



INSTITUTO POLITÉCNICO NACIONAL

CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN

**RECONOCIMIENTO DE
IMÁGENES DEL
LENGUAJE
DE SEÑAS MEXICANO**

T E S I S

QUE PARA OBTENER EL GRADO DE
MAESTRO EN CIENCIAS DE LA COMPUTACIÓN

P R E S E N T A

ING. FAUSTO PÁVEL PRIEGO PÉREZ



CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN

DIRECTORES DE TESIS

DR. JESÚS MANUEL OLIVARES CEJA
DR. JOSÉ FÉLIX SERRANO TALAMANTES

MÉXICO, D. F.

DICIEMBRE 2012



INSTITUTO POLITÉCNICO NACIONAL SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México, D.F. siendo las 16:00 horas del día 10 del mes de diciembre de 2012 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación del:

Centro de Investigación en Computación

para examinar la tesis titulada:

“Reconocimiento de Imágenes del lenguaje de señas mexicano”

Presentada por el alumno:

PRIEGO

Apellido paterno

PÉREZ

Apellido materno

FAUSTO PÁVEL

Nombre(s)

Con registro:

A	1	1	0	8	7	5
---	---	---	---	---	---	---

aspirante de: **MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN**

Después de intercambiar opiniones los miembros de la Comisión manifestaron **APROBAR LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA

Directores de Tesis

Dr. Jesús Manuel Olivares Ceja

M. en C. José Félix Serrano Tejamantes

Dr. Edgardo Manuel Felipe Riverón

Dr. Adolfo Guzmán Arenas

Dr. Juan Humberto Sossa Azuela

Dr. Ricardo Barrón Fernández

PRESIDENTE DEL COLEGIO DE PROFESORES

Dr. Luis Alfonso Villa Vargas



INSTITUTO POLITÉCNICO NACIONAL
CENTRO DE INVESTIGACIÓN
EN COMPUTACIÓN
DIRECCIÓN



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de México, Distrito Federal el día 12 del mes de Diciembre del año 2012, el que suscribe Fausto Pável Priego Pérez alumno del Programa de Maestría en Ciencias de la Computación con número de registro A110875, adscrito al laboratorio de Base de Datos y Tecnología de Software, manifiesta que es autor intelectual del presente trabajo de Tesis bajo la dirección del Dr. Jesús Manuel Olivares Ceja y del Dr. José Félix Serrano Talamantes y cede los derechos del trabajo intitulado “Reconocimiento de imágenes del Lenguaje de Señas Mexicano”, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expreso del autor y/o director del trabajo. Este puede ser obtenido escribiendo a la siguiente dirección pavelpriego@gmail.com. Si el permiso se otorga, el usuario deberá dar el agradecimiento correspondiente y citar la fuente del mismo.

A handwritten signature in blue ink, appearing to read 'Fausto Pável Priego Pérez', is positioned above a horizontal line.

Fausto Pável Priego Pérez

Se ha encontrado que las personas con problemas auditivos severos o hipoacusia están marginadas de muchas actividades sociales, incluyendo las académicas y laborales. Es por esto, que en esta tesis se propone un sistema de información que reconoce imágenes de manos que representan símbolos estáticos del alfabeto del Lenguaje de Señas Mexicano (LSM) que corresponde al estándar internacional ISO 639-2 sgn-MX.

El sistema contiene dos módulos principales: aprendizaje y reconocimiento. El reconocimiento es aplicado constantemente a imágenes que se obtienen del dispositivo Kinect; cada una se compara con los patrones almacenados en la base de conocimientos del sistema. El módulo de aprendizaje entra en funcionamiento cuando se detecta un nuevo patrón, es decir, cuando la similitud entre la imagen obtenida y el patrón almacenado es menor al 90 por ciento.

En este trabajo se usa el dispositivo Kinect con una frecuencia de muestreo de una imagen por segundo, que es el tiempo requerido para procesar y reconocer cada entrada que representa una seña del alfabeto del LSM.

Los resultados obtenidos permiten proponer este sistema como una alternativa para la comunicación entre personas con problemas severos auditivos y personas normales.

It has been found that people with severe hearing loss problems are also mute and that in many cases are marginalized in many social activities, including academic and jobs.

In this proposal, it is offered an alternative to solve the communication problem of people with severe hearing loss. In this thesis is developed an information system to recognize images of human hands representing signs that belong to the Mexican Sign Language (LSM) which correspond to the ISO 639-2 sgn-MX international standard.

The system contains two main modules: learning and recognition. Recognition is applied constantly to images obtained from a Kinect device by comparing them with patterns in the knowledge base. The learning module is activated when a new pattern is detected (when the similarity between an image and a pattern is less than 90 %).

Kinect has been used with a sampling frequency of one image per second, based on the time required to process and recognize each image that represent an LSM sign. The results obtained, let to propose this system as an alternative for communication among people with severe hearing loss and ordinary people.

Agradecimientos

A Dios. A mi madre por ser un ejemplo de vida, a mi familia, al Dr. Manuel Leonardo Mera Hernández, a Elizabeth López Romero y a mis amigos.

Al Doc Fernando Galindo Soria por su conocimiento que me aportó, su apoyo y consejo.

A mis directores de tesis, Dr. Jesús Manuel Olivares Ceja y Dr. José Félix Serrano Talamantes.

A mis sinodales por su paciencia, enseñanza y comentarios para mejorar este trabajo.

A CONACYT por el apoyo económico prestado para concluir mis estudios y al personal del Centro de Investigación en Computación.

A mis compañeros de generación por las experiencias obtenidas y a todos los profesores que impartieron las asignaturas que asistí por la oportunidad, paciencia, sabiduría, instrucción y consejo.

A la Ing. Diana Naim Rivera Aguilar, Yelitze Ramírez Martínez, M. en C. Roberto Velázquez Cabrera, M. en C. José Luis Arévalo, Ing. Migdaled López Juárez, M. en C. German Téllez Castillo, Dr. Jesús Figueroa Nazuno, Dr. Marco Antonio Moreno Armendáriz, M. en C. Antonio Ramírez Ramírez, Dr. Salvador Godoy Calderón y M. en C. Xóchitl Raquel Wong Cohén por su apoyo y consejo.

Al Departamento de Tecnologías Educativas del CIC, por la amabilidad y eficiencia en el desarrollo de su trabajo.

Esta tesis fue apoyada por el proyecto SIP20121031 "Organización de documentos textuales basada en información oculta en el corpus".

RESUMEN	I
ABSTRACT	III
AGRADECIMIENTOS	V
ÍNDICE DE FIGURAS	XI
ACRÓNIMOS, ABREVIATURAS Y SIGLAS	XIII
GLOSARIO DE TÉRMINOS	XV
INTRODUCCIÓN	1
1. PROBLEMÁTICA	3
1.1 OBJETIVOS	6
1.1.1 OBJETIVO GENERAL	6
1.1.2 OBJETIVOS ESPECÍFICOS	7
1.2 JUSTIFICACIÓN	7
1.2.1 BENEFICIOS ESPERADOS	8
1.2.2 ALCANCES	9
2. ESTADO DEL ARTE	11
2.1 LENGUAJE DE SEÑAS MEXICANO	11
2.2 UNA INTERFAZ PARA EL RECONOCIMIENTO DE SEÑAS GESTUALES PARA PERSONAS HIPOACÚSICAS	13
2.3 SISTEMA PARA EL RECONOCIMIENTO DEL ALFABETO DACTILOLÓGICO	13
2.4 RECONOCIMIENTO DE GESTOS MANUALES UTILIZANDO KINECT (YI LI)	14
2.5 RECONOCIMIENTO DE GESTOS MANUALES UTILIZANDO KINECT (HENG DU)	14
2.6 APLICACIÓN DE LOS SISTEMAS EVOLUTIVOS EN EL ANÁLISIS DE ESPECTROS DE RAYOS GAMMA	15
2.7 SISTEMA EVOLUTIVO PARA EL DIAGNÓSTICO DE FALLAS EN MÁQUINAS ROTATORIAS	16
3. MARCO TEÓRICO	17
3.1 MODELO DE COLOR RGB	17
3.2 MORFOLOGÍA MATEMÁTICA	18
3.2.1 ELEMENTO DE ESTRUCTURA	18
3.2.2 DILATACIÓN BINARIA \oplus	18
3.2.2.1 Propiedades de la dilatación	19
3.2.3 EROSIÓN BINARIA \ominus	19

3.2.3.1	Propiedades de la erosión	20
3.2.4	APERTURA BINARIA °	20
3.2.4.1	Propiedades de la apertura	21
3.2.5	CLAUSURA BINARIA •	21
3.2.5.1	Propiedades de la clausura	21
3.2.6	SOMBRERO DE COPA BLANCO WTH	22
3.2.6.1	Propiedades del Sombrero de Copa Blanco	22
3.3	HISTOGRAMA	22
3.4	MÉTODOS DE SEGMENTACIÓN BASADOS EN EL UMBRALADO	23
3.4.1	MÉTODO DE UMBRALADO DE KITTLER E ILLINWORTH	24
3.5	ALGORITMO DE ETIQUETADO DE COMPONENTES CONEXAS ITERATIVO	26
3.6	MOMENTOS DE HU	27
3.6.1	CENTRO DE GRAVEDAD DEL OBJETO	27
3.6.2	EJES PRINCIPALES DE LA ELIPSE	28
3.6.3	MOMENTOS INVARIANTES DE HU	28
3.6.4	SESGO PROYECTIVO	30
3.6.5	EXCENTRICIDAD	30
3.6.5.1	Valores de excentricidad en secciones cónicas	31
3.7	REDONDEZ	31
3.8	CLÁUSULAS DE HORN	31
3.9	PERCEPCIÓN DE PROFUNDIDAD	32
3.9.1	MAPA PROFUNDIDAD KINECT	33
3.10	MATRIZ EVOLUTIVA	34
4.	DESARROLLO DE LOS MÉTODOS PROPUESTOS	37
<hr/>		
4.1	CARACTERÍSTICAS DEL PRIMER MÉTODO DE SOLUCIÓN PROPUESTO	37
4.2	ARQUITECTURA DEL PRIMER MÉTODO DE SOLUCIÓN PROPUESTO	37
4.2.1	CAPTURA DE LA IMAGEN	37
4.2.2	CORRECCIÓN DE LA ILUMINACIÓN	39
4.2.3	SEGMENTACIÓN	40
4.2.4	FILTRO MORFOLÓGICO	40
4.2.5	CLASIFICACIÓN Y RECONOCIMIENTO	42
4.2.6	PRESENTACIÓN	42
4.3	CARACTERÍSTICAS DEL SEGUNDO MÉTODO DE SOLUCIÓN PROPUESTO	42
4.4	ARQUITECTURA DEL SEGUNDO MÉTODO DE SOLUCIÓN PROPUESTO	43
4.4.1	CAPTURA DE LA IMAGEN	43
4.4.2	SEGMENTACIÓN DE LA MANO	45
4.4.3	ESCALAMIENTO DE LA MANO	47
4.4.4	MATRICES EVOLUTIVAS	48
4.4.4.1	Proceso de aprendizaje-reconocimiento sin reforzamiento	49
4.4.4.1.1	Proceso de reconocimiento de la matriz evolutiva sin reforzamiento.	50
4.4.4.1.2	Aprendizaje de la matriz evolutiva sin reforzamiento	51
4.4.4.1.3	Ejemplo matriz evolutiva sin reforzamiento	51
4.4.4.2	Proceso de aprendizaje-reconocimiento con reforzamiento	53
4.4.4.2.1	Proceso de reconocimiento de la matriz evolutiva con reforzamiento.	54
4.4.4.2.2	Aprendizaje de la matriz evolutiva con reforzamiento	55

4.4.4.3	Reforzamiento	55
4.4.4.4	Olvido	57
4.4.4.5	Ejemplo del proceso de reconocimiento de una letra del LSM usando matrices evolutivas.	57
4.4.4.5.1	Vector de entrada	57
4.4.4.5.2	Reconocimiento-aprendizaje del patrón	60
5.	RESULTADOS OBTENIDOS	63
5.1	PRIMER MÉTODO DE SOLUCIÓN PROPUESTO	63
5.1.1	CONSIDERACIONES	63
5.1.2	OPERACIÓN DEL MÉTODO	64
5.1.3	RESULTADOS	65
5.1.4	VENTAJAS Y DESVENTAJAS DEL PRIMER MÉTODO	66
5.2	SEGUNDO MÉTODO DE SOLUCIÓN PROPUESTO	66
5.2.1	CONSIDERACIONES	67
5.2.2	OPERACIÓN DEL MÉTODO	67
5.2.3	RESULTADOS	68
5.2.4	VENTAJAS Y DESVENTAJAS DEL SEGUNDO MÉTODO	71
5.3	CONTRIBUCIONES	72
5.4	CONCLUSIONES	73
5.5	TRABAJO A FUTURO	74
REFERENCIAS		75

<i>Figura 1.1 Etapas fundamentales del Sistema de Reconocimiento Automatizado de Objetos</i>	4
<i>Figura 1.2 Problemas a considerar en la extracción de rasgos.</i>	6
<i>Figura 1.3 Porcentaje de la población por tipo de discapacidad</i>	8
<i>Figura 1.4 Escolaridad de personas con hipoacusia</i>	9
<i>Figura 2.1 Alfabeto del LSM</i>	12
<i>Figura 2.2 Patrón de la seña k de la LSE</i>	13
<i>Figura 2.3 Contorno ponderado</i>	13
<i>Figura 2.4 Rasgos visuales</i>	14
<i>Figura 2.5 Detección de punta de los dedos</i>	15
<i>Figura 2.6 Huella caracterizada multicanal del radioisótopo ⁶⁰Co</i>	15
<i>Figura 2.7 Espectro de una máquina rotatoria</i>	16
<i>Figura 3.1 Cubo unitario del modelo de color RGB</i>	17
<i>Figura 3.2 Dilatación</i>	19
<i>Figura 3.3 Erosión</i>	20
<i>Figura 3.4 Apertura</i>	20
<i>Figura 3.5 Clausura</i>	21
<i>Figura 3.6 Corrección de iluminación no uniforme</i>	22
<i>Figura 3.7 Histograma</i>	23
<i>Figura 3.8 Ejemplo de diferentes métodos de umbralado</i>	25
<i>Figura 3.9 Ejemplo de algunos rasgos extraídos de los momentos invariantes de Hu</i>	29
<i>Figura 3.10 Secciones cónicas y excentricidad</i>	31
<i>Figura 3.11 Nube de puntos</i>	33
<i>Figura 3.12 Imagen de profundidad en pseudocolor</i>	33
<i>Figura 3.13 Modelo de red neuronal de una capa</i>	34
<i>Figura 3.14 Representación matricial del modelo de la figura 3.13</i>	35
<i>Figura 4.1 Etapas del reconocedor del alfabeto del LSM del primer método de solución propuesto</i>	38
<i>Figura 4.2 Imágenes de entrada del primer método de solución propuesto</i>	39
<i>Figura 4.3 Corrección de la iluminación</i>	39
<i>Figura 4.4 Métodos de umbralado probados en el primer método propuesto</i>	41
<i>Figura 4.5 Imagen binaria segmentada con objetos indeseables</i>	41
<i>Figura 4.6 Datos visuales de la forma y orientación de la mano</i>	42
<i>Figura 4.7 Presentación del resultado del primer método propuesto</i>	43
<i>Figura 4.8 Etapas del reconocedor del alfabeto del LSM del segundo método de solución propuesto</i>	44
<i>Figura 4.9 Imagen en pseudocolor de un mapa de profundidad</i>	45
<i>Figura 4.10 Segmentado por profundidad</i>	46
<i>Figura 4.11 Visualización del usuario de la zona segmentada</i>	46
<i>Figura 4.12 Imagen binaria del objeto de interés</i>	47
<i>Figura 4.13 Una matriz evolutiva.</i>	48
<i>Figura 4.14 Una matriz evolutiva sin reforzamiento</i>	49
<i>Figura 4.15 Representación gráfica de una matriz evolutiva de patrones sin reforzamiento</i>	49
<i>Figura 4.16 Resultado del proceso de aprendizaje en una matriz evolutiva sin reforzamiento</i>	52
<i>Figura 4.17 Una matriz evolutiva con reforzamiento</i>	53
<i>Figura 4.18 Representación gráfica de una matriz evolutiva de patrones con reforzamiento</i>	53
<i>Figura 4.19 Representación gráfica de una matriz evolutiva de patrones reforzados</i>	56
<i>Figura 4.20 Un patrón reforzado</i>	56
<i>Figura 4.21 Un patrón evolucionado</i>	57
<i>Figura 4.22 Ejemplo de patrones de letras almacenados en la base de datos</i>	58
<i>Figura 4.23 Un patrón de la letra A</i>	58
<i>Figura 4.24 Ajuste del patrón de la letra A en una retícula de 5 x 5 celdas</i>	59
<i>Figura 4.25 Discretización binaria de la letra A</i>	59

<i>Figura 4.26 Vector del patrón de la letra A ajustada</i>	60
<i>Figura 4.27 Un patrón de la matriz evolutiva reforzado</i>	60
<i>Figura 5.1 Imagen de la letra D</i>	64
<i>Figura 5.2 Imagen resultante del proceso de umbralado de la imagen de la letra D mediante el método de Kittler e Illinworth</i>	64
<i>Figura 5.3 Visualización del ángulo y el contorno de la letra D</i>	65
<i>Figura 5.4 Resultado del reconocimiento de la letra D en forma de texto</i>	65
<i>Figura 5.5 Ejemplo de operación del aprendizaje-reconocimiento del segundo método de solución propuesto</i>	69
<i>Figura 5.6 Diferencias entre patrones de las letras</i>	70
<i>Figura 5.7 Comparación del número de señas reconocidas.</i>	71

ASL: Lenguaje de Signos Americano (American Sign Language).

EE: Elemento de estructura.

FPS: Frames Per Second (cuadros por segundo o imágenes por segundo).

ININ: Instituto Nacional de Investigaciones Nucleares.

IR: Infrarrojo.

JPG: Es un formato de compresión de imágenes, tanto en color como en escala de grises. JPG son las siglas de Joint Photographic Experts Group creadores del formato.

LED: Diodo de material semiconductor que emite luz al aplicar una tensión eléctrica, y que se utiliza para componer imágenes en pantallas de diversos instrumentos, como calculadoras, relojes digitales (Light-Emitting Diode).

LSE: Lengua de Signos Española.

LSM: Lenguaje de Señas Mexicano.

RGB: Red, Green, Blue.

RDI: Región De Interés, también conocido por sus siglas en ingles ROI.

ROI: Región Of Interest.

SRAO: Sistema para el Reconocimiento Automatizado de Objetos.

WTH: Sombrero de Copa Blanco (White Top Hat).

Clúster: Técnica en la cual se utiliza la capacidad de procesamiento de un conjunto o conglomerado de computadoras que se comportan como una única computadora con una capacidad mayor en procesamiento y recursos de hardware.

Entropía: En la teoría de la información, es el grado de incertidumbre que existe sobre un conjunto de datos.

Excentricidad: Es un parámetro que determina el grado de desviación de una sección cónica con respecto a una circunferencia.

Evolución: Cambio de forma, desarrollo o transformación de las ideas o de las teorías, mudanza de conducta, de propósito o de actitud, desarrollo de las cosas o de los organismos, por medio del cual pasan gradualmente de un estado a otro [1].

Evolucionar: Mudanza de conducta, de propósito o de actitud, dicho de un organismo o de otra cosa. Desenvolverse, desarrollarse, pasando de un estado a otro [1].

Evolución biológica: Proceso continuo de transformación de las especies a través de cambios producidos en sucesivas generaciones [1].

Hipoacusia: Disminución de la capacidad auditiva; pérdida parcial de la capacidad auditiva.

Hipoacúsico: Persona que no oye, o que no oye bien.

Imagen de profundidad: Imagen obtenida con base en una nube de puntos; la imagen obtenida del Kinect usa el espectro de la luz infrarroja para su captura.

Kinect: Dispositivo creado por la compañía Primesense. Este dispositivo permite capturar imágenes en el modelo de color RGB, y que provee una imagen de profundidad utilizando luz infrarroja.

Normalizar: Ajustar algo a una norma o intervalo; procesar la información para que dos conjuntos diferentes de datos sean comparables.

Píxel: Elemento discreto de una imagen digital, cuyo valor indica la intensidad del color o del nivel de gris de ese punto en la imagen.

Patrón: Representación numérica de un objeto dado.

Segmentación: Consiste en subdividir una imagen en sus respectivas regiones u objetos, sin que ningún píxel pertenezca a más de una región u objeto.

Umbral: Valor o nivel de gris usado para segmentar los valores de intensidad de una imagen en dos o más regiones diferentes.

Umbralado: Proceso en el que dada una imagen en escala de grises, se genera una nueva imagen digital en la cual solamente existen dos intensidades.

Actualmente, existen en el mundo millones de personas con algún tipo de discapacidad que enfrentan problemas de discriminación, por ejemplo, en las actividades académicas [2] y en diversas oportunidades de empleo [3].

En México, el censo del año 2010 indica que la población con algún tipo de discapacidad era de 5'739,270 personas [4]; este sector representa el 5.1% del total de la población. Aproximadamente el 35.5% de esta población no tuvo acceso o no logró aprobar ningún grado dentro del sistema educativo por causa de su discapacidad [5]. El promedio de escolaridad de este grupo de personas es menor al cuarto grado de primaria. En este grupo de personas están consideradas aquellas que padecen hipoacusia aguda.

Algunos de los problemas de discapacidad mencionados, pueden ser parcial o totalmente resueltos mediante prótesis, código Braille o con la ayuda de animales entrenados para guiarlos. En el caso de las personas con hipoacusia se han desarrollado los lenguajes de señas que pueden ser utilizados incluso por personas sin discapacidad y enseñarse desde la instrucción primaria como parte de la cultura básica en los niños para permitirles la interacción con personas que presentan problemas del oído o del habla.

En esta tesis se propone el desarrollo de una herramienta de software para ayudar a las personas con hipoacusia aguda, mediante el reconocimiento del alfabeto del Lenguaje de Señas Mexicano (LSM) usando imágenes de profundidad obtenidas de la cámara infrarroja del dispositivo Kinect.

En esta tesis se utiliza el término Matriz Evolutiva [6] [7] [8] [9] como una forma de representación de una red neuronal de una capa, en la cual cada neurona recibe las entradas de los píxeles que forman las imágenes de manos que representan señas del alfabeto del Lenguaje de Señas Mexicano. Se considera evolutiva en el sentido de la transformación constante de las neuronas que la forman y de la representación de cada patrón, conforme se capturan imágenes de entrada.

La palabra evolutivo de acuerdo con la Real Academia Española [1], se refiere al “desarrollo de las cosas o de los organismos, por medio del cual pasan gradualmente de un estado a otro” y también a una “doctrina que explica todos los fenómenos, cósmicos, físicos y mentales, por transformaciones sucesivas...”. Como la evolución por efecto de los elementos de la naturaleza, como un arroyo transforma un paisaje en un cañón o la lluvia transforma un arroyo en un río caudaloso. Como una tormenta tropical se transforma en un huracán.

Este documento está estructurado de la siguiente manera:

En el capítulo 1 se describen los problemas relacionados con el reconocimiento de imágenes y algunas consideraciones del funcionamiento del dispositivo Kinect. La sección 1.1 presenta el objetivo general y los objetivos específicos. La sección 1.2 presenta la justificación de la tesis junto con los beneficios esperados.

En el capítulo 2 se presenta la evolución de los lenguajes de señas y algunos trabajos relacionados con el tema de esta tesis.

En el capítulo 3 se describe el marco teórico relacionado con los dos métodos de solución propuestos en esta tesis.

En el capítulo 4 se presenta el análisis, diseño y los módulos de los dos métodos propuestos en esta tesis.

En el capítulo 5 se describen los resultados experimentales obtenidos.

Al final se presentan las conclusiones del trabajo realizado en esta tesis y las referencias.

Los lenguajes de señas tienen signos estáticos y dinámicos. Varios autores han desarrollado propuestas para reconocer imágenes de alfabetos de lenguajes de señas, enfocándose en los signos estáticos. Entre estos trabajos, se encuentra la propuesta de García Incertis [11] que reconoce algunas letras del alfabeto de la Lengua de Signos Española; por su parte, Razo Gil [12] reconoce imágenes de la mayoría de las letras del alfabeto del Lenguaje de Señas Mexicano.

El reconocimiento de los lenguajes de señas enfrenta problemas característicos de los Sistemas de Reconocimiento Automatizado de Objetos (SRAO) [13], en sus diferentes etapas (figura 1.1):

1. Captura de la imagen.
2. Preprocesamiento de la imagen.
3. Segmentación de las manos.
4. Detección y extracción de características relevantes de las manos.
5. Reconocimiento e interpretación de la seña expresada con las manos.

En la etapa de la captura de la imagen se tienen varios problemas. El primer problema es elegir un dispositivo adecuado a la aplicación considerando la velocidad de muestreo y la resolución de la imagen capturada, dado que existen diferentes tipos de cámaras, algunas con propósitos muy específicos.

El segundo problema es la configuración propia del dispositivo pues varía debido a que cada fabricante establece su propia forma de configurar.

El tercer problema es la selección y obtención de las mejores condiciones de operación del dispositivo de captura, entre estas las que deben considerarse: ubicación, iluminación, enfoque, encuadre, distancias, velocidad de captura. Por ejemplo en los sistemas para control de calidad en bandas transportadoras de productos se debe usar una frecuencia de muestreo rápida, con el fin de sincronizar la captura, con el momento en que la banda transportadora posiciona el producto en el área de encuadre de la cámara. Un problema común en el reconocimiento de objetos en imágenes es provocado por los cambios de luminosidad, como consecuencia de las diferentes condiciones de iluminación que causa variaciones en los colores, dificultando el proceso de segmentación. Por ejemplo, en la detección de la piel hay variaciones como consecuencia de las diferentes condiciones de iluminación en el modelo de color RGB; para resolver este problema, generalmente se transforman las imágenes capturadas a otro modelo de

color que tenga mayor tolerancia a los cambios de luminosidad, como el modelo de color HSV.

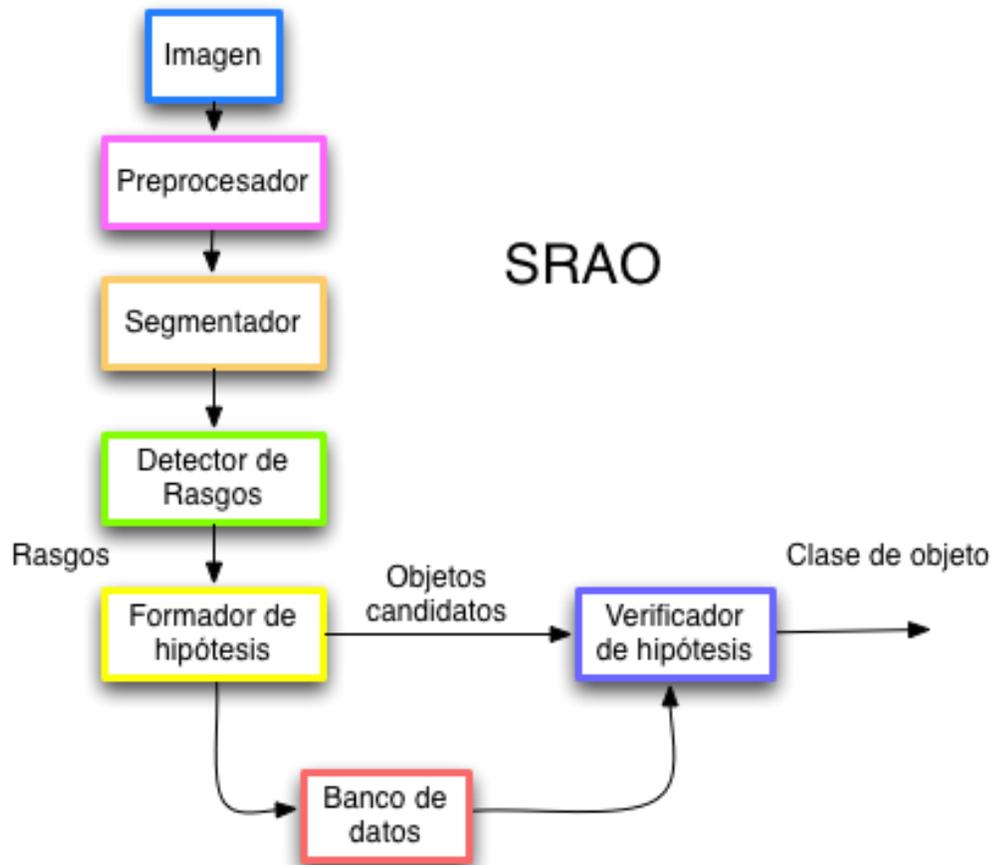


Figura 1.1 Etapas fundamentales del Sistema de Reconocimiento Automatizado de Objetos

El cuarto problema, es la transferencia de las imágenes del dispositivo de procesamiento de información y la compatibilidad del formato de salida del dispositivo, así como el tipo conectores de salida.

También existen problemas generales inherentes a las aplicaciones de visión por computadora como la iluminación artificial o natural, factores naturales en aplicaciones con condiciones no controlados como el clima. Estos problemas requieren del preprocesamiento especializado para un buen desempeño del sistema. En cuanto a la posición de los objetos existen variaciones generadas por las transformaciones geométricas [9] que crean distorsiones del objeto las cuales dificultan la extracción de características relevantes para el sistema. Las transformaciones geométricas las clasifica [14] como sigue:

1. Lineales

- Traslación (T)
- Rotación (R)
- Escalado isotrópico (S)
- Euclidea T + R
- Similitud T + R + S
- Afín
 - Similitud + S anisotrópico + Deformación (*shear*)
- Proyectiva

2. Polinómicas

3. Generales

Es difícil obtener las condiciones óptimas de trabajo, por lo que generalmente la imagen adquirida no tiene las mejores características, siendo necesario el preprocesamiento de la imagen, por lo que el tiempo de estos procesos debe de considerarse. En ocasiones se usa un clúster de computadoras y aprovechando unidades de procesamiento gráfico para mejorar, restaurar o filtrar ruido en las imágenes; también se usa para efectuar cambios de modelo de color o para procesar una imagen que permita que se identifique una mano.

El problema en la etapa de segmentación, es obtener una imagen donde los objetos candidatos sean fácilmente identificables; en esta etapa existen problemas como la determinación de umbrales, detección de la región de interés y problemas debido al ruido.

Para enfrentar el problema de luminosidad y de la segmentación de objetos de interés, se utiliza en esta tesis al dispositivo Kinect, que para estos problemas se aprovecha su cámara infrarroja.

El problema del reconocimiento de signos, se aborda de diferentes maneras, con sistemas basados en seguimiento del objeto, filtros de Kalman, gradientes basados en flujo óptico, historia del movimiento en imágenes, histograma orientado a gradientes, no orientados al seguimiento de objetos mediante análisis de componentes principales, distorsión de tiempo dinámico. Existen también, métodos enfocados en la detección de dedos mediante cadenas de Markov o métodos de esqueletizado, entre otros [15].



Figura 1.2 Problemas a considerar en la extracción de rasgos.

En la etapa de extracción de rasgos, el problema es determinar aquellos rasgos que son relevantes y caractericen la mano para permitir discriminar su significado sin comprometer la eficiencia en el proceso general. Dado que en la etapa de adquisición y por la naturaleza del problema se enfrentan algunos tipos de problemas generados por transformaciones geométricas como la traslación, escalamiento, rotación y deformaciones del objeto, se requieren condiciones particulares de operación con el propósito de tener invariantes y tolerancias ante estas transformaciones (figura 1.2).

En la etapa de reconocimiento se tienen problemas tales como una correcta clasificación, la interpretación, el almacenamiento de los patrones, la búsqueda y el indexado.

1.1 Objetivos

En esta sección se presenta el objetivo general y los objetivos específicos para lograr el reconocimiento de imágenes del alfabeto del Lenguaje de Señas Mexicano.

1.1.1 Objetivo general

Con el desarrollo de esta tesis se realizará una herramienta computacional que procese imágenes obtenidas de un dispositivo que detecta profundidad mediante una cámara infrarroja y que reconozca en las imágenes capturadas, las 21 señas estáticas del alfabeto del Lenguaje de Señas Mexicano y la base inicial de cuatro señas dinámicas.

1.1.2 Objetivos específicos

El objetivo general se alcanza con base en los objetivos específicos siguientes:

1. Capturar imágenes en forma de mapas de profundidad mediante el dispositivo Kinect.
2. Aislar la mano de las imágenes de profundidad, mediante la segmentación para lo cual, se establecen los límites inferior y superior de la profundidad de interés; con las ventajas de reducir el problemas de luminosidad y mejorar la eficiencia en el procesamiento.
3. Normalizar la imagen a un tamaño de patrón específico para obtener tolerancia a escala.
4. Asignar el significado a las manos que representan una seña del alfabeto del Lenguaje de Señas Mexicano, con base en la comparación con una base de conocimiento en el que se almacenan los patrones.
5. Ampliar la base de conocimiento cuando el patrón obtenido se desconozca o cuando tenga una diferencia mayor que el 10% respecto a los patrones existentes.
6. Crear una interfaz para facilitar la visualización de las señas de entrada.

1.2 Justificación

Cuando una persona presenta hipoacusia leve o aguda, se ve en la necesidad de recurrir a alternativas que le ayuden a mejorar su comunicación; mediante herramientas tecnológicas o con algún tratamiento médico o quirúrgico. Otra opción para comunicarse, es el uso de los lenguajes de señas.

Cuando se usan los lenguajes de señas, se necesita un modelo bilingüe de educación, donde se maneja el lenguaje de señas asociándolo con acciones y/o palabras escritas.

El lenguaje de señas no es una lengua universal [16]; cada país tiene su propia lengua de señas y como consecuencia de esto, si interactúan personas de dos países diferentes para lograr una comunicación efectiva se requiere de un intérprete o traductor. En el caso de México se conoce como Lenguaje de Señas Mexicano (LSM) ISO 639-2 sgn-MX [17]. Si la persona desconoce el LSM, se requiere también de un intérprete que genere sonidos a partir del reconocimiento de las señas y que exprese con señas las frases que se le comuniquen mediante sonidos.

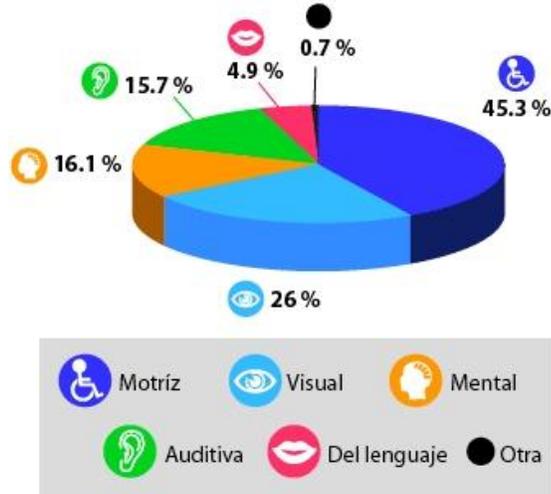


Figura 1.3 Porcentaje de la población por tipo de discapacidad

Se sabe que en México hay 5'739,270 personas con algún tipo de discapacidad física o mental [2], o con un problema de salud de largo plazo, que les impide realizar con plenitud algunas actividades que llevarían a cabo en condiciones normales. Este volumen representa 5.1 % de la población total del país [4]. En la figura 1.3 se muestra el porcentaje de personas que presentan algún tipo de discapacidad.

En relación con el nivel de escolaridad muestra que las personas con discapacidad auditiva de más de 15 años, apenas 5.9 % ha completado la educación básica, 4.4 % tenían educación media superior y 3.2 % lograron estudios superiores. En contraste, en E.U.A., el 30 % de las personas con algún tipo de discapacidad cuenta con carrera universitaria o superior [3]. En México la deserción escolar de personas con alguna discapacidad es alta y las escuelas especializadas son escasas.

En la figura 1.4 se observa la distribución porcentual de la población de personas con hipoacusia con más de 15 años de edad, por sexo según el nivel de instrucción y su promedio de escolaridad [5].

1.2.1 Beneficios esperados

Se espera que el sistema desarrollado sea útil a las personas que tienen problemas de audición severa. El sistema tiene un uso potencial como herramienta para la enseñanza del alfabeto del LSM, sirviendo como un medio de validación de la ejecución correcta de las señas expresadas por las personas que estén aprendiendo este alfabeto, tengan o no hipoacusia, como los familiares de las personas con problemas de audición.

Nivel de instrucción	General	Hombres	Mujeres
Sin instrucción	39.4	35.7	44.0
Primaria incompleta	28.7	31.0	25.9
Primaria completa	14.3	14.7	13.9
Secundaria incompleta	2.1	2.4	1.9
Secundaria completa	5.9	6.2	5.6
Media superior	4.4	4.3	4.4
Superior	3.0	3.6	2.2
Posgrado	0.2	0.3	0.2
No especificado	2.0	1.8	1.9
Total	100.0	100.0	100.0
Promedio de escolaridad	3.4	3.6	3.1

Figura 1.4 Escolaridad de personas con hipoacusia

1.2.2 Alcances

Los alcances del sistema desarrollado en esta tesis, se listan a continuación:

1. El sistema propuesto se enfoca hacia el reconocimiento de las señas estáticas del alfabeto del Lenguaje de Señas Mexicano.
2. El ambiente de operación requiere de ubicaciones donde no se reciban rayos solares directamente, debido a que afectan la captura correcta de las imágenes en el espectro infrarrojo usado por el dispositivo Kinect, por lo que en ausencia de luz solar, la captura de las imágenes del dispositivo infrarrojo sufren menos alteraciones.
3. El sistema reconoce las señas de la mano dominante si se trata de una persona zurda o diestra, dado que el sistema aprende los patrones que se le presentan.
4. El sistema reconoce imágenes de las manos capturadas con el dispositivo Kinect, con una resolución entre 100 x 100 píxeles y 400 x 400 píxeles.

En este capítulo se describe en la sección 2.1 el Lenguaje de Señas Mexicano hacia el que se enfoca esta tesis.

En las secciones 2.2 a 2.5 se comentan brevemente algunos de los trabajos de investigación que se han desarrollado en el mundo, usando diferentes herramientas tecnológicas, para el reconocimiento de alfabetos de señas y gestos expresados con la mano.

En las secciones 2.6 y 2.7 se incluyen dos trabajos que muestran la aplicación de las matrices evolutivas para el reconocimiento de patrones.

2.1 Lenguaje de Señas Mexicano

El lenguaje de señas es probablemente una de las primeras formas de comunicación utilizadas por los humanos antes de la invención del lenguaje hablado. En 1620, Juan Pablo de Bonet en España, escribió el primer libro para enseñar el lenguaje de señas para las personas sordomudas. En 1755 Charles Michele de L'Epee en París, Francia fundó la primera escuela en el mundo para personas con discapacidad auditiva. En 1778 Samuel Heinicke de Leipzig, fundó la primera escuela en Alemania para sordomudos. En los Estados Unidos los nativos tenían un sistema de comunicación entre ellos, parecido al de la gente sordomuda. Thomas Hopkins Gallaudet, con el fin de ayudar a su vecina Alice Cogswell, viajó a Francia en 1815 para estudiar métodos de comunicación para personas sordomudas y en consecuencia en 1817 Gallaudet fundó en Hartford, Connecticut, la primera escuela para gente sorda en los Estados Unidos de América.

El sistema de Lenguaje de Signos Americano está basado en el sistema francés. El lenguaje de Señas Mexicano, así como el Lenguaje de Señas Americano tienen sus raíces en la propuesta de Charles Michele de L'Epee. LSM engloba diferentes enfoques lingüísticos fundados en México.

El alfabeto del Lenguaje de Señas Mexicano está formado por señas estáticas y dinámicas (figura 2.1) [10]. Las señas dinámicas muestran flechas de trayectoria que indican la dirección del movimiento.

El alfabeto del LSM que se produce mediante la ejecución de señas usando una mano base y una mano dominante. La mano dominante es la izquierda para las personas zurdas y la derecha para las personas diestras.

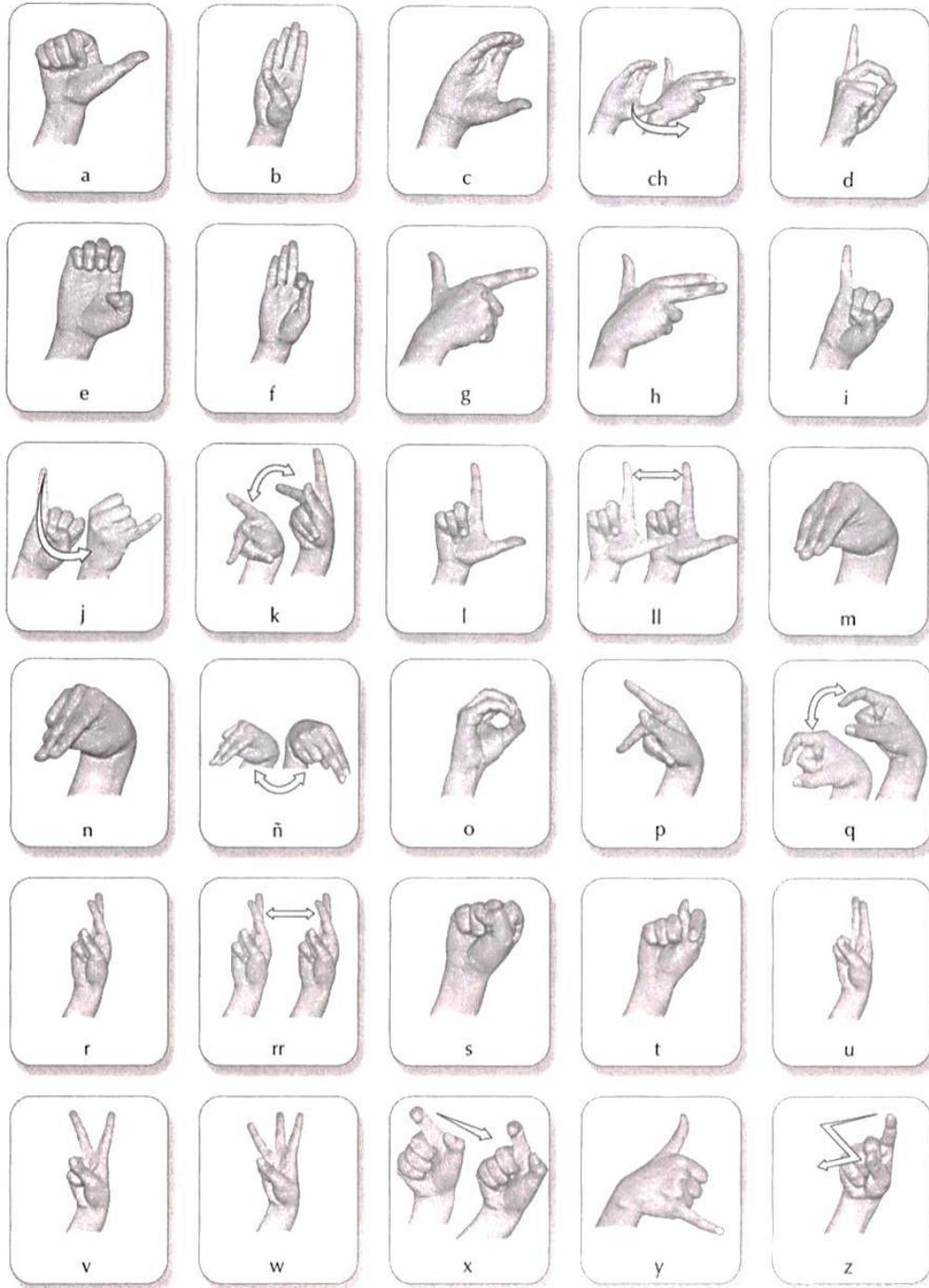


Figura 2.1 Alfabeto del LSM

2.2 Una interfaz para el reconocimiento de señas gestuales para personas hipoacúsicas

(García Incertis, Gómez García-Bermejo, Zalama Casanova 2006) [11].

En este trabajo se describe el reconocimiento de la Lengua de Signos Española mediante la identificación de la mano cubierta con un guante de color azul. Los contornos de la mano son extraídos de la imagen a través de un segmentado mediante el modelo de color HSV.

Después de la extracción, el contorno se muestrea y se obtiene una huella basada en la distancia del arco (figura 2.2). Los puntos resultantes de dicho muestreo se comparan con un diccionario de signos usando cuatro criterios de distancia. En los resultados obtenidos se reporta el reconocimiento de 19 letras del alfabeto de la Lengua de Signos Española (LSE).

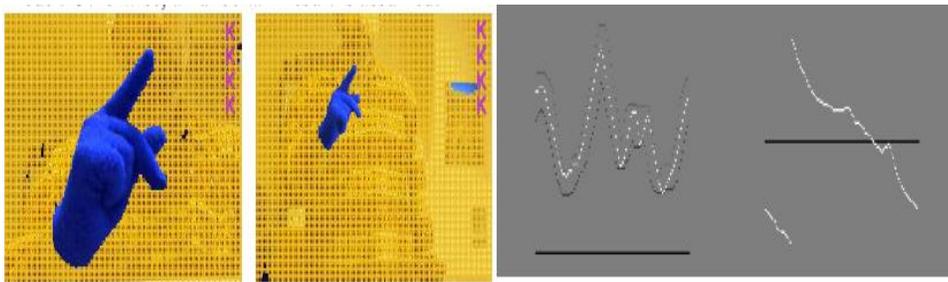


Figura 2.2 Patrón de la seña k de la LSE

2.3 Sistema para el reconocimiento del alfabeto dactilológico

(Razo-Gil 2009) [12].

En este trabajo se captura una seña, se realiza un umbralado por el método de Otsu para segmentar la mano, se extrae el contorno, y el código de contorno resultante se pondera, (figura 2.3), se clasifica el símbolo y se compara mediante un algoritmo de votación para obtener el significado del signo en caso de ser reconocido.

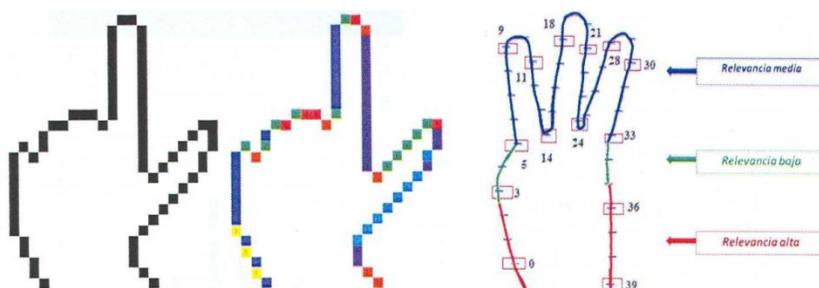


Figura 2.3 Contorno ponderado

2.4 Reconocimiento de gestos manuales utilizando Kinect (Yi Li)

(Yi Li 2012) [18].

En esta propuesta se realiza primero la extracción de las manos utilizando el dispositivo Kinect, después se realiza el cálculo de la envolvente convexa usando el algoritmo de Graham. Se obtiene el contorno de la mano y los defectos de convexidad para identificar y obtener un vector para cada uno de los dedos. Finalmente el reconocimiento de cada seña se clasifica por medio del conteo de dedos, por el dedo específico detectado y por una comparación de vectores de cada dedo. Algunas de las señas reconocidas se pueden observar en la figura 2.4; este algoritmo reporta 9 señas reconocidas.

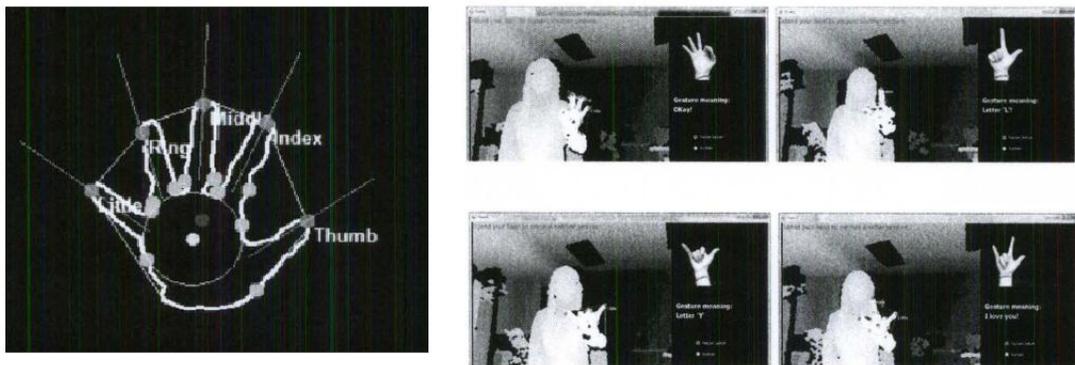


Figura 2.4 Rasgos visuales

2.5 Reconocimiento de gestos manuales utilizando Kinect (Heng Du)

(Heng Du, TszHang 2011) [19].

En esta propuesta se obtiene una imagen del mapa de profundidad del dispositivo Kinect para segmentar un área de interés y a partir de ella continuar el procesamiento usando tonos de gris. Una vez eliminado el fondo y aislada la mano, se elimina el ruido usando un filtro, después se extrae el contorno de la mano y se realiza una aproximación poligonal a la línea del contorno con el fin de encontrar los defectos de convexidad. Los defectos se filtran para detectar las puntas de los dedos como se muestra en la figura 2.5. El reconocimiento se hace por medio del conteo de dedos y los resultados reportan la detección de 5 señas de la mano.



Figura 2.5 Detección de punta de los dedos

2.6 Aplicación de los sistemas evolutivos en el análisis de espectros de rayos gamma

(Torres, Longoria, Rojas 1998)[20].

En esta propuesta se utilizan matrices evolutivas para analizar y reconocer patrones característicos obtenidos de señales de un sistema de espectroscopía especializado, el cual analiza los elementos fotónicos generados por la irradiación con neutrones en un reactor nuclear. El sistema genera huellas o patrones de espectros de rayos gamma que posteriormente puede identificar y responder a las diferentes señales aprendidas. En la figura 2.6 se muestra la huella obtenida con ocho espectros del radioisótopo ^{60}Co .

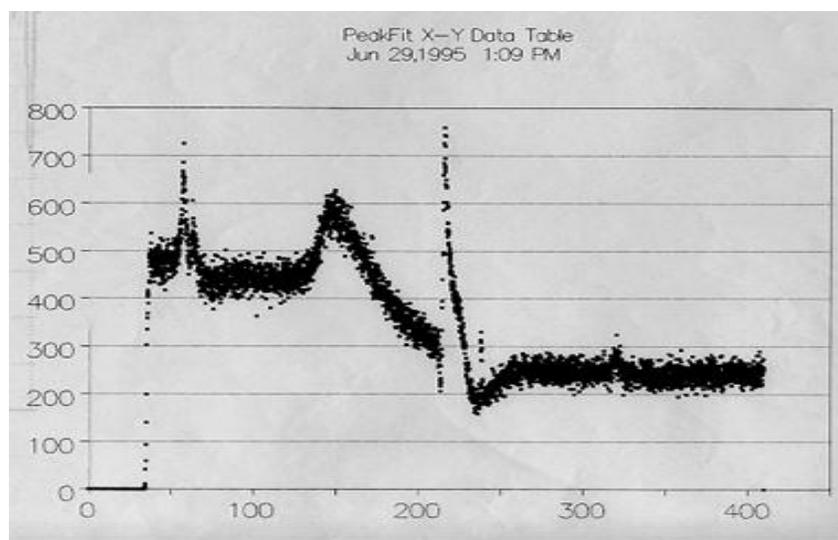


Figura 2.6 Huella caracterizada multicanal del radioisótopo ^{60}Co

2.7 Sistema evolutivo para el diagnóstico de fallas en máquinas rotatorias

(De la Cruz, Longoria, Carrillo 1998)[21].

Este trabajo se enfoca en el diagnóstico de fallas en máquinas rotatorias, para identificar fallas como las que se producen por el desbalanceo en motores eléctricos. Este sistema crea y reconoce diferentes huellas generadas por los espectros de señales vibratorias, obtenidos mediante el muestreo usando acelerómetros, amplificadores sensitivos y un analizador dinámico de señales. El reconocimiento se realiza usando matrices evolutivas como representación de una red neuronal. Con esta propuesta se reconocen condiciones anormales en máquinas rotatorias. En la figura 2.7 se muestra la huella de un ventilador con un desbalanceo inducido de 0.915 gramos.

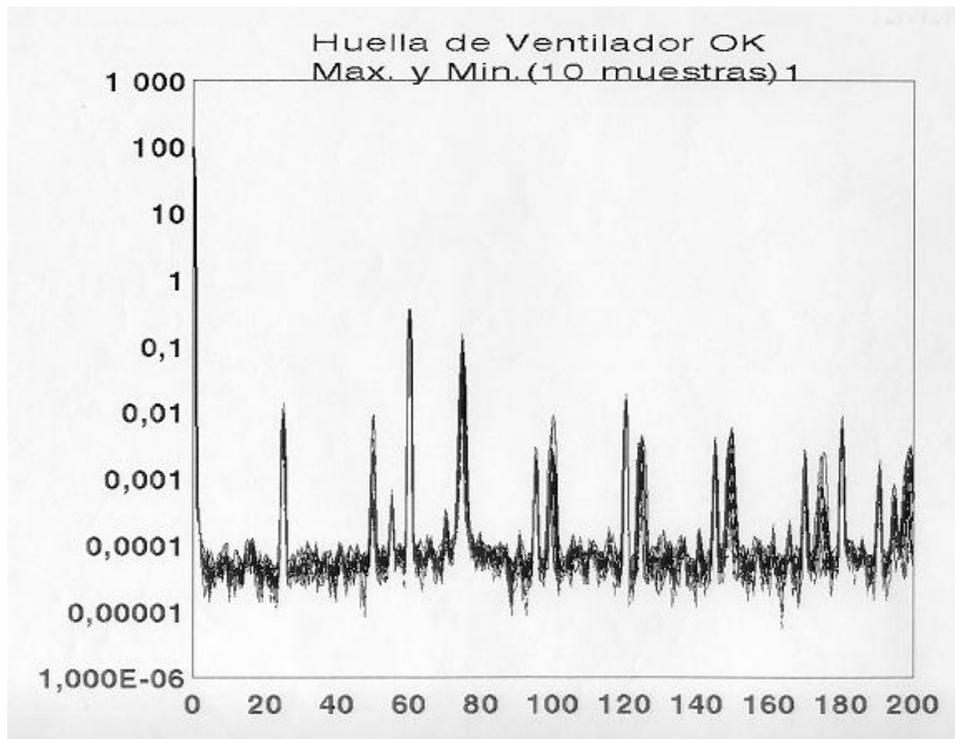


Figura 2.7 Espectro de una máquina rotatoria

En este capítulo se presentan los conceptos utilizados en el desarrollo de los métodos propuestos para el reconocimiento del alfabeto del Lenguaje de Señas Mexicano.

3.1 Modelo de color RGB

El sistema RGB emplea un sistema de coordenadas cartesianas y está basado en la síntesis aditiva de colores; es decir, un color se obtiene dependiendo de las diferentes proporciones de las componentes roja, azul y verde (RGB). Este modelo de color se puede representar como un cubo unitario (figura 3.1). La diagonal principal que contiene la misma cantidad de RGB, representa los niveles de grises. El negro está situado en el origen ubicado en las coordenadas $(0, 0, 0)$, el blanco está en la coordenada del punto $(1, 1, 1)$.

Todos los colores quedan definidos como puntos que se encuentran en el cubo o en el interior del mismo. Las imágenes en este modelo de color están formadas por tres planos independientes, es decir un plano correspondiente para cada color.

El uso de este modelo de color en las imágenes que se procesan puede implicar más restricciones o consideraciones en la etapa de captura de la imagen, pues no hay separación entre la información cromática y la información de luminancia [23] [25] [26].

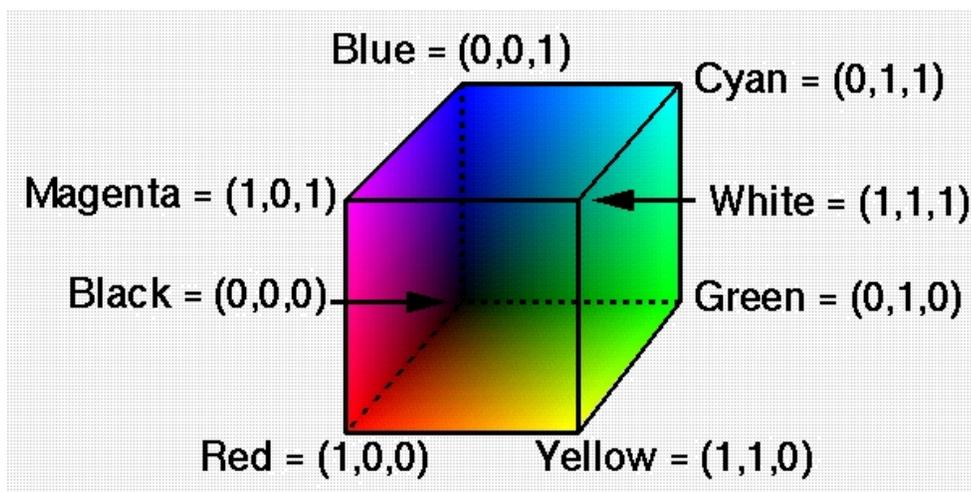


Figura 3.1 Cubo unitario del modelo de color RGB

3.2 Morfología matemática

La Morfología matemática [25] [26] [27] [28] [29], es una teoría que contiene un conjunto robusto de técnicas de procesado no lineal para el análisis y tratamiento de las estructuras geométricas.

Las bases teóricas de la morfología matemática se deben al científico alemán nacido en Rusia, Hermann Minkowski (1864-1909). Entre los muchos y variados logros a lo largo de su fructífera vida profesional, trabajó en el tema de las figuras convexas y las relaciones entre sus formas, siendo precisamente en este campo donde creó las bases matemáticas fundamentales para la Morfología Matemática.

Minkowski fue el primer ser humano al que se le ocurrió sumar formas. Siendo A y B dos conjuntos cualesquiera, sobre cuyos elementos esté bien definida la operación binaria suma (+), el conjunto de la suma de las dos formas A y B contiene todos los elementos que resultan de sumar cada uno de los elementos del conjunto A con todos y cada uno de los elementos del conjunto B.

$$A+B=\{x=a+b, \text{ con } a \in A \text{ y } b \in B\}$$

3.2.1 Elemento de estructura

Al patrón de ajuste que se utiliza para examinar las estructuras geométricas de una imagen en Morfología Matemática se le conoce como elemento de estructura E.E.

El elemento de estructura (EE) se maneja de modo análogo a una máscara de convolución, aunque en este caso se trabaja con operaciones sobre conjuntos y la convolución se basa en operaciones aritméticas.

El elemento de estructura puede tener cualquier tamaño y forma ya sea horizontal, vertical, redonda, cuadrada etc.

Es importante definir en cada operación la referencia del elemento de estructura; generalmente el punto está situado en el centro para conservar simetría en el elemento de estructura [25] [28] [30] [32].

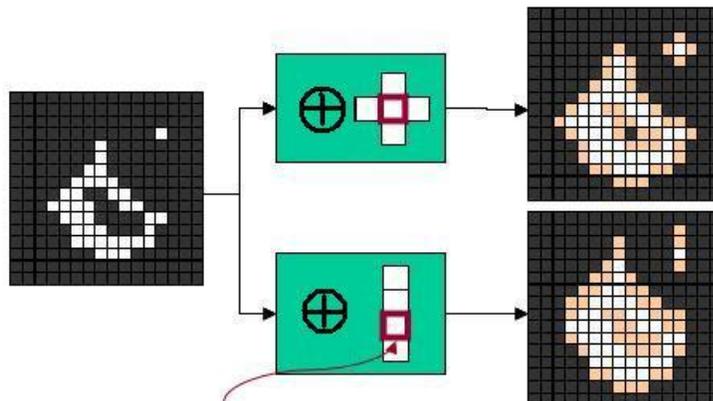
3.2.2 Dilatación binaria \oplus

Dilatación de la imagen A por el Elemento de Estructura C. En este caso A no es una imagen, sino un conjunto, el resultado de la dilatación del conjunto A, también es un conjunto. Sin embargo, por brevedad, en lo sucesivo

utilizaremos expresiones como "la imagen A" para referirnos a "la imagen A(x, y) definida por el conjunto A" [25] [28] [30] [32].

$$\delta_C(A) = A \oplus C = \{x | (\hat{C})_x \cap A \neq \emptyset\}$$

La salida de la dilatación es el conjunto de puntos barridos por el centro del elemento de estructura donde coincide con algún punto de A. Como se observa en la figura 3.2, la dilatación se puede interpretar como el resultado de reemplazar cada píxel blanco de la imagen original por una réplica del elemento de estructura.



(En esta figura y las siguientes, el origen de coordenadas del EE aparece destacado)

Figura 3.2 Dilatación

3.2.2.1 Propiedades de la dilatación

1. La dilatación es extensiva.
2. Rellena algunos agujeros que son más pequeños que el elemento de estructura.

3.2.3 Erosión binaria \ominus

Erosión de la imagen A por el Elemento de Estructura C:

$$\varepsilon_C(A) = A \ominus C = \{x | C_x \subset A\}$$

La salida de la erosión, es el conjunto de puntos barridos por el centro del elemento de estructura donde se cumpla que todos los puntos de C estaban contenidos en A [32].

Elimina grupos de píxeles donde el elemento de estructura no cabe, como se muestra en la figura 3.3.

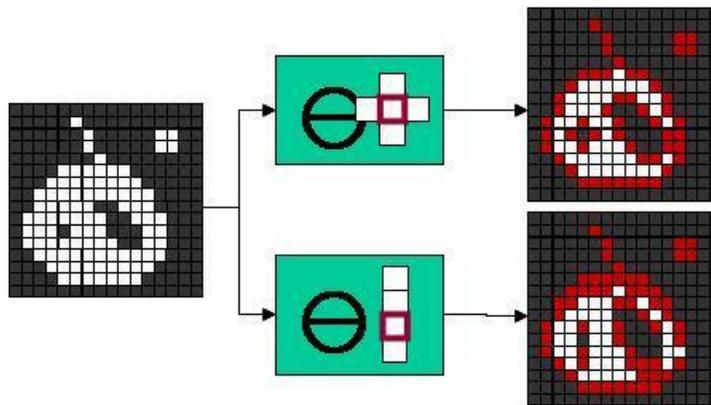


Figura 3.3 Erosión

3.2.3.1 Propiedades de la erosión

1. La erosión es antiextensiva.
2. Elimina pequeñas islas y protuberancias.

3.2.4 Apertura binaria °

Es la composición de un operador de erosión seguido de otro operador de dilatación con el mismo elemento de estructura.

$$\gamma_B(A) = A \circ B = (A \ominus B) \oplus B = \delta_B(\varepsilon_B(A))$$

Se obtiene desplazando el elemento de estructura B por el interior del conjunto y eliminando las zonas por las que B no puede pasar [32], en este caso los píxeles eliminados se resaltan en rojo, como se muestra en la figura 3.4.

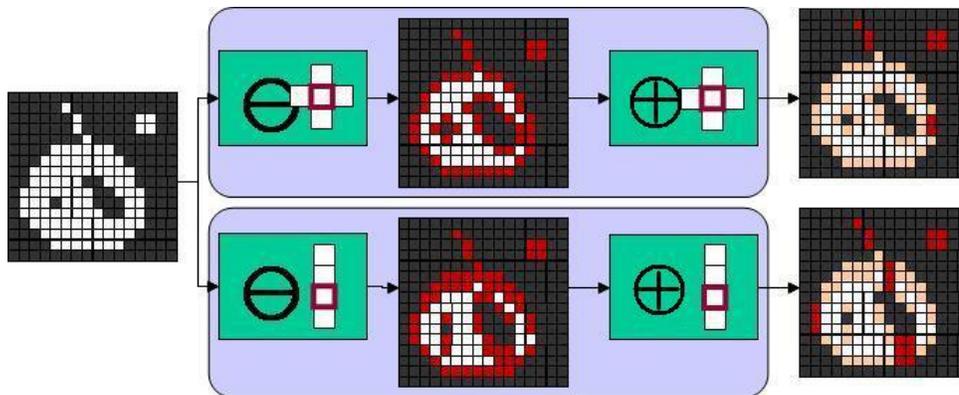


Figura 3.4 Apertura

3.2.4.1 Propiedades de la apertura

1. Es un operador antiextensivo por lo tanto $A^\circ B$ es un subconjunto de A .
2. Es un operador idempotente por lo tanto $(A^\circ B)^\circ B = A^\circ B$.
3. Este operador alisa contornos, es decir, redondea las esquinas donde no quepa el elemento de estructura, elimina las protuberancias y separa objetos en puntos estrechos.

3.2.5 Clausura binaria •

Es la composición de un operador de dilatación seguido de otro operador de erosión con el mismo elemento de estructura [30] [32].

$$\varphi_B(A) = A \bullet B = (A \oplus B) \ominus B = \varepsilon_B(\delta_B(A))$$

Se obtiene desplazando el elemento de estructura B por el exterior del conjunto y rellenando las zonas por las que B no puede pasar, en este caso los píxeles añadidos se resaltan en rojo, como se muestra en la figura 3.5.

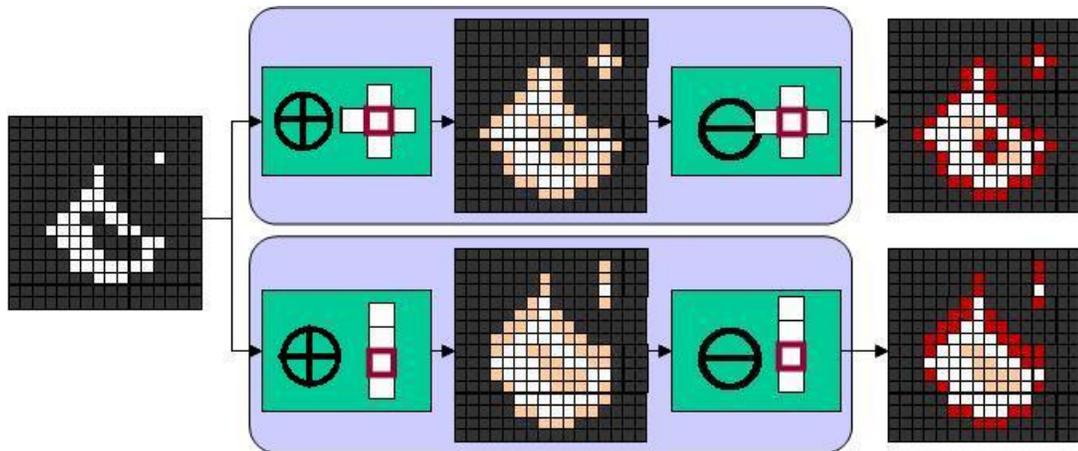


Figura 3.5 Clausura

3.2.5.1 Propiedades de la clausura

1. Es un operador extensivo por lo tanto $A \subset A \bullet B$.
2. Es un operador idempotente por lo tanto $(A \bullet B) \bullet B = A \bullet B$
3. Este operador tiende a alisar porciones del contorno, fusiona grietas estrechas, rellena agujeros pequeños y conecta objetos vecinos.

3.2.6 Sombrero de Copa Blanco WTH

El sombrero de copa blanco o White Top Hat de una imagen es la diferencia entre la imagen original denotada como f y la apertura en niveles de grises de dicha imagen [25] [28].

$$WTH(f) = f - \gamma(f)$$

Esta operación puede ser utilizada para corregir problemas de iluminación no uniforme [25] [28] como se muestra un ejemplo en la figura 3.6.

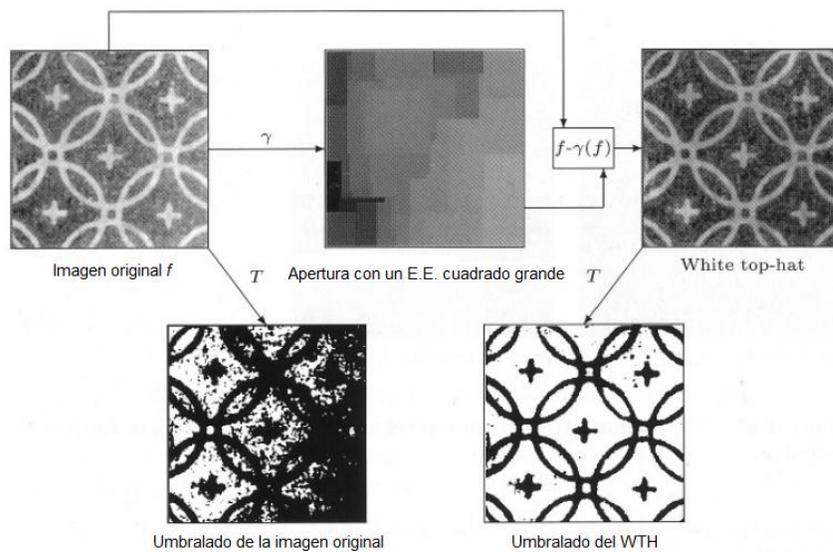


Figura 3.6 Corrección de iluminación no uniforme

3.2.6.1 Propiedades del Sombrero de Copa Blanco

1. Es un operador antiextensivo.
2. Es un operador idempotente.

3.3 Histograma

El histograma de una imagen digital se representa generalmente como un gráfico de barras en el que el eje coordenado x , representa los distintos niveles de intensidad de la imagen, y el eje coordenado y es la frecuencia con el que cada color está presente en la imagen.

El histograma de una imagen con L niveles de intensidad en el intervalo cerrado $[0, L-1]$, se denota como $P(r_k)$ y es una función discreta de la forma:

$$P(r_k) = \frac{n_k}{N}$$

donde r_k es el k -ésimo nivel de gris, n_k es el número de píxeles en la imagen con el nivel de intensidad r_k y N el número total de píxeles en la imagen. Un ejemplo de histograma [13] se puede ver en la figura 3.7.

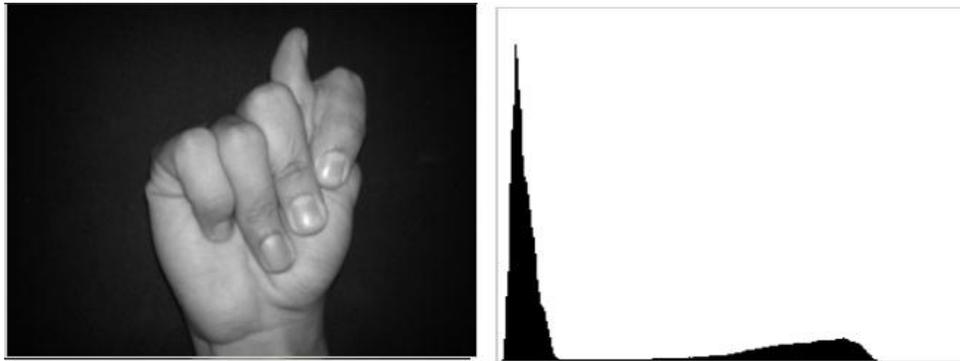


Figura 3.7 Histograma

3.4 Métodos de segmentación basados en el umbralado

En muchas aplicaciones de procesamiento de imágenes, los niveles de gris de los píxeles pertenecientes a los objetos son sustancialmente diferentes a los niveles de gris de los píxeles que pertenecen al fondo. Para estos casos el umbralado se convierte en una simple, pero eficaz herramienta para separar a los objetos del fondo.

El resultado de la operación de umbralado es una imagen binaria donde uno de los estados representa el objeto de interés, como lo es un texto impreso, una leyenda, una pieza defectuosa, una mano, etc., mientras que el estado complementario corresponde al fondo de la imagen.

Dependiendo de la aplicación, el primer valor representado por nivel de gris es 0 negro en el caso del texto y 1 para el fondo de la más alta luminancia por el papel del documento.

Varios factores como el ruido, la iluminación ambiental, contrastes inadecuados, tamaño del objeto no acorde a la escena complican la operación de umbralado.

Sezgin y Sankur [31] clasificaron los métodos de umbralado en seis categorías de acuerdo a la información en que se enfoca cada uno de estos métodos.

1. Métodos basados en la forma del histograma, donde características como las crestas, valles y curvaturas suavizadas se analizan.
2. Métodos basados en agrupamiento, donde las muestras de un nivel de gris son agrupadas en dos partes, una parte como fondo y otra parte como el objeto de interés, o alternativamente como una mezcla de dos Gaussianas.
3. Métodos basados en la entropía resultan en algoritmos que usan la entropía del objeto y de las regiones de fondo, la entropía cruzada entre el original y la imagen binarizada, etc.
4. Métodos basados en atributos que buscan una medida de similitud entre el nivel de gris y la imagen binaria resultante, como la similitud en las formas difusas y /o la correlación entre píxeles.
5. Los métodos espaciales utilizan la distribución de probabilidad de orden alto y / o la correlación entre los píxeles.
6. Métodos locales que adaptan el valor de umbral en cada píxel a las características locales de la imagen.

En la figura 3.8, se muestra el resultado de algunos métodos de umbralado, la imagen original y el histograma de dicha imagen.

3.4.1 Método de umbralado de Kittler e Illinworth

También llamado método del error mínimo [13] [31]. En este método, el histograma es visto como un estimado de la función de densidad de probabilidad $P(g)$ de la población mezclada consistente de los niveles de gris de los píxeles del objeto y del fondo.

Se asume para las dos componentes de la mezcla $p(g|0)$ y $p(g|1)$, una distribución normal con parámetros para las medias μ_0 , μ_1 y desviaciones estándar σ_0 , σ_1 y a priori P_0 y P_1 , es decir:

$$p(g) = \sum_{i=1}^1 P_i \cdot p(g_i)$$

donde:

$$p(g) = \frac{1}{2\pi\sigma_i} \exp\left(-\frac{g - \mu_i}{2\sigma_i^2}\right)$$

El valor de umbral puede ser hallado al resolver la siguiente ecuación cuadrática:

$$\frac{(g - \mu_0)^2}{\sigma_0^2} + \ln \sigma_0^2 - 2 \ln P_0 = \frac{(g - \mu_1)^2}{\sigma_1^2} + \ln \sigma_1^2 - 2 \ln P_1$$

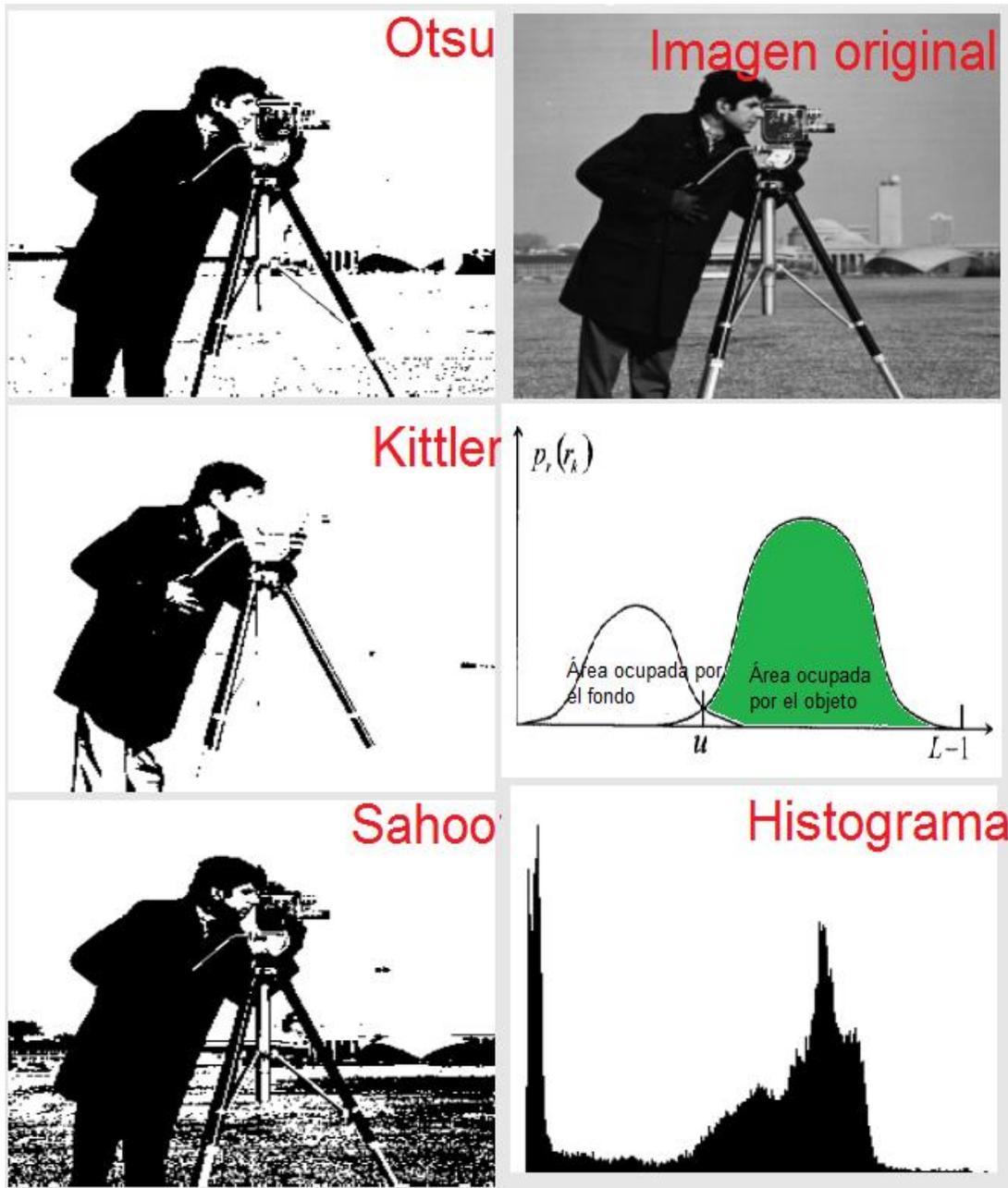


Figura 3.8 Ejemplo de diferentes métodos de umbralado

Los parámetros asociados con la densidad mixta $p(g)$ son usualmente desconocidos. Sin embargo, pueden ser estimados a partir del histograma, mediante las fórmulas siguientes:

$$P_0(K) = \sum_{g=0}^k h(g)$$

$$P_1(K) = \sum_{g=k+1}^{255} h(g)$$

$$\mu_0(K) = \frac{\sum_{g=0}^k g \cdot h(g)}{P_0(k)}$$

$$\mu_1(K) = \frac{\sum_{g=k+1}^{255} g \cdot h(g)}{P_1(k)}$$

$$\sigma_0^2 = \frac{\sum_{g=0}^k (g - \mu_0(k))^2 \cdot h(g)}{P_0(k)}$$

$$\sigma_1^2 = \frac{\sum_{g=k+1}^{255} (g - \mu_1(k))^2 \cdot h(g)}{P_1(k)}$$

Estos parámetros deben ser sustituidos en la siguiente función criterio:

$$J(k) = 1 + [P_0(k) \ln \sigma_0 + P_1(k) \ln \sigma_1] - 2[P_0(k) \ln P_0(k) + P_1(k) \ln P_1(k)]$$

El valor umbral k se determina al minimizar la función anterior:

$$k = \arg \min_{k \in [0, L-1]} \{J(k)\}$$

3.5 Algoritmo de etiquetado de componentes conexas iterativo

Una de las operaciones más comunes en visión artificial, es la de encontrar las componentes conectadas dentro de una imagen [13]. Los puntos en una componente conectada forman candidatos para la representación de un objeto, por lo que el etiquetado de componentes conexas permite encontrar el número de objetos que hay en una imagen. Las etiquetas puestas a los objetos pueden ser mostradas en las imágenes a través de colores, niveles de gris, números, etc.

Este algoritmo iterativo no usa almacenamiento auxiliar para crear una imagen etiquetada a partir de una imagen binaria. Consta de tres pasos básicos, uno de etiquetado inicial, uno de propagación de arriba hacia abajo de etiquetas y, finalmente, uno de propagado de etiquetas de abajo hacia arriba.

Sea $b(x, y)$ la función que representa una imagen binaria y $e(x, y)$ la imagen resultante, entonces los pasos son los siguientes:

1. Barrer $b(x, y)$ hasta encontrar un pixel de tipo objeto (con valor 1 o L-1) aún no etiquetado y asignarle una nueva etiqueta E. Esto da como resultado la imagen $e_1(x, y)$.
2. Barrer $b(x, y)$ de arriba hacia abajo hasta encontrar un pixel etiquetado y propagar su etiqueta a sus vecinos, según la métrica elegida. Esto da como resultado la imagen $e_2(x, y)$.
3. Barrer $e_2(x, y)$ de abajo hacia arriba hasta encontrar un píxel etiquetado y propagar su etiqueta a sus vecinos, según la métrica elegida. Esto da como resultado la imagen $e_{\text{final}}(x, y)$.

3.6 Momentos de Hu

Tomemos $f(x, y)$ como la intensidad del punto (x, y) en una región [13] [23] [24]. El momento de orden $(p + q)$ para la región se define como:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x, y)$$

donde la sumatoria se toma sobre todas las coordenadas espaciales (x, y) de puntos de la región. El momento central de orden $(p + q)$ se calcula con la fórmula:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y)$$

3.6.1 Centro de gravedad del objeto

El centro de gravedad del objeto [13] [23] [24] se calcula usando la fórmula siguiente:

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad \bar{y} = \frac{m_{01}}{m_{00}}$$

3.6.2 Ejes principales de la elipse

Los momentos de orden dos son usados para determinar los ejes principales del objeto [13]. Los ejes principales pueden describirse como el par de ejes alrededor de los cuales se localizan los ejes menor y mayor. En términos de los momentos, las orientaciones de los ejes principales, ϕ , vienen dadas como:

$$\phi = \frac{1}{2} \tan^{-1} \left[\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right]$$

Los momentos centrales normalizados de orden $(p+q)$ se definen como:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu'_{pq}} \quad \text{donde} \quad \gamma = \frac{p+q}{2} + 1 \quad \text{para} \quad (p+q) = 2, 3, \dots$$

3.6.3 Momentos invariantes de Hu

El siguiente conjunto de momentos invariantes propuesto por HU (1962) [13] [23] [24], se puede obtener usando únicamente los momentos centrales normalizados de órdenes 2 y 3.

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

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

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

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

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

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

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

Para que todos los momentos contribuyan por igual en la función de discriminación, y por tanto estén dentro del mismo orden de magnitud, a veces es necesario proceder a una normalización del tipo:

$$\phi_n = \text{absln}(\text{abs}(\phi_n))$$

Este conjunto de momentos resulta ser invariante a la traslación, la rotación el cambio de escala (Hu 1962), en la figura 3.9 se muestra un sistema reconecedor de objetos que discrimina 5 diferentes tipos de objetos.

La cantidad de objetos en la imagen, el ángulo de rotación y la posición del objeto en la imagen son aleatorios, se extraen rasgos geométricos obtenidos del cálculo de los momentos invariantes Hu para lograr el reconocimiento de todos los objetos del banco de 114 imágenes.

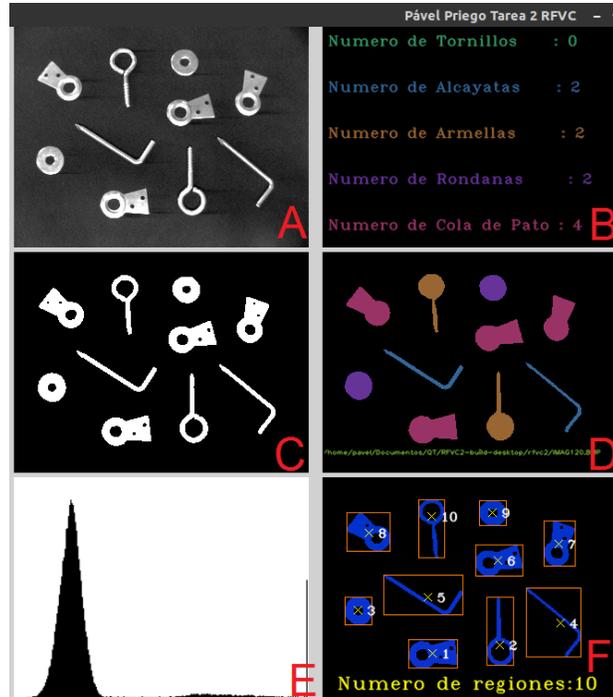


Figura 3.9 Ejemplo de algunos rasgos extraídos de los momentos invariantes de Hu

Las imágenes de la figura 3.9 son las siguientes:

1. La Imagen A muestra la imagen original.
2. La Imagen B muestra cuales y cuantos son los objetos reconocidos.
3. La Imagen C muestra la imagen resultado del segmentado con el método de umbralado de Kittler e Illinworth.
4. La imagen D muestra los objetos coloreados con un pseudocolor para cada clase.
5. La imagen E muestra el histograma de la imagen.
6. La Imagen F muestra el resultado del conteo de regiones, el centro de gravedad de cada región, calculado con los momentos de Hu y muestra el número de etiqueta de cada región.

3.6.4 Sesgo Proyectivo

Los momentos H_u de orden tres, describen el sesgo de las proyecciones del objeto [13]. Los coeficientes de sesgo de las proyecciones de un objeto sobre los ejes x , y ; calculándose con la fórmula:

$$Sk_x = \frac{\mu_{30}}{\mu_{20}^{3/2}} \quad Sk_y = \frac{\mu_{03}}{\mu_{02}^{3/2}}$$

Los signos de los coeficientes indican qué lado de una proyección se encuentra sesgada respecto a un eje, como se indica a continuación:

Si Sk_x Bias proyectivo x es:

- Positivo, el sesgo es a la izquierda del eje y .
- 0, el sesgo es simétrico alrededor de y .
- Negativo, el sesgo es a la derecha de y .

Si Sk_y Bias proyectivo y es:

- Positivo, el sesgo es abajo del eje x .
- 0, el sesgo es simétrico alrededor de x .
- Negativo, el sesgo es arriba de x .

3.6.5 Excentricidad

Los ejes mayor y menor de una región se definen en términos de su frontera, y son útiles para obtener la orientación del objeto. El cociente de las longitudes de estos ejes, llamado excentricidad de la región, es también un descriptor global importante de la forma del objeto [13] [23] [24]. Este parámetro determina el grado de desviación de una sección cónica con respecto a una circunferencia como se muestra en la figura 3.10.

Las longitudes del semieje mayor y semieje menor de la elipse circundante de un objeto pueden ser calculados como:

$$(a,b) = \frac{2[\mu_{20} + \mu_{02} \pm (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2]}{\mu_{00}}$$

En términos de estos valores, se puede calcular el llamado factor de excentricidad de tal elipse como:

$$\varepsilon = \frac{a^2 + b^2}{a^2}$$

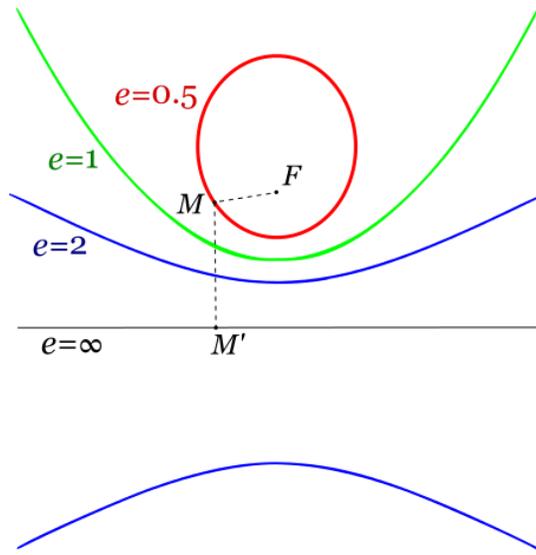


Figura 3.10 Secciones cónicas y excentricidad

3.6.5.1 Valores de excentricidad en secciones cónicas

La excentricidad de una circunferencia es cero ($\varepsilon = 0$).

La excentricidad de una elipse es mayor que cero y menor que 1 ($0 < \varepsilon < 1$).

La excentricidad de una parábola es uno ($\varepsilon = 1$).

La excentricidad de una hipérbola es mayor que 1 ($\varepsilon > 1$).

3.7 Redondez

Un objeto redondo perfecto tendrá una redondez de 1, mientras que los objetos rectos y alargados se acercarán a cero. Este rasgo se obtiene de la relación entre el área calculada y el área de un círculo.

$$\text{Redondez} = \frac{4 \cdot \pi \cdot \text{área}}{\text{Perímetro}^2}$$

3.8 Cláusulas de Horn

Dada una cláusula cualquiera, $L_1 \vee L_2 \vee \dots \vee L_n$, podemos transformarla en una sentencia equivalente con forma de condicional. Para ello, escribimos primero los literales negativos y luego los positivos:

$$\neg p_1 \vee \neg p_2 \vee \dots \vee \neg p_k \vee q_1 \vee q_2 \vee \dots \vee q_m$$

Por una de las leyes de D' Morgan, esta sentencia es equivalente a:

$$\neg(p_1 \wedge p_2 \wedge \dots \wedge p_k) \vee q_1 \vee q_2 \vee \dots \vee q_m$$

y usando otra equivalencia, la sentencia anterior equivale a:

$$(p_1 \wedge p_2 \wedge \dots \wedge p_k) \Rightarrow (q_1 \vee q_2 \vee \dots \vee q_m)$$

Hay de dos tipos: las que tienen un literal positivo (con cabeza) y las que no lo tienen.

Las cláusulas determinadas, o cláusulas de Horn *con cabeza* son las que sólo tienen un literal positivo:

$$\neg(p_1 \vee \neg p_2 \vee \dots \vee \neg p_k \vee q) = (p_1 \wedge p_2 \wedge \dots \wedge p_k \Rightarrow q)$$

Caso particular son las que no tienen más que ese literal positivo, que representan hechos, es decir, conocimiento factual. Los objetivos determinados, o *cláusulas de Horn sin cabeza* son las que no tienen ningún literal positivo:

$$(\neg p_1 \vee \neg p_2 \vee \dots \vee \neg p_k) \equiv \neg(p_1 \wedge p_2 \wedge \dots \wedge p_k)$$

En este caso, las fórmulas atómicas p_1, p_2, \dots, p_k son inconsistentes, es decir no es posible que sean todas verdaderas. Proviene de la negación de conclusiones que el sistema deductivo debe obtener mediante resolución y refutación.

3.9 Percepción de profundidad

A través de la visión binocular somos capaces de interactuar en un mundo tridimensional, pues podemos apreciar las distancias y el volumen de los objetos en el entorno que nos rodea.

Nuestros ojos, debido a su separación, obtienen dos imágenes con pequeñas diferencias entre ellas, conocidas como disparidad. Nuestro cerebro procesa las diferencias entre ambas imágenes y las interpreta de tal forma que percibimos la sensación de profundidad, lejanía o cercanía de los objetos que nos rodean. Este proceso se denomina estereopsis [22].

3.9.1 Mapa profundidad Kinect

El Kinect es un dispositivo capaz de medir la distancia que existe entre la cámara que percibe los rayos infrarrojos que el dispositivo emite y su entorno lo que crea una nube de puntos (ver figura 3.11) que son capturados y procesados internamente por el dispositivo Kinect produciendo una matriz de valores fraccionarios que se conoce como mapa de profundidad.

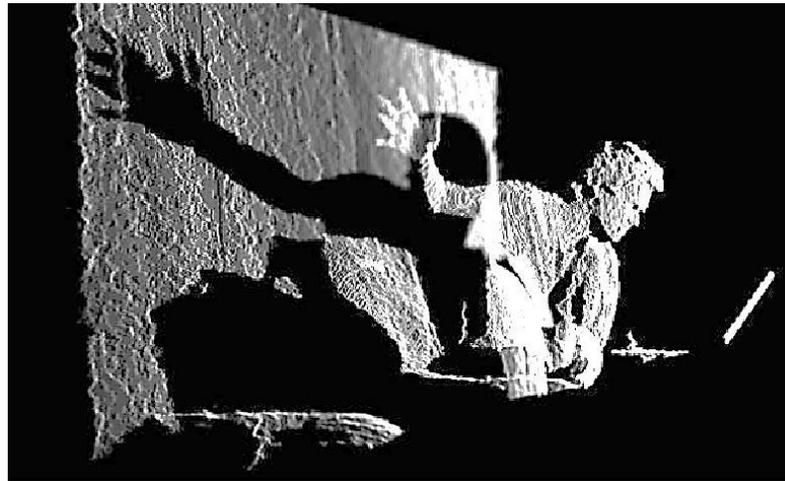


Figura 3.11 Nube de puntos

Esta matriz no tiene una representación visual directa al no pertenecer a los modelos de color visibles por el ojo humano, por lo que para poder tener una representación visual en pseudocolor, las distancias son transformadas a un modelo de color generalmente en escala de grises como se puede observar en la figura 3.12.

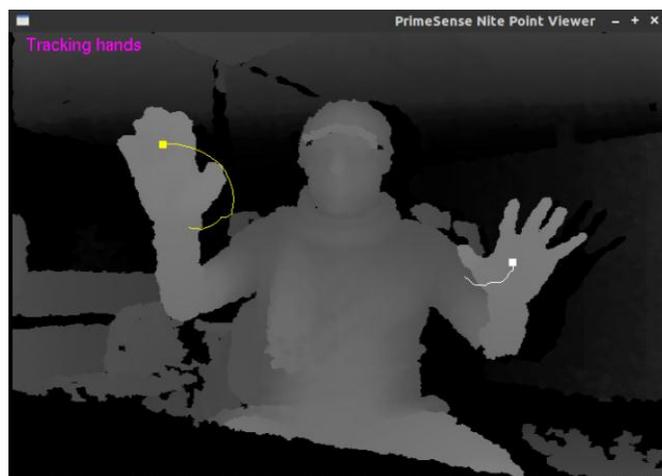


Figura 3.12 Imagen de profundidad en pseudocolor

3.10 Matriz Evolutiva

Una Matriz Evolutiva [6] [7] [8] [9] es un modelo con el que se representa una red neuronal artificial de una capa, en el que cada neurona recibe todos los pixeles x_i que forman una imagen binaria $X = (x_1, x_2, \dots, x_n)$. (figura 3.13). Cada neurona n_j es similar al modelo de neurona de Mc Culloc y Pitts.

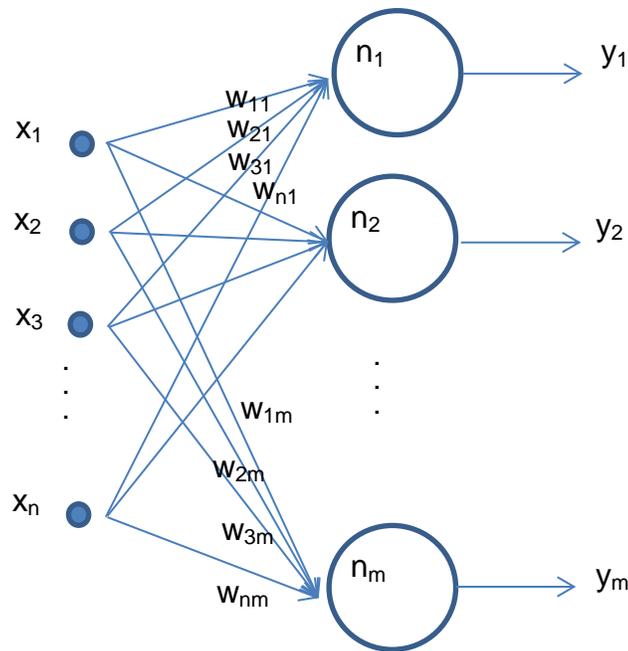


Figura 3.13 Modelo de red neuronal de una capa

donde:

x_i : pixel i de una imagen de entrada

w_{ij} : peso sináptico de la entrada del pixel i a la neurona j

n_j : neurona j

y_j : salida de la neurona j

Cada neurona n_j tiene una combinación característica de pesos sinápticos $W = (w_{1j}, w_{2j}, \dots, w_{nj})$. Con la sumatoria de los productos $w_{ij} * x_i$ se obtiene un valor que se le denomina semejanza de la entrada X con el patrón representado en la neurona n_j . El valor obtenido de la semejanza se compara con un umbral θ que se determina con base en la aplicación. Si en varias neuronas se obtiene un valor y_j que alcanza o supera el umbral θ definido, la decisión se obtiene con el máximo y_j obtenido de entre los y_j candidatos. Si ninguna de las neuronas sirve para identificar la imagen de entrada, entonces se crea una nueva neurona con sus pesos sinápticos iguales a la entrada que hasta el momento fue desconocida.

El modelo de neurona de la figura 3.11 se puede representar mediante la matriz evolutiva de la figura 3.14.

$$\begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1n} & h_1 \\ w_{21} & w_{22} & \cdots & w_{2n} & h_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ w_{m1} & w_{m2} & \cdots & w_{mn} & h_m \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ -1 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

Figura 3.14 Representación matricial del modelo de la figura 3.13

Cada fila de la matriz representa el conjunto de los pesos sinápticos w_{ij} de cada neurona n_j especializada para reconocer un patrón. Cada columna es un peso asociado para una dendrita o entrada específica x_i . X representa el conjunto de entradas y el cálculo de la salida del soma de cada neurona se expresa en el vector columna siguiente:

$$\begin{pmatrix} y_1 = \left(\sum_{i=1}^n x_i \cdot w_{1i} \right) - h_1 \\ y_2 = \left(\sum_{i=1}^n x_i \cdot w_{2i} \right) - h_2 \\ \vdots \\ y_m = \left(\sum_{i=1}^n x_i \cdot w_{mi} \right) - h_m \end{pmatrix}$$

4. Desarrollo de los métodos propuestos

En esta tesis se probaron dos métodos para solucionar el problema del reconocimiento de las imágenes del alfabeto del Lenguaje de Señas Mexicano. El primero de ellos, se describe en la sección 4.2 y se enfoca en el cálculo de invariantes para identificar las imágenes y el segundo método, descrito en la sección 4.4, utiliza una matriz evolutiva en el proceso de identificación de las señas.

4.1 Características del primer método de solución propuesto

Las características del primer método de solución propuesto son las siguientes.

1. Enfocado en invariantes a transformaciones.
2. Uso de una cámara digital de 8 megapíxeles.
3. Condiciones de iluminación controladas.
4. La escena contiene fondo negro.
5. Uso de imágenes estáticas.
6. Las letras del alfabeto son ejecutadas con la mano derecha.

4.2 Arquitectura del primer método de solución propuesto

El primer método de solución propuesto se compone de siete etapas, las cuales se muestran en la figura 4.1.

4.2.1 Captura de la imagen

Se realiza mediante una cámara digital integrada a un Smartphone Sony de 8 mega píxeles; el formato de la imagen capturada es jpg de color, la iluminación es proporcionada por la lámpara LED propia del dispositivo, la imagen de salida es convertida a una matriz de tres canales de color azul, rojo y verde con una profundidad de 255 niveles en cada color.

Las imágenes capturadas mostradas en la figura 4.2 tienen rotaciones naturales realizadas por el posicionamiento del individuo, es decir, no se rotó la imagen artificialmente; la distancia entre el objeto y la cámara varía entre 30 centímetros a 1 metro; la posición de las manos en las imágenes son aleatorias; las imágenes pertenecen al autor y a una mujer adulta con diferente altura, complexión, tamaño de la mano y color de piel.



Figura 4.1 Etapas del reconocedor del alfabeto del LSM del primer método de solución propuesto

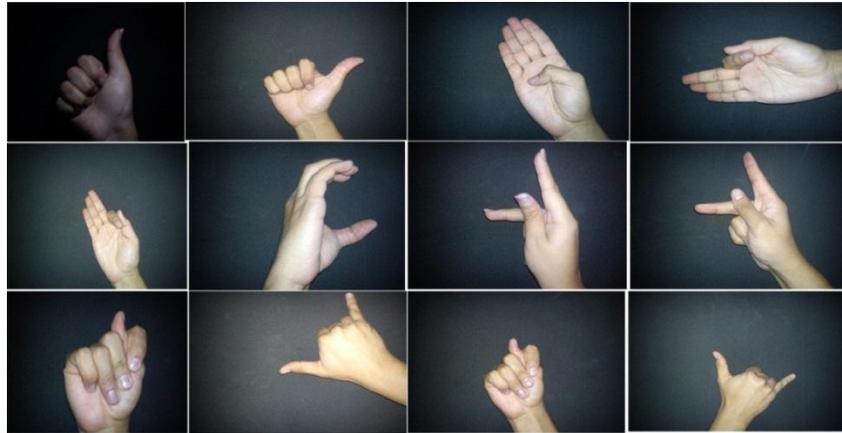


Figura 4.2 Imágenes de entrada del primer método de solución propuesto

4.2.2 Corrección de la iluminación

La corrección de la iluminación [28], se realiza mediante un proceso morfológico de sombrero de copa blanco (WTH). Este proceso se usa para atenuar objetos claros sobre fondos oscuros y requiere de un elemento de estructura acorde a la distribución de iluminación. Para este proceso se utilizó un elemento de estructura de forma cuadrada y de un tamaño grande con respecto a la imagen que se procesa, es decir, el tamaño es del 20% con respecto al tamaño de la imagen original.

El resultado parcial de la operación WTH es una imagen con bloques claros correspondientes al gradiente de iluminación. La imagen original es mejorada mediante la resta de la imagen de apertura, permitiendo que las zonas blancas de la imagen se atenúen, de tal forma que no dificulten procesos de segmentado posteriores.

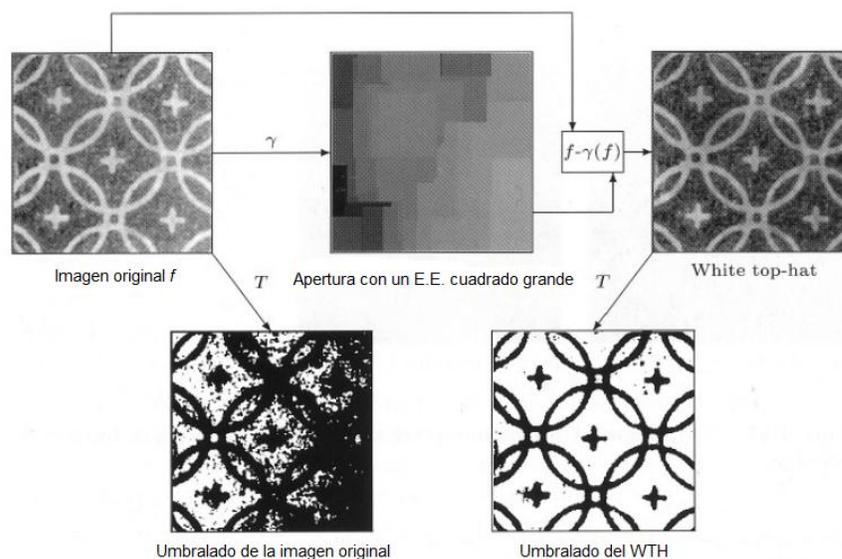


Figura 4.3 Corrección de la iluminación

Esta operación en imágenes en colores requiere que primero cada canal de color de la matriz sea separado en sus respectivos canales RGB y posteriormente hacer la operación para cada canal de color, el resultado de la operación de cada canal de color de la imagen se reintegra en una imagen en colores RGB.

4.2.3 Segmentación

Se realizaron pruebas con tres distintos algoritmos para binarizar imágenes: el método de Otsu, el método de Kittler e Illinworth y el método de Sahoo [13] [31] como se muestra en la figura 4.4.

Los resultados con varias de las imágenes capturadas hicieron evidente, que el método de Kittler e Illinworth realiza la segmentación del objeto de interés con menor ruido y resalta más las características de la forma de la mano, por lo que este método es el que se utiliza para el segmentado.

Las imágenes mostradas en la figura 4.4, son las siguientes:

1. La Imagen A muestra la imagen original.
2. La imagen B muestra el histograma de la imagen.
3. La Imagen C muestra la imagen transformada a escala de grises de la cual se parte para realizar el umbralado.
4. Imagen binarizada por el método de Otsu.
5. Imagen binarizada por el método de Kittler e Illinworth.
6. Imagen binarizada por el método de Sahoo.

4.2.4 Filtro morfológico

La operación apertura remueve las estructuras brillantes de la imagen, mientras que la operación de clausura, tiene el mismo efecto de filtrado pero afecta a las estructuras oscuras de la imagen. Si la imagen está corrompida por ruido de funciones simétricas, se puede usar una combinación secuencial como puede ser una operación de apertura seguido por la operación clausura dual o viceversa; la elección de qué operación se realiza primero depende de los objetos de la imagen que deben ser extraídos. En la figura 4.5 se observan algunos artefactos presentes en la imagen que son removidos eficientemente por un filtro secuencial.

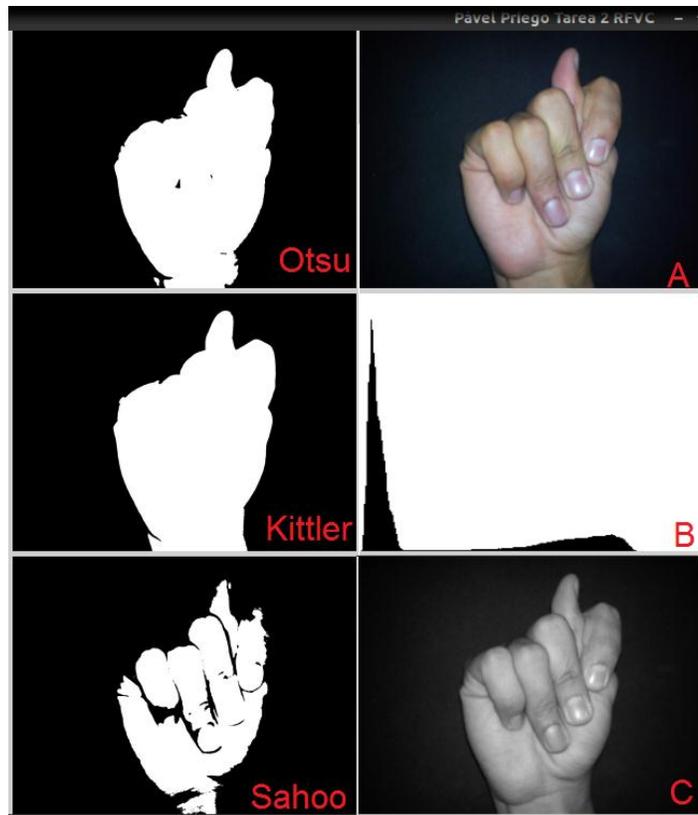


Figura 4.4 Métodos de umbralado probados en el primer método propuesto

Debido al tamaño la imagen procesada el elemento de estructura elegido es de tamaño 40 x 40, denotado por $\phi_{40}[\gamma_{40}(f)]$; después de realizar la operación filtro secuencial de apertura - clausura se eliminan los artefactos (figura 4.5) y se suavizan los bordes del objeto.

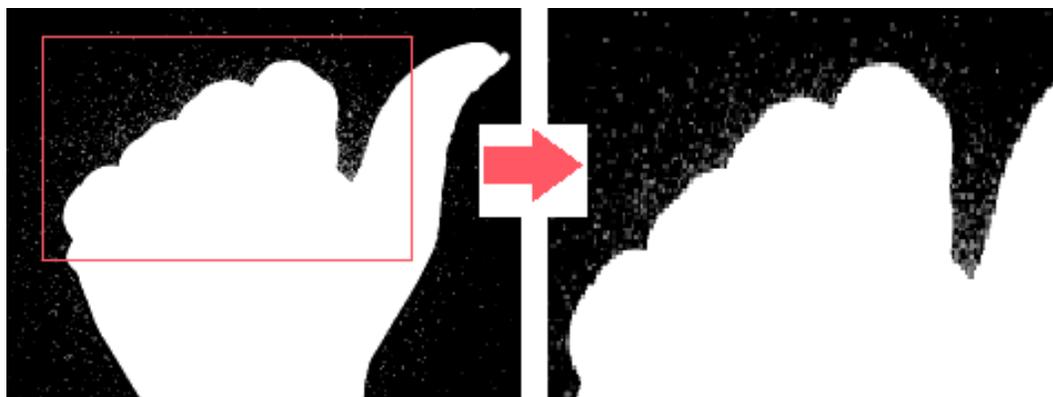


Figura 4.5 Imagen binaria segmentada con objetos indeseables

4.2.5 Clasificación y reconocimiento

Los valores representativos de cada letra se obtienen mediante los rangos de los valores de cada rasgo, como se observa en la figura 4.6. En el caso del ángulo se tiene en cuenta la ambigüedad al usar los momentos de Hu para el cálculo; se realiza el cálculo del sesgo proyectivo como consideración adicional.

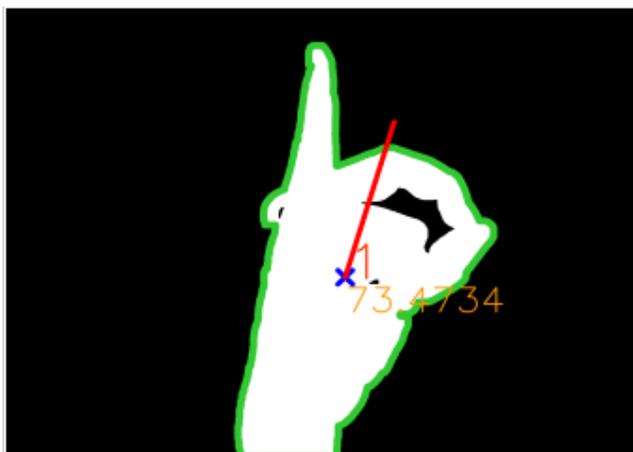


Figura 4.6 Datos visuales de la forma y orientación de la mano

La clasificación se realiza mediante un Sistema Experto; las reglas de este sistema están en la forma normal conjuntiva o cláusulas de Horn.

4.2.6 Presentación

La figura 4.7 muestra el despliegue en pantalla del significado de la seña reconocida la cual se muestra como un texto. En la interfaz se visualiza la imagen original, la imagen binaria segmentada por el método de umbralado de Kittler e Illinworth y los rasgos visuales del contorno, el ángulo y el centro de gravedad del objeto.

4.3 Características del segundo método de solución propuesto

Las características del segundo método de solución propuesto en esta tesis son:

1. Enfoque en tolerancias a transformaciones.
2. Alta tolerancia a los cambios de luz por el uso de cámara infrarroja.
3. Uso de mapas de profundidad.
4. Uso de secuencia de imágenes.
5. Proceso de reforzamiento.
6. Proceso de olvido.

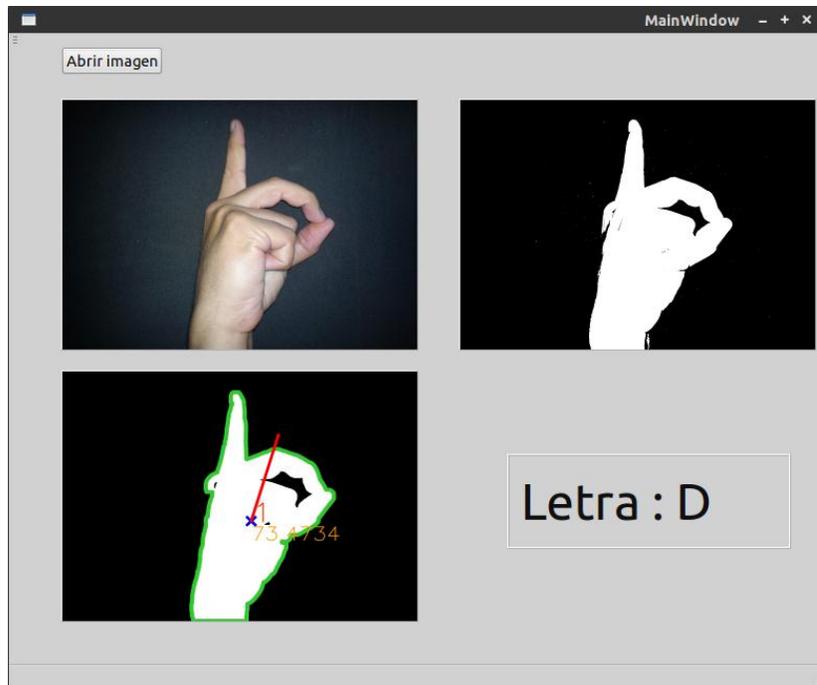


Figura 4.7 Presentación del resultado del primer método propuesto

4.4 Arquitectura del segundo método de solución propuesto

El segundo método de solución propuesto consta de siete etapas mostradas en la figura 4.8.

4.4.1 Captura de la imagen

Se obtiene una secuencia de mapas de profundidad de 640 por 480 píxeles del dispositivo Microsoft Kinect; la imagen es producida por la reconstrucción de la nube de puntos del emisor de infrarrojos del dispositivo; los valores de cada píxel representan distancias en milímetros.

El dispositivo es capaz de capturar 30 imágenes por segundo con una resolución de 640 x 480 píxeles por cada imagen. Se obtienen dos Matrices del dispositivo en cada muestreo, una de ellas es una imagen RGB y la otra es un mapa de profundidad el cual sólo contiene datos de distancia en cada píxel de la imagen.

Una imagen en pseudocolor del mapa de profundidad se puede apreciar en la figura 4.9.

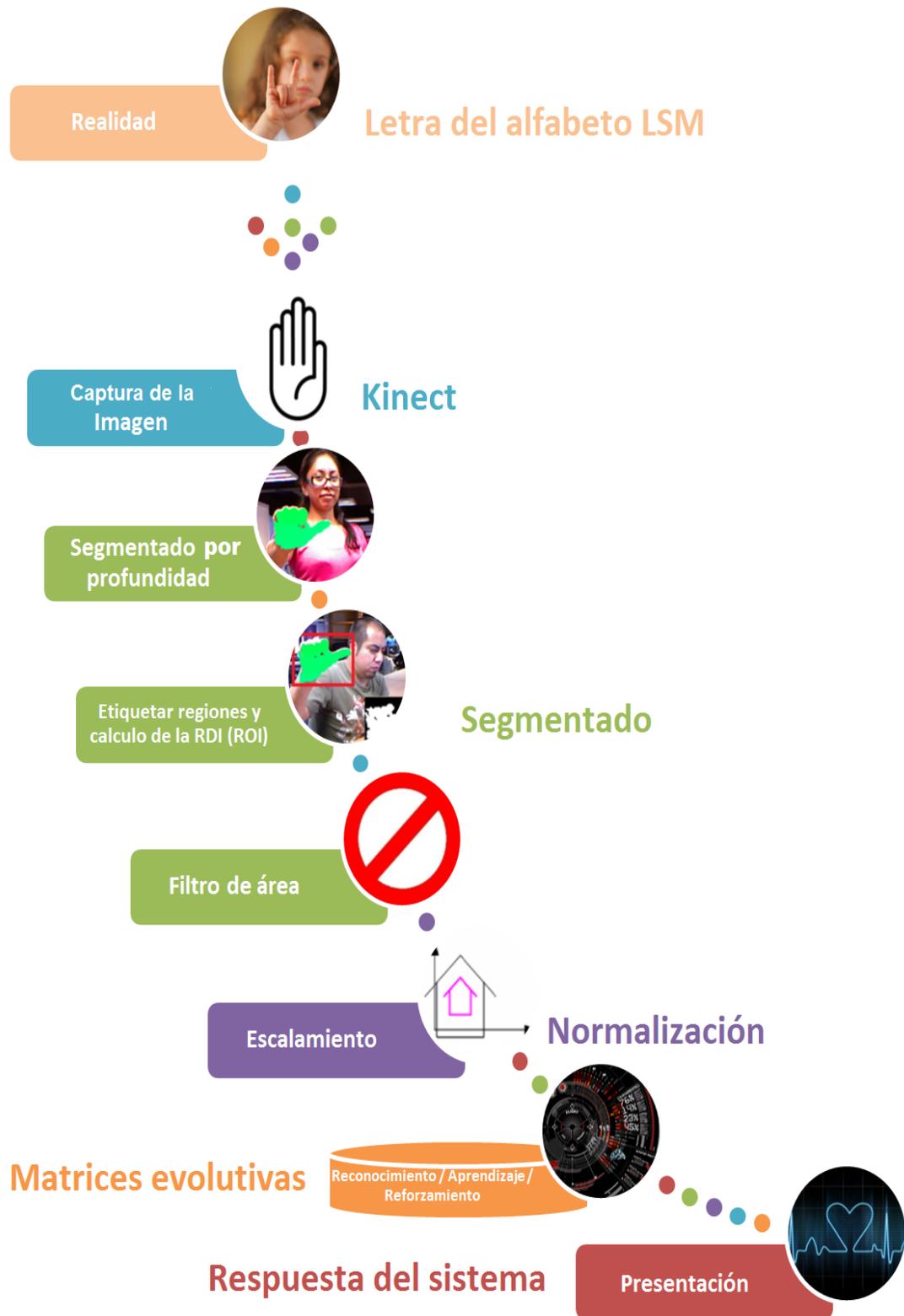


Figura 4.8 Etapas del reconocedor del alfabeto del LSM del segundo método de solución propuesto

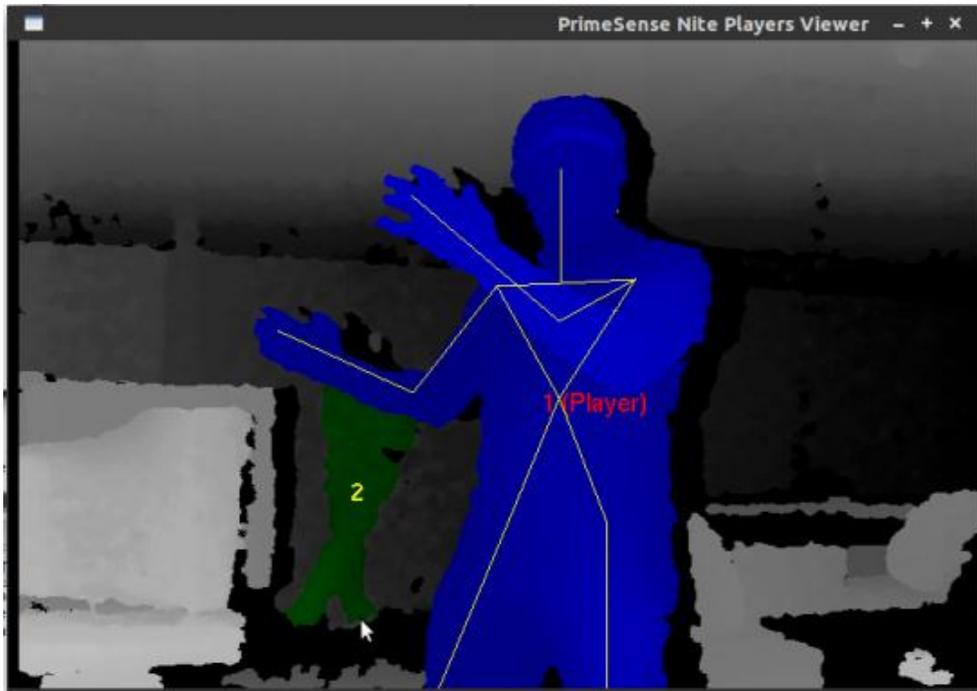


Figura 4.9 Imagen en pseudocolor de un mapa de profundidad

4.4.2 Segmentación de la mano

La imagen obtenida con la cámara infrarroja del dispositivo Kinect es usada para realizar un segmentado basado en la distancia que existe entre el objeto y el dispositivo. En este proceso se crea un umbral de reconocimiento de 200 mm en donde se espera encontrar el objeto de interés y aislarlo del fondo. El umbral requiere dos parámetros, la distancia frontal y la distancia del fondo.

La distancia frontal es aquella situada entre el objeto y el dispositivo Kinect. La distancia que está por delante del objeto se discrimina.

La distancia de fondo es aquella situada atrás del objeto, es decir, aquello ubicado por detrás del objeto, se le considera fondo y se discrimina.

Estos dos parámetros pueden modificarse por el usuario usando las barras de desplazamiento mostradas en la parte inferior de la figura 4.10. En este trabajo, la distancia frontal calculada con base en experimentos con los que se puede capturar mejor una mano es de 400mm. La distancia del fondo se encontró adecuada en 600mm. El umbral de captura es de 200mm.

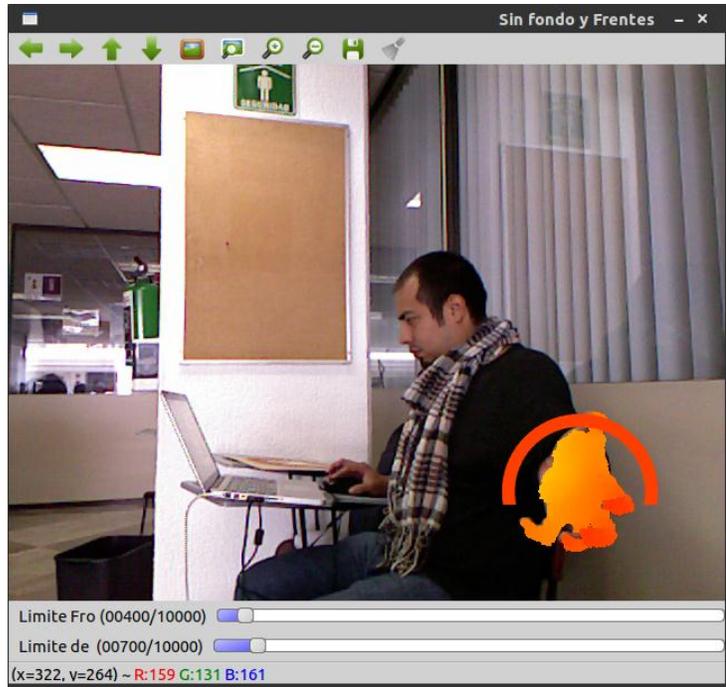


Figura 4.10 Segmentado por profundidad

La cámara infrarroja permite una tolerancia a la luz indirecta del sol y a condiciones de oscuridad, el resultado de la etapa de segmentado por profundidad se muestra en la figura 4.11. Las barras de desplazamiento en la parte inferior de la imagen el límite donde se inicia y superior rango de distancia donde el objeto (mostrado en tono naranja figura 4.10) en esa zona se representa sobre la imagen RGB con el fin de ayudar al usuario a identificar las señas que se ejecutan.

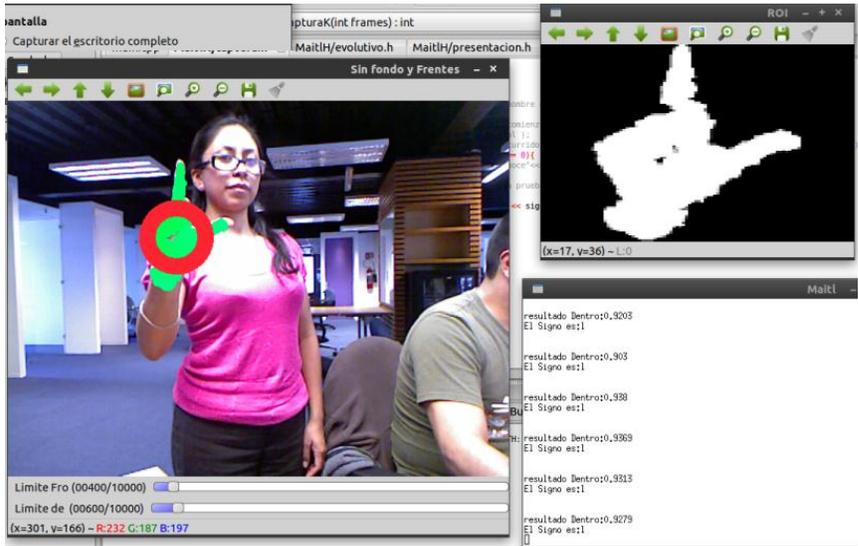


Figura 4.11 Visualización del usuario de la zona segmentada

El segmentado por profundidad representa una ventaja con respecto a transformaciones de modelos de color, por ejemplo del modelo de color RGB al HSV, usado frecuentemente para facilitar la segmentación de tonalidades de color como es el color de la piel y permitiendo una mejor tolerancia a los cambios de luminosidad.

La segunda etapa de segmentación consiste en etiquetar todos los objetos encontrados. En la ventana de trabajo para cada una de estos objetos candidatos se calcula su área y el paralelogramo mínimo que contiene al objeto, con el fin de crear una región de interés y consumir menos recursos de cómputo.

Si ningún objeto se encuentra en la imagen, no se realizan cálculos. Solamente se realizan cálculos hasta encontrar uno o más objetos candidatos. La imagen resultante es una imagen binaria, donde el nivel del fondo es 0 y los píxeles que contienen información tienen valor de 1.

La etapa final de segmentación se realiza mediante un filtro de área, discriminando objetos con un tamaño menor a 100 por 100 píxeles o un tamaño mayor a 400 por 400 píxeles. En la figura 4.12 se muestra una imagen de una seña obtenida en este módulo. La imagen presenta un objeto 4-conectado y por lo tanto, no existe ruido o artefactos que no estén conectados a la región.

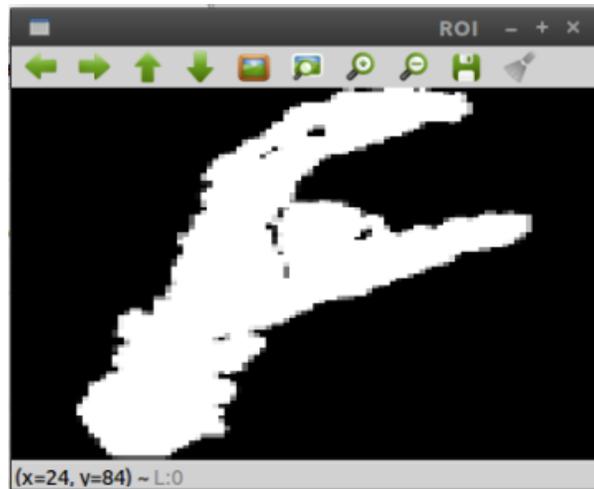


Figura 4.12 Imagen binaria del objeto de interés

4.4.3 Escalamiento de la mano

Las imágenes filtradas pueden tener formas con una elongación mayor horizontalmente o verticalmente. Para corregir las elongaciones, en este módulo se transforma la imagen a una forma cuadrada, considerando el máximo del tamaño del objeto en el eje horizontal x o en el eje vertical y. La

figura cuadrada siempre tiene tamaño de lado $\max\{X, Y\}$ y en esta etapa se ajustan: la forma, el tamaño y la posición del objeto con el objetivo de minimizar la variación entre imágenes de la misma clase. Este proceso se realiza mediante un ajuste a un tamaño de patrón definido de 200 x 200 píxeles.

Cuando el paralelogramo tiene forma rectangular, se hace un ajuste usando el máximo entre el tamaño X y el de Y . Con este dato se definen las nuevas dimensiones del objeto. Con la diferencia entre el mínimo de X o Y se completa con píxeles de fondo la parte más angosta del rectángulo en ambos extremos del paralelogramo mínimo que envuelve al objeto. Como consecuencia el objeto se centra en el nuevo paralelogramo cuadrado.

El paralelogramo obtenido se escala al tamaño del patrón definido con el fin de minimizar la variación de tamaño de las manos.

4.4.4 Matrices evolutivas

En esta tesis se usa una matriz evolutiva (figura 4.13) para almacenar los patrones del conjunto de letras del alfabeto del LSM y hacer posible su identificación mediante un programa de computadora.

	X_1	X_2	X_3	X_4	X_5	...	X_n	h	Significado
Patrón 1	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	...	X_{1n}	h_1	"objeto β "
Patrón 2	X_{21}	X_{22}	X_{23}	X_{24}	X_{25}	...	X_{2n}	h_2	"objeto α "
Patrón 3	X_{31}	X_{32}	X_{33}	X_{34}	X_{35}	...	X_{3n}	h_3	"objeto β "
...						...			
Patrón m	X_{m1}	X_{m2}	X_{m3}	X_{m4}	X_{m5}	...	X_{mn}	H_m	"objeto μ "

Figura 4.13 Una matriz evolutiva.

donde:

x_{ij} : Es un elemento del patrón almacenado en la matriz evolutiva (columnas).

h : *Factor de coincidencia*, es el número de coincidencias o reforzamientos del patrón j .

n : Es el número de elementos de cada patrón (X) de la matriz evolutiva.

m : Es el número de patrones almacenados en la matriz (filas).

Significado: Es el significado de cada patrón (fila). En este modelo dos patrones pueden tener el mismo significado.

Los procesos de la matriz evolutiva son:

1. Proceso de aprendizaje - reconocimiento.
 - a. Sin reforzamiento.
 - b. Con reforzamiento.
2. Proceso de reforzamiento.
3. Proceso de olvido.

4.4.4.1 Proceso de aprendizaje-reconocimiento sin reforzamiento

La matriz evolutiva sin reforzamiento tiene la particularidad de no incluir el factor de coincidencia h en sus elementos, como se observa en la figura 4.14.

X_1	X_2	X_3	X_4	X_5	...	X_n	Significado
0	0	0	0	0	...	0	A
1	1	1	1	0	...	0	U
1	1	0	0	0	...	1	K

Figura 4.14 Una matriz evolutiva sin reforzamiento

La figura 4.15 presenta de forma gráfica una matriz evolutiva sin reforzamiento.

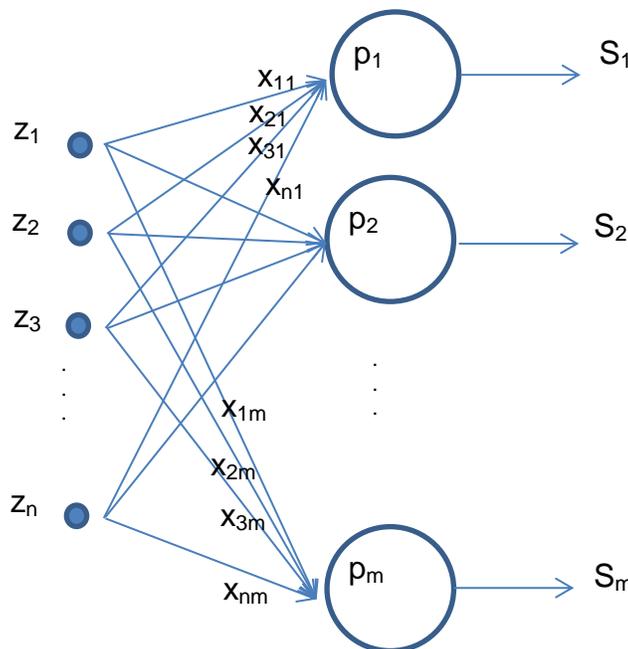


Figura 4.15 Representación gráfica de una matriz evolutiva de patrones sin reforzamiento

donde:

Z_i Es un pixel de la imagen de entrada.

X_{ij} Es un elemento del patrón almacenado en la matriz evolutiva.

P_j Es la neurona especializada para reconocer el patrón j de la matriz evolutiva.

S_j Es la respuesta de la neurona j .

n : es el número de elementos que tiene el vector de entrada (Z) y coincide con el número de elementos de cada patrón (X) de la matriz evolutiva.

m : es el número de patrones de la matriz evolutiva.

4.4.4.11 Proceso de reconocimiento de la matriz evolutiva sin reforzamiento.

Para efectuar el proceso de reconocimiento entre el vector de entrada y el patrón almacenado en la base de datos se utiliza un criterio de semejanza. La distancia utilizada para obtener el factor de semejanza entre patrones α de la matriz evolutiva sin reforzamiento es el siguiente:

$$\alpha = \frac{\sum_{i=1}^n [1 - |Z_i - X_i|]}{n}$$

Si se requiere saber el porcentaje de semejanza se calcularía de la siguiente forma:

$$\text{Porcentaje de semejanza} = \alpha \cdot 100\%$$

Para calcular el factor de semejanza cuando se tienen varios patrones, se usa el factor α_j , expresado con la siguiente ecuación:

$$\alpha_j = y_j = \frac{\sum_{i=1}^n [1 - |Z_i - X_{ij}|]}{n}$$

La respuesta $\bar{\alpha}$ es el vector de semejanzas de la matriz evolutiva; con m patrones se obtiene:

$$\bar{\alpha} = \begin{pmatrix} y_1 = \frac{\sum_{i=1}^n [1 - |Z_i - X_{i1}|]}{n} \\ y_2 = \frac{\sum_{i=1}^n [1 - |Z_i - X_{i2}|]}{n} \\ \vdots \\ y_m = \frac{\sum_{i=1}^n [1 - |Z_i - X_{im}|]}{n} \end{pmatrix}$$

El criterio de decisión para identificar el patrón más semejante en la matriz evolutiva sin reforzamiento es:

$$f(\bar{\alpha}) = \begin{cases} \text{Verdadero, si } \arg \max(\bar{\alpha}) > \theta \\ \text{Falso, si } \arg \max(\bar{\alpha}) \leq \theta \end{cases}$$

donde θ es el valor del umbral de similitud o de semejanza establecido por el usuario.

4.4.4.12 Aprendizaje de la matriz evolutiva sin reforzamiento

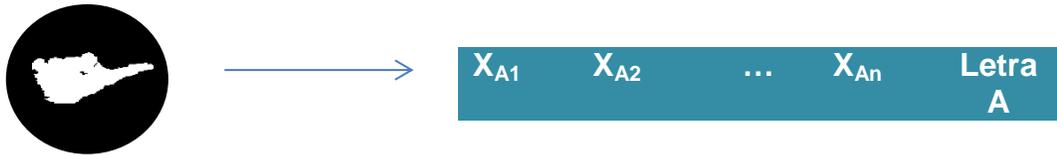
Si $\alpha \leq \theta$ se solicita al usuario que indique el significado del símbolo desconocido para anexar en la matriz evolutiva el patrón hasta ahora desconocido y su significado. Cada X_i del patrón nuevo se copia de cada Z_i del vector de entrada.

4.4.4.13 Ejemplo matriz evolutiva sin reforzamiento

Para ejemplificar el proceso de aprendizaje-reconocimiento suponemos que la matriz evolutiva no tiene elementos, entonces, cuando un nuevo objeto, en este caso es la letra A de la figura 4.16, se introduce, el patrón se transforma a un vector cuyo valor de cada elemento se almacena a la matriz evolutiva junto con el significado asignado por el usuario.

Si posteriormente, otra imagen de la letra A se introduce, se efectúa el proceso de reconocimiento. Si no se cumple el umbral de similitud entonces se añade el nuevo vector en la matriz evolutiva, obteniendo dos patrones de la letra A con el mismo significado (figura 4.16). Esto se muestra a continuación con base en dos eventos.

Primer evento: se procesa una seña de la letra A.



Segundo evento: se procesa otra seña de la letra A con una variación que no supera el umbral de similitud.



Como resultado de los eventos se almacenan dos patrones con el significado de la letra A en la matriz evolutiva.

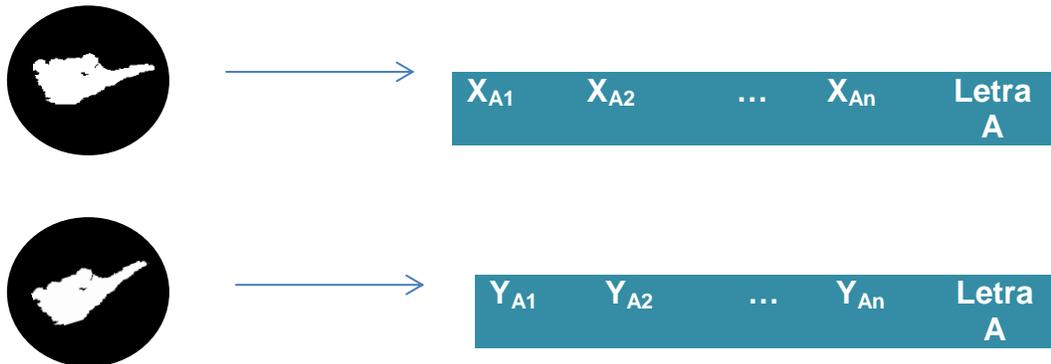


Figura 4.16 Resultado del proceso de aprendizaje en una matriz evolutiva sin reforzamiento

Con este modelo, cuando una imagen no satisface el umbral de similitud con los patrones almacenados en la matriz evolutiva, se aprende el nuevo patrón y se agrega en la base de conocimiento, se podría llegar a tener varios patrones con el mismo significado (figura 4.16) pero con variaciones muy pequeñas en su estructura incrementando el cálculo en la comparación para hacer el reconocimiento de una letra.

4.4.4.2 Proceso de aprendizaje-reconocimiento con reforzamiento

La matriz evolutiva incluye el factor de coincidencia o de reforzamiento h ; los elementos de la una matriz de 25 elementos por patrón se pueden observar en la figura 4.17.

X_1	X_2	X_3	X_4	X_5	...	X_{25}	h	Significado
1	0	1	0	2	...	0	7	A
2	3	3	3	2	...	0	3	U
2	4	0	0	0	...	1	4	K

Figura 4.17 Una matriz evolutiva con reforzamiento

La figura 4.18 presenta de forma gráfica una matriz evolutiva con reforzamiento.

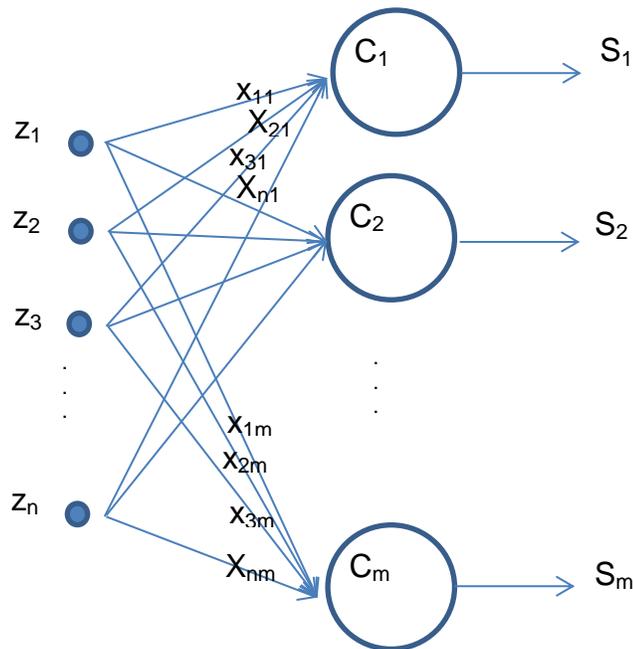


Figura 4.18 Representación gráfica de una matriz evolutiva de patrones con reforzamiento

donde:

Z_i Es un pixel de la imagen de entrada.

X_{ij} Es un elemento de la clase j almacenada en la matriz evolutiva.

C_j Es la neurona especializada para reconocer los patrones pertenecientes a la clase j de la matriz evolutiva.

S_j Es la respuesta de la neurona j .

n : es el número de elementos que tiene el vector de entrada (Z) y coincide con el número de elementos de cada patrón (X) de la matriz evolutiva.

m : es el número de neuronas o de clases de la matriz evolutiva.

4.4.4.21 Proceso de reconocimiento de la matriz evolutiva con reforzamiento.

Para poder efectuar el reconocimiento el vector de entrada se compara contra cada uno de las clases almacenados en la matriz evolutiva, figura 4.17, la cual consiste de filas de patrones almacenados con columnas de elementos (X_i) un umbral h y el significado de cada elemento de la clase de la misma. El umbral indica cuantos patrones han coincidido con el que se muestra ahí.

Para obtener el factor de semejanza de cada clase se usa la fórmula siguiente:

$$\alpha = \frac{\sum_{i=1}^n [1 - |Z_i - \frac{X_i}{h}|]}{n}$$

Para calcular el factor de semejanza cuando se tienen varias clases, se usa el factor α_j , expresado con la siguiente ecuación:

$$\alpha_j = y_j = \frac{\sum_{i=1}^n [1 - |Z_i - \frac{X_{ij}}{h_j}|]}{n}$$

Como se puede observar para poder comparar el vector de entrada con cada clase de la matriz evolutiva se efectúa la normalización siguiente:

$$\frac{X_{ij}}{h_j}$$

La respuesta $\bar{\alpha}$, es el conjunto de factores de semejanza obtenidos de la matriz evolutiva con m clases:

$$\bar{\alpha} = \begin{pmatrix} y_1 = \frac{\sum_{i=1}^n [1 - |Z_i - \frac{X_{i1}}{h_1}|]}{n} \\ y_2 = \frac{\sum_{i=1}^n [1 - |Z_i - \frac{X_{i2}}{h_2}|]}{n} \\ \vdots \\ y_m = \frac{\sum_{i=1}^n [1 - |Z_i - \frac{X_{im}}{h_m}|]}{n} \end{pmatrix}$$

El criterio de decisión para el reconocimiento en la matriz evolutiva con reforzamiento es:

$$f(\bar{\alpha}) = \begin{cases} \text{Verdadero, si } \arg \max(\bar{\alpha}) > \theta \\ \text{Falso, si } \arg \max(\bar{\alpha}) \leq \theta \end{cases}$$

donde θ es el valor del umbral de similitud o de semejanza establecido por el usuario.

4.4.4.2.2 Aprendizaje de la matriz evolutiva con reforzamiento

Si $\alpha \leq \theta$ se solicita al usuario que indique el significado del símbolo desconocido y se anexa en la matriz evolutiva, si lo adiciona, el significado, es digitado por el usuario y cada X_i del patrón nuevo se copia de cada Z_i del vector de entrada y se inicializa con valor uno el factor de coincidencia h .

4.4.4.3 Reforzamiento

Cuando se tienen pequeñas variaciones en un mismo patrón, en nuestro caso letras del alfabeto del LSM, se utiliza un proceso llamado reforzamiento con el fin de agrupar en una misma clase la información de varios patrones con pequeñas variaciones de similitud (figura 4.19), ya que tienen el mismo significado en la matriz evolutiva.

Cada uno de estos patrones cumple con el criterio de decisión del umbral de similitud; sin embargo las variaciones son capturadas debido al proceso de reforzamiento dando origen a clases de letras.

El reforzamiento se hace sumando el elemento X_i de la clase j de la matriz almacenada con el valor del elemento Z_i correspondiente a la imagen de entrada es decir $z_i + x_{ij}$ e incrementando en una unidad el valor de coincidencias h .

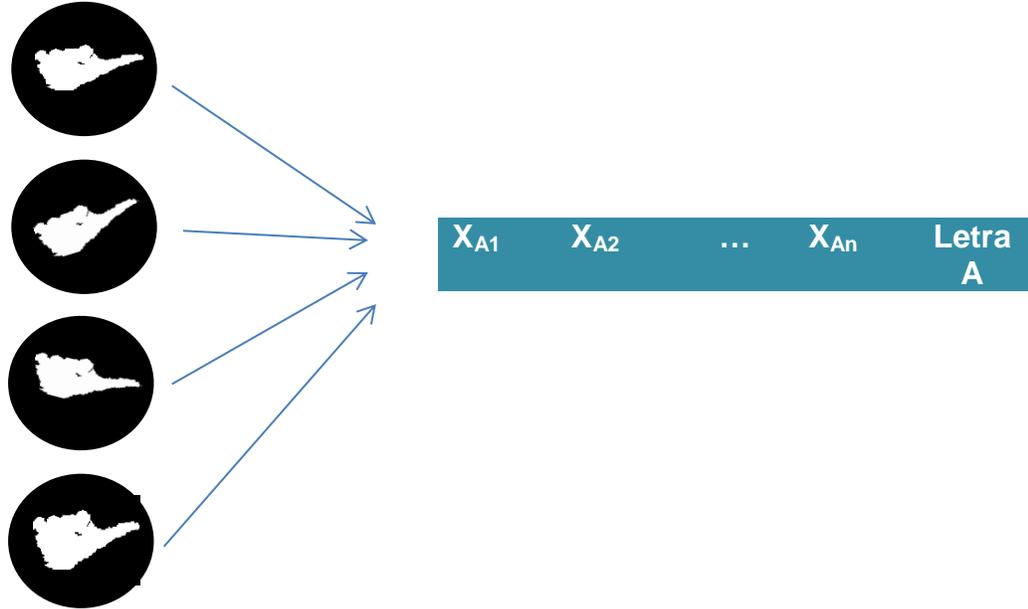


Figura 4.19 Representación gráfica de una matriz evolutiva de patrones reforzados

El proceso de reforzamiento en los patrones de las matrices evolutivas, causa que algunos elementos del vector que representa el patrón tengan valores más altos que otros, como se observa en la figura 4.20, generando zonas que se caracterizan de la siguiente manera:

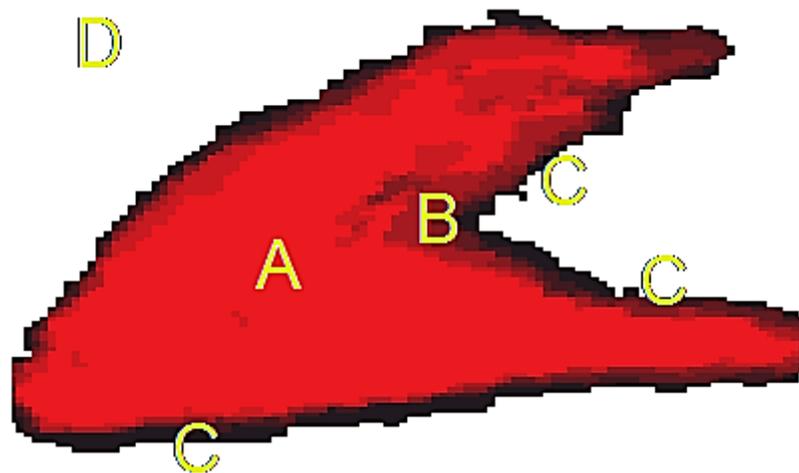


Figura 4.20 Un patrón reforzado

- A. Zona de mayor ocurrencia del patrón.
- B. Zona de tolerancia a transformaciones del objeto.
- C. Zona de ruido.
- D. Fondo.

4.4.4.4 Olvido

Después de un proceso de reconocimiento-reforzamiento, los patrones de la matriz evolutiva pueden tener un factor de coincidencias (h) con valores muy grandes, como 500 o mucho más; en estos casos lo que puede hacerse es restar un factor (por ejemplo 200); este factor puede ser el valor promedio de los elementos de la clase o ser elegido de acuerdo al problema por el usuario. Los elementos que resulten negativos se asignan con cero o con el valor mínimo de los niveles de la imagen; de esta forma algunos puntos que dentro del patrón se consideran poco relevantes (ver imagen C de la figura 4.21) se eliminan, quedando únicamente los puntos que se consideran relevantes del patrón.

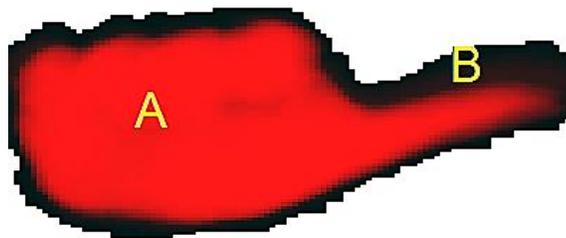


Figura 4.21 Un patrón evolucionado

4.4.4.5 Ejemplo del proceso de reconocimiento de una letra del LSM usando matrices evolutivas.

Los patrones de las clases almacenadas en la matriz evolutiva son vectores que en esta tesis tienen $200 \times 200 = 40,000$ elementos cada uno; se obtiene por colocar las filas de la matriz una enfrente de la otra. Esto se ejemplifica suponiendo que se tienen imágenes con 25 elementos, después de haber realizado el proceso de normalización. El ejemplo que se muestra es de una mano que representa la letra A del Lenguaje de Señas Mexicano, figura 4.23. Se usa un umbral de similitud $\theta = 95\%$.

4.4.4.5.1 Vector de entrada

Durante el proceso de aprendizaje-reconocimiento cada imagen es tratada como un vector de píxeles para ser comparada con los patrones

almacenados en la base de conocimiento. En la figura 4.22 se muestra en forma que se facilita la visualización al lector, un ejemplo de patrones de letras almacenados en la matriz evolutiva.

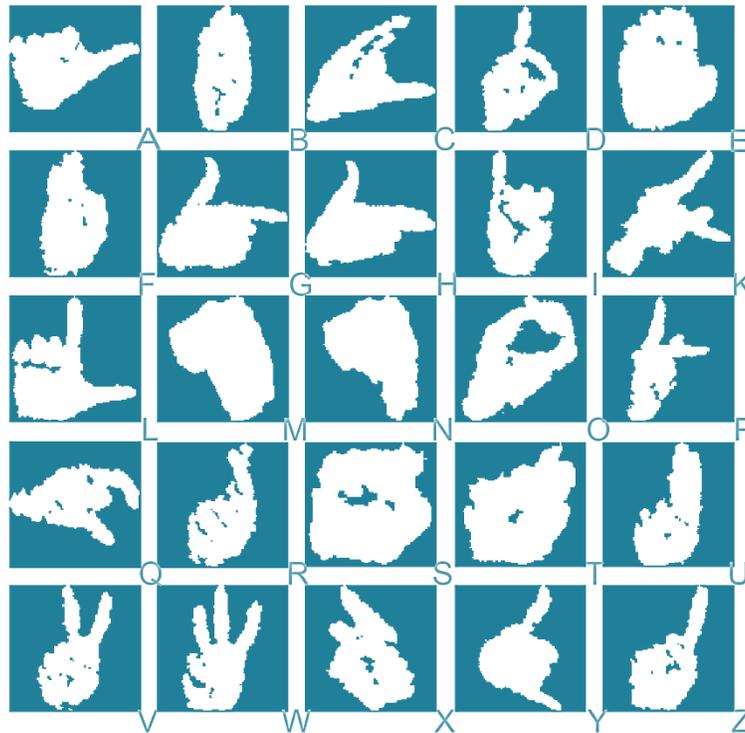


Figura 4.22 Ejemplo de patrones de letras almacenados en la base de datos

Para comparar la imagen con los patrones almacenados se ajusta la imagen de la figura 4.23 a un patrón de 5 x 5 píxeles como se observa en la figura 4.24.



Figura 4.23 Un patrón de la letra A

La imagen de entrada se normaliza para tener valores en el intervalo $[0,1]$; entonces si la celda tiene 50% de ocupación el nivel es 1, en otro caso el nivel es 0. En la figura 4.25 se muestran los cuadrantes considerados para

su transformación a la forma vectorial con unos y ceros mostrados en la figura 4.26.

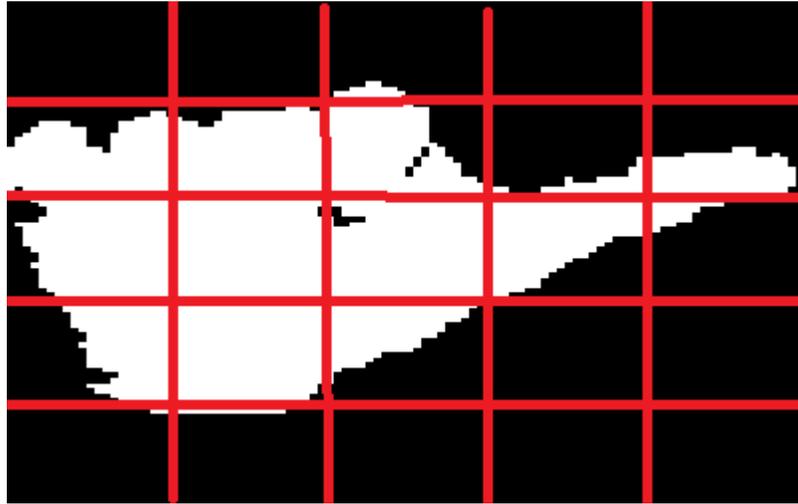


Figura 4.24 Ajuste del patrón de la letra A en una retícula de 5 x 5 celdas

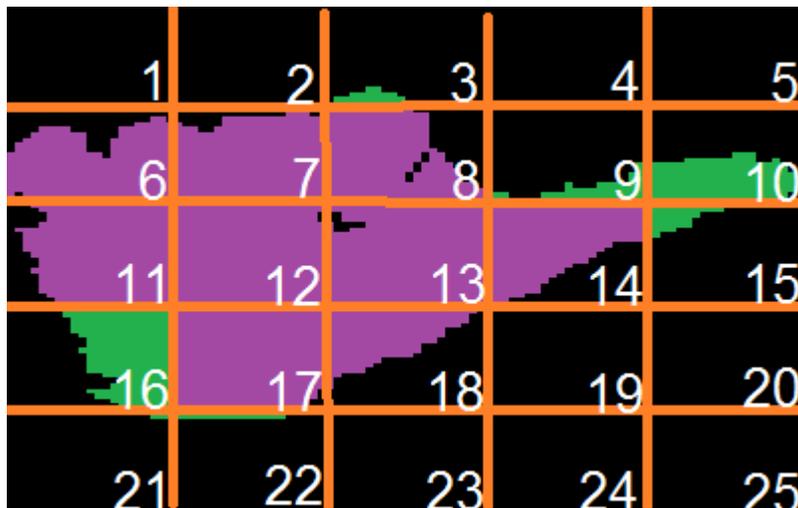


Figura 4.25 Discretización binaria de la letra A

La matriz se transforma a un vector colocando cada fila una enfrente de la otra como se muestra en la figura 4.26.



0 0 0 0 0 1 1 1 0 0 1 1 1 1 0 0 1 1 0 0 0 0 0 0 0

Figura 4.26 Vector del patrón de la letra A ajustada

4.4.4.5.2 Reconocimiento-aprendizaje del patrón

Un ejemplo del cálculo del reconocimiento-aprendizaje de una imagen es el siguiente: supóngase que se tiene en la entrada el vector del patrón A con elementos Z_i (figura 4.26) y en la matriz evolutiva se tiene un patrón digitalizado llamado A con elementos X_i (figura 4.27), un factor de coincidencias $h = 7$ y un umbral de similitud $\theta = 95\%$.



Figura 4.27 Un patrón de la matriz evolutiva reforzado

La semejanza entre la imagen y el patrón se calcula de la siguiente forma:

$$\alpha = \frac{\sum_{i=1}^{25} [1 - |Z_i - (\frac{X_i}{h})|]}{n}$$

donde:

Z_i es un pixel de la imagen de entrada.

X_i es un elemento del patrón almacenado en la matriz evolutiva.

h es el número de coincidencias o reforzamientos del patrón en este caso $h = 7$.

Expandiendo la expresión anterior para los 25 elementos se obtiene el cálculo siguiente:

$$\alpha = \frac{\left(1 - |Z_1 - (\frac{X_1}{7})|\right) + \left(1 - |Z_2 - (\frac{X_2}{7})|\right) + \dots + \left(1 - |Z_{25} - (\frac{X_{25}}{7})|\right)}{n}$$

Sustituyendo los valores de la imagen de entrada y del patrón para calcular la semejanza, se obtiene:

$$\alpha = \frac{22.0006}{25}$$

$$\alpha = 0.88$$

Por lo tanto, el porcentaje de similitud es de 88%

Para cada clase de la matriz evolutiva se obtiene el factor α_j . De los α_j calculados, se selecciona el mayor y se compara contra el umbral de similitud θ . Si $\alpha_j > \theta$ entonces se refuerza el patrón existente en la matriz evolutiva sumando el valor de cada elemento del valor de entrada Z_i a cada elemento de la clase de la matriz evolutiva X_i e incrementando uno al umbral h . Si $\alpha_j \leq \theta$ se solicita al usuario que indique el significado del símbolo desconocido y se anexa como un nuevo patrón en la matriz evolutiva, el valor del factor de coincidencia h se inicializa con el valor uno y el significado es el que digitó el usuario.

Dado que $\theta = 95\%$ entonces no se cumple el umbral de similitud porque el porcentaje de similitud obtenido es de 88% por lo tanto, se pide al usuario que digite el significado de la entrada. La imagen procesada se adiciona como un renglón más en la matriz evolutiva.

5. Resultados obtenidos

En este capítulo se comentan los resultados obtenidos de la implantación de los dos métodos de solución propuestos para el reconocimiento del alfabeto del Lenguaje de Señas Mexicano.

Los resultados obtenidos muestran que con el segundo método se reconocen 25 señas del alfabeto del LSM, identificadas en imágenes capturadas mediante el sensor de profundidad del dispositivo Kinect. Con el primer método se reconocen 20 señas del mismo alfabeto, en este caso se usa una cámara fotográfica para capturar las imágenes que se reconocen.

En cada método se presenta su descripción, consideraciones, la forma de operación y se comentan los resultados obtenidos.

5.1 Primer método de solución propuesto

El primer método se basa en la extracción de características que describen las propiedades de las señas de las manos contenidas en imágenes capturadas con una cámara fotográfica. A cada característica se le llama rasgo descriptor.

Se seleccionaron rasgos descriptores con capacidad discriminante, fiabilidad y que fueran capaces de ofrecer invariantes a transformaciones. Inicialmente se usaron los momentos invariantes de Hu pero debido a que en algunas señas se detectó ambigüedad, por ejemplo, la letra P y la letra K son similares y únicamente se diferencian en el ángulo en el que se visualizan, ocurre algo similar con las letras L y G, las letras U, I y Z. Para resolver la ambigüedad fue necesario utilizar otros rasgos que permitieran discriminar la forma y el ángulo; estos son: el factor de excentricidad, ángulo del elipsoide y la redondez del objeto.

5.1.1 Consideraciones

El primer método requiere que el fondo de la imagen capturada sea oscuro y que la ejecución de la seña del usuario sea realizada con la mano derecha, debido a que los ángulos son diferentes para la mano izquierda. La distancia entre el objeto y la cámara se estableció en 50cm, dado que la iluminación en esa distancia fue la más adecuada para la cámara que se usó (sin demasiado brillo, ni imágenes opacas). Las imágenes capturadas incluyen la mano hasta el inicio de la muñeca, por lo que es necesario el uso de mangas y ropa con un color semejante al del fondo con el objetivo de que el antebrazo no sea capturado en la imagen porque afecta el proceso de reconocimiento de la seña.

La cámara fotográfica utilizada es de la marca Sony modelo Xperia X10 con una resolución de 8 megapíxeles.

En este método las imágenes se almacenan manualmente en una computadora y el usuario selecciona la imagen a procesar.

En este método los rasgos de los patrones de las letras que se reconocen están en el código del programa a diferencia del segundo método que almacena los patrones reconocidos en la memoria RAM.

5.1.2 Operación del método

La operación del primer método requiere de la captura inicial de las imágenes con señas del alfabeto del LSM.

El usuario selecciona manualmente el archivo de la imagen a procesar, en la figura 5.1 se seleccionó la seña de la letra D.



Figura 5.1 Imagen de la letra D

A la imagen seleccionada se le aplica el proceso de umbralado usando el método de Kittler e Illinworth, dando como resultado la imagen de la figura 5.2. En esta etapa es posible que se presenten algunos artefactos indeseables en la imagen, como puntos blancos en el fondo negro.



Figura 5.2 Imagen resultante del proceso de umbralado de la imagen de la letra D mediante el método de Kittler e Illinworth

La eliminación de los artefactos se realiza mediante el uso de filtro secuencial morfológico de apertura – clausura. Al objeto resultante se le calcula su contorno y el ángulo del elipsoide que lo envuelve. El resultado de estos dos procesos se observa en la figura 5.3.

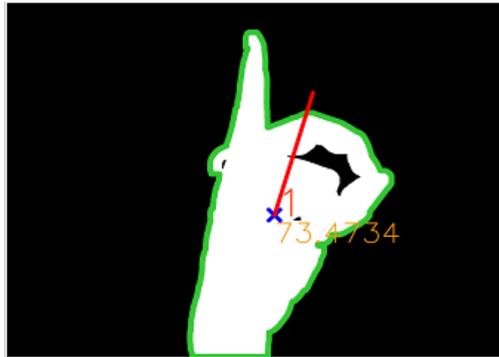


Figura 5.3 Visualización del ángulo y el contorno de la letra D

El resultado del proceso de reconocimiento se presenta en forma de texto en un recuadro (figura 5.4).



Figura 5.4 Resultado del reconocimiento de la letra D en forma de texto

5.1.3 Resultados

Con este primer método se reconocieron 20 de las señas del alfabeto del Lenguaje de Señas Mexicano, correspondientes a las letras:

A, B, C, D, E, F, G, I, K, L, M, O, Q, R, S, T, V, W, X, Y.

Se probaron 53 imágenes, en promedio tres imágenes por letra con deformaciones aleatorias.

Las señas de las letras H, P, U y Z se confundieron con los patrones de las letras G, K e I, la seña de la letra N se confunde con la letra M; por lo tanto se consideran como no reconocidas. La letra J tampoco se reconoce porque es una forma dinámica de la letra I. En este método, las señas no reconocidas fueron consecuencia de la aplicación de las invariantes de rotación y de escalamiento, propias del método.

El problema de rotación de las letras Q y C, así como de las letras L y G tratado en [12], fue resuelto con este método, mediante el cálculo del ángulo del semieje más cercano al eje coordenado X de la elipse que envuelve a la mano y la desambiguación de los 180 grados de este eje mediante el cálculo del sesgo proyectivo [13].

5.1.4 Ventajas y desventajas del primer método

Este método tiene como ventajas las siguientes:

1. El método usa las imágenes de una cámara fotográfica estándar.
2. No requiere uso de guantes.

Las desventajas del método que se detectan son:

1. El método requiere de un fondo oscuro.
2. Se requiere de una manga de color del fondo
3. Las señas que se reconocen se efectúan únicamente con la mano derecha.
4. Reconoce solamente 20 señas del alfabeto del LSM

5.2 Segundo método de solución propuesto

El segundo método utiliza la cámara infrarroja de un dispositivo Kinect para facilitar el proceso de segmentado de la mano del ejecutante de las señas del alfabeto del LSM.

Este método utiliza una técnica que consta de un proceso de reconocimiento–aprendizaje, usando patrones almacenados en matrices evolutivas. Este proceso se apoya en tres procesos: el proceso de normalización, el proceso de reforzamiento y el proceso de olvido.

El proceso de normalización se realiza con el propósito de atenuar variaciones por escalamiento de la mano ajustar el patrón a un tamaño estándar establecido en 200 x 200 píxeles.

El proceso de reforzamiento acumula la información de los patrones que tienen un valor de similitud igual o superior al umbral establecido. Con

este proceso se obtiene un patrón común con zonas ponderadas. Las pequeñas variaciones de los patrones que se acumulan se deben a diferencias en los ángulos o debido a deformaciones en su ejecución.

El proceso de olvido permite eliminar las zonas del patrón que tienen pocas ocurrencias y que se consideran como poco significativas.

Las pruebas realizadas permitieron detectar que estos procesos le proporciona al segundo método tolerancia a transformaciones como: rotación, escalamiento, translación y otras deformaciones ocasionadas por la morfología de las manos de las personas.

Para facilitar el proceso de reconocimiento, este método requiere que se personalice la base de conocimientos para cada usuario, con el objetivo de eliminar la ambigüedad entre patrones de diferentes personas, por ejemplo, si el método se entrena para una persona con dedos delgados, puede llegar a confundir la seña de la letra H con la seña de la letra G ejecutada por una persona con dedos de la mano más gruesos

5.2.1 Consideraciones

El segundo método utiliza información de un flujo de video obtenido del sensor de profundidad del dispositivo Kinect. El flujo está formado por secuencias de mapas de profundidad. Como consecuencia los movimientos rápidos de la mano pueden provocar desenfoque por movimiento (*blur motion*). El desenfoque afecta la definición de la mano y dificulta su correcto reconocimiento.

La ventana de captura de la mano usando el dispositivo Kinect se estableció en 200mm. La ventana se puede colocar desde los 400mm hasta una distancia máxima de 800mm.

Las señas de entrada son ejecutadas por el usuario con la mano dominante generalmente la mano derecha, aunque es posible en este método que el programa se entrene y reconozca las señas ejecutadas con la mano izquierda sin cambiar el código de los programas. Si se quisieran almacenar los patrones de ambas manos se tendrían problemas de ambigüedad, por ejemplo, con las letras Z e I.

En este método, los patrones se procesan en memoria RAM y la persistencia de patrones se hace en archivos binarios que se almacenan en el disco duro.

5.2.2 Operación del método

Para la operación del segundo método es importante establecer la distancia y el ángulo de operación del dispositivo de captura Kinect debido a que la propuesta es tolerante a transformaciones, no invariante a ellas.

En la figura 5.5 se muestra un ejemplo del proceso de aprendizaje y reconocimiento del segundo enfoque. Se asume que la base de conocimiento está vacía al principio del proceso de reconocimiento.

El primer renglón muestra la seña de la letra A que se aprende porque no existe ningún patrón que tenga similitud mayor o igual que el 90%. El umbral de 90% se determinó con base en diferentes experimentos.

En el segundo renglón otra persona muestra la seña de la letra A, obteniendo una similitud del 92% con respecto a la letra A previamente aprendida, por lo que no se requirió de un nuevo aprendizaje, aunque sí ocurre un reforzamiento.

En el renglón tres, la segunda persona vuelve a mostrar el signo de la letra A; el sistema reconoce la letra A con un porcentaje de similitud de 90%, por lo que solamente ocurre un reforzamiento.

En el cuarto renglón la primera persona muestra la seña de la letra B obteniendo una similitud de 60% con respecto al patrón de la letra A, por lo que no satisface el umbral de, mayor o igual a 90 %, por lo tanto se invoca al módulo de aprendizaje y se aprende la seña de la letra B.

En la quinta imagen la segunda persona muestra la seña de la letra C obteniendo una similitud con respecto al patrón de la letra A de 68% y 70% con respecto al patrón de la letra B, por lo que, se invoca al módulo de aprendizaje y se adquiere la seña de la letra C como un nuevo patrón.

5.2.3 Resultados

El segundo método desarrollado en esta tesis para reconocer el alfabeto del Lenguaje de Señas Mexicano resuelve los problemas presentados en el primer método porque sus algoritmos se adecuan a los patrones de las imágenes, facilitando su comparación con los patrones almacenados. Este método permite tolerancias a transformaciones geométricas propias del problema.

Con base en pruebas experimentales se construyó una tabla de diferencias entre los patrones de las letras del LSM que reconoce el segundo método (figura 5.6). En la tabla se muestra el porcentaje de la diferencia entre la letra de la fila con la letra de la columna. La diferencia entre las mismas letras es cero y ocurre en la diagonal principal. Con sombreado azul se indican los niveles de diferencia que representan los patrones con mayor similitud, como el caso de la letra G y la H, que tienen una diferencia de solamente 11% y el caso de la letra B y la F que tienen una diferencia de 11%, entre otros. Por lo tanto, la semejanza deberá ser al menos de 90%, es decir, una diferencia menor a 11%.

Imagen de entrada	Patrón detectado	% Similitud	Respuesta
		Ninguna	Aprende "A"
		92% "A"	"A"
		90% "A"	"A"
		60% "A"	Aprende "B"
		68% "A" 70% "B"	Aprende "C"

Figura 5.5 Ejemplo de operación del aprendizaje-reconocimiento del segundo método de solución propuesto

La comparación de los resultados obtenidos en los métodos desarrollados en esta tesis, con los reportados en los trabajos del estado del arte, se presentan en la figura 5.7. Los resultados están separados en dos columnas, la primera considera los trabajos relacionados con las señas del LSM y la segunda, los trabajos de gestos ejecutados con las manos y las señas de otros lenguajes.

En los trabajos de los gestos diferentes al LSM, el trabajo que más señas reconoce es el de la referencia [11] con 19 señas del LSE. Este trabajo reporta el requerimiento de condiciones especiales de operación, estas son: iluminación adecuada y el uso de guantes de color azul. Le sigue el trabajo de la referencia [18] que reconoce nueve gestos expresados con las manos usando el dispositivo Kinect, aunque no pertenecen a un lenguaje

de señas. El tercer trabajo es el de la referencia [19] que reconoce cinco gestos de los dedos de las manos, para lo cual utiliza el dispositivo Kinect.

En los resultados solo se encontró, hasta el momento, un trabajo que reconoce el LSM [12], el cual reporta que reconoce 25 letras con un porcentaje de reconocimiento promedio del 80%. Este trabajo requiere condiciones controladas de iluminación, distancia fija entre la cámara y el objeto y el uso de fondo de color negro.

	a	b	c	d	e	f	g	h	i	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
a	0	0.35	0.31	0.33	0.35	0.31	0.31	0.28	0.33	0.34	0.32	0.40	0.40	0.32	0.26	0.22	0.39	0.45	0.35	0.43	0.35	0.33	0.33	0.33	0.39
b	0.35	0	0.30	0.18	0.22	0.11	0.28	0.32	0.16	0.30	0.33	0.30	0.30	0.33	0.23	0.29	0.16	0.41	0.34	0.16	0.16	0.16	0.20	0.24	0.21
c	0.31	0.30	0	0.28	0.29	0.30	0.23	0.22	0.31	0.19	0.30	0.41	0.41	0.28	0.29	0.30	0.31	0.41	0.29	0.32	0.32	0.31	0.29	0.24	0.31
d	0.33	0.18	0.28	0	0.27	0.14	0.24	0.27	0.17	0.25	0.26	0.41	0.40	0.32	0.18	0.34	0.17	0.46	0.34	0.18	0.17	0.16	0.22	0.21	0.17
e	0.35	0.22	0.29	0.27	0	0.19	0.26	0.28	0.26	0.31	0.31	0.23	0.26	0.30	0.32	0.36	0.28	0.23	0.22	0.28	0.29	0.29	0.22	0.28	0.32
f	0.31	0.11	0.30	0.14	0.19	0	0.22	0.26	0.12	0.29	0.28	0.32	0.32	0.31	0.19	0.31	0.18	0.40	0.32	0.19	0.16	0.14	0.16	0.24	0.19
g	0.31	0.28	0.23	0.24	0.26	0.22	0	0.11	0.26	0.25	0.24	0.43	0.45	0.25	0.26	0.38	0.27	0.38	0.30	0.32	0.26	0.27	0.26	0.25	0.26
h	0.28	0.32	0.22	0.27	0.28	0.26	0.11	0	0.27	0.25	0.28	0.42	0.44	0.27	0.30	0.36	0.31	0.40	0.31	0.34	0.33	0.32	0.28	0.25	0.28
i	0.33	0.16	0.31	0.17	0.26	0.12	0.26	0.27	0	0.29	0.33	0.36	0.37	0.36	0.21	0.30	0.15	0.44	0.33	0.18	0.17	0.17	0.18	0.25	0.16
k	0.34	0.30	0.19	0.25	0.31	0.29	0.25	0.25	0.29	0	0.32	0.40	0.40	0.25	0.25	0.36	0.29	0.44	0.31	0.30	0.30	0.29	0.31	0.22	0.27
l	0.32	0.33	0.30	0.26	0.31	0.28	0.24	0.28	0.33	0.32	0	0.48	0.51	0.36	0.29	0.40	0.36	0.41	0.32	0.37	0.31	0.28	0.30	0.34	0.33
m	0.40	0.30	0.41	0.41	0.23	0.32	0.43	0.42	0.36	0.40	0.48	0	0.11	0.39	0.43	0.33	0.38	0.30	0.32	0.36	0.39	0.39	0.30	0.32	0.44
n	0.40	0.30	0.41	0.40	0.26	0.32	0.45	0.44	0.37	0.40	0.51	0.11	0	0.38	0.42	0.30	0.38	0.34	0.34	0.36	0.39	0.39	0.31	0.33	0.43
o	0.32	0.33	0.28	0.32	0.30	0.31	0.25	0.27	0.36	0.25	0.36	0.39	0.38	0	0.32	0.38	0.35	0.33	0.28	0.35	0.33	0.32	0.37	0.32	0.32
p	0.26	0.23	0.29	0.18	0.32	0.19	0.26	0.30	0.21	0.25	0.29	0.43	0.42	0.32	0	0.31	0.25	0.52	0.39	0.28	0.21	0.21	0.27	0.29	0.23
q	0.22	0.29	0.30	0.34	0.36	0.31	0.38	0.36	0.30	0.36	0.40	0.33	0.30	0.38	0.31	0	0.34	0.47	0.33	0.35	0.34	0.33	0.30	0.30	0.35
r	0.39	0.16	0.31	0.17	0.28	0.18	0.27	0.31	0.15	0.29	0.36	0.38	0.38	0.35	0.25	0.34	0	0.43	0.33	0.11	0.16	0.21	0.25	0.22	0.14
s	0.45	0.41	0.41	0.46	0.23	0.40	0.38	0.40	0.44	0.44	0.41	0.30	0.34	0.33	0.52	0.47	0.43	0	0.21	0.41	0.44	0.45	0.40	0.46	0.44
t	0.35	0.34	0.29	0.34	0.22	0.32	0.30	0.31	0.33	0.31	0.32	0.32	0.34	0.28	0.39	0.33	0.33	0.21	0	0.31	0.34	0.32	0.30	0.31	0.31
u	0.43	0.16	0.32	0.18	0.28	0.19	0.32	0.34	0.18	0.30	0.37	0.36	0.36	0.35	0.28	0.35	0.11	0.41	0.31	0	0.16	0.20	0.25	0.24	0.13
v	0.35	0.16	0.32	0.17	0.29	0.16	0.26	0.33	0.17	0.30	0.31	0.39	0.39	0.33	0.21	0.34	0.16	0.44	0.34	0.16	0	0.11	0.24	0.26	0.11
w	0.33	0.16	0.31	0.16	0.29	0.14	0.27	0.32	0.17	0.29	0.28	0.39	0.39	0.32	0.21	0.33	0.21	0.45	0.32	0.20	0.11	0	0.25	0.30	0.15
x	0.33	0.20	0.29	0.22	0.22	0.16	0.26	0.28	0.18	0.31	0.30	0.30	0.31	0.37	0.27	0.30	0.25	0.40	0.30	0.25	0.24	0.25	0	0.22	0.26
y	0.30	0.24	0.24	0.21	0.28	0.24	0.25	0.25	0.25	0.22	0.34	0.32	0.33	0.32	0.29	0.30	0.22	0.46	0.31	0.24	0.26	0.30	0.22	0	0.27
z	0.39	0.21	0.31	0.17	0.32	0.19	0.26	0.28	0.16	0.27	0.33	0.44	0.43	0.32	0.23	0.35	0.14	0.44	0.31	0.13	0.11	0.15	0.26	0.27	0

Figura 5.6 Diferencias entre patrones de las letras

En la figura 5.6 se muestra el comportamiento de la similitud entre los diferentes patrones de las letras del alfabeto del LSM. Los valores en la tabla serán similares aun cuando distintas personas sean las que entrenen al segundo método. Los patrones con mayor sensibilidad a la ambigüedad son los que aparecen en las intersecciones sombreadas con color azul y tipografía cursiva, debido a que la similitud entre ellas es alta. Por ejemplo, el caso de los patrones de las letras G y H que tienen una diferencia de 11%.

El primer método propuesto en esta tesis reconoce 20 letras del LSM con un porcentaje de reconocimiento del 100%. Las condiciones de operación son similares a las de la referencia [12].

Con el segundo método propuesto en esta tesis se reconocen 25 letras del LSM con porcentaje de similitud igual o superior al 90%. Este método aprovecha el dispositivo Kinect para aumentar la eficiencia en la segmentación y contar con tolerancia a condiciones de iluminación. Además, se aprovechan las matrices evolutivas para alcanzar los porcentajes de similitud reportados, logrando dar mayor flexibilidad a la forma de operación y funcionamiento respecto a algunas de las restricciones de operación de varios trabajos reportados.

Trabajo de investigación	Número de señas del LSM reconocidas	Gestos manuales reconocidos
Una interfaz para el reconocimiento de señas gestuales para personas hipoacúsicas [11]	-	19
Sistema para el reconocimiento del alfabeto Dactilológico [12]	25*	-
Reconocimiento de gestos manuales utilizando Kinect (Yi Li) [18]	-	9
Reconocimiento de gestos manuales utilizando Kinect (Heng Du) [19]	-	5
Reconocimiento de imágenes del LSM 1er método de solución propuesto	20	-
Reconocimiento de imágenes del LSM 2do método de solución propuesto [33]	25	-

*bajo ciertas condiciones.

Figura 5.7 Comparación del número de señas reconocidas.

5.2.4 Ventajas y desventajas del segundo método

El segundo método presenta las siguientes ventajas:

1. Invariancia la luminosidad porque aprovecha la cámara infrarroja del dispositivo Kinect.
2. No se requiere de entrenamiento previo, porque es capaz de aprender patrones en línea y constituir su base de conocimientos desde el primer patrón.
3. Los procesos de normalización, reforzamiento y olvido proporcionan tolerancia a rotaciones y deformaciones de las señas.
4. No se requiere de guantes, ropa o accesorios especiales.
5. No se requiere de un fondo especial.

6. El método puede utilizarse con otros sensores de profundidad similares, por ejemplo:
 - ASUS WAVI Xtion PRO.
 - ASUS WAVI Xtion PRO Live.
 - PMD [vision].
 - Panasonic D-Imager.
 - Leap-motion.

Las desventajas del método que se detectan son:

1. Se requiere de un dispositivo especial.
2. No atiende el reconocimiento de señas dinámicas.
3. Cuando las manos de las personas son muy diferentes entre sí, se requiere personalizar la base de conocimientos para evitar ambigüedades.

5.3 Contribuciones

Las contribuciones del desarrollo de esta tesis son:

1. La puesta en práctica y aprovechamiento de la tecnología del dispositivo Kinect como una forma alternativa para llevar a cabo el proceso de segmentado y filtrado de imágenes.
2. Un enfoque alternativo para realizar el reconocimiento de lenguaje de señas respecto a los existentes en la literatura.
3. Dos distancias para la obtención del factor de semejanza en las matrices evolutivas.
4. Un método de solución basado en matrices evolutivas aplicado al reconocimiento del alfabeto del Lenguaje de Señas Mexicano que aprovecha los mapas de profundidad.
5. Un sistema capaz de adaptarse a los usuarios con el fin de que no necesiten un experto o una capacitación especializada para entrenarlo. Por lo tanto se provee un sistema más intuitivo y fácil de utilizar.
6. Una ponencia en 12th edition of the International Congress on Computer Science (CORE 2012) [33].

5.4 Conclusiones

En esta tesis se desarrollaron dos métodos para reconocer el alfabeto del Lenguaje de Señas Mexicano.

Con el primer método se reconocen 20 señas del LSM, presenta invariantes a rotaciones, escalamiento y traslaciones. Este método requiere de fondo negro y condiciones adecuadas de iluminación, que las señas del alfabeto del LSM efectuadas se ejecuten con la mano derecha y se fije la distancia de operación en 50cm entre el objeto y la cámara.

Con el primer método se detectó que uno de los problemas para segmentar la mano, es encontrar el punto donde se encuentra la muñeca, porque el antebrazo afecta directamente en el proceso de reconocimiento de las señas.

La combinación de los momentos invariantes de Hu, con los otros rasgos, permitieron discriminar las letras que son similares ante rotaciones.

El primer método es sensible a cambios en las condiciones de luminosidad.

Con el segundo método se reconocieron 25 señas del LSM, presenta tolerancia a las rotaciones y deformaciones en la ejecución de las señas. Es invariante a la traslación y también a los cambios de luminosidad. No requiere de aprendizaje previo.

La imagen capturada con la cámara infrarroja del dispositivo Kinect representa una ventaja, porque hace eficiente el segmentado de la mano, que es el objeto de interés en esta tesis. Las pruebas realizadas permitieron detectar que el dispositivo Kinect para este problema es deseable que tenga mejor resolución para la captura de las imágenes de la mano, debido a que el grado de discretización es en milímetros y se estima que para efectos de reconocer los detalles de la mano se debe trabajar con fracciones de milímetro. El dispositivo Kinect no puede distinguir las transiciones entre señas ejecutadas rápidamente pues se produce un efecto de desenfoque por movimiento conocido como *motion blur*.

El uso de la matriz evolutiva y los procesos de normalización, reforzamiento y olvido permiten mayor eficiencia en la detección de imágenes con respecto al primer método. El reforzamiento permite reducir el número de patrones con los que se hace la comparación para hacer el reconocimiento de las señas de la mano.

El proceso de aprendizaje-reconocimiento del segundo método requiere definir un umbral de similitud mínimo para discriminar entre las señas del alfabeto del LSM con base en los experimentos y dada la similitud de algunas señas se detectó que el umbral debe de ser menor o igual al 10% para garantizar una adecuada discriminación entre los patrones.

Una misma base de conocimiento puede servir para distintas personas cuando tengan similitud en la morfología de sus respectivas manos y que la ejecución de las señas sea similar.

5.5 Trabajo futuro

Con base en la experiencia en el desarrollo de esta tesis, se proponen los trabajos a futuro siguientes:

1. Combinar el primer y segundo métodos para aprovechar las ventajas de ambos métodos.
2. Ampliar el trabajo de la presente tesis para resolver las señas dinámicas, para lo cual deben considerarse elementos como la cara, el torso y la identificación de la posición de la seña en uno de los 12 cuadrantes del LSM.
3. Utilizar diferentes unidades de computo para aumentar la eficiencia en el procesamiento porque actualmente se procesa solamente una imagen por segundo usando un procesador Core 2 duo a 2.4Ghz y 4 Gb de RAM.
4. Se requiere desarrollar un método eficiente de indexación y búsqueda para el acceso a cantidades de cientos de patrones.
5. Aplicar el segundo método de solución al reconocimiento de otros tipos de objetos de interés.
6. Se recomienda integrar un proceso de deconvolución para resolver el problema del desenfoque (*motion blur*).

Referencias

- [1] Real Academia Española, <http://lema.rae.es/drae/>, sitio web 2012
- [2] INEGI Características de las personas con discapacidad auditiva, http://www.inegi.org.mx/prod_serv/contenidos/espanol/bvinegi/productos/censos/poblacion/2000/discapacidad/auditiva_i.pdf, sitio web 2011
- [3] Americans with disabilities, <http://www.census.gov/prod/2008pubs/p70-117.pdf>, sitio web 2011
- [4] Personas con discapacidad y estadísticas de las personas sordas, <http://cuentame.inegi.org.mx/poblacion/discapacidad.aspx?tema=P>, sitio web 2011
- [5] INEGI Las personas con discapacidad en México: una visión censal, http://www.inegi.org.mx/prod_serv/contenidos/espanol/bvinegi/productos/censos/poblacion/2000/discapacidad/discapacidad2004.pdf, sitio web 2011
- [6] Fernando Galindo Soria, Sistemas evolutivos en *Boletín de política informática*, México, 1986
- [7] Fernando Galindo Soria, Una representación matricial para sistemas evolutivos en *Teoría y práctica de los sistemas evolutivos* (editor Jesús Manuel Olivares Ceja), México, 1998
- [8] Fernando Galindo Soria, Sistemas evolutivos: nuevo paradigma de la informática, en *Memorias del congreso TEC-COMP 91*, México, 1991
- [9] Fernando Galindo Soria, Sistemas evolutivos, espacios de caos y mecanismos de olvido en matrices evolutivas (comunicación personal), México, 2012
- [10] Luis Armando López García, Rosa María Rodríguez Cervantes, María Guadalupe Zamora Martínez, Susana San Esteban Sosa, *Mis manos que hablan, lengua de señas para sordos*, 1ra. edición, editorial Trillas, México, 2010, pp 305
- [11] Isaac García Incertis, Hand gesture recognition for deaf people interfacing en *The 18th international conference on pattern recognition (ICPR'06)*, IEEE, 2006
- [12] Laura Jeanine Razo Gil, *Sistema para el reconocimiento del alfabeto dactilológico* (Tesis de maestría), Centro de Investigación en Computación del Instituto Politécnico Nacional, México, 2009
- [13] Juan Humberto Sossa Azuela, *Rasgos descriptores para el reconocimiento de objetos*, Instituto Politécnico Nacional, México, 2006, pp 245
- [14] Transformaciones geométricas, <http://www.vision.uji.es/courses/Doctorado/FVC/FVC-T10-Geometricas-4p.pdf/>, sitio web 2011
- [15] Helen Cooper, Brian Holt, Richard Bowden, Sign language recognition, University Of Surrey, Guildford 2011
- [16] Karla Faurot, Dianne Dellinger, Andy Eatough, Steve Parkhurst, Lenguaje de signos mexicano - La identidad como lenguaje del sistema de signos mexicano, www.sil.org/mexico/lenguajes-de-signos/G009e-Identidad-mfs.pdf, sitio web 2011
- [17] Lengua de señas mexicana, http://es.wikipedia.org/wiki/Lengua_de_señas_mexicana, sitio web 2011
- [18] Yi Li, hand gesture recognition using Kinect en *3rd International Conference on Software Engineering and Service Science (IC-SESS2012)*, IEEE, 2012

- [19] Heng Du, TszHang To, *Hand gesture recognition using Kinect*, Boston University Technical report No. ECE-2011-04, 2012
- [20] Luis E. Torres Hernández, Luis C. Longoria G., Antonio Rojas Salinas, Aplicación de los sistemas evolutivos en el análisis de espectros de rayos gamma, Instituto Nacional de Investigaciones Nucleares, Departamento de Experimentación Nuclear, México, 1998.
- [21] Eduardo De la Cruz Sánchez, Luis C. Longoria Gándara, Rodolfo A. Carrillo Mendoza, Sistema evolutivo para el diagnóstico de fallas en máquinas rotatorias, Instituto Nacional de Investigaciones Nucleares, Departamento de Experimentación Nuclear, México, 1998.
- [22] Ogle, Kenneth N., *Researches in binocular vision*, Oxford, England: W. B. Saunders, Inglaterra, 1950
- [23] Gonzalo Pajares, Jesús de la Cruz, *Visión por computador, imágenes digitales y aplicaciones*, 2da edición, Alfaomega, 2008, pp 768
- [24] Gonzalo Pajares, Jesús de la Cruz, *Ejercicios resueltos de visión por computador*, 1ra reimpresión, Alfaomega, 2008, pp 354
- [25] Edgardo Felipe Riverón, Notas de clases en análisis de imágenes y morfología matemática (Curso académico), CIC-IPN, México, 2012
- [26] Rafael C. Gonzalez, Richard E. Woods, *Digital image processing*, 3ra edición, Prentice Hall, 2008, pp 976
- [27] J. L. Díaz de León-Santiago y C. Yáñez-Márquez, *Introducción a la morfología matemática de conjuntos*, CIC-IPN-UNAM-FCE., México, 2003
- [28] P. Soille, *Morphological image analysis*, 2da edición, Springer, 2002, pp 391
- [29] Cecilia Albortante Morato, *Separación de las redes vasculares arterial y venosa en imágenes oftalmoscópicas de retina humana* (Tesis de maestría), Centro de Investigación en Computación del Instituto Politécnico Nacional, México, 2009
- [30] José Luis Alba, et al, *Morfología matemática*, www.gts.tsc.uvigo.es/pi/Morfologia_matematica.pdf, Universidad de Vigo, sitio web, 2012
- [31] Mehmet Sezgin, Bulent Sankur, Survey over image thresholding techniques and quantitative performance evaluation en *Journal of Electronic Imaging* Vol 13(1), 2004
- [32] Operaciones morfológicas, http://www.tsc.uc3m.es/imagen/Curso_ProcesadoMorfologico/Contenido/Operaciones/OperacionesMorfologicas.html, Universidad Carlos III de Madrid, sitio web 2012
- [33] Fausto Pável Priego Pérez, Jesús Manuel Olivares Ceja, José Félix Serrano Talamantes, Diana Naim Aguilar Rivera, Reconocimiento de imágenes del Lenguaje de Señas Mexicano, en *Advances in Computer Science and Engineering*, volumen 58, pp 57-68, ISSN: 1870-4069, noviembre 2012, México, D.F.