

SISTEMA DE RECONOCIMIENTO DE IMÁGENES COMO INTÉRPRETE DEL LENGUAJE DE SEÑAS

Ordóñez Ortiz Edgar

e_ordonez@ieee.org

Escuela Superior Politécnica de Chimborazo
Riobamba-Ecuador

ABSTRACT.- The present work raises a project applied to the recognition of objects in images by means of the Transformed one of Distance using Matlab. The Distance of the Transformed one is a mathematical tool used to represent the information of objects be already of visual or acoustic type; It is a simple but effective method for the recognition of objects in accordance with his form and position inside the image. Nevertheless, it will be projected further away implementing an Integral System of Recognition of the American Language of Signs (ASL) in Real-time.

PALABRAS CLAVES: Distancia de Transformada, Matlab, Reconocimiento de imágenes, ASL.

I. INTRODUCCIÓN

Independiente del tipo de tecnología es indudable la importancia en el mundo de hoy del material audiovisual en donde dado su natural relación con los sentidos hace mas intuitiva su percepción por las personas.

Esta **percepción** en sí no es más que un reconocimiento en la que se establece previamente un modelo gráfico estadístico (SGM *Statical Graphical Models* [1]) basado en la información que para nosotros es familiar, Ver fig. 1.

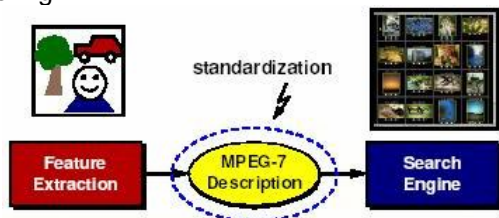


Figura 1. MPEG7, Un proyecto que se basa en la descripción de esquemas gráficos

Apoyado en este principio, partimos de un modelo de reconocimiento básico pero muy útil enfocado a imágenes bidimensionales, como es la Distancia de la Transformada de una función. Este proceso calcula la distancia de un

punto $p(x,y)$ o de un grupo de puntos con respecto a otro u otros tomados como referencia. El proceso se facilita ya que Matlab ofrece funciones que permiten agilizar este proceso:

```
distgc = distance(-15,0,60,150)
distgc =
    129.9712
```

Una vez establecido el proceso, se identifica la referencia, esta es la Forma de los objetos. La forma de un objeto siempre queda representado por su contorno. Entre los métodos matemáticos, el de **Canny**[7] permite obtener el contorno de un objeto en una imagen (Ver fig. 2);



Figura 2. Detección de Bordes con el método Canny

Una imagen al ser representada digitalmente como una matriz (es decir un grupo de puntos) y por ende también su contorno, permite calcular su distancia. Si se compara esta distancia, con la de referencia se puede verificar si son iguales (distancia cero) o similares (distancia mayor que cero) estableciéndose el reconocimiento del objeto.

II. DESARROLLO DE LA APLICACIÓN

Se tomará para el desarrollo de la aplicación un Programa para el Reconocimiento de Lenguajes de Señas en Tiempo Real.

El modelo gráfico a utilizar sin importar el procesamiento que apliquemos en la ejecución se basa en el mismo principio que en algún momento se planteara para estándares como MPEG7 ó MPEG21, Ver fig. 3.

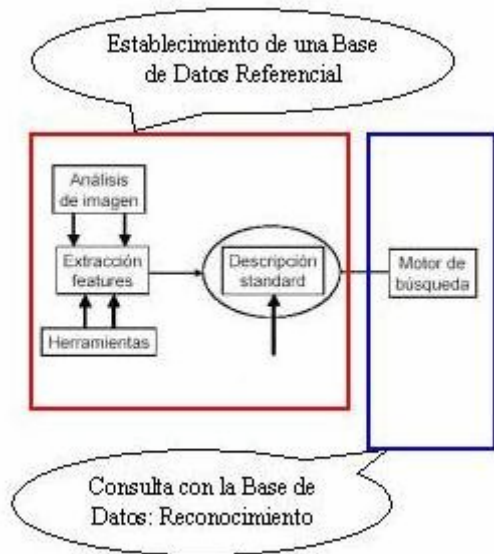


Figura 3. Modelo Gráfico usado en la aplicación.

La identificación del modelo previo a su diseño comprende el establecimiento de las etapas de:

- Adquisición de la imagen
- Extracción de Características
- Definición de esas Características
- Reconocimiento requerido por un usuario.

A. Adquisición de la Imagen, el análisis se limita a cualquier imagen bidimensional (ver fig. 13); en la fig. 4 muestra ejemplos de imágenes bidimensionales.

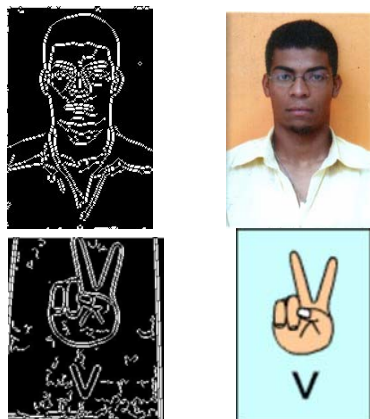


Fig. 4. Proceso de Adquisición de imágenes bidimensionales

La adquisición de esta imagen se realizará por medio de una Webcam tanto como en la creación de la base de datos como en el reconocimiento; se usará un toolbox en Matlab llamado **VFM**, ver Fig. 5 y Fig. 6.



Figura 5. Toolbox VFM (*Visión for Matlab*) desarrollado por Farzad Pezeshkpour permite importar la imagen hacia una ventana en Matlab y obtenerla en formato Binario.



Figura 6. Ejemplos de la Adquisición, la calidad de las imágenes dependerá de la Webcam y condiciones de luz.

La adquisición de la imagen presenta un problema ya que para el reconocimiento de la imagen esta debe mantener la misma posición y condiciones de luminosidad homogéneas; estas condiciones son muy difíciles de mantener por las características de la Webcam así que se adquieren las fotos en tiempo real pero desde imágenes montadas en un pedestal e iluminadas artificialmente para mantener las condiciones¹, Ver Fig. 7.

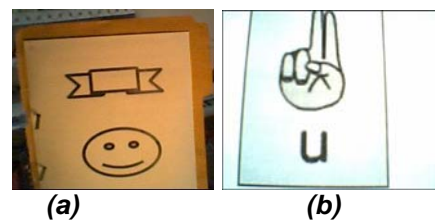


Figura 7. (a) Imagen iluminada con luz amarilla
(b) Imagen usando luz Fluorescente

Con la foto adquirida; se inicia el aprendizaje ó creación de la base de datos a partir de estas imágenes. Las distintas etapas que se usarán son [4]:

B. Detección de Contornos, primero se convierte la imagen RGB (refiriéndose a una imagen a color Red Green Blue) a imagen de intensidad; ver Fig. 8.

¹ Si embargo si está aplicado en forma real, Ver las conclusiones

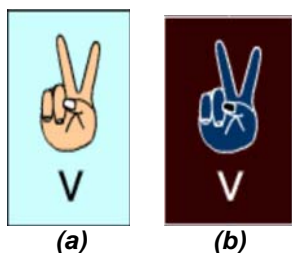


Figura 8. (a) Imagen RGB a procesar. (b) Conversión de la Imagen en imagen de Intensidad

Luego se obtiene el contorno de la Imagen (ver Fig.9), para este caso se emplea el método Canny; este método ofrece ventajas sobre otros (Método Sobel, Método Prewitt, Método Roberts, Método Cruce por Cero, Método Laplaciano-Gaussiano) porque se deben de realizar elecciones de los valores umbral (corte) y del tamaño de las máscaras a utilizar (el tamaño condiciona el grado de suavizado, el cual puede afectar a las detecciones por cruce por cero y al máximo gradiente sobre una imagen en el método Roberts o Laplaciano).



Figura 9. Obtención del contorno del elemento

Para mejorar el nivel del contorno de la imagen; se usa procedimientos morfológicos (abrir, cerrar, encoger, adelgazar, esbozar, engrosar, rellenar, etc.) que permiten aumentar o remover píxeles de la imagen binaria de acuerdo a ciertas reglas que dependen del patrón de vecindad del píxel que se modifica.

C. Selección del Objeto, El comando **bwselect** permite seleccionar objetos dentro de una imagen binaria, en este caso, el comando muestra la imagen en pantalla y permite seleccionar unas coordenadas de referencia con el ratón, (Ver Fig. 10).



Figura 10. Selección y Procesamiento del Elemento

En esta parte se obtiene la información de referencia.

D. Calculo de la Transformada, tanto la imagen como el contorno se representa digitalmente en matrices; terminado los pasos anteriores las imágenes ingresadas representan la base de datos referencial y cada una de ellas se almacena en una matriz distinta.

La distancia del grupo de puntos $p_1(x,y), p_2(x,y), \dots, p_N(x,y)$ al objeto de referencia es

$$\frac{1}{N} \sum d_a(P_i) \quad (1)$$

La transformada de distancia es el método de representación que aplica la función de distancia en todos los píxeles de la imagen tomando como referencia el contorno del objeto que se quiere reconocer. El valor de distancia que toma cada píxel es el menor valor que se obtiene de una comparación de un operador asimétrico. La forma del operador se muestra en la Fig.11 [4].

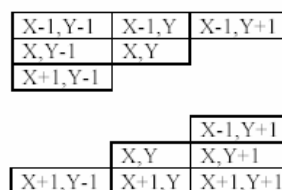


Figura 11. Operador Asimétrico de la Transformada de un arreglo matricial.

Se llevan a cabo dos barridos en la matriz en direcciones opuestas. Para cualquier píxel, la menor distancia es la que tiene con su vecindad horizontal o vertical. Al comparar cualquier píxel y su operador, se involucra la distancia entre los píxeles que se compara y el valor del píxel de referencia como se muestra en la Ecuación (2),

$$p_r(x,y) = D_{1r} + p_1(x_1, y_1) \quad (2)$$

Donde p_r es el píxel central en el operador, D_{1r} es la distancia que existe entre el píxel central y el píxel con el que se compara. Se calcula la distancia entre dos puntos $p_1(c_1, r_1)$ y $p_r(c_r, r_r)$; de los píxeles vecinos:

$$D_{1,2} = \sqrt{(c_r - c_1)^2 + (r_r - r_1)^2} \quad (3)$$

Donde p_1 es el píxel que se compara.

III. RECONOCIMIENTO DE IMAGENES

Establecida la base datos; esta no solo contiene la información del contorno de cada una de las fotos sino también de su posición, orientación y proporcionalidad por lo que para el reconocimiento únicamente se adquirirá la imagen en tiempo real y verificará si dicha imagen se encuentra en la base de datos; **aquí yace la importancia** de mantener la misma posición en el momento de obtener y reconocer la foto, Ver Fig.12.

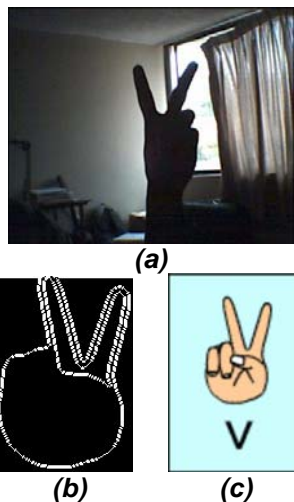


Figura 12. (a) Obtenemos la imagen para el reconocimiento.(b) Al estar en la mismas coordenadas el sistema no necesita volver a “entrenarse”.(c) La menor distancia indicará el elemento reconocido, en este caso coincide con la foto entrenada.

De no coincidir el objeto con los que se encuentran en la base de datos; indicará la forma más parecida.

IV. ANÁLISIS MATEMÁTICO

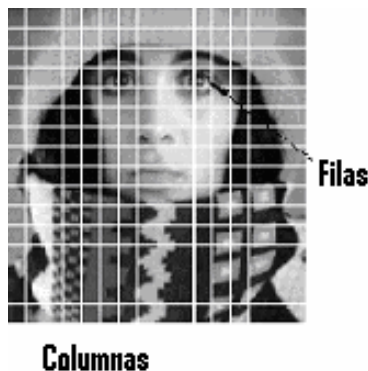


Figura 13. Representación digital de una imagen en un arreglo Matricial

$$Pixel_RGB = f(R,G,B)$$

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \dots & f(0,M-1) \\ f(1,0) & f(1,1) & \dots & f(1,M-1) \\ \vdots & \vdots & & \vdots \\ f(N-1,0) & f(N-1,1) & \dots & f(N-1,M-1) \end{bmatrix} \quad (4)$$

A. Gradiente.-

El gradiente es un vector, en donde sus componentes miden la rapidez en que los valores de los pixels cambian en la distancia y en las direcciones x e y. dx y dy son las distancias en las direcciones x e y respectivamente, en términos de número de pixels entre dos puntos.

$$\frac{\partial f(x,y)}{\partial x} = \Delta x = \frac{f(x+d_x, y) - f(x,y)}{dx} \quad (5)$$

$$\frac{\partial f(x,y)}{\partial y} = \Delta y = \frac{f(x, d_y + y) - f(x,y)}{dy} \quad (6)$$

$$M = \sqrt{\Delta_x^2 + \Delta_y^2} \quad (7)$$

$$\theta = \tan^{-1} \left(\frac{\Delta_y}{\Delta_x} \right) \quad (8)$$

Con el fin de detectar la presencia de una discontinuidad en el gradiente, debemos calcular el cambio en el gradiente en el punto (i,j). Esto se puede hacer referenciando la medida aportada por la magnitud del gradiente y su dirección.

$$\Delta_x = f(i+1, j) - f(i, j) \quad (9)$$

$$\Delta_y = f(i, j+1) - f(i, j) \quad (10)$$

Para la implementación y computación del gradiente se utilizan máscaras o filtros que representan o equivalen a dichas ecuaciones. En este caso, computar el gradiente sobre toda una imagen con las condiciones de que $dx=dy=1$ consiste en convolucionar la imagen con unas máscaras del tipo que se muestran en la Fig.14.

$$\Delta x = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} \quad \Delta y = \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix}$$

Figura 14. Máscaras del tipo $dx=dy=1$ usadas para convolucionar la imagen en el procesamiento.

$$\frac{\partial f(x, y)}{\partial x} = \Delta_x = \frac{f(x + d_x, y) - f(x, y)}{d_x} \quad (11)$$

$$\frac{\partial f(x, y)}{\partial y} = \Delta_y = \frac{f(x, d_y + y) - f(x, y)}{d_y} \quad (12)$$

$$\Delta_x = f(i + 1, j) - f(i, j) \quad (13)$$

$$\Delta_y = f(i, j + 1) - f(i, j)$$

$$\nabla^2 f(x, y) = \frac{\partial^2 f(x, y)}{\partial x^2} + \frac{\partial^2 f(x, y)}{\partial y^2} \quad (14)$$

Como las operaciones son lineales podemos combinar ambas operaciones de diferentes formas:

- Suavizado de la imagen y cálculo de la 2 derivada.
- Convolución de la imagen original utilizando el laplaciano del Gaussiano (**Operador LoG**).

$$\nabla^2(G \otimes I) = \nabla^2 G \otimes I \quad (15)$$

B. Transformada de Fourier en Imágenes:

$$F(\omega) = \int f(x) e^{-2\pi i \omega x} dx \quad (16)$$

$$F(\omega) = \int_0^X A e^{-2\pi i \omega x} dx$$

$$F(\omega) = \frac{-A}{2\pi i \omega} e^{-2\pi i \omega x} \Big|_0^X$$

$$F(\omega) = \frac{-A}{2\pi i \omega} \left[e^{-2\pi i \omega X} - 1 \right]$$

$$F(\omega) = \frac{-A}{2\pi i \omega} \left[e^{\pi i \omega X} - e^{-2\pi i \omega X} \right]$$

$$F(\omega) = \frac{-A}{\pi \omega} \text{sen}(\pi \omega X) e^{-\pi} \quad (17)$$

$$F\{F(u, v)\} = f(x, y) =$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u) \exp[j2\pi i(ux + vy)] dudv \quad (18)$$

$$F\{f(x, y)\} = F(u, v)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \exp[-j2\pi i(ux + vy)] dx dy \quad (19)$$

$$F\{f(x, y)\} = F(u, v) =$$

$$= \frac{1}{MN} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x, y) \exp \left[-j2\pi i \left(\frac{ux}{N} + \frac{vy}{M} \right) \right] \quad (20)$$

$$F\{f(u, v)\} = F(x, y) =$$

$$= \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} F(u, v) \exp \left[-j2\pi i \left(\frac{ux}{N} + \frac{vy}{M} \right) \right] \quad (21)$$

V. RESULTADOS DE LAS PRUEBAS

Tabla I. Comportamiento frente a distintas consideraciones de iluminación. La luz blanca deteriora la percepción de color, sin embargo la luz amarilla cuando se refleja en el lente de la Webcam hace muy deficiente la captura de la imagen. La luz natural mantuvo una baja tasa de errores pero no se puede mantener las condiciones de iluminación perennemente.

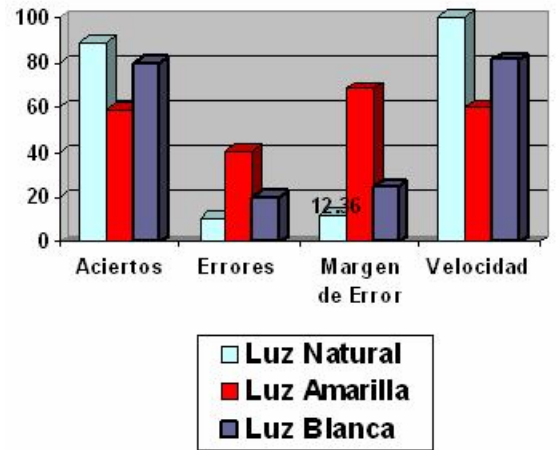
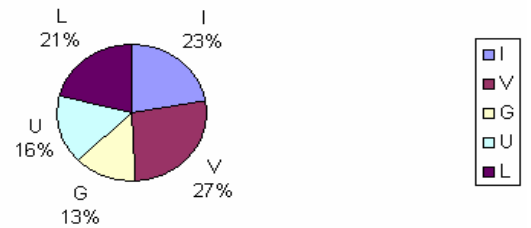


Tabla II. Una imagen que presenta una forma más irregular que otra requiere más tiempo para su reconocimiento. Aquí se demuestra con 5 letras del Lenguaje de Señas.

Reconocimiento



VI. CONCLUSIONES

Este proyecto se desarrolló dentro de una concepción de Prototipo Evolutivo por Desarrollo Incremental y se muestra bajo un perfil tecnológico multidisciplinario y cuya portabilidad permitió en las pruebas de campo aplicarlo a nivel de prototipo en sistemas como control biométrico, Información multimedia como Bibliotecas Virtuales para estimulación temprana en niños o buscadores en Internet basados en imágenes. Entre sus posibles aplicaciones, recientes Análisis de Morfología Biológica determinan que se puede descubrir alteraciones o mutaciones Celulares con la observación y evaluación de su forma y tamaño, parámetros para los cuales podríamos tender un puente de cooperación.

Dadas las limitaciones que presenta una webcam para la realización de las pruebas sobre el prototipo interactuando directamente con la mano, **también** se construyó un modelo en el cual las imágenes que son procesadas se adquieren desde una posición estática y en condiciones de luminosidad preestablecidas mediante fotografías dispuestas; se pudo evidenciar que el margen de error disminuyó notablemente basado en la carencia de movimiento involuntario del elemento fotografiado, sin embargo el principio es el mismo y ambos casos los resultados cumplen las expectativas planteadas. La revisión de la literatura muestra que la anterior descripción es la que han adoptado la mayoría de universidades para los prototipos documentados sobre tratamiento de imágenes.

Para minimizar el error así como la pérdida de calidad en el caso de las imágenes se puede usar cámaras especiales con tarjetas de procesamiento de video integrada que permiten fijar coordenadas en el marco de la foto. El porcentaje de error depende del ruido que presente el contorno a reconocer. Entiéndase a ruido al nivel de interferencia que presenta una imagen por factores como sombras, proyecciones de luz incoherentes, etc. A mayor ruido, será mayor la distancia de la Transformada entre los objetos aún si la forma y la posición son iguales.

El Reconocimiento de patrones de Voz con la transformada sigue el mismo formato de aplicación, se basa en la comparación de un registro de voz referencial con otro.

VII. BIBLIOGRAFÍA

- [1] BILMES, Jeff, BARTELS, Chris, Graphical Model Architectures for Speech Recognition. IEEE Signal Processing Magazine, September 2005.
- [2] WANG Ye, DENG Li, ACERO Alex. Spoken Language Understanding. IEEE Signal Processing Magazine, September 2005.
- [3] GARCÍA Ramos, Román Compresión de imágenes fijas en Matlab a través de DCT y Wavelet, Ciencias con Especialidad en Ingeniería Electrónica. UPLA –P.
- [4] MENDIETA Domínguez, Denise, Reconocimiento de Objetos Bidimensionales en Imágenes mediante la Transformada de Distancia utilizando Matlab. Ingeniería en Electrónica y Comunicaciones, UPLA –P.
- [5] CARCEDO Y Franco, Ana; Programa de Segmentación de Regiones en imágenes médicas en Matlab. Ingeniería en Electrónica y Comunicaciones, UPLA –P.
- [6] FLORES Pablo, RODRÍGUEZ Marcelo, Analizador de Imágenes Médicas, Universidad de la República, Uruguay.
- [7] SUMITHA, Lalendra Balasuriya, Frontal View Human Face Detection and Recognition, Department of Statistics and Computer Science University of Colombo, Sri Lanka, May 2000
- [8] MPEG Requirements Group, "mpeg-7 requirements document", doc. iso/mpeg n2727, mpeg Seoul meeting, march 1999
- [9] Pontificia Universidad Católica de Perú, Laboratorio IEEE 253 de Comunicaciones. Artículo sobre "Video Digital".
- [10] SYMEONIDIS, Klimis. Hand Gesture Recognition Using Neural Networks, School of Electronic and Electrical Engineering, UNIS.
- [11] PEREZ Emiliano, GALLEGO David, ShADEs 3D: Un Toolbox de Matlab para la Adquisición de Superficies Tridimensionales. Universidad de Extremadura
- [12] PART-ENANDER, Sjoberg A. The MATLAB Handbook, E., Addison-Wesley, New York, 1996.

Páginas Web Visitadas.

- <http://www.electronicimaging.org/program/06/conferences/index>. Aplicaciones de Medicina con Imágenes.
- <http://www.sys.uea.ac.uk/~fuzz/vfm/default.html> Página del Toolbox Vision for Matlab.
- <http://mpeg-industry.com>. Página acerca de las Aplicaciones.

VIII. BIOGRAFÍA



EDGAR ORDÓÑEZ ORTIZ
(1982), Egresado de
Tecnología en Informática
Aplicado y Estudiante del
9no Semestre de Ingeniería
Electrónica en Computación
de la Escuela Superior
Politécnica del Chimborazo

en Riobamba Ecuador; Miembro IEEE de la
Sociedad de Comunicaciones y Broadcasting
de la Rama IEEE ESPOCH, Desarrollador de
Aplicaciones para la Facultad de Informática y
Electrónica, miembro del Equipo Bicampeón
Nacional de Robótica por la ESPOCH,
participante con mención Honorífica de la VIII
Feria de Ciencia y Tecnología y III Concurso
Nacional de Innovación Fenacyt organizado por
FUNDACYT, participante del XXII Congreso
Estudiantil de Ingeniería Eléctrica y Electrónica
INGELECTRA 2006 en Valdivia Chile,
postulante para Andescon 2006.