ARQUITECTURA PARA EL RECONOCIMIENTO DE EMOCIONES BASADO EN CARACTERÍSTICAS FACIALES

Elsa Irene Herrera Santiago

Instituto Tecnológico Superior de Misantla

eihs_1683@hotmail.com

Resumen

La interacción diaria con diferentes seres humanos, es parte del desarrollo integral del ser humano. No solo los gestos faciales codifican nuestro estado de ánimo, sino que las características físicas de nuestra cara nos identifican como individuos únicos en un mundo superpoblado. Pero qué pasaría si diseñáramos una arquitectura capaz de realizar una de las actividades físicas más comunes en nosotros como es el reconocimiento de emociones. En este artículo se presenta la propuesta de una arquitectura capaz de conseguir el reconocimiento de emociones a través de algoritmos especificados a dicho trabajo.

Palabra(s) Clave(s): eigenface, emociones, pca, svm, viola&jones, wavelet.

1. Introducción

Actualmente, el reconocer la emoción en una persona es una actividad cotidiana de los seres humanos, resulta hasta simple mirar a una persona y predecir su comportamiento de acuerdo a la emoción que reconocemos en ella. Esta acción puede ser riesgosa y hasta engañosa, ya que una emoción depende de diferentes factores propios de los humanos como son el envejecimiento, expresiones faciales, cambios de iluminación, puntos de vista inducidos por el movimiento del cuerpo y oclusiones. Sin embargo, lo que para el cerebro resulta ser una actividad rápida y sencilla, para la visión artificial resulta ser una operación mucho más compleja basada en estímulos visuales, detalles y matices de los rostros. Desde el siglo antepasado, ha sido interesante construir un sistema automático para la

identificación de individuos a través de su cara, de tal manera que, en 1889 Sir Galton buscaba fórmulas matemáticas para solucionar el problema.

Es por ello que, en los últimos 20 años de estudio, muchos investigadores establecen el reconocimiento de caras y emociones como un tema lejos de resolverse, ya que el replicar esta conducta humana tan cotidiana parece ser un problema computacional muy complicado. Dicho tema ha dado la pauta para un gran número de artículos publicados en revistas y conferencias enfocados al área de visión artificial.

En esta tesis se propone una técnica para el reconocimiento de emociones creando una arquitectura capaz de contemplar métodos específicos para la solución del problema, es por ello que se basará en el Algoritmo de Viola-Jones para la obtención de componentes de una región especificada (en este caso el rostro humano) tales como cejas, ojos, nariz, boca, etc.

Con lo anterior, se aplica una transformada Wavelet, con la cual se pueden crear a base de Wavelet lineales una o varias Wavelets ortonormales fundamentadas en la resolución y las escalas de la imagen para la obtención de un plano preciso. Esto lleva a la elección de un método de análisis de componentes principales o PCA (Sirovich y Kirby, 1987; Kirby & Sirovich, 1990) el cual está basado típicamente en dos fases: formación y clasificación. Pero aunado a esto, se puede constituir un clasificador de emociones como es SVM, que ayuda a clasificar las características específicas conforme a similitudes o diferencias y crear así un vector a comparar con las plantillas creadas de emociones.

Dichos métodos serán los constituyentes de la arquitectura propia a desarrollar para el reconocimiento de emociones, como base de múltiples investigaciones. Sin embargo, no se pretende encasillar su uso a una sola área, sino dar la flexibilidad de implementación en diferentes ramas del área de visión artificial.

2. Desarrollo

La metodología se basa en la propuesta de una arquitectura para el reconocimiento de emociones, ésta se divide en seis etapas (figura 1).

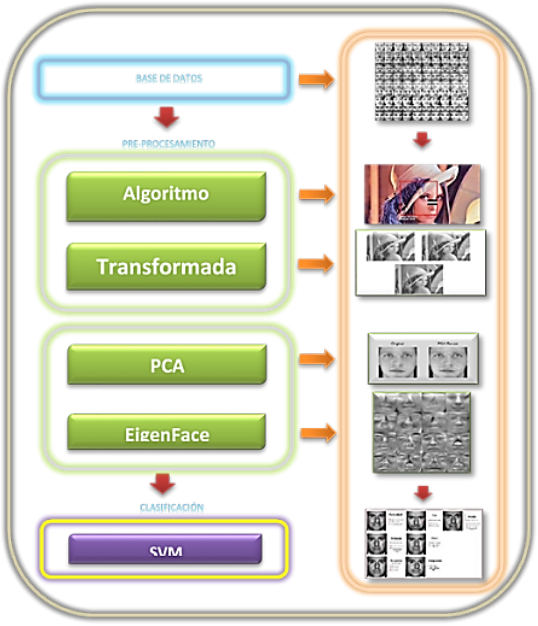

Figura 1 Diagrama de bloques.

La primera etapa se realiza la elección de la base de datos de conocimiento, en donde se tengan diversas muestras de caras o rostros a realizar la detección y clasificación de emociones, se pretende que las imágenes capturen las seis emociones de estudio de una persona, es decir tener seis imágenes por persona.

La segunda etapa es la detección de caras dentro de imágenes. A partir de la detección con el algoritmo Viola-Jones es cuando se puede crear cualquier sistema que analice la información contenida en las caras: ojos, nariz, boca, cejas, y pómulos. La detección facial se encarga de determinar si hay o no alguna cara en una imagen dada y, en caso de que exista, de extraer la localización y el contenido de dicha cara y sus elementos contenidos.

En la tercera etapa se evalúa factores como la iluminación, postura, distancia, simetría y laterabilidad. Detectando la dimensionalidad del rostro de la persona, con la finalidad de establecer a través de una transformada Wavelet un plano de la imagen, dimensionando las características del rostro cambiando texturas y resolución para la obtención completa de elementos y la determinación de una cara y una no cara.

En la cuarta etapa se definen las singularidades dentro de un espacio dimensionado, dichos elementos a identificar serían extraídos por el algoritmo PCA,

obteniendo características de acuerdo a sus similitudes o diferencias. Se establecen características y no diferencias, lo cual conlleva una reducción de la dimensionalidad de la imagen. Por último, se realiza la comparación de plantillas de las emociones con las de un vector basado en características específicas de acuerdo a la descomposición de la imagen.

En la quinta etapa se incorporó a la arquitectura de reconocimiento de emociones un sistema basado en el algoritmo básico de Eigenfaces. Dicho algoritmo será incorporado en la fase de entrenamiento dentro de la arquitectura con el fin de reducir y obtener los componentes que serán los necesarios para la reconstrucción de la imagen y la clasificación de una emoción.

En la sexta etapa se realiza la comparación de plantillas de emociones de acuerdo a un vector creado con características específicas conforme a la descomposición de la imagen. Por último, se crean clases de comparación con características específicas.

Selección de Variables

La base de datos de estudio pertenece a la Universidad Técnica de Munich (Alemania), y está conformada por imágenes de rostros de 18 personas que muestran las 6 emociones básicas definidas por Eckman y Friesen (alegría, sorpresa, enojo, tristeza, desagrado y miedo). Esta base de datos fue generada como parte del proyecto FG-NET (Face and Gesture Recognition Research Network), cada imagen es de 320x240 píxeles a 8 bits y está en formato JPEG.

Esta base de datos sobrelleva el paradigma que presentan algunas bases de datos en donde muestran emociones distintas a las naturales, debido a que en esta base de datos se pide a las personas reaccionar lo más natural o comportarse lo más espontáneo posible mientras se les estimula con videos o imágenes.

Extracción de características faciales con Viola-Jones

La segunda fase del sistema de detección de emociones consiste en aplicar el algoritmo de Viola-Jones para desarrollar un detector facial/rasgos en tiempo real, el cual proporciona un alto porcentaje de acierto la imagen y posición de uno de los

ojos, boca, nariz y pómulo, de los cuales se extraerán las posiciones de los rasgos en ellos.

Primero se establece el obtener la cara de la persona, y para ello se utiliza el algoritmo de Viola-Jones (figura 2), una explicación básica de lo que hace este algoritmo es ésta:

- Se transforma la imagen a escala de grises.
- Recorre la imagen a procesar mediante ventanas de 24×24 pixeles a diferentes escalas.
- Para cada una de estas imágenes obtiene una serie de características, que son los resultados de la diferencia de los valores de sus pixeles entre áreas.\

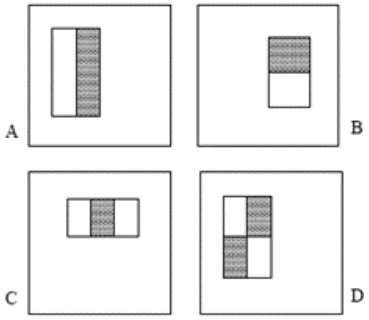

Figura 2 Características utilizadas en el algoritmo Viola-Jones para la detección de caras.

Una vez obtenida la cara del sujeto, se procesa la imagen de la cara para obtener la posición de los ojos, nariz, boca y pómulos. Para ello, se utiliza nuevamente el algoritmo de Viola-Jones anteriormente mencionado pero entrenado en este caso para detectar los rasgos humanos (figura 3).

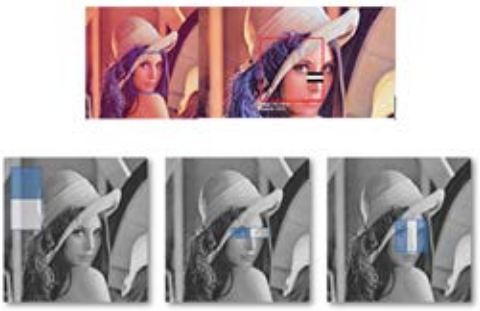

Figura 3 Cara detectada mediante algoritmo Viola&Jones.

Reducción dimensionalidad imagen con Transformada de Wavelet

Una vez localizados los elementos característicos de la imagen, la tercera fase comprende una técnica usando Transformada Wavelet, la cual ayuda a localizar pequeñas variaciones en imágenes que el ojo humano no detecta a simple vista. Por ello se segmenta la imagen en detalles a diferentes niveles de resolución y orientaciones, sin repeticiones, comenzando a realizar la similitud cuadro por cuadro, localizando las que se asemejen.

Pero esto se obtiene factores determinados por la misma transformada de Wavelet de paquetes, dichos paquetes se basan en una resolución múltiple ortogonormal compuesta por Wavelet lineales, que proporcionan la información espacial, orientación y posición de la imagen contemplando la textura de la misma a diferentes escalas y resolución. Dichos factores son resultados del contexto que se le dará a la imagen de manera física para la experimentación, dicho contexto será determinado por el diseñador, implementando variables como iluminación, postura, distancia, simetría y laterabilidad.

Con lo anterior, el concepto de ruido como ese efecto “indeseable” es el elemento que se debe eliminar de nuestras imágenes, éste consiste en la aparición aleatoria de señales ajenas a la imagen original, especialmente apreciable en las zonas de sombra de la imagen, donde se puede hacer visible la separación del contexto que rodea, el elemento caro (figura 4).

Figura 4 Discriminación de datos, como entorno del elemento cara, por medio del cambio de texturas de la imagen.

Tomando lo anterior, se puede dar respuesta al por qué se utilizan los Wavelets de Morlet para la detección de rostros en este trabajo, y es que está en el hecho de

que son buenos extractores de características de las imágenes, forman una especie de firma del rostro, ya que se basan en la respuesta de cómo las células de la corteza visual tienen campos receptivos, las cuales son pequeñas regiones altamente estructuradas; investigaciones por parte de los científicos Hubel y Wiesel describieron a esas células como detectores de bordes, algunos investigaciones más recientes como la hecha por Jhones y Palmer mostraron que el comportamiento de la respuestas de estas células en los gatos, correspondían a medidas locales en la frecuencia. También se notó que la respuesta dependía de la frecuencia y orientación del campo visual. En los experimentos hechos por Jhones y Palmers, la respuesta de estas células primarias fue medida con electrodos, mientras que los campos receptivos de estas células primarias fueron medidos por proyectar estímulos parecidos a un punto en una ventana homogénea. En otro trabajo, dos científicos Pollen y Ronner (1974) examinan la relación de fase entre células adyacentes de la corteza visual de los gatos, ellos concluyen que las células de un par de células adyacentes tienen ciertas simetrías definidas por frecuencias similares, magnitudes similares, y especificaciones similares de las direcciones: Una de ambas tiene simetría par y otra simetría impar. Esto permite modelar ambos campos receptivos de los pares de células por funciones de valor complejo.

Sin embargo, el algoritmo es muy limitado, pues sólo trabaja con imágenes de tamaño pequeñas, y busca rostros con una medida estándar, por lo tanto, tendrá mal desempeño en imágenes grandes, también tendrá mal desempeño si se buscan localizar imágenes con rostros muy grandes, o imágenes con rostros muy pequeños. Se puede seguir investigando para proponer un algoritmo más robusto que busque las áreas candidatas a ser rostro, pero en imágenes a diferentes niveles de resolución para poder reconocer rostros pequeños y rostros muy grandes (figura 5).

Obtención de Componentes Principales con PCA

El éxito de esta arquitectura para el reconocimiento de emociones depende fuertemente de los elementos utilizados para representar las imágenes para su

posterior clasificación. El patrón que represente a una imagen debe estar compuesto por los elementos más sobresalientes de ella, lo cual es obtenido en esta cuarta etapa, permitiendo reducir la cantidad de datos usados en el proceso de clasificación, y aumentar la diferencia entre ellas para que actúe como un poderoso discriminante o clasificador. Una de las técnicas más utilizadas para seleccionar un subconjunto de elementos que cumpla con esas condiciones es el Análisis de Componente Principal (PCA), la cual genera un conjunto de vectores ortonormales que maximizan la dispersión entre todas las muestras proyectadas, reduciendo al mismo tiempo su dimensión (figura 6).

Figura 5 Compresión de imágenes con Transformada Wavelet.

Figura 6 Obtención de componentes principales.

2.5. Reducción de componentes principales con EigenFace

Tanto la etapa de entrenamiento como la de reconocimiento de la emoción utilizan una base de rostros compuesta por tres conjuntos de imágenes (20 y 40 personas por grupos de prueba) de dimensión NxN.

Esta quinta fase consiste del proceso de entrenamiento, la cual conlleva los siguientes pasos:

- Cada imagen es reorganizada como un vector de dimensión N^2 cuyo valor es construido como la concatenación de cada una de las filas de la imagen, formando así una matriz de N^2xM.
- Se obtiene el rostro promedio.
- El rostro promedio obtenido es restado a cada una de las imágenes M obteniendo un nuevo conjunto de vectores que conforman la matriz de N^2 xM.

En este punto se buscan los autovectores de la matriz de covarianza de dimensión N^2xN^2. Estos vectores propios son los vectores ortonormales usados para construir la representación de las imágenes. El tamaño de la matriz hace intratable este paso por el espacio y el tiempo requerido). Para resolver este problema se buscan los autovectores de la matriz de covarianza. Debido a su gran dimensión, éstos deben ser aproximados a través de los vectores propios de la matriz de covarianza reducida.

En este trabajo se propone un nuevo método que consiste en formar una imagen de menor tamaño que permita obtenerlos directamente (figura 7).

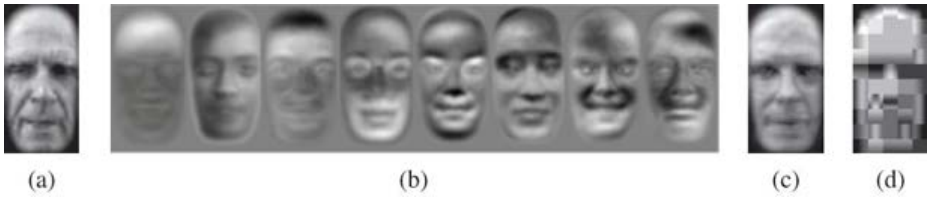

Figura 7 Ejemplo de un eigenface donde se muestran las diferentes clases.

Cada imagen es dividida en bloques de PxP. Cada uno de estos bloques se promedia y se obtiene una nueva imagen de DxD, que se obtiene de reemplazar cada bloque por su promedio.

Cada imagen es reorganizada como un vector de dimensión 2D cuyo valor es construido como la concatenación de cada una de las filas de la imagen, formando así una matriz de D^2xM:

- Se obtiene el rostro promedio.
- El rostro promedio obtenido es restado a cada una de las imágenes obteniendo un nuevo conjunto de vectores que conforman la matriz.

- Se obtiene la matriz de covarianza de dimensiones D2 x D2.
- Se obtienen los autovectores, los que ordenados de mayor a menor según sus correspondientes autovalores, conforman la matriz.
- Se obtiene un patrón.

Clasificación de emociones con SVM

La última fase se basa en las dos ideas fundamentales para la construcción de un clasificador SVM, las cuales son la transformación del espacio de entrada en un espacio de alta dimensión y la localización en dicho espacio de un hiperplano separador óptimo. La transformación inicial se realiza mediante la elección de una función kernel adecuada. La ventaja de trabajar en un espacio de alta dimensión radica en que las clases consideradas serán linealmente separables con alta probabilidad, por tanto, encontrar un hiperplano separador óptimo será poco costoso desde el punto de vista computacional. Además, dicho hiperplano vendrá determinado por unas pocas observaciones, denominadas, vectores soporte por ser las únicas de las que depende la forma del hiperplano.

Una de las principales dificultades en la aplicación de este método radica en la elección adecuada de la función kernel. Es decir, construir la función de transformación del espacio original a un espacio de alta dimensión es un punto crucial para el buen funcionamiento del clasificador.

La forma final de la regla de clasificación para un clasificador binario (dos clases, +1 y -1) son parámetros aprendidos por el clasificador durante el proceso de entrenamiento, por lo tanto, nuestro patrón obtenido con el algoritmo eigenface, será el valor de la función kernel. Si la función es mayor que un umbral entonces la emoción estimada para un punto x será una +1 y será - 1 en caso contrario.

En el problema del reconocimiento de emociones es típico trabajar con más de dos emociones. Suponiendo que el número de emociones consideradas es n. Es necesario llevar a cabo una generalización del clasificador binario al caso multiclase. En este caso, se dispone de n clasificadores, es decir, n valores de la regla de clasificación para cada objeto. En ambos casos para determinar la emoción que corresponde a cada objeto se realiza una ponderación sobre todas las reglas

de clasificación disponibles. Las SVMs han demostrado ser un método muy efectivo en la clasificación de expresiones faciales espontáneas como describe Bartlett, et al. (2001).

3. Resultados

La tabla 1, muestra los resultados del reconocimiento de 6 emociones prototipo.

Tabla 1 Resultados finales.

EMOCIÓN	*DATOS DE PRUEBA*	*IMAGEN*	*POSITIVOS*	*FALSOS POSITIVOS*	*EFICIENCIA*
Alegría	24		19	2	95%
	48		39	5	97.5%
Sorpresa	24		19	3	95%
	48		39	8	97.5%
Miedo	24		18	3	90%
	48		37	7	92.5%
Enojo	24		19	1	95%
	48		39	3	97.5%
Disgusto	24		19	2	95%
	48		39	5	97.5%
Tristeza	24		19	1	95%
	48		39	3	97.5%
EFICIENCIA EN MUESTRA DE 20					**94.16%**
FICIENCIA EN MUESTRA DE 40					**96.66%**

En la tabla 1 se consignan los resultados de la arquitectura propuesta. Es de observarse que la expresión de miedo es la más difícil de reconocer con un 92.5%

debido a que en algunas ocasiones ésta se manifiesta como una expresión ambigua. Esta observación concuerda con lo encontrado en el estudio realizado por Ehrlich en donde se expone que la expresión de miedo puede ser una excepción de las clases básicas de emociones derivadas, ya que esta emoción está restringida al proceso de selección forzada mediante el cual se seleccionan las categorías.

4. Discusión

Como trabajo a futuro se pretende enfocarse en la etapa de reducción de dimencionalidad, puesto que en el tema de reconocimiento de emociones este es el punto medular que afecta en general las arquitecturas.

Dentro de la arquitectura expuesta, se puede mejorar la dimencionalidad de la matriz de confusión generada a partir de los componentes principales obtenidos, de dichos componentes actualmente se realiza una reducción directa de dicha matriz ocasionando la perdida de información ocasional, sin embargo, si la reducción de la matriz de realiza normalizada, esa pérdida no existiría.

Por lo anterior se pretende obtener un algoritmo EigenFace normalizado y no directo, puesto que optimizaríamos nuestros porcentajes de resultados de acuerdo a cada emoción.

5. Conclusiones

La creación de una arquitectura segmentada por diferentes algoritmos como son Viola-Jones, Transformada de Wavelet, Algoritmo de Componentes Principales, EigenFace y Máquina de Vector Soporte para el manejo de una imagen, ayuda a la obtención de componentes específicos, descriptivos y característicos que permiten la construcción de rasgos faciales, que alimentaran al clasificador obteniendo con ello el reconocimiento de emociones.

En particular, el estudio de patrones de pixeles es aplicado a las regiones detectadas como ojos por el algoritmo de Viola y Jones. Posteriormente se aplica el análisis de componentes principales o PCA (del inglés Principal Component Analysis) y se seleccionan las características críticas, las cuales son utilizadas

como parámetros de entrada para un clasificador SVM. De esta manera, se logra reducir el número de falsos aciertos/positivos mejorando, por tanto, la tasa de acierto global de sistema.

Gracias a dicha arquitectura, el algoritmo de Viola-Jones, además de ofrecer buenos resultados, ofrece también una tasa de análisis muy elevada, de más de 90% de eficiencia, haciendo del mismo una herramienta útil para sistemas de obtención de características faciales en tiempo real.

En resumen, queda claro que se ha obtenido una importante ventaja en la codificación de la imagen mediante la representación que proporciona la transformada wavelet. No obstante, se ha comprobado que las componentes de alta frecuencia de la transformada wavelet pueden codificarse de forma mucho más burda que los componentes de baja frecuencia sin que aparezcan pérdidas aparentes en la calidad de la imagen. Por ello a diferencia de otros autores, la aplicación de una Transformada Wavelet como parte del pre-procesamiento de la imagen es de ayuda para la minimización de la imagen. Quizás, uno de los grandes problemas de los sistemas de detección de emociones reside precisamente en la capacidad para caracterizar aquello que no corresponde a una emoción. Por ello, se propone usar una máquina de vectores soporte o SVM (del inglés Support Vector Machine) de ochos clases para así evitar dicho problema. Adicionalmente, utilizar las auto-caras para caracterizar las imágenes, este modelo obtiene una tasa de acierto superior al 90%.

6. Bibliografía y Referencias

[1] Zhao, W.; Chellastra, R.; Rosenfeld, A. y Phillips, PJ (2003). Reconocimiento facial: A Literatura Encuesta, ACM Computing Surveys, Vol. 35, No. 4, diciembre 2003, pp 399-458.

[2] Zhou, S.; Krueger, V. & Chellapa, R. (2003). Reconocimiento probabilístico de rostros humanos de Video, Visión por Computador y la Interpretación de Imágenes, Vol. 91, 2003, pp 214-245.

[3] P. Viola and M. Jones, “Robust real-time face detection”. In Proc. Of IEEE Workshop on Statistical and Computational Thories of Vision, 2001.

[4] Coifman, R. R. & Meyer, Y. (1990). Bases ortonormales paquetes wavelet, preparativos para la imprenta.

[5] Liu C. C., P. y D. Dai Yan H. (2007). Paquetes wavelet discriminante Local coordina para la cara reconocimiento, Journal of Machine Learning Investigación, Vol. 8 (mayo de 2007) 1165-1195.

[6] Sirovich, L. & Kirby, M. (1987). Procedimiento de pocas dimensiones para la caracterización de Rostros Humanos, Revista de la Sociedad Americana de Óptica A, Vol. 4 (3), (marzo de 1987), 519 - 524, 1.084-7.529.

[7] Yang, Y., Lu, B. L. (2006). Predicción de proteínas subcelulares Multi-Lugares con Min-Max Modular Vector Apoyo a máquina, en Memorias del Tercer Simposio Internacional en redes neuronales (ISNN 2006).

[8] Guevara Díaz J. (2013). Detección de rostros por medio de las wavelets de morlet.

[9] Guevara M. L., Echeverry J. D. & Ardila Urueña W. (2008). Detección de rostros en imágenes digitales usando clasificadores en cascada.

[10] Serrano A, Conde C., De Diego I. M., Cabello E., Bai L. & Shen L. (2007). Parallel gabor pca with fusion of svm scores for face verification.