

# Diseño de un Sistema de Reconocimiento de Gestos No Móviles mediante el Procesamiento Digital de Imágenes

V. Valencia Sierra y B. Villa Brochero.

**Abstract**— in this document we will present the design process and implementation of static hand gesture recognition system using Matlab. The System can show the image of the acquired letter and the translation of this sign in the Colombian Sign Language. This System was developed to recognize 20 static letters from the Colombian Sign Languages. A Graphical User Interface was implemented in order to show better the information and to have a simple use for the user.

**Resumen**— En este artículo se muestra proceso de diseño e implementación de un sistema de reconocimiento de gestos no móviles mediante el entorno de Matlab; por medio del cual se visualiza la imagen de la letra adquirida, junto con la traducción en el lenguaje de señas colombiano. Este sistema se desarrolló para reconocer las 20 letras no móviles del lenguaje de señas colombiano mediante técnicas de procesamiento de imágenes digitales. Se implementó una interfaz gráfica en Matlab para una mejor visualización del sistema y para su fácil uso por parte del usuario.

**Index Terms**— Colombian Sign Language, Image Processing, Image Segmentation, Gray Scale, Graphical Interface, Gesture Recognition, Static Recognition, Key Points, Matlab, SIFT

**Palabras Claves**— Escala de Grises, Interfaz Gráfica, Lenguaje de Señas Colombiano, Matlab, procesamiento de imágenes, puntos característicos, reconocimiento de gestos, reconocimiento estadístico, segmentación de imágenes, SIFT

## I. INTRODUCCIÓN

EN este artículo se presenta el desarrollo de un sistema de reconocimiento de gestos no móviles del lenguaje de señas Colombiano, mediante el procesamiento digital de imágenes. El Sistema se divide principalmente en cuatro fases: adquisición

de imagen, procesamiento de la imagen, reconocimiento de la imagen y muestra de resultado.

Según estadísticas arrojadas por el Departamento Administrativo Nacional de Estadística (DANE), en Colombia el 6,4% de los habitantes tiene una limitación permanente; de éste grupo el 17,3% de las personas, tienen limitaciones permanentes para oír, es decir, que de cada 100 colombianos con limitación, 17 presentan algún tipo de discapacidad permanente auditiva. [1]

La comunicación mediante el lenguaje de señas es casi imprescindible en el día a día de un ser humano que sufre de limitación auditiva, pues a partir de ella puede entrar a relacionarse con las personas que le rodean y a la vez con su entorno, por lo que se hace un reto aprender cada una de las señas que conforman su sistema de comunicación, y más aún para las personas sin discapacidad al momento de lograr descifrar el mensaje proporcionado por esta población. Es por esto que surge la necesidad de desarrollar una herramienta que permita reconocer los gestos que son realizados por personas con discapacidad auditiva, con el fin de brindar una herramienta didáctica que permita a las personas que están aprendiendo éste tipo de lenguaje (discapacitados o no discapacitados); lo puedan hacer de una manera más interactiva y amigable con su entorno.

*El lenguaje de Señas se caracteriza por ser visual y corporal, es decir la comunicación se establece con el cuerpo en un espacio determinado.* [2]

El lenguaje de Señas Colombiano (LSC) fue reconocido en el año 1996 y es el lenguaje oficial utilizado por la comunidad sorda del país. El LSC consta de 27 señas que representan a cada una de las letras del alfabeto, mediante la unión de varias señas se puede lograr expresar palabras, frases e ideas completas. En la implementación de nuestros sistemas solamente utilizaremos 20 señas no móviles del LSC debido a que esta investigación se basa principalmente en el análisis y tratamiento de imágenes. La figura 1 contiene todos los caracteres que representan el alfabeto en el LSC.

Por otro lado, el procesamiento de imágenes consiste en todos los métodos y técnicas implementados para extraer información de la imagen. La segmentación es una de esas técnicas la cual básicamente busca separar el fondo del objeto

V. Valencia Sierra, estudiante de décimo semestre de Ingeniería Mecatrónica en la Universidad Autónoma del Caribe (Barranquilla, Colombia). (E-mail: valeria.valencia@uautonoma.edu.co).

B. Villa Brochero, estudiante de décimo semestre de Ingeniería Mecatrónica en la Universidad Autónoma del Caribe (Barranquilla, Colombia). (E-mail: betsy.villa@uautonoma.edu.co).

de interés. Las imágenes en su mayoría están compuestas por zonas que tienen características similares (nivel de gris, textura, momentos, etc.). Generalmente estas zonas corresponden a objetos de la imagen. *La segmentación de una imagen consiste en la división o partición de la imagen en varias zonas o regiones homogéneas y disjuntas a partir de su contorno, su conectividad, o en términos de un conjunto de características de los píxeles de la imagen que permitan discriminar unas regiones de otras.* Entre las características utilizadas en la segmentación encontramos: los tonos de gris, la textura, los momentos, la magnitud del gradiente, la dirección de los bordes, entre otras. [3]

Actualmente existen varios enfoques de reconocimiento de objetos, de los cuales los más conocidos son:

- *Reconocimiento estadístico:* Se basa en la determinación y uso de funciones de probabilidad.

- *Reconocimiento sintáctico:* Analiza la estructura de los objetos, ej: el esqueleto.

- *Reconocimiento de redes neuronales:* Constituye la técnica más reciente y trata de imitar el funcionamiento de los sistemas biológicos. [4]

En el desarrollo del sistema de reconocimiento de gestos no móviles se utilizó el reconocimiento estadístico mediante el cual, para la detección de gesto de la mano, fue necesario hallar los puntos clave característicos de cada imagen con el fin de poder compararlos con la imagen de entrada y finalmente si existe una coincidencia mostrar la imagen de salida obtenida. Este método se conoce como SIFT (Scale Invariant Features Transform) introducido por David Lowe en 1999, es considerado una de las mejores técnicas de procesamiento de imágenes, ya que la información obtenida de la imagen se transforma en coordenadas invariantes de escala, rotación y luminosidad. Lo primero que se realiza es obtener un conjunto de puntos de la imagen, los cuales serán denominados keypoints, de acuerdo al paso por cada etapa, el número de keypoints se irá reduciendo y quedarán los más importantes para ser usados en la comparación. Su idea principal es la transformación de la imagen a una representación compuesta de “puntos de interés”. Esos puntos contienen la información característica de la imagen que luego son usados para la detección de muestras.

El algoritmo se realiza mediante 4 etapas:

1. Construcción de Pirámides de Scale-Space: En esta etapa se representa la imagen en diferentes escalas y tamaños. Mediante la función de diferencia gaussiana se identifican los posibles puntos de interés que son invariantes a escala y orientación.
2. Localización de puntos clave: En esta etapa se localizan los puntos que se mantienen invariantes en escala. Esto se realiza mediante la comparación de cada pixel con su pixel vecino. Los puntos clave son seleccionados en base a las medidas de su estabilidad.
3. Asignación de orientación: En esta en esta etapa paso se asigna a cada punto clave una dirección de acuerdo a las direcciones del gradiente y a la zona que rodea dicho punto.

4. Descriptor de Puntos clave: Esta etapa consiste en calcular un descriptor para la región de la imagen local que sea fácilmente identificable, sin embargo tan invariable como sea posible a las variaciones restantes, tales como el cambio en la iluminación o el punto de vista 3D.

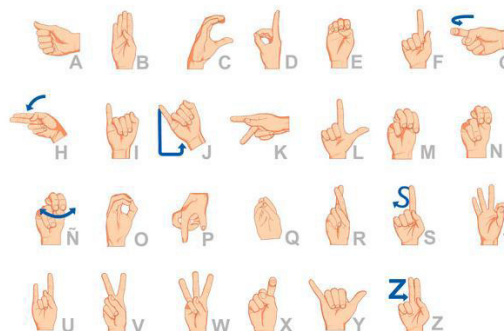


Fig.1 Caracteres del lenguaje de señas colombiano. Tomado de: Fenascal, Lengua de Señas Colombiana, Segundo Tomo, Santafé de Bogotá, 1996, pág.VII

## II. ANTECEDENTES

Realizando una revisión de trabajos relacionados con sistemas de reconocimiento de señas, se encuentran trabajos interesantes como el desarrollado por Herigert, Bouchet y Pianetti, quienes por medio de Java y con el uso de diferentes imágenes, realizaron un software, utilizando una base de datos en SQL, para llevar a cabo el reconocimiento de los objetos, sin importar variantes que éstos puedan tener en cuanto a escala y rotación, o iluminación, basándose en el algoritmo SIFT, y desarrollando una interfaz en donde se visualizan las acciones que realiza el método para la búsqueda de las imágenes coincidentes. De lo cual se puede concluir que el algoritmo arroja buenos resultados, para la búsqueda de puntos comunes entre dos imágenes, sin importar las características que pueden variar en ellas, mencionadas anteriormente. [5]

Chiguano, junto con Moreno y Corrales, desarrollan un sistema que traduce el lenguaje de señas e implementan un entrenador del mismo. El sistema se desarrolló en LabView, concluyendo que el desempeño del sistema depende en gran medida de las condiciones del ambiente bajo el cual se realice la adquisición de la imagen, puesto que éste último al ser controlado, arrojaba una imagen más limpia facilitando el tratamiento de la misma pero reduciendo la flexibilidad de la aplicación para su uso en cualquier condición. [6]

Por último, Ordóñez, presenta un método para el tratamiento de imágenes y sonidos en Matlab por medio de la distancia de transformada de una función que resulta ser una herramienta matemática bastante útil para el reconocimiento de objetos, siempre y cuando sea aplicada de forma correcta. La aplicación desarrollada contaba con etapas de adquisición, extracción de características de la imagen, definición de dichas características y por último el reconocimiento. Por medio de la transformada se obtenían las matrices características de las imágenes que más adelante entrarían al sistema de traducción. [7]

Como se puede observar en los trabajos analizados, es importante iniciar con un tratamiento de imágenes, para luego continuar con el proceso de reconocimiento y utilización del sistema, con el fin de garantizar resultados y minimizar probabilidades de error en la aplicación diseñada.

### III. METODOLOGIAS

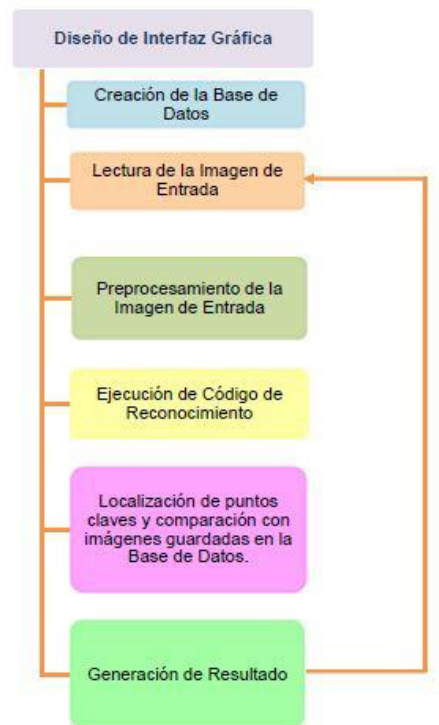


Fig.2 Diseño Metodológico del Sistema de Reconocimiento de gestos no móviles.

En la figura 2. Se muestra todas las etapas utilizadas para el diseño y la ejecución del sistema de reconocimiento de gestos no móviles. Como entorno de programación y diseño de la interfaz se utilizó Matlab, primero se realizó la captura de imágenes y la creación de la base de datos del sistema. La captura de imágenes de prueba se da a partir de la cámara digital Canon PowerShot A4000 IS Azul la cual nos brinda una alta calidad de las fotografías capturadas teniendo en cuenta sus 16 Megapíxeles, y Zoom Óptico 8x.

La imagen de entrada del sistema es sometida a un pre procesamiento para disminuir el ruido de esta imagen y también se hace un cambio en las dimensiones de la imagen para disminuir el costo computacional.

#### A. Interfaz Grafica

La interfaz se desarrolla por medio de la GUIDE contenida en el entorno de Matlab, la cual consta de una ventana, que se distribuye de la siguiente forma:

1. Encabezado: Nombre del Proyecto, Estudiantes, y Universidad

2. Captura de Imagen: Caja para la visualización de la imagen a tomar por parte de la cámara, y la cual posee los siguientes comandos:

2.1 Encender cámara: Comando utilizado para que la cámara aparezca en el cuadro principal donde se encuentra el logo más grande la Universidad Autónoma del Caribe.

2.2 Apagar cámara: Comando utilizado para dejar de utilizarla cámara.

2.3 Capturar: Comando utilizado para realizar captura de imágenes.

2.4 Abrir imagen: Comando utilizado para abrir las imágenes capturadas.

3. Visualización de caracteres del Lenguaje de Señas Colombiano e imagen capturada de entrada al algoritmo principal.

4. Visualización de Imagen de Salida.

5. Información del Proyecto.

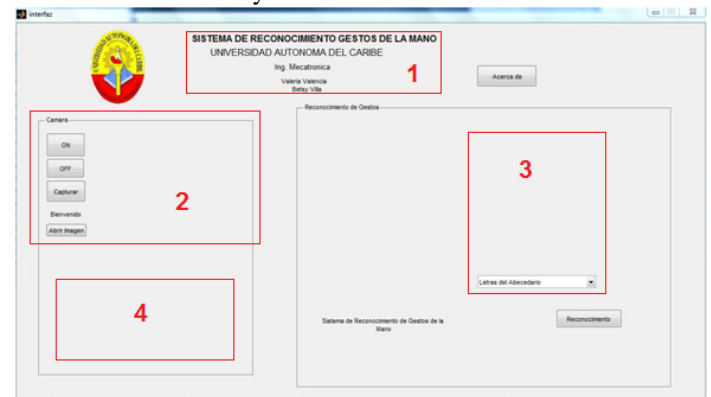


Fig.3 Distribución de la Interfaz Gráfica desarrollada en Matlab

#### B. Función de Reconocimiento

La función de reconocimiento recibe la imagen de entrada la convierte en escala de grises y mediante la técnica de SIFT halla los puntos de interés de la imagen de entrada para que estos puedan ser comparado con los puntos de interés de las imágenes que se encuentran en la base de datos del sistema, al final la función si encuentra dos imágenes con puntos de interés iguales o similares significa que ha encontrado la señal y letra correspondiente al lenguaje de señas colombiano. Si la función no encuentra ninguna relación entre los puntos de interés de la imagen de entrada y las que se encuentran en la base de datos simplemente el sistema arrojará que no ha sido posible identificar la señal y que lo intente nuevamente.

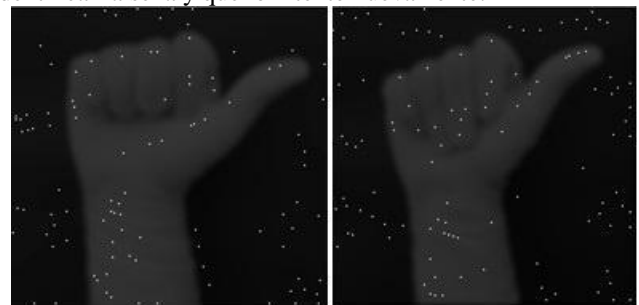


Fig.4 Imagen en Base de Datos vs Imagen de Entrada, puntos característicos obtenidos.

#### IV. IMPLEMENTACIÓN

Durante la realización de pruebas al sistema de reconocimiento de gestos no móviles del lenguaje de señas colombiano, se ejecuta inicialmente el algoritmo principal, en el cual se encuentra el código desarrollado para la Interfaz de Usuario y en el que a su vez se lee la imagen de entrada que puede estar guardada en el ordenador, o capturada por medio de la cámara del mismo con dimensiones 320x240.

Una vez elegida la imagen, ésta se despliega en la parte superior de la interfaz, como se observa en la Figura 5.



Fig.5 Visualización de imagen de entrada

Posteriormente se ejecuta el proceso de reconocimiento, pulsando un botón llamado de la misma forma, con el fin de encontrar los puntos claves o característicos de la imagen de entrada y las imágenes que se encuentran en la base de datos, y realizando una comparación donde se determina cuál es la imagen encontrada en la base de datos que se asocia a la imagen ingresada al sistema. Después de ejecutar este algoritmo, se reconoce la imagen de entrada, en el LSC, mediante la comparación con todas las imágenes de la bases de datos y el resultado es la representación ASCII equivalente de la misma.

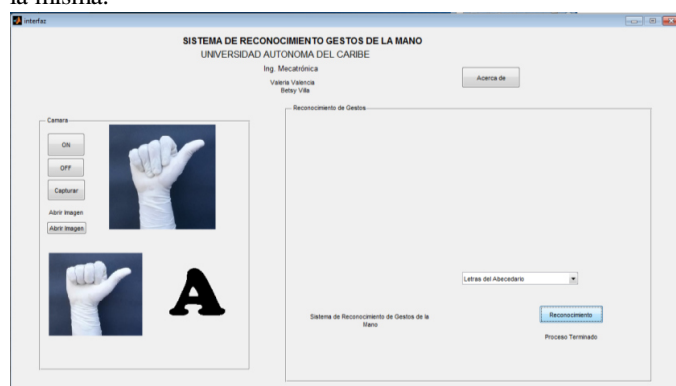


Fig.6 Resultado Obtenido luego de realizar el proceso de reconocimiento

#### V. RESULTADOS

Para esta implementación, se realizan dos pruebas, donde se utilizan dos bases de datos las cuales contienen los patrones de los gestos no móviles utilizados en el Lenguaje de Señas Colombiano. Las imágenes utilizadas son en formato de archivo ".jpg". Para la aplicación experimental del método propuesto tomamos los 20 signos no móviles del LSC como consulta para la imagen de entrada. Las imágenes contenidas

en la base de datos son: A, B, C, D, E, F, I, K, L, M, N, O, P, Q, R, T, U, V, W, Y.

N°	Imagen	Letra	Porcentaje de Reconocimiento	Porcentaje de Error
1		A	40%	60%
2		B	100%	0%
3		C	70%	30%
4		D	50%	50%
5		E	50%	50%
6		F	50%	50%
7		I	40%	60%
8		K	60%	40%
9		L	20%	80%
10		M	70%	30%
11		N	70%	30%
12		O	50%	50%
13		P	10%	90%
14		Q	20%	80%
15		R	50%	50%
16		T	60%	40%
17		U	40%	60%

Fig.7 Resultado Prueba 1

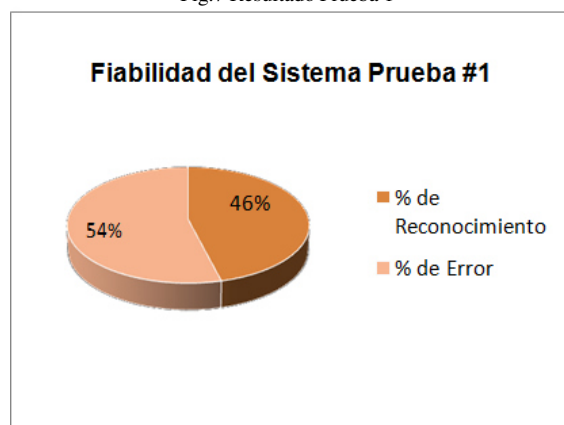


Fig.8 Grafica del porcentaje de Fiabilidad del Sistema en la primera prueba

Durante la segunda prueba se capturan imágenes utilizando un guante quirúrgico, con el fin de estandarizar el sistema, para minimizar los errores arrojados en la primera prueba, con 2 personas de tez blanca y una persona de tez morena, generando los siguientes resultados:















N°	Imagen	Letra	Porcentaje de Reconocimiento	Porcentaje de Error
1		A	90%	10%
2		B	60%	40%
3		C	100%	0%
4		D	90%	10%
5		E	100%	0%
6		F	30%	70%
7		I	70%	30%
8		K	100%	0%
9		L	60%	40%
10		M	100%	0%
11		N	30%	70%
12		O	60%	40%
13		P	100%	0%
14		Q	70%	30%

Fig.7 Resultado Prueba 1

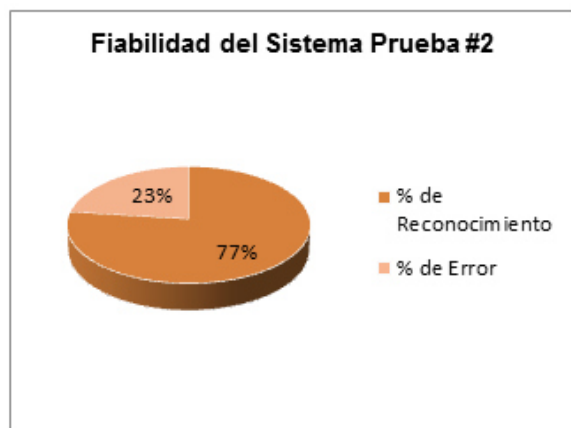


Fig.10 Grafica del porcentaje de Fiabilidad del Sistema en la segunda prueba

## VI. CONCLUSIÓN

Actualmente, la implementación de aplicaciones utilizando el procesamiento digital de imágenes es de gran auge, debido al sin número de análisis que se pueden realizar por medio de la captura de imágenes o videos, para solventar distintas necesidades.

El algoritmo de reconocimiento se encarga de extraer características intrínsecas de las imágenes de entrada y base de datos, que resulten ser claves al momento de realizar la comparación que permita determinar la señal correspondiente al Lenguaje de Señas Colombiano. Además, se encarga de recibir una imagen de una señal del Lenguaje de Señas Colombiano dada por el usuario, y hallar, con el mejor grado de certeza posible, en una base de datos.

Para el reconocimiento preciso y fiable, las características extraídas de la imagen de entrada deben ser detectables incluso en los cambios en la escala de la imagen, el ruido y la iluminación. Se encuentra una mejor respuesta del sistema, con la utilización de un elemento que estandarice la mano con respecto a los factores de color, utilizando un guante, en este caso, tipo quirúrgico.

Como trabajo a futuro se recomienda optimizar el código de reconocimiento, para detectar los caracteres del LSC, en tiempo real, que permitan construir una especie de traductor online, para discapacitados

## REFERENCIAS

- [1] Colombia, Departamento Administrativo Nacional De Estadística, Censo General 2005, *Discapacidad: personas con limitaciones permanentes*. Bogotá. DANE. 2006. [www.dane.gov.co/files/censo2005/discapacidad.pdf](http://www.dane.gov.co/files/censo2005/discapacidad.pdf)
- [2] *Lengua de Señas Colombiana*, Segundo Tomo, Santafé de Bogotá, 1996, pág. VII
- [3] A. Naranjo. *Aplicación de DSP's para la Transcripción de Lenguaje de Señas a Texto*. Universidad Técnica Ambato. Ambato, Ecuador. 2014 [http://repo.uta.edu.ec/bitstream/handle/123456789/7793/Tesis\\_t902ec.pdf?sequence=1](http://repo.uta.edu.ec/bitstream/handle/123456789/7793/Tesis_t902ec.pdf?sequence=1)
- [4] D. Paredes. *Seguimiento y Caracterización de Componentes del Rostro para la Detección de Expresiones Faciales*. Centro Nacional de Investigación y Desarrollo Tecnológico. Cuernavaca, México, 2009. <http://www.cenidet.edu.mx/subplan/biblio/seleccion/Tesis/MC%20Danie1%20A%20Paredes%20Soto%202009.pdf>
- [5] M. Herigert, N.Bouchet, I.Pianetti. *Reconocimiento de Imágenes mediante Scale Invariant Feature Transformation (SIFT)*. Universidad Tecnológica Nacional- Facultad Regional Concepción del Uruguay. Uruguay, 2009.
- [6] E. Chiguano, *Diseño e implementación de un sistema traductor de lenguaje de señas de manos a un lenguaje de texto mediante visión artificial en un ambiente controlado*. Escuela Politécnica Nacional. Ecuador.
- [7] E. Ordoñez. *Sistema de reconocimiento de imágenes como intérprete de lenguaje de señas*. Escuela Superior Politécnica de Chimborazo. Ecuador

## BIBLIOGRAFÍA



**Valeria Valencia Sierra** (Barranquilla, Julio 10 del 1992) Egresada destacada de la escuela Normal Superior del Distrito de Barranquilla y actualmente es estudiante de decimo semestre del programa de Ingeniería Mecatrónica en la Universidad Autónoma del Caribe.



**Betsy Villa Brochero** (Barranquilla, Octubre 20 del 1992) Egresada destacada de la Escuela Normal Superior del Distrito de Barranquilla, realizo estudio de Ingles como segunda lengua en la Universidad del Atlántico (Barranquilla, 2009) y Alemán (Barranquilla 2013), actualmente es estudiante de decimo semestre del programa de Ingeniería Mecatrónica en la Universidad Autónoma del Caribe