

Alineación facial para reconocer emociones humanas

Ing. M. Jiménez-Vázquez¹, Dr. J.A. Montero-Valverde², Dra. M. Martínez-Arroyo³, MTI. J. Carranza-Gómez⁴

Resumen- La expresión facial es una de las formas visuales de comunicación que más utilizan las personas para su interacción. El reconocimiento automático de las emociones humanas a través del análisis facial es una manera natural de interacción en una amplia variedad de aplicaciones. La alineación facial es una etapa importante en este proceso, ya que de ahí derivan las características utilizadas durante el aprendizaje de los modelos utilizados para la clasificación. En este trabajo se muestran las primeras etapas realizadas en el proceso de reconocer las emociones humanas resaltando la importancia de la detección facial y su posterior alineación como etapas relevantes del proceso.

Palabras clave- alineación fotométrica, alineación geométrica, detección facial, reconocimiento de patrones, visión artificial.

Introducción

En la sociedad humana, las emociones desempeñan un papel muy importante. Las emociones son los indicadores más confiables para socializar. Ellas determinan la manera de pensar, de comportarnos y de comunicarnos con los demás [Picard y otros] [8]. [Fernández y otros][13], Por lo tanto, es comprensible pensar que las nuevas generaciones de máquinas deben tener la habilidad de entender las emociones humanas [Ekman] [1], [Pantic y otros] [2], para poder interactuar de forma natural y flexible entre los humanos y las computadoras [Cowie y otros] [3]. Por ejemplo, resulta de gran utilidad que un vehículo interactúe de alguna manera con el conductor cuando reconozca que no está concentrado en la conducción o que éste está aburrido, otro ejemplo de aplicación sería en la utilización de una herramienta de enseñanza que identifique como interactuar con el usuario cuando detecte que éste no está interesado o poniendo atención en el tema [Cowie y otros] [3]. Una etapa importante en este proceso consiste en identificar de manera confiable el rostro humano y determinar algunas regiones de interés, tal como los ojos, con el fin de realizar la alineación del rostro propuesta por [Kazemi y otros] [14] como un paso previo a la extracción de características que determinen la expresión facial. En este trabajo se describe la detección del rostro y la alineación del mismo aplicando dos enfoques: a) alineación geométrica y b) alineación fotométrica. Así mismo, se muestran los resultados obtenidos de esta etapa. El resto del artículo está organizado como se indica. En la Sección II se presenta la metodología utilizada para la detección facial; en la Sección III se muestran los resultados parciales; en la Sección IV se indican las conclusiones; el trabajo futuro se menciona en la Sección V; finalmente las referencias son listadas en la Sección VI.

Metodología para el reconocimiento del rostro

La metodología utilizada en este trabajo con el fin de reconocer de manera automática las seis emociones humanas básicas [1] [2], se integra de tres etapas como se muestra en la figura:

1. Preprocesamiento de la imagen.
2. Extracción y selección de características.
3. Clasificación.

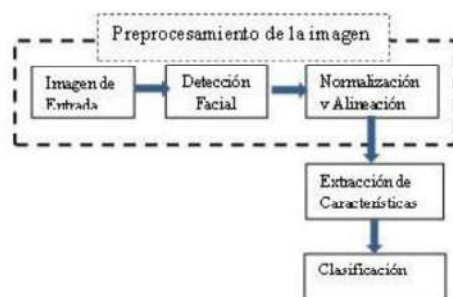


Figura 1.-Etapas para el Reconocimiento de expresiones faciales

¹ Jiménez-Vázquez es alumno de la Maestría en Sistemas Computacionales del I. T. de Acapulco.
mario_jv@hotmail.com

A continuación se va a describir la que se encuentra marcada con líneas punteadas en la figura 1.

Preprocesamiento de la imagen. Esta etapa comprende la captura de la imagen por medio de algún dispositivo electrónico, la detección del rostro y la alineación del mismo.

Detección del Rostro.-La detección del rostro se lleva a cabo aplicando el algoritmo descrito por Viola y Jones [7]. Este algoritmo utiliza una imagen integral para extraer características de forma rápida y precisa, debido a que no trabaja directamente con los valores de intensidad de los píxeles, sino que lo hace a través de una imagen acumulativa que se va formando a través de operaciones básicas que se realizan a medida que se va deslizando en la imagen. La figura 2.a muestra la aplicación de este proceso con el fin de obtener la imagen integral a partir de la Imagen original ($Im(x,y)$). La imagen integral realiza un desplazamiento de izquierda a derecha y de arriba hacia abajo en la imagen realizando la suma de los píxeles en la localización x,y , a medida que se desplaza (figura 2b), con el fin de obtener la imagen integral aplicando la expresión (1).

En la figura 2.b muestra unos filtros que son utilizados para la extracción de características en una imagen. La suma de los píxeles que se encuentran dentro de los rectángulos blancos se sustraen de la suma de píxeles en los rectángulos grises. Las características de dos rectángulos se muestran en la figura 2b.



Figura 2. Imagen integral y convolución de filtros en el rostro

$$II(X,Y) = \sum_{x' \leq x; y' \leq y} Im(X',Y') \quad 1$$

Donde:

$II(X,Y)$.-Representa la imagen integral
 $Im(X',Y')$.-Representa la imagen original

Condiciones:

x' .-Menor o Igual a x
 y' .-Menor o Igual a y

Extracción de características para facial.- Cuando se trabaja con imágenes, las características de cada objeto se extraen seleccionando la región de interés en la figura y representando ésta a través de patrones que la describen de cierta manera al aplicar unas funciones para su procesamiento. La extracción de características es un paso importante en el reconocimiento de patrones debido a que cuando éstas son procesadas proporcionan atributos importantes que son utilizados para representar a los objetos de tal manera que se pueden agrupar en una determinada clase [Schapire y otros] [9].

En el reconocimiento facial la extracción de características es aplicada a la imagen utilizando los filtros con bases Haar. Estos filtros son calculados eficientemente sobre la imagen integral y son selectivos en la orientación espacial y la frecuencia, además permiten ser modificados en escala y orientación de acuerdo a las necesidades requeridas, es decir, si se requiere agrandar la imagen se utiliza un múltiplo y si se requiere minimizar se utiliza un divisor en la escala. En el caso de la detección del rostro, se utilizan los filtros con bases Haar. Cuando se aplican los filtros Haar, éstos realizan una codificación de diferencia de intensidades en la imagen y no en los píxeles que contiene debido a que éstos trabajan con valores (0,255), generando características de contornos, puntos y líneas, mediante la captura de contraste entre las regiones donde se apliquen los filtros, como se muestra en la figura 2.b.

Alineación del Rostro.-La alineación es un método que consiste en estandarizar el tamaño, el color, y algunas veces los puntos de referencias que contiene la imagen que se va a someter a este proceso, sin provocar en ella distorsiones que la alteren. La normalización en una imagen puede ser de tipos vertical y horizontal.

En este trabajo se realiza la normalización a través de la técnica de utilizada por [Tan y Trigs] [12] y la técnica de ecualización de histogramas utilizada por [González y otros] [15], que consiste en poner los píxeles en un nivel de gris, eliminando el ruido y estandarizando su longitud y tamaño, para su tratamiento.

Para realizar el acondicionamiento, normalización y alineación de la cara, se debe seguir un procedimiento de mediante el uso de algunas técnicas de transformación de la imagen, para tener un mejor resultado al momento de manipular los píxeles que componen, por lo que sólo mencionaremos a continuación los métodos que serán utilizados en este trabajo.

Para llevar a cabo la alineación facial existen dos enfoques: a).-**alineación geométrica [14]**, que toma los puntos de las referencias de la cara (obtenidos por el detector de los puntos de referencias) y los ajusta para que queden alineados horizontalmente. En la **alineación fotométrica[12]**, se trata de ajustar el brillo de la imagen detectada, mediante el uso de una técnica, para obtener un mejor resultado en el manejo de los píxeles. Se van a considerar tres enfoques diferentes de normalización con el fin de seleccionar el que ofrezca mejores resultados con el clasificador. La alineación fotométrica ayudará a eliminar el ruido provocado por el exceso o falta de brillo en la imagen, dejándola de manera homogénea en la intensidad de gris. Después de pasar por el proceso de normalización se va a probar con el clasificador para ver cuál de estas tres normalizaciones proporciona mejores resultados.

- a).-Ecualización del histograma.
- b).-Método de Tan y Trigs.
- c).-La conversión a escala de grises de la imagen sin utilizar una técnica.

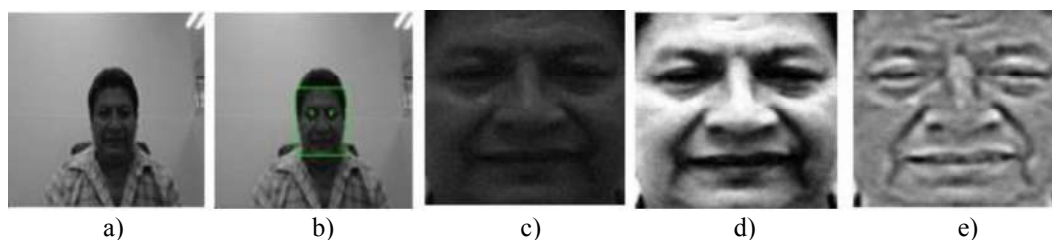


Figura 3.- Detección y normalización del rostro aplicando diferentes técnicas

Una vez que se tiene la imagen en escala de grises, el siguiente paso a seguir es, llevar a cabo la localización del rostro en la imagen. La posición precisa del rostro en la imagen es importante para llevar a cabo el análisis facial. En este punto se requiere buscar la región del rostro en la imagen; para esto, se debe normalizar la imagen de la cara, se pueden seguir diferentes reglas, como la distancia entre las pupilas, la posición de la nariz, o la distancia entre las comisuras de los labios. También se debe definir el tamaño de las imágenes y la gama de colores. La figura 3.c, muestra un rostro delimitado sin alinear.

Delimitación de la imagen.- Cuando se captura una imagen están presentes muchos objetos que contienen información innecesaria que podría alterar el contenido de ésta, en este caso, se debe delimitar el contorno facial de la imagen (rostro) para observar que están presentes todos y cada uno de los componentes faciales (ojos, boca, nariz, cejas, frente). Esto se realiza con la finalidad de dejar solamente la cara en la imagen como se muestra en la figura 3.c, dejando a un lado los otros elementos componentes del rostro como: orejas, pelo, cuero cabelludo y cualquier otro objeto (aretes o algún tatuaje), que pueda estar presente en el rostro durante el procesamiento.

De esta manera se precedió a eliminar los píxeles que rodean el contorno facial dejando solamente los píxeles que contienen el rostro, con se muestra en la figura 2.c, 2.d, 2e.

Este punto es muy importante debido a que la imagen debe estar completamente despejada de cualquier objeto que pueda proporcionar información inadecuada que altere el contenido de la misma en el siguiente proceso.

A continuación se detallan otros métodos donde se aplican las otras técnicas de normalización. Estas técnicas se aplican para tener un mejor manejo de los píxeles que componen la imagen.

Aplicación de la Técnica de Ecuación del Histograma.-En la transformación, todos los píxeles de un mismo nivel de gris se transformarán a otro nivel de gris, y el histograma se distribuirá en todo el rango disponible separando en lo posible las ocupaciones de cada nivel. El resultado de la ecuación maximiza el contraste de una imagen sin perder información de tipo estructural, esto es, conservando su entropía.

En la figura 3.d, se muestra una imagen aplicando la técnica de normalización a través de ecuación de histograma. Es notorio que ésta es una de las mejores técnicas para realizar la extracción de características en la normalización de imágenes

Aplicación de la Técnica de Tan y Trigs.-El método de Tan y Trigs se basa prácticamente en aplicar dos enfoques principales en la normalización de las imágenes, estos métodos son:

Descriptores basados en características geométricas y

Descriptores basados en la apariencia.

La técnica utilizada por Tan y Trigs se centra en la utilización de los patrones binarios locales (LBP) y sus generalizaciones [Tan y Trigs] [12]. Los Patrones Binarios Locales (LBP), son un descriptor no paramétrico de textura de imagen local computacionalmente eficiente [Fasel y otros] [6].

Las características de Los Patrones Binarios Locales (LBP) son invariantes a los cambios de nivel de grises monotónicos por diseño y, por lo tanto, generalmente se considera que no requieren preprocesamiento de imágenes antes de su uso. La fiabilidad de los patrones binarios locales (LBP), es que disminuye significativamente bajo grandes variaciones de iluminación. Otra limitación que tiene el uso de Los Patrones Binarios Locales, es su sensibilidad al ruido aleatorio y de cuantificación en regiones de imagen casi uniformes tales como la frente y las mejillas.

En la figura 3.e, se muestra una imagen aplicando la técnica de normalización a través de la aplicación de Tan y Trigs, basada en los patrones locales binarios.

Obtención facial.- La técnica de *Boosting* fue introducida por [Schapire y otros] [9], este es un método de clasificación que combina varios clasificadores básicos para formar un único clasificador más complejo y preciso. La idea se basa en la afirmación de que varios clasificadores sencillos que se desarrollan con los filtros Haar, cada uno de ellos con una precisión ligeramente superior a una clasificación aleatoria, pueden combinarse para formar un clasificador de mayor precisión, siempre y cuando se disponga de un número suficiente de muestras de entrenamiento. La aplicación de clasificadores en cascada ha permitido obtener buenos resultados, estos son reportados en los trabajos realizados [Viola y Jones][7], [Papageorgiou y otros][10], [Castrillón y otros][11].

Para aplicar la técnica de *boosting* primero se debe establecer un algoritmo de aprendizaje sencillo (clasificador débil o base), que será llamado repetidas veces para crear diversos clasificadores base. Para el entrenamiento de los clasificadores base se emplea, en cada iteración, un subconjunto diferente de muestras de entrenamiento y una distribución de pesos diferente sobre las muestras [9]. Entre mayor sea el número de muestras de entrenamiento mayor será la precisión en la clasificación de las características. Finalmente, estos clasificadores base se combinan en un único clasificador que se espera sea mucho más preciso que cualquiera de los clasificadores base por separado.

Resultados Parciales

A continuación se muestran los resultados parciales de la detección del rostro a través de una imagen facial, así mismo, a estas imágenes se le aplican la técnica de ecuación de histogramas y la técnica de Tan y Trigs antes descritas para llevar a cabo el manejo de los píxeles que integran la imagen, con la finalidad de tener una

estandarización en cada una de ellas. Si no se aplicara una de estas técnicas a las imágenes, se corre el riesgo de tener una información poco confiable debido a que las imágenes podrían presentar ruido, que manifiesta a través de la falta o exceso de brillo, es decir, podrían estar muy oscuras o muy claras.

Es de gran importancia aplicar la técnica que mejor visualice la imagen para tener los mejores resultados en la detección de los estados emocionales en dicha imagen.



Figura 4.-Imagen original convertida a escala de grises y detección del rostro



Figura 5.- Normalización y alineación del rostro aplicando las técnicas descritas

Conclusiones

La combinación de la técnica de Boosting con los filtros de base Haar para realizar la detección del rostro en una imagen digital para llevar a cabo la extracción de características faciales, proporcionan un alto porcentaje para la detección requiriendo bajo tiempo de procesamiento en una gran cantidad de muestras.

La reducción del tiempo empleado en la localización de los puntos de referencia, se debe a que el método desarrollado por Viola y Jones emplea el desplazamiento de la imagen integral en áreas de interés eliminando áreas que no contienen información relevante para el procesamiento.

La aplicación de diferentes técnicas de normalización y alineación del rostro, proporcionan una visión más amplia para el procesamiento de imágenes faciales, obteniendo mejores resultados en la detección de rostros.

Teniendo la imagen facial después de haber aplicado la técnica de ecualización del histograma o la técnica de Tan y Trigs, la siguiente etapa es extraer las características aplicando el algoritmo de la Máquina de Soporte de Vectores (MVS). Este algoritmo analizará las características detectadas en el rostro, analizando la suma de los vectores que componen el rostro, los cuales serán sometidos al proceso de aprendizaje por medio de muestras que alimentarán el algoritmo, para reconocer cada uno de los estados emocionales.

Referencias Bibliográficas

- [1] P. Ekman, "Emotion in the Human Face". Cambridge Univ. Press, 1982.
- [2] M. Pantic, L.J.M. Rothkrantz, "Automatic analysis of facial expressions: the state of the art", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424-1445, Dec. 2000.
- [3] R. Cowie, E. Douglas-Cowie, K. Karpouzis, G. Caridakis, M. Wallace and S. Kollias, "Recognition of Emotional States in Natural Human-Computer Interaction," in *Multimodal User Interfaces*, Springer Berlin Heidelberg, 2008.
- [9] P. Ekman and W.V. Friesen, "Constants across cultures in the face and emotions," *J. Personality Social Psychology*, vol. 17, no. 2, pp. 124-129, 1971.
- [4] P. Ekman and W.V. Friesen, "Facial Action Coding System (FACS): Manual". Palo Alto: Consulting Psychologists Press, 1978.
- [5] MPEG Video and SNHC, "Text of ISO/IEC FDIS 14 496-3: Audio," in *Atlantic City MPEG Mtg.*, Oct. 1998, Doc. ISO/MPEG N2503.

- [6] B. Fasel and J. Luttin, "Automatic Facial Expression Analysis: a survey," *Pattern Recognition*, vol. 36, no. 1, pp. 259-275, 2003.
- [7] VIOLA, Paul y JONES, Michael. *Robust Real-Time Face Detection*. *International Journal of Computer Vision*, Vol.57, No.2, pp 137-154, 2004.
- [8] R.W. Picard, E. Vyzas, J. Healey, "Toward Machine Emotional Intelligence: Analysis of Affective Physiological State," in *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175- 1191, 2001.
- [9] SCHAPIRE, R and FREUND, Y. *A decision theoretic generalization of on-line learning and application to boosting*. AT&T Bell Laboratories. USA, 1995.
- [10] PAPAGEORGIOU,C., OREN,M. and POGGIO,T. *A general frame-work for object detection*. Proceedings of International Conference on Computer Vision, enero de 1998, Bombay, India.
- [11] CASTRILLÓN, M., DÉNIZ, O., GUERRA, C., and HERNÁNDEZ, M. "*ENCARA2: Real-time detection of multiple faces at different resolutions in video streams*". *ELSEVIER, Journal Visual Communication and Image Representation*. Article in press (2007).
- [12] Xiaoyang Tan and Bill Triggs "Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions". INRIA & Laboratoire Jean Kuntzmann, 655 avenue de l'Europe, Montbonnot 38330, France.
- [13] J.L. Fernández, J.M. Gómez, V.F. Muñoz, I. García, D. Melgar, C. Vara and A. García, "Human-Machine Interface Evaluation in a Computer Assisted Surgical System," in *IEEE International Conference on Robotics and Automation*, pp. 231-236, New Orleans, 2004.
- [14] Vahid Kazemi and Josephine Sullivan, CVPR 2014. One Millisecond Face Alignment with an Ensemble of Regression Trees.
- [15] Gonzalez, R. C. , and Woods, P., *Digital Image Processing*. Addison Wesley, 2002.