

DETECCIÓN DE ROSTROS EN IMÁGENES DIGITALES USANDO CLASIFICADORES EN CASCADA

Faces Detection in Digital Images Using Cascade Classifiers

RESUMEN

En este artículo se presenta la implementación y aplicación de un método de detección de rostros y características faciales sobre imágenes digitales, basado en la extracción de características con filtros de base Haar y clasificadores en cascada que permitió un porcentaje de detección del rostro y de los ojos del 100% y 92% respectivamente en el mejor de los casos, con un tiempo de procesamiento bajo.

PALABRAS CLAVES: Detección de rostros, imagen integral, base Haar, *boosting*.

ABSTRACT

This article presents the implementation and application of a method to detect faces and facial features on digital images, based on Haar base filters feature extraction, and cascade classifiers, this techniques presents a 100% and 92% percentage for face and eyes detection respectively for the best of all cases, using a low processing time.

KEYWORDS: *Face detect, integral image, base Haar, boosting.*

MARTA LUCÍA GUEVARA

Ingeniera Electricista.
Profesor Catedrático
Universidad Tecnológica de Pereira
marluega@ohm.utp.edu.co

JULIAN DAVID ECHEVERRY

Ingeniero Electrónico, M. Sc.
Profesor Auxiliar
Universidad Tecnológica de Pereira
jdec@utp.edu.co

WILLIAM ARDILA URUEÑA

Licenciado en Física, M. Sc.
Profesor Titular
Universidad Tecnológica de Pereira
williamar@utp.edu.co

Grupo de investigación en Control e Instrumentación

1. INTRODUCCIÓN

El rostro humano es un objeto dinámico que tiene un alto grado de variabilidad en su apariencia lo cual hace que su detección sea un problema difícil de tratar en visión por computador. Inicialmente el problema de detección del rostro en los sistemas de reconocimiento no recibió la atención necesaria y se partió de que el rostro ya había sido detectado, fue solo en la década de los ochenta que surgieron los primeros algoritmos, basados en técnicas heurísticas y antropométricas, y en la década de los noventa cuando el desarrollo de algoritmos de detección rostros inició su crecimiento [1], proponiéndose una gran variedad de técnicas, desde algoritmos básicos de detección de bordes hasta algoritmos compuestos de alto nivel que utilizan métodos avanzados de reconocimiento de patrones.

Estas técnicas de detección se han abordado desde diferentes enfoques: *Enfoques basados en rasgos faciales o características locales*, en los que se buscan determinados elementos que componen el rostro, como los ojos, la nariz, la boca, *Enfoques holísticos o basados en la imagen*, en este caso los métodos trabajan con la imagen completa o zonas concretas de la misma de la cual se

extraen características que puedan representar el objeto buscado, *Enfoques híbridos*, estos métodos usan tanto la información local como la global para la detección, basándose en el hecho de que el sistema de percepción humano distingue tanto las características locales como globales del rostro [1, 2].

Alrededor de los enfoques mencionados se han planteado diferentes trabajos como los de [3, 4, 5], en los que se usa la información del color de la piel para realizar la detección, obteniendo resultados alrededor del 90%, [6] usa redes neuronales para segmentar el rostro alcanzando porcentajes de detección entre 77,9% y 90,3% para las diferentes configuraciones de la red, [1] usa una base Haar para la extracción de características y *Adaboost* para la selección de estas y clasificación, alcanzando un porcentaje de detección del 94,1%, este método propuesto por Paul Viola y Michael Jones, es uno de los métodos más usados hoy en día ya que ha permitido segmentar múltiples rostros en una imagen con tiempos de procesamiento bajos. Alrededor de esta investigación se han realizado otros trabajos como los de [8, 9, 10, 11], que han aplicado el método para segmentar rostros y han adicionado otras etapas como la detección de

características faciales (ojos y boca), corrección de pose y seguimiento del rostro.

Este documento presenta la implementación de una técnica de detección de rostros frontales como etapa inicial de un sistema automático para el reconocimiento de emociones partir del análisis del movimiento y deformación del rostro.

La estructura de este trabajo se dividió así: en la sección 2 se presenta la metodología a seguir, en la sección 3 se presenta el marco conceptual que sustenta la metodología, en la sección 4 se hará una descripción de la base de datos usada, de las pruebas realizadas y los resultados, y por último se presentaran las conclusiones.

2. METODOLOGÍA

La metodología se basada en la propuesta en [7]. Esta se divide en tres etapas como se muestra en la figura 1: en la primera se realiza una transformación de la imagen generando una nueva llamada imagen integral, en la segunda etapa se realiza la extracción de características usando filtros con base Haar, y por último se usa *boosting* para la construcción de clasificadores en cascada

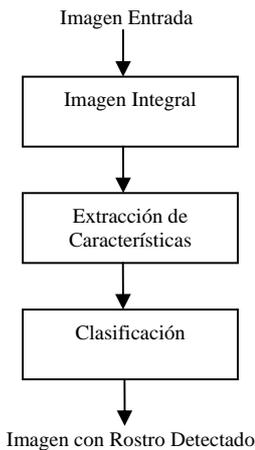


Figura 1. Diagrama de bloques

3. MARCO CONCEPTUAL

Imagen Integral

Esta nueva representación de una imagen fue introducida en [7]. Esta imagen permite extraer de forma rápida características a diferentes escalas ya que no se trabaja directamente con los valores de intensidad si no con una imagen acumulativa que se construye a partir de operaciones básicas.



Figura 2. Imagen integral

La imagen integral (ver figura 2), en la localización x, y , contiene la suma de los píxeles de la parte superior izquierda de la imagen y se puede calcular como se indica a continuación:

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

Donde $II(x,y)$ es la imagen integral e $Im(x,y)$ es la imagen original.

Extracción de Características

En imágenes las características de cada objeto se extraen al aplicar ciertas funciones que permitan la representación y descripción de los objetos de interés de la imagen (patrones). La extracción de características es un paso en el reconocimiento de patrones en el cual las medidas u observaciones son procesadas para encontrar a tributos que puedan ser usados para asignar los objetos a determinada clase [12].

En la metodología seguida, la extracción de características es realizada aplicando a la imagen filtros con bases Haar. Estos filtros pueden ser calculados eficientemente sobre la imagen integral, son selectivos en la orientación espacial y frecuencia, y permiten ser modificados en escala y orientación. En la figura 3, se muestran algunos de los filtros usados para la extracción de características.

Los filtros con bases Haar, realizan una codificación de diferencia de intensidades en la imagen, generando características de contornos, puntos y líneas, mediante la captura de contraste entre regiones.

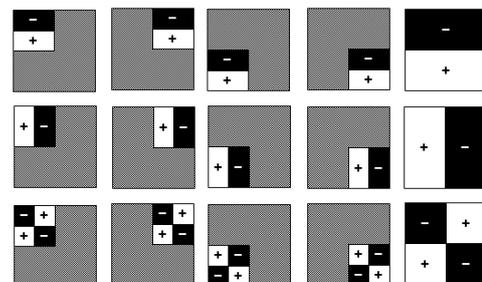


Figura 3. Filtros Haar rotados, trasladados y con cambios de escala

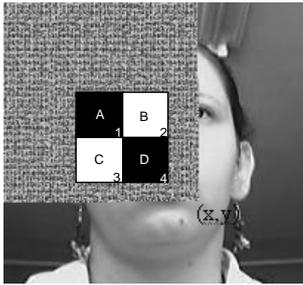


Figura 4. Convolución del filtro Haar con una imagen integral

En la figura 4, se muestra la convolución de un filtro Haar con la imagen integral. De esta operación se puede extraer una característica en un tiempo constante sobre la imagen integral adicionando y sustrayendo los valores de los vértices para cada rectángulo. Para mayor claridad, en la figura la suma de los píxeles que forman el rectángulo D se puede calcular como:

$$sumD = (4 + 1) - (2 + 3) \tag{2}$$

Donde 1, 2, 3, 4, son los valores dados en la imagen integral en dichas localizaciones.

Clasificación

Esta etapa dentro del algoritmo de detección se encarga de asignar un conjunto de características dado a una clase con la que se encuentra una mayor similitud, de acuerdo a un modelo inducido durante el entrenamiento [13].

Boosting fue introducido por [13], este es un método de clasificación que combina varios clasificadores básicos para formar un único clasificador más complejo y preciso. La idea se basa en la afirmación de que varios clasificadores sencillos, cada uno de ellos con una precisión ligeramente superior a una clasificación aleatoria, pueden combinarse para formar un clasificador de mayor precisión, siempre y cuando se disponga de un número suficiente de muestras de entrenamiento. La aplicación de clasificadores en cascada ha permitido obtener buenos resultados, estos son reportados en [7], [8], [14]. En la figura 5, se muestra un esquema de un clasificador en cascada.

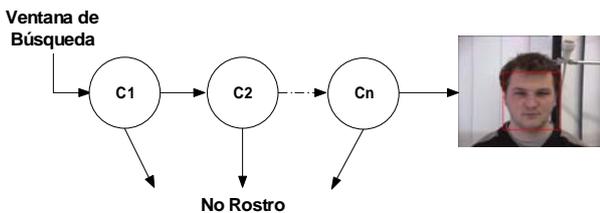


Figura 5. Clasificador en cascada

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 de entrenamiento [13]. Finalmente, estos clasificadores base se combinan en un único clasificador que se espera sea mucho más preciso que cualquiera de los clasificadores base por separado.

En función de los clasificadores base que se utilicen, las distribuciones que se empleen para entrenarlos y el modo de combinarlos, podrán crearse distintas clases del algoritmo genérico de *boosting*. El algoritmo de *boosting* empleado por Viola y Jones en su trabajo es conocido como *AdaBoost*.

A continuación se muestra el algoritmo usado en la metodología. Este se encuentra planteado en [13].

- Dado un conjunto de imágenes $(x_1, y_1), \dots, (x_n, y_n)$ donde $y_i = 0, 1$ para muestras negativas y positivas respectivamente.

- Inicializar los pesos $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ para $y_i = 0, 1$, donde m es el número de muestras negativas y l es el número de muestras positivas.

- Para $t = 1, \dots, T$:

1. Normalizar los pesos $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$

2. Seleccionar el mejor clasificador base respecto al peso del error:

$$\epsilon_t = \min_{f,p,\theta} \sum_i w_i |h(x_i, f, p, \theta) - y_i|$$

3. Definir $h_t(x) = h(x, f_t, p_t, \theta_t)$ donde f_t, p_t y θ_t son usadas para minimizar ϵ_t .

4. Actualizar los pesos: $w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$ donde $e_i = 0$ si la muestra x_i es clasificada correctamente, o $e_i = 1$ en otro caso, con $\beta_t = \frac{\epsilon_t}{1 - \epsilon_t}$.

5. El clasificador robusto final es:

$$C(x) = \begin{cases} 1 & \sum_{i=1}^T \alpha_i h_i(x) \geq \frac{1}{2} \sum_{i=1}^T \alpha_i \\ 0 & \text{en otro caso} \end{cases}$$

4. PRUEBAS Y RESULTADOS

Base de datos FEETUM

Pertenece a la Universidad Técnica de Munich (Alemania). Esta conformada por imágenes del rostro de 18 personas que muestran las 6 emociones básicas definidas por Eckman y Friesen. Esta base de datos a sido generada como parte del proyecto FG-NET (*Face and Gesture Recognition Research Network*), cada foto se encuentra en formato JPEG a 8 bits con un tamaño de 320x240 píxeles.

Pruebas

La prueba se realizó sobre 50 imágenes pertenecientes a dos personas de la base de datos FEETUM. A cada imagen se le aplicó el método de detección de rostros y se midió el tiempo de procesamiento por imagen. Los resultados se muestran a continuación:

	Persona 1	Persona 2
No de Imágenes	25	25
No de Detecciones	24	25
No de Fallos	1	0
% de Detección	96%	100%
Tiempo de Detección Promedio	40.24ms	38.45ms

Tabla 1. Resultados de detección del rostro.

El porcentaje de detección en este artículo se calcula como la relación de la cantidad de rostros detectados sobre el número de imágenes de usadas.

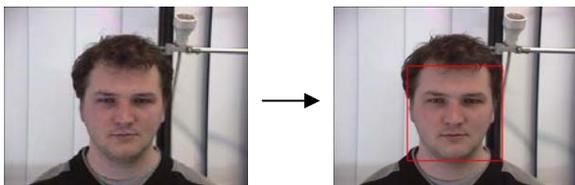


Figura 6. Ejemplo de detección del rostro para la persona 1.

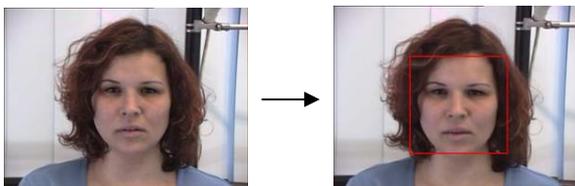


Figura 7. Ejemplo de detección del rostro para la persona 2.

Teniendo en cuenta que el proyecto se enfoca en el reconocimiento de expresiones faciales y para esto es necesario centrar el análisis en regiones específicas del

rostro como los ojos y la boca, y al ver la versatilidad de este algoritmo y su posible extensión a la detección de otras regiones, este fue entrenado para detectar características del rostro (ojos). Los resultados obtenidos para la detección de los ojos se muestran a continuación:

	Persona 1	Persona 2
No de Imágenes	25	25
No de Detecciones	22	23
No de Fallos	3	2
% de Detección	88%	92%
Tiempo de Detección Promedio	8.76ms	8.84ms

Tabla 2. Resultados de detección de ojos.

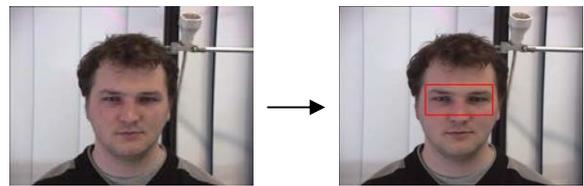


Figura 8. Ejemplo de detección de los ojos para la persona 1.

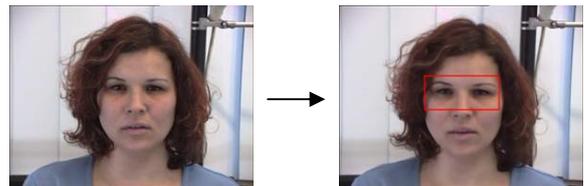


Figura 9. Ejemplo de detección de los ojos para la persona 2.

5. CONCLUSIONES

Los altos porcentajes de detección y bajos tiempos de procesamiento demuestran la efectividad de la combinación de técnicas de *boosting* con filtros de base Haar, para la detección del rostro y características faciales y dan la posibilidad de llevar el sistema a una aplicación en tiempo real.

La reducción del desempeño del método en la detección de ojos se atribuye a que el algoritmo debe realizar la búsqueda sobre toda la imagen. Esto se puede mejorar aplicado el método únicamente sobre la región que ha sido detectado como rostro, ya que se estaría descartando información del fondo, y además se disminuiría el tiempo de procesamiento.

6. AGRADECIMIENTOS

Este trabajo se desarrollo en el marco del proyecto de investigación: "IMPLEMENTACIÓN Y EFECTIVIDAD DE UN SISTEMA BASADO EN INTELIGENCIA

ARTIFICIAL COMO HERRAMIENTA PARA EL TRATAMIENTO PSICOLÓGICO DE PERSONAS CON TRASTORNO DE ESTRÉS POSTRAUMÁTICO", financiado por Colciencias. Código 111037019600 y la Universidad Tecnológica de Pereira. Código 511-3-243-08.

7. BIBLIOGRAFÍA

- [1] RAMÍREZ, Carlos Alejo, PÉREZ, Manuel David. Detección de caras y análisis de expresiones faciales.
- [2] LECUMBERRY, R., Federico. Cálculo de disparidad y segmentación de objetos en secuencias de video. Tesis de maestría en ingeniería eléctrica. Universidad de la república Montevideo, Uruguay. 3 de agosto de 2005.
- [3] LI, Yadong, GOSHTASBY, Ardeshir y GARCÍA, Oscar. *Detecting and tracking human faces in videos*. Wright State University, 2000.
- [4] FERIS, Rogério, EMÍDIO DE CAMPOS, Teófilo y MARCONDES, César. *Detection and tracking of facial features in video sequences. Lecture Notes in Artificial Intelligence*, Vol. 1793, pp. 197-206, 2000.
- [5] HABILI, Nariman. *Automatic segmentation of the face and hands in sign language video sequences*. Technical report, Adelaide University, Australia, 2001.
- [6] ROWLEY, Henry, BALUJA, Shumeet, and KANADE, Takeo. *Neural network-based face detection*. IEEE, 1998.
- [7] VIOLA, Paul y JONES, Michael. *Robust Real-Time Face Detection. International Journal of Computer Vision*, Vol.57, No.2, pp 137-154, 2004.
- [8] CASTRILLÓN, M., DÉNIZ, O., GUERRA, C., and HERNÁNDEZ, M. "ENCARA2: Real-time detection of multiple faces at different resolutions in video streams". *ELSEVIER, Journal Visual Communication and Image Representation*. Article in press (2007).
- [9] BARTLETT, Marian Stewart et all. *Real time face detection and facial expression recognition: development and applications to human computer interaction*. University of California, USA, 2004.
- [10] SILAPACHOTE, P, KARUPPIAH, D y HANSON, A. *Feature selecting using adaboost for face expression recognition*. University of Massachusetts Amherst, USA.
- [11] BELAROUSSI, Rachid and MILGRAM, Maurice. *Face detecting and skin color based tracking: a comparative study*. University Pierre an Maire Curie. Paris, Francia, 2007.
- [12] IEEE *Standar Glossary for Image Processing and Pattern recognition Terminology*. Published by the Institute of Electrical and Electronics Engineers, New York, USA, 1990.
- [13] SCHAPIRE, R and FREUND, Y. *A decision theoretic generalization of on-line learning and application to boosting*. AT&T Bell Laboratories. USA, 1995.
- [14] PAPAGEORGIOU,C., OREN,M. and POGGIO,T. *A general frame-work for object detection*. Proceedings of International Conference on Computer Vision, enero de 1998, Bombay, India.

