

Universidad de las Ciencias Informáticas

Facultad 9



# **Propuesta de una técnica para la detección de imágenes por bordes o contornos en secuencias de video.**

**Trabajo de Diploma para optar por el título de Ingeniero en Ciencias Informáticas**

**AUTOR:** Reinier Quintero Pérez

**TUTOR:** Ing. Yanio Hernández Heredia

Ciudad de la Habana, Julio de 2010

Año 52 de la Revolución



“El futuro de Cuba tiene que ser necesariamente un futuro de hombres de ciencia, de hombres de pensamientos...”

Fidel Castro Ruz

## **Dedicatoria**

A mis padres, abuelos, en especial a Zaida Filiú Brooks madre incondicional por su ejemplo y amor infinito, por estar conmigo en todos los momentos de mi vida y por ser ella la razón de seguir adelante cada día. A mi novia Cecilia por haber estado a mi lado en los momentos malos, buenos y difíciles con tanto amor y cariño, en fin a mi familia y amigos por la fé y la confianza depositada.

## **Agradecimientos**

Agradezco a todas las personas que de una forma u otra lograron formarme como ser humano y como profesional, a mis dos madres y novia que siempre han estado conmigo en los momentos más difíciles de la carrera y a lo largo de mi vida estudiantil gracias por el cariño que he tenido siempre, a mí tío y papá por depositar toda su confianza en mí, a la Revolución una deuda impagable por haberme dado esta oportunidad de prepararme como profesional. A mi tutor Yanio y a Frank por el apoyo incondicional, siempre con la voluntad de trabajar y sirviendo de guía en cada idea, en cada fundamentación.

# **Declaración de Autoría**

Declaro que soy el único autor de este trabajo y autorizo al Departamento productivo Señales Digitales de la Universidad de las Ciencias Informáticas a hacer uso del mismo en su beneficio.

Para que así conste firmo la presente a los 29 días del mes de noviembre del año 2009.

Reinier Quintero Pérez

Yanio Hernández Heredia

## **Datos de contacto**

Ing. Yanio Hernández Heredia (email [yhernandez@uci.cu](mailto:yhernandez@uci.cu))

Profesor graduado de Ingeniero en Ciencias Informáticas de la facultad 9 y jefe del área de gestión y procesamiento de medias del Departamento Señales Digitales.

## **Opinión del tutor**

## **Resumen**

En la actualidad la visión Computacional se muestra como una técnica que facilitara las actividades que pasan desapercibidas o no se es capaz de captar por descuido o falta de información, en este sentido son varios los campos que se beneficiaran por el uso de este paradigma para contar con información que facilite operaciones, indique detalles no disponibles a simple vista, etc. Mediante esta investigación se presenta una propuesta de una técnica de reconocimiento de imágenes por bordes o contornos en secuencias de video para mejorar la gestión y procesamiento de medias en las aplicaciones del Departamento productivo Señales Digitales de la Universidad de Ciencias Informáticas. La investigación muestra las distintas técnicas de visión computacional relacionadas a la detección de imágenes por bordes o contornos y se centra fundamentalmente en la búsqueda de una heurística para el procesamiento de fotogramas y reconocimiento de patrones en los mismos.

## **Palabras claves**

Detección por bordes, detección por contornos, visión computacional, fotograma, píxel.

# Abstract

At present the Computer Vision is shown as a technique to facilitate the activities that go unnoticed or is not able to capture through carelessness or lack of information, in this sense are several fields that will benefit from the use of this paradigm to have information that will facilitate transactions, provide details unavailable to the naked eye, and so on. Through this research presents a proposal of an image recognition technique by edges or contours in video sequences to improve the management and media processing applications Digital Signal Department of Computer Science University. Research shows the different techniques of Computer Vision related to the detection of edges or contours imaging and focuses primarily on a heuristic search for pattern recognition in image sequences.

# Contenido

<b>Introducción</b> .....	1
<b>Capítulo 1 Fundamentación Teórica</b> .....	7
1.1 ¿Qué son las técnicas de detección de imágenes por bordes o contornos? .....	7
1.2 Estado del arte de las técnicas de detección de imágenes por bordes o contornos .....	8
1.3 Análisis de las técnicas existentes .....	17
1.3.1 Técnica de Canny .....	19
1.3.2 Matching Jerárquico de Chamfer .....	21
1.3.3 Contornos activos .....	23
1.4 Herramientas para el procesamiento de imágenes .....	27
1.4.1 C++ .....	28
1.4.2 MATLAB .....	29
Conclusiones .....	29
<b>Capítulo 2 Solución propuesta</b> .....	30
2.1 Selección del método propuesto .....	30
2.2 Preprocesamiento .....	32
2.3 Segmentación .....	33
2.4 Extracción de características .....	35
2.5 Reconocimiento de patrones .....	36
Conclusiones .....	38
<b>Capítulo 3 Resultados obtenidos</b> .....	39
3.1 Dominio del modelo propuesto.....	39
3.2 Preprocesamiento .....	40
3.3 Segmentación .....	43
3.4 Extracción de características y reconocimiento de patrones .....	46
Conclusiones .....	48
<b>Conclusiones Generales</b> .....	49
<b>Recomendaciones</b> .....	50
<b>Referencias Bibliográficas</b> .....	51
<b>Anexos</b> .....	52
<b>Glosario</b> .....	62

# Índice de figuras

Figura 1 Proceso de obtención del gradiente (4).....	20
Figura 2 Supresión no máxima (4).....	20
Figura 3 Histéresis de umbral (4).....	21
Figura 4 Matching Jerárquico de Chamfer (18).....	23
Figura 5 Escala de grises.....	41
Figura 6 Reducción de ruido con un filtro pasa bajo.....	42
Figura 7 Reducción de ruido con un filtro de mediana.....	43
Figura 8 Segmentación de la imagen.....	43
Figura 9 Binarización con umbral Otsu.....	44
Figura 10 Binarización con umbral arbitrario.....	45
Figura 11 Binarización de la imagen.....	45
Figura 12 Operaciones morfológicas.....	46
Figura 13 Proceso de extracción de características y reconocimiento.....	47
Figura 14 Ruido Detector (9).....	52
Figura 15 Ruido Sal y pimienta (9).....	52
Figura 16 Imagen digital (9).....	53
Figura 17 Fotograma (9).....	53
Figura 18 Script del preprocesamiento a escala de grises.....	54
Figura 19 Script de filtrado de una imagen usando un filtro pasa bajo.....	55
Figura 20 Script de filtrado de una imagen usando un filtro de mediana de máscara 3x3.....	56
Figura 21 Script de filtrado de una imagen usando un filtro de mediana de máscara 3x3.....	56
Figura 22 Script de formulación de un umbral usando el criterio de Otsu (11).....	57
Figura 23 Script del proceso de umbralización.....	58
Figura 24 Script del proceso de binarización de la imagen.....	59
Figura 25 Script del proceso de morfología.....	59
Figura 26 Script del proceso de segmentación.....	60
Figura 27 Script del proceso de extracción de características.....	60
Figura 28 Script de la simulación de la red perceptron del toolbox de MATLAB 7.0.....	61

# Introducción

Las Tecnologías de la Informática y las Comunicaciones (TIC) han proporcionado al hombre una serie de herramientas de gran potencial para facilitar el desarrollo de sus actividades sociales, la mayoría de estas están soportadas por medios informáticos y redes de computadoras que permiten intercambiar un gran flujo de datos entre diferentes plataformas. Entre ellos están las medias que no son más que contenedores de información que portan imágenes, video y audio digital, cuya función no es únicamente informativa, recreativa o publicitaria, sino que también se comparan, argumentan y plantean soluciones, donde las imágenes digitales constituyen un componente esencial ya que son usadas tanto para darle veracidad al tema que aborda la media, como hacer más atractivo el contenido de las mismas a la vista de los usuarios, el concepto de imagen digital puede definirse como una matriz de  $M \times N$  elementos que se obtienen mediante la discretización de la imagen analógica, tanto en las dimensiones geométricas, mediante la generación de celdas por muestreo equiespaciado de la superficie, como en sentido radiométrico, mediante la cuantificación, cada elemento de la matriz se denomina píxel<sup>1</sup> **(1)**.

Dentro de los fotogramas se pueden encontrar 2 tipos de imágenes digitales: en escala de grises cuando se representa con una sola banda de color el valor de la intensidad de un píxel mediante una matriz de  $M \times N$  elementos numéricos, siendo este número la intensidad luminosa en un determinado punto o píxel, cuando se requiere de la combinación de 3 bandas de color (rojo, verde, azul) para representar el valor de la intensidad de un píxel, se está en presencia del modelo de colores RGB<sup>2</sup> o imagen en escala RGB, el mismo está formado por 3 matrices de  $M \times N$  elementos, siendo este número la intensidad luminosa en cada una de las bandas espectrales de cada punto o píxel **(2)**. Ambas son útiles para facilitar el trabajo con las herramientas de procesamiento de fotogramas y obtener una mayor calidad en los mismos en cuanto a brillantez, claridad y resolución. Con la digitalización de las imágenes ha sido posible mejorar la calidad del video, lograr un mayor realismo en la media del que se puede apreciar en una pantalla más grande, con una señal más robusta frente a interferencias, ruido, propagación, muy superior a la del video analógico. El video digital es el resultado de la combinación de secuencias de imágenes, audio, texto y movimiento que se obtienen por muestreo y cuantificación de la señal de video analógica que no es más

---

<sup>1</sup> del inglés picture element, "elemento de la imagen", menor unidad en la que se descompone una imagen digital.

<sup>2</sup> del inglés Red, Blue, Green, sistema de síntesis aditiva basado en rojo, azul y verde.

que la representación de la información como un flujo continuo de datos análogos para mostrarse en una pantalla de televisión compuesto por escenas, que representan un conjunto de tomas relacionadas entre sí **(3)**.

En los primeros años de Internet era muy difícil lograr difundir las medias por la web, o de lograrse se hacía con muy poca calidad, con el desarrollo y el aumento del ancho de banda de las redes de telecomunicación aumentaron considerablemente los sistemas que proveen servicios, similares a los de la televisión donde el usuario puede ver la información que se está transmitiendo en el instante que accede o puede optar por otras alternativas como interactuar con el flujo de datos recibido a través del video bajo demanda. El ámbito de las medias se ha convertido en uno de los campos más prometedores y de constante evolución, de manera creciente las compañías están entendiendo que dado el rápido incremento en el número de ellas en internet, la forma para aprovechar mejor esa oportunidad y obtener beneficios es convertirse en un contenedor que facilite el acceso a la misma. Una de las tantas alternativas que podría facilitar estos servicios sería brindar la opción de realizar una búsqueda mediante imágenes extraídas de la media, proporcionándole a los usuarios un criterio de búsqueda mucho más sencillo y encontrar el contenido que pueda adecuarse a sus necesidades específicas, la tecnología de búsqueda en medias se está convirtiendo en algo imprescindible debido a que son mayores las necesidades de los consumidores que tienen acceso a estos medios y dentro de este marco tecnológico, la digitalización y tratamiento de medias ha irrumpido con gran fuerza, sin embargo no es hasta hace poco que se ha superado las barreras computacionales que supone su manipulación y análisis.

En la Universidad de Ciencias Informáticas el Departamento de Señales Digitales es el encargado de brindar los servicios que contienen medias y garantizar la calidad de las mismas a través del área de Gestión y Procesamiento de medias, actualmente el contenido de la misma se procesa de forma automática mediante herramientas de edición con las cuales se extraen los fotogramas<sup>3</sup> a procesar para realizar operaciones de búsqueda con imágenes, resúmenes de información y acceder a lugares específicos dentro de ella. Este proceso de extracción de fotogramas requiere una continua determinación de la posición y orientación de los objetos en cada imagen dentro de una secuencia de imágenes, lo que resulta ser muy complicado debido a las distintas formas de objetos contenidos en las mismas, además la calidad de las imágenes se ve afectada por problemas de edición y ciertos factores inherentes al

---

<sup>3</sup> conjunto de imágenes individuales de una secuencia o animación.

procesamiento de las mismas como el ruido, el cual puede estar presente de dos formas: independiente y dependiente a los datos de la imagen.

El ruido independiente de los datos de la imagen es el que se puede describir mediante un modelo aditivo o filtro, como el ruido detector, donde cada píxel en la imagen es la suma del valor real del píxel más un valor de ruido aleatorio con distribución gaussiana<sup>4</sup>, un ejemplo de ello es el ruido gaussiano, el cual produce pequeñas variaciones en la imagen, generalmente se debe a ruido en los digitalizadores, perturbación en la transmisión, etc y se considera que el valor final del píxel sería el ideal más una cantidad correspondiente al error que puede describirse como una variable aleatoria gaussiana. Por otra parte el ruido dependiente de los datos de la imagen es aquel que se puede modelar mediante un modelo multiplicativo<sup>5</sup> o no lineal por lo que se asume en lo posible por la complejidad matemática de calcular el mismo que el ruido es dependiente de los datos, un ejemplo de ello es el “salpicado” (a menudo referido como ruido de picos de intensidad, de pecas o de sal y pimienta) donde la imagen es adquirida de manera directa en formato digital y ese mecanismo de adquisición o errores en la transmisión electrónica de sus datos son las fuentes de ruido. Ejemplo de este es el ruido impulsional, cuyo valor que toma el píxel no tiene relación con el valor ideal, sino con el ruido que toma los valores muy altos o bajos denominados (puntos blancos o negros).

Con la inexistencia de un método completamente factible en el Departamento productivo Señales Digitales para procesar correctamente los fotogramas, trae consigo que la gestión y procesamiento de medias no se realice de forma correcta, la media no tenga la calidad requerida en cuanto a los servicios de búsqueda y contenido y se almacene con errores. Entre las alternativas para una posible solución se encuentran las técnicas de detección de imágenes por bordes o contornos, estas proporcionan vías de gran potencial para que el procesamiento de fotogramas sea menos complejo y se realice de forma eficiente, mediante el reconocimiento y comparación de objetos con patrones, facilitando información de detalles que no están disponibles a simple vista en zonas de interés de manera dinámica, ayudando a mejorar la calidad de las imágenes para su posterior utilización o interpretación, por ejemplo remover defectos como el ruido, problemas por movimiento o desenfoque, mejoran ciertas propiedades de la misma como el color,

---

<sup>4</sup> distribución de probabilidad de variable continua que con más frecuencia aparece en fenómenos reales.

<sup>5</sup> método de pronóstico que estima los parámetros del modelo y los usa para generar el pronóstico.

contraste y estructura, lo que facilita el trabajo con herramientas de edición de poca calidad porque no dependen de la intensidad de la luz que incide en la señal para que sea visible.

Dado la **situación problemática** planteada anteriormente se evidencia la necesidad de reconocer imágenes por bordes o contornos en secuencias de videos de forma automática, con vista a darle solución al **problema** planteado queda definido como **objeto de estudio** las técnicas de reconocimiento de imágenes por bordes o contornos, enmarcado en el **campo de acción** de las técnicas de reconocimiento de imágenes por bordes o contornos en secuencias de video para el área de gestión y procesamiento de medias en las aplicaciones de reconocimiento de imágenes del Departamento productivo Señales Digitales.

El **objetivo general** de la investigación consiste en proponer una técnica para el proceso de detección de imágenes por bordes o contornos en secuencias de video para el Departamento productivo Señales Digitales y para darle cumplimiento se definieron las siguientes **tareas**:

- Caracterizar las técnicas de reconocimiento de imágenes por bordes o contornos en secuencias de video existentes.
- Identificar los problemas actuales en la aplicación de técnicas de reconocimiento de imágenes por bordes o contornos en secuencias de video en el Departamento productivo Señales Digitales.
- Evaluar el contenido de la información obtenida de la caracterización de las técnicas de reconocimiento de imágenes por bordes o contornos en secuencias de video.
- Diagnosticar las experiencias de la aplicación de técnicas de detección de imágenes por bordes o contornos en secuencias de video en Cuba y el mundo.
- Diseñar una propuesta de una técnica de detección de imágenes por bordes o contornos en secuencias de video.
- Validar la técnica propuesta a través del experimento.

Para darle cumplimiento a las tareas se emplearon los siguientes métodos científicos de investigación:

### **Métodos teóricos**

- **Analítico -sintético:** Ha permitido definir la propuesta de la técnica de detección de imágenes por bordes o contornos en secuencias de video, a partir de propiciar el conocimiento las teorías y documentos que ocupan el centro de la investigación extrayendo los aspectos más significativos relacionados con el objeto de estudio de la misma.
- **Inductivo-deductivo:** Me permitió llegar a un conocimiento global y esencial sobre las características de las técnicas de detección de imágenes por bordes o contornos encontradas.
- **Histórico-Lógico:** Sirvió como fuente de ayuda para una mejor comprensión de la evolución de las técnicas de reconocimiento de imágenes por bordes o contornos a lo largo del desarrollo del mundo de la visión computacional.

### **Métodos Empíricos**

- El **experimento** permitió validar la información de la técnica propuesta.

El trabajo consta de 3 capítulos donde se realiza un estudio crítico y descriptivo de las técnicas de reconocimiento de imágenes por bordes o contornos, además de las secciones de Conclusiones generales, Recomendaciones, Referencias bibliográficas, Glosario de términos y Anexos.

**Capítulo 1 Fundamentación teórica:** trata los temas relacionados con la fundamentación teórica que justifica la investigación, se analiza el estado actual del tema a nivel nacional e internacional así como las nuevas tendencias, tecnologías empleadas en la solución y otros temas de interés.

**Capítulo 2 Solución propuesta:** se expone todo lo referente a la solución propuesta, se realiza un estudio de su eficiencia mediante un diseño de experimentos en la herramienta MATLAB<sup>6</sup>.

---

<sup>6</sup> del inglés Matrix Laboratory, laboratorio de matrices.

**Capítulo 3 Resultados obtenidos:** muestra los resultados obtenidos después de la aplicación de la solución propuesta, aportando conclusiones respecto al uso de la misma así como recomendaciones para trabajos futuros.

Los Anexos incorporados al final del documento contienen elementos que apoyan la fundamentación teórica de la investigación así como una muestra de los resultados obtenidos y se definen claramente en el Glosario de términos los conceptos relacionados con el tema de la tesis u otro tema que sea de difícil comprensión.

# Capítulo 1 Fundamentación Teórica

Los temas vinculados con las técnicas de detección de imágenes por bordes o contornos proceden de diferentes campos informáticos, por lo que han surgido diferentes definiciones, problemas y resultados. En este capítulo se abordan los conceptos y temas relacionados al proceso de detección de imágenes por bordes o contornos, con el objetivo de ubicar el contexto en el que se enmarca la investigación, se hace un estudio detallado de las técnicas existentes en la actualidad en el ámbito nacional e internacional y se describen las mismas teniendo en cuenta las principales tendencias en el mundo de las Ciencias Informáticas para dar tratamiento a problemas similares a los de la investigación. En el área de procesado digital de fotogramas es de interés obtener una descripción de los objetos que forman las imágenes, este tipo de información es utilizada para la segmentación, seguimiento y reconocimiento de imágenes así como ayudar a mejorar la calidad de las mismas, sin embargo es complicado obtenerla por las diferencias que existen en las formas de los objetos y la complejidad de las técnicas utilizadas.

## 1.1 ¿Qué son las técnicas de detección de imágenes por bordes o contornos?

En los últimos años, varios sistemas prototipos se han propuesto abordar diferentes aspectos de la información contenida en las imágenes, como son la similitud de formas y las relaciones semánticas entre objetos de la imagen mediante los bordes y contornos de la misma. Los bordes y contornos de la imagen son características que ofrecen información detallada y relevante de la misma, estos suministran una valiosa descripción sobre las fronteras de los objetos y pueden ser utilizados para obtener una partición de la imagen digital en regiones coherentes como paso previo al análisis de su contenido y procesamiento.

Los bordes de una imagen digital son transiciones entre dos niveles de gris significativamente distintos, que forman zonas donde existe cambios de un valor a otro en las intensidades de la imagen que delimitan los objetos, el fondo y los contornos entre ellos proporcionando una descripción matemática de los objetos en las imágenes **(4)**. Los contornos son zonas de píxeles donde existe un cambio brusco de nivel de gris, representados como una colección ordenada de  $n$  puntos o curvas libres<sup>7</sup>, que iterativamente se aproximan a la forma del objeto, modelando las fronteras entre una imagen, el fondo y el resto de la

---

<sup>7</sup> se tratan los conceptos de curvas libres, flexibles o de seguimiento como el conjunto de curvas que forman el contorno global.

imagen mediante combinaciones, originando cualquier forma imaginable, delimitando las zonas de los objetos en dos partes, una acotada (el contorno y su interior) y otra infinita (el fondo), creándose un sub-mundo gráfico particular en cada forma definida resaltando las propiedades de la imagen. La línea base de un contorno define la complejidad de éste y sus propiedades, cuando una línea se cierra sobre sí misma el contorno creado determina un espacio interno, creándose una tensión entre este espacio y sus límites, otorgando a la línea creadora un gran poder de atracción visual siendo esta la base de la detección por contornos para el procesamiento de imágenes, ofreciendo soluciones que no requieren un procesamiento posterior y son directamente interpretables lo que garantiza que si se emplea el modelo correcto la presencia de falsos negativos o falsos positivos sea pequeña **(5)**. Con el incremento del tamaño de las colecciones de imágenes digitales surge una necesidad básica: la búsqueda y obtención de información relevante para el procesamiento de las mismas.

Los métodos tradicionales para la obtención de este tipo de información se muestran ineficientes e inadecuados por dos motivos: la cantidad de trabajo que requiere asociar, identificar o seguir un objeto en una imagen no es fácil y además no es subjetivo, por otra parte la imprecisión a la hora de detectar las formas de los objetos, muchas aplicaciones requieren que la representación de los objetos sea invariante a la traslación, rotación y escalado mientras que otras no. Para la solución a estos problemas se crearon las técnicas de reconocimiento de imágenes por bordes o contornos, que no son más que los algoritmos encargados de encontrar y seguir objetos contenidos en las imágenes a través de los bordes y contornos de la misma, aportando información de interés de los mismos mediante criterios de búsquedas capaces de manejar imprecisiones y falta de información en las descripciones propuestas por el observador, permitiendo describir el contenido en las imágenes sobre la base de las características y/o representaciones de interés con la pretensión de abordar el diseño para los sistemas de reconocimiento de imágenes y mejorar la calidad de las mismas.

## **1.2 Estado del arte de las técnicas de detección de imágenes por bordes o contornos**

Los inicios de las técnicas de reconocimiento de imágenes vienen unido a la graficación por computadora, esta es una de las áreas más interesantes de las ciencias de la computación cuyo objetivo principal es establecer los principios, técnicas y algoritmos para la generación y manipulación de imágenes mediante

un ordenador, permitiendo la interacción de los mismos con el hombre. Uno de los mayores problemas de la graficación y el reconocimiento de imágenes ha sido la subjetividad con que el ser humano analiza dichas imágenes, en muchos casos la experiencia profesional es un factor determinante en el resultado obtenido, pero normalmente se requiere un aprendizaje bastante sistemático para alcanzar la fiabilidad, sin embargo en los casos que el profesional tiene un grado elevado de experiencia, en muchos tipos de imágenes la detección de bordes y contornos resulta ser una tarea muy complicada debido a la escasa calidad en las imágenes, ocultación de los objetos contenidos en las mismas, propiedades difícilmente visibles al ojo humano debido a su reducido tamaño (micro-estructuras), subjetividad en el proceso de interpretación de la imagen tras el cansancio, desconcentración e inexistencia de una búsqueda sistemática y exhaustiva, unido a esto la complejidad de los algoritmos relacionados a estos procesos y el alto costo computacional de los mismos y la poca capacidad de graficar, procesar y realizar cálculos en imágenes de una manera más sencilla.

Con la evolución de los componentes electrónicos y sistemas digitales además de los dispositivos computacionales, la habilidad de generar y procesar imágenes por medio de una computadora se ha hecho una necesidad y en la actualidad las técnicas de reconocimiento de imágenes por bordes o contornos son una herramienta eficaz en cuanto a graficación, creación y procesamiento por su manera de realizar cada uno de estos procesos de una forma más simple, rápida y económica. Se ha venido utilizando diferentes aproximaciones a la temática del procesamiento de imágenes, en las que es posible distinguir la búsqueda de primitivas visuales de bajo nivel (incluyendo características como texturas y color), pasando por estrategias de aprendizaje y agrupamiento jerárquico e incluyendo la capacidad de manipular objetos genéricos en contextos y configuraciones no predeterminadas.

La visión por computadoras<sup>8</sup> está estrechamente relacionada con el procesamiento de imágenes aunque existe una gran diferencia entre dos enfoques. El procesamiento de imágenes trata sobre cómo mejorar una imagen para la interpretación visual de una persona, pero la visión computacional es una rama de la inteligencia artificial que tiene como objetivo modelar matemáticamente los procesos de percepción visual en los seres vivos y generar programas que permiten simular estas capacidades visuales por medio de los ordenadores, esto facilita realizar a partir de secuencias de imágenes de la escena una descripción de que

---

<sup>8</sup> visión para reconocer y localizar objetos mediante el procesamiento de imágenes por una computadora.

es útil para el observador y que está libre de información irrelevante, tomar decisiones útiles acerca de los objetos físicos reales del mundo de escenas en base a imágenes adquiridas digitalmente, mediante la construcción de descriptores de la escena a través de herramientas informáticas, en base a características relevantes contenidas en una imagen **(6)**.

Es posible distinguir tres tipos de descriptores que permiten un análisis de las imágenes a tres niveles con diferente capacidad de abstracción y que son:

- Nivel 1 primitivas básicas: En este caso la posibilidad de establecer relaciones entre imágenes se realizan en base a valores de color de los puntos, a las texturas que se pueden observar y a los bordes o contornos que son posibles de detectar con diferentes operadores **(7)**.
- Nivel 2 sintáctico: Denominado “reconocimiento de patrones” por aquí se unen a las anteriores informaciones para buscar regiones de mayor entidad y comprobar si se ajustan a reglas de descripción de elementos tipo **(7)**.
- Nivel 3 semántico: Llamado “reconocimiento de objetos” en que la complejidad es un poco mayor al permitir que existan imprecisiones entre la definición del objeto de la realidad a buscar y las posibles transformaciones en el paso a dos dimensiones, las variaciones en las condiciones de adquisición de la imagen, entre otros **(7)**.

Las técnicas de reconocimiento de imágenes basadas en visión computacional tradicionalmente realizan un pre-proceso en la imagen, seguido de un proceso de segmentación para separar los objetos o regiones de interés del fondo de la imagen, un proceso de extracción de características encaminado a obtener un conjunto de descriptores y utilizar estos para mejorar la calidad de la imagen y extraer los objetos y regiones de interés de la misma y por último realizar el reconocimiento de la imagen con las características relevantes tomadas de la imagen. En la adquisición de estas existen factores físicos difíciles de controlar, entre los que se pueden citar los niveles bajos de iluminación, la reflexión sobre los objetos y el ruido aleatorio, que hacen que las imágenes no presenten siempre una buena calidad para su procesamiento, por estas razones es necesario un mejoramiento de la imagen utilizando un preprocesamiento, este no es más que los procedimientos de manipulación y análisis de la información contenida en una imagen digital por medio de un computador, con el objetivo de corregir defectos y

mejorar de la calidad visual de la misma como preparación a un procesamiento adecuado para poder obtener regiones segmentadas y garantizar que estas sean candidatas a ser objetos o regiones de interés dentro de la imagen.

El preprocesamiento en imágenes puede realizarse a través de técnicas de reducción del área de trabajo, reducción de ruido y realce de contrastes.

- Las técnicas de reducción del área de trabajo consisten en simplificar la imagen a la región que corresponda al objeto de interés, para reducir tiempo de procesamiento de las etapas posteriores y eliminar objetos que no aportan información al proceso.
- Las técnicas de reducción de ruido consisten en técnicas de filtrado y suavizado de imágenes en un entorno local de un píxel determinado.
- Las técnicas de realce de contraste se basan fundamentalmente en aumentar la separación del nivel de gris del fondo y los objetos.

Para obtener buenos resultados en la fase del preprocesamiento se recomiendan las técnicas de reducción de ruido debido a que las mismas se llevan a cabo a través de filtros que retienen tanto como sea posible las características de la imagen, permitiendo que en etapas posteriores como la de extracción de características y clasificación tengan mayores porcentajes de acierto **(8)**. Las técnicas que tratan de suprimir el ruido se denominan técnicas de mejora de la imagen, entre ellas se encuentran las técnicas de procesamiento local, las cuales son ampliamente utilizadas debido a que trabajan sobre la imagen realizando cálculos sencillos en un entorno local de un píxel determinado, dentro de este grupo de técnicas se hallan las máscaras o filtros espaciales. El filtrado es una técnica para modificar o mejorar la imagen resaltando o atenuando algunas características de la misma, es una operación de vecindad porque calcula el valor del píxel dado en la imagen procesada mediante un algoritmo que toma en cuenta los valores de los píxeles de la vecindad de la imagen original.

Se han definido una amplia variedad de filtros para diversas aplicaciones que se agrupan en las siguientes categorías: filtros lineales y no lineales. Los filtros lineales actúan calculando un nuevo valor para cada

píxel a partir de los coeficientes<sup>9</sup>  $w$  de la máscara o filtro y de la información de nivel de gris del entorno del píxel a procesar, cada uno de estos coeficientes se multiplica por el píxel vecino correspondiente para hallar, finalmente, el nuevo valor del píxel  $z$  después de sumar todos los términos como se muestra en la siguiente ecuación, donde  $mn$  son el número total de coeficientes de la máscara,  $w_i$  son coeficientes de la máscara y  $z_i$  son los valores de niveles de gris de la imagen correspondientes a estos coeficientes **(9)**.

$$z = w_1z_1 + w_2z_2 + \dots + w_mz_n = \sum_{i=1}^{mn} w_i z_i$$

Entre los filtros lineales más importantes se encuentra el filtro pasa bajo, el cual corresponde a un filtro caracterizado por permitir el paso de frecuencias más bajas y atenuar las frecuencias más altas de la imagen en el dominio de Fourier<sup>10</sup>, los componentes de baja frecuencia de la imagen son los que determinan la intensidad media y el contraste global de la imagen. Este tipo de filtro se utiliza para la eliminación de ruido como el gaussiano, pero conllevan a una disminución de la información de los bordes de la imagen. Los filtros no lineales no usan explícitamente el coeficiente de la manera descrita en la ecuación 1, sino que realizan operaciones sobre un entorno del píxel.

Llevar la imagen a procesar a escala de grises se considera también parte del preprocesamiento para hacerlo más simple, obteniendo una imagen de mejor calidad donde se pueden tener más detalles acerca de los bordes y contornos de la misma. En computación una escala de grises es una escala empleada en las imágenes digitales en las que el valor de cada píxel posee un valor a una graduación de la mezcla de blanco y negro (gris) en diferentes proporciones, dependiendo de la cantidad de luz, el ojo humano puede interpretarlo como otro color, cuando las imágenes se representan en escalas de este tipo, están compuestas por sombras de grises, que van desde el color negro más profundo variando gradualmente en intensidades de grises hasta llegar al color blanco **(10)**.

Este tipo de análisis es efectivo para conjuntos de imágenes en las que los descriptores satisfacen cualquier consulta permitida, pero todas estas se caracterizan porque el motivo de las mismas es único, ocupan gran parte de la escena, es fácilmente distinguible del fondo y se pueden normalizar los valores de intensidad, tamaño de objeto o su posición, pero no es suficiente en aquellas colecciones de imágenes

---

<sup>9</sup> básicamente una máscara es una matriz 2D pequeña, cuyos coeficientes determinan la naturaleza del proceso que ésta lleva a cabo.

<sup>10</sup> función que hace corresponder a una función  $f$  con valores complejos y definidos en la recta, con otra función  $g$ .

con grandes riquezas de detalles que no es representable de forma compacta por los descriptores cuyas características son: la falta de control sobre las condiciones con que se obtienen las imágenes (posición de la cámara, movimiento, iluminación, etc.), la gran variabilidad de las características de los objetos de interés (color, tamaño, posición, entre otros), y la existencia de partes ocultas, por lo que se requiere un proceso de segmentación de la imagen en dependencia de la forma de los objetos dividido en 3 etapas fundamentalmente: umbralización, binarización y operaciones morfológicas.

La umbralización es una técnica de segmentación ampliamente utilizada para el reconocimiento y procesamiento de imágenes, se emplea cuando hay una clara diferencia entre los objetos de interés a extraer de la imagen respecto al fondo de la escena. La mayoría de las técnicas de umbralización se basan en estadísticas sobre el histograma<sup>11</sup> unidimensional de una imagen, los cuales constituyen las bases para el desarrollo de numerosas técnicas de procesamiento de imágenes ya que proveen estadísticas, el realce de las mismas y la determinación de umbrales para su binarización. Para la localizar los umbrales se pueden usar procedimientos paramétricos y no paramétricos **(11)**:

- En los procedimientos paramétricos, la distribución de los niveles de gris de una clase de objeto lleva a encontrar umbrales.
- En los procedimientos no paramétricos los umbrales se obtienen de una forma óptima de acuerdo a algún criterio.

Si bien hay diferentes métodos para hallar un umbral, la mayoría de ellos no proporcionan buenos resultados cuando se trabajan con imágenes del mundo real, debido a la presencia de ruido, histogramas pobres o una iluminación inadecuada, es por esto que la mayor parte de los algoritmos para reconocer imágenes están escritos a partir de imágenes binarias, lo que permite reducir el volumen de datos a tratar en una imagen en escala de grises o una a color. La binarización de una imagen digital consiste en convertir la imagen digital en blanco y negro, de tal manera que se preserven las propiedades esenciales de la imagen, este paso es muy importante puesto que se trata de una primera segmentación de la imagen, con la cual se pueden extraer los objetos respecto al fondo de la escena.

---

<sup>11</sup> función que muestra, para cada nivel de gris, el número de píxeles de la imagen que tienen ese nivel de gris.

La morfología<sup>12</sup> matemática es una herramienta que se utiliza para la simplificación de los datos de la imagen, preservar sus características esenciales o eliminar aspectos irrelevantes de la misma, teniendo en cuenta la identificación y descomposición de los objetos, la extracción de rasgos y la localización de defectos que estén relacionados con las formas de los mismos, supresión de ruido, simplificación de formas y destacar estructuras de objetos dentro de las imágenes. En este proceso existen dos operaciones fundamentales, donde la mayoría de los algoritmos morfológicos están basados en las mismas: dilatación y erosión. La dilatación es una de ellas y se utiliza principalmente para resaltar el tamaño del objeto de interés en la imagen, sin embargo la erosión generalmente disminuye el tamaño de los objetos que no son de interés en la imagen **(12)**.

La tarea de encontrar correspondencias entre imágenes u objetos de diferentes escenas es parte de muchas aplicaciones informáticas basadas en visión computacional, el registro de imágenes y reconocimiento de objetos son solo algunas, en estas tareas juegan un papel muy importante los detectores de características de la imagen que no son más que descriptores basados en propiedades de las mismas hacia regiones u objetos de interés previamente definidos en un proceso de segmentación. La extracción de características requiere de detectores que definan bien los “puntos de interés” seleccionados tales como bordes, texturas y contornos, el detector debe contar con la repetitividad para reconocer los distintos puntos de interés bajo diferentes condiciones de visión, debe ser capaz de realizar el barrido de cada punto de forma distinta y al mismo tiempo robusto al ruido, que permita la detección de errores y la deformación geométrica en la imagen **(13)**. Existen una gran variedad de detectores que han sido propuestos para el proceso de extracción de características, pero con el objetivo de abordar acerca de los más completos y realizar una comparación en cuanto a la reducción de tiempo de procesamiento y complejidad se hace referencia a los más utilizados.

El detector *SIFT*<sup>13</sup> es un algoritmo de visión por computador, para detectar y describir características locales en las imágenes mediante la localización de puntos distintivos, en las mismas, calculando el descriptor de los mismos. Fue inicialmente usado en tareas de reconocimiento teniendo como principal ventaja que los puntos de interés extraídos de los mismos son invariantes a escala, rotación y parcialmente invariantes a cambios de punto de vista e iluminación, extrayendo las características a partir

---

<sup>12</sup> disciplina matemática que estudia la generación y las propiedades de la forma.

<sup>13</sup> del inglés Scale Invariant Feature Transform, función de la escala-invariante a transformar.

de una pirámide de imágenes con muestreo entre cada nivel. Los puntos clave de los objetos primero se extraen de un conjunto de imágenes de referencia y se almacena en una base de datos, luego el objeto es reconocido en una nueva imagen por separado comparando cada característica de la nueva imagen a esta base de datos, realizando la búsqueda de candidatos por características basadas en la distancia euclídea de la función de sus vectores **(14)**.

Sin embargo el algoritmo *SURFT*<sup>14</sup> es un detector invariante a rotación y a escala, con elevada precisión en la extracción de características y su bajo coste computacional, supera a otros métodos existentes de acuerdo a la repetitividad para el reconocimiento de puntos, robustez y distinción de los descriptores donde se puede realizar un cálculo y comparación mucho más rápido. El principio de funcionamiento del algoritmo es igual que el *SIFT* pero con una mayor complejidad, realizando la extracción de puntos fuertes y orientaciones del gradiente para reducir el efecto de las variaciones geométricas de la imagen **(15)**.

Para resolver problemas más complejos como la detección de múltiples objetos o regiones de interés en imágenes son usados estos detectores en una colección de datos que mediante sus características definen su estructura y propiedades y son representados en forma de tabla (dataset). La mayoría de estos algoritmos de extracción de características requieren como paso preliminar, un costoso proceso de búsqueda y reconocimiento de patrones en imágenes por diferentes escalas en diferentes posiciones por lo que se basan en redes neuronales artificiales, estas no son más que procesadores distribuidos en paralelos que poseen una tendencia natural para almacenar conocimiento de forma experimental y lo hace disponible para su uso, el mismo es adquirido a través de un proceso de aprendizaje y de una forma jerárquica intenta interactuar con los objetos para reconocer patrones de la misma forma que lo hace el sistema nervioso biológico **(16)**.

Dentro de las redes neuronales, las más utilizadas son la redes perceptron multicapas, compuesta por una serie de nodos que componen la capa de entrada, un conjunto de una o más capas ocultas de neuronas que no son más que las capas donde se genera cualquier tipo de representación interna de información, pero que no tenga relación directa con la información de entrada ni con la salida y una capa de neuronas de salida, donde la señal de entrada se propaga hacia delante desde la capa de entrada hasta la oculta

---

<sup>14</sup> del inglés Speeded Up Robust Features, acelerar características robustas.

hasta la salida. Este tipo de configuración se conoce como MLP<sup>15</sup>, que consiste en una red neuronal artificial modelo donde las conexiones entre las unidades no forman un ciclo dirigido y que establece los mapas de datos de entrada en un conjunto de salida correspondiente, el hecho de que se utilice para resolver con éxito una multitud de problemas se debe a que esta utiliza el algoritmo de aprendizaje que actualmente está más extendido conocido como *back propagation*, por lo que se basa también a la corrección del error. Básicamente este algoritmo consiste en dos pasadas a través de las diferentes capas de la red, una hacia adelante y otra hacia atrás, en la pasada hacia adelante se aplica en la capa de entrada un patrón o vector de entrada, este propaga su efecto a través de las diferentes capas y como consecuencia produce un vector salida donde no se modifican los pesos sinápticos<sup>16</sup> de las neuronas, sin embargo durante la pasada hacia atrás de acuerdo con la regla de corrección del error mediante la comparación de la señal de salida con la señal deseada y como resultado se obtiene una señal de error que se propaga en sentido contrario a través de la red modificando sus pesos, de forma tal que al pasar nuevamente el vector entrada hacia delante, la respuesta obtenida se asemeje más a la salida esperada. Este algoritmo es un método eficiente para entrenar este tipo de redes, aunque no es capaz de resolver todos los problemas, su importancia está en la capacidad de auto-adaptar los pesos de las neuronas intermediarias para aprender la relación que existe entre los conjuntos de vectores o patrones de entrada y su correspondiente salida y poder aplicar esa relación después del entrenamiento a nuevos vectores de entradas imperfectos o con ruido **(16)**.

La red perceptron multicapas posee tres características fundamentales:

- El modelo de cada neurona incluye una función no lineal, continua y derivable donde  $uk$  es la suma total de la actividad interna en la neurona  $k$  (la señal de entrada) y  $yk$  la salida que se produce en la neurona.

$$yk = \frac{1}{1 + \exp(-uk)}$$

---

<sup>15</sup> del inglés Multiplayer Perceptrons, perceptron multicapa.

<sup>16</sup> peso de su interconexión con otra neurona.

- La red contiene una o más capas ocultas de neuronas que no forman parte ni de la entrada ni de la salida, las cuales capacitan a la red para aprender progresivamente cualquier correspondencia entre la entrada y la salida, y almacenar internamente la información.
- La red poseen un gran número de conexiones, estas vienen determinadas por los pesos de la red, lo que un cambio en las conexiones equivale un cambio en los pesos.

La combinación de estas características, hace que la habilidad de la red para aprender sea muy potente, en este tipo de tarea hay un número fijo de categorías en las cuales las muestras de entrada deben clasificarse, para ello se requiere una fase de entrenamiento en la que se le presenta a la red los patrones que debe aprender y las categorías en cual clasificarlos, para después presentarle otro patrón de prueba desconocido pero que pertenezca a algunas de las categorías aprendidas y esta debe decidir a cual se parece más. Este comportamiento hace que sea difícil conocer a priori la respuesta de la red debido a dos motivos: el comportamiento lineal de las neuronas las cuales están muy interconectadas lo que hace difícil un análisis teórico de la red y la existencia de neuronas ocultas que impide “ver” el cómo se lleva a cabo el proceso de aprendizaje y determinar cuáles son las características que ayudan a mejorar el mismo. La ventaja de usar redes neuronales esta en el hecho de que se pueden separar regiones de decisión tan complicadas como se deseen, dependiendo del número de neuronas y capas que contenga la red lo que ayuda a resolver problemas de reconocimiento y clasificación de regiones u objetos de interés de alta complejidad (16).

### **1.3 Análisis de las técnicas existentes**

La detección de imágenes por bordes o contornos en imágenes digitales es un campo de investigación relativamente joven en Cuba y el mundo, aunque está en constante desarrollo, no es hasta hace algunos años que se supera la barrera de complejidad computacional que supone el procesamiento de imágenes, propiciando esto la proliferación de novedosas técnicas en este marco tecnológico. Actualmente se pueden identificar dos tendencias fundamentales en las técnicas vinculadas a la detección de imágenes digitales por bordes o contornos: los métodos basados en características y los métodos basados en modelos deformables. Los métodos basados en características son operaciones basadas en la identificación, seguimiento y usualmente en esquemas de matching de las características de la imagen

tales como puntos, líneas, curvas, borde y contornos, a través de filtros<sup>17</sup>, algoritmos o máscaras, los cuales resaltan determinadas características de la imagen, modifican el aspecto de la misma o eliminan el ruido, obteniendo como resultado una nueva imagen o “mapa característico”. Estos mapas son el comienzo del proceso de reconocimiento de imágenes por bordes o contornos, ellos mejoran las características de la imagen pero no proporcionan un agrupamiento lógico de los elementos de la imagen que se mueven lo que son altamente sensibles a la oclusión<sup>18</sup>, el proceso de detección de características requiere de decisiones a tomar en cada píxel analizado y a bajo nivel es efectivo en un solo punto, lo que no surte efecto por ejemplo a la hora de identificar estructuras geométricas **(17)**.

Los métodos basados en modelos deformables se basan en el seguimiento de los contornos de las imágenes en movimiento en una secuencia de imágenes mediante el análisis visual de las imágenes esperadas, estas técnicas extraen los contornos de interés basándose en modelos que utilizan información geométrica de la forma de los objetos, son mucho más robustas frente a la presencia de ruido y otros elementos que afectan la calidad de la imagen o dificultan el proceso de detección, permiten segmentar imágenes más complejas y sus soluciones proporcionadas no requieren procesamiento posterior y son directamente interpretables. Estos métodos están diseñados por procesos enfocados a funciones específicas, basadas en visión activa, donde los datos de la imagen son obtenidos en una manera sistemática para asegurar que las características de los objetos están presentes en la información, pero muchos de ellos permanecen sin interpretación hasta que una tarea específica los involucra tomando como salida una colección de imágenes intrínsecas.

Este enfoque trae consigo dos grandes ventajas, primeramente ya no es necesario agrupar características de la imagen ya que simplemente el área relevante de la imagen es comparada con la apariencia que se desea, por otra parte como se realiza un análisis restringido a una “región de interés” disminuye el costo computacional del sistema. La visión activa tiene como conocimiento a priori el conocimiento geométrico de los tipos u objetos de las imágenes, que permite que en vez de esperar que propiedades geométricas deseables de los contornos provengan de los datos de la imagen estas propiedades son impuestas desde el principio mediante un modelo elástico de curvas flexibles que se ajusta a los datos de la imagen

---

<sup>17</sup> convolución de una matriz con respecto a un píxel y la vecindad de este.

<sup>18</sup> ambigüedad a la hora de darle seguimiento a las características de la imagen.

facilitando la interpretación visual de imágenes, estableciendo las técnicas para dicho análisis, sintetizar y enfocar el procesamiento que es aplicado a cada imagen dentro de la secuencia de imágenes (17).

### 1.3.1 Técnica de Canny

Dentro de los métodos basados en características de la imagen está la técnica de Canny, algoritmo considerado uno de los mejores métodos de detección de contornos por su gran adaptabilidad para ser aplicado a diversos tipos de imágenes, además de no disminuir su calidad ante la presencia de ruido en la imagen original. Sin embargo se debe tener cuidado en no aplicar un suavizado excesivo a la imagen, puesto que se pudieran perder detalles de la imagen y provocar un pésimo resultado final. Esta técnica mediante el empleo de máscaras de convolución y en el cálculo de la primera derivada realiza el tratamiento de píxeles de imágenes de forma discreta para trabajar dichas máscaras, toma el valor de todas las regiones donde no varía la intensidad, por lo que un cambio de intensidad representa un cambio brusco en la primera derivada, y así se detecta un borde.

Para aplicar la técnica de Canny se debe seguir los siguientes pasos:

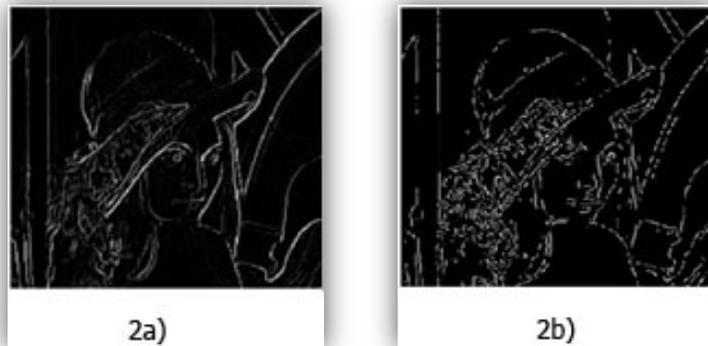
- Obtención del vector gradiente.
- Proceso de supresión no máxima.
- Proceso de histéresis de umbral.

La obtención del gradiente no es más el cálculo de la magnitud y orientación del vector gradiente en cada píxel, para ellos se aplica un filtro gaussiano a la imagen original (figura 1a) con el objetivo de suavizar la imagen y tratar de eliminar el posible ruido existente, una vez terminada de suavizar la imagen se logra para cada píxel la magnitud y módulo del gradiente obteniendo así dos imágenes como lo muestra la figura 1b y 1c.



**Figura 1 Proceso de obtención del gradiente (4).**

El proceso de supresión no máxima consiste en lograr el adelgazamiento del ancho de los bordes obtenidos por el gradiente hasta lograr bordes de un píxel de ancho. Las dos imágenes generadas en la obtención del gradiente sirven de entrada para generar una imagen con los bordes adelgazados como se muestra en la figura 2a, para ello se considera encontrar la dirección para cada píxel que mejor se aproxime a la dirección del ángulo de gradiente, posteriormente se observa si el valor de la magnitud del gradiente es más pequeño que al menos uno de sus dos vecinos en la dirección del ángulo obtenida, de ser así se le asigna el valor 0 que significa píxel bloqueado, o sea que se desecha dicho píxel en la supresión no máxima, en caso contrario se le asigna el valor que tenga la magnitud del gradiente (ver figura 2b).



**Figura 2 Supresión no máxima (4).**

El último paso sería el proceso de histéresis de umbral, el mismo se utiliza para reducir la probabilidad de aparición de contornos falsos ya que la imagen obtenida en el paso anterior puede contener máximos locales creados por el ruido. El proceso de histéresis (ver figura 3a) consiste en tomar esta imagen, la orientación de los puntos de borde de la imagen y dos umbrales, el primero más pequeño que el segundo, para cada punto de la imagen se debe localizar el siguiente punto de borde no explorado que sea mayor al segundo umbral y a partir de dicho punto seguir las cadenas de máximos locales conectados en ambas direcciones perpendiculares a la normal del borde siempre que sean mayores que el primer umbral. Así se marcan todos los puntos explorados y se almacena la lista de todos los puntos en el contorno conectado logrando eliminar los posibles máximos locales producidos por el ruido como lo muestra la figura 3b.

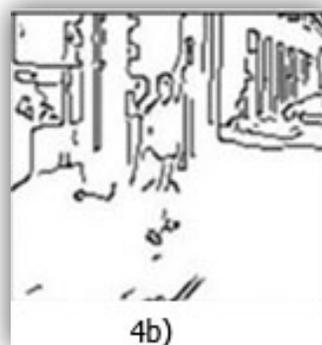


Figura 3 Histéresis de umbral (4).

### 1.3.2 Matching Jerárquico de Chamfer

El matching por contornos usando el algoritmo de Chamfer es otra técnica que forma parte de los métodos basados en características, su principio se basa en igualar patrones (ver figura 4a) de manera aproximada y con un alto grado de acierto, efectividad y velocidad de procesamiento, realizando el matching de los puntos de bordes (ver figura 4b) u otras características de bajo nivel de un patrón, verificando si este está presente en la imagen digital, logrando no solo el mejor calce en la imagen sino una medida de confianza en el proceso de detección.

La técnica se basa en la búsqueda del o de los mínimos factor distancia entre cada posición de la imagen y el template<sup>19</sup> o plantilla, con la ventaja de que no hace una correlación con todos los puntos del template en cada posición del área de búsqueda, sino con los puntos de bordes del plantilla, con los cuales calcula un factor distancia entre la transformación distancia de los bordes de la imagen, iniciando el proceso de matching en una resolución piramidal de la imagen para obtener a partir de la imagen original otras idénticas con menos resolución a medida que se sube de nivel, el proceso de detección por bordes o contornos anexado a la resolución piramidal permite ahorrar tiempo en el procesamiento de las imágenes, a medida que la cantidad de template aumenta se pueden procesar mas imágenes por unidad de tiempo, luego las plantillas son agrupados basándose en las características similares existentes entre ellos con la finalidad de minimizar la cantidad de patrones a utilizar en el proceso y con esto el tiempo de procesamiento, la utilización de este agrupamiento en el proceso de matching hace necesario crear un template prototipo (ver figura 4c) para cada grupo, el cual contiene las características originales perteneciente al mismo, se vuelve a iniciar el proceso de matching con los template prototipos si uno de estos provoca un calce entonces se procede a realizar este proceso con cada una de las plantillas pertenecientes al grupo hasta que se tenga éxito en el matching o en el peor de los casos se descarte esa solución y se procede al proceso de matching nuevamente hasta encontrar los bordes de la imagen **(18)**.



---

<sup>19</sup> patrones de identificación que contienen factores de similitud entre regiones de la imagen y el área de investigación.



**Figura 4 Matching Jerárquico de Chamfer (18).**

En el reconocimiento de patrones como lo muestra la figura 4d buscando las características radiométricas relacionadas con sus vecinos próximos, la heurística de matching Jerárquico de Chamfer es un método de correlación eficaz, sin embargo su principal desventaja radica en cómo correlacionar o aprender apropiadamente los diferentes umbrales para cada uno de los niveles de profundidad de la resolución piramidal para una búsqueda de patrones exitosa, así como el cálculo de los valores de la distancia de Chamfer dependen del tamaño de la imagen que se emplea como entrada en el proceso de detección, obteniendo valores diferentes si el factor es calculado en una imagen completa contra una región de ella en una misma región, el formato y resolución de las imágenes es importante ya que de esto depende que el proceso de detección sea eficiente.

### **1.3.3 Contornos activos**

El nombre de contorno activo se debe a la teoría física de la elasticidad, estos son considerados como cuerpos elásticos que responde a fuerzas aplicadas bajo ciertas restricciones, su geometría permite una amplia cobertura de formas que emplea diversas representaciones geométricas que involucran varios grados de libertad (curvas y superficies). Partiendo de su formulación matemática se distinguen principalmente dos clasificaciones de contornos activos, los paramétricos y los geométricos. Los contornos activos paramétricos son representados en forma explícita por funciones de contorno parametrizadas (curvas o superficies) para su deformación, la cual se rige por propiedades intrínsecas, formuladas explícitamente (19), los objetos o regiones de interés resultantes al aplicar este modelo son directamente interpretables (no requieren procesamiento posterior), cada contorno activo paramétrico permite definir una

sola región u objeto de interés por lo que se requiere un contorno inicial para cada uno, donde el tiempo de procesamiento dependerá directamente del número de puntos del contorno.

Los contornos activos geométricos son representados implícitamente como conjuntos de nivel de una función dinámica con una dimensión más para su deformación, la cual depende explícitamente de sus propiedades geométricas **(20)**, provee una segmentación indistinta para una o varias regiones u objetos de interés, no existen problemas con el número de puntos del contorno, pero su tiempo de procesamiento depende directamente del tamaño de la imagen. A continuación se ofrecen la descripción de los contornos activos o técnicas de reconocimientos de imágenes por bordes o contornos basadas en modelos deformables existentes.

## Snake

La técnica snake es un contorno activo paramétrico, definido por una curva spline minimizadora de energía, guiada por fuerzas<sup>20</sup> restrictivas externas e influenciada por fuerzas de la imagen que tienden a localizarse en características de ésta como líneas y bordes, evolucionando de manera dinámica hacia los contornos relevantes de la imagen. Esta técnica en principio no está determinada para la búsqueda automática de contornos prominentes en las imágenes, sino para la refinación de la solución propuesta por otra serie de mecanismos, o sea hace evolucionar el contorno obtenido por dicho método hasta lograr el mínimo local apropiado, las snake son mecanismos para dar cierto grado de conocimiento a priori a la interpretación de la imagen a bajo nivel, en el lugar de esperar que las propiedades deseables de los contornos como son continuidad y suavidad de los datos de imagen provengan de los datos de imagen, estas propiedades son impuestas desde el principio mediante un modelo elástico de curva flexible que posteriormente se ajusta a los datos de la imagen, variando los parámetros de la elasticidad de la curva, con lo que se puede controlar la cantidad de información geométrica que se asume.

Su principio se basa en la localización de bordes, líneas y contornos subjetivos, así como en el seguimiento de dichos contornos durante el movimiento de forma variada en la imagen, en esta técnica tanto la conectividad de los contornos como la presencia de los bordes en la imagen van a afectar al

---

<sup>20</sup> en terminología de contornos activos la fuerza viene dada por el gradiente de la energía que sirve para poner restricciones de suavidad, es decir para regularizar la solución.

funcional de energía y por lo tanto afecta los detalles de la estructura del contorno localmente óptimo, por lo que el snake permite la interacción entre mecanismos de alto nivel introducidos por el usuario con el modelo de contorno para llevarlo hacia el mínimo local apropiado. Para que este contorno activo sea útil es necesario definir funcionales energéticos que lo atraigan hacia las características relevantes de la imagen (líneas y bordes), según como se ajusten estas características se puede modificar el comportamiento del mismo con respecto a la imagen y obtener un mejor resultado **(17)**.

Uno de los potenciales energéticos más sencillos es la intensidad de la imagen por sí misma que está determinado por la energía de las líneas, en caso de que la imagen tenga la misma energía en varias líneas el snake dependiendo del valor de la intensidad de la imagen sujeto a las otras restricciones se verá atraído a los contornos cercanos a regiones claras u oscuras de la imagen. Para detectar los bordes de la imagen con mayor facilidad se eleva el módulo del gradiente de la imagen con signo negativo, así el Snake se enfoca a las zonas de elevado gradiente en la imagen y evitar de cierta forma caer en mínimos locales dado por estructuras como el ruido. Para ello se elige una escala, se desenfoca la imagen para eliminar las estructuras a menor escala antes de determinar el gradiente de la imagen, si la escala es muy grande el contorno activo se verá atraído por los bordes de una forma muy aproximada desde posiciones bastantes alejadas, a escalas pequeñas se verá atraído a las micro-estructuras o ruidos más cercanos, pero no podría alcanzar los bordes reales de las regiones a escalas mayores, es por eso que se comienza este proceso a escalas mayores para obtener con el mismo una primera solución próxima y después se disminuye la escala para ir refinando la solución. La técnica snake es una nueva forma de contorno activo que extiende el área de aplicación de estos modelos, su ventaja radica en la capacidad de hacer uso del conocimiento a priori incierto, así como una descripción verbal de las características del objeto por un humano experto, teniendo en cuenta una descripción más detallada del objeto, cuyo acercamiento puede mejorar la detección del borde en imágenes de mala calidad.

## **Plantillas deformables**

Las plantillas deformables o template son figuras geométricas predefinidas que se deforman hasta conseguir cerrar el objeto deseado **(17)**, es un contorno activo paramétrico que surge porque existía un problema con la técnica snake y era que sus conocimientos a priori eran muy pobres y las formas favorecidas eran muy elementales (líneas y puntos).

Las mismas constan de 3 elementos básicos:

- Un modelo geométrico parametrizado para las características incluyendo probabilidades para los parámetros, correspondiendo a una medida geométrica adecuada.
- Un modelo de representación para determinar cómo una plantilla deformable de geometría específica dará origen a intensidades específicas de la imagen, esto puede ser expresado como una medida de representación adecuada.
- Un algoritmo que utiliza las medidas geométricas y de representación adecuadas para equiparar la plantilla con la imagen.

La metodología para configurar de forma lineal una plantilla deformable es obtener el espacio de tipos<sup>21</sup> **(17)**, usando la restricción a una parametrización lineal, esto ofrece ventajas a la hora de simplificar los algoritmos de ajuste y evita problemas a la hora de buscar el mínimo local, son mecanismos muy potentes para buscar estructuras conocidas en una imagen, garantizando que el modelo geométrico sea más específico, mediante la construcción de un conjunto de curvas flexibles que conforman el contorno global, con un conjunto de parámetros que controlen los tamaños de las diferentes partes y los ángulos con las que se unen, ajustándose de manera similar a un snake buscando el contorno a determinar para minimizar su energía. La principal desventaja de esta técnica es que tiene que delimitarse su aplicación a casos predefinidos, de lo contrario no puede realizar su objetivo, por ejemplo en el caso particular en el que se quiera detectar el iris del ojo y se tenga una plantilla con figura geométrica cuadrada, entonces la operación de segmentación será mucho más difícil si se utilizara una figura geométrica circular.

## Contornos dinámicos

Los contornos dinámicos son curvas de seguimiento que utilizan información dinámica a priori **(5)**. Pueden ser aplicados estáticamente a imágenes simples o dinámicamente a secuencias temporales de imágenes para contemplar sus movimientos y deformaciones, teniendo como base las plantillas deformables, este define formas parametrizadas y a partir de ahí define ecuaciones dinámicas para dichos parámetros. Esta técnica se convierte en un proceso de dos fases, en el cual el modelo dinámico es usado para predecir y

---

<sup>21</sup> subespacio para restringir el espacio de trabajo a contornos que sean de interés.

extrapolar el movimiento de un tiempo discreto al próximo, la posición predicha en cada paso es refinada usando las características de la imagen. Para realizar el seguimiento de objetos identificados con figuras predefinidas usando contornos dinámicos se generan un conjunto de curvas B-spline (diccionario de figuras), con el objetivo de asociar un objeto encontrado en la imagen con la figura correspondiente del diccionario y se define una medida de similitud que permite realizar esta asociación, este proceso de seguimiento en una secuencia de imágenes se realiza durante un intervalo de tiempo, determinando la ubicación y la orientación del mismo en ese lapso.

## **1.4 Herramientas para el procesamiento de imágenes**

Muchos años de investigación han demostrado que lograr que un ordenador pueda interpretar y procesar imágenes de forma automática es un proceso complejo de realizar, sin embargo el desarrollo tecnológico en las capacidades de procesamiento y memoria en las computadoras ha hecho posible que este proceso sea un poco más flexible ante la creciente necesidad de procesar imágenes automáticamente. En el entorno del procesamiento de imágenes se deben tomar todos los elementos que se tienen y seleccionar aquellos que resultan apropiados para llevar cabo este proceso, las imágenes u objetos que son utilizadas deben ser procesadas a través de diferentes métodos, los cuales pueden ser tanto experimentales buscando resultados más novedosos, o los que buscan acercarse a las formas de expresión más conocidas como el clasicismo, impresionismo, etc. Las formas de procesar imágenes pueden producir una ilimitada cantidad de combinaciones y resultados ya que están ligadas a la creatividad y experimentación con cada método en dependencia de la herramienta utilizada, para ello es necesario conocer las características y posibilidades de las técnicas con que cuentan las mismas para resolver el problema y enriquecer la solución propuesta. El creciente interés en el desarrollo de estas herramientas en sistemas de visión automática mediante computadoras ha sido basado en un esquema sistemático de 3 etapas:

- Procesamiento a bajo nivel mediante el trabajo directo con las imágenes para extraer las propiedades de las mismas como bordes, gradiente, color, etc.
- Procesamiento de nivel intermedio agrupando los elementos obtenidos en el bajo nivel, para obtener por ejemplo contornos y regiones, con el propósito de segmentación.

- Procesamiento de alto nivel con la interpretación de los procesos antes obtenidos en los niveles inferiores y se utilizan conocimiento a priori<sup>22</sup> del dominio.

Enfocando en el procesamiento de imágenes en los tres niveles se encuentran la librería OpenCV desarrollada en el lenguaje de programación C++, el toolbox de MATLAB, entre otras. Todos son lenguajes de alto nivel que expresan algoritmos para el procesamiento de imágenes de una manera adecuada a la capacidad cognitiva humana, en lugar de la capacidad ejecutora de las máquinas, generando un programa más sencillo y comprensible así como escribir un código válido para diversas máquinas y posiblemente diversas plataformas. Entre los algoritmos que se pueden encontrar en estos lenguajes para el procesamiento de imágenes se encuentran: mejora de propiedades de la imagen como contraste y color, transformaciones geométricas (rotación, traslación, escalado), filtros (suavizado de bordes y morfología), transformación binaria, detección de bordes y contornos, posición y coincidencias de formas, reconocimiento de modelos geométricos (líneas, círculos), formas y caracteres y exportación de imágenes y video compatibles con varios fabricantes.

### 1.4.1 C++

C++ es un lenguaje de programación diseñado a mediados de los años 1980 con la intención de extender al exitoso lenguaje de programación C con mecanismos que permitan la manipulación de objetos, posteriormente se añadieron facilidades de programación genérica, con la particularidad de redefinir los operadores (sobrecarga de operadores) y de poder crear nuevos tipos de datos que se comporten como tipos fundamentales. Ofrece una serie de librerías para el procesamiento de imágenes entre las que están CIMG y las OpenCV, esta última es una de las más completas, contienen código abierto que incluye funcionalidades para el preprocesamiento y derivación de la imagen, operaciones de convolución, así como funciones para la realización y uso de un filtrado robusto, brindando la facilidad de realizar una aplicación distribuible sin necesidad de pagar licencia **(21)**.

---

<sup>22</sup> tipo de conocimiento que es independiente de toda experiencia, más específico al tema es construir un conjunto de características de la imagen que formen el contorno global deseado.

## 1.4.2 MATLAB

MATLAB es un software matemático que ofrece un entorno de desarrollo integrado (IDE) con un lenguaje de programación propio, que permite la representación de datos, funciones e imágenes mediante la manipulación de matrices, la implementación de algoritmos, la creación de interfaces de usuario (GUI) y la comunicación con programas en otros lenguajes y con otros dispositivos hardware. Su nombre surge de matrix laboratory, dado que el elemento básico de dato es una matriz, la cual no requiere de dimensionamiento, lo que permite resolver problemas en una fracción de tiempo, del que nos tomaría al escribir un programa en cualquier lenguaje como, Pascal, Fortran, C, etc. Tiene un ambiente de cómputo, de alta ejecución numérica y de visualización, que integra el análisis numérico, cálculo de matrices, procesamiento de imágenes y diseño gráfico, en un ambiente sencillo de utilizar, donde los problemas y sus soluciones son expresadas justamente como están escritas a diferencia de la programación tradicional, cuenta con varias familias de soluciones para aplicaciones específicas llamadas toolbox, que no son más que colecciones de funciones utilizadas para resolver alguna clase particular de problema, las áreas en donde los toolbox están disponibles incluyen el procesamiento y reconocimiento de imágenes, diseño de sistemas de control, la simulación de sistemas dinámicos, la identificación de sistemas, redes neuronales artificiales y otros **(22)**.

## Conclusiones

Los métodos basados en características son útiles cuando las imágenes son sencillas, el objeto y el fondo son uniformes, la presencia de ruido y otros elementos no resultan importantes, pero en situaciones más complejas en que es necesario realizar un procesamiento posterior a un nivel más alto, para interpretar, reconocer, clasificar, procesar imágenes segmentadas a bajo nivel y eliminar los falsos positivos y negativos son poco factibles. Una solución satisfactoria consiste en utilizar técnicas que sean más robustas, que modelen las fronteras entre un objeto, el fondo y el resto de la imagen y permitan extraer los contornos de los objetos de interés, basándose en modelos que utilizan información a priori de las imágenes. El uso de contornos activos con información a priori de las imágenes para el seguimiento de las mismas en movimiento, permite independizar los problemas de ruido en zonas que no son de interés, lo que mantiene una coherencia en la trayectoria, y la precisión en la determinación de la posición.

## **Capítulo 2 Solución propuesta**

Este capítulo está enfocado en la propuesta de la técnica para el reconocimiento de imágenes por bordes o contornos, así como la herramienta a utilizar para validar la misma, queda evidenciado como será el proceso de aplicación de la técnica así como una breve descripción de los algoritmos utilizados, en aras de alcanzar una alta precisión en el reconocimiento, lo cual a menudo implica alto costo computacional y por otro lado con la aplicación del modelo correcto lograr reducir el tiempo de procesamiento a fin de que más imágenes se puedan procesar en un tiempo limitado. Teniendo en cuenta estos requisitos se explican en el capítulo como la solución propuesta da cumplimiento a los mismos, proporcionando una detallada descripción para su correcta aplicación.

### **2.1 Selección del método propuesto**

Dentro de las dos tendencias vinculadas a las técnicas de detección de imágenes por bordes o contornos, para el diseño de la propuesta se seleccionaron los métodos basados en modelos deformables, debido a que los mismos permiten diseñar procesos enfocados a tareas específicas que cuenten con los criterios necesarios para la solución de problemas, en cuanto atenuación de ruido, reconocimiento de características y procesamiento, proporcionando técnicas mucho más robustas frente a la presencia de ruido o otros factores que elevan la complejidad del proceso de detección, las cuales segmentan imágenes más complejas y sus soluciones no requieren procesamiento posterior y son directamente interpretable.

Dentro de los métodos basados en modelos deformables se escoge la técnica snake como propuesta para el reconocimiento automático de imágenes por bordes o contornos en secuencias de video, esta técnica es una nueva forma de contorno activo que extiende el área de aplicación de estos modelos, su ventaja radica en la capacidad de hacer uso del conocimiento a priori incierto, así como una descripción verbal de las características del objeto por un humano experto, teniendo en cuenta una descripción más detallada del objeto, cuyo acercamiento puede mejorar la detección en imágenes de mala calidad, ofreciendo versatilidad para distintos tipos de escenarios, ya sea con ruido o con objetos difíciles de trabajar, brindando una descripción matemática del área de interés sin importar su forma con la finalidad de distinguir objetos específicos dentro de la misma mediante la segmentación de regiones, busca mínimos

locales definidos por funcionales de energía definiendo variables de interés (intensidades de la imagen) que permitan experimentar con esta propiedad, obteniendo figuras específicas para su procesamiento de alto nivel donde los objetos de interés no deben tener ninguna forma predefinida para su extracción a la que se le incorpora información morfológica, mediante parámetros asociados a su geometría haciendo uso de MATLAB 7.0.

Se propone validar la solución en esta herramienta debido a que la geometría de los contornos activos permite una amplia cobertura de formas que emplea diversas representaciones geométricas a través de curvas y superficies y es más fácil representar y adaptarse al modelo con la misma, además de que es un software muy cómodo de ensayo para el experimento, facilita una serie de funcionalidades para el procesamiento de imágenes, por lo que su uso aprovecha la representación matemática para cada objeto o región de interés que proporcionan los contornos activos, ajustándolos dinámicamente sobre la imagen, teniendo la ventaja de resolver problemas en fracción de tiempo, del que nos tomaría al desarrollar un programa en otras herramientas para la programación. Haciendo uso del toolbox de MATLAB 7.0 la adaptación del contorno activo propuesto hacia los objetos o regiones de interés está determinado por su ajuste, mediante fuerzas y restricciones que determinan la evolución del contorno en función de las características de la imagen, el cual es estimado y calculado usando un conjunto de funciones de dicha herramienta.

La formulación del ajuste del contorno se basa en un criterio de balance de energía, en términos de las distintas propiedades definidas, tanto del contorno como de la imagen a procesar, se minimiza la energía total en el contorno al deformarlo desde un estado inicial hasta el nivel mínimo que corresponde con el ajuste óptimo a la zona de interés. Con la búsqueda de ese ajuste se derivan los términos del funcional de energía a un enfoque que aproxima la solución óptima, para lograr el mismo en primer lugar se definieron las fuentes de energía que afectaban al contorno, como restricciones de formas mediante la umbralización y suavidad a través del filtro de mediana, estos términos se estimaron para formar el funcional de energía del contorno, el cual consta de las propiedades del contorno (energía interna) así como las propiedades de la imagen (energía externa), de acuerdo con la estimación realizada el modelo se ajusta dinámicamente, donde su energía variará según este se deforma, evaluando los términos de energía interna y externa definido por el funcional. Por ser la técnica Snake una técnica basada en visión computacional, la misma parte de diferentes escenarios que rigen los pasos a seguir teniendo en cuenta

las propiedades y el ajuste del mismo a las necesidades planteadas para su correcta aplicación quedan definidas las siguientes etapas: preprocesamiento, segmentación, extracción de características y reconocimiento.

## 2.2 Preprocesamiento

El preprocesamiento consta de 2 etapas fundamentalmente: la primera etapa se lleva la imagen a escala de grises mediante algoritmos de conversión de tipos a transformaciones como color a gris, con el objetivo de extraer una imagen de mejor calidad donde se pueden obtener más detalles acerca de los bordes y contornos de la misma. Primeramente se representa en un solo tono de gris el conjunto de colores de la imagen y se pone cada color en una intensidad, para ello se utiliza un algoritmo de conversión de color a gris predefinido por el toolbox de MATLAB 7.0 mediante la función *rgb2gray*, esta convierte la imagen digital en una imagen matricial bidimensional de  $M \times N$ , donde  $N$  representa el número de píxeles de ancho y  $M$  los píxeles de largo, cada elemento de la imagen tiene un valor 0 (negro) a 255 (blanco), logrando que se pueda leer en 3 planos de color RGB.

En la segunda etapa se propone realizar el preprocesamiento de la imagen a través de una técnica de reducción de ruido empleando algoritmos de análisis y mejora, para ello se emplea un filtro de mediana con el objetivo de retener las características de interés de la imagen, atenuar el ruido existente y que las probabilidades de acierto en el proceso de extracción de características sean mayores. Este tipo de filtrado es una técnica no lineal de mejora de la señal, utilizada para operar sobre la imagen cambiando el valor de intensidad de cada píxel analizado mediante máscaras de convolución, aplicando un suavizado en la imágenes toma un entorno de un píxel de la imagen y ordena los niveles de gris del entorno de menor a mayor, obteniendo aquel que sea la mediana estadística de la muestra de valores, es decir devuelve aquel valor de gris que cumpla con la mitad de los valores de la muestra son superiores a él, y la otra mitad son inferiores a el mismo, teniendo en cuenta los valores de intensidad de los píxeles en la vecindad establecida y como principal función hacer que los puntos de intensidades muy distintas se hagan muy parecidos a sus vecinos, eliminando puntos de muy alta y muy baja intensidad para la eliminación de ruido, este filtro con respecto a otros operadores como los filtros pasa bajo mantiene la información de los bordes de la imagen, aprovechan estas características para adaptarse bien a ciertos

tipos de ruido como el impulsivo, sin embargo su mayor inconveniente es que es un proceso lento en comparación con otros filtros lo que implica un coste computacional más elevado.

## 2.3 Segmentación

La generación de metodologías para la extracción de información a partir de regiones u objetos de interés de la imagen posibilita una mejor comprensión del procesamiento de la misma, la segmentación es una importante tarea hacia un sistema de análisis de la imagen a alto nivel, esta permite la cuantificación y visualización de las regiones u objetos de interés y de su calidad depende en gran medida el resultado final de la interpretación automática de una secuencia de imágenes. Aunque no existe un método universal para realizar la segmentación de una imagen, este proceso está íntimamente ligado a la tarea a resolver estos problemas y finaliza el mismo cuando satisface las expectativas del observador. La segmentación realizada separa los objetos o regiones de interés del resto de la imagen, mediante criterios de proximidad, similitud o continuidad, buscando construir un conjunto de píxeles, regiones o contornos similares en intensidad, color, textura u otra característica, llegando a definir un conjunto de estructuras o descriptores de la escena del objeto o región de interés, se centra principalmente en la extracción de estructuras u objetos de la imagen de interés, como son: la detección de contornos y regiones permitiendo la descomposición de la imagen en diferentes entidades lógicas, las cuales deben ser lo suficientemente invariables e independientes y lo suficientemente significativas para su reconocimiento, para ello el proceso de segmentación propuesto se divide en 3 etapas: umbralización, binarización y morfología matemática.

Para el proceso de umbralización de la imagen se propone un umbral calculado por el método óptimo de Otsu, este es un procedimiento no paramétrico basado en el criterio de Otsu, que elige el umbral óptimo maximizando la varianza<sup>23</sup> entre clases de objetos mediante una búsqueda exhaustiva, su importancia radica en que es automático, por lo que no necesita supervisión humana, ni información previa de la imagen antes su procesamiento, sus principios encuentran similitudes entre los píxeles pertenecientes a un objeto y sus diferencias respecto al resto, por lo que este método garantiza que la escena se caracterice por un fondo uniforme y por objetos parecidos. La aplicación de este umbral se realiza

---

<sup>23</sup> muestran la variabilidad de una distribución, indicando por medio de un número, si las diferentes puntuaciones de una variable están muy alejadas de la media de dicha variable aleatoria.

teniendo en cuenta los siguientes pasos: después de haber aplicado el preprocesamiento a la imagen se procede a la umbralización de la misma con el algoritmo propuesto, para ello se calcularon la distribución de probabilidades de los niveles de gris así como el momento acumulado de orden cero y de orden uno con el objetivo de definir la distribución de intensidades de la imagen, las clases de objetos contenidas en la misma y garantizar que la escena contenga un fondo uniforme y por objetos parecidos. Luego se etiquetan los píxeles correspondientes a las regiones u objetos de interés con el valor 1 y con 0 aquellos que conforman el fondo logrando separar las regiones u objetos de interés del fondo de la imagen, dando lugar así a una imagen binarizada. Para la binarización de la imagen se deben comparar los niveles de gris presentes en la misma con el valor del umbral obtenido en el proceso de umbralización dentro de los niveles de gris obtenidos por el mismo, de tal forma que todos lo que sean menores al umbral calculado se convertirán en negro y todos los mayores en blanco.

Una vez definidos y separados las regiones u objetos de interés en la imagen se aplican operaciones morfológicas para la simplificación de los datos de la misma, así como preservar sus características o eliminar otros aspectos irrelevantes. El primer paso del proceso morfológico se representan las regiones u objetos de interés de la imagen a través de conjuntos de coordenadas discretas para representar el conjunto de píxeles o puntos que conforman la imagen y definir el elemento estructural de la misma que no es más que el ajuste o modificación en una zona de la imagen sin que afecte al resto, protegiendo la zona que no debe variar, una vez definido el elemento estructural en dependencia del mismo se procede a aumentar el tamaño y cantidad de los objetos de la imagen con la unión de píxeles relacionados mediante el algoritmo de dilatación tomando cada píxel del objeto con valor 1 y cambiando al valor 1 todos aquellos píxeles pertenecientes al fondo de la imagen que tienen un elemento estructural con valor 4 y 8 con el píxel del objeto, es decir poner con el valor 1 los píxeles del fondo vecino a los píxeles del objeto.

Como último paso se eliminan detalles irrelevantes en la imagen en términos de tamaño mediante la erosión, seleccionando un elemento estructural cuyo tamaño sea un poco menor que el área de interés que se quiere conservar y tomando cada píxel que tiene un elemento estructural con los píxeles vecinos del fondo y cambiarlos a 0, es decir poner con valor 0 los píxeles del objeto vecinos a los píxeles del fondo. En la bibliografía consultada se puede encontrar una forma bastante genérica de implementar los algoritmos de erosión y dilatación, los cuales están hechos con instrucciones ya definidas en los diferentes

lenguajes de programación abordados, en este caso se usó MATLAB 7.0 para implementar dichos algoritmos mediante la instrucción *imdilate*, *imerode*, *imopen*, *imclose*.

## 2.4 Extracción de características

Este epígrafe tiene como fin estudiar y proponer el diseño del procesamiento básico para el proceso de extracción de características mediante un algoritmo que determine ciertas propiedades de la imagen para aplicarlo posteriormente a una secuencia de imágenes a través de la herramienta MATLAB 7.0. Para la aplicación del mismo se debe tener la imagen en formato binario de intensidad 0 y 1 (255), que contendrá la información de regiones u objetos de interés, posteriormente se le generará un algoritmo en esta herramienta que procese dicha imagen binarizada para obtener las expresiones analíticas que representen las características más significativas para su extracción. La forma y el color son propiedades importantes para la extracción de características en la imagen, este proceso constituye uno de los retos más importantes en las aplicaciones de reconocimiento de imágenes debido a la complicada tarea de extraer las mismas en una imagen donde la forma de los objetos no sea tan importante como las características. El propósito de la extracción de características es comparar el método de detección con el fin de su correcta aplicación para el reconocimiento de imágenes, por lo que es necesario detectar puntos estables del entorno que sean persistentes a lo largo de un extenso número de imágenes consecutivas, de esta manera se mejoraría el reconocimiento de regiones u objetos de interés en la imagen, realizándose una asociación correcta entre las características de la misma con la influencia de pequeños cambios de puntos de vista.

Con el estudio comparativo de los diferentes detectores de características abordados, se compararon los alcances de los descriptores y se determinó que el algoritmo *SURFT* es más rápido que el *SIFT*, por lo que se propone la aplicación del mismo para el proceso de extracción de características de la imagen como paso previo al reconocimiento de objetos o regiones de interés de las misma. La precisión del método de extracción de características será calculado usando las mismas técnicas y algoritmos propuestos, para la extracción se contiene solamente en la imagen el objeto o la región de interés como resultado del proceso de segmentación, luego se puede proceder a realizar la extracción de los mismos a través de un conjunto de datos de imágenes o a partir de una imagen de ensayo para finalmente mediante

el uso de una red neuronal multicapa que tiene como entrada el vector característico de cada objeto o región de interés se procede a reconocer los diferentes tipos de objetos.

El proceso de extracción de características de la imagen propuesto se basa principalmente en el seguimiento de puntos, para cada fotograma extraído se extraen primeramente los puntos de interés a través del algoritmo detector propuesto, como la imagen esta en 2D<sup>24</sup> se realiza a partir de una matriz homográfica, calculada en cada par de imágenes consecutiva, esto viene dado a un conjunto de puntos en un plano que llamado plano de referencia, se busca su correspondencia en el plano imagen aun plano homográfico, también conocido como transformación proyectiva en el plano, el cuales descrito mediante una matriz H de rango 3 x 3 llamada matriz homográfica la cual determina la transformación de los puntos en el plano de referencia a puntos en el plano imagen **(23)**. Esta matriz puede ser calculada seleccionando cuatro correspondencias de puntos coplanares<sup>25</sup> entre las imágenes, dado un punto detectado en una de las imágenes, se predice su posición en la imagen consecutiva, si esta posición se sitúa a una distancia menor que dos píxeles se considera el punto correspondiente. El hecho de utilizar como entrada para el seguimiento de puntos un gran número de correspondencias permite hacer el cálculo de la matriz homográfica más preciso. Una vez obtenido los puntos se procede a trabajar con funciones del toolbox de visión de MATLAB 7.0 para determinar las características de la imagen que definan la posición de los objetos a identificar, en caso de que las características a extraer sean líneas rectas se usa la transformada de Hough mediante la función *SPhough*, esta función puede ser llamada empleando varios parámetros por lo que para adquirir una descripción más completa de los mismos se recomienda utilizar la ayuda de MATLAB 7.0. En el caso de que las características a extraer presenten otra geometría, en dependencia de la misma se define y calcula un área de los objetos contenidos así como su orientación, para esto se hace uso de la función *bwlabel*.

## 2.5 Reconocimiento de patrones

Después de realizar el preprocesamiento de la imagen, segmentación y extracción de características de la misma se propone para realizar el reconocimiento de regiones u objetos de interés una red neuronal perceptron multicapas que posee el toolbox de MATLAB 7.0, la cual mediante el algoritmo *back*

---

<sup>24</sup> 2D o en dos dimensiones.

<sup>25</sup> son puntos que yacen en el mismo plano.

*propagation* encuentra una representación interna que le permite generar las salidas deseadas durante el entrenamiento, y posteriormente durante el funcionamiento es capaz de generar salidas para entradas que no les fueron mostradas durante el aprendizaje pero que se asemejen a algunas de las ya aprendidas, teniendo como datos de entrada las relaciones de características y como datos de salida el número de patrones encontrados en la imagen.

El proceso de reconocimiento de la red parte de dos operaciones: entrenar la misma y simularla. Antes de realizar dichas operaciones, primeramente se debe definir el problema que la red neuronal debe solucionar para proporcionarle a la misma la información necesaria para que pueda llevar la tarea con éxito, definiéndole los patrones de aprendizaje que va a utilizar en el proceso de entrenamiento, en el caso de la red multicapa de MATLAB 7.0 esto se hace definiendo dos matrices una para las entradas y otras para las salidas donde cada patrón de aprendizaje se define por columnas, luego se procede a entrenar la misma mediante el algoritmo de auto-aprendizaje *back propagation* con 2 pasadas a través de la capa de la red uno hacia delante y otra hacia atrás, solo se le debe indicar los parámetros de aprendizaje de la red mediante la instrucción  $testinput = C([1:2],:)$ .

Una vez definido el problema y con este los patrones de entrenamiento, se procede a inicializar la red neuronal, en el caso de la red propuesta se usa la función *initp*, primeramente se crea un objeto para almacenar la red creada mediante la función *newp* ( $[x],y$ ) donde  $y$  es el número de neuronas que tiene la red en la capa de salida y  $x$  es una matriz que representa el rango del valor de la entrada de la red neuronal cuyo número de filas de esta matriz MATLAB 7.0 lo utiliza para definir el número de entradas que tendrá la red neuronal. Finalmente se procede a entrenar la red mediante el comando *train* el cual implementa un entrenamiento con la regla de aprendizaje tipo perceptron, a medida que la red se va entrenando MATLAB 7.0 grafica la manera como va evolucionando al transcurrir las iteraciones de aprendizaje para después validar la misma, este proceso se realiza una vez la red entrenada donde se valida si el comportamiento de la misma es correcto o no, para ello se usa la función *sim*( $x,z$ ), donde  $x$  es red perceptron entrenada anteriormente y  $z$  es el patrón de prueba y se procede a simular la red con una imagen de prueba, se realiza el preprocesamiento y segmentación de la misma y se introduce en la red neuronal aprendida.

## Conclusiones

Para el proceso de detección de imágenes por bordes o contornos se propone los contornos activos, mediante el ajuste de los mismos y considerando los parámetros de morfología, se estiman los indicadores de transición entre objetos y regiones de interés y el fondo de la imagen, para esto se definieron funciones detectoras de bordes que cuantifican los cambios entre los niveles de intensidad para los píxeles de la imagen a través del vector gradiente de la misma, el mismo es un tipo particular de fuerza de los contornos activos paramétricos utilizados para mejorar la convergencia del proceso de ajuste. En el procesamiento de la imagen se tiene en cuenta primeramente un preprocesamiento de la misma a través de una técnica de reducción de ruido, así como la selección de un umbral apropiado para la binarización de la imagen usando criterios morfológicos que ayuden al proceso de ajuste del contorno activo. La extracción de características de la imagen y el reconocimiento en la misma se propone utilizar la red perceptron que contiene MATLAB 7.0 basada en un modelo de reconocimiento de regiones u objetos según las características extraídas de los mismos.

## **Capítulo 3 Resultados obtenidos**

En este capítulo se abordan los resultados obtenidos después de haber aplicado la propuesta así como las ventajas de este modelo con respecto a otros, llevando una secuencia cronológica en su presentación para tener al final otras alternativas para su aplicación. El enfoque principal consiste en asegurar en los experimentos realizados que el modelo propuesto es adecuado para la detección de imágenes por bordes o contornos de forma automática en secuencias de video, se presenta un estudio del modelo activo y a través de los resultados obtenidos se detallan los pasos necesarios para el reconocimiento de imágenes aplicando el mismo, es cual tiene como principal ventaja su gran adaptabilidad para poder ser aplicado a diversos tipos de imágenes.

### **3.1 Dominio del modelo propuesto**

La investigación se enmarca en las técnicas de reconocimiento de imágenes por bordes o contornos, donde se muestra un perfil tecnológico para el posterior desarrollo de un sistema basado en visión computacional para el área de gestión y procesamiento de medias del Departamento productivo Señales Digitales del centro de desarrollo GEYSED. El modelo propuesto permitirá profundizar la adquisición de conocimientos para el procesamiento de imágenes aplicadas a procesos en donde sea necesario realizar una recopilación de información proveniente del mundo real, aunque el experimento se enmarca en un estudio en imágenes extraídas de toolbox de MATLAB 7.0 para el reconocimiento de patrones semejantes a rostros de personas, también se concibió para procesar imágenes extraídas en secuencias de video mediante técnicas de inteligencia computacional. Sin embargo al realizar las primeras investigaciones se determinó que para este tipo de aplicaciones, específicamente para imágenes en movimiento en continuo cambio, esta técnica consume mucho tiempo de cómputo durante el procesamiento de dichas imágenes, por lo que se orienta profundizar en los algoritmos computacionales referentes al procesamiento de imágenes mediante técnicas de reconocimiento de imágenes por bordes o contornos que involucran acelerar las etapas de adquisición, procesamiento, extracción de características y reconocimiento de patrones.

La investigación no brinda un modelo óptimo o general para todo tipo de aplicaciones de reconocimiento y procesamiento de imágenes, pero si demuestra que la introducción de sistemas basados en visión por

computadora es un objetivo factible de implementar, mediante el desarrollo de interfaces como canal de comunicación desde el usuario a la computadora para el proceso de mejoramiento de la calidad de la imagen, su interpretación, extracción de características y procesamiento, por lo que este trabajo define que para cada aplicación se deberá determinar las técnicas adecuadas para una extracción eficiente de las regiones u objetos de interés dentro de la imagen. La propuesta diseñada es capaz de incorporar nuevas tareas o adaptarse a cambios utilizando las otras técnicas existentes, sin requerir demasiado trabajo para su comprensión aunque es indispensable realizar inicialmente una exploración de la imagen al área de interés antes de realizar el proceso de detección, para eliminar componentes que no aportan información relevante al proceso de detección y reducir el tiempo de cálculo de las etapas posteriores, para esto la umbralización óptima según el criterio de Otsu es el método apropiado. Al comparar las técnicas propuestas para la reducción de ruido, se demuestra que se logra un mejor resultado aplicando un filtro de mediana para el suavizado de la imagen, la combinación de las técnicas de detección de ruido abordadas mejoran los resultados obtenidos que al aplicar una técnica individualmente, haciendo más fácil visualizar las características de interés de la imagen.

## 3.2 Preprocesamiento

En el experimento se tuvo en cuenta para la adquisición de las imágenes una serie de formatos que están actualmente en uso, por citar algunos ejemplos: GIF, JPG, TIF y PPM son los más utilizados. Esta larga lista de formatos que existe constituye un problema debido a que se pueden interpretar de formas diferentes y se debe contar con una biblioteca que admita al menos una lectura de cada uno de ellos como la librería que posee el toolbox de MATLAB 7.0 para el procesamiento de imágenes de diferentes formatos. Para realizar el experimento se trabajará con los formatos JPG, PNG y TIF y teniendo en cuenta que pueden existir diferentes tamaños de imágenes para el mismo se redimensiona la imagen a un tamaño adecuado al método que se va a utilizar. En la fase de preprocesamiento de la imagen digital se tienen en cuenta los procedimientos a realizar en la solución propuesta, para ello en la primera etapa se toma una imagen de la colección de imágenes que posee el toolbox de MATLAB 7.0 como lo muestra la figura 5a, se adquiere la misma mediante la función *imread* y luego con la función  $[f,c]=size()$  se convierte en una imagen matricial, para representar en un solo tono de gris los colores observados y se pone cada color en una intensidad para después con la función *rgb2gray* llevarla a escala de grises, en la figura 18 del anexo se muestra el script desarrollado en MATLAB 7.0 para realizar este proceso y los resultados

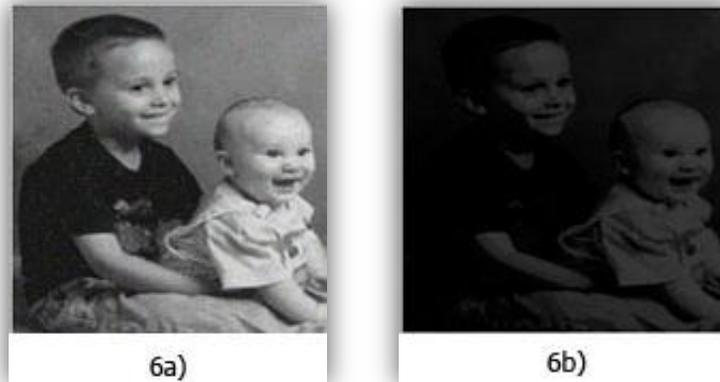
obtenidos se muestran a continuación en la figura 5b. Cabe aclarar que en estos resultados la descripción RGB proporcionada de un color de la imagen hace referencia a la descomposición del color en términos de la intensidad de los colores con que se forma el rojo, el verde y el azul. Esta representación de la imagen es un modelo de color basado en la síntesis aditiva y con el que es posible representar un color mediante la mezcla por adición de los tres colores de luz primarios, mediante los valores RGB pueden mostrarse colores notablemente diferentes que aunque utilicen un mismo modelo de color, los espacios de colores varían considerablemente.



**Figura 5 Escala de grises.**

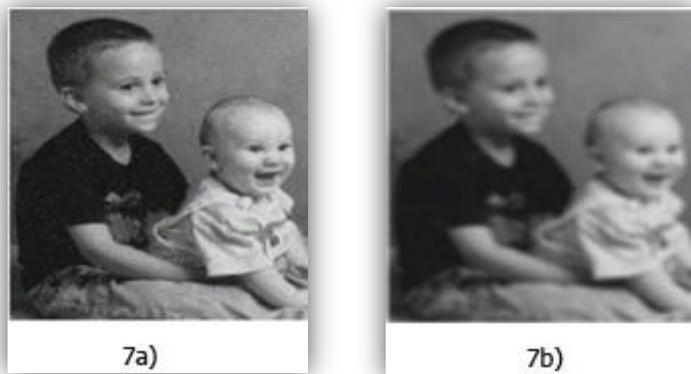
A continuación se muestran los resultados de la segunda etapa del preprocesamiento tras aplicar técnicas de reducción de ruido como el filtro lineal pasa bajo con convolución de una imagen con una máscara predefinida y un filtro no lineal de mediana mediante una operación no lineal con los píxeles entorno de vecindad con el objetivo de eliminar o atenuar el ruido o efectos espurios que puedan presentarse en la imagen y escoger la mejor técnica para realizar el filtrado. En este ejemplo como lo muestra la figura 6, la imagen en escala de grises obtenida en el paso anterior se le añade un ruido del tipo “sal y pimienta” (ver figura 6a) y se le aplica posteriormente un filtro ideal pasa bajo (ver figura 6b) en el dominio de la transformada discreta de Fourier, con el objetivo de descartar cualquier residuo de ruido de cualquier tipo haciendo más restringido el dominio de la frecuencia del filtro pasa bajo se le multiplica esta transformada. El resultado obtenido tras aplicar el filtro pasa bajo es considerado satisfactorio para aquellos casos donde se desee realizar un filtrado general de la imagen en aras de mejorar su calidad y atenuar la presencia de

ruido, pero no es eficiente cuando se desea darle un procesamiento posterior a la imagen en cuanto a detección y extracción de sus características porque este tipo de filtro atenúa durante el suavizado de la imagen los bordes y otras propiedades de la misma.



**Figura 6 Reducción de ruido con un filtro pasa bajo.**

Para la atenuación de ruido como indica la propuesta se aplica a la imagen en escala de grises con ruido de sal y pimienta, un filtro de mediana, a continuación se exponen dos resultados tras aplicar el mismo con una máscara de convolución de 3x3 (figura 7a) y otra de 9x9 (figura 7b). Se puede apreciar en los resultados obtenidos la importancia de definir el tamaño de la máscara, cuanto mayor sea esta se consigue una mayor reducción de ruido, pero en cambio se produce una mayor difuminación de bordes en el área de interés que se desea conservar. La metodología aplicada para el preprocesamiento de imágenes presenta buenos resultados como se puede observar, pero es indispensable realizar inicialmente una reducción de la imagen al área de interés para la eliminación de componentes que no aportan información al proceso y reducir el tiempo de cálculo de las etapas posteriores, para esto con el uso del filtro de mediana se satisfacen estos requerimientos mediante la máscara escogida.



**Figura 7 Reducción de ruido con un filtro de mediana.**

### **3.3 Segmentación**

La segmentación realizada es el resultado del proceso de umbralización, binarización y morfología de la imagen, en la que se separaron los objetos o regiones de interés del resto de la misma mediante los criterios definidos, llegándose a formar un conjunto de estructuras o descriptores de la escena del objeto o región de interés mediante la detección de bordes y contornos de la misma, aplicando para su reconocimiento los contornos activos. En el experimento se eliminan zonas de área pequeña, descartando que pudieran ser rostros de personas como muestra la figura 8 los resultados obtenidos después de este proceso.



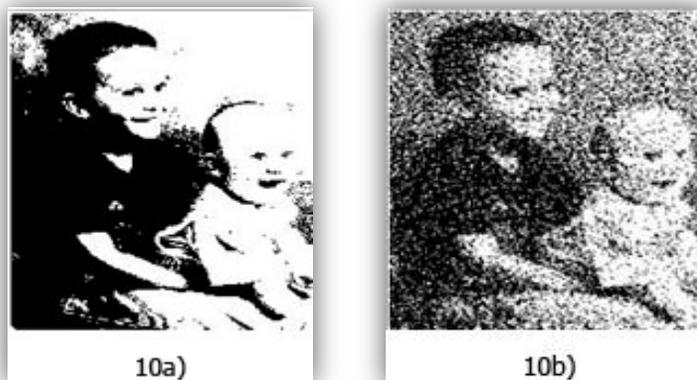
**Figura 8 Segmentación de la imagen.**

En la etapa de umbralización de la imagen para probar la robustez del umbral propuesto se realizó un experimento de prueba para comparar el umbral propuesto con un umbral arbitrario, para ello se escogió la imagen en escala de grises obtenida del preprocesamiento y se le añadió un ruido gaussiano, luego se binariza la imagen con ruido con los dos umbrales y se comparan los resultados. Como se puede apreciar en los mismos, la imagen sin ruido binarizada con el umbral calculado por el método Otsu (ver figura 9a) mantiene rasgos semejantes a los de la imagen con ruido gaussiano binarizada usando el mismo método (figura 9b).



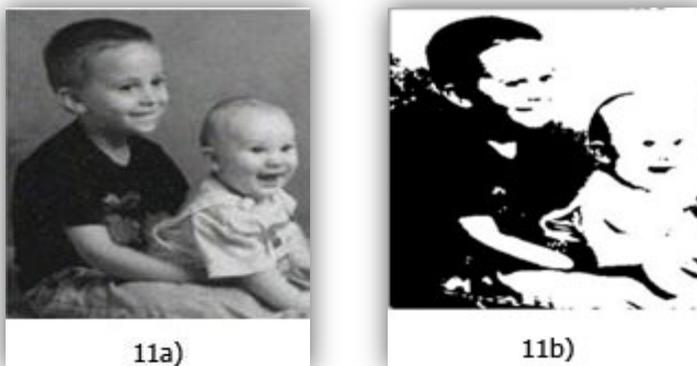
**Figura 9 Binarización con umbral Otsu.**

Sin embargo la imagen resultante tras binarizar la imagen sin ruido con el umbral arbitrario (figura 10a) difiere en gran medida de la obtenida luego de aplicar el mismo proceso a la misma imagen pero con ruido gaussiano (figura 10b), por lo que no se logra definir bien la binarización de la misma por la presencia del ruido. De esta manera se comprueba la efectividad del umbral propuesto en la búsqueda de un umbral óptimo con el que se pueda binarizar una imagen correctamente en tiempo real y que estén inmersas a todo tipo de perturbaciones. Si bien mediante este experimento se puede apreciar las ventajas del método de Otsu, se puede extender su utilidad para aquellos casos en que la iluminación no es uniforme o es en general inadecuada.



**Figura 10 Binarización con umbral arbitrario.**

En la segunda etapa de la segmentación, se compararon los niveles de gris presentes en la imagen “E\_grises\_filtrada.tif” (ver figura 11a), con el umbral obtenido aplicando el criterio de Otsu, teniendo en cuenta que si el nivel de gris de la imagen es menor que el umbral predeterminado se le asigna al píxel de la imagen binarizada el valor 0 (color negro), y si es mayor, se le asigna el blanco, obteniendo de esta manera la imagen en blanco y negro “imagen\_binarizada.tif” como muestra la figura 11b.



**Figura 11 Binarización de la imagen.**

En el proceso morfológico como primer paso se representa la imagen binarizada “imagen\_binarizada.tif” como una función de dos variables discretas  $a[m,n]$  que pueden tomar dos valores “0” o “1”, dependiendo del nivel de gris de la imagen, este conjunto de coordenadas discretas corresponde a todos los puntos o píxeles que pertenecen a la imagen, por lo que tras aplicar morfología matemática mediante estos

conjuntos de coordenadas se representan los objetos de la imagen. Las principales operaciones morfológicas utilizadas son la dilatación (ver figura 12a) y la erosión (ver figura 12b), donde se logra separar e identificar los principales objetos o regiones de interés en la imagen.

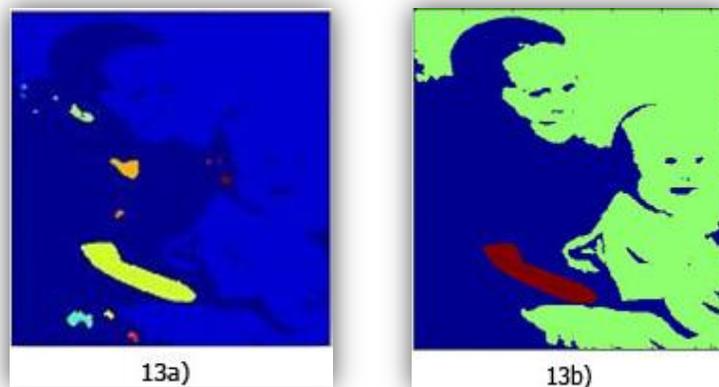


Figura 12 Operaciones morfológicas.

### 3.4 Extracción de características y reconocimiento de patrones

Tradicionalmente se recomienda en el experimento un histograma de color por cada componente rojo, verde y azul para caracterizar una imagen. El problema en optar esta opción es que un histograma global preserva la información de color contenida en una imagen, pero no la localización de este color en la misma, debido a que este método no proporciona el conocimiento acerca de donde están los colores concentrados en dos imágenes teniendo similares histogramas de color pueden tener diferentes semánticas, es por eso que se recomienda realizar el experimento para el proceso de extracción de características en base a la textura y formas de la imagen aunque los algoritmos basados en estas propiedades tienen una sustancial limitación. El algoritmo de representación basado en formas solo es efectivo cuando la imagen posee un fondo uniforme y el de textura no es apropiado para imágenes que no contengan las mismas. La representación de características en la imagen juega un papel muy importante en el algoritmo propuesto para la extracción de características de la misma, pues permite contar con un descriptor de la imagen para su comparación mediante la asociación de un vector de características que incluye información de la imagen.

Para experimentar con el detector propuesto para la extracción de características se ha realizado una evaluación del mismo en base a los requerimientos deseables: condiciones de forma y relación de aspecto<sup>26</sup> de la imagen, mediante los resultados obtenidos se muestra que *SURFT* es el detector que mejor se comporta según los requerimientos evaluados. La evaluación del detector se realiza a partir de una secuencia de imágenes que contiene el toolbox de MATLAB 7.0 por la ausencia de un componente de estructuras de datos, se toma para la evaluación del detector *SURFT* la imagen mostrada en la figura 13a, se analiza su repetitividad y se define el área de los objetos mediante el conteo de objetos encontrados, para ello se tiene en cuenta la longitud del eje mayor y el eje menor del objeto mediante su cantidad de píxeles, su centro de masa y la orientación del objeto en la imagen, para un mayor entendimiento del algoritmo se muestra en la figura 27 del anexo el script correspondiente al proceso de extracción de características. Basado en las condiciones de la forma de los objetos en las imágenes y la relación de aspecto de la misma se hace uso de la red perceptron que posee el toolbox de MATLAB 7.0, la cual permite la segmentación de la imagen y el reconocimiento de los patrones definidos, para el entrenamiento de la red se propone usar el algoritmo *back propagation* teniendo una entrada y una salida para los datos. Una vez entrenada la red se simula la misma con una imagen extraída del toolbox de imágenes, se preprocesa, se segmenta y de acuerdo con las características extraídas se define un patrón a reconocer que este caso sería los rostros de personas, obteniendo como salida el número de patrones encontrados como lo muestra en la figura 13b.



**Figura 13 Proceso de extracción de características y reconocimiento.**

<sup>26</sup> proporción entre la anchura y altura de la imagen.

## **Conclusiones**

Se ha presentado el experimento con un pequeño sistema basado en visión computacional para la detección de rostros de personas en una imagen desarrollado en la herramienta MATLAB 7.0, el mismo se basa en un modelo de reconocimiento de regiones mediante la red perceptron del toolbox de esta herramienta, entrenado con algunas imágenes, la cual aprende los patrones de número de personas según las características de forma y relación de aspecto en la imagen. Para trabajos futuros se puede implementar un sistema de visión con cámaras de video para determinar el número de personas donde sea indispensable este tipo de sistema, también tras aplicar la propuesta a los fotogramas extraídos para la mejora del procesamiento de los mismos y se pueden definir criterios de búsqueda para extraer información relacionada con la media a partir de los patrones encontrados.

## Conclusiones Generales

Con el desarrollo de la investigación se caracterizaron las técnicas de reconocimiento de imágenes por bordes o contornos existentes, lo que permitió definir una propuesta para el procesamiento de fotogramas e imágenes. Para formular la propuesta se realizó un estudio selectivo de las técnicas abordadas en la investigación, que permitieran modelar y definir las fronteras entre un objeto, el fondo y el resto de la imagen para extraer los datos de interés de la misma, así como independizar los problemas de ruido en zonas de interés manteniendo coherencia en la trayectoria y precisión en la determinación de la posición de estos datos. Al lograr los objetivos trazados, el diseño la propuesta se enmarcó en los contornos dinámicos, una técnica basada en visión activa que se validó mediante un modelo de reconocimiento de regiones para la detección de personas, desarrollado en MATLAB 7.0. Para ello se tuvo en cuenta primeramente un preprocesamiento de la imagen a través de una técnica de reducción de ruido, la selección de un criterio de umbral para la segmentación y para el reconocimiento de patrones el uso de la red perceptron multinivel que posee el toolbox de dicha herramienta. Los resultados obtenidos se consideran satisfactorios en cuanto a velocidad de procesamiento y extracción de características, factibilidad de implementación, tiempo y recursos. Cabe resaltar que este modelo aun está en fase experimental pero es completamente funcional para la extracción básica de características en secuencias de imágenes.

## Recomendaciones

- Profundizar en las técnicas de detección por bordes o contornos de imágenes y el campo de la visión computacional, con vista a desarrollar sistemas de reconocimiento automático de patrones y control de calidad en el preprocesamiento de imágenes para la gestión y procesamiento de medias.
- Mejorar el modelo propuesto de detección de imágenes, en aras de obtener mejores resultados superiores en cuanto a índice de precisión y detección de características.
- Investigar sobre otras formas de realizar la detección de imágenes a fin de garantizar mejores resultados.
- Profundizar los conocimientos acerca de la herramienta MATLAB 7.0 en el área de gestión y procesamiento de medias.

# Referencias Bibliográficas

1. Gonzalez, Rafael C, Richard E Woods. *Digital Image Processing*. 2002.
2. Molina, Rafael. *Introducción al procesamiento y análisis de imágenes digitales*. Granada : s.n., 1998.
3. Bimbo, Alberto del. *Base de Datos de Imágenes y Video*. Ciudad de la Habana : s.n., 2005.
4. Rebaza, Jorge Valverde. *Detección de bordes mediante el algoritmo de Canny*. s.l. : Escuela Académico Profesional de Informática.
5. sandoval, Ing omar. *CONTORNOS ACTIVOS.GRADIENTE DESCENDIENTE*. Mexico : CENTRO DE INVESTIGACIONES EN INGENIERIA Y CIENCIAS APLICADA.
6. Martin, Sanz Pajarez. *Visión por computador, imágenes digitales*. España : s.n., 2001.
7. Manuel Agustín Melchor, Jose Miguel Valiente González. *Bases de datos para Multimedia: Recuperación por Contenido*.
8. G, ÁLVAREZ. *PREPROCESAMIENTO DE IMÁGENES APLICADAS A MAMOGRAFÍAS DIGITALES*. Colombia : Scientia Et Technica, 2006. 31.
9. Hoyos, Marcela Hernández. *Procesamiento de imágenes digitales*.
10. Jimmy Wales. es.wikipedia.org. *wikipedia*. [En línea] ewikipedia.org, 18 de marzo de 2010. [Citado el: 20 de abril de 2010.] [http://es.wikipedia.org/wiki/Escala\\_de\\_grises](http://es.wikipedia.org/wiki/Escala_de_grises). 3.0.
11. Artificial, Cátedra de Visión. *Segmentación por Umbralización - Método de otzu*.
12. *Operaciones Morfológicas en Imágenes Binarias*.
13. Ballesta, Mónica. *EVALUACION DE DETECTORES DE PUNTOS DE INTERÉS PARA SLAM VISUAL*. 2007.
14. Lowe, David. *David Lowe's patente para el algoritmo SIFT*. 1999.
15. Ess, Andreas. *Visión por Computador y Comprensión de la imagen (CVIU)*. 2008.
16. Rubira, Emiliano Aldabas. *Introduccion al reconocimiento de patrones mediante redes neuronales*. Barcelona : s.n.
17. García, Elizabeth González. *CONTORNOS ACTIVOS*. Matanzas.Cuba : s.n.
18. Borgfors, G. *Hierarchical Chamfer Matching*. A. 1988.
19. Kass, M. *Snakes: active contour models*. 1988.
20. Osher, S. *Fronts propagating with curvature dependent speed*.
21. Rodriguez, Leandro Jiménez. *Procesamiento de imágenes biológicas*. bellaterra : s.n., 2008.
22. Goering, Richard. *Matlab edges closer to electronic design automation world*. 2004.
23. Sotoca, J. M. *Estimación de movimiento*.
24. Cristián Arriagada García, Diego Aracena-Pizarro. *DETECCIÓN Y RECONOCIMIENTO DE SEÑALES*. Chile : s.n., 2007. 2.
25. Zheng, Q. *Automatic feature point extraction and tracking in image sequences for arbitrary camera motion..* 1995.

# Anexos



Figura 14 Ruido Detector (9).



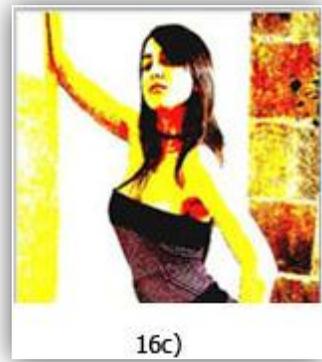
Figura 15 Ruido Sal y pimienta (9).



16a)



16b)



16c)

Figura 16 Imagen digital (9).



Figura 17 Fotograma (9).

---

```
% En este script se presentan los pasos de la propuesta para la detección
% de imágenes por bordes o contornos de forma automática...

% Universidad de las Ciencias Informáticas
% Departamento de Señales Digitales

close all, clear all
imagen =imread('imagen_original.jpg','jpg');%cargar la imagen original
[f,c]=size(imagen);%se convierte la imagen digital en una imagen matricial
i=rgb2gray(imagen)% se convierte a escala de grises
imshow(uint8(i)) ,title('Imagen en escala de grises');% muestro la imagen en escala de grises
```

**Figura 18 Script del preprocesamiento a escala de grises.**

```

% En este script se presentan el filtrado de la imagen con ruido del tipo
% sal y pimienta en el dominio de la frecuencia de la Transformada de
% Fourier...
close all, clear all
A=double(imread('E_grises.tif','tif'));
[f,c,l]=size(A);
% Se calcula la transformada discreta de Fourier a través de la función fft2
DFT_A=fft2(A);
% Se calcula el centro el espectro de la frecuencia a través de la función fftshift
DFT_A=fftshift(DFT_A);
% Construcción de un filtro ideal pasabajos
DO=f/1/4;
PasabajosIdeal=zeros(f,c,l);
for i=1:f
    for j=1:c
        for k=1:l
            D=sqrt((i-f/2)^2+(j-c/2)^2+(k-l/2)^2);
            if D<DO
                PasabajosIdeal(i,j)=1;
            end
        end
    end
end
end
% Se multiplica la transformada de la imagen por el filtro ideal
DFT_A_filtrada=DFT_A.*PasabajosIdeal;
% Se aplica la DFT inversa
A_filtrada=ifft2(DFT_A_filtrada);
% Se toma el módulo para descartar cualquier residuo del tipo complejo
A_filtrada=abs(A_filtrada);
% Muestro la imagen filtrada
figure, imshow(uint8(A_filtrada)), title('Imagen filtrada')

```

**Figura 19 Script de filtrado de una imagen usando un filtro pasa bajo.**

```

% En este script se muestra el filtro de mediana empleado en la solución
% para la atenuación de ruido, con las diferentes máscaras de convolución
close all, clear all
% Se carga la imagen en escala de grises
A=double(imread('E_grises.tif','tif'));
% Se le añade ruido del tipo sal y pimienta con desviación estandar de 0.01
fn=imnoise(A,'salt & pepper',0.01);
% Se crea una máscara de convolución con la función fspecial de 3x3
h1=fspecial('average');

% Se aplica el filtro de mediana con la función imfilter
% a la imagen con ruido de sal y pimienta con la máscara de convolución de 3x3
medial=imfilter(fn,h1);
% Se muestra la imagen filtrada con una máscara de convolución de 3x3
imshow(uint8(medial),title('Filtro de mediana 3X3'));

```

**Figura 20 Script de filtrado de una imagen usando un filtro de mediana de máscara 3x3.**

```

% En este script se muestra el filtro de mediana empleado en la solución
% para la atenuación de ruido, con las diferentes máscaras de convolución
close all, clear all
% Se carga la imagen en escala de grises
A=double(imread('E_grises.tif','tif'));
% Se le añade ruido del tipo sal y pimienta con desviación estandar de 0.01
fn=imnoise(A,'salt & pepper',0.01);
%Se crea una máscara de convolución con la función fspecial de 9x9
h2=fspecial('average',[9,9]);
% Se aplica el filtro de mediana con la función imfilter
% a la imagen con ruido de sal y pimienta con la máscara de convolución de
% 9x9
media2=imfilter(fn,h2);
% Se muestra la imagen filtrada con una máscara de convolución de 9x9
imshow(uint8(media2),title('Filtro de media 9X9'));

```

**Figura 21 Script de filtrado de una imagen usando un filtro de mediana de máscara 3x3.**

```

%En este script se implementa un umbral basado en el criterio de Otsu
function umbralOptimo = umbralOtsu(Imagen)
Imagen = fix(double(Imagen));
[f,c]=size(Imagen);
%Se calcula la distribucion de probabilidad de los niveles de gris
P=zeros(1,256);
for i=1:f*c
P(Imagen(i)+1) = P(Imagen(i)+1) + 1;
end
P = P / (f*c);
%Se calcula del momento acumulado de orden cero y del momento acumulado de
%primer orden segun Otsu
omega = zeros(1,256);
omega(1) = P(1);
mu = zeros(1,256);
for i=2:256
omega(i) = omega(i-1) + P(i);
mu(i) = mu(i-1) + (i-1)*P(i);
end
%Se calcula del umbral óptimo segun Otsu
sigmaB2max = 0;
sigmaB2 = 0;

% Se determina laIntensidad media total de la imagen
mut = mu(256);
umbralOptimo = 0;
for t=1:256
    omega1 = omega(t);
    omega2 = 1 - omega1;
    if omega1~=0 & omega2~=0
        mu1 = mu(t) / omega1;
        mu2 = (mut-mu(t)) / omega2;
        sigmaB2 = omega1 * (mu1-mut)^2 + omega2 * (mu2 - mut)^2;
        if sigmaB2 > sigmaB2max
            sigmaB2max = sigmaB2;
            umbralOptimo = t-1;
        end
    end
end
end
end

```

Figura 22 Script de formulación de un umbral usando el criterio de Otsu (11).

```

% En este script se compara el umbral propuesto(umbralOtsu) para el proceso de
% binarización, con otro umbral arbitrario
clear all, close all
% Se carga la imagen con ruido de sal y pimienta
A=double(imread('E_grises.tif','tif'));
% Se Binariza la imagen obtenida usando la función umbralOtsu(ver Figura 18 del anexo)
Botsu = A > umbralOtsu(A);
% Se Binariza otra vez la imagen pero usando un umbral global arbitrario,
% por ejemplo,25
Bglobal = A > 25;
%Se Agrega a la imagen original ruido gaussiano de media 0 y varianza 20:
r=20*randn(size(A));
Ar=fix(min(max(A+r,0),255));
%Se Binariza despues la imagen con ruido gaussiano usando tanto el umbral dado por el
% método de Otsu como el umbral arbitrario usado antes
Botsur = A > umbralOtsu(Ar);
Bglobalr = Ar > 25;
%Se muestra las imágenes binarias resultantes de los dos métodos de umbralización,
%para el caso de la imagen sin ruido, y con ruido gaussiano:
title('Imagen binarizada con umbral Otsu')
title('Imagen con ruido binarizada con umbral Otsu')
figure, subplot(1,2,1), imshow(Bglobal)
title('Imagen binarizada con umbral arbitrario')
subplot(1,2,2), imshow(Bglobalr)
title('Imagen con ruido binarizada con umbral arbitrario')

```

**Figura 23 Script del proceso de umbralización.**

```

% En este script se muestra el proceso de binarización de la imagen
% usando el umbral óptimo de Otsu
% Se le asigna a una variable umbral el umbral de Otsu obtenido en el paso
% anterior
umbral = umbralOtsu(Af);
% Se compara el umbral de otsu con los niveles de gris de la imagen, si estos son menores que el
% umbral predeterminado, se le asigna al píxel de la imagen binarizada el
% valor 0 (color negro), y si es mayor, se le asigna el blanco
Ab = (Af > umbral);
% Se muestra la imagen binarizada
figure, imshow(Ab), title('Imagen Binarizada')

```

Figura 24 Script del proceso de binarización de la imagen.

```

% En este script se aplican operaciones morfológicas a la imagen binarizada
% obtenida en el paso anterior
% Se representa la imagen binarizada como una función de dos variables
% discretas, para representar los objetos de la imagen
EE1 = ones(5);
EE2 = [0 0 1 0 0;
       0 1 1 1 0;
       1 1 1 1 1;
       0 1 1 1 0;
       0 0 1 0 0];

% Se aplica operaciones de dilatación mediante la función imdilate
Am = imdilate(Ab, EE1);

% Se aplica operaciones de erosión mediante la función imerode
Am = imerode(Ab, EE1);

% Se muestran los resultados tras aplicar dichas operaciones morfológicas
figure, imshow(Am), title('Imagen Dilatada')
figure, imshow(Am), title('Imagen Erosionada')

```

Figura 25 Script del proceso de morfología.

```

%En este script se procede a realizar el descarte de objetos pequeños
% Se define un límite o área mínima de búsqueda
areaMinima = 500;
% Se itera hasta la cantidad de objetos encontrados comparando con el área
% mínima
for i=1:cantidadObjetos
    [fO,cO] = find(Ae==i);
    cantidadPixel = length(fO);
    if cantidadPixel < areaMinima
        for j=1:length(fO)
            for k=1:length(cO)
                Am(fO(j),cO(k)) = 0;
            end
        end
    end
end
end
% se procede a contar la cantidad de objetos candidatos a rostros de
% personas
Ae = bwlabel(Am,4);
cantidadObjetos = max(max(Ae));
figure, imagesc(Ae), title(strcat('Objetos candidatos : ',num2str(cantidadObjetos)))
disp(strcat('Cantidad de Objetos Encontrados luego del descarte por area minima: ',num2str(cantidadObjetos)))

```

**Figura 26 Script del proceso de segmentación.**

```

% En este script se muestra el proceso de extracción de características mediante el
% modelo propuesto
% Se define el área de los objetos mediante la función regionprops
aux = regionprops(Ae, 'Area');
vectorArea = [aux.Area];
% Se define la longitud del eje mayor del contorno activo(cantidad de píxeles)
aux = regionprops(Ae, 'MajorAxisLength');
vectorEjeMayor = [aux.MajorAxisLength];
% Longitud del eje menor del objeto (cantidad de píxeles):
aux = regionprops(Ae, 'MinorAxisLength');
vectorEjeMenor = [aux.MinorAxisLength];
% Se define el centroide del objeto (centro de masa):
aux = regionprops(Ae, 'Centroid');
vectorCentroide = [aux.Centroid];
% Se define la orientación de cada objeto:
aux = regionprops(Ae, 'Orientation');
vectorOrientacion = [aux.Orientation];

```

**Figura 27 Script del proceso de extracción de características.**

```

%En este script se presenta la red neuronal aprendida para el
%reconocimiento del numero de personas de la imagen
%
% Universidad de las Ciencias Informáticas
% Departamento de Señales Digitales
% Red neuronal Artificial
Im=PreProcesar('Simular',1);% Comprende la etapa de preprocesamiento,
%segmentación y extracción de características
I=Segmentar(Im);
C=ExtraerCaracteristicas(I,1,1);
testinput = C([1:2],:); % Para 2 características
result = sim(Red,testinput)% se procede a simular la red mediante el comando sim
[n,m]=size(result);
% Calcular Numero Personas
num_personas=0;
for j=1:m;
if result(1,j)==1;
num_personas=num_personas+1;
end
end
num_personas
figure, plotdata(testinput, result, 'o', '*')

```

**Figura 28**Script de la simulación de la red perceptron del toolbox de MATLAB 7.0.

# Glosario

<b>Término</b>	<b>Definición</b>
<b>Resolución Piramidal</b>	Algoritmo para minimizar el tiempo de demora del proceso de matching de contornos aplicando la técnica de Chamfer, permitiendo realizar la operación de matching desde una imagen de menor tamaño a otra de mayor tamaño en correlación a la técnica de detección por borde o contorno desde un mayor nivel hasta la original que es la de menor nivel <b>(24)</b> .
<b>Distribución de probabilidad</b>	Función que asigna a cada suceso definido sobre la variable aleatoria la probabilidad de que dicho suceso ocurra.
<b>Transformación distancia</b>	Proceso en el cual a partir de una imagen binaria, con píxeles de borde y fondo, se crea otra, donde cada píxel posee un valor que indica la distancia al píxel más cercano a un borde <b>(24)</b> .
<b>Variable continua</b>	Es aquella que puede tomar cualquiera de los infinitos valores existentes dentro de un intervalo.
<b>Reflectancia</b>	Medida de la capacidad de una superficie para reflejar energía electromagnética en una determinada longitud de onda <b>(5)</b> .
<b>Convolución</b>	Acción de desplazar el filtro píxel a píxel, iniciando en la posición en la matriz hasta llegar a la penúltima posición <b>(4)</b> .

<b>Energía y Fuerza</b>	En general la fuerza viene dada por el gradiente de energía de la imagen y estas sirven para regularizar la solución <b>(17)</b> .
<b>Umbral</b>	Rango para separar los objetos de una imagen que nos interesen del resto, y definir que píxeles conforman los objetos que se buscan y qué píxeles son sólo el entorno de estos objetos.
<b>Suavizado</b>	Técnica que se obtiene mediante el promedio de los valores de intensidad de los píxeles en el entorno de vecindad con una máscara de convolución para eliminar el posible ruido existente <b>(4)</b> .
<b>B-Spline</b>	Función polinomial a trozos de grado $k-1$ , definida por una secuencia de puntos de control llamados nodos, con los cuales los valores de la función son controlados y donde $k$ es el orden de la más alta derivada <b>(25)</b> .
<b>Spline</b>	Función polinomial capaz de representar eficientemente los contornos de una curva en una imagen <b>(17)</b> .
<b>Fuerzas restrictivas</b>	Fuerzas de la imagen que añaden características de alto nivel que hacen que la técnica "Snake" valla hacia el mínimo local apropiado <b>(4)</b> .
<b>Filtro gaussiano</b>	Técnica de reducción de ruido diseñada para mejorar la imagen donde se presentan ruido.
<b>Mínimo local</b>	Resultado tras evolucionar el contorno propuesto hasta el último local más cercano, es decir modelarlo hasta lograr el contorno deseado <b>(5)</b>

<b>Parametrización lineal</b>	Forma de representar una curva o superficie de una función vectorial como una imagen. Su importancia radica en que permite tratar como funciones a curvas que no lo son <b>(5)</b> .
<b>Filtrado</b>	Proceso para la eliminación de ruido en una imagen. <b>(4)</b>
<b>Gradiente</b>	Valor dado por la dirección de más rápido cambio de intensidad de la imagen. <b>(4)</b>
<b>Máscara de convolución</b>	Técnica para hacer un ajuste o modificación en una zona de la imagen sin que afecte al resto, protegiendo la zona que no debe variar <b>(4)</b> .