Proc. Computer Vision and Pattern Recognition, CVPR-97*, pages 561–567, Puerto Rico.

Chang, J. Y. and Chen, J. L. (2001). Automated facial expression recognition system using neural networks. *Journal of the Chinese Institute of Engineers*, 24(3):345–356.

Cohen, I., Sebe, N., Garg, B. A., Chen, C. L. S., and A, T. S. H. (2003). Facial expression recognition from video sequences: Temporal and static modeling. In *Computer Vision and Image Understanding*, pages 160–187.

Cohn, J. F., Ambadar, Z., and Ekman, P. (2005). Running head: Facial action coding system. University of Pittsburgh and University of California at San Francisco. Observer-Based Measurement of Facial Expression with the Facial Action Coding System.

Ekman, P. (1993). Facial expression and emotion. 48(4):384–392.

Ekman, P. and Friesen, W. (1978). The facial action coding system: A technique for measurement of facial movement.

Essa, I. (1995). Analysis, interpretation, and synthesis of facial expressions.

Fagertun, J. and Stegmann, M. B. (2005). The imm frontal face database. *Informatics and Mathematical Modelling*.

Horn, B. (1986). *Robot Vision*. Ed Mc Graw-Hill.

Hsu, C. W., Chang, C. C., and Lin, C. J. (2008). *A Practical Guide to Support Vector Classification*. Departament of Computer Science, National Taiwan University, Taipei 106, Taiwan.

Izaguirre, A. Z., Gonzales, J. L. M., and Lopez, L. A. (2001). Determinación del movimiento a partir de secuencias de imágenes.

Kobayashi, H. and Hara, F. (1992). Recognition of mixed facial expressions by neural network. *IEEE International workshop on Robot and Human Communication*, pages 387–391.

Romero, I. O. (2002). Detección y seguimiento de objetos en imágenes infrarrojas usando información temporal. Master's thesis, Instituto nacional de astrofisica óptica y electrónica.

Rosenblum, M., Yacoob, Y., and Davis, L. S. (1994). Human emotion recognition from motion using a radial basis function network architecture. Technical Report CS-TR-3304.

Viola, P. and Jones, M. (2002). Robust real-time object detection. *International Journal of Computer Vision - to appear*.

Wu, T. F., Lin, C. J., and Weng, R. C. (2003). Probability estimates for multi-class classification by pairwise coupling.

Yacoob, Y. and Davis, L. (1994). Recognizing Human Facial Expressions. Technical Report CS-TR-3265.