ISSN 1870-4069

Reconocimiento de expresiones faciales con base en la dinámica de puntos de referencia faciales

E. Morales-Vargas, C.A. Reyes-Garcia, Hayde Peregrina-Barreto

Instituto Nacional de Astrofísica Óptica y Electrónica,
División de Ciencias Computacionales, Tonantzintla, Puebla,
México

emoralesv@inaoep.mx, kargaxxi@inaoep.mx, hperegrina@inaoep.mx

Resumen. Las expresiones faciales permiten a las personas comunicar emociones, y es prácticamente lo primero que observamos al interactuar con alguien. En el área de computación, el reconocimiento de expresiones faciales es importante debido a que su análisis tiene aplicación directa en áreas como psicología, medicina, educación, entre otras. En este articulo se presenta el proceso de diseño de un sistema para el reconocimiento de expresiones faciales utilizando la dinámica de puntos de referencia ubicados en el rostro, su implementación, experimentos realizados y algunos de los resultados obtenidos hasta el momento.

Palabras clave: Expresiones faciales, clasificación, máquinas de soporte vectorial,modelos activos de apariencia.

Facial Expressions Recognition Based on Facial Landmarks Dynamics

Abstract. Facial expressions allow people to communicate emotions, is practically the first thing that we observe when interacting with someone. In computer science, facial expressions recognition is important because their analysis has direct application in areas such psychology, medicine, education, among others. In this paper is presented the design process of a facial expressions recognition system which uses facial landmarks dynamics, its implementation, experiments performed and some of the results obtained until now.

Keywords: facial expressions, classification, support vector machines, active appearance models.

1. Introducción

Existen diversas áreas en las que las expresiones faciales son estudiadas, entre ellas se incluyen la psicología, neurociencia, educación, o sociología [15,19]. Existe

una fuerte evidencia que soporta el hecho de que existen siete emociones básicas que tienen asociadas una expresión facial, que pueden ser: enojo, desprecio, disgusto, miedo, felicidad, tristeza o sorpresa [17,8,2,16].

Fig. 1. Expresiones faciales básicas, de izquierda a derecha: enojo, disgusto, desprecio, felicidad, miedo, tristeza y sorpresa. Imágenes tomadas de la base de datos CK+ [12].

Se han propuesto varias metodologías para el reconocimiento de expresiones faciales con un enfoque estático. Los enfoques estáticos utilizan descriptores locales en un solo recuadro de una secuencia para extraer características. Un operador utilizado en varios trabajos es Patrones Binarios Locales (LBP) y sus mejoras, el cual se utiliza para describir la textura en una imagen y posteriormente utilizan como clasificador Máquinas de Soporte Vectorial (SVM) [18,14,20].

Los enfoques dinámicos para el reconocimiento de expresiones faciales toman como referencia la diferencia entre estado neutral y la representación de una expresión facial, extrayendo como característica discriminatoria la dinámica del rostro. [12] utiliza el desplazamiento de puntos de referencia faciales obtenidos a través de Modelos Activos de Apariencia (AAM), y el trabajo propuesto por [9] utiliza los ángulos que se forman en el rostro para realizar el reconocimiento.

En este artículo nos enfocamos en la generación de un vector de características que reconozca expresiones faciales de manera robusta en secuencias de imágenes que sea lo mas compacto y simple posible con respecto a trabajos presentados anteriormente, capturando la dinámica del rostro mediante el desplazamiento y la dirección de puntos de referencia faciales.

2. Sistema propuesto

El sistema propuesto utiliza la base de datos CK+ [12] la cual contiene un conjunto de secuencias de imágenes en las que se utilizó AAM para estimar la forma del rostro o puntos de referencia faciales, que se define como un conjunto de coordenadas. La forma del rostro del estado neutral y de la expresión facial

se alinean para reducir el ruido y posteriormente se extraen características que describen la dinámica de los puntos de referencia faciales. Se utilizó SVM para clasificar las expresiones faciales. En la Fig. 2 se puede observar el sistema completo y en las secciones siguientes se describe cada una de las etapas.

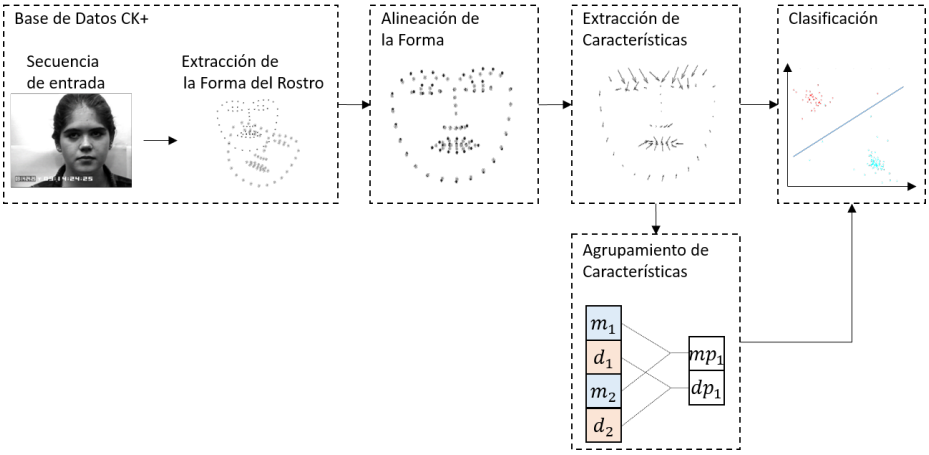


Fig. 2. Metodología general propuesta, base de datos tomada de [12].

2.1. Alineación de la forma del rostro

Al utilizar los puntos de referencia faciales existen factores que se deben tomar en cuenta, es necesario remover el efecto del tamaño, orientación y ubicación de las coordenadas para reducir el ruido que se introduce al sistema [3].Uno de los métodos utilizados en la literatura para alinear la forma del rostro es el análisis de Procrustes [4,12], en nuestro estudio se utilizaron transformaciones afines para reducir la orientación y la variación espacial, y posteriormente los valores de las coordenadas se normalizan en un rango entre 0 y 1 en términos del estado neutral para evitar perder las deformaciones causadas por el movimiento del rostro.

2.2. Extracción de características

Una vez que se ha alineado la forma del rostro del estado neutral de una secuencia con su expresión facial es posible extraer información correspondiente a la dinámica de los puntos de referencia faciales de manera adecuada. Para hacer distinción entre las coordenadas en una secuencia entre el estado neutral y su expresión facial, re-definiremos la forma del rostro **s** de la siguiente manera:

- Estado neutral: $sn = [xn_1, yn_1, xn_2, yn_2..., xn_n, yn_n]$.

(a) Forma sin alinear. **(b)** Procrustes. **(c)** Transformación afín.

Fig. 3. En a(a) se puede observar que la forma del rostro en estado neutral y la representación facial pueden estar afectadas por ubicación, orientación y tamaño, en (b) se puede observar el resultado que se obtiene al realizar el análisis de Procrustes a ambas formas y en (c) se observa el resultado al realizar un conjunto de transformaciones afines para así preservar información del movimiento de las cejas y mandíbula que se puede perder con el análisis de Procrustes.

- Expresión facial: $se = [xe_1, ye_1, xe_2, ye_2..., xe_n, ye_n]$.

La primera etapa de la caracterización consiste en obtener el desplazamiento horizontal y vertical del rostro, para esto se extrae el estado neutral a la expresión facial, dando como resultado el desplazamiento, lo cual se realiza con las ecuaciones 1 y 2:

$$\Delta x_i = xe_i - xn_i, \tag{1}$$

$$\Delta y_i = ye_i - yn_i. \tag{2}$$

Con el desplazamiento horizontal y vertical de los puntos de referencia es posible calcular la magnitud del movimiento con la ecuación 3 y la dirección con la ecuación 4.

$$m_i = \sqrt{(\Delta x_1)^2 + (\Delta y_1)^2)}, \tag{3}$$

$$d_i = \tan^{-1}(\frac{\Delta x_i}{\Delta y_i}). \tag{4}$$

La caracterización base consiste en la concatenación de la intensidad y magnitud del movimiento de los puntos de referencia faciales, la cual se define de la siguiente manera: $c = [m_1, d_1, m_2, d_2, ..., m_n, d_n]$.

2.3. Agrupación

Diversos trabajos utilizan el agrupamiento de características para combinar resultados de varios descriptores o para agrupar características de una región. Algunos métodos populares que utilizan la agrupación son la Transformada de Características Invariante a Escala (SIFT) [11], el Histograma de Gradientes Orientados (HOG) [5], entre otros. La agrupación ayuda a producir una representación más estable de un grupo de características inestables [10]. De manera general, el agrupamiento de características consiste en transformar la

representación de características en una nueva representación más útil y estable que preserva solo información relevante [1]. La dificultad al agrupar radica en identificar cuales características corresponden a cada grupo.

Se redujo el vector de características mediante operaciones de agrupación $f = \{f_a(x), f_m(x)\}$ donde $x \subset c$ a un nueva nueva representación con 22 valores. El criterio de agrupación consiste en seleccionar y agrupar las características

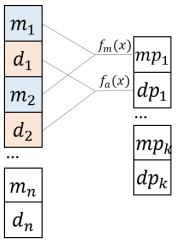


Fig. 4. Diagrama de agrupación, del lado izquierdo se puede observar el vector de características sin agrupar, varias características se agrupan en una sola mediante y se concatenan en una nueva representación mediante operaciones de agrupación.

correspondientes a las diferentes áreas del rostro: interior de cejas, exterior de cejas, parpados, nariz, labio superior, labio inferior, esquina derecha del labio, esquina izquierda del labio y mandíbula [7]:

$$f_a(x) = \frac{1}{|x|} \sum_{i=1}^{|x|} x_i, \tag{5}$$

$$f_m(x) = max(x). \tag{6}$$

3. Datos

Se realizaron experimentos en la base de datos CK+ [12] la cual contiene 327 secuencias de imágenes. En la Tabla 1 se puede observar el número de secuencias para cada expresión facial. Cada secuencia comienza en estado neutral y termina con la representación de una expresión facial. Un juez experto, manualmente codificó las secuencias de imágenes mediante el Sistema de Codificación Facial (FACS) [6], asignó una etiqueta la cual indica qué expresión facial se percibe en la secuencia. Para cada imagen, los autores obtuvieron un conjunto de 68 coordenadas que describen la forma del rostro mediante AAM, un algoritmo basado en gradiente descendente propuesto en [13]. La forma del rostro es un conjunto de **n** coordenadas, donde cada coordenada pertenece a un vértice de la forma del rostro, la cual se encuentra definida por $s = [x_1, y_1, x_2, y_2..., x_n, y_n]$.

Tabla 1. Frecuencia de las muestras para cada expresión facial de la base de datos CK+ [12].

Emoción	N
Enojo(En)	45
Disgusto(Dis)	59
Desprecio(Des)	18
Felicidad(Fel)	69
Miedo(Mi)	25
Tristeza(Tris)	28
Sorpresa(Sor)	83

4. Experimentos y resultados

Las pruebas del sistema propuesto fueron hechas utilizando la base de datos extendida de Cohn-Kanade (CK+) [12], la cual contiene secuencias de imágenes de personas actuando las siete expresiones faciales básicas (Fig. 1). Debido a las clases desbalanceadas y siguiendo el marco de referencia propuesto en [12], el sistema propuesto se validó utilizando la estrategia de dejar un sujeto fuera (*leave one out*) y se presentan la exactitud promedio y la exactitud ponderada por expresión facial para poder realizar un análisis de las consecuencias que esto conlleva. El clasificador utilizado fué Máquinas de Soporte Vectorial.

En la Tabla 2 se presenta la matriz de confusión que muestra los resultados correspondientes utilizando la dinámica de puntos de referencia faciales, sin agrupar características. El porcentaje de clasificación para esta configuración es de 93.5 % y corresponde al promedio pesado de la diagonal. En la Tabla 3 se muestra la matriz de confusión con los resultados correspondientes a la clasificación cuando se agrupan las características, con esta configuración se obtuvo un 92.3 % de exactitud.

Tabla 2. Matriz de confusión correspondiente al reconocimiento de expresiones faciales con un enfoque dinámico sin agrupación de características.

		Predicción						
		En	Des	Dis	Mi	Fel	Tris	Sor
Valor real	En	**91.1**	0	6.7	0	0	2.2	0
	Des	0	**83.3**	0	0	0	16.7	0
	Dis	3.4	1.7	**94.8**	0	0	0	0
	Mi	0	0	0	**88.8**	8	4	0
	Fel	0	1.4	0	1.4	**98.6**	0	0
	Tris	3.6	0	0	0	0	**96.4**	0
	Sor	0	1.2	0	4.8	0	1.2	**92.8**

Los porcentajes de clasificación individuales para cada expresión facial muestran que nuestro sistema en sus dos configuraciones presenta un desempeño

Tabla 3. Matriz de confusión correspondiente al reconocimiento de expresiones faciales con un enfoque dinámico y agrupación de características.

		Predicción						
		En	Des	Dis	Mi	Fel	Tris	Sor
Valor real	En	**91.1**	0	4.4	0	0	4.4	0
	Des	5.6	**83.3**	0	0	0	11.1	0
	Dis	3.4	0	**96.6**	0	0	0	0
	Mi	0	0	0	**84**	12	0	4
	Fel	0	1.4	0	1.4	**97.1**	0	0
	Tris	10.7	3.6	0	3.6	0	**82.1**	0
	Sor	0	1.2	0	3.6	0	1.2	**94**

aceptable con respecto a los trabajos relacionado con vectores de características que subjetivamente se han descrito como pequeños.

5. Discusión

En la tabla 4 se presenta una comparación entre los trabajos fuertemente relacionados con nuestra propuesta.[12] comienza con una etapa de normalización en la cual se extrae el ruido de la ubicación, escala y rotación de lo puntos de referencia faciales utilizando la superposición o análisis de Procrustes. Con estos datos se extrae la Forma Normalizada de Similaridad ($SPTS$) la cual se refiere a la forma del rostro después de la normalización del estado neutral, obteniendo un vector con 136 valores que describe el desplazamiento de los puntos de referencia del rostro.

Por otra parte, [9] propone un descriptor basado en capturar los cambios de 560 ángulos obtenidos a partir de la combinación entre los 68 puntos de referencia en el rosto. Los autores mencionan que su enfoque es independiente de la pose del rostro debido a que solo se mide la variación del movimiento. El descriptor utilizado puede tener tres valores discretos dependiendo de la magnitud de la diferencia de los ángulos entre la expresión facial y el estado neutral. En el trabajo de [9] también se utilizo la base de datos de [12] por lo que es directamente comparable con la metodología propuesta.

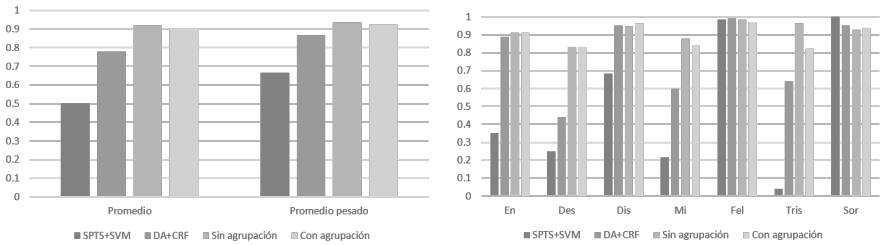
La metodología propuesta hace uso de transformaciones afines en lugar de utilizar Procrustes[12], método de referencia para la normalización de los puntos que describen al rostro. Como se puede observar en la Tabla 4, el uso de dichas transformaciones afines permite conservar mejor la información del movimiento del rostro al comparar con el estado neutral. En los trabajos relacionados se requiere analizar un mayor número de direcciones [9] y el movimiento sólo tiene una descripción horizontal y vertical [12]. La metodología propuesta en este trabajo permite calcular directamente (entre puntos correspondientes) en qué dirección se produjo el movimiento de un determinado punto del rostro y brinda una mejor descripción de dicho movimiento. Lo anterior permite también mejorar la exactitud del reconocimiento de las expresiones.

Tabla 4. Comparación entre diferentes trabajos relacionados.

Trabajo	Tamaño del vector	Exactitud promedio	Exactitud pesada
SPTS+SVM [12]	136	50.4	66.7
DA+CRF [9]	560	78	86.9
Propuesto S.A.	136	92.1	93.6
Propuesto C.A.	22	89.7	92.3

Al realizar la presente investigación se buscó encontrar un vector de características que describa el movimiento del rostro con el menor número de valores posibles, es por eso que se optó por adoptar un esquema de agrupamiento. Si bien comparado con la configuración sin agrupamiento (S. A.), la metodología con agrupamiento (C. A.) no aumenta el porcentaje de clasificación, lo cual puede ser debido a la selección de los puntos de referencia escogidos para representar cada zona de interés del rostro, si reduce la dimencionalidad del vector de características, de 136 valores a 22. Esto permite generar una representación simple y compacta que reduce el procesamiento en comparación con los trabajos relacionados.

En la Fig. 5 se pueden observar la comparación entre nuestro método con sus dos configuraciones: con agrupación y sin agrupación, *SPTS+SVM* [12] y *DA+CRF* [9]. Las clases desbalanceadas afectan el desempeño del clasificador debido a que si en las clases con un número pequeño de muestras ocurre un acierto o error estos afectan la exactitud drásticamente, es por esto que también se presenta el promedio ponderado o pesado por cada expresión facial, el cual asigna un peso al resultado de cada expresión facial dependiendo del número de muestras que se tienen por clase, otra forma de calcular este promedio pesado es obteniendo la exactitud general del clasificador.



(a) Exactitud general. **(b)** Exactitud por expresión facial.

Fig. 5. En (a) se observa la comparación del promedio pesado del porcentaje de reconocimiento entre nuestro método con sus dos variantes, *SPTS+SVM*, y *DA+CRF*. En (b) se observa la comparación del porcentaje de reconocimiento por expresión facial entre las variantes de nuestro método y métodos de la literatura.

6. Conclusiones

En este artículo se presentó el proceso de diseño de un sistema simple y rápido para el reconocimiento de expresiones faciales el cual se basa en la dinámica de puntos de referencia faciales, con los cuales se calcula la magnitud y la dirección del movimiento en el rostro. Se agruparon los resultados del descriptor de movimiento en regiones de interés del rostro y se realizó una comparación entre los resultados obtenidos al clasificar antes y después de la agrupación.

Agrupar los valores de la representación de la dinámica del rostro y clasificar con SVM alcanzó un porcentaje de reconocimiento de 92.3 %, generar una representación mediante agrupación no aumentó el porcentaje de reconocimiento sin agrupar debido a la perdida de información que conlleva el utilizar el enfoque de agrupación empleado pero es importante remarcar que existe una diferencia significativa, debido a que al se agrupan los valores se obtiene una nueva representación compacta y simple que logra discriminar entre las expresiones faciales de una manera similar a su contraparte.

Agradecimientos. El autor E. Morales-Vargas agradece al Consejo Nacional de Ciencia y Tecnología (CONACyT) por al apoyo a esta investigación a través de la beca #702647. Los autores agradecen al proyecto MX14MO06 "Técnicas de análisis y clasificación de voz y expresiones faciales: aplicación a las enfermedades neurológicas en recién nacidos y adultos"del programa ejecutivo de cooperación científica y tecnológica México-Italia financiado por AMEXID de la SRE y el Ministerio de Asuntos Exteriores de Italia.

Referencias

1. Boureau, Y.L., Ponce, J., Lecun, Y.: A Theoretical Analysis of Feature Pooling in Visual Recognition. In: 27th International Conference on Machine Learning, Haifa, Israel (2010)
2. Burrows, A.M., Waller, B.M., Parr, L.A., Bonar, C.J.: Muscles of facial expression in the chimpanzee (Pan troglodytes): descriptive, comparative and phylogenetic contexts. Journal of Anatomy 208(2), 153–167 (Feb 2006), http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2100197/
3. Cohn, J.F., Zlochower, A.J., Lien, J., Kanade, T.: Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. Psychophysiology 36(1), 35–43 (Jan 1999), http://onlinelibrary.wiley.com/doi/10.1017/S0048577299971184/abstract
4. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(6), 681–685 (Jun 2001)
5. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). vol. 1, pp. 886–893 vol. 1 (Jun 2005)
6. Ekman, P., Friesen, W.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press (1978)

7. Ekman, P.: Facial expression and emotion. American Psychologist 48(4), 384–392 (1993)
8. Galati, D., Miceli, R., Sini, B.: Judging and coding facial expression of emotions in congenitally blind children. International Journal of Behavioral Development 25(3), 268–278 (May 2001), http://dx.doi.org/10.1080/01650250042000393
9. Iglesias, F., Negri, P., Buemi, M.E., Acevedo, D., Mejail, M.: Facial expression recognition: a comparison between static and dynamic approaches. In: International Conference on Pattern Recognition Systems (ICPRS-16). pp. 1–6 (Apr 2016)
10. Krig, S.: Feature Learning and Deep Learning Architecture Survey. In: Computer Vision Metrics, pp. 375–514. Springer International Publishing (2016), http://link.springer.com/chapter/10.1007/978-3-319-33762-3_10, dOI: 10.1007/978-3-319-33762-3_10
11. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60(2), 91–110 (Nov 2004), https://link.springer.com/article/10.1023/B:VISI.0000029664.99615.94
12. Lucey, P., Cohn, J.F., Kanade, T., Saragih, J., Ambadar, Z., Matthews, I.: The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops. pp. 94–101 (Jun 2010)
13. Matthews, I., Baker, S.: Active Appearance Models Revisited. International Journal of Computer Vision 60(2), 135–164 (Nov 2004), http://link.springer.com/article/10.1023/B%3AVISI.0000029666.37597.d3
14. Mohammadi, M.R., Fatemizadeh, E.: Fuzzy local binary patterns: A comparison between Min-Max and Dot-Sum operators in the application of facial expression recognition. In: 2013 8th Iranian Conference on Machine Vision and Image Processing (MVIP). pp. 315–319 (Sep 2013)
15. Pantic, M., Rothkrantz, L.J.M.: Automatic analysis of facial expressions: the state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(12), 1424–1445 (Dec 2000)
16. Park, S., Kim, D.: Spontaneous facial expression classification with facial motion vectors. In: 8th IEEE International Conference on Automatic Face Gesture Recognition, 2008. FG '08. pp. 1–6 (Sep 2008)
17. Peleg, G., Katzir, G., Peleg, O., Kamara, M., Brodsky, L., Hel-Or, H., Keren, D., Nevo, E.: Hereditary family signature of facial expression. Proceedings of the National Academy of Sciences 103(43), 15921–15926 (Oct 2006), http://www.pnas.org/content/103/43/15921
18. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on Local Binary Patterns: A comprehensive study. Image and Vision Computing 27(6), 803–816 (May 2009), http://www.sciencedirect.com/science/article/pii/S0262885608001844
19. Zeng, Z., Pantic, M., Roisman, G.I., Huang, T.S.: A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(1), 39–58 (Jan 2009)
20. Zhao, G., Pietikainen, M.: Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(6), 915–928 (Jun 2007)