

Una Técnica para la Localización de Ojos Humanos en una Imagen Bidimensional

Luna Esteban Bruce Leroy,
Mendoza Campa Juan Manuel,
Parra Bautista Roberto,
ESIME Culhuacan - IPN
M. en C. Aguilar Jáuregui María Elena
Profesora del CIC-IPN

El presente trabajo propone una técnica para la localización de ojos de personas en una imagen bidimensional. Los vectores característicos del ojo humano se obtuvieron a partir de una base de datos de imágenes de 300 X 300 píxeles.

El sistema fue desarrollado bajo una plataforma C++ e implementado en una PC con un procesador Pentium IV a 2.40 GHz.

INTRODUCCIÓN

Dentro del campo de la visión artificial, la detección y el rastreo de ojos juegan un papel muy importante ya que con estas funciones se puede identificar a una persona y/o interpretar su estado emocional. Por lo tanto, es crucial un sistema de detección y/o rastreo de ojos eficiente para poder tener una mejor interacción humano – computadora.

A pesar de muchos esfuerzos todavía no existen métodos robustos y 100% precisos para detectar y rastrear ojos humanos.

El desarrollo actual se puede dividir en dos categorías: los que se basan sobre imágenes en línea y los de fuera de línea.

Este trabajo se enfoca en la segunda categoría, ya que fue necesario considerar la complejidad de los algoritmos y los tiempos de procesamiento.

Aún cuando existe una gran variedad de tipos de ojos humanos en cuanto a rasgos se refiere, es posible obtener sus características propias o descriptores, los cuales permiten afirmar que en verdad se trata de un ojo y no de algún objeto parecido en cuanto a forma.

EXTRACCIÓN DE CARACTERÍSTICAS

La forma de los ojos brinda características muy importantes para su reconocimiento. Debido a que estas formas son muy variadas, en la metodología propuesta la segmentación se realizó de forma manual, como se muestra en la **figura1**.

Los formatos de imágenes que se evaluaron fueron: bmp (Bitmap), jpeg (Joint Photographic Expert Group) y gif (Graphics Interchange Format) debido a que son los más usados en computación. Bmp consiste en un formato sin pérdida de información; por otra parte, jpeg es un formato de compresión de imagen que no produce pérdida de color aunque la compresión elimina datos, mientras que gif es un formato con un alto grado de compresión y mayor pérdida de color^[2].

Para lograr mejores resultados en cuanto al funcionamiento del sistema se utilizó el formato bmp para la extracción de características.

Aunque el sistema visualiza las imágenes en color verdadero, el trabajo del sistema se basa únicamente en el canal verde que es el que presenta información con menos ruido.

Para la selección de la información

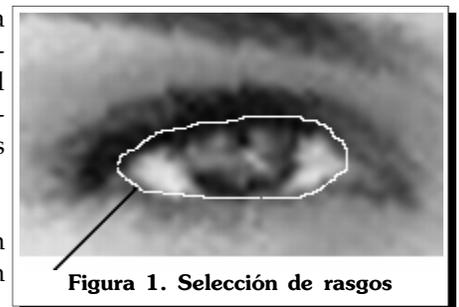


Figura 1. Selección de rasgos

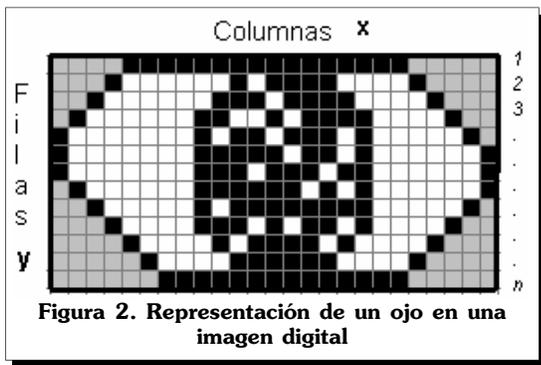


Figura 2. Representación de un ojo en una imagen digital

cuantitativa de nuestro interés se deben obtener los valores de brillo de cada pixel en sus coordenadas (x, y); la manera de realizar esto es mediante el uso de una matriz bidimensional (x, y). Considerando que la forma del ojo es casi elíptica, es difícil representar la intensidad de brillo de los píxeles en forma matricial.

La figura 2 representa la forma (aproximada) de un ojo en una imagen digital, en la que se muestran las coordenadas (x, y) de la misma. Los píxeles en color gris representan ruido o información inútil para describir un ojo, por consiguiente se consideran como un valor X (sin importancia) en la matriz. Dado esto, fue necesario transformar la matriz resultante en un vector siguiendo los criterios siguientes:

1. Se consideran los píxeles pertenecientes o internos al borde.
2. Se descartan los píxeles externos al borde.
3. Los primeros valores del vector corresponden a los píxeles que radican en la primera fila y que obedecen al criterio número 1.
4. Los valores posteriores del vector corresponden a los valores obtenidos a partir de la n fila y así sucesivamente hasta la última fila que obedezca al criterio 1.

$$\text{OJO} = [a_{11}, \dots, a_{1m}, a_{21}, \dots, a_{2m}, \dots, a_{n1}, a_{nm}]$$

Donde:

- a → cualquier valor de intensidad de brillo.
- n → las filas pertenecientes o internas al borde.
- m → son las columnas pertenecientes o internas al borde.

Una vez obtenido el vector se obtiene el histograma o función de densidad que corresponde a una función discreta $p(r_k) = n_k/n$, donde r_k es el k-ésimo nivel de gris, n_k es el número de píxeles de la imagen con ese nivel de gris, n es el número total de píxeles de la imagen y k es el rango de niveles de gris [0,255] [1]. Como se mencionó

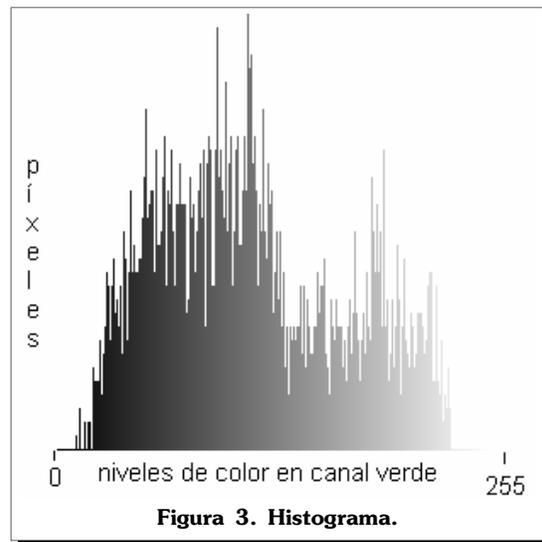


Figura 3. Histograma.

anteriormente, para este trabajo se utiliza el canal verde de las imágenes RGB en lugar de los niveles de gris, pero el rango sigue siendo el mismo.

La figura 3 muestra el histograma del área del ojo seleccionado.

MOMENTOS INVARIANTES

Estos momentos codifican la superficie del objeto y son invariantes a traslación, rotación y escala. Estos momentos de orden p+q de una imagen I(x, y) se definen como:

$$M_{p,q} = \sum_{x=1}^N \sum_{y=1}^M x^p y^q I(x, y)$$

Para realizar una descripción del objeto independientemente de su posición se utilizan los momentos centrales que se obtienen a partir de la fórmula siguiente:

$$\mu_{p,q} = \sum_{x=1}^N \sum_{y=1}^M (x - \bar{x})^p (y - \bar{y})^q I(x, y)$$

Los momentos centrales normalizados se utilizan para describir al objeto, independientemente del tamaño que tengan, y están definidos por la fórmula:

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu_{0,0}^\alpha} \quad \alpha = \frac{p+q}{2} + 1$$

Posteriormente se extraen los momentos invariantes de Hu por medio de las fórmulas siguientes:

$$\phi_1 = \eta_{20} + \eta_{02}$$

$$\phi_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2$$

$$\phi_3 = (\eta_{30} + 3\eta_{12})^2 + (3\eta_{21} + \eta_{03})^2$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

$$\phi_7 = (3\eta_{21} - \eta_{30})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

Wu Yan^[3] propone añadir información del borde y del histograma del objeto para obtener características que son invariantes a cambios en el nivel de gris de la imagen^[4]. Estos momentos se obtienen a partir de los momentos unidimensionales de la imagen y se definen como:

$$\beta_1 = \eta_4 / \eta_2^2$$

$$\beta_2 = \eta_5 / \eta_2 \eta_3$$

$$\beta_3 = \eta_6 / \eta_2 \eta_4$$

De esta forma, el vector característico que va a representar un ojo esta formado por 4 momentos de Hu y los 3 momentos de Yan.

En cada uno de los vectores característicos no se consideraron los tres últimos momentos de Hu debido a que tienen valores muy pequeños en comparación con los restantes, por tal motivo se optó por descartarlos.

Vector característico = $[\phi_1, \phi_2, \phi_3, \phi_4, \beta_1, \beta_2, \beta_3]$

RED NEURONAL DE RETROPROPAGACIÓN

Una de las aplicaciones más extendidas de la red neuronal de retropropagación (BPN, Back Propagation Network) es el reconocimiento o clasificación de patrones^[5].

Para éste trabajo la red diseñada se compone de siete neuronas de entrada debido a que el vector característico tiene siete momentos y una neurona de salida con valor igual a 1 (describe a un ojo).

Para el entrenamiento de la red neuronal se varía el número de capas ocultas y el número de neuronas por capa con el propósito de verificar que arquitectura es más eficiente para disminuir el error de aprendizaje (figura 4).

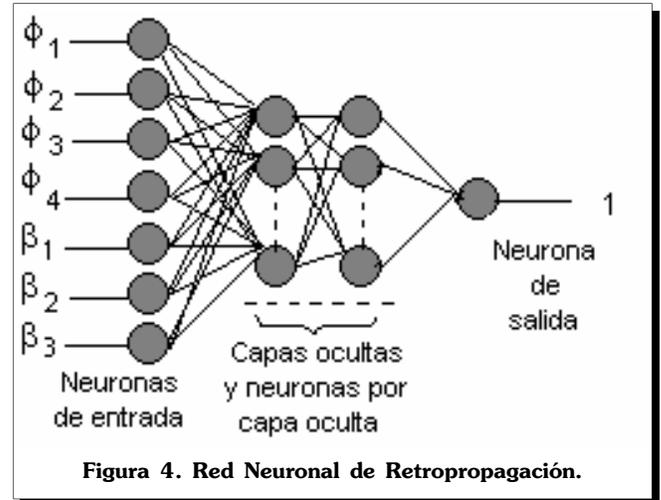


Figura 4. Red Neuronal de Retropropagación.

En lo que respecta a la fase de prueba de la BPN, se utilizarán imágenes de igual tamaño que las imágenes que sirvieron como muestra para la extracción de características; para ello será necesario realizar un barrido sobre la imagen pixel por pixel hasta que la red encuentre las coordenadas (x, y) de la imagen que describan a un ojo.

Las imágenes a utilizar para la prueba del sistema serán en formato bmp para la evaluación de la eficiencia del mismo; posteriormente se emplearan los formatos de compresión para comparar los resultados obtenidos tomando en cuenta diferentes tamaños de imagen.

CONCLUSIONES

Para el funcionamiento óptimo del sistema se evalúa el número de capas ocultas necesarias que se emplean en la red neuronal, con esto se esperan resultados satisfactorios y una reducción de costos en cuanto a tiempo de procesamiento y complejidad de programación.

El uso de este sistema puede servir como una base para el desarrollo de diversos sistemas que se basen en

la extracción de características de rostros humanos y el reconocimiento de los patrones resultantes en imágenes digitales. Las aplicaciones posibles de este sistema pueden ser el reconocimiento de personas y en un futuro la autenticación de las mismas, esto a partir de sus ojos.

REFERENCIAS

- [1] González, R. C., Woods, R. E.; *"Tratamiento Digital de Imágenes"*, Addison Wesley, 1996.
- [2] Margulis, Dan; *"Professional Photoshop, Color Correction, Retouching and Image manipulation with Adobe Photoshop"*, Wiley, 1995.
- [3] Wu Yan, Ding Mingyue, Ann Peng Jiexiong; *"Research on using moment invariants in scene matching"*, Institute of Image Recognition and Artificial Intelligence, HuaZhong. University of Science and Technology, WuHan, Hubei, China, 430074).
- [4] Ming-Kuei Hu; *"Visual pattern recognition by moment invariants"*, Information Theory, IEEE Transactions on, Volume: 8, Issue: 2, Feb 1962, Pages:179 – 187.
- [5] José R. Hilera, Víctor J. Martínez; *"Redes Neuronales Artificiales, fundamentos, modelos y aplicaciones"*, Alfaomega Ra-ma, 2000.
- [6] Kennet R. Castleman; *"Digital Image Processing"*, Prentice Hall, 1996.
- [7] Maria Petrou, Panagiota Bosdogianni; *"Image Processing, The fundamentals"*, Wiley, 1999.
- [8] A. Pérez, M. L. Córdoba, A. García, R. Méndez, M. L. Muñoz, J. L. Pedraza, Sánchez; *"A Precise Eye-Gaze Detection and Tracking System"*, Departamento de Arquitectura y Tecnología de Sistemas Informáticos (DATSI), Universidad Politécnica de Madrid.
- [9] D. Davies, P. Palmer, & M. Mirmehdi; *"Detection and Tracking of Very Small Low Contrast Objects"*, School of Electrical Engineering, Information Theory & Mathematics, University Surrey, England.
- [10] T. D'Orazio, M. Leo, A. Distante; *"Eye detection in face images for a driver vigilance system"*, CNR-ISSIA, Italia.
- [11] R. Thilak kumar, S. Kumar Raja and A. G. Ramakrishnan; *"Eye Detection using color cues and projection functions"*, Department of Electrical Engineering, Indian Institute of Science.
- [12] Aguilar Ma. Elena, Sossa Humberto, *"Detección y Localización de Rostros Humanos en Imágenes Usando descriptores invariantes y Redes Neuronales"*, Informe Técnico, serie Azul, CIC-IPN, 2000.
- [13] Christian Bird and Bhaskar Shrestha; *"Tracking the Human Eye"*, University of Vermont.