

Identificación de emociones básicas.

JIMÉNEZ-VÁZQUEZ Mario†*, MONTERO-VALVERDE José Antonio, MARTÍNEZ-ARROYO Miriam, DE LA CRUZ-GÁMEZ, Eduardo.

Instituto Tecnológico de Acapulco.

Recibido: septiembre, 22, 2018; Aceptado Febrero 9, 2019.

Resumen

En este trabajo se muestra una metodología para reconocer las emociones humanas básicas (contento, enojo, neutro y sorpresa) a través del análisis facial, el cual está desarrollado de la siguiente manera: primera etapa; reconocimiento del rostro en una imagen, segunda etapa; extracción de características principales, tercera etapa; reconocimiento de la expresión facial. Por el momento se exponen los primeros resultados utilizando imágenes faciales, los cuales resultan satisfactorios. En la sociedad las emociones humanas desempeñan un papel muy importante en las personas, sin importar el género de éstas, se manifiestan de alguna forma mostrando diferentes estados de ánimo. Las emociones determinan la manera de pensar, de comportarnos y de comunicarnos con los demás. Por lo tanto, es comprensible pensar que las nuevas generaciones de máquinas deben tener la habilidad de entender las emociones humanas para poder interactuar de forma natural y flexible con los humanos.

Por ejemplo, resulta de gran utilidad que un vehículo interactúe con el conductor cuando reconozca a través de la expresión facial que no está concentrado en la conducción, o que una herramienta de enseñanza, determine cómo interactuar con el usuario cuando reconozca que no está atento. En estos ejemplos se muestra la importancia la expresión facial.

Palabras clave: Emociones humanas, HOG (Histogramas de Gradientes Orientados), SVM (Máquina de Vectores de Soporte (SVM), Detección facial.

Abstract

This paper shows a methodology to recognize basic human emotions (contentment, anger, neutrality and surprise) through facial analysis, which is developed as follows: first stage; face recognition in an image, second stage; extraction of main characteristics, third stage; recognition of facial expression. At the moment the first results are exposed using facial images, which are satisfactory.

In society, human emotions play a very important role in people, regardless of their gender, they manifest in some way showing different moods. Emotions determine the way we think, behave and communicate with others. Therefore, it is understandable to think that new generations of machines must have the ability to understand human emotions in order to interact naturally and flexibly with humans.

For example, it is very useful for a vehicle to interact with the driver when it recognizes through the facial expression that it is not focused on driving, or that a teaching tool determines how to interact with the user when he recognizes that he is not attentive. In these examples, the importance of facial expression is shown.

Keywords: Human Emotions, HOG (Oriented Gradient Histograms), SVM (Support Vectors Machine (SVM), Face Detection.

Citación: JIMÉNEZ-VÁZQUEZ Mario†*, MONTERO-VALVERDE José Antonio, MARTÍNEZ-ARROYO Miriam, DE LA CRUZ-GÁMEZ, Eduardo. Identificación de emociones básicas. Foro de Estudios sobre Guerrero 2018, Mayo 2019- Abril 2020 Vol.5 No.6 902-911.

*Correspondencia al Autor (mario_jv@hotmail.com)

† Investigador contribuyendo como primer autor.

I.- Introducción

Se puede dar como un hecho, que en un futuro cercano las contribuciones que hacen la inteligencia artificial y la robótica con el desarrollo de las aplicaciones, van a permitir la creación de máquinas más capacitadas para realizar cada vez un mayor número de funciones en la sociedad, su influencia va a ser notable en una variedad de dominios tales como: asistentes quirúrgicos, guiar vehículos en campos de batalla, asistentes en hogares, apoyo en oficinas, y asistentes en contextos educacionales, entre otros.

En la sociedad humana, las emociones desempeñan un rol muy importante. Las emociones son los indicadores más confiables de socializar. Ellas determinan la manera de pensar, de comportarnos y de comunicarnos con los demás. Por lo tanto, es comprensible pensar que las nuevas generaciones de máquinas deben tener la habilidad de entender las emociones humanas para poder interactuar de forma natural y flexible entre los humanos y las computadoras. 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, o una herramienta de enseñanza que identifique como interactuar con el usuario cuando detecte que éste no está poniendo atención en el tema. Las emociones faciales se manifiestan de distintas formas en las personas, siendo las expresiones faciales una de las más importantes para la comunicación entre las personas.

Por lo tanto, en los casos anteriores se considera que la interacción entre las personas y las computadoras se realiza como resultado del análisis de la expresión facial. Las expresiones faciales son relevantes en la comunicación diaria. El rostro humano describe un gran número de tareas cognitivas a través de las cejas, ojos, nariz, boca y mentón.

En este contexto, las personas generalmente levantan las cejas cuando están siendo inquisitivos o expresan sorpresa, y las bajan cuando están hablando en voz baja, o bien levantan las cejas y los labios para mostrar un comportamiento de duda, asimismo fruncen las cejas y labios cuando muestran enojo. Investigaciones realizadas en psicología social han mostrado que la transmisión de mensajes en las conversaciones puede ser dominada más por las expresiones faciales que por las palabras expresadas (Cowie, y otros, 2008).

En este sentido, una mayor comprensión del estado emocional humano por parte de las máquinas podría generar nuevas expectativas de nuestro comportamiento y actuar en consecuencia, proporcionando nuevas alternativas de solución a algunos problemas. Por lo tanto, el reconocimiento del estado emocional humano se considera un área de investigación relevante en varias comunidades científicas de inteligencia artificial por la variedad de aplicaciones que involucra.

Recientemente, la investigación relacionada con los estados afectivos ha sido ampliamente estudiada y existe una creencia creciente de que proveer a las computadoras con la capacidad de entender los estados emocionales de las personas es una tarea importante (Pantic, Pentland, Nijholt, & Huang, 2007), (Robinson & Kaliouby, 2009.). Se cree que, con el fin de conseguir progresos en el futuro en las interacciones humano-máquina es necesario que éstas puedan reconocer el estado emocional de los usuarios.

Con el desarrollo de esta investigación, se pretende demostrar que analizando algunos puntos del rostro tales como, los ojos, las cejas y la boca, es posible reconocer de manera confiable estas emociones. La salida del sistema es el reconocimiento de una emoción, misma que será manejada en una etapa futura para apoyar el proceso de toma de decisiones en una máquina.

El resto del artículo está organizado como se indica. En la Sección II se abordan los trabajos relacionados con el reconocimiento automático de emociones humanas, la metodología utilizada se muestra en la Sección III, los resultados obtenidos se indican en la Sección IV, las conclusiones se mencionan en la Sección V, la contribución de este trabajo se indica en la sección VI. Finalmente, las referencias son listadas en la Sección VII.

Objetivos

Detectar el rostro de una persona en imágenes faciales obtenidas con la cámara web en ambientes reales.

Detectar los estados emocionales en las imágenes faciales a través de las expresiones faciales.

Clasificar los estados emocionales.

II.-Trabajos relacionados

Hasta hace poco tiempo todos los trabajos que se desarrollaban para el reconocimiento de las emociones humanas se basaban en dos sistemas estándares los cuales se describirán brevemente a continuación:

El Sistema de Codificación de Acción Facial (FACS), desarrollado por Ekman y Friesen en 1977 (Ekman & Friesen, 1978) y

Los Parámetros de Animación Facial (FAPs) Estos parámetros son una parte del estándar MPEG-4, Codificación híbrido natural sintética (SNHC), desarrollado en 1998 (Video, 1998).

El Sistema de Codificación de Acción Facial (FACS)

Este sistema está basado en observaciones realizadas por observadores. Solo que estas observaciones que se realizaban no eran lo suficientemente confiables porque no se pueden considerar como una ciencia exacta para dar una conclusión de un estudio realizado (Cowie, y otros, 2008).

Las limitaciones que tienen los observadores pueden ser superadas mediante la representación de las expresiones faciales y sus comportamientos en términos de un conjunto fijo de parámetros faciales. Con un marco de este tipo en su lugar, solamente estos parámetros individuales tienen que ser observados sin tener en cuenta el comportamiento facial como un todo. Para resolver estos problemas, Ekman y Friesen (Ekman & Friesen, 1978), desarrollaron el sistema de codificación de acción facial (FACS), que desde entonces se ha convertido en el estándar de facto.

La Codificación de Acción Facial es un enfoque basado en el músculo. Se trata de identificar los diversos músculos faciales que individualmente o en grupos causan cambios en los comportamientos faciales. Estos cambios en la cara y el subyacente cambio en los músculos (uno o más) se llaman unidades de acción (UA). El Sistema de Codificación de Acción Facial (FACS) se compone de varias de tales unidades de acción.

Los Parámetros de Animación Facial (FAP)

Los investigadores de la animación por computadora se ven en la necesidad de enfrentar problemas similares de reconocimiento de expresión del rostro que enfrentaron días antes de que apareciera el FACS. No había ninguna norma unificadora y casi cada sistema de animación que fue desarrollado tenía su propio conjunto definido de parámetros. Como se ha señalado por Pandzic y Forchheimer (Panzic & Forchheimer, 2003), los esfuerzos de los investigadores de animación y gráficos estaban más centrados en los movimientos faciales que los parámetros causados, en lugar de los esfuerzos para elegir el mejor conjunto de parámetros.

Este enfoque hace inservibles los sistemas a través de dominios. Para hacer frente a estos problemas y proporcionar una parametrización de control facial estandarizada, el Moving Pictures Experts Group (MPEG) presentó las especificaciones de animación facial (FA) en el estándar MPEG-4. La versión 1 del estándar MPEG-4 (junto con la especificación FA) se convirtió en el estándar internacional en 1999.

El estándar MPEG-4 es compatible con la animación facial, proporcionando parámetros de animación facial (FAPs). Cowie et al. Indican la relación entre los MPEG-4 FAP y FACS UA: "MPEG-4, centrándose principalmente en la síntesis de la expresión facial y la animación, define los parámetros de animación facial (FAPs) que están fuertemente relacionados con las Unidades de Acción (UA), el núcleo de la FACS "(página 125 de (Cowie, y otros, 2008)). Para entender mejor esta relación entre el FAP y la UA, se menciona una breve introducción a algunas de las normas y terminologías que son relevantes para el reconocimiento de expresión facial MPEG-4.

El MPEG-4 define un modelo de la cara en su estado neutro para tener un conjunto específico de propiedades como a) todos los músculos de la cara se relajan; b) los párpados son tangentes al iris; c) la pupila es un tercio del diámetro del iris y así sucesivamente.

Las características clave como la separación del ojo, diámetro del iris, etc. se definen como neutral en este modelo de cara.

III.- Metodología de Desarrollo

La metodología utilizada en este trabajo se muestra en la figura 1.1, con el fin de reconocer de manera automática cuatro emociones humanas básicas (Ekman & Friesen, Constants across cultures in the face and emotion., 1971), (Pantic & Patras, Detecting facial actions and their temporal segments in nearly frontal-view face image sequences. In Systems, Man and Cybernetics., 2005).

Como se observa, la metodología consta de cinco etapas. Una breve descripción de la misma se ofrece a continuación. En la etapa 1 se obtiene el rostro de una persona utilizando la cámara de una computadora, la imagen se toma bajo condiciones ambientales. En la etapa 2 se identifica el rostro de una persona en la imagen tomada con anterioridad, en esta fase se aplica el algoritmo propuesto por Viola y Jones (Viola & Jones, 2004). Una vez que el rostro es detectado en la imagen se procede a la alineación aplicando los Promedios de Filtros Sintéticos Exactos (ASEF) (Felzenswalb, MacAllester, & Ramanan, 2008), en la etapa 3. Para realizar extracción de las características que representan las diferentes emociones faciales, esta tarea se realiza aplicando la técnica de Histograma de Gradientes Orientados (HOG) (Dalal & Triggs, 2005.), esta tarea se lleva a cabo en la etapa 4. El aprendizaje del modelo basado en las Máquinas de Vectores Soporte (MVS) (Cortes & Vapnik, 1995), utilizando las características seleccionadas en el paso anterior se realiza en la etapa 5, en esta etapa se tienen que considerar el cincuenta por ciento de las imágenes para el entrenamiento y cincuenta por ciento de imágenes para para la evaluación del modelo.

1.-Imagen de Entrada. 2. Detección Facial. 3.Normalización y alineación. 4. Extracción de Características. 5. Clasificación.

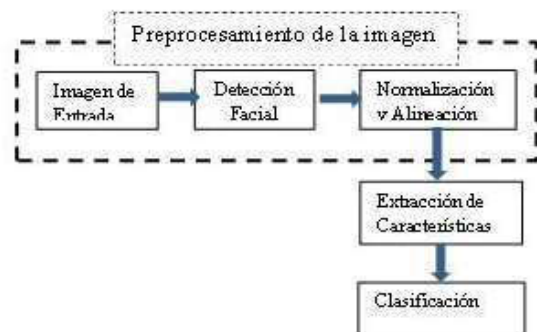


Figura 1. Etapas para el Reconocimiento de expresiones faciales.

Detección del Rostro.-La detección del rostro se lleva a cabo aplicando el algoritmo descrito por Viola y Jones (Viola & Jones, 2004).

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 del descrito por Viola y Jones 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 deslaza (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. Aplicación de la imagen integral y convolución de filtros en el rostro.

$$I(x, y) = \sum_{x' \leq x; y' \leq y} I(x', y') \quad 1$$

Donde:

Condiciones:

I(x, y).-Representa la imagen integral
 x'.-Menor o Igual a x

Im (X',Y').-Representa la imagen original
 y'.-Menor o Igual a y

Extracción de características para detectar el rostro

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, que a través de la imagen integral recorrerá la imagen facial de izquierda a derecha y de arriba hacia abajo seleccionando la información útil, es decir, la información que contiene los atributos que describen las características del rostro, y desechando la información que no sea de utilidad (los atributos que no contienen características del rostro). 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 4.5, donde se puede observar claramente que se está detectando un rostro en el recuadro que se marca en color verde en la imagen facial obtenida.

Obtención del Rostro

La técnica de *Boosting* fue introducida por Schapire y Freund (Freund & Schapire, 1997), este es un método de clasificación que utiliza varios clasificadores básicos para formar un único clasificador más complejo y preciso.

Los fundamentos se basan en que varios clasificadores sencillos que se desarrollan, cada uno de ellos con una precisión ligeramente superior en una clasificación aleatoria de los ejemplos de entrenamiento, pueden combinarse para formar un clasificador que sea de mejor 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 en las muestras de entrenamiento, entre mayor sea el número de muestras, habrá mayor precisión en los resultados obtenidos, como se muestra en los trabajos realizados por Viola y Jones (Viola & Jones, 2004).

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 (Freund & Schapire, 1997). 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 es mucho más preciso que cualquiera de los clasificadores base por separado. Como resultado de la combinación de los filtros Haar, la técnica de Boosting y el algoritmo de Viola y Jones, se muestra la imagen de la figura 3.

Localización de los Ojos

Generalmente los algoritmos de búsqueda de ojos utilizan las coordenadas (x, y) para ubicar los píxeles del centro de los ojos izquierdo y derecho en las imágenes frontales.

Para que esto sea verdadero, el algoritmo debe devolver la ubicación del ojo proporcionando cierta tolerancia, medida típicamente como una fracción de la distancia interocular, es decir, la distancia entre los centros de los ojos en el rostro. Específicamente en este trabajo, se propone la utilización de la clase de filtros denominada Promedio de Filtros Sintéticos Exactos (ASEF) (Felzenswalb, MacAllester, & Ramanan, 2008), por dos razones importantes en el desarrollo de este trabajo. En primer lugar, se especifica una superficie de respuesta de correlación completa para cada instancia de entrenamiento durante la construcción del filtro. En segundo lugar, el resultado de los filtros utilizados en cada imagen de entrenamiento se promedia para mostrar el objeto.



Figura 3. Ubicación del rostro y localización de los ojos
Delimitación de la Imagen

Después de haber detectado el rostro en la imagen de entrada, el siguiente paso es 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 despejado el rostro de la imagen que se muestra en la figura 4, 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 y que puedan alterar el contenido de la imagen facial.

Este punto es muy importante debido a que la imagen debe estar completamente despejada de cualquier objeto que pueda proporcionar información inadecuada que interfiera en el siguiente proceso, debido a que esto ocasionaría que los resultados que se obtengan no tendrían la información real para la extracción de características.

El rostro delimitado y alineado durante la fase de preprocesamiento se muestra en la figura 4.



Figura 4. Rostro delimitado obtenido de la imagen de entrada. (Imagen original obtenida con la cámara web).

Descriptor de Características

Un descriptor de características es una representación de una imagen que la simplifica al extraer información útil y descartar información irrelevante. Típicamente, un descriptor de características convierte una imagen 3D a un vector (conjunto de características de longitud n), en el descriptor de características HOG.

En el descriptor de características de HOG, la distribución (histogramas) de las direcciones de los gradientes (gradientes orientados) se utilizan como características. Los gradientes (derivados x e y) de una imagen son útiles porque la magnitud de los gradientes es grande alrededor de los bordes y esquinas (regiones de cambios abruptos de intensidad) y sabemos que los bordes y esquinas contienen mucha más información sobre la forma del objeto que las regiones planas del mismo.

En una investigación realizada por Dalal y Triggs (Dalal & Triggs, 2005.), encontraron que se pueden definir vectores de características de baja dimensión que son sensibles al contraste. En dichos estudios se ha encontrado que los rendimientos en algunas categorías de objetos mejoran el uso de características sensibles al contraste, mientras que algunas categorías se benefician del contraste de características insensibles. Por lo tanto, en la práctica, se utilizan vectores de características que incluyen ambos contrastes sensitivos y no sensitivos.

Como se ha mencionado anteriormente, en la práctica se puede utilizar una proyección analítica utilizando vectores dimensionales. En este trabajo de investigación se utilizan 108 vectores dimensionales, definidos por 27 sumas sobre diferentes normalizaciones, uno para cada canal de orientación, (**9 insensibles al contraste y 18 sensibles al contraste**) y **4 dimensiones que capturan la energía del gradiente general en bloques de diez celdas (i, j)**. Por lo tanto, el mapa de características final tiene un vector de **31 dimensiones** (Felzenswalb, MacAllester, & Ramanan, 2008).

El vector final de características se calcula con los siguientes datos:

W = ancho de la imagen entre el tamaño de la celda = $80 / 8 = 10$

H = alto de la imagen entre el tamaño de la celda = $96 / 8 = 12$

HOG = dimensión del vector de características de HOG = 31

Por lo tanto, la fórmula utilizada es la siguiente:

Tamaño final del vector de características = $W \times H \times HOG$

Tamaño final del vector de características = $10 \times 12 \times 31 = 3720$.

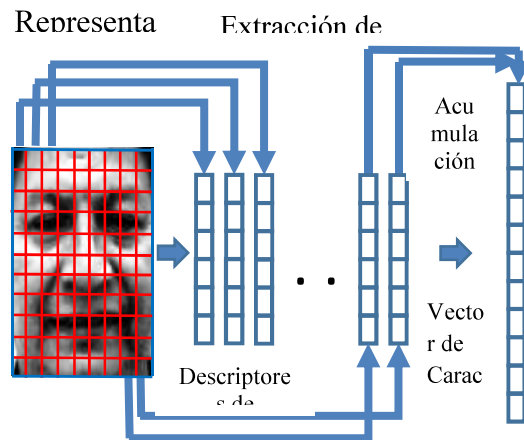


Figura 5. Rostro dividido en celdas para obtener los descriptores de características que se forman utilizando HOG, acumulados para formar el vector de características final.

Máquina de Vectores de Soporte

Una Máquina de Soporte Vectorial (MVS), es un sistema de aprendizaje automático que se ha utilizado para resolver problemas de clasificación y regresión de manera muy eficiente (Cortes & Vapnik, 1995), estas máquinas son capaces de clasificar muestras en dos posibles conjuntos de información “positiva” y “negativa”, que en este caso cuando realiza la detección de una imagen facial corresponden a “rostro” y “no rostro”. Desechando la información que no resulta útil para la detección del rostro (información negativa) respectivamente. Para realizar este proceso, se requiere de un entrenamiento previo de la máquina con esta información, por lo que se le introduce ejemplos de información “positiva” y “negativa”, que corresponden a las imágenes faciales para que realice dicha clasificación.

A cada uno de los vectores finales que se formaron con la suma de todos los descriptores de características de los bloques de celdas que componen la imagen se les asigna un número de etiqueta (correspondiente a cada imagen) para llevar a cabo su identificación y clasificación para que posteriormente alimenten el algoritmo de aprendizaje.

El proceso de alimentación de la Máquina de Vectores de Soporte con los parámetros (valores numéricos) que representan a cada uno de los vectores finales de características que describen los estados emocionales de las imágenes almacenadas en la base de datos, se realiza para clasificar a éstas con la finalidad de reconocer el estado emocional detectado en ella.

IV.-Resultados Obtenidos

Las pruebas que se realizaron para validar el preprocesamiento de las imágenes ayudan al clasificador (MVS) de expresiones faciales, debido a que la imagen que se captura con la cámara web puede ser afectada por los cambios de iluminación del medio ambiente.

En las pruebas realizadas aplicando la técnica de Tan y Triggs se utilizaron 470 imágenes con el estado emocional de contento, 210 imágenes con el estado emocional de enojado, 160 imágenes con el estado emocional de neutro y 112 imágenes con el estado emocional de sorpresa. Con la utilización de estas imágenes aplicando la técnica de normalización de Tan y Triggs, obtuvo un resultado de 92% de efectividad en el reconocimiento de las expresiones faciales.

Los porcentajes de reconocimiento de las expresiones faciales en las imágenes pueden mejorar si se realiza una buena base de datos de imágenes, esto es, evitando tener imágenes con expresiones ambiguas.

La Tabla 1. Muestra de los resultados que se obtuvieron al clasificar cuatro estados emocionales básicos aplicando esta técnica de normalización.

En este caso la confiabilidad del clasificador es del 92% mostrando un error del 8%, el incremento de la confiabilidad se debe a que las características que interesan para formar el vector utilizado en las etapas de entrenamiento y evaluación fueron resaltadas durante la normalización de la imagen.

| | Contento | Enojado | Neutral | Sorpresa |
|----------|----------|---------|---------|----------|
| Contento | 221 | 3 | 1 | 13 |
| Enojado | 3 | 217 | 12 | 6 |
| Neutral | 5 | 9 | 223 | 1 |
| Sorpresa | 12 | 1 | 7 | 218 |

Tabla 1.- Resultados obtenidos en el reconocimiento de los estados emocionales detectados y reconocidos por el clasificador MVS.

V.- Contribución del Proyecto

Las tecnologías computacionales avanzan muy rápidamente y el procesamiento de imágenes no es la excepción a estos avances. Los investigadores han realizado algunos sistemas utilizando diferentes técnicas como las redes neuronales y los algoritmos genéticos para el reconocimiento de objetos incluyendo el reconocimiento facial, pero no se ha utilizado la técnica del histograma de gradientes orientados en combinación con las máquinas de vectores de soporte para el reconocimiento facial y en éstos los estados emocionales de las personas, por lo tanto, esta es una nueva forma de hacer el reconocimiento facial.

En este proyecto se hace una combinación de algunos métodos para realizar el reconocimiento facial primeramente utilizando la técnica de la imagen integral y los filtros de Haar y después la técnica de histogramas de gradientes orientados haciendo uso del descriptor de características.

Finalmente se utiliza la técnica de la máquina de vectores de soporte que se alimenta con el vector de características para clasificar la información de las imágenes para detectar los estados emocionales. El uso y combinación de estas técnicas para la detección facial en un conjunto de imágenes producen un resultado de 92%, del reconocimiento de los estados emocionales en dichas imágenes, mejorando los resultados que se han obtenido con el uso de otras técnicas mencionadas anteriormente.

VI.- Conclusiones

En este trabajo se mostró una arquitectura basada en histogramas de gradientes orientados y máquinas de vectores de soporte para realizar esta tarea. Aunque estas técnicas ya se han utilizado de forma conjunta en otras aplicaciones (reconocimiento de personas, seguimiento de objetos) no se habían enfocado al reconocimiento del estado emocional humano. En este trabajo se utilizaron un total de 952 imágenes, es decir, participaron 238 personas, la mayoría estudiantes del Instituto Tecnológico de Acapulco, a cada uno de ellos se les pidió que actuaran de forma natural al mostrar sus emociones. Asimismo, las imágenes fueron capturadas en condiciones ambientales normales. Por lo tanto, se concluye lo siguiente.

1.- La arquitectura planteada inicialmente, basada en utilizar la técnica de HOG para la extracción de características integrada con las MVS's como mecanismo de clasificación genera resultados satisfactorios cuando se trabaja con imágenes obtenidas bajo condiciones no controladas.

2.- Los clasificadores basados en MVS's realmente obtienen resultados satisfactorios cuando trabajan con cantidades limitadas de datos.

VII.- Referencias

- Cowie, R., Douglas-Cowie, E., Karpouzis, K., Caridakis, G., Wallace, M., y Kollias, S. (2008). Reconocimiento de estados emocionales en la interacción natural humano-computadora. En *interfaces de usuario multimodales* (pp.119-153). Springer, Berlín, Heidelberg.
- Ekman, P., Friesen, W. V., & Ellsworth, P. (2013). *Emotion in the human face: Guidelines for research and an integration of findings*. Elsevier.
- Pantic, M., Pentland, A., Nijholt, A., & Huang, T. S. (2007). Human computing and machine understanding of human behavior: a survey. In *Artificial Intelligence for Human Computing*(pp. 47-71). Springer, Berlin, Heidelberg.
- Robinson, P., & El Kaliouby, R. (2009). Computation of emotions in man and machines.
- Ekamn, P., & Friesen, W. (1978). Facial action coding system (FACS): manual.
- Video, M. P. E. G. (1998). SNHC, "Text of ISO/IEC FDIS 14 496-3: Audio,". *Atlantic City MPEG Mtg.*
- Pandzic, I. S., & Forchheimer, R. (Eds.). (2003). *MPEG-4 facial animation: the standard, implementation and applications*. John Wiley & Sons.
- Viola, P., y Jones, MJ (2004). Robusta detección de rostro en tiempo real. *Revista internacional de visión por computador* , 57 (2), 137-154.
- Felzenszwalb, P., McAllester, D., & Ramanan, D. (2008, June). A discriminatively trained, multiscale, deformable part model. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (pp. 1-8). IEEE.
- Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (Vol. 1, pp. 886-893). IEEE.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
- Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, 17(2), 124.
- Pantic, M., & Patras, I. (2005, October). Detecting facial actions and their temporal segments in nearly frontal-view face image sequences. In *Systems, Man and Cybernetics, 2005 IEEE International Conference on* (Vol. 4, pp.3358- 3363). IEEE.
- Freund, Y., & Schapire, RE (1997). Una generalización teórica de la decisión del aprendizaje en línea y una aplicación para impulsar. *Diario de informática y ciencias del sistema* , 55 (1), 119-139.