Revista Cubana de Ciencias Informáticas
Vol. 9, No. 1, Enero-Marzo, 2015
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 45-62

Tipo de artículo: Artículo original
Temática: Inteligencia Artificial
Recibido: 14/11/2014 | Aceptado: 19/01/2015

Análisis y propuesta de selección de rasgos para el Reconocimiento de Expresiones Faciales

Research and proposal of features selection for Facial Expression Recognition

Lic. Odín Castañeda Cao [1*], Dra. María Matilde García Lorenzo [1]

[1] Grupo Inteligencia Artificial, Centro de Estudios de Informática, Universidad Central Marta Abreu de Las Villas, Carretera a Camajuaní Km 5 ½ Santa Clara Villa Clara Cuba C.P: 54830. Correo-e: mmgarcia@uclv.edu.cu

* Autor para la correspondencia: odincc@gmail.com

Resumen

En este artículo se presentó un estudio de trabajos existentes en la temática de Reconocimiento de Expresiones Faciales, específicamente en la etapa de la selección de rasgos para la clasificación. Se abordó la línea de estudios de P. Ekman como uno de las más aceptadas y continuadas, así como algunas propuestas alternativas. Se tuvieron en cuenta los aportes positivos y los principales inconvenientes de las diferentes soluciones revisadas y finalmente se presentó una propuesta de sistema para el reconocimiento de emociones. Se diseñó un experimento para probar la propuesta de rasgos realizada en la que se obtuvieron resultados positivos.

Palabras clave: clasificación de emociones, reconocimiento de expresiones faciales, selección de rasgos.

Abstract

In this paper was presented a study of existing works on the subject of Face Expression Recognition, specifically in the sub problem of features selection for the classification. The line of studies of P. Ekman was approached as one of the most accepted and developed, as well as some alternative proposals. Both positive contributions and main drawbacks were taken into account from the different researched solutions, and finally a proposal of a system for

emotions recognition was presented. An experiment was designed to test the efficiency of the proposed set of patterns with positive results obtained.
Keywords: *emotion classification, facial expression recognition, features selection.*

Introducción

En la computación afectiva, campo que viene desarrollándose cada vez con más interés en los años recientes, un elemento crítico es el de definir que parámetros van a caracterizar una emoción. El proceso de involucrar las emociones con la computación demanda un estudio de estas, y de establecer categorías para identificarlas.

Como plantea Rosalind W. Picard del Media Lab del Instituto de Tecnología de Massachusetts (Picard, 2003), las emociones dependen de muchos factores físicos al igual que las condiciones meteorológicas, en estas últimas se miden valores de presión atmosférica, humedad del aire, temperaturas, y luego se utilizan algoritmos para predecir cuerpos atmosféricos como tornados o ventiscas, que no son más que condiciones extremas de estos valores, pero para todos los valores intermedios solo existen vagas clasificaciones del tipo "tiempo aceptable" o "parcialmente nublado". Con las emociones la situación es similar, solo para valores de un conjunto de contracciones musculares que sobrepasen ciertos umbrales decimos que se ha manifestado una expresión que denota la presencia de una emoción.

Para que las emociones lleguen a ser datos que un sistema pueda utilizar, es necesario un primer paso de recopilación de información, y un segundo paso de procesamiento de esta información y su clasificación, para más tarde aplicar este resultado según la tarea del sistema en cuestión. El primer paso de recopilación de información es un problema en sí mismo, pues para esto habría que determinar qué tipo de información define una emoción.

Aunque existen trabajos que incluyen medidas de parámetros como actividad eléctrica muscular, conductividad en la piel, pulsaciones del volumen de sangre y respiración como parámetros de actividad nerviosa (Picard, 2003), la mayor parte de los trabajos de clasificación de emociones se han enfocado en información recopilada del procesamiento de imágenes de rostros.

La expresión del rostro es un reflejo bastante certero del estado emocional del individuo según N. Dailey (Dailey *et al.,* 2002), y el acceso a estas imágenes se ajusta a la mayoría de las aplicaciones para las que se ha concebido el problema de clasificar emociones. Poner sensores en diversos puntos del organismo provee mayor información, pero esto es solo practicable en un ambiente de investigación de laboratorio y no en la mayoría de escenarios donde se desearía aplicar esta técnica.

El problema a resolver está en determinar cuáles son los atributos o características más apropiadas a considerar para la clasificación. A continuación se presenta una reseña de los principales trabajos que abordan esta temática y finalmente se hace una propuesta de rasgos para la clasificación.

Materiales y métodos

La mayor parte de los trabajos en esta área se han enfocado en el análisis de rostros para clasificar emociones. Este problema trae implícito el problema de seleccionar los mejores patrones para la segunda etapa que es la clasificación en sí, o lo que es lo mismo, identificar cuáles son las características que mejor identifican una emoción reflejada en un rostro.

Las tres etapas principales son: detectar el rostro, extraer los atributos y clasificarlos. Esta fase del estudio parte del supuesto de contar con una entrada de datos donde se enfoca el rostro, ya sea para una imagen fija o en movimiento. El problema de localizar el rostro en una imagen con más elementos o el prerrequisito de trabajar con imágenes de rostros centralizados pertenecen a un paso anterior y se considera un pre procesamiento ya alcanzado en este punto.

En el 1978, Paul Ekman y Wallace V. Friesen desarrollaron el Sistema de Codificación de Acciones Faciales (FACS) (Pantic 2006). Este consiste en un profundo estudio de los músculos faciales y una representación de las contracciones de estos en un conjunto de valores llamado Unidades de Acción (AU). Una unidad de acción es, por ejemplo, el levantamiento de la parte interior de la ceja (AU1), y otra el levantamiento de la mejilla (AU6). Las unidades no necesariamente tienen una equivalencia con un músculo, para una unidad de acción pueden intervenir varios músculos faciales, o un mismo músculo determinar por sí solo más de una unidad. El sistema incluye valores como intensidad, duración y asimetría para cada movimiento. Este estudio se continuó y en el 2002 los propios autores hicieron el lanzamiento de la última revisión.

El sistema original consiste de 44 unidades, y requiere mediciones en 3D, lo cual complejiza un procesamiento en tiempo real. Otros autores (Kawakami, Yamada, 1995) presentaron una simplificación del uso de solo 17 unidades, pero continuaba el requerimiento de análisis 3D.

K. Khorasani usó la imagen del rostro con la llamada clasificación "neutral" para comparar los rasgos extraídos de esta contra los de la que se quiere clasificar (Khorasani, 2004), esta variante por supuesto implica el obvio inconveniente de la necesidad de contar de antemano con la imagen del rostro "neutral" y su procesamiento.

Otros trabajos (Azcarate, 2005) cuentan con el inconveniente de tener que fijar a mano marcadores en la imagen para que el sistema le pueda seguir el rastro a los movimientos, los cuales son además analizados en una secuencia de imágenes. Con limitantes parecidas, Cohen (*et. al*, 2003) trabajan con una entrada continua de video la cual es preparada en la imagen inicial, marcando puntos como los bordes de los ojos y la boca, y luego hacen coincidir la imagen del rostro neutral con estas marcas para construir una malla 3D. Con la malla extraen los movimientos de estructuras e identifican los rasgos y las dependencias entre estos. Utilizan marcas temporales para ir procesando el desarrollo de una expresión hasta clasificarla en el clímax de su proyección.

Por otra parte, se tienen propuestas de dividir la imagen en las llamadas Regiones de Interés (De Silva, 2010). En este caso, las estructuras que componen el rostro son figuras geométricamente complejas, y el enfoque y tratamiento particularizado de cada una de ellas ha sido una técnica con resultados positivos. Áreas comunes en la selección de estas regiones han sido los ojos y la boca, intuitivamente se entiende que estas son estructuras muy expresivas dentro del rostro, además del hecho de que en torno a ellas giran la mayoría de las unidades de acción. Estudios han arrojado como dato que las estructuras faciales superiores (ojos, cejas) parecen tener más peso en la clasificación que las estructuras inferiores (boca, mejillas). (Pantic, 2000)

M. Karthigayan y M. Rizon (2008) trabajaron los ojos y la boca aproximados con construcciones de elipses. Para esto, utilizaron un algoritmo genético con el objetivo de encontrar las aristas de las elipses que mejor aproximaban las imágenes, usando una elipse regular para los ojos y una elipse irregular o compuesta (de 2 aristas menores) para la boca, por la diferencia entre el labio superior y el inferior. Además aplicaron una técnica de contorno sobre una escala negro y blanco para la definición de las estructuras y encontraron con pruebas calculadas a mano de que las clases que

querían obtener presentaban diferentes valores de las aristas de las elipses, por lo que este conjunto serviría de entrada al problema de clasificación.

El grupo de investigación de emociones de la universidad de California en el 2004 (Khorasani, 2004) utilizó 102 marcadores en el rostro del sujeto con los cuales midieron los parámetros de movimientos faciales. Se utilizaron estrictas características: la grabación con tres cámaras siguiendo los marcadores, condiciones de iluminación y sonido idóneas (utilizó también entrada de datos auditiva), y el sujeto era una actriz profesional la cual representó las distintas emociones a clasificar. Para el procesamiento de rasgos utilizaron un algoritmo específico a su escenario: normalizaron cada marcador respecto a la nariz y utilizaron una imagen neutral (otra dependencia más) contra la cual comparaban rotaciones de los marcadores divididos en áreas para obtener los valores que definirían los rasgos.

En la propuesta de A. Azcarate y F. Hageloh (2005), también partieron de un trabajo que requería una ubicación precisa a mano de una malla de alambre (wireframe) en 3D sobre la imagen 2D, a la cual había que inicializarla especificándole puntos como las esquinas de los ojos y de la boca. De esta forma, trabajaron una idea para reducir este inconveniente y se basan en el trabajo de R. Lienhart y J. Maydt (2002) que propone la utilización de las llamadas características tipo Haar para la detección de objetos en una imagen digital, inspiradas a su vez en las funciones ortonormales de la secuencia Haar propuesta por el matemático Alfred Haar en el 1909.

La idea consistía en introducir un algoritmo que usara las características tipo Haar para identificar las estructuras del rostro permitiendo así la ubicación aproximada de la malla de alambre para el inicio del procesamiento, aunque todavía requería que un usuario pusiera ciertos marcadores. Utilizan la llamada imagen integral, que es una información adicional a la imagen original, en la cual se almacena en cada pixel ii(x,y) la suma de las intensidades de los pixeles del rectángulo desde la posición (0,0) hasta el pixel en cuestión en la escala a gris de la imagen original i(x,y).

Luego con cálculos de regiones rectangulares que facilita la imagen integral, se detectan las características Haar existentes, optimizado además, por el algoritmo de Adaboost, el cual reduce considerablemente las estructuras notables a analizar.

Esta utilización de las características Haar con el cálculo de la imagen integral y Adaboost ha sido explotada en varios trabajos de detección de objetos en imágenes digitales (Jorgensen, 2006) (Hoiem, 2009).

Otro algoritmo muy utilizado en esta rama es la transformada de coseno discreta (DCT por sus siglas en inglés). DCT es un algoritmo altamente explotado en el área de compresión de imágenes: es, de hecho, el algoritmo que da lugar a la compresión dentro de los formatos de fichero estándares jpeg, mpeg y sus derivados.

Existen muchas variantes de la DCT (Strang, 1999), pero la más usada es la DCT-II bidimensional por sus propiedades de suavidad en los extremos de los intervalos.

El principio de compresión detrás de las DCT está basado en las series de Fourier. Calculando los coeficientes para los distintos vectores de la base, se logra la aproximación.

La base es en teoría infinita, y depende del problema en cuestión el nivel de error suficiente para detener la aproximación, o lo que es lo mismo, la cantidad de coeficientes a utilizar (y por tanto calcular).

El funcionamiento de la DCT se basa en este mismo principio. La imagen digital, típicamente representada con tres componentes (YCbCr o en caso del modelo anterior RGB), se trabaja cada componente por separado y en bloques, se normaliza y se calculan los coeficientes de la DCT, y por último se discriminan los coeficientes de menor impacto (se sustituyen por 0). Este tipo de compresión se denomina compresión con pérdida (lossy compression) pues hay pérdida de información al discriminar vectores de funciones que aportaban una mejor aproximación, pero que fue considerada no necesaria para las condiciones del problema. El ejemplo de esto se ve cuando nos encontramos una imagen de baja calidad en la que aparecen los llamados macro bloques o pixeles gigantes, esto es producto de la poca cantidad de coeficientes utilizados. La parte positiva es que se logra representar una imagen, con un error determinado, usando una cantidad de información excesivamente inferior a la contenida en la imagen original; en términos computacionales: se utiliza mucha menos memoria de almacenamiento, y en términos de extracción de patrones de una imagen: se trabaja con un conjunto mucho menor de datos para representar la imagen.

Estas características favorables fueron explotadas por L. Ma y K. Khorasani (2004), quienes usaron los coeficientes más significativos de la DCT como la entrada de su clasificador, que era una red neuronal. Ellos además hicieron

cierto pre procesamiento, restándole una imagen neutral a la que se iba a clasificar, además de tener todas estas ya normalizadas con dimensiones y características específicas. Después de aplicar la DCT, hicieron varias pruebas con distintas dimensiones de la matriz de coeficientes significativos a seleccionar, de los obtenidos con la DCT, para quedarse con la óptima y utilizar esta matriz como su conjunto de patrones que determinarían la clasificación.

Esta variante se ha usado frecuentemente en trabajos de reconocimiento de expresiones en imágenes rostros (Pan, 2000), pues constituye un procedimiento computacionalmente eficiente de trabajar con una representación de la imagen. Sin embargo, DCT es una variante algorítmica de reducir cualquier imagen, bien se podría usar en problemas generales de identificación en imágenes, este acercamiento con DCT no explota la idea de que lo que se trabajan son emociones y la búsqueda, por tanto, de elementos concretos que caractericen estas.

Una propuesta similar fue hecha por M. N. Dailey y G. W. Cottrell (2002), donde utilizaron los llamados filtros Gabor como patrones de la imagen para la entrada del clasificador. El principio de los filtros Gabor es compartido con la DCT en los trabajos de Fourier. Estos son también funciones con componentes trigonométricos para los que se definen diferentes orientaciones angulares y de conjunto constituyen una representación optimizada de la imagen. J. G. Daugman, físico y profesor de reconocimiento de patrones en la universidad de Cambridge, descubrió que los filtros Gabor podían modelar células de la corteza visual de cerebros de mamíferos.

Una de las técnicas de extracción de rasgos más comunes en reconocimiento de emociones faciales son la extracción basada en píxeles (Subramanian *et al.*, 2012), los filtros Gabor, las transformadas de Curvelet y los patrones binarios locales que fue la técnica que utilizaron. Todos estos algoritmos son muy efectivos en procesamientos de imágenes en general, pero no aprovechan la información de que están trabajando sobre rostros o tratando de clasificar emociones.

Z. Zeng y M. Pantic (2009) presentan un trabajo donde tratan de ir más lejos y critican los enfoques simples de reconocimientos de expresiones en el rostro y los trabajos orientados a las clases de Ekman. Proponen usar más elementos e información del contexto, e involucran otras ramas como la psicología y la lingüística. Modelan la emoción con un esquema multidimensional con aristas como la evaluación que se mueve de negativo a positivo, la activación que refleja la voluntad o ánimo de manifestación, el control, etc. Defienden la inclusión de información auditiva, y en la visual la importancia de movimientos temporales que llaman claves en la expresión de una emoción.

En el trabajo de B. Fasel (*et al.*, 2002) también proponen la inclusión de información auditiva, y de resultados de estudios psicológicos. Habla de la necesidad de analizar un rostro en diferentes ángulos y con variables niveles de iluminación. En la revisión de estudios reconocen el valor del estándar de Ekman y mencionan los desventajosos requisitos del uso de técnicas de pre procesamiento como el empleo de una imagen neutral para calcular diferencia y de marcadores manuales para la extracción de rasgos.

En otro trabajo de M. Pantic con M. Valstar (*et al.*, 2004) se enfocan en el subproblema de detectar y clasificar la ocurrencia de una Unidad de Acción del estándar FACS de Ekman. Requieren como entrada para este análisis la llamada plantilla temporal, que es una imagen construida a partir de una secuencia de imágenes del rostro, la cual tiene información de donde y cuando ocurrió una alteración, en este caso, un movimiento facial. Utilizan una combinación de kNN con un sistema experto de reglas, logran clasificar quince Unidades de Acción y alcanzan un 76,2% de aciertos con la base de datos Cohn-Kanade. Como desventaja, requieren de marcas manuales de puntos en las imágenes para su normalización.

M. Pantic (2000) también presentó un resumen de los avances en el campo del reconocimiento de expresiones faciales, y lo que pretendía ser una guía para el desarrollo en esta línea. Además del análisis de rostros, habla también de información de voz, gestos de manos y expresión corporal. Se concentra en revisar los avances en los trabajos de automatizar expresiones faciales en imágenes fijas y en secuencias de imágenes.

Divide los enfoques de extraer rasgos en tres formas: tomando el rostro completo (holístico), tomando el rostro como un conjunto de componentes como ojos, boca, cejas, etc. (analítico), o una combinación de estos dos enfoques (hibrido).

En enfoques analíticos se presentan modelaciones del rostro basándose en puntos distribuidos en el rostro, con cierta similitud a los puntos de las Unidades de Acción de Ekman. En los holísticos se ponen de ejemplo las mallas 3D y modelos espacio-temporales de movimientos (para secuencias de imágenes). Para los híbridos plantean que normalmente se utilizan puntos para determinar la posición inicial de alguna plantilla.

M. Pantic identifica otros tres problemas: 1ro: el sistema debe ser capaz de analizar sujetos de cualquier género, edad o etnia, y además el hecho de que cada persona tiene su propio umbral de intensidad en sus expresiones; aquí dice que

los trabajos que usan una imagen "neutral" como referencia para comparar tienen ventaja. 2do: es importante entender que el lenguaje corporal es dependiente de la situación, aunque es difícil obtener computacionalmente el contexto en que aparece la expresión, por lo que este tópico es ignorado por los sistemas existentes. 3ro: existe en estos momentos un creciente estudio psicológico que plantea que el tiempo en las expresiones faciales es un factor crítico en la interpretación de las expresiones.

M. Valstar y M. Pantic (*et al.*, 2011) presentan un intento de universalización de los resultados en el área, identificando el problema de la diversidad de enfoques, conjuntos de rasgos y bases de datos utilizadas para los entrenamientos de los clasificadores. El trabajo propone dos retos para los algoritmos y clasificadores, de unidades de acción y de emociones en expresiones faciales. Reconocen como las bases de casos más utilizadas la Cohn-Kanade (Kanade, 2000) (Cohn, 2010), la JAFFE y la MMI (Valstar, 2010).

Resumen investigativo

El modelo FACS y las Unidades de Acción de Ekman se han establecido como un estándar reconocido y eficiente para el estudio de los movimientos y las expresiones faciales así como su propuesta de siete clases de emociones. Algunos autores (Pantic, 2000) atacan el modelo de las siete clases llamándolo ambiguo por estar expresado lingüísticamente, y que no existe una correlación determinante entre las Unidades de Acción y las emociones, por ejemplo alzar las cejas y sonreír puede presentarse tanto en una expresión de sorpresa como en una de alegría, pero es sin dudas el esquema más descriptivo de los movimientos faciales y ha sido punto de partida para muchos trabajos en el área.

A pesar de haberse alcanzado altos porcentajes de aciertos en la clasificación de emociones, los sistemas computacionales siempre estarán acotados por la interpretación e identificación que se les den a estas. Aunque los mecanismos humanos para detectar rostros son muy robustos, no se puede decir lo mismo de la interpretación de rostros. De acuerdo con un estudio (Pantic, 2000), un observador puede clasificar las siete emociones básicas con un acierto de 87%, con factores mediáticos como la familiaridad de los rostros, la familiaridad de la personalidad, la experiencia general con distintas emociones, la atención dada al rostro, y las pistas no-visuales como el contexto en que se muestra la expresión. De aquí los intentos de muchos autores de incorporar más elementos de entrada de datos para la clasificación, pero estos hacen dependiente el sistema de pre requisitos que no sean aplicables en un entorno real.

Un sistema ideal debe ser capaz de hacer todas las etapas automáticamente (sin que medie ningún humano), saber manejar condiciones de la foto como iluminación, orientación e inclinación del rostro, variación de tamaño, ruido y desenfoque en la calidad de la imagen, obstáculos en el rostro como el pelo y espejuelos, distinguir todas las expresiones y detectar las 44 AU de FACS, tener aprendizaje adaptativo, operar en tiempo real y asignar etiquetas de interpretación cuantificadas y múltiples. (Pantic, 2000)

De los trabajos analizados, ninguno cumple con todos los requisitos. Los requisitos más incumplidos son manejar distintos niveles de iluminación y orientación, identificar todas las posibles expresiones, identificar las 44 acciones faciales, y contar con un proceso de aprendizaje adaptativo.

En un entorno real de aplicación de la computación afectiva como por ejemplo computación educacional, se puede contar con una cámara dirigida directamente al rostro, por lo que factores como orientación e iluminación no serían de prioridad a cubrir pero sí es necesario que el sistema sea completamente automático y prescinda de marcas y puntos situados a mano previo el análisis.

En dependencia de la aplicación en cuestión se puede valorar si es factible obtener la imagen del rostro neutral, la cual ha demostrado elevar considerablemente los resultados de acierto en la clasificación.

Aunque muchos trabajos han alcanzado resultados utilizando técnicas de compresión de imágenes generales como Haar-Adaboost y filtros Gabor, para la selección de los rasgos tendría sentido utilizar elementos relacionados con las emociones reflejadas en el rostro, que es el campo de trabajo. El uso de un modelo derivado del sistema FACS y de las Unidades de Acción de Ekman ha sido lo más frecuente y ha mostrado altos niveles de acierto en la clasificación. Combinar esto con técnicas de regiones de interés también ha mostrado resultados elevados.

Como clasificadores se utilizaron con éxito las redes neuronales, algoritmos bayesianos, kNN y máquinas de soporte vectorial (SVM), así como algoritmos compuestos y derivados de estos.

Un último elemento es la base de casos sobre la cual entrenar y probar el sistema. La comunidad científica dedicada a este campo aún no ha llegado a un consenso para establecer una base de datos estándar con la cual se pueda comparar

la efectividad de los distintos modelos propuestos. Las necesidades para cada sistema varían, pero la base de datos de Cohn-Kanade ha sido la base más utilizada y recomendada por los trabajos en el área debido al volumen de sujetos y confiabilidad de la clasificación, que cuenta con codificación de las Unidades de Acción, cuenta con secuencias de imágenes partiendo de la neutra que permiten el trabajo tanto para sistemas de imágenes fijas como para imágenes en progresión, y tiene un balance adecuado en los sujetos que la componen en términos de sexo, etnia y edad. Aún así, en la experiencia de los autores de este trabajo, se detectaron varias inconsistencias en esta base de datos.

Diseño de un sistema de selección de rasgos para clasificar emociones

Los estudios de Ekman fueron asumidos y asimilada su propuesta de las 7 emociones universales. Su modelo de las Unidades de Acción y los trabajos derivados de estas por otros autores fueron llamativos en esta propuesta para utilizar una variante de estos como la entrada al clasificador.

Entre los objetivos principales que se desea alcanzar está el trabajo en tiempo real, lo más independiente posible de pre procesamientos y precondiciones. Se busca descartar la mayor cantidad de restricciones y condiciones en las imágenes de entrada, así como la simplificación del conjunto de entrada de rasgos.

En vista a eliminar imprecisiones como la distancia del sujeto a la cámara, zoom del lente, e inclinación del rostro, se procede a normalizar la entrada de rasgos utilizando puntos que no puedan ser alterados por expresiones. Se valoraron puntos que estuvieran vinculados a la estructura ósea y que no estuvieran sujetos a ninguno de los movimientos musculares del rostro. Después de considerar empíricamente varios, el interior de los ojos, los lagrimales, fueron seleccionados, estando fijados en el interior de la cavidad ocular. Todas las medidas que constituyen la entrada de rasgos, son normalizadas por la distancia entre estos puntos.

Con esta normalización se cuenta con un valor para cada rostro el cual no varía independientemente de la emoción que se esté reflejando. Dividiendo cada valor de las unidades de acción de un rostro por su factor de normalización, se obtiene una aproximación en los valores obtenidos de distintos rostros en intensidades de emociones similares. Esto, además, descarta cualquier error introducido por diferencias de distancias en que sean tomadas las imágenes, o lo que es lo mismo, el tamaño de la misma. El establecimiento de un sistema de coordenadas basadas en estos dos puntos normaliza también el grado de inclinación de la imagen.

Revista Cubana de Ciencias Informáticas
Vol. 9, No. 1, Enero-Marzo, 2015
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 45-62

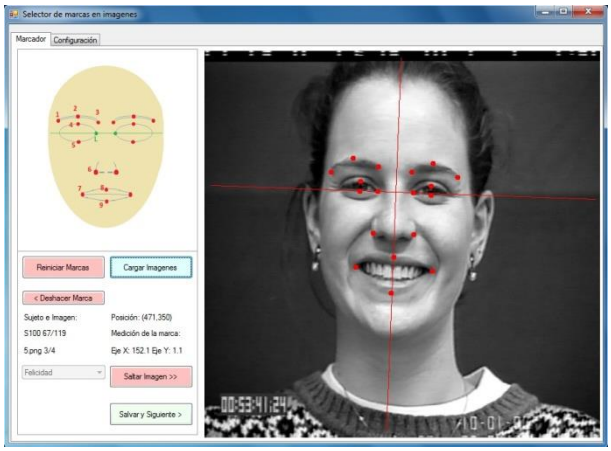

Figura 1: Propuesta de rasgos

Se analizan sólo imágenes tomadas exactamente de frente y además se separan los clasificadores según los rasgos raciales ya que las diferencias marcadas entre etnias pueden confundir al clasificador. Para resolver esto pudiera contarse con varios clasificadores, uno por cada grupo, y dar como resultado la clasificación más ocurrente, valorando el costo computacional de este procedimiento; de lo contrario se hace necesaria la preselección del clasificador especializado en la etnia en cuestión.

Las unidades de acción que se trabajan se dividen en 4 grupos: cejas, ojos, nariz y boca Fig. 1. Los puntos de la nariz (6I y 6D) contribuyen de conjunto con los puntos de extremos de la boca (7I y 7D) para recoger la información equivalente a las unidades de acción de Ekman relativas al movimiento de las mejillas.

Los movimientos de las unidades de acción en esta propuesta son sustituidos por las mediciones de ubicación de los puntos dados en el rostro. Los puntos del 1 al 7 y el L tienen sus versiones izquierdas y derechas. El 8 y 9 (labio superior y labio inferior respectivamente) son centrales. Los valores que se trabajan son las distancias de los puntos a los ejes determinados con centro en el punto medio entre los puntos L izquierdo y derecho, Fig. 1. Por último, los valores son normalizados dividiéndose por la distancia entre los puntos L, que es el factor de normalización de la imagen.

Resultados y discusión

Se implementó una aplicación para la captación manual de los puntos con los que se calculan los valores del conjunto de rasgos Fig. 1 y la generación de este conjunto según opciones.

En la etapa de clasificación, para probar el conjunto de rasgos, se utilizó la aplicación WEKA por su prestigioso arsenal de herramientas para ejecutar experimentos y análisis con diferentes clasificadores.

Se realizó una primera prueba con la base de datos JAFFE, pero al hacer una selección de las imágenes con mejor calidad y que se estimaron tenían una apropiada representación por los actores y una clasificación estimada confiable, se obtuvieron niveles de representación de cada clase muy bajos y disparejos, que afectaron el entrenamiento y la clasificación. Los mismos autores de JAFFE advierten de un índice pobre de representación y clasificación en algunos casos. Aún con estas condiciones, el nivel de acierto medio de clasificación con un Perceptrón Multicapas fue de un 78%, con la clase Sorpresa en 100% y Felicidad, Tristeza y Disgusto por encima del 80%.

Se utilizó la base de datos de Cohn-Kanade (Kanade, 2000) (Cohn, 2010), la cual es de mayor reconocimiento (Pantic, 2000), en la que también se tuvo que hacer una selección de los casos a utilizar pues habían inconsistencias en los datos, como imágenes sin etiquetar y representaciones de clases evidentemente mal registradas. Se trabajó con un conjunto de 342 instancias de expresiones de 119 sujetos, y se prepararon dos pruebas de entrenamientos: una de clasificación independiente, y otra con información de imagen neutral asociada.

En la variante de clasificación individual se trabaja directamente con la posición de los puntos normalizada relativa a los puntos de control centros del sistema, L-Izquierdo y L-Derecho. En el caso de la imagen neutral asociada, para cada elemento de la muestra se le tenían en cuenta los valores de los atributos para la expresión neutral, los cuales eran substraídos a cada uno de los conjuntos de las otras clases y esta diferencia constituían los valores de los atributos de entrada.

En las corridas con el WEKA de validaciones cruzadas a diez iteraciones arrojaron valores de aciertos para un Perceptrón Multicapas de 79.53% para la prueba de clasificación individual, y de un 88.89% para la prueba con imagen neutral asociada. El resultado superior con imagen neutral asociada es esperado pues los rasgos se conforman

con información adicional. La clase de mejor clasificación fue Felicidad con 97.5%, seguida por Sorpresa y Tristeza con 96.4% y 93.3% respectivamente.

El perceptrón utilizado, tanto en la variante con imagen neutral asociada como en la de clasificación individual, estaba formado con una capa oculta de 18 nodos, 31 atributos de entrada, y la capa de salida con un nodo para cada una de las siete clases. Los pesos de las dendritas se generaron aleatoriamente al inicio y se actualizaron con una ganancia de 0.3 y un momento de 0.2.

En la variante de imagen neutral asociada se clasificaron correctamente 304 de las 342 instancias, para un acierto de 88.89%.

Tabla 1: Matriz de confusión para la variante con imagen neutral.

a	b	c	d	e	f	g		Acierto
79	1	0	0	0	0	1	a = Felicidad	0.975
0	28	0	0	0	1	1	b = Tristeza	0.933
1	0	80	1	0	1	0	c = Sorpresa	0.964
0	1	0	37	5	0	4	d = Ira	0.787
0	0	0	7	51	0	0	e = Disgusto	0.879
1	2	3	1	0	17	1	f = Miedo	0.68
0	3	0	2	1	0	12	g = Desprecio	0.667

En la variante de clasificación individual se clasificaron correctamente 272 de las 342, para un 79.532% de acierto promedio.

Tabla 2: Matriz de confusión para la variante de clasificación individual.

a	b	c	d	e	f	g		Acierto
77	0	1	1	2	0	0	a = Felicidad	0.951
0	19	0	3	0	4	4	b = Tristeza	0.633
1	0	80	0	1	1	0	c = Sorpresa	0.964
1	5	0	31	7	1	2	d = Ira	0.66
1	1	0	7	48	1	0	e = Disgusto	0.828
4	3	3	1	0	10	4	f = Miedo	0.4
3	4	0	3	0	1	7	g = Desprecio	0.389

Varios factores hacen que sea de baja relevancia una comparación de resultados en esta área. Muchos autores eligen un subconjunto de clases de emociones para clasificar, y cada trabajo tiene sus condiciones de captación de rasgos, requiriendo pre procesamientos específicos. Lo más importante que hace poco apropiada una comparación es el conjunto de datos. Si bien existen bases de datos como la Cohn-Kanade con un alto reconocimiento para trabajos en el

área, aún no se ha logrado un conjunto de muestras que sirva de estándar lo cual es crítico para una comparación de niveles de eficiencia. La mayoría de los autores conforman su propio conjunto de muestras en lugar de utilizar uno pre-publicado.

Tabla 3: Ejemplos de resultados de trabajos de clasificación de emociones en expresiones faciales

Autores	Método	Conjunto de casos	Acierto
G. J. Edwards	Análisis de componentes principales con distancia de Mahalanobis	200 imágenes, 25 sujetos	74%
H. Hong	Galerías personalizadas y Correspondencia de grafo elástico	175 imágenes, 25 sujetos	81%
C. L. Huang	Análisis de componentes principales con Clasificador de distancia mínima	90 imágenes, 15 sujetos	84.5%
M. J. Lyons	Análisis de componentes principales y Asignación Latente de Dirichlet sobre grafos etiquetados	193 imágenes, 9 sujetos femeninos japoneses	75 – 92%
F. Hara	Redes Neuronales 234x50x6 con aprendizaje de propagación hacia atrás	90 imágenes, 15 sujetos	85%
C. Padgett	Redes Neuronales 15x10x7 con aprendizaje de propagación hacia atrás	84 fotos de Ekman	86%
Z. Zhang	Redes Neuronales 646x7x7 con aprendizaje Rprop	213 imágenes, 9 sujetos femeninos japoneses	90%
J. Zhao	Redes Neuronales 10x10x3 con aprendizaje de propagación hacia atrás	94 fotos de Ekman	100%
M. Pantic	Sistema experto de reglas	265 imágenes de vistas dobles, 8 sujetos	91%

En esta investigación llevada a cabo por M. Pantic de resultados alcanzados por trabajos con objetivos y condiciones similares en el área se observan niveles de acierto comparables a los 79.532 y 88.89 obtenidos en las dos variantes para la propuesta de rasgos planteada, usando un clasificador de redes neuronales clásico sin alteraciones y una base de casos pública.

Conclusiones

De acuerdo a los resultados alcanzados en las pruebas de clasificadores del conjunto de rasgos se determinó que la línea de trabajo es adecuada y se debe continuar desarrollando. La inmediata área de atención será la continua revisión del conjunto de rasgos, eliminando o agregando, teniendo en cuenta su peso en la clasificación.

Se proyectará también el desarrollo de la aplicación modular, con una interfaz de intercambio para la capa de captación de los datos de la imagen y la construcción de los rasgos, así como la capa de entrenamiento y clasificación, con vistas a contar con una aplicación capacitada para ser explotada en un ambiente real, como en el campo de Educación Asistida dentro de la Computación Afectiva.

Referencias

- A. JORGENSEN: AdaBoost and Histograms for Fast Face Detection. Master of Science Thesis, School of Engineering Physics Royal Institute of Technology, Stockholm, Sweden. 2006.
- AZCARATE, AITOR, FELIX HAGELOH, KOEN VAN DE SANDE, AND ROBERTO VALENTI. "Automatic facial emotion recognition." Universiteit van Amsterdam. 2005.
- B.FASEL, JUERGEN LUETTIN: Automatic facial expression analysis: a survey. Pattern recognition, 2003, 36 (1), p. 259-275.
- COHEN, N. SEBE: Facial expression recognition from video sequences: temporal and static modeling. Computer Vision and Image Understanding, 2003, 91 (1) p. 160-187.
- D. HOIEM, A. A. EFROS: Geometric context from a single image. En: Tenth IEEE International Conference. Computer Vision. ICCV 2005. IEEE p. 654-661
- DE SILVA: Evaluation of Facial Expressions of Web Users. Engineering Project Submitted as part requirement for B.Eng (Hons), School Of Engineering And Advanced Technology, Massey University. 2010.
- G. STRANG: The Discrete Cosine Transform. SIAM review, 1999, 41 (1) p. 135-147.
- K. SUBRAMANIAN, S. SURESH, R. VENKATESH BABU: Meta-Cognitive Neuro-Fuzzy Inference System for Human Emotion Recognition. En: The 2012 International Joint Conference on Neural Networks (IJCNN). Brisbane, QLD: IEEE, 2012, pp. 1-7
- KANADE, T., COHN, J. F., & TIAN, Y. Comprehensive database for facial expression analysis. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France, 2000, p. 46-53.
- L. MA, K. KHORASANI: Facial Expression Recognition using constructive feedforward neural networks. En: IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2004, 34(3), p. 1588-1595

- LUCEY, P., COHN, J. F., KANADE, T., SARAGIH, J., AMBADAR, Z., & MATTHEWS, I. (2010). The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression. Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis (CVPR4HB 2010), San Francisco, USA, p. 94-101.
- M. KARTHIGAYAN, M. RIZON: Genetic Algorithm and Neural Network for Face Emotion Recognition. Affective Computing (2008): 57-68.
- M. VALSTAR AND M. PANTIC: Induced disgust, happiness and surprise: An addition to the MMI Facial Expression Database. Proc. 3rd Intern. Workshop on EMOTION (satellite of LREC): Corpora for Research on Emotion and Affect, p. 65 (2010)
- MAJA PANTIC, I. PATRAS. Dynamics of Facial Expression: Recognition of Facial Actions and Their Temporal Segments from Face Profile Image Sequences. IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics, 36 (2), p. 433-449. (2006)
- MAJA PANTIC, Student Member, IEEE, and Leon J.M. Rothkrantz: Automatic Analysis of Facial Expressions: The State of the Art. En: IEEE Transactions on Pattern Analysis and Machine Intelligence, Diciembre 2000, 22 (12), p. 1424-1445
- MATTHEW N. DAILEY, GARRISON W. COTTRELL: EMPATH: A Neural Network that Categorizes Facial Expressions. Journal of cognitive neuroscience, 2002, 14 (8), p. 1158-1173.
- MARK HALL, EIBE FRANK, GEOFFREY HOLMES, BERNHARD PFAHRINGER, PETER REUTEMANN, IAN H. WITTEN; The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1. (2009)
- MICHAEL J. LYONS, SHIGERU AKEMASTU, MIYUKI KAMACHI, JIRO GYOBA. Coding Facial Expressions with Gabor Wavelets, 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200-205 (1998) (Japanese Female Facial Expression (JAFFE) Database, Disponible en: http://www.face-rec.org/databases/)
- MICHEL F. VALSTAR, BIHAN JIANG, MARC MEHU, MAJA PANTIC, AND KLAUS SCHERER: The First Facial Expression Recognition and Analysis Challenge. En: IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG 2011), IEEE 2011, pp. 921-926
- MICHEL VALSTAR, IOANNIS PATRAS AND MAJA PANTIC: Facial Action Unit Recognition Using Temporal Templates. En: 13th IEEE International Workshop on Robot and Human Interactive Communication. ROMAN, IEEE, 2004. pp. 253-258.
- PICARD: Affective Computing: challenges. International Journal of Human-Computer Studies, 2003, 59(1), p. 55-64.

- R. LIENHART, J. MAYDT: An extended set of Haar-like features for rapid object detection. En: 2002 International Conference on Image Processing. 2002. Proceedings. (Vol. 1, pp. I-900). IEEE.
- Z. PAN, A. G. RUST: Image redundancy reduction for neural network classification using discrete cosine transforms. En: Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks, 2000. IJCNN 2000, (Vol. 3, pp. 149-154). IEEE.
- ZHIHONG ZENG, MAJA PANTIC, GLENN I. ROISMAN, THOMAS S. HUANG: A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. En: IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(1), p. 39-58