

**REGISTRO DE TRANSEÚNTES EN TIEMPO REAL UTILIZANDO UN
SISTEMA DE VISIÓN ARTIFICIAL SOBRE UN AMBIENTE CONTROLADO**

JOAN MAURICIO VALENCIA LÓPEZ

MAURICIO ABRIL CAÑAS

**UNIVERSIDAD TECNOLÓGICA DE PEREIRA
PROGRAMA DE INGENIERÍA ELÉCTRICA
PEREIRA
2007**

**REGISTRO DE TRANSEÚNTES EN TIEMPO REAL UTILIZANDO UN
SISTEMA DE VISIÓN ARTIFICIAL SOBRE UN AMBIENTE CONTROLADO**

JOAN MAURICIO VALENCIA LÓPEZ

MAURICIO ABRIL CAÑAS

**Trabajo de grado presentado como requisito para optar al título de ingeniero
electricista**

**Directora
Ing. Bonie Johana Restrepo Cuestas**

**UNIVERSIDAD TECNOLÓGICA DE PEREIRA
PROGRAMA DE INGENIERÍA ELÉCTRICA
PEREIRA
2007**

Nota de aceptación:

Firma del presidente del jurado

Firma del jurado

Dedicatoria.

A mis padres, que gracias a sus consejos, esfuerzo, dedicación, y responsabilidad, me han guiado por el buen camino hacia el cumplimiento de mis metas.

Joan Mauricio Valencia L.

A mi mamá y a mi abuela por su dedicación.

Mauricio Abril Cañas.

Resumen

El presente trabajo muestra el desarrollo de un sistema de visión artificial libre de parámetros que trabaja en línea o fuera de línea, utilizado para contar transeúntes sobre un ambiente controlado a partir de una secuencia de imágenes capturada con una cámara web sin calibración, sobre un área de $9m^2$ partiendo de la idea que el fondo es estático.

Parte fundamental de la solución del problema se basa en la búsqueda y localización de los transeúntes, que se realiza por medio de la detección de movimiento utilizando sustracción del fondo. Luego, se realiza la segmentación de cada transeúntes en el área de registro por medio de la técnica de umbralización automática que entrega una imagen binaria; este método es lo suficientemente robusto para adaptarse a cambios suaves de iluminación. Se realiza un filtrado espacial no lineal para eliminar píxeles aislados y separar regiones que permitan definir el área de cada región que será etiquetada. A continuación se realiza un seguimiento de las coordenadas del centroide para cada transeúnte por medio de una predicción de movimiento que identifica durante todo el recorrido a cada uno de los objetos de interés como la misma etiqueta en la secuencia de imágenes.

A partir de este seguimiento se identifica la dirección en que se desplazan los elementos de interés (de arriba a abajo y viceversa); permitiendo así obtener un registro general del número de transeúntes que han circulado por el área de conteo.

El sistema está en capacidad de trabajar sobre una base de datos (fuera de línea) ó a partir de la adquisición directa de vídeo desde una cámara web (en línea).

La validación de los resultados se realiza encontrando la matriz de confusión a la base de datos obtenida, de tal forma que se pueda observar la respuesta del sistema a los casos especiales que se puedan presentar.

Agradecimientos

Al MS.c Germán Andres Holguin por su orientación y ayuda en la realización de este proyecto.

A todos los integrantes del grupo de Investigación en Instrumentación y Control, especialmente a la Ing. Bonie Johana Restrepo por su asesoría y dedicación, al Ing. Jason Edwin Molina, MS.c Cristian Guarnizo, MS.c Eduardo Giraldo, MS.c Julian David Echeverry, Ing. Andres Felipe López, Ing. Gustavo Betancourt, Ing. Ricardo Linares, Ing. Damián Alberto Álvarez, Ing. Marta Lucia Guevara, y al Ing. Santiago Sánchez Acevedo.

Y a todos nuestros compañeros de estudio que de una u otra forma hicieron que este proyecto saliera adelante.

Introducción

En nuestros días, existe una creciente demanda por las aplicaciones de control de acceso; Las necesidades de seguridad y supervisión en sitios públicos como centros comerciales, museos, estadios, escuelas, edificios y sistemas de transporte masivo, obligan a realizar inversiones muy altas en soluciones como guardias de seguridad, torniquetes registradores o sensores fotovoltaicos que no garantizan una alta confiabilidad.

Esta actividad en la mayoría de los casos es desempeñada por personas que generan errores en el conteo, proporcionando inexactitud en la medida; debido a que este proceso requiere un alto grado de disponibilidad y adaptación del operario al ambiente, convirtiéndola así en una función monótona y rutinaria.

Por tanto, en la búsqueda de una solución se opta por automatizar el proceso por medio de un sistema de visión artificial. Estos sistemas son altamente utilizados en aplicaciones como la detección y seguimiento de objetos, análisis de terrenos, análisis de imágenes satelitales, navegación de robots móviles, reconocimiento de rostros y huellas dactilares, control y calidad de productos, detección de cuerpos extraños en el ser humano, inspección de motores, entre otras; que convierten a esta estrategia de solución en una herramienta con la que se busca reducir los tiempos de trabajo y los errores sistemáticos.

Aunque el ser humano se puede concentrar, por periodos cortos de tiempo, en eventos que considere importantes, la cantidad de información que se procesa e interpreta es usualmente limitada. Sin embargo, la concepción de visión artificial en el monitoreo y la vigilancia no se limita a reemplazar al operario, sino que le sirve como una herramienta de procesamiento de información que extienda su percepción y razonamiento.

Los sistemas de visión artificial presentan tres grandes aspectos a considerar: los de visión (procesamiento de las imágenes), los de adquisición (posición y estabilización del instrumento de adquisición y ambiente donde se trabaja) y por último los aspectos dependientes de la lente u óptica empleada, dado que la correcta implementación de cada uno de estas etapas indican el éxito o el fracaso de la etapa posterior en el proceso, debido a que se encuentran estrechamente relacionadas.

Se busca desarrollar entonces una metodología, que resuelva el problema del registro automático de personas en un recinto o ambiente controlado, que permita trabajar en tiempo real con un bajo costo computacional; en donde debe estar incluido el tiempo de

procesamiento y respuesta del sistema, que debe ser lo suficientemente pequeño para que el proceso sea confiable y no requiera de hardware elaborado.

Objetivos

OBJETIVO GENERAL

Desarrollar una metodología para la detección de transeúntes, identificación y seguimiento de trayectorias de movimiento en un espacio determinado y controlado, con el fin de mantener un registro de ingreso y salida en tiempo real utilizando un sistema de visión artificial.

OBJETIVOS ESPECÍFICOS

- Implementar un sistema de visión artificial para la detección de personas en un ambiente controlado.
- Implementar una metodología para el seguimiento de los objetivos reconocidos en la secuencia de vídeo.
- Desarrollar un algoritmo para mantener en tiempo real el registro de eventos relevantes ocurridos en la secuencia de vídeo.

CONTENIDO

1. Introducción a los sistemas de visión artificial.	1
1.1. Sistema de visión artificial.	1
1.2. Elementos que integran un sistema de visión artificial.	2
1.2.1. Restricciones de la escena.	2
1.2.2. Adquisición de imágenes.	3
1.2.3. Preprocesamiento.	3
1.2.4. Segmentación.	4
1.2.5. Extracción de características.	6
1.2.6. Clasificación o interpretación.	6
1.2.7. Acción.	7
2. Fundamentos teóricos.	8
2.1. Imágenes digitales.	8
2.1.1. Tipos de imágenes.	8
2.1.2. Características de una imagen en niveles de gris.	12
2.2. Preprocesamiento.	14
2.2.1. Filtrado espacial.	15
2.2.2. Filtrado en el dominio de la frecuencia.	18
2.3. Detección de movimiento.	19
2.3.1. Métodos basado en la estimación del fondo.	20
2.4. Segmentación	20
2.4.1. Métodos basados en histogramas.	21
2.4.2. Métodos basados en regiones y contornos.	22
2.5. Seguimiento en imágenes.	23
2.5.1. Determinación de movimiento empleando flujo óptico.	23
2.5.2. Determinación del movimiento por segmentación de imágenes.	26
3. Arquitectura del sistema.	32
3.1. Adquisición de la imagen RGB.	33
3.2. Preprocesamiento.	33
3.2.1. RGB a niveles de gris.	34
3.3. Detección de movimiento.	34
3.3.1. Captura del fondo de la imagen.	35
3.3.2. Sustracción del fondo.	35
3.4. Segmentación.	35

3.4.1. Binarización.	35
3.4.2. Filtrado espacial no lineal.	36
3.5. Extracción de características.	37
3.5.1. Etiquetado de los objetos.	37
3.5.2. Seguimiento (tracking).	38
3.5.3. Registro.	39
4. Implementación del sistema y análisis de resultados.	40
4.1. Resultados experimentales.	42
5. Conclusiones	47
6. Discusiones y trabajos futuros.	49
A. Referencias de las figuras	51

LISTA DE FIGURAS

1.1. Modelo genérico de un sistema de visión artificial.	2
1.2. Representación de una imagen digital.	3
1.3. Reducción de ruido gaussiano utilizando filtrado lineal.	4
1.4. Segmentación de una imagen usando umbralización estática.	5
1.5. Segmentación orientada a bordes y regiones.	5
2.1. Representación de una imagen a color	9
2.2. Modelo de color RGB.	9
2.3. Espacio de color de Hering.	10
2.4. Modelos de color.	10
2.5. Imagen en niveles de gris.	11
2.6. Imagen binaria.	11
2.7. Resolución de las imágenes.	13
2.8. Igualación del histograma.	13
2.9. Plantilla 3 x 3 de promediado filtro pasa bajo.	15
2.10. Promedio ponderado de píxeles.	16
2.11. Tipos de conectividad.	17
2.12. Aplicación de un elemento estructural.	18
2.13. Parámetros del histograma.	21
2.14. Segmentación de imágenes utilizando segmentación en líneas, bloques y grupos.	27
2.15. Diferencia de imágenes, actualización del fondo.	28
2.16. Diferencia de imágenes, actualización del fondo.	30
3.1. Diagrama de bloques.	32
3.2. Intervalos de tiempo disponible para realizar el procesamiento de las imágenes.	33
3.3. Cambio de imágenes RGB a niveles de gris.	34
3.4. Histograma bimodal.	36
3.5. Binarización mediante umbralización dinámica.	36
3.6. Ubicación del centroide y etiqueta de un transeúnte en una imagen binaria.	38
3.7. Seguimiento de un objeto usando un método de predicción de movimiento.	39
4.1. Ubicación de la fuente de adquisición.	41
4.2. Panel frontal.	41
4.3. Registro de persona que sube y que baja.	43
4.4. Registro en ambos sentidos en un mismo instante de tiempo.	44
4.5. Registro de varias personas.	44

4.6. Fp presentado por el sistema.	45
4.7. Fn presentado debido a un translope entre transeúntes.	46

LISTA DE TABLAS

4.1. Tiempos de procesamiento empleados por el sistema	42
4.2. Matriz de confusión para 30 y 15 frames por segundo.	43
A.1. Referencias bibliográficas de figuras.	51

1. INTRODUCCIÓN A LOS SISTEMAS DE VISIÓN ARTIFICIAL.

En la actualidad la visión artificial se ha convertido en una herramienta utilizada en gran parte de trabajos de inspección industrial y de vigilancia en general; pues, lo que busca es determinar en el menor tiempo posible, que grandes bloques de producción cumplan estándares fijos en los que se disminuya en lo posible los errores de tipo humano.

En la búsqueda de la automatización de estos problemas, se encuentran dos campos distintos que son, la visión computacional y la visión artificial. La visión computacional surgió con la idea de conectar una cámara de vídeo a un computador, para que las imágenes capturadas fueran interpretadas y así obtener una representación simbólica de los objetos presentes en las escenas analizadas [1]. Esto con el propósito de brindar al sistema de visión, ver y entender el mundo en tiempo real, sin ninguna intervención humana. Mientras que la visión artificial cumple la misma función pero adaptándose a las limitaciones tecnológicas que impiden imitar fielmente la visión humana.

1.1. Sistema de visión artificial.

Un sistema de visión artificial puede ser definido como un sistema que permite adquirir y analizar imágenes de forma automática, en las que se encuentran los datos necesarios para controlar un proceso o actividad definida con anterioridad. La finalidad de éste sistema es imitar en lo posible el sistema de visión humana [2].

Para considerar que un sistema sea de visión artificial debe cumplir con las siguientes tareas:

- Realizar captura y análisis de imágenes, dado que el sistema debe realizar estas actividades en conjunto de tal forma que el grado de procesamiento dependerá en gran medida de la calidad de la información adquirida.
- Debe tener definido el problema previamente, es decir, establecer las variables que se desean analizar o cuantificar dentro de la solución del problema.
- El sistema debe estar capacitado para obtener los datos de interés, ya que dentro de su implementación se deben definir algoritmos que delimiten los datos de tal forma que el costo computacional sea reducido.

- Por último, el sistema debe ser automático, lo que implica que el sistema de visión actúa de forma autónoma al momento de realizar los análisis de la información presente en las imágenes obtenidas.

1.2. Elementos que integran un sistema de visión artificial.

Un sistema de visión artificial (ver figura 1.1), se puede representar a partir de un modelo que cumpla con las siguientes funciones fundamentales: establecer que restricciones debe cumplir el ambiente en donde se desean adquirir las imágenes; capturar, extraer, analizar y reconocer características propias de la imagen que permitan tomar decisiones de acuerdo a la información obtenida; esto con el fin de que el sistema de visión artificial realice una tarea específica.

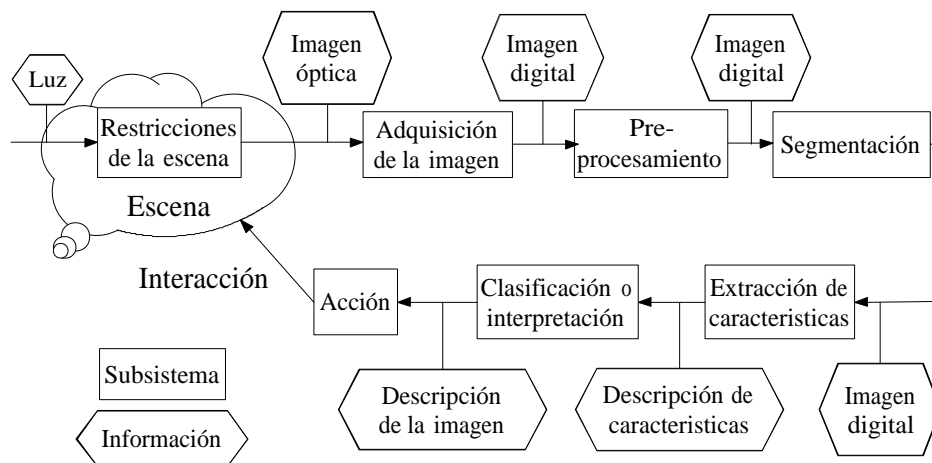


Figura 1.1. Modelo genérico de un sistema de visión artificial.

El modelo anterior se encuentra dividido por subsistemas que generan diferentes tipos de información que se encuentra distribuida desde el momento de la adquisición de la imagen, hasta la toma de una decisión que genere la interacción directa o indirecta con el mundo real [3]. Cada una de las partes que conforman este modelo se explica a continuación.

1.2.1. Restricciones de la escena.

El objetivo de imponer restricciones en la escena es reducir la complejidad del problema limitándolo a un ambiente controlado, puesto que en el ambiente de estudio existen variables aleatorias como son: el nivel de luminosidad, sombras, traslapes entre objetos, oclusiones, entre otros, que influyen en el éxito o fracaso de un sistema de visión artificial. En el caso del nivel de iluminación, este puede ser una limitante para lograr un buen contraste del objeto de estudio con el fondo, de tal manera que se dificulta el procesamiento y análisis

posterior de las imágenes adquiridas. El problema radica en que el nivel de iluminación de la escena se determina de forma experimental, puesto que no existe una técnica que indique de forma precisa el número y disposición angular de las luminarias que se deben colocar para determinada aplicación.

1.2.2. Adquisición de imágenes.

La adquisición de imágenes involucra el proceso de transformación de estímulos de luz que reflejan los objetos observados y que inciden en los foto-sensores de una cámara de vídeo, los cuales son convertidos a valores digitales almacenados en la memoria del computador para su posterior análisis.

Normalmente la adquisición se realiza con Dispositivos de Carga Acoplada *CCD* por sus siglas en inglés (**Charge Couple Device**), que se encargan de transformar la velocidad de la luz que incide en ellos a cargas eléctricas, para así generar una señal eléctrica de vídeo similar a la de una videograbadora, que será digitalizada por un convertidor Analógico/Digital con el fin de obtener un marco base o **Frame Grabber**, que genera una imagen digital interpretable por la computadora designada, es decir, valores numéricos que representan los elementos de la imagen conocidos como colores [3]. Véase figura 1.2.

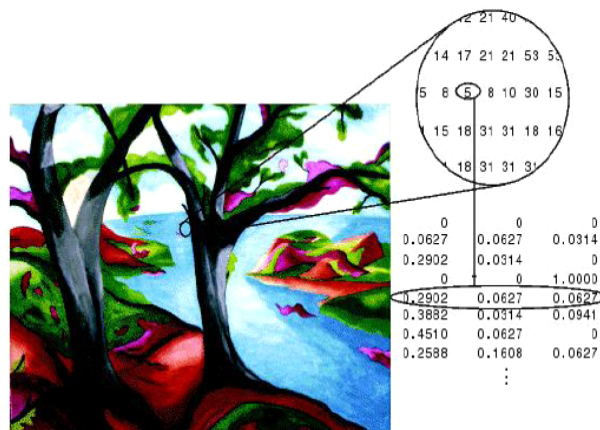
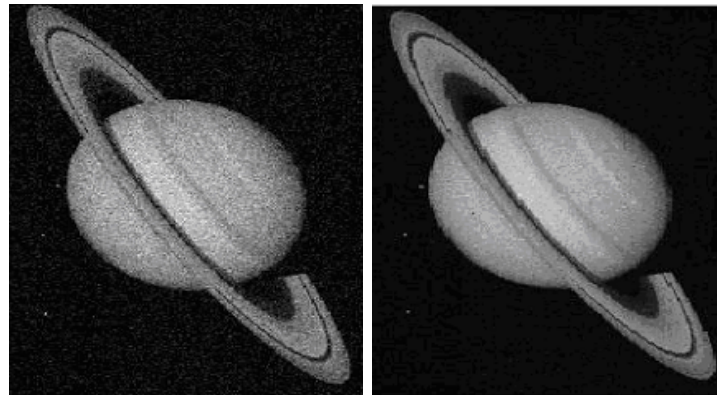


Figura 1.2. Representación de una imagen digital.

1.2.3. Preprocesamiento.

La función primordial de este subsistema es reconstruir la imagen original en un formato digital lo más fielmente posible, dado que en el proceso de adquisición la imagen puede sufrir pérdidas en calidad, distorsión geométrica, desenfoque y además la información recibida puede contener ruido.

El preprocesamiento, o mejor conocido como procesamiento de bajo nivel, busca solucionar de la mejor manera estos problemas utilizando técnicas de suavizado y realzado. El primero se encamina hacia la supresión del ruido introducido durante la captura de la imagen, mientras que el segundo está encaminado a eliminar falsos reflejos y sombras que dificulten la extracción de la información para el procesamiento de la imagen. Algunas de las técnicas empleadas para realizar esta labor son el ajuste de contraste, expansión del histograma, filtrado, entre otras técnicas que permitan eliminar los imperfectos presentes al momento de adquirir la imagen [3]; véase figura 1.3, en donde, la figura (a) ilustra una imagen que presenta ruido aditivo ó comúnmente llamado ruido gaussiano y la figura (b) Representa la misma imagen preprocesada a partir de un filtro Wiener (filtro lineal).



(a) Imagen con ruido gaussiano.

(b) Imagen filtrada.

Figura 1.3. Reducción de ruido gaussiano utilizando filtrado lineal.

1.2.4. Segmentación.

Es el subsistema encargado de dividir una imagen de entrada en sus partes constituyentes u objetos, a partir de los que se extraen una serie de parámetros que posibiliten su clasificación; de tal forma que permitan al sistema diferenciar a los objetos de interés de los que no lo son.

Al momento de segmentar una imagen se deben buscar propiedades específicas de cada uno de los objetos de estudio, dado que en general lo que se busca es diferenciar los objetos del fondo de la imagen ó de otros objetos. El sistema debe estar configurado con los parámetros a identificar que permitan una correcta partición de la regiones ó segmentos de la imagen que faciliten los procesos siguientes. Véase figura 1.4, donde (a) muestra la imagen original en niveles de gris, mientras que en (b) Se ilustra la misma imagen después de realizarse una segmentación, utilizando una umbralización estática, que entrega una imagen de tipo binaria.

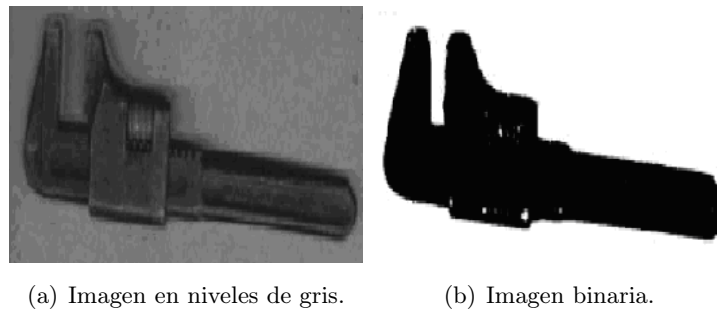


Figura 1.4. Segmentación de una imagen usando umbralización estática.

En general la segmentación es uno de los problemas más difíciles de resolver; ésta etapa del proceso determina el éxito o el fracaso del análisis. Por esta razón, lo que se busca es obtener una segmentación robusta. En general los algoritmos de segmentación se basan en dos principios fundamentales: *discontinuidad* (orientada a bordes) y *similitud* (orientada a regiones). En donde una región se define como un área de la imagen en la que sus píxeles poseen propiedades similares de intensidad, color, textura entre otros; mientras que un borde es una línea que separa dos regiones que poseen diferentes propiedades (discontinuidad). Véase figura 1.5 donde se observan la segmentación orientada a bordes y orientada a regiones.

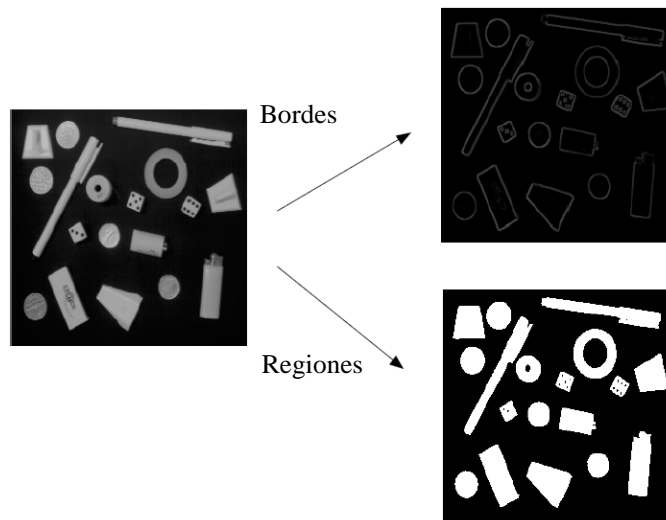


Figura 1.5. Segmentación orientada a bordes y regiones.

Para la detección de bordes, basándose en el hecho de discontinuidad, se proponen los siguientes tipos de operadores: primera derivada, segunda derivada y operadores morfológicos; mientras que para la detección de regiones se realizan técnicas basadas en el hecho de la similitud como son: binarización basada en la umbralización, crecimiento de regiones mediante la adición de píxeles, división de regiones y similitud de textura, color o niveles de gris [4].

Idealmente para lograr una segmentación apropiada, la imagen debe tener las siguientes características [5]:

- Las regiones deben ser uniformes y homogéneas con respecto a una característica específica.
- El interior de las regiones debe ser simple y sin demasiados orificios pequeños.
- Las regiones adyacentes deben tener valores significativamente diferentes con respecto a la característica a la que son uniformes.
- El borde de cada región debe ser simple, sin discontinuidades pronunciadas.

1.2.5. Extracción de características.

El subsistema de extracción de características es el encargado de distinguir y describir las características de los objetos presentes en la imagen, de forma tal que permita su identificación de modo acertado.

La caracterización se realiza partir de dos tipos de descriptores: los descriptores de bordes, cuyo objetivo es la identificación de bordes mediante el ajuste de rectas, curvas, funciones polinómicas y códigos encadenados; y los descriptores de regiones, encaminados a obtener propiedades tales como: color, textura, superficie, posición, nivel de intensidad y estructura geométrica sin importar su orientación [6]. La elección del descriptor que se desea utilizar depende del tipo de segmentación utilizado.

1.2.6. Clasificación o interpretación.

Este subsistema es el que indica a que grupo o clases predefinidas pertenece un objeto extraído de una imagen. En donde la similitud de todas o algunas características entregadas por los descriptores del objeto, son las que brindan los criterios de ubicación dentro de un grupo.

En la mayoría de los casos esta clasificación se realiza por medio de comparación de plantillas, modelos estadísticos o en su defecto utilizando varias configuraciones de redes neuronales que brinden un alto grado de acierto. Dado que dependiendo de la complejidad del problema, el clasificador debe estar en la capacidad de tomar la mejor decisión, que defina en que grupo o clase se debe ubicar la característica extraída que describe el objeto dentro de la imagen. Cabe resaltar que la elección del clasificador depende en gran medida de la complejidad del problema y del grado de exactitud que se desea tener [6].

1.2.7. Acción.

Es el subsistema final que se encarga de interactuar de forma directa o indirecta con el mundo físico, ya que este entrega la información obtenida de la imagen procesada a un usuario ó en su defecto brinda una orden a un equipo (robot), para que ejecute un trabajo dependiendo de los resultados obtenidos. Esta acción, debe estar definida con anterioridad y depende en gran medida de la función que desempeñe el sistema de visión artificial [6].

2. FUNDAMENTOS TEÓRICOS.

En este capítulo se describen las técnicas de procesamiento de imágenes y algunos conceptos básicos utilizados en los sistemas de visión artificial; enfocados a la detección de objetos móviles.

2.1. Imágenes digitales.

Una imagen digital es una función de intensidad de luz bi-dimensional representada por la función $f(x,y)$, donde (x,y) representan las coordenadas espaciales y $f(\cdot)$ es el valor en cualquier punto del brillo ó intensidad (nivel de gris de la imagen). Mientras que en una imagen multicolor f es un vector, en el que cada componente indica el brillo de la imagen en el punto (x,y) de la banda de color a la que corresponde; donde cada elemento de la serie es llamado *píxel* derivado del término **picture element**. [7] [8].

$$f(x,y) = \begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0,N-1) \\ f(1,0) & f(1,1) & \cdots & f(1,N-1) \\ \vdots & \vdots & & \vdots \\ f(N-1,0) & f(N-1,1) & \cdots & f(N-1,N-1) \end{bmatrix} \quad (2.1)$$

2.1.1. Tipos de imágenes.

Imágenes a color.

Son imágenes compuestas por cuatro valores almacenados en tres planos de píxeles o matrices véase figura 2.1, que se encuentran definidos por los modelos de color, entre los que se encuentran: el modelo RGB (**red, green, blue**), el modelo de color de Hering, el modelo HSL (**hue, saturation y luminance**), el modelo HSV (**hue, saturation, value**) y el diagrama de cromaticidad establecido por la CIE (**Comisión Internationale de l'Eclairage**) [4] [7].

En el modelo RGB cada color aparece en sus componentes espectrales primarias rojo, verde, azul. La razón por la que se toman estos tres colores tiene que ver con la teoría tricromática de color que indica que tres estímulos cromáticos diferentes, combinados en diferentes proporciones, pueden producir el efecto equivalente a muchos más estímulos que una persona percibirá como muchos colores diferentes.

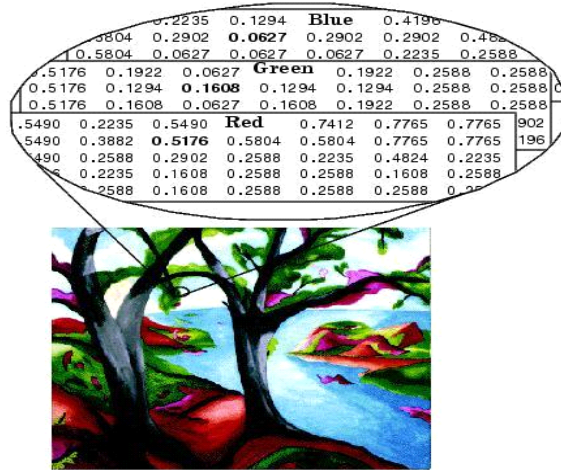


Figura 2.1. Representación de una imagen a color

Este modelo está ubicado en el sistema de coordenadas cartesianas positivas del espacio tridimensional. El sub-espacio de color de interés es el tetraedro mostrado en la figura 2.2, en el que los valores RGB están en tres vértices; mientras que el cian, magenta y amarillo se sitúan en otros tres vértices, el negro corresponde al origen y el blanco en el vértice más alejado del origen. En este modelo, los niveles de gris se extienden desde el negro hasta el blanco a lo largo de la diagonal que une estos dos puntos y los colores son vectores normalizados definidos desde el origen [4] [7].

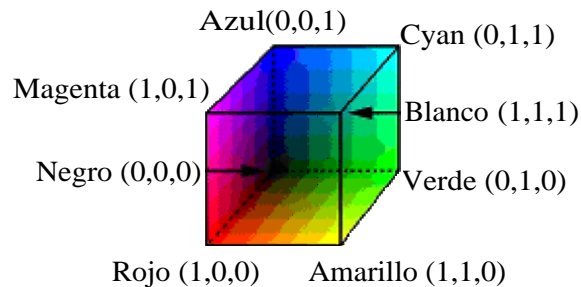


Figura 2.2. Modelo de color RGB.

Por otra parte el modelo de color de Hering utiliza las partes negativas de los ejes x , y del espacio cartesiano tridimensional pero de una forma particular. Basado en su teoría de percepción de color, Hering estableció parejas de colores opuestos, el amarillo opuesto al azul y el rojo opuesto al verde. Al igual que en el modelo RGB y en cualquier modelo de color, este permite asignar los valores que designan las características de color de la imagen como

son la intensidad y el brillo (eje z) [9], véase figura 2.3.

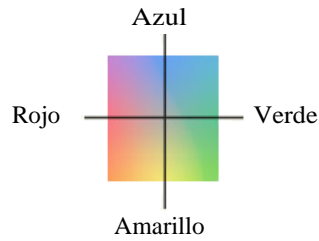


Figura 2.3. Espacio de color de Hering.

Existe también un modelo de color basado en el diagrama de cromaticidad establecido por la CIE, que está formado por un plano en el que se distribuye el espectro de luz visible para los seres humanos (aproximadamente entre 380nm (violeta) y 700nm (rojo) en longitud de onda), que se ubican en el diagrama alrededor de una línea curva en el diagrama de cromaticidad. Dichos colores son los colores puros del espectro (colores saturados). Dado que cualquier otro punto no exactamente en el borde pero dentro del diagrama representa una mezcla de colores del espectro, que entre más se acerque al centro del diagrama (color blanco) será menos saturado dada la influencia de la luz blanca [4] [9] [10].

En el modelo de color HSL ó HSI, que se representa como un doble cono, los ápices corresponden al blanco y el negro, los parámetros angulares son los colores, la distancia desde el eje a las esquinas del cono corresponde a la saturación y la distancia a lo largo del eje blanco-negro corresponde a la luminosidad. Mientras que en el modelo de color HSV ó HSB, la intensidad varía de negro a blanco en un único cono, en el que se puede identificar tres parámetros que son color, saturación y brillo, véase figura 2.4 [9].

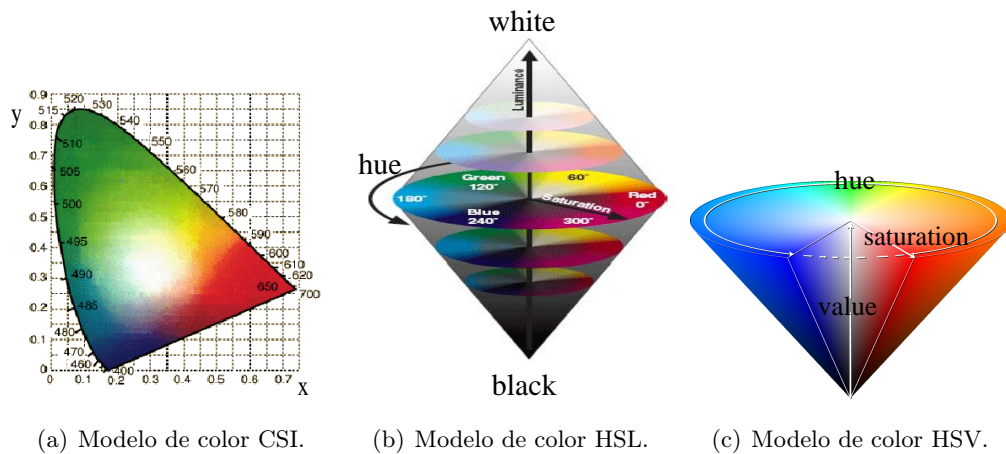


Figura 2.4. Modelos de color.

Cabe destacar que la aplicación de alguno de estos modelos de color depende más de la característica que se desee resaltar en una imagen a color y del problema que se busca solucionar.

Imágenes en niveles de gris.

Son imágenes compuestas de un simple plano o matriz de píxeles, en donde sus valores están representados entre $[0, 255]$ para una imagen de 8-bit, en donde el 0 representa el color negro y el 255 el color blanco, dado que en ese intervalo se ubican todas las intensidades ó grises disponibles en la imagen, véase figura 2.5 [7] [11].

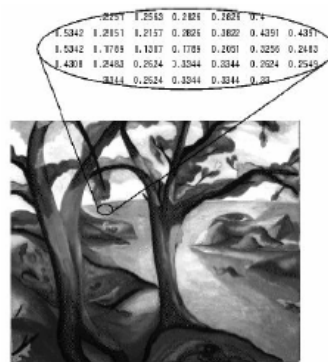


Figura 2.5. Imagen en niveles de gris.

Imágenes binarias.

Son aquellas donde cada píxel asume valores discretos de unos o ceros. Esencialmente estos valores corresponden a encendido o apagado; en donde una imagen binaria es almacenada con un orden lógico de ceros o píxeles apagados y unos o píxeles encendidos [7], véase figura 2.6.

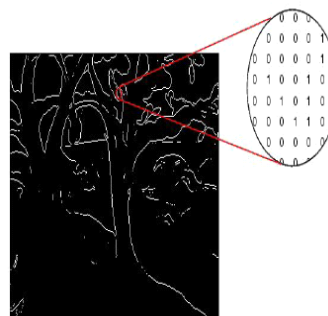


Figura 2.6. Imagen binaria.

2.1.2. Características de una imagen en niveles de gris.

Tamaño de las imágenes: Las dimensiones de una imagen se representan por medio de los pares ordenados de píxeles (M,N) , donde M representa el número de filas de y N el número de columnas. En la mayoría de los casos se opta por tener M igual a N dado que la igualdad entre filas y columnas simplifica el cálculo y presenta grandes ventajas computacionales, cuando las imágenes a procesar tienen tamaños en potencias de dos (128 x 128, 256 x 256, 512 x 512). De tal forma que al momento de almacenar estas imágenes el número de **bits** necesarios para ser almacenadas se calcula con la siguiente expresión:

$$b = N \times N \times m \quad (2.2)$$

Donde m representa los distintos niveles de gris de la imagen, por ejemplo para una imagen de 512 x 512 con 256 niveles de gris (2^8) $m=8$ son necesarios 2.097.152 **bits** o 262.144 **bytes**; es por esto que a mayor número de **bytes** empleados para el almacenamiento de una imagen, mayor será su resolución [11].

Resolución de la imagen: La resolución de una imagen expresa cuanto detalle el observador puede ver en ella. Esta característica depende directamente de los valores de N y de m , dado que tener un mayor número de píxeles y niveles de gris en una imagen, da como resultado una mejor visualización de sus detalles; pero a su vez incrementa el tamaño de almacenamiento. Véase figura 2.7, en donde (a) muestra la imagen original y (b) la misma imagen pero solo utilizando el 2.78% de píxeles de la imagen original. Mientras que en (d) se muestra la representación de una imagen pero disminuyendo m (niveles de gris) de la imagen mostrada en (c).

Histograma: El histograma de una imagen es una función discreta que representa el número de píxeles en la imagen ubicados en el eje y , en función de los niveles de intensidad ó gris presentes en la imagen ubicados en el eje x . En donde el tipo de distribución encontrada en el histograma representa el grado de contraste presente en la imagen; si los datos presentan una distribución uniforme en todo el rango del niveles de gris, se obtendrá un alto contraste que permitirá observar mejor los objetos presentes en la imagen, véase figura 2.8(b). En caso contrario, cuando el histograma muestra que los píxeles se encuentran concentrados solo en algunos valores de los niveles de gris el contraste será bajo. Véase figura 2.8(a) [11].



(a) Imagen original.



(b) Imagen con el 2.78% de píxeles de la imagen original.

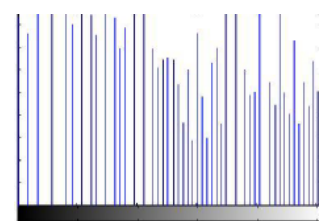
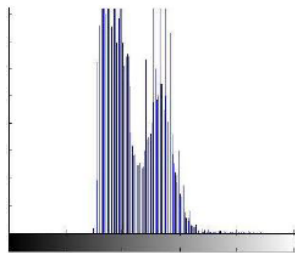


(c) Imagen con $m=8$.



(d) Imagen con $m=2$.

Figura 2.7. Resolución de las imágenes.



(a) Imagen original con su respectivo histograma.

(b) Imagen después de realizada una igualación del histograma.

Figura 2.8. Igualación del histograma.

Al proceso de distribuir uniformemente todos los píxeles sobre los niveles de gris disponibles para la aplicación, se le conoce comúnmente como *igualación del histograma* [4] [12].

2.2. Preprocesamiento.

Todas las imágenes tienen una cantidad de ruido (valores distorsionados), bien debidos al CCD de la cámara, al medio de la transmisión de la señal ó al mismo ruido del sub muestreo, entre otros factores; el tipo de ruido presente en las imágenes puede clasificarse en las siguientes categorías:

- **Gaussiano:** Este tipo de ruido produce pequeñas variaciones en la imagen, siendo producido en la mayoría de los casos por las diferentes ganancias en el sensor, ruido en el digitalizador, perturbación en la transmisión y demás. Se considera siempre que el valor final del píxel sería el ideal más una cantidad correspondiente al error, que puede describirse como una variable gaussiana.
- **Impulsional:** Este tipo de ruido afecta principalmente el valor del píxel, debido a que el valor que toma este no tiene relación con el valor ideal, sino con el valor del ruido que generalmente posee valores muy altos o bajos.
- **Frecuencial:** Es la imagen resultante de la multiplicación entre la imagen ideal y otra señal, en donde la interferencia está caracterizada por ser una señal senoidal de frecuencia determinada.
- **Multiplicativo:** La imagen es fruto de la multiplicación de dos señales de tipo aleatorio.

De la teoría aplicada al tratamiento digital de señales es sabido, que el filtrado de imágenes se puede realizar tanto en el dominio de la frecuencia como en el dominio espacial en el instante en que un sistema lo requiera. En el dominio de la frecuencia la operación de filtrado se reduce a una simple multiplicación de transformadas, mientras que en el dominio espacial el filtrado resulta ser una operación de convolución; en donde el filtro utilizado puede verse como un mecanismo de cambio ó transformación de una señal de entrada a la que se le aplica una función conocida como función de transferencia, para obtener una función de salida; en donde la señal es una función de una o dos variables independientes [13].

Dada la variedad de ruido y los problemas que este puede causar en el contenido de información de una imagen que conforma una secuencia de vídeo, se hace necesario implementar distintas técnicas que reduzcan este fenómeno haciendo especial énfasis en las imágenes en niveles de gris, dado su facilidad de estudio y su velocidad de procesamiento.

2.2.1. Filtrado espacial.

Filtrado espacial lineal: Esta técnica utiliza la función $f(I_1(x,y))$ con el fin de obtener la imagen I_2 en la que cada píxel de I_1 es sustituido por un promedio ponderado de píxeles vecinos [14], como se enuncia en la siguiente expresión matemática.

$$I_2 = f(I_1(x,y)) = \sum_{i,j \in W} \alpha(i,j) I_1(X-i, Y-j) \quad (2.3)$$

En esta ecuación X, Y hacen referencia a la posición de la imagen en la que el filtro se encuentra realizando su función, mientras que W se extiende al conjunto de píxeles cercanos utilizando cálculos entre las vecindades, y $\alpha(i,j)$ son los coeficientes de ponderación del filtro.

Una interpretación equivalente al filtrado espacial es convolucionar la imagen con la respuesta impulsional del filtro, que puede representarse mediante una plantilla ó máscara que en casos particulares posee todos los coeficientes iguales con los que se realiza un promedio [12].

Como ejemplo, se tiene W de tamaño 3 x 3, representada por la siguiente plantilla, véase figura 2.9.

1/9	1/9	1/9
1/9	1/9	1/9
1/9	1/9	1/9

Figura 2.9. Plantilla 3 x 3 de promediado filtro pasa bajo.

En donde los valores representados en la plantilla anterior son utilizados de forma que puedan ser aplicados directamente sobre una imagen con el fin de suavizar o eliminar una característica [12], entregando como resultado la siguiente expresión.

$$I_2 = 1/9(I_1(x,y) + I_1(x-1,y-1) + I_1(x,y-1) + I_1(x+1,y-1) + I_1(x-1,y) + I_1(x+1,y) + I_1(x-1,y+1) + I_1(x,y+1) + I_1(x+1,y+1)) \quad (2.4)$$

En general el filtrado de una imagen $M \times N$ consiste en calcular una imagen $I_2(x,y)$ para todos los $M \times N$ píxeles mediante la ecuación 2.1, ó análogamente desplazar la plantilla

barriendo todas las posiciones de la imagen, dando para cada una de ellas una salida igual a la suma del producto de todos los coeficientes por el valor de los píxeles que se encuentran dentro de la máscara utilizada [13]. En ocasiones se suelen dar distintas ponderaciones a los diferentes píxeles que serán utilizados con el fin de conformar filtros lineales como son, los filtros paso bajo, los filtros paso alto, los filtros pasa banda, entre otros [12].

Para mostrar mejor la aplicación de una máscara sobre una imagen, véase figura 2.10. En la que se muestra como los coeficientes de la máscara multiplican y luego suman los valores que rodean al punto central donde actúa la máscara, con el fin de establecer su valor (este proceso se realiza para cada píxel de la imagen).

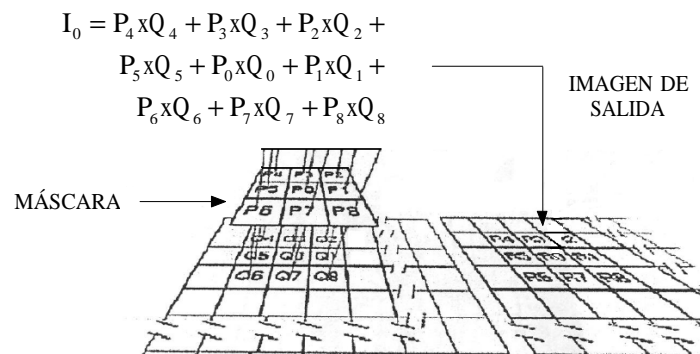


Figura 2.10. Promedio ponderado de píxeles.

Filtrado espacial no lineal: Dentro de estos filtros también se encuentran los filtros espaciales no lineales que operan sobre entornos. Sin embargo, su operación se basa directamente en los valores de los píxeles en el entorno en consideración. Unos ejemplos de filtros no lineales habituales son los filtros mínimo, máximo y de mediana que son conocidos como filtros de rango [11].

Filtro de mediana: Tiene un efecto de difuminado de la imagen, y permite realizar una eliminación de ruido impulsional de forma eficaz, calculando la media de de los píxeles en los que opera el filtro

Filtro de máximo: También llamado erosión que se representa por el símbolo (\ominus), se emplea para buscar los puntos más brillantes de una imagen, debido a la propiedad que posee de adelgazar líneas.

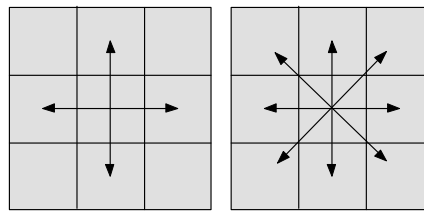
$$A \ominus B = \{x \in E^N / x + b \in A \text{ para todo } b \in B\} \quad (2.5)$$

En donde A es el valor que toma el píxel dentro de la imagen, mientras que B es el valor que conforma la máscara binaria conocida como elemento estructural, empleado para realizar las operaciones morfológicas.

Filtro de mínimo: También llamado dilatación que se representa por el símbolo (\oplus) , se emplea de forma contraria al anterior ya que posee la propiedad ensanchar las líneas negras de la imagen.

$$A \oplus B = \{a + b : a \in A \text{ y } b \in B\} \quad (2.6)$$

Para implementar estos filtros es necesario especificar la vecindad ó conectividad de cada píxel, ver figura 2.11, En donde (a) muestra una conectividad 4 y (b) Una conectividad 8. Generalmente esto se hace indicando el número de píxel de un bloque que rodea al píxel actual considerando este como el centro del bloque [11] [4]. Básicamente funciona como un filtro lineal pero la operación que se realiza sobre cada píxel es no lineal y específica de cada filtro [7].



(a) Conectividad 4. (b) Conectividad 8.

Figura 2.11. Tipos de conectividad.

En la siguiente figura 2.12, se ilustra como un elemento estructural de conectividad 4 compuesto por valores de 1 y 0 y centrado en P_0 , actúa sobre una imagen.

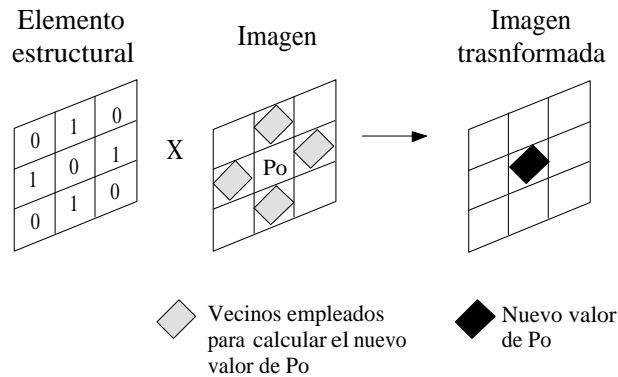


Figura 2.12. Aplicación de un elemento estructural.

Otra clasificación de los filtros espaciales puede hacerse basándose en su finalidad; es el caso de los filtros de realce **sharpening** que se utilizan para realzar pequeños detalles en la imagen y los filtros de suavizado **smoothing** que se aplican para difuminar la imagen. A su vez se tienen los filtros diferenciales que se componen de varios tipos de máscaras (Laplaciano, Prewitt, Sobel, y demás.), cuya aplicación principal es la detección de bordes [8].

2.2.2. Filtrado en el dominio de la frecuencia.

Al momento de adquirir una imagen con un sensor digital, esta puede ser representada como una señal discreta, que permite obtener sus componentes en el dominio de la frecuencia espacial (las altas y bajas frecuencias que la conforman) [13]. Es por esto que a las imágenes se les pueden aplicar transformadas tales como la Transformada Discreta Directa de Fourier Bidimensional, que en el caso de las imágenes en niveles de gris, $I(x,y)$ viene dada por la ecuación:

$$I(u,v) = F\{I(x,y)\} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} I(x,y) e^{-j\frac{2\pi ux}{N}} e^{-j\frac{2\pi vy}{N}} \quad \forall u, v = 0, 1, \dots, N-1 \quad (2.7)$$

Donde $F\{I(x,y)\}$ representa la aplicación de la transformada de Fourier a la imagen $I(x,y)$. Esta ecuación se aplica por lo general a imágenes cuadradas de tamaño $N \times N$ para reducir costo computacional.

En el dominio de la frecuencia también se encuentran los filtros Butterworth determinados por la expresión general:

$$H(u,v) = \frac{1}{1 + [D(u,v)/D_0]^{2n}} \quad (2.8)$$

filtro pasa bajo de Butterworth de orden n y con frecuencia de corte D_0 .

$$H(u,v) = \frac{1}{1 + [D_0/D(u,v)]^{2n}} \quad (2.9)$$

filtro pasa alto de Butterworth de orden n y con frecuencia de corte D_0 .

Lo que caracteriza estos filtros es que en el dominio de la frecuencia, su respuesta en filtrado es bastante plana lo que permite el análisis de imágenes, pero debido a que su caída de filtrado es lenta, se hace necesario buscar filtros de orden superior [4] [13].

2.3. Detección de movimiento.

El análisis de una secuencia de imágenes se puede ver desde dos puntos de vista, que corresponden a los dos enfoques principales para imágenes en movimiento; el primero de ellos es el enfoque continuo, donde se considera la secuencia de imágenes como un flujo de intensidades cambiantes y el segundo punto de vista es el discreto, aquí se considera la secuencia de imágenes como un conjunto de diferentes imágenes estáticas [6].

El estudio de las características debidas al movimiento, generalmente ayudan a la extracción de los objetos presentes en las imágenes; debido a que los puntos del mismo objeto presentan unas velocidades semejantes. Existen distintos tipos de movimiento que son:

- Movimiento de la cámara.
- Movimiento de los objetos.
- Cambios en la iluminación.
- Cambios en la estructura, forma o tamaño del objeto.

Al momento de analizar el movimiento son tres los pasos que hay que considerar; el primer paso es determinar en que zonas de la imagen existe algún objeto que esté moviéndose. En esta primera etapa no es importante la exactitud, sólo detectar las futuras zonas de interés en las que se realizará un análisis más detallado. En la segunda etapa la atención sólo está en

aquellas zonas donde se haya detectado actividad, para así extraer la información que pueda ser usada para el reconocimiento de objetos, análisis de movimiento y demás. En la última etapa se tiene en cuenta el conocimiento previo del problema en concreto; así, por ejemplo la información obtenida de un objeto pequeño que se mueve lentamente cerca de la cámara debe ser igual a la de otro de igual forma que se mueve rápidamente pero lejos de ella [15].

2.3.1. Métodos basado en la estimación del fondo.

Independientemente del tipo de imágenes, de la secuencia de vídeo con la que se trabaja y de las características ambientales, la sustracción del fondo es un método comúnmente utilizado para detectar y segmentar regiones en movimiento en tiempo real, dado que este método provee la información más completa del objeto, pero a su vez es extremadamente sensible a los cambios dinámicos en la escena, como son cambios de iluminación y movimientos de la cámara; dado que el modelo realiza la comparación de cada nueva imagen con un modelo del fondo de la escena [16] [17].

Para realizar la estimación inicial del fondo se pueden utilizar dos procedimientos diferentes: de forma supervisada, realizando una captura de la escena cuando no hay objetos móviles; o de forma no supervisada, que el sistema actualice la imagen cada vez que detecte que no existe un peatón dentro del área de interés [16].

En el caso de trabajar con un fondo estático se puede modelar como una constante o empleando el filtro de media; este método funciona para casos simples, ya que para casos donde el fondo se encuentra en movimiento como en el caso de árboles siendo movidos por el viento, se aplican métodos más sofisticados tales como modelos estadísticos para cada píxel, la desventaja de estos métodos es el elevado costo computacional que impide su aplicación en sistemas que funcionen en tiempo real. Aunque algunos métodos intentan reducir este problema realizando la estimación de fondo, fuera de línea (fase de entrenamiento), estos sistemas carecen de lograr una adaptación a los cambios en el fondo o a cambios en las condiciones de luz mientras están operando [17].

2.4. Segmentación

Dentro de los métodos de segmentación clásicos aparecen tres grandes grupos, en función de la estrategia que se emplee para realizar la segmentación; en primer lugar se aplica la del conocimiento global de toda la imagen o de una región, dado que si se posee información sobre la región (nivel de gris, posición, textura), separar la información de interés del resto de la imagen; seguidamente se encuentran los métodos basados en bordes o fronteras, en los que se buscan las zonas de cambio de características entre dos regiones; y finalmente se encuentran los métodos basados en histogramas [4].

En general los métodos de segmentación se pueden realizar de tres formas; manualmente en donde el usuario realiza la segmentación con la ayuda de una herramienta informática; de forma semi-automática o interactiva, en la que el ordenador realiza el proceso, pero el usuario interviene en determinados momentos para definir parámetros o corregir resultados. Por último, de forma automática, en donde el ordenador realiza el segmentado de la información sin intervención del usuario durante la ejecución de la aplicación.

2.4.1. Métodos basados en histogramas.

La umbralización a partir de histogramas consiste en elegir el nivel de gris que separa los valores correspondientes al objeto y al fondo, el problema está en encontrar el umbral óptimo que realice esta función ya para imágenes que poseen contornos suaves o presentan ruido aleatorio, el histograma no presenta lóbulos (agrupaciones de píxeles cerca a un nivel de gris) claramente definidos [4].

Para localizar el umbral óptimo entonces se aplican técnicas tales como: **clustering**, **entropy**, **metric**, **moments**, **interclass variance** [4] [12].

La umbralización automática basada en histogramas **clustering**, (véase figura 2.13), clasifica los píxeles estadísticamente, sin tener en cuenta su situación espacial; es decir, no se emplea información de regiones o bordes, sólo información de intensidad de cada punto, es por ello que ésta técnica funciona adecuadamente solo para imágenes que contienen pocos objetos y bajas cantidades de ruido [18]. A diferencia de la umbralización manual, el cálculo del umbral k que separa las regiones de la imagen se expresa con la siguiente ecuación:

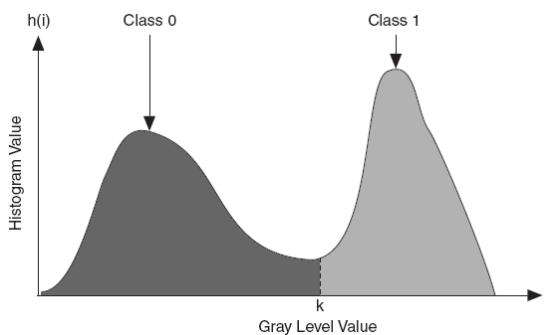


Figura 2.13. Parámetros del histograma.

$$k = \frac{(u1 + u2)}{2} \quad (2.10)$$

Donde $u1$ es el valor máximo de todos los niveles de gris entre 0 y k , e $u2$ es el valor máximo de todos los niveles de gris entre $k + 1$ y 255. Esta técnica se adapta a los cambios suaves de iluminación que puede variar a lo largo de la secuencia de imágenes.

2.4.2. Métodos basados en regiones y contornos.

Para segmentar una imagen por medio de regiones y contornos, se aplica lo que comúnmente se conoce como binarización por medio de umbralización, que consiste en definir un conjunto de valores (umbrales) que determinan la pertenencia de un píxel a un grupo de píxeles que se puedan diferenciar visualmente de otros. Básicamente la técnica de binarización mediante un umbral consiste en convertir una imagen en niveles de gris en una imagen de dos intensidades (blanco y negro), determinando de manera manual o automática un valor de umbral óptimo que separe los objetos del fondo de una imagen [12]. Las dificultades que surgen al elegir el umbral obedecen a factores tales como bajos contrastes, alta relación de la señal al ruido, patrones complejos y modalidades variables en el histograma de niveles de grises. A continuación se describen los tres tipos de segmentación basada en umbralización.

Umbralización fija: La técnica de segmentación basada en un umbral fijo se puede emplear en imágenes donde existe suficiente contraste entre los objetos que se desean separar y el fondo, en donde establecer un umbral que divida las regiones de interés puede realizarse de forma sencilla. Para poder establecer el umbral T adecuado, se debe tener información de la intensidad de cada píxel de los objetos de interés [12] [19].

La fijación del umbral se considera una operación que implica pruebas con respecto a una función T de la forma:

$$T = T\{x, y, A(x, y), f(x, y)\} \quad (2.11)$$

Donde $f(x, y)$ es el nivel de gris del punto (x, y) , y $A(x, y)$ representa alguna propiedad local en el vecindario de este punto. En donde una imagen umbralizada $g(x, y)$ se define como:

$$g(x, y) = \begin{cases} 1 & \text{si } f(x, y) \geq T, \\ 0 & \text{si } f(x, y) < T \end{cases} \quad (2.12)$$

La efectividad de la técnica de segmentación basada en umbral fijo, como se puede observar, depende de la correcta elección de un valor T . Dado que este método ignora las nociones de

proximidad, conectividad, por lo que el uso de este método suele estar limitado a imágenes donde las variaciones de la iluminación y presencia de ruido, están muy controladas [19] [16].

Multiumbralización: Esta técnica de umbralización consiste en la elección de múltiples umbrales. Este método se basa en extraer los mínimos relevantes del histograma y usarlos como umbrales para segmentar la imagen, de forma que en la imagen resultante contenga para cada región una etiqueta diferente [19].

Umbralización dinámica o adaptativa: En la mayoría de las imágenes que se encuentran en condiciones reales ocurre que píxeles del mismo objeto, que en teoría tienen la misma tonalidad, tiene niveles de gris diferentes, esto por que la iluminación no es uniforme o el objeto se conforma por diferentes niveles de gris. Partiendo del principio de la segmentación basada en umbralización fija, la umbralización dinámica permite resolver este problema, haciendo que el valor del umbral varíe en función de las características locales de la imagen [19]. En donde el procedimiento para aplicar una umbralización dinámica consiste en:

- Dividir la imagen original $I(i, j)$ en sub imágenes $I_k(i, j)$ donde los cambios de iluminación no sean tan fuertes.
- Determinar independientemente el umbral T_k para cada sub-imagen.
- Si en alguna sub-imagen es imposible determinar el umbral, éste se calcula mediante la interpolación de los valores de los umbrales de las sub-imágenes vecinas.
- Procesar cada sub-imagen con respecto a su umbral local.

2.5. Seguimiento en imágenes.

La parte más importante en cuanto al desarrollo del sistema de seguimiento de objetos, es el procesado y análisis de la imagen (clasificación de características), ya que a partir de el se colocan en marcha todos los demás procesos del sistema como son: el seguimiento de trayectorias y la interacción con el mundo físico [6].

En la búsqueda de una solución al problema se busca la optimización del programa que ejecute el el seguimiento, dado que este debe cumplir con las especificaciones de tiempo real de la aplicación; para así lograr una implementación rápida y eficaz. A continuación se explicarán algunos de los métodos comúnmente usados en el seguimiento de objetos.

2.5.1. Determinación de movimiento empleando flujo óptico.

El estudio de seguimiento de objetos en secuencias de imágenes, se conoce como **tracking**; el cual se puede desarrollar aplicando el concepto de *flujo óptico*, desarrollado por Gibson

en 1930 que se encarga de describir la dirección y velocidad del movimiento de los puntos en la imagen. Su teoría básica proporcionó el punto de partida para gran parte del trabajo realizado en visión por computadora. En publicaciones posteriores Gibson, habla del gradiente de deformación de la imagen a lo largo del movimiento del observador, concepto que fue bautizado como flujo óptico o simplemente flujo. En la actualidad el flujo óptico se define como el registro de velocidades en el plano de la imagen debidas al movimiento del observador, al movimiento de objetos en la escena ó al movimiento aparente dado por los cambios de intensidad entre cada imagen, en la que se generan falsos movimientos de los objetos o del observador. [20].

Para entender mejor este concepto se muestran a continuación los métodos más comunes por medio de los que se obtiene flujo óptico, necesarios para el estudio del seguimiento de objetos en secuencias de imágenes. Se debe destacar que la mayoría de estos métodos poseen un costo computacionalmente elevado y son inaplicables a algoritmos en tiempo real sin hardware especializado [21].

Métodos basados en características.

El seguimiento de objetos por medio de características, consiste en encontrar en cada una de las imágenes de la secuencia las características estipuladas inicialmente como son los lados, esquinas o algún otro tipo de estructura reconocible de los objetos, conocidos comúnmente como patrones que serán seguidas a lo largo de la secuencia de imágenes [22].

Originalmente se tienen dos etapas dentro de este tipo de tratamiento; la primera consiste en la detección de las características de los objetos en la imagen, mientras que la segunda etapa consiste en seguir a lo largo de cada una de las imágenes la misma característica. El problema radica en que cada etapa maneja sus problemas individualmente, como se muestra a continuación:

- Dentro de la primera etapa, correspondiente a la detección de características; se requiere que el detector sea capaz de determinar las características establecidas de forma rápida y confiable. Dado que estos requisitos satisfacen el correcto funcionamiento de las tareas realizadas por las etapas posteriores del sistema, y más cuando el sistema debe cumplir un requerimiento de tiempo (funcionar en tiempo real).
- Por lo que respecta a la correspondencia entre las características detectadas en cada una de las imágenes, el problema surge al tratar de encontrar la característica equivalente que se presenta entre la imagen actual y la imagen anterior, con el fin de obtener el desplazamiento que esta realizó.

La ventaja que presenta este método es que al tener una menor cantidad de información (características a seguir), el costo computacional en cuanto a tiempo de cálculo se reduce

notablemente; por esta razón para solucionar un problema se deben escoger con anterioridad las características más relevantes de identificación (forma, área, perímetro, color y demás).

Métodos basados en gradiente.

Los métodos basados en el gradiente explotan la relación entre el espacio y las transiciones de valores en el tiempo. Estos métodos utilizan distintas clases de técnicas de estimación diferencial aplicadas al cálculo del movimiento de cada píxel de la imagen, puesto que las relaciones resultantes entre ellos pueden ser usadas para segmentar las imágenes. Un ejemplo, es la segmentación de los puntos en la imagen que presentan la misma velocidad [22]. En este caso la estimación de movimiento en las imágenes está basada en la observación de los cambios de intensidad de brillo (valores de los píxeles en la imagen) con respecto al tiempo, registrándose como velocidades de cambio de las características de brillo de la imagen.

Es importante tener presente que los vectores de movimiento difieren de acuerdo a la perspectiva del movimiento en 3D en el plano de la imagen, por lo que para aplicaciones donde se requieran grandes rangos de precisión no son recomendables estos métodos basados en gradiente. Además la técnica de detección de movimiento por flujo óptico presenta el problema de que las discontinuidades en los campos del flujo óptico son debidas a las discontinuidades de brillo ocasionadas por la presencia de ruido, patrones de reflejos de las superficies del objeto en movimiento y posibles oclusiones o traslapes entre los objetos de movimiento.

El seguimiento de objetos por medio de flujo óptico resulta altamente costoso en lo que a tiempo de cálculo se refiere; por lo que los algoritmos de seguimiento que se basan en este fundamento son por lo general llevados a implementaciones en hardware para que pueda ser aplicados en el seguimiento de objetos en tiempo real.

Métodos basados en correlación.

La correlación es la medida de la relación de igualdad que hay entre dos conjuntos de datos, los valores que toma indican el grado de similitud de ellos, de esta forma 0 indica la ausencia de relación y 1 indica una relación de igualdad completa. La correlación aplicada en el análisis de imágenes, más específicamente en el seguimiento de objetos, consiste en comparar secciones de una imagen con secciones en la imagen siguiente, por esto comúnmente este método es considerado como parte integrante de los métodos anteriormente descritos; por ésta razón algunos autores no lo consideran como un método en si [23], debido a que la correlación no considera ninguna reducción de información previa, también presenta el inconveniente de ser computacionalmente costoso [10] [22].

2.5.2. Determinación del movimiento por segmentación de imágenes.

El método de determinación de movimiento en imágenes utilizando flujo óptico, está limitado en el mundo real, puesto que los objetos con frecuencia se ocultan unos a otros, aparecen y desaparecen objetos en los bordes del campo de visión, la iluminación varía y no está garantizada la suavidad de la iluminación de los objetos [22] [24].

Es por esto que se busca la determinación del movimiento desde otro punto de vista, que consiste en separar la imagen en partes y luego determinar el movimiento de cada uno de estas partes, con algoritmos generalmente denominados como segmentación de imágenes.

A diferencia de los métodos de determinación de movimiento en las imágenes por los gradientes espacio-temporales, los métodos de determinación por segmentación de imágenes son mucho más variados en su desarrollo. Sin embargo, podemos establecer tres métodos básicos:

- Algoritmos de segmentación en líneas: Extraen el contorno de los objetos modelando estos como segmentos de líneas, de tal forma que permitan determinar los parámetros de movimiento en base al desplazamiento de los segmentos entre cada una de las imágenes sucesivas. Este algoritmo extrae en primer lugar los contornos de los objetos contenidos en la imagen; usando operadores por convolución y operadores morfológicos detectores de bordes. Posteriormente determinan las coordenadas de cada uno de los segmentos de líneas que conforman los objetos en las imágenes, para así determinar la correspondencia entre los segmentos de las imágenes consecutivas, con el fin de determinar el movimiento de aquellos segmentos cuya posición ha cambiado [6].
- Algoritmos de segmentación en bloques: Modelan las imágenes dividiéndolas en rectángulos u otros polígonos de forma conocida y de tamaños variables que se adaptan a las áreas del nivel de gris o color encontradas en la imagen. El movimiento se determina en base al desplazamiento de los centros de los rectángulos ocurridos en la sucesión de imágenes. Por lo general este algoritmo divide las imágenes en una serie de bloques, para luego volverlos a dividir en bloques más pequeños que deben cumplir con ciertas restricciones en cuanto a la dispersión de niveles de gris o color dentro de los cuadros. El proceso de división continúa hasta cumplir con dichas restricciones; luego se procede a buscar las correspondencias entre los bloques de dos imágenes consecutivas; por último, se determina el movimiento de aquellos bloques que han cambiado de posición [6].
- Algoritmos de segmentación en grupos (clusters): Modelan las imágenes como áreas amorfas adaptadas a áreas de la imagen de características homogéneas en cuanto a color, nivel de gris o textura, para así determinar el movimiento en base al desplazamiento de los centros de masa de cada una de las áreas correspondientes entre dos imágenes consecutivas. Estos algoritmos de segmentación utilizan procesos de simplificación de las

imágenes en áreas de color o textura uniforme, tales como los operadores morfológicos, de tal forma que las imágenes resultantes estén divididas en áreas uniformes, para así proceder a determinar los centros de cada área, con el fin de hallar las correspondencias entre las áreas de las imágenes consecutivas, y por último, determinar el movimiento entre los centros de áreas correspondientes de dos imágenes consecutivas [6].

En la figura 2.14 se observa el resultado de aplicar cada uno de los tres algoritmos a la misma imagen.

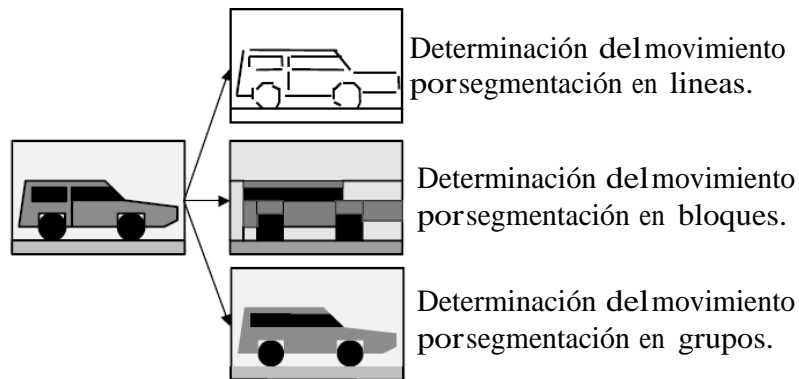


Figura 2.14. Segmentación de imágenes utilizando segmentación en líneas, bloques y grupos.

Todos los algoritmos de segmentación tienen la característica común de generar un campo vectorial de flujo óptico poco denso: un vector por segmento, cuadro o área. Es posible asignar el mismo vector a todos los píxeles dentro de un cuadro o área, pero es una generalización que puede no ser útil en todos los casos. También los algoritmos de segmentación presentan el problema de tener alto costo computacional, puesto que los procesos de buscar las correspondencias entre segmentos, bloques o áreas de una imagen a otra no son sencillos. Además cuando una imagen está compuesta por múltiples objetos, las segmentaciones suelen ser numerosas y por tanto, el cálculo extremadamente laborioso [24].

Seguimiento en base a diferencia.

La idea principal de la cual deriva el nombre, es la de realizar la diferencia (resta) de los valores de cada píxel de la imagen actual, con los valores de cada píxel de la imagen anterior. Como es de suponer esto no representa mayor complicación, ya que si las imágenes son idénticas la diferencia es nula, en caso contrario se presentarán valores diferentes de cero representando existencia de movimiento. Este método es inmune a los cambios de iluminación pero presenta la desventaja de que si el movimiento es muy pequeño se puede tomar como ruido de las imágenes y no ser detectado. El algoritmo se simplifica en la figura 2.15 [6].

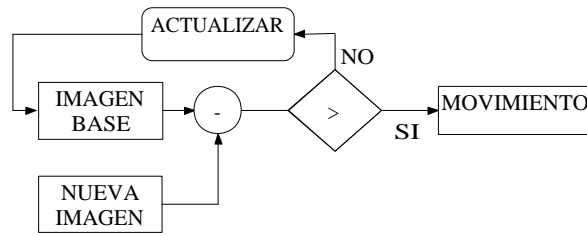


Figura 2.15. Diferencia de imágenes, actualización del fondo.

Ahora se tendrá no una imagen sino una secuencia representada por $f(x,y,t)$. Como se indicó el método para determinar el movimiento será la diferencia entre dos imágenes. Luego la imagen que indicará el campo de velocidades entre los instantes i y j será:

$$V_{i,j} = \begin{cases} 1 & \text{si } f(x,y,i) - f(x,y,j) \geq T \\ 0 & \text{en cualquier otro caso} \end{cases} \quad (2.13)$$

El sentido de aplicar un valor de umbralización es minimizar la influencia del ruido en ambas imágenes; teniendo presente que los movimientos pequeños no son detectados. También deberá aplicarse un filtrado ya que la influencia del ruido provoca píxeles aislados que hay que eliminar. Puede tomarse entonces la condición de realizar una erosión, la cual elimina los píxeles que no se encuentren conectados a sus cuatro u ocho vecinos (eliminar píxeles aislados).

Este método presenta la gran ventaja de la inmunidad a los cambios de iluminación ya que el tiempo transcurrido entre la toma de las dos imágenes suele ser muy pequeño. En el caso de tener una imagen en que la todos los objetos que la conforman fueran estáticos podría tomarse como referencia. Si la diferencia está por encima de un umbral se dice que existe movimiento; si no, el cambio se debe a la iluminación y ese valor sustituye el anterior en la imagen base.

Los sistemas que utilizan este principio son muy utilizados para la detección de movimiento, aplicándose en el mayor de los casos en sistemas de seguridad, pero siendo conjugados con otros tipos de técnicas que ayuden a solventar los problemas que este presenta como son: la susceptibilidad al ruido, las vibraciones y los movimientos de la cámara. Como consecuencia el algoritmo en base a diferencia se complica y depende de otros procedimientos.

Seguimiento en base a centroide.

La determinación de la posición de un objeto a lo largo de la secuencia de imágenes puede estar dada por el centroide de dicho objeto, que es obtenido mediante un análisis de agrupamiento, dependiente de operaciones de segmentación y reconocimiento previas [6].

Para que el seguimiento de objetos en base a centroide sea aceptable deben realizarse por lo general otro tipo de procedimientos previos, como los de umbralización, segmentación y reconocimiento, dado que estos al mismo tiempo contribuyen con el mejoramiento de la definición de los objetos y por ende en su seguimiento, pero aumentando substancialmente su costo computacional.

El método funciona a partir de realizar la segmentación, de tal forma que se pueda calcular el centroide de la región que represente al objeto a seguir durante toda la secuencia de imágenes, mediante la determinación de sus momentos utilizando la siguiente fórmula:

$$\dot{G}_k 0 = \sum_{i=1}^m \sum_{j=1}^n \dot{G}_k(i, j) \quad (2.14)$$

Donde $\dot{G}_k(i, j)$ es el conjunto de valores umbralizados de la región en la imagen, m y n son el número de filas y columnas de la región. Teniendo el resultado de la fórmula anterior es posible determinar las coordenadas del centroide mediante:

$$\begin{aligned} \hat{\delta}_x(k) &= \frac{1}{\dot{G}_k 0} \sum_{i=1}^m \sum_{j=1}^n x_j \dot{G}_k(i, j) \\ \hat{\delta}_y(k) &= \frac{1}{\dot{G}_k 0} \sum_{i=1}^m \sum_{j=1}^n y_i \dot{G}_k(i, j) \end{aligned} \quad (2.15)$$

Seguimiento por compuerta (Gate).

Este método es llamado así por haber sido primero desarrollado en hardware analógico. Su objetivo es procesar picos detectados por escáner donde la compuerta es el nivel del umbral que indica ya sea la ausencia o la presencia del blanco. La compuerta también es conocida como ventana de seguimiento ver figura 2.16, la cual puede variar de acuerdo a los cambios en el tamaño del objeto a seguir [6].

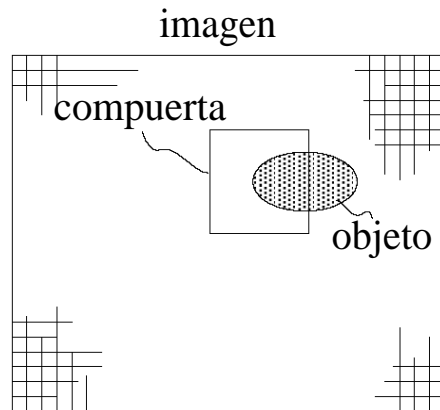


Figura 2.16. Diferencia de imágenes, actualización del fondo.

El algoritmo comienza designando los objetos a seguir, que pueden ser dados por una búsqueda automática, por el usuario o por alguna otra fuente externa que suministre las coordenadas de posición del objeto. Posteriormente a la designación se realiza un barrido de la imagen, que puede iniciar desde la esquina superior derecha, de izquierda a derecha y en orden descendente, o bien iniciar en el centro y en forma circular de adentro hacia afuera de la imagen, o en forma aleatoria.

El seguimiento se realiza con la comparación de parámetros del blanco, en el sentido de maximizar el número de píxeles dentro de la ventana con el mismo tipo de parámetros. Los parámetros a los que hace referencia pueden ser generados a partir de valores tan simples como valores binarios o tan complejos como algún conjunto de características. Así, los parámetros son calculados para cada píxel dentro de las diferentes ventanas y comparados con las estadísticas de la imagen anterior. Teniendo el resultado de los parámetros, la ventana de seguimiento puede ser ajustada para mantener el mayor número de píxeles dentro de ella. Dentro de los inconvenientes de este tipo de procedimiento se encuentra la susceptibilidad al ruido y las oclusiones [6].

Seguimiento en base a correlación.

El seguimiento basado en correlación es la búsqueda de la similitud de una imagen con la imagen siguiente. El seguimiento en base a correlación consiste en estimar los desplazamientos del objeto conforme se mueve dentro del cuadro de la imagen, mediante el cálculo de la correlación entre la imagen de referencia y la imagen actual. Donde la expresión matemática de la correlación para dos conjuntos de valores $\vec{x} = |x_1, x_2, \dots, x_n|$ y $\vec{y} = |y_1, y_2, \dots, y_n|$ se representa de la siguiente forma:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (2.16)$$

Donde x, y son valores pertenecientes a dos vectores con medias aritméticas \bar{x}, \bar{y} respectivamente.

El seguimiento en base a correlación presenta dos importantes ventajas: se comporta aceptablemente en ambientes con ruido y es factible hacer el seguimiento aun cuando el campo de vista o imagen se mueva [6].

3. ARQUITECTURA DEL SISTEMA.

En el presente capítulo se describirá de forma detallada cada uno de los subsistemas que comprenden el diagrama de bloques 3.1, utilizados para el desarrollo del sistema de visión artificial que cumple la función de realizar un registro de transeúntes en tiempo real sobre un ambiente controlado.

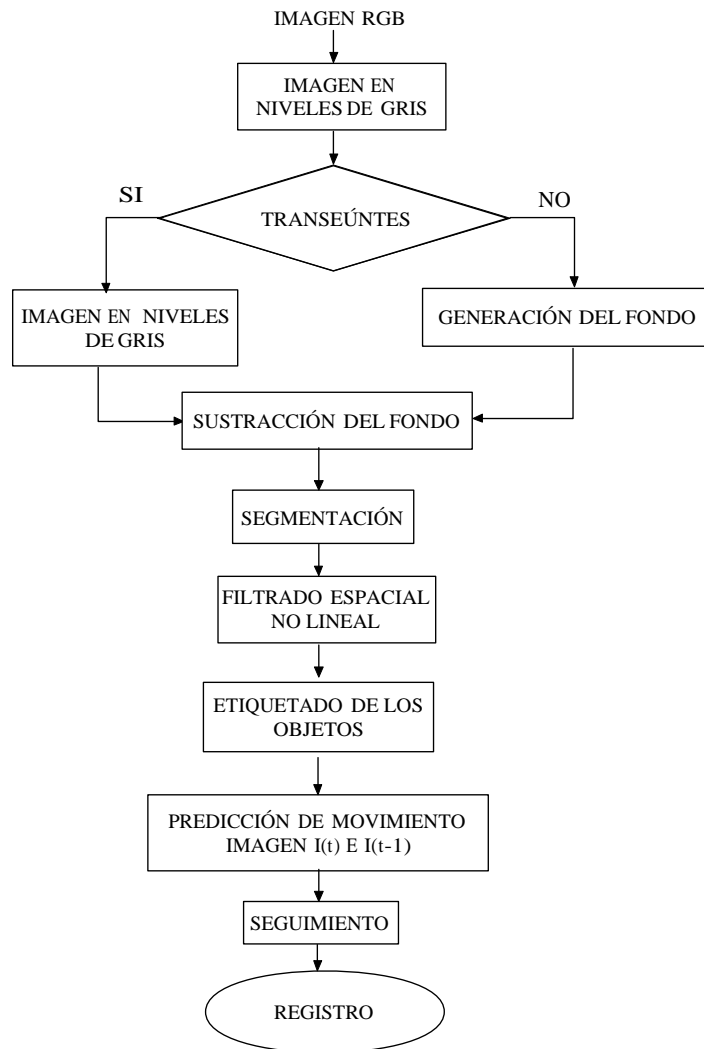


Figura 3.1. Diagrama de bloques.

3.1. Adquisición de la imagen RGB.

Para la adquisición se empleó una cámara **VÍDEO BLASTER WEBCAM PLUS CREATIVE** de 24 bits (16,7 millones de colores) sin calibración, que para este caso captura imágenes de un tamaño de 320 píxeles de largo por 240 píxeles de ancho en formato RGB, a una tasa de muestreo de 30 **frames** por segundo.

Como se enunció en el capítulo 1, el objetivo de imponer restricciones en la escena es el de reducir la complejidad del problema, minimizando en lo posible variables aleatorias que influyan en la respuesta del sistema, con el fin de proporcionar datos rápidos y confiables.

Es importante tener presente que en la adquisición de los datos, la tasa de muestreo ó imágenes por segundo que se deseen adquirir, cumple un factor determinante en los resultados obtenidos; los **frames** por segundo determinan el tiempo disponible para realizar el procesamiento en tiempo real. Cuando se habla de procesamiento en tiempo real, la señal procesada o resultado debe ser entregada antes de que se cumpla un tiempo fijado por la tasa de muestreo, puesto que el algoritmo debe realizar todo el procesamiento para un **frame** antes de que ingrese el siguiente, garantizando que no existan pérdidas de información [25] [26].

La tasa de muestreo depende directamente de la aplicación que se desee desarrollar, dado que no es lo mismo registrar una persona en movimiento que un automóvil. En la figura 3.2 se muestra el intervalo de tiempo disponible entre imágenes para realizar el procesamiento de vídeo, donde r es el intervalo de tiempo disponible, e I es la imagen a procesar.

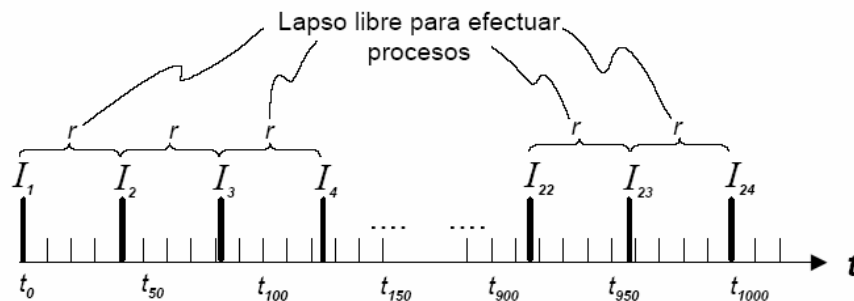


Figura 3.2. Intervalos de tiempo disponible para realizar el procesamiento de las imágenes.

3.2. Preprocesamiento.

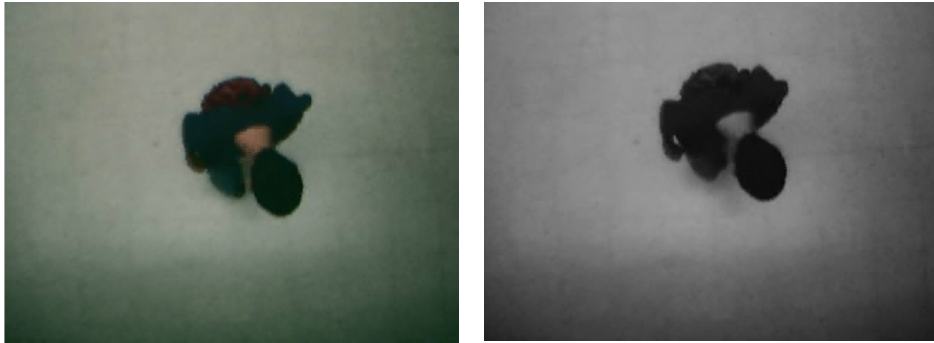
En la etapa de preprocesamiento se busca eliminar el ruido producido en la adquisición de las imágenes, como por ejemplo el producido por los cambios de iluminación y los objetos que no son de interés como objetos inmóviles ú objetos que no sean transeúntes.

3.2.1. RGB a niveles de gris.

Después de realizada la adquisición a cada imagen se le realiza un filtrado que permita obtener la imagen en niveles de gris $[0, 255]$, para reducir el costo computacional, pues se trabaja únicamente con una matriz y no con 3 como en el caso de imágenes en formato RGB. La figura 3.3 (a) muestra la imagen original en RGB, mientras que en (b) se ilustra la misma imagen después de ser convertida a niveles de gris; mediante la utilización del *Escalamiento de la norma Euclidiana*, que se realiza por medio de la siguiente expresión.

$$f(x,y) = \sqrt{\frac{1}{3}(R^2 + G^2 + B^2)} \quad (3.1)$$

Donde R, G y B son los valores de los píxeles en cada uno de los planos de color rojo, verde y azul respectivamente y $f(x,y)$ es su respectivo valor en niveles de gris (ver numeral 2.2.1). En donde el factor de escalamiento $\sqrt{1/3}$ asegura que el rango de los valores de los píxeles en la imagen de salida siempre van a estar limitados por el rango de los píxeles de la imagen de entrada, dando como resultado una representación de la misma imagen pero con valores que se almacenan en una sola matriz (ver sección 2.1.1) [17].



(a) Imagen RGB.

(b) Imagen en niveles de gris.

Figura 3.3. Cambio de imágenes RGB a niveles de gris.

3.3. Detección de movimiento.

En esta etapa se detecta el movimiento producido por los transeúntes dentro del área de conteo por medio de la realización de los siguientes procesos:

3.3.1. Captura del fondo de la imagen.

La captura de la imagen de fondo ó en inglés **background** se realiza de forma supervisada en un instante de tiempo t , en donde los transeúnte que cruzan por el área de conteo son nulos, de tal forma que esta imagen permita al sistema tener una representación de los objetos que en un momento inicial formen parte del fondo (objetos estáticos), ver sección 2.3.1 [17].

3.3.2. Sustracción del fondo.

Para la sustracción del fondo se realiza la diferencia entre la imagen actual I y el fondo I_f ; esta diferencia de imágenes se realiza píxel a píxel, es decir el píxel ubicado en la posición (M, N) de la imagen actual es restado con el píxel de la posición (M, N) de la imagen del fondo. Permitiendo centrar la atención en los objetos de interés y a su vez eliminando los problemas existentes relacionados con cambios de iluminación suaves.

En la operación de restado de las dos imágenes es posible que se obtengan valores negativos, por esta razón a cada valor de píxel de la imagen resultante se le aplica el valor absoluto (ver ecuación 3.2), de tal forma que estos valores queden por encima del nivel del umbral que se escogerá para su posterior segmentación, esto en el caso de personas que presenten un nivel de gris muy inferior al nivel de gris del fondo (ropa oscura).

$$I_r = |I(M, N) - I_f(M, N)| \quad (3.2)$$

3.4. Segmentación.

3.4.1. Binarización.

Para la aplicación correspondiente se realizó la binarización mediante la implementación de una segmentación automática basada en histogramas **clustering**, ver sección 2.4.1. Esta técnica funciona adecuadamente para imágenes que tienen pocos elementos y baja cantidad de ruido ya que agrupa los píxeles según su intensidad y no su situación espacial.

Otra característica de las imágenes a las que se les realiza la segmentación, es que tienen histogramas bimodales, en los que se puede diferenciar claramente dos lóbulos, el de la izquierda que representan el fondo o **background** y el de la derecha que indica el primer plano o foreground, ver figura 3.4.

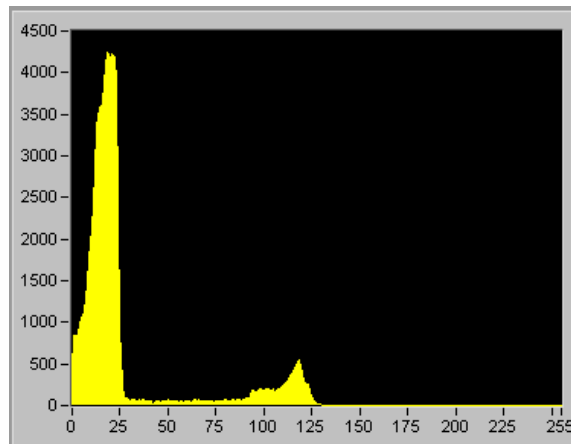
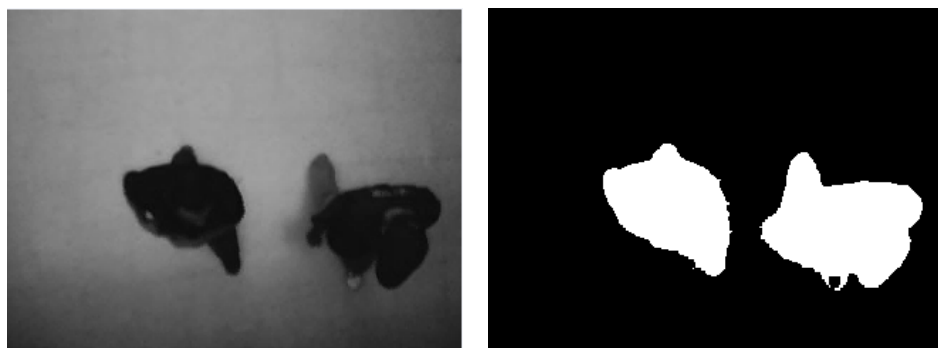


Figura 3.4. Histograma bimodal.

En estos histogramas se tienen dos grupos que pueden ser diferenciados uno de otro por medio de la selección de un umbral óptimo, que agrupe a cada pixel de la imagen dependiendo del nivel de gris que estos tengan [7] y se representa por la ecuación 2.10.



(a) Imagen en niveles de grises.

(b) Imagen binarizada con el umbral escogido.

Figura 3.5. Binarización mediante umbralización dinámica.

3.4.2. Filtrado espacial no lineal.

Estos filtros no se pueden representar en forma de ecuación matricial a diferencia del filtro lineal, dado que éste tipo de filtro carece de un formato general. Cada tipo de filtrado es un algoritmo distinto que depende de las características que se deseen extraer [4].

En la aplicación se emplearon dos filtros de rango; inicialmente se aplicó el filtro de mínimo (dilatación), seguido de un filtro de máximo (erosión), empleando un elemento estructurante

de unos y tamaño 3 x 3 con una conectividad de 8 (ver sección 2.2.1), creando así una operación morfológica conocida como cierre (**closing**) representada por el símbolo (\bullet). En esta operación la dilatación inicial elimina los detalles oscuros e introduce más brillo en la imagen (suaviza el contorno); mientras que la erosión oscurece de nuevo la imagen sin volver a introducir los detalles eliminados por la dilatación y a su vez separa regiones que se encuentran levemente cerca una de otra. La propiedad más sobresaliente de estos filtros es que son idempotentes, es decir, la aplicación consecutiva de una dilatación o erosión no cambia el resultado de la primera aplicación [4]. En la ecuación 3.4 se representa el proceso efectuado por un cierre en donde se observa que se aplica la dilatación ecuación 2.5 y luego al resultado se le aplica la erosión ecuación 2.4 independientemente.

$$A \bullet B = A^B = (A \oplus B) \ominus B \quad (3.3)$$

3.5. Extracción de características.

Como se enunció en el capítulo 1 la función primordial de este subsistema es distinguir y describir las características de los objetos presentes en la imagen; es por esto que para la aplicación diseñada se utilizó un descriptor de regiones, que permite conocer propiedades tales como momentos de inercia, centros de masas, coordenadas del centro de masa, entre otros valores que brindan información detallada de cada uno de los transeúntes que se encuentren dentro del área de conteo.

3.5.1. Etiquetado de los objetos.

Para el desarrollo de la aplicación se buscó etiquetar los transeúntes utilizando un método conocido como BLOB (**B**inary **L**arge **O**bjects), que clasifica los elementos binarios por medio de una agrupación determinada de píxeles conectados, ya sea que posean valores de unos o ceros.

Para esta aplicación el sistema etiqueta a un transeúnte, si este tiene un valor comprendido entre 4000 y 7000 píxeles conectados con valor igual a uno. Se asume este valor para evitar que objetos pequeños o demasiado grandes (ruido) que no sean transeúntes sean registrados. Como se expresa en la siguiente ecuación.

$$g(x,y) = \begin{cases} \text{no etiqueta} & \text{si } A < 4000, \\ \text{etiqueta} & \text{si } 4000 \leq A \leq 7000, \\ \text{no etiqueta} & \text{si } A > 7000 \end{cases} \quad (3.4)$$

Donde A es el número de píxeles conectados con valor igual a uno.

Cumplido el parámetro de conectividad, el sistema procede a colocar las etiquetas a cada uno de los transeúntes que tengan su área total dentro de la escena. A partir de este etiquetado que consiste en colocar un recuadro ó **bounding box** alrededor del transeúnte se procede a calcular el valor de sus momentos empleando la ecuación 2.14 para poder encontrar las coordenadas de los centroides mediante la ecuación 2.15. Dado que este valor se utilizará para realizar la etapa de predicción de movimiento y los posteriores etiquetados. En la figura 3.6 se observa la posición del centroide en un transeúnte encerrado dentro de una etiqueta.

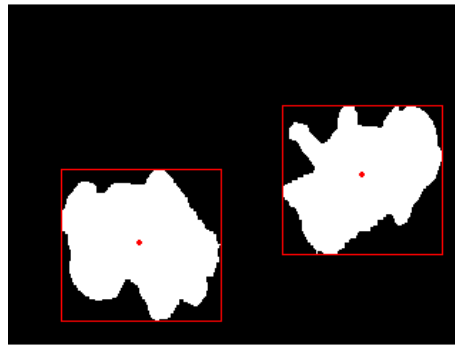


Figura 3.6. Ubicación del centroide y etiqueta de un transeúnte en una imagen binaria.

3.5.2. Seguimiento (tracking).

El seguimiento empleado en este proyecto se basa en la utilización de un método de predicción de movimiento que funciona identificando las coordenadas de los centros de masas de los transeúntes en las imágenes $I(t)$ (imagen actual) e $I(t + 1)$ (imagen anterior), de tal forma que el sistema logre predecir la nueva posición del mismo objeto en la imagen $I(t + 1)$ (imagen siguiente), utilizando la información de la velocidad pasada del objeto en movimiento, de tal forma que se logre identificar como el mismo objeto en cada imagen, aplicando la siguiente ecuación:

$$\hat{v}(t + 1) = (1 - a)v(t) + a\hat{v}(t) \quad (3.5)$$

Donde a es un coeficiente de amortiguamiento igual a 0.3, $v(t)$ y $\hat{v}(t+1)$ son las mediciones del vector de desplazamiento en un tiempo $t+1$ respectivamente.

En la figura 3.7 se describe el procedimiento utilizado por el método de predicción de movimiento. En donde la primera parte consiste en estimar la próxima posición $P(t+1)$ del

objeto actual en un tiempo t , la cual se estima cambiando las coordenadas del centro de masa actual $C(t)$ por $\hat{v}(t+1)$. De tal forma que el sistema defina un área circular de búsqueda cuyo centro se encuentra ubicado $P(t+1)$ y con un radio predefinido, de tal forma que la búsqueda de las coordenadas de la imagen I+1 incidan dentro del área circular de búsqueda; donde $C(t+1)$ representa las coordenadas del centro de masa del objeto detectado. El proceso es iterativo, dado que el sistema relaciona a cada transeúnte como el mismo cada vez que detecta las coordenadas dentro de la circunferencia trazada con anterioridad de radio 60 píxeles. De tal forma que el proceso se aplica independientemente para cada uno de los transeúntes que circulan por el área de conteo [27].

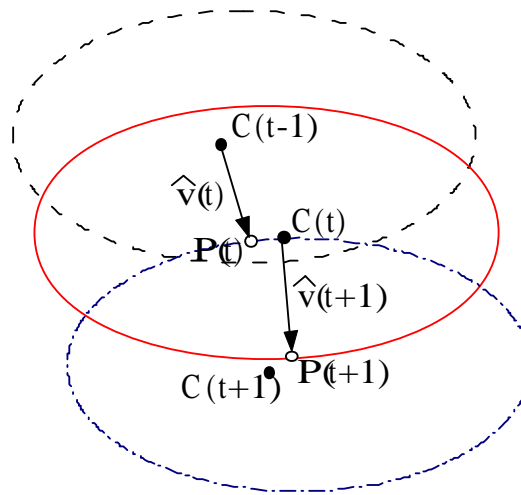


Figura 3.7. Seguimiento de un objeto usando un método de predicción de movimiento.

3.5.3. Registro.

El sistema entrega al usuario en tiempo real la información sobre el número de transeúntes que han circulado por el área de conteo, discriminándolos por la dirección de recorrido (de arriba hacia abajo y viceversa). Además realiza un registro global del número de transeúntes que cruzaron el área de conteo.

4. IMPLEMENTACIÓN DEL SISTEMA Y ANÁLISIS DE RESULTADOS.

La implementación del sistema se realizó en un computador **Pentium IV** con procesador de 2.8 GHz y 512 MB de RAM corriendo sobre **Windows XP**. El sistema de desarrollo de software utilizado es LabVIEW versión 8.0 de **National Instruments**.

Para la aplicación desarrollada se requieren tener instalados los módulos de **Vision and Motion** que se mencionan a continuación.

- **NI-IMAQ for USB Cameras:** Permite trabajar con cámaras web USB que sean compatibles con el **NI Assistant**, **LabVIEW** y NI-Visión de IMAQ. Las cámaras pueden operar a diferentes resoluciones y velocidades de muestreo, dependiendo de sus capacidades [26]. Este módulo se requiere cuando se desea trabajar en línea, es decir cuando se adquiere el vídeo directamente desde la cámara web
- **NI-VISION 8.0 Development Module:** Proporciona las herramientas necesarias para realizar el procesamiento de las imágenes a estudiar [7].

La implementación del software se realizó a partir de una máquina de estados en el sistema de desarrollo LabVIEW 8.0. Este es un lenguaje de programación de alto nivel que expresa algoritmos de forma adecuada a la capacidad cognitiva humana; que además proporciona una interfaz gráfica para dibujar diagramas de estado en su forma natural, es decir el software está en capacidad de generar automáticamente el código de flujo de datos correspondiente.

Para la adquisición de los vídeos la **webcam** se ubicó en una posición estática a 3.5 metros sobre el nivel del piso y sin inclinación alguna, en una posición acimutal como se observa en la figura 4.1. Dicha ubicación permite obtener un área de conteo de $9m^2$ de dimensiones 3 m de largo por 3 m de ancho, en la que un máximo de tres transeúntes pueden aparecer a la misma vez por la parte superior o inferior del área de conteo. El fondo de esta área es clara; esta característica proporciona un buen contraste entre el fondo y los transeúntes. Los cambios del nivel de iluminación para este caso son suaves en toda el área de conteo; a su vez se asume que por la escena los únicos objetos móviles son transeúntes.

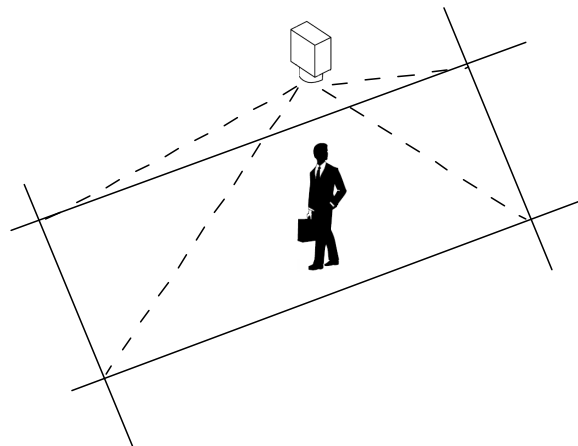


Figura 4.1. Ubicación de la fuente de adquisición.

El sistema tiene la funcionalidad de trabajar en línea (adquiriendo con la webcam) o fuera de línea (utilizando una base de datos). Además, cuando se trabaja en línea el sistema permite guardar el vídeo; de tal forma que se puede crear una base de datos para probar el algoritmo fuera de línea. A continuación se muestra el panel frontal del sistema, que entrega en donde se observa que el usuario solo tiene la posibilidad de actualizar el fondo. Véase figura 4.2.



Figura 4.2. Panel frontal.

4.1. Resultados experimentales.

La secuencia de vídeo utilizada para probar el funcionamiento del sistema de visión artificial tiene una duración de 4:00 minutos; adquirida a una resolución de 320 x 240 píxeles, a 30 **frames** por segundo, en formato .avi y trabajando con imágenes sin compresión.

Para encontrar la velocidad de procesamiento del sistema, el algoritmo se probó para tasas de muestro de 30, 15 y 7 **frames** por segundo, con diferente número de transeúntes que circulaban por el área de conteo, obteniendo los siguientes resultados.

Número de transeúntes	Tiempo \approx (ms)
1	29
2	30
3	33

Tabla 4.1. Tiempos de procesamiento empleados por el sistema

Debido a las particularidades que dependen de la aplicación, del ambiente de trabajo y de la adquisición del vídeo, no es posible utilizar bases de datos existentes para realizar un análisis comparativo de resultados. En el momento de cuantificar los resultados obtenidos de la aplicación se opta por buscar casos especiales dentro de la bases de datos que permitan definir cuales son las posibles causas de fallo en la detección de los objetos de interés. Para la validación de la metodología propuesta se utiliza la matriz de confusión que definen las diferentes respuestas que entrega el sistema [28].

- **Fp:** Grupo de píxeles conectados que son detectados como movimiento, pero que no corresponden a ningún transeúnte (falsa alarma).
- **Fn:** Grupo de píxeles conectados que son detectados como movimiento, pero no son identificados por el sistema como transeúntes.
- **Vp:** Grupo de píxeles conectados que son detectados como movimiento, que son identificados como transeúntes por el sistema.
- **Vn:** Grupo de píxeles no conectados, que no son detectados por el sistema como transeúntes.

Para tasas de muestro de 30 y 15 **frames** por segundo se obtuvieron los siguientes resultados, discriminando la dirección de cruce de los transeúntes en el área de conteo de arriba hacia

abajo y viceversa.

	Vp	Fp	Fn	Real
Suben	39	1	0	39
Bajan	40	0	2	42

Tabla 4.2. Matriz de confusión para 30 y 15 **frames** por segundo.

El sistema no registra **Vn** dado las restricciones impuestas en la escena, dado que se asume que todos los objetos que atraviesan el área de conteo son transeúntes. Además con una velocidad de muestreo igual ó menor a 7 **frames** por segundo el sistema registro pérdidas de información debidas a transeúntes que pasaban a altas velocidades (corriendo o saltando).

A continuación se ilustran diversos casos en donde el sistema funciona de forma correcta, mostrando a su vez las condiciones en que el sistema falla.

Algunos **Vp** presentados por el sistema son los siguientes:

En las figura 4.3 se observa que el sistema registra los transeúntes que circulan en ambos sentidos registrando primero el transeúnte que pasa de abajo hacia arriba y luego el que va en sentido contrario. Mientras que en la figura 4.4 el sistema registra los dos transeúntes cuando se encuentran cruzando el área de conteo en un mismo instante.

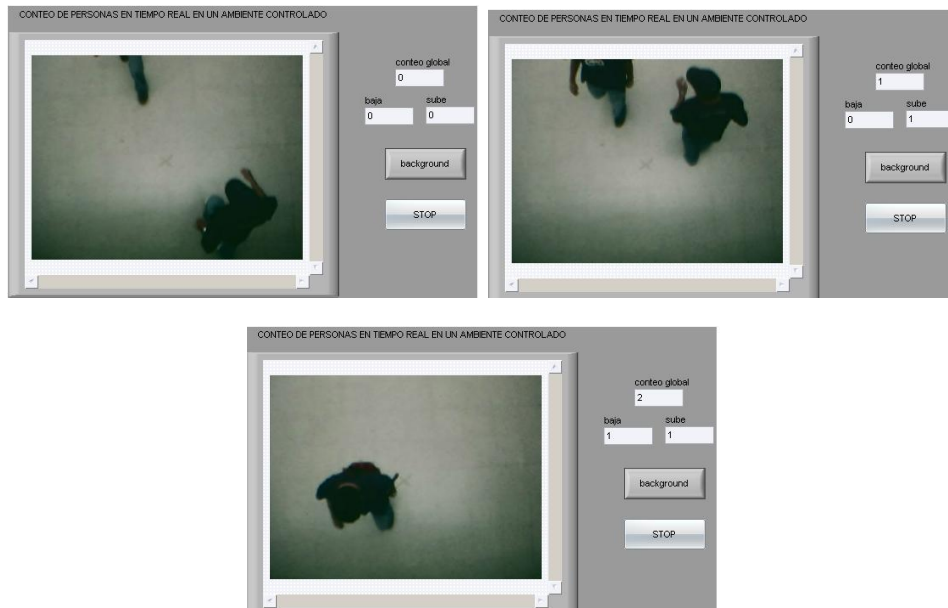


Figura 4.3. Registro de persona que sube y que baja.



Figura 4.4. Registro en ambos sentidos en un mismo instante de tiempo.

En el siguiente caso se ilustra la cantidad máxima de personas que pueden estar en el área de conteo, en donde se ilustra que el sistema registra al transeúnte que pasa de abajo, hacia arriba, sin importar los dos transeúntes que se encuentran estáticos, los cuales pasados un instante t son registrados.

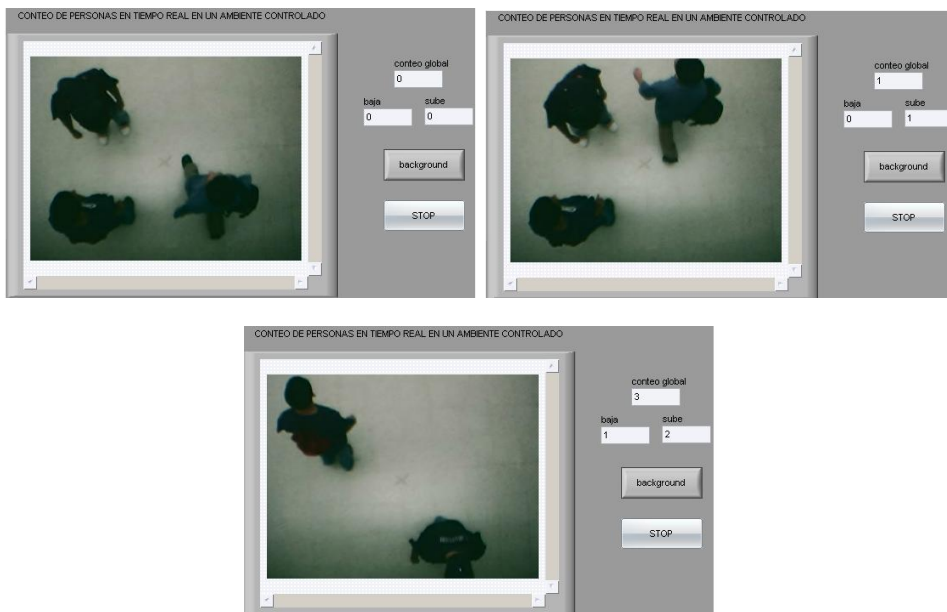


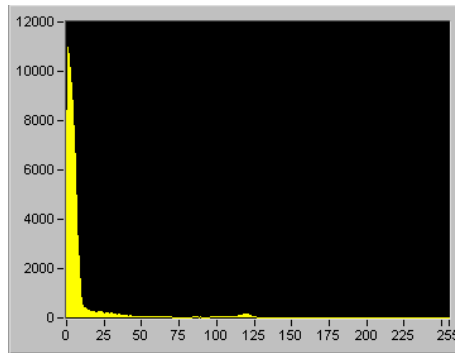
Figura 4.5. Registro de varias personas.

El **Fp** obtenido en los resultados se debe al bajo contraste entre la ropa del transeúnte y el fondo ya que el nivel de gris es menor que el umbral escogido para realizar la segmentación; por tanto el sistema parte al transeúnte en dos bloques de píxeles, resultando así 2 etiquetas, ver figura 4.6, en donde (a) ilustra la imagen capturada en RGB, (b) el etiquetado y (c) muestra el histograma en donde se observa el bajo contrastes que existe entre el **background** y el **foreground**.



(a) Imagen capturada en RGB.

(b) Etiquetado realizado por el sistema.



(c) Histograma.

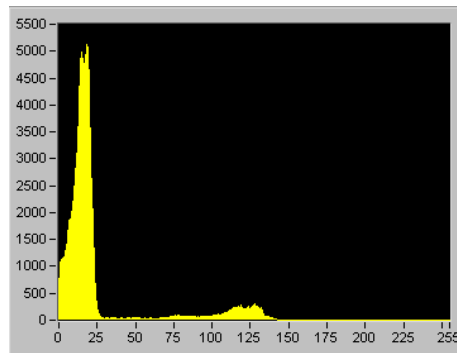
Figura 4.6. Fp presentado por el sistema.

Los **F_n** se deben al traslape entre dos personas, ya que una agrupación mayor a 10.000 píxeles conectados no es etiquetada como transeúnte. Como se especificó en los objetivos del proyecto, el sistema no está en capacidad de solucionar esta situación. Véase figura 4.7, en donde (a) ilustra la imagen capturada en RGB, (b) el etiquetado y (c) el histograma en donde se observa que el objeto presenta un umbral diferenciable pero no es etiquetado debido a su gran agrupación de píxeles.



(a) Imagen capturada en RGB.

(b) Resultado entregado por el sistema.



(c) Histograma.

Figura 4.7. Fn presentado debido a un traslapo entre transeúntes.

5. CONCLUSIONES

Se propone una metodología basada en un sistema de visión artificial para la detección de transeúntes; que fue probada a partir de una base de datos obtenida en el ambiente controlado descrito en el capítulo 3. El sistema propuesto está en capacidad de trabajar realizando la adquisición de vídeo directamente desde una cámara web sin requerimientos de calibración o procedimientos de puesta a punto complicados. Su diseño se basa en el estudio de las características del problema, inicialmente las condiciones del ambiente de estudio, permitiendo encontrar las técnicas apropiadas que en conjunto dieron solución eficiente al trabajo planteado.

La aplicación se caracteriza por tener un funcionamiento autónomo, libre de parámetros de ajuste, es decir el usuario no manipula en ningún momento variables durante la ejecución de la aplicación sino que se establecen automáticamente. Además se tiene un bajo costo en la etapa de adquisición, porque se demostró que por medio de una cámara de baja calidad (webcam) el sistema operaba en forma satisfactoria, y el procesamiento de imágenes se realiza en tiempo real, permitiendo un análisis de resultados sin pérdidas de información.

Utilizar una umbralización automática a partir de histogramas en la etapa de segmentación permite que la elección del umbral se adapte a los cambios de contraste de la escena y sea robusto a cambios de iluminación y además objetos que presenten niveles de gris diferentes sean segmentados de forma óptima; lo que no se hubiera logrado con una umbralización estática que fija un valor para todas las imágenes que conforman la secuencia sin tener en cuenta las variaciones de contraste.

La técnica de sustracción de fondo permitió mejorar la detección de movimiento en comparación con técnicas que utilizan únicamente las imágenes actual y anterior (Choi [22]). Pues con la última existen pérdidas de información, mientras que la técnica utilizada en la aplicación permite segmentar objetos de interés estacionarios en la escena. La tasa de muestreo de la adquisición de vídeo (15 **frames** por segundo) permitió tener información suficiente para registrar los transeúntes que pasan por el área de conteo sin importar la velocidad con que la atraviesen; igualando así las tasas de muestro empleada en otras investigaciones Siebel [17] y Seemann [19].

Se implementó una metodología para el seguimiento de los objetivos reconocidos en la secuencia de vídeo (Tracking), por medio de la predicción de movimiento de sus centroides dado que los transeúntes no presentaban un patrón igual entre **frame** y **frame**, que

permitieran reconocerlos, como área, perímetro, forma, y demás; en cada una de las imágenes el mismo objeto a una etiqueta. El seguimiento y la detección permiten tener un registro del número de personas que circulaban en el área de conteo discriminado por la dirección de cruce (arriba hacia abajo y viceversa).

El sistema presentó errores debido a los traslapes presentados en las imágenes, como se definió en el anteproyecto, este no era un problema que se buscaba resolver. Otro caso especial en donde se presentan errores es cuando no existe una diferencia considerable en el contraste entre el fondo y el transeúnte, lo cual se puede solucionar implementando una técnica que permita agrupar píxeles que no se encuentren conectados, pero que pertenezcan a un mismo transeúnte (agrupar píxeles cercanos no conectados).

6. DISCUSIONES Y TRABAJOS FUTUROS.

En el caso de buscar la solución a las pérdidas de información debidas a traslapes se deberían utilizar técnicas, que en primer lugar solucionen este problema por medio de segmentación más robustas que logre definir correctamente los contornos de los objetos traslapados, como son segmentación jerárquica en color, modelamiento de transeúntes, visión estéreo (multicámaras); y en segundo lugar técnicas que mejoren el seguimiento de los transeúntes como son, filtrado de Kalman (que se aplica para la predicción de movimiento), vectores de correspondencia, **Tracking** en base a colores, modelos de seguimiento **interframe**, es decir, trabajar con imágenes de distintos instantes de tiempo y relacionarlas entre si.

Si se desea implementar el algoritmo en ambientes exterior, se debe buscar un modelamiento de fondo que sea robusto a objetos que pertenezcan al fondo pero estén en movimiento, a los cambios fuertes de iluminación, sombras debidas a objetos que no son de interés árboles, automóviles y demás eventualidades propias de escenarios fuera de cualquier recinto cerrado. Si otros objetos que no sean transeúntes entran en la escena, el sistema realizaría el seguimiento como si este lo fuera. Esta dificultad puede ser manejada realizando una extracción de características que permita una clasificación basada en la silueta, posición, dirección de movimiento del objeto que se ha empezado a seguir.

La metodología podría ser utilizada en aplicaciones como son, la detección de movimiento en sitios prohibidos, recolección de información para realizar **marketing** (zonas de tránsito más frecuentadas), cuantificar la densidad peatonal, medir distancias entre objetos en movimiento, detección de comportamientos sospechosos, medición de la dirección dominante a determinadas horas del día en un sitio determinado (estaciones de transportes masivos), detección de objetos abandonados en lugares tales como, la sala de espera de una estación o un aeropuerto, estadios, colegios entre otros, reconocimiento de patrones de actividad de individuos o grupos.

Anexos

A. REFERENCIAS DE LAS FIGURAS

A continuación se muestra la tabla con las referencias bibliográficas de las paginas web y libros de los que se adquirieron las figuras incluidas dentro de la tesis.

REFERENCIAS BIBLIOGRÁFICAS	
Número de la figura	Referencia
1.1 2.15	R. L. López, <i>Un sistema de visión para el registro automático de vehículos basado en la lectura de números vin</i> , Mayo 2002.
1.2 1.3 1.4 2.1 2.5 2.6 2.8	<i>Image Processing toolbox of Matlab</i> , disponible en www.mathworks.com
2.2 2.3 2.4	G. O. Hernández, <i>El significado de los colores - Capitulo 4</i> . México: Editorial Trillas, 2001
2.7	M. Petrou and P. Bosdogianni, <i>Image Processing The Fundamentals</i> . England: John Wiley and Sons, 1999. disponible en www.wiley.com
2.10	O. I. H. Martinez, <i>Diseño de un algoritmo para la detección y extracción de cuerpos en movimiento en secuencias de imágenes</i> , Septiembre 2005.
2.12	National Instruments, <i>Imaq-vision user manual</i> , 2005. disponible en www.ni.com
2.13	D. C. P. Albores, <i>Seguimiento de objetos por medio de visión activa</i> , Ph.D. dissertation, Instituto Nacional de Astrofísica, Óptica y Electrónica, Mayo 2002.
3.3	J. A. F. Muñoz, <i>Estudio comparativo de las técnicas de procesamiento digital de imágenes</i> , Julio 1999.
3.8	B. D. C. Jae Won Kim, Kang Sun Choi and S. J. Ko, <i>Real time vision based people counting system for the security door</i> , Korea 2002.

Tabla A.1. Referencias bibliográficas de figuras.

BIBLIOGRAFÍA

- [1] D. Mark and W. H. Freeman, *Vision*. Estados Unidos: McGraw-Hill, 1982.
- [2] B. G. Batchelor and P. F. Whelan, “Intelligent vision system for industry,” 1997.
- [3] R. L. López, “Un sistema de visión para el registro automático de vehículos basado en la lectura de números vin,” Mayo 2002.
- [4] G. Pajares and J. D. la Cruz, *Visión por computador Imágenes Digitales y Aplicaciones*. España: Rama, 2001.
- [5] G. J. Awcock and R. Thomas, *Applied Image Processing*. Estados Unidos: McGraw-Hill, 1995.
- [6] D. C. P. Albores, “Seguimiento de objetos por medio de visión activa,” Ph.D. dissertation, Instituto Nacional de Astrofísica, Óptica y Electrónica, Mayo 2002.
- [7] N. Instruments, “Imaq-vision user manual,” 2005.
- [8] M. Petrou and P. Bosdogianni, *Image Processing The Fundamentals*. England: John Wiley and Sons, 1999.
- [9] G. O. Hernández, *El significado de los colores - Capítulo 4*. México: Editorial Trillas, 2001.
- [10] J. Bigun, *Vision with Direction*. Sweden: Springer-Verlang Berlin, 2005. [Online]. Available: www.hh.se/staff/josef
- [11] A. Bovik, *Handbook of Image and Video Processing*. Austin, Texas: University of Texas, 2000.
- [12] D. A. Álvarez and M. L. Guevara, “Detección de microcalcificaciones en mamografías digitales,” Marzo 2006.
- [13] A. J. M. Penín and M. J. L. Touriño, “Identificación y catalogación de imágenes de interfaz,” Ph.D. dissertation, Universidad De Vigo, Septiembre 2003.
- [14] M. Faundez, “Tratamiento digital de voz e imagen y aplicación a la multimedia,” *Alfaomega Grupo Editor*, vol. 1ra edición, pp. 120–121, Mexico D.F. 2001.
- [15] O. I. H. Martinez, “Diseño de un algoritmo para la detección y extracción de cuerpos en movimiento en secuencias de imágenes,” Septiembre 2005.

- [16] A. R. M. Islas, “Estimación de fondo y primer plano en secuencias de imágenes para la detección de objetos en movimiento,” Febrero 2003.
- [17] N. Siebel, “Design and implementation of people tracking algorithms for visual surveillance applications,” Marzo 2003.
- [18] T. Seemann, “Digital image processing using local segmentation,” Ph.D. dissertation, School of Computer Science and Software Engineering Faculty of Information Technology Monash University, Australia Abril del 2002.
- [19] L. E. J. Copoya, “Algoritmo de segmentación local dinámica aplicado a la caracterización de partículas,” Febrero 2003.
- [20] R. Jain, R. Kasturi, and B. G. Schuurack, *Machine Vision*. USA: McGraw Hill, 1995.
- [21] J. M. Sotoca, “Estimación de movimiento,” Junio 2002.
- [22] A. Z. Izaguirre, “Visión artificial dinámica determinación de movimiento a partir de secuencias de imágenes,” Ph.D. dissertation, Universidad Del Pais Vasco, Septiembre 1998.
- [23] S. M. Smith, “Finding optic flow.” [Online]. Available: <http://www.fmrib.ac.uk/steve/review/review/node1.html>
- [24] J. A. F. Muñoz, “Estudio comparativo de las técnicas de procesamiento digital de imágenes,” Julio 1999.
- [25] G. Ling and K. S. Yuan, “Multimedia image and video processing,” Junio 2001.
- [26] N. Instruments, “Ni-imaq for usb cameras,” 2005.
- [27] B. D. C. Jae Won Kim, Kang Sun Choi and S. J. Ko, “Real time vision based people counting system for the security door,” Korea 2002.
- [28] M. P. Rita Cucchiara, Costantino Grana and A. Prati, “Detecting moving objects, ghosts, and shadows in video streams iee transactions on pattern analysis and machine intelligence, vol.25, no.10,” Octubre 2003.