

DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA
DE MONITOREO DE PERFILES DE OBJETOS
UTILIZANDO VISIÓN ARTIFICIAL

CARLOS ANDRÉS MARÍN LOSADA
2002200853
JHON JAIRO TRUJILLO SÁNCHEZ
2002101253

UNIVERSIDAD SURCOLOMBIANA
FACULTAD DE INGENIERÍA
INGENIERÍA ELECTRÓNICA
NEIVA
2010

DISEÑO E IMPLEMENTACIÓN DE UN SISTEMA
DE MONITOREO DE PERFILES DE OBJETOS
UTILIZANDO VISIÓN ARTIFICIAL

CARLOS ANDRÉS MARÍN LOSADA
2002200853
JHON JAIRO TRUJILLO SÁNCHEZ
2002101253

Trabajo de grado presentado para optar
al título de Ingeniero Electrónico

Director
M.Sc. JOSÉ DE JESÚS SALGADO PATRÓN
Ingeniero Electrónico

UNIVERSIDAD SURCOLOMBIANA
FACULTAD DE INGENIERÍA
INGENIERÍA ELECTRÓNICA
NEIVA
2010

CONTENIDO

	pág.
RESUMEN	13
ABSTRACT	14
INTRODUCCIÓN	15
1. VISIÓN ARTIFICIAL	17
2. IMAGEN DIGITAL	18
2.1 IMÁGENES A COLOR	18
2.2 FORMATOS GRÁFICOS DIGITALES	19
2.3 FORMATOS NUMÉRICOS	20
2.4 PROCESAMIENTO DIGITAL DE IMÁGENES	20
3. CONFIGURACIÓN DE UN SISTEMA DE CAPTURA	21
3.1 PARÁMETROS FUNDAMENTALES DE UN SISTEMA DE CAPTURA	21
3.1.1 Resolución	21
3.1.2 Campo de visión	21
3.1.3 Distancia de trabajo	21
3.1.4 Profundidad de campo	21
3.1.5 Contraste	21
3.2 ILUMINACIÓN	22

3.3	CONSIDERACIONES DE MOVIMIENTO	23
4.	GENERALIDADES DEL ANÁLISIS DE IMÁGENES DINÁMICAS	24
4.1	ETAPAS DE UN SISTEMA DE VISIÓN COMPUTACIONAL	24
4.1.1	Adquisición del <i>frame</i> de vídeo	26
4.1.2	Pre-procesado de video	26
4.1.2.1	Ajustes geométricos	26
4.1.2.2	Ajustes radiométricos o de intensidad	27
4.1.2.3	Derivada de la imagen	27
4.1.2.4	Profundidad de la escena	27
4.1.3	Modelado de fondo	27
4.1.3.1	Modelo de representación del fondo	29
4.1.3.2	Modelo de inicialización de fondo	32
4.1.3.3	Modelo de adaptación del fondo	32
4.1.4	Detección de objetos de primer plano	33
4.1.4.1	Umbral de clasificación	33
4.1.5	Técnicas de post-procesado	34
4.1.6	Representación y descripción	35
4.1.7	Reconocimiento e interpretación	35
5.	DESCRIPCIÓN GENERAL DEL SISTEMA	36
5.1	PROCESO DE ADQUISICIÓN DE VIDEO	36
5.2	ESTIMACIÓN Y SUSTRACCIÓN DE FONDO	36
5.3	SEGMENTACIÓN DE OBJETOS EN MOVIMIENTO	37

5.4	CLASIFICACIÓN AUTOMÁTICA DE OBJETOS	37
5.5	SEGUIMIENTO DE OBJETOS EN MOVIMIENTO	38
5.6	REGISTRO Y CONTEO DE OBJETOS	38
5.7	ESPECIFICACIÓN DE REQUISITOS	38
5.7.1	Especificaciones funcionales	38
5.7.2	Especificaciones no funcionales	39
6.	DESARROLLO FUNCIONAL DEL SISTEMA	40
6.1	CONSIDERACIONES INICIALES DEL SISTEMA	40
6.1.1	Consideraciones en la adquisición de video	40
6.1.2	Consideraciones del entorno	41
6.1.3	Consideración de movimiento de los peatones	41
6.2	PROCESO DE ADQUISICIÓN DE VIDEO DIGITAL	41
6.3	ESTIMACIÓN Y SUSTRACCIÓN DE FONDO	44
6.3.1	Diferenciación de <i>frames</i> adyacentes	45
6.3.2	Método de múltiples gaussianas	46
6.3.3	Filtro de mediana aproximado	47
6.3.4	Filtro de mediana aproximado selectivo	48
6.3.5	Normalización de la luminancia	49
6.4	SEGMENTACIÓN DE OBJETOS EN MOVIMIENTO	50
6.5	CLASIFICACIÓN AUTOMÁTICA DE OBJETOS	52
6.6	SEGUIMIENTO DE OBJETOS EN MOVIMIENTO	53
6.6.1	Método de la distancia euclidiana	54

6.6.2	Método del filtro de Kalman	56
6.7	REGISTRO Y CONTEO DE OBJETOS	57
7.	PRUEBAS Y ANÁLISIS DE RESULTADOS	60
7.1	ESTIMACIÓN Y SUSTRACCIÓN DE FONDO	60
7.1.1	Método de diferencia de <i>frame</i> adyacente	60
7.1.2	Método de múltiples gaussianas	62
7.1.3	Filtro de mediana aproximado	64
7.1.4	Filtro de mediana aproximado selectivo	66
7.1.5	Normalización de la luminancia para el filtro de mediana aproximado selectivo	68
7.2	SEGUIMIENTO DE OBJETOS EN MOVIMIENTO	68
7.2.1	Método de la distancia euclidiana	68
7.2.2	Método del filtro de Kalman	69
7.3	ANÁLISIS DE RENDIMIENTO Y CONFIABILIDAD DEL SISTEMA	71
8.	CONCLUSIONES Y RECOMENDACIONES	75
	BIBLIOGRAFÍA	78
	ANEXOS	80

Nota de Aceptación

Firma del Presidente del Jurado

Firma del Jurado

Firma del Jurado

Neiva, 12 de Octubre de 2010.

AGRADECIMIENTOS

Siempre existirán personas cuyas ideas y pensamientos coadyuvaron al desarrollo y finalización de este trabajo de grado, a todo ellos agradecemos su apoyo y conocimiento aportado.

Expresamos nuestros más sinceros agradecimientos a:

Dios, el pensamiento que siempre nos acompaña en todos los momentos de nuestra existencia.

El Ingeniero José de Jesús Salgado Patrón director de este proyecto, por sus oportunas correcciones y grandes aportes.

Los Ingenieros Javier Humberto Rubio Plazas y Vladimir Mosquera Cerquera quienes con sus aportes académicos despejaron las inquietudes que surgieron en el desarrollo del proyecto.

El Ingeniero William Cabrera Castro por la oportuna colaboración en la revisión del documento.

La Universidad Surcolombiana por permitirnos realizar el trabajo de campo en sus instalaciones.

Nuestros familiares por su apoyo incondicional y paciencia.

*A mis padres, Edilberto y Alba Deny por el apoyo, la solidaridad, la confianza y el cariño
que siempre me han brindado.*

*A mis hermanos Nelson, Edilzuri, Andrea y Alejandra quienes me colaboraron y me
brindaron sus consejos cuando sentí nublar mis pensamientos
ante las adversidades de la vida.*

*A mi novia, Luz Dary quien con su compañía, hizo que mis días
fueran más divertidos y enriquecedores.*

Carlos Andrés Marín Losada

A Dios por la esperanza que me mueve y el amor que me da felicidad.

A mi madre, Gilma Sánchez por su amor, comprensión, apoyo incondicional y paciencia.

*A mi padre, que siendo su deseo con firmeza y constancia, hizo que me formara siguiendo los
mejores principios.*

*A mis hermanos, porque en ellos encontré los consejos, que fueron el aliento en cada momento que
sentí desfallecer.*

A mis amigos, compañeros de sueños.

Jhon Jairo Trujillo Sánchez

GLOSARIO

BACKGROUND: son las regiones que conforman el fondo de una imagen, en procesamiento de video se refiere a la parte estática de una imagen.

BITMAP: es un formato gráfico digital empleado por los sistemas operativos Windows e IBM para representar imágenes.

FOREGROUND: son las regiones u objetos que componen el primer plano o frente de una imagen, en procesamiento digital de video se refiere a los objetos en movimiento.

FRAME: fotograma o cuadro de imagen que compone una secuencia de video.

MoG: mezcla de múltiples gaussianas.

SG: simple gaussiana.

TOOLBOX: es el conjunto de herramientas, métodos y técnicas de programación, empleadas para desarrollar una tarea específica.

LISTA DE FIGURAS

	pág.
Figura 1. Imagen en modo RGB	19
Figura 2. Parámetros fundamentales de un sistema de captura	22
Figura 3. Diagrama de bloques de un sistema de visión artificial	24
Figura 4. Etapas de un algoritmo de sustracción de fondo fijo	25
Figura 5. (a) Inicialización de fondo sin objetos en movimiento, (b) <i>Frame</i> actual, (c) Adaptación de fondo y (d) Detección correcta de un objeto en movimiento	28
Figura 6. (a) Inicialización de fondo con objetos en movimiento, (b) <i>Frame</i> actual, (c) Adaptación de fondo con inicialización incorrecta y (d) Detección incorrecta de objetos en movimiento	29
Figura 7. Máscara de discriminación de frente	34
Figura 8. Esquema de ubicación de la cámara de video	41
Figura 9. Fases de la adquisición de datos de video	44
Figura 10. Diagrama de flujo del método de <i>frames</i> adyacentes	46
Figura 11. Diagrama de flujo del algoritmo de filtro de mediana aproximado	48
Figura 12. Diagrama de flujo del algoritmo de filtro de mediana selectivo	49
Figura 13. Diagrama de Flujo del bloque de normalización de luminancia	50
Figura 14. Diagrama de flujo para seguimiento de objetos	55
Figura 15. Diagrama de flujo para el Filtro de Kalman	57
Figura 16. Registro y conteo de objetos	58
Figura 17. Diagrama de flujo de la etapa de conteo	59
Figura 18. Detección de objetos en movimiento utilizando la diferencia de <i>frame</i> adyacente (<i>Frame 36, 38,40</i>)	61
Figura 19. Ocultamiento del objeto en estado estacionario (<i>Frame 57 59 60</i>)	61
Figura 20. Detección inadecuada debido a cambios bruscos de luz y camuflaje (<i>Frame 103, 105, 107</i>)	62
Figura 21. Estimación del modelo de fondo y extracción de los objetos en movimiento del primer plano (<i>Frame 33 36 39</i>)	63
Figura 22. Actualización inadecuada del modelo de fondo (<i>Frame 101,103</i>)	63
Figura 23. Alta sensibilidad del modelo de fondo debido a cambios bruscos de luz (<i>Frame 118</i>)	64
Figura 24. Estimación del modelo de fondo por el método de mediana aproximada (<i>Frame 95, 99</i>)	65
Figura 25. Estimación errónea del modelo de fondo debido a cambios bruscos de luz (<i>Frame 118, 120</i>)	65
Figura 26. Actualización inadecuada del modelo de fondo (<i>Frame 249, 253</i>)	66
Figura 27. Estimación del modelo de fondo con actualización	

selectiva de regiones (<i>Frame</i> 95, 99)	67
Figura 28. Inconvenientes de adaptación del modelo de fondo debido a cambios bruscos de luz. (<i>Frame</i> 118, 122)	67
Figura 29. Corrección de los efectos generados por cambios bruscos de luz (<i>Frame</i> 118, 120, 122)	68
Figura 30. Seguimiento de un peatón utilizando el filtro de kalman (<i>Frame</i> 34, 36, 38)	69
Figura 31. Diagrama de flujo del filtro de kalman	70
Figura 32. Ejecución del sistema de conteo con segmentación de un transeúnte	73
Figura 33. Software de conteo monitoreando dos transeúntes	74
Figura 34. Ejecución del sistema de conteo con traslape de transeúntes	74

RESUMEN

En este proyecto de grado se presenta el diseño, la implementación y la verificación de un sistema de visión artificial para la detección y conteo de transeúntes en un entorno no controlado en tiempo real. El sistema diseñado está libre de parámetros de configuración y se implementa con una sola cámara en posición estática, para localizar y monitorear en tiempo real los transeúntes en la zona de visión, determinar su dirección y mantener un registro discriminado de las personas que se desplazan en direcciones opuestas.

El software de la aplicación se desarrolló bajo la plataforma de Matlab[®], el cual realiza la captura digital de video, el procesamiento de la secuencias de *frames*, la caracterización e interpretación de los transeúntes en el frame y el registro del conteo. La interfaz grafica de usuario se desarrolló mediante la herramienta gráfica de Matlab[®] (*GUIDE*).

Para el diseño del algoritmo se investigaron diferentes técnicas de segmentación de objetos en movimiento, se implementaron las técnicas más robustas y eficientes, teniendo en cuenta las limitaciones en velocidad de procesamiento del equipo de cómputo empleado. El objetivo de la segmentación es distinguir aquellas zonas diferentes al fondo, para ello es de importancia que en la inicialización del fondo no haya objetos en movimiento, los cuales son considerados como objetos de frente. Entre las técnicas estudiadas, las más usadas son la diferencia de *frames* adyacentes, el método de múltiples gaussianas, el filtro de mediana aproximado y el filtro de mediana aproximado selectivo. Los mejores resultados se obtuvieron con el filtro de mediana aproximado selectivo, ya que permite la actualización del fondo pixel a pixel y su selectividad evita que objetos de interés pasen a ser parte del fondo.

El filtro de mediana aproximado selectivo se adapta fácilmente a cambios suaves de iluminación del entorno. Las restricciones para el funcionamiento correcto del algoritmo radican en obtener un fondo inicial de la escena completamente estático y libre de objetos en movimiento.

El algoritmo que realiza el proceso de seguimiento de los transeúntes y etiquetado se diseñó a partir del principio de distancia euclidiana. El objetivo del algoritmo es discernir si el centroide corresponde al mismo transeúnte o no, en caso verdadero al transeúnte se le asigna la misma etiqueta, en caso contrario al peatón se le asigna una nueva etiqueta, facilitando su identificación y conteo.

ABSTRACT

This research project presents the design, the implementation and the verification of an artificial system of vision for the detection and counting of pedestrians in a not controlled real time environment. The designed system is free of configuration parameters and is implemented by only one camera in static position, to locate and to monitor in real time the pedestrians in the zone of vision, to determine their direction and to keep a discriminated record of the people who move in opposite directions.

The software of application was developed under the platform Matlab, which capture in digital way video, the processing of frames stream, the characterization and interpretation of the pedestrians in the frame and the counting register. The user's graphical interface developed by means of Matlab's graphical tool (GUIDE).

For the design of the algorithm there were investigated different techniques of segmentation of objects in movement, there were implemented the most robust and efficient techniques, having in it counts the limitations in speed of processing of the equipment of computer used. The aim of the segmentation is to distinguish those zones different from the background, for it performs importance that in the initialization of the background there are no objects in movement, which are considered to be objects of foreground. Among the studied techniques, the most used they are the differentiation of adjacent frames, the method mixture of gaussians, the approximate median filter and the approximate selective median filter. The best results were obtained by the approximate selective median filter due to the fact it allows the update of the background allows pixel to pixel and its selective character prevents that objects of interest appear in the background.

The approximate selective median filter adapts easily to soft changes of lighting of the environment. The restrictions for the correct functioning of the algorithm take root in obtaining an initial completely static background of the scene and freely of objects in movement.

The algorithm that realizes the process of follow-up of the pedestrians and labeling was designed from the technique of Euclidean distance. The aim of the algorithm is to discern if the centroide corresponds to the same pedestrians or not, in real case the same label is assigned to the pedestrian, in case opposite to the pedestrian a new label is assigned to him, facilitating his identification and count.

INTRODUCCIÓN

El monitoreo de objetos en movimiento es una herramienta importante para el diseño de aplicaciones en seguridad, control de acceso, control de multitudes en interiores o exteriores, control de inventarios, conteo de productos, selección y registros de productos, control de calidad, registros estadísticos en centros comerciales, en edificios públicos, museos, bibliotecas o centros recreativos. Se puede considerar como una etapa inicial para aplicaciones mucho más complejas como reconocimiento de características particulares en peatones, seguimiento de peatones sospechosos, identificación de transeúntes con actitudes delincuenciales, seguimiento de presos en patios de alta seguridad, entre otras.

Los sistemas de video vigilancia tienen cada vez mayor demanda. La presencia de numerosas cámaras de seguridad en cualquier entorno urbano es un hecho. Según Sandoval¹, gracias al avance técnico y funcional se ha logrado que instalar cámaras de captura de video no precise de altas inversiones económicas y, por ende, la mayoría de bancos, estaciones, aeropuertos, centros comerciales y parqueaderos incorporen en sus instalaciones un sistema de medianas prestaciones para la seguridad basada en video vigilancia.

La innovación tecnológica es un factor apremiante en el mundo actual. En este ámbito la electrónica y la inteligencia artificial, son algunas de las ciencias de más rápida evolución, pues están directamente implicadas en áreas de mucho interés para el ser humano como son la medicina, la gestión de recursos naturales, la comunicación y la gestión de la información.

El trabajo de grado que aquí se introduce se enmarca en el área de la visión artificial. Esta disciplina tiene como finalidad la extracción del mundo físico a partir de imágenes, utilizando para ello una cámara y un computador². El objetivo de un sistema de visión artificial es actuar sobre una representación de una realidad que le proporciona información sobre brillo, colores, formas, texturas, movimiento u otra característica relevante que permite obtener una interpretación semántica de los objetos presentes en dicha realidad.

Durante los últimos años se han realizado gran cantidad de investigaciones y trabajos en el área del procesamiento de la silueta humana en imágenes con el objetivo de obtener mayor precisión y confiabilidad en los procesos de estimación del conteo de peatones en interiores y exteriores.

¹ SANDOVAL C. Seguimiento de objetos en secuencias de video. Universidad Politécnica de Valencia. Departamento de Comunicaciones. 2003

² VÉLEZ J. F. MORENO A. B. SÁNCHEZ A y ESTEBAN J L. Visión por computador. Universidad Rey Juan Carlos. 2003

A medida que la cantidad de métodos presentes en la literatura aumentan, el problema general aun no se ha resuelto y la mayor dificultad radica en que no existe un método robusto para la correcta detección de objetos con siluetas irregulares.

El conteo de perfiles de objetos es una importante tarea en los sistemas automáticos de video vigilancia; puede ser una sub-tarea en una etapa determinada de un procesamiento o puede ser una tarea primaria. La estimación robusta de la cantidad de objetos, para el caso específicamente de personas puede responder preguntas tales como ¿cuántas personas estuvieron dentro de una habitación entre dos tiempos determinados?, ¿cuántas personas transitaron por un pasillo en determinado horario? o ¿determinar el flujo aproximado de turistas que salen o ingresan a la ciudad por las diferentes terminales?

De las diferentes técnicas halladas en la literatura para tratar este problema, se seleccionó la técnica basada en la extracción de patrones de reconocimiento de objetos. Dentro de los patrones más comunes se tiene la detección de bordes, la textura, el centro de masa, entre otros. Para este proyecto, específicamente se tuvo en cuenta los patrones de reconocimiento que aportan información sobre el tamaño (área) y la posición (centroide) del objeto de interés.

1. LA VISIÓN ARTIFICIAL

La visión artificial, se describe como un proceso cuya finalidad es proporcionar sentido de la vista, práctico y económico a cualquier tipo de sistema. Los sistemas de visión involucran la utilización de la tecnología existente de la forma más efectiva para dotar de cierto grado de autonomía a un sistema en aplicaciones específicas, explotando al máximo las restricciones del problema a resolver.

Este es un sistema de visión creado con el objetivo de superar las limitaciones del sistema de percepción humano en el sentido de atención, vigilancia y monitoreo; 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 monitoreo y vigilancia no se limita a reemplazar el sistema humano, sino que sirve como una herramienta de procesamiento de información que extienda su percepción y razonamiento³. El objetivo principal de la visión artificial es extraer las características de una imagen para su descripción e interpretación por un sistema de cómputo.

³ PLATANIOTIS, K. N. y REGAZZONI, C. S. Visual-Centric Surveillance Networks and Services. IEEE Signal Processing Magazine. Vol 22. Ed. 2005

2. IMAGEN DIGITAL

La imagen digital está formada por un conjunto definido de puntos llamados píxeles que forman una matriz con filas y columnas. Cada píxel de la imagen almacena la información de su tono o luminosidad, donde el tono negro es de valor 0 y el blanco de valor 1 en formato binario.

Generalmente se reconocen dos tipos de imágenes, en escala de grises y en color. En la primera, cada punto de la imagen se almacena en un *byte*, donde su valor numérico representa su tono que puede oscilar entre el blanco (255) y el negro (0), esto quiere decir que es una imagen donde existen 256 tonos de gris. En el segundo caso, una imagen de color se representa por sus componentes rojo, verde y azul por síntesis aditiva, cada componente es una imagen con 256 tonos pero asociada a cada uno de los colores primarios, que al visionarla el ordenador la colorea adecuadamente, la suma de las tres componentes por síntesis aditiva forma la imagen final donde cada píxel va definido por 3 *bytes*, el rojo, el verde y el azul, con una gama de $16^3 = 2.16$ colores posibles.

2.1 IMÁGENES A COLOR

Cuando se habla de imágenes de color real, se refiere a imágenes capturadas por cámaras de video, equipadas con sensores sensibles a la longitud de onda del espectro de luz visible.

El modo de color expresa la cantidad máxima de datos de color que se pueden almacenar en un determinado formato de archivo gráfico. Se considera el modo de color como el contenedor en que se coloca la información sobre cada píxel de una imagen.

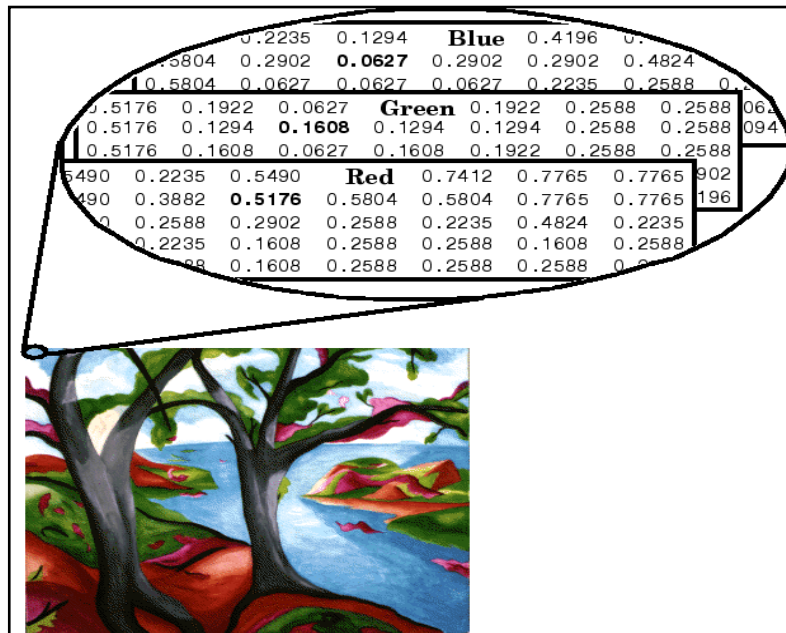
El modo *bitmap* o monocromático, corresponde a una profundidad de color de un bit, ofrece una imagen monocromática formada exclusivamente por colores blanco y negro puros, sin tonos intermedios entre ellos.

El modo escala de grises, maneja un solo canal para trabajar con imágenes monocromáticas de 256 tonos de gris, entre el blanco y el negro; admite cualquier formato de grabación.

El modo color indexado, se denomina así, porque tiene un solo canal de color de 8 bits, por lo que sólo se puede obtener con él un máximo de 256 colores. En este modo, la gama de colores de la imagen se adecua a una paleta con un número restringido de ellos, por lo que puede resultar útil para trabajar con algunos formatos que sólo admiten la paleta de colores.

El modo *RGB* se utiliza en todos los procesos en los que el color se obtiene por mezcla aditiva de luces como en televisión, pantallas gráficas, iluminación artificial entre otros. En todos estos dispositivos, la gama completa de colores se obtiene a partir de la mezcla de tres colores primarios, como se observa en la Figura 1.

Figura 1. Imagen en modo RGB



Fuente: Manual de usuario del *toolbox* de procesamiento de imágenes de matlab®

2.2 FORMATOS GRÁFICOS DIGITALES

Existe una gran variedad de formatos para las imágenes digitales dentro del campo de la informática. Algunos son dependientes del sistema operativo, mientras que otros son multiplataforma independientes del formato. Existen ciertas aplicaciones que permiten reconocer y traducir formatos de otros sistemas, por lo que se tiene un alto nivel de compatibilidad.

El formato *JPEG* (*Joint Photographic Experts Group*), se trata de un formato *bitmap* con distintos niveles de comprensión y pérdida de calidad. Su principal ventaja es que permite reducir el tamaño de las imágenes, soporta cualquier resolución espacial y cualquier profundidad de bits, por lo que es el más apropiado para fotografía digital, la mayoría de cámaras digitales operan internamente con este formato.

2.3 FORMATOS NUMÉRICOS

La calidad en toda imagen de mapa de bits está condicionada por la cantidad de bits utilizada para representar cada pixel, por el sistema de coordenadas o espacio de color empleado para describir los colores numéricamente y la tecnología empleada en el sistema de captura.

Desde el punto de vista numérico es posible trabajar con datos que tienen distintas escalas según el tipo de formato numérico que soporte el software de gestión. En Matlab[®], es posible trabajar con los siguientes formatos numéricos:

- *uint8* (8 bits). Establece un arreglo de números enteros en el cual la escala de grises o de color de un pixel esta en el rango de [0,255]; es decir, cada pixel puede contener valores en el rango de 0 a 255 con un total de $2^8=256$ valores.
- *Double*. Contiene un arreglo de valores enteros y no enteros, el cual la escala de grises o de color de un pixel esta en el rango de cero a uno.
- *Logical*. Contiene un arreglo únicamente de ceros y unos, generalmente una imagen en este formato se conoce como una imagen binaria, se emplea como mascara de extracción de los objetos o zonas que componen la imagen.

2.4 PROCESAMIENTO DIGITAL DE IMÁGENES

Es el conjunto de técnicas que se aplican a las imágenes digitales con el objetivo de mejorar la calidad o facilitar la búsqueda de información, usando como herramienta principal una computadora; el procesamiento digital de imágenes es una forma de procesamiento digital de señales en dos y tres dimensiones.

El estudio del procesamiento digital de imágenes se basa en dos áreas de aplicación primordiales, el mejoramiento de la calidad de la información contenida en una imagen y el procesamiento de los datos contenidos en un escenario a través de una maquina de percepción autónoma.

3. CONFIGURACIÓN DE UN SISTEMA DE CAPTURA

Para una aplicación de procesamiento de imágenes, el primer paso es preparar el ambiente de captura, el entorno en el cual se toman las imágenes, luego se configura el hardware apropiadamente. La caracterización, el desarrollo y las funcionalidades del sistema final, son concertadas con las prioridades del usuario, en este orden de ideas se identifican tres pautas muy importantes; los parámetros de captura, la iluminación y el movimiento.

3.1 PARÁMETROS FUNDAMENTALES DE UN SISTEMA DE CAPTURA

A continuación se describen los parámetros más importantes en un sistema de captura, tales como: resolución, campo de visión, distancia de trabajo, profundidad de campo y contraste.

3.1.1 Resolución. Indica el número de píxeles contenidos en cada unidad de longitud; como la resolución mide el número de píxeles por longitud, se deduce que a mayor resolución, mayor número de puntos de imagen en el mismo espacio y, por consiguiente, mayor definición o nitidez de la imagen.

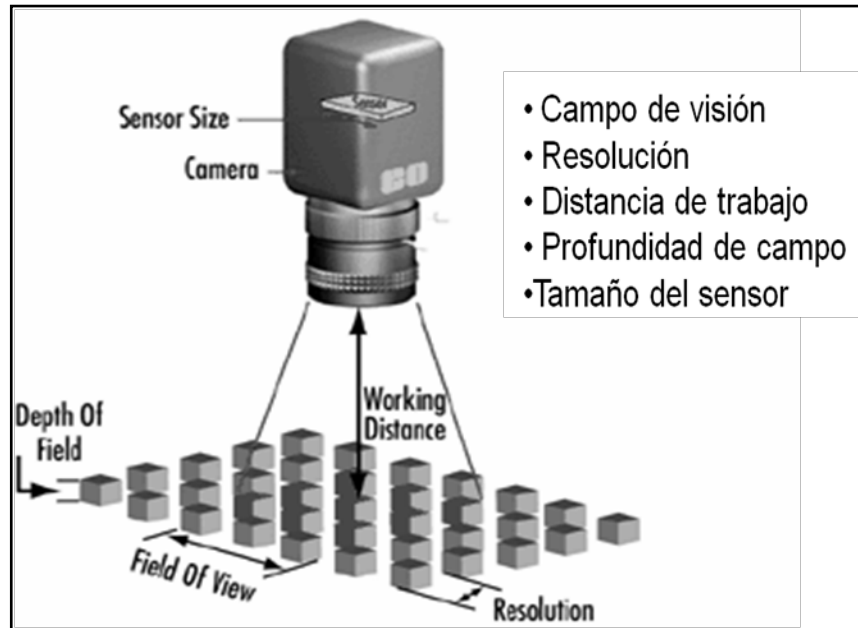
3.1.2 Campo de visión. Es el área de inspección que la cámara eventualmente puede captar.

3.1.3 Distancia de trabajo. Es la distancia medida desde el frente del lente de la cámara hasta el objeto que se desea inspeccionar.

3.1.4 Profundidad de campo. Es la distancia focal en la cual la imagen captada por la cámara es nítida (está enfocada).

3.1.5 Contraste. Se define como la diferencia relativa en intensidad entre un píxel de la imagen y sus píxeles vecinos. La resolución y el contraste son factores estrechamente relacionados que contribuyen a la calidad de la imagen. El contraste define las diferencias en los valores de intensidad entre el objeto de interés y el fondo. El sistema de captura debe estar bien calibrado para que en la imagen se distingan los objetos del fondo, sin embargo, empleando técnicas de iluminación apropiadas se puede mejorar el contraste de la escena.

Figura 2. Parámetros fundamentales de un sistema de captura



Fuente: <http://www.slideshare.net/omarspp/imagen-fundamentos-i>

3.2 ILUMINACIÓN

Uno de los aspectos más importantes del ambiente de captura es la iluminación, ya que con imágenes adquiridas bajo condiciones de iluminación apropiadas el software de procesamiento resuelve las tareas de forma más eficiente y sin errores; una de las ventajas de una buena iluminación, es el aumento del contraste entre los objetos y el fondo de la imagen, además cuando se controla la iluminación se minimizan los errores de segmentación y falsa detección debido a variables externas como la luz ambiente.

Los tipos comunes de fuentes de iluminación incluyen el halógeno, *LED*, fluorescente y láser; dependiendo del tipo de iluminación que se seleccione se puede determinar el éxito o fracaso de la aplicación; ya que una iluminación inapropiada puede causar sombras y luces intensas que degradan la calidad de la imagen.

3.3 CONSIDERACIONES DE MOVIMIENTO

En algunas aplicaciones, el objeto bajo inspección se mueve o alguna parte de él lo hace. En otras aplicaciones, se necesita mover la cámara para posicionarla en diferentes lugares.

Las imágenes adquiridas en aplicaciones que involucran movimiento pueden parecer borrosas cuando se excede la velocidad de respuesta de la cámara, por lo cual es necesario disminuir la velocidad de movimiento de los objetos o cambiar la cámara por una que tenga mayor velocidad de respuesta.

4. GENERALIDADES DEL ANÁLISIS DE IMÁGENES DINÁMICAS

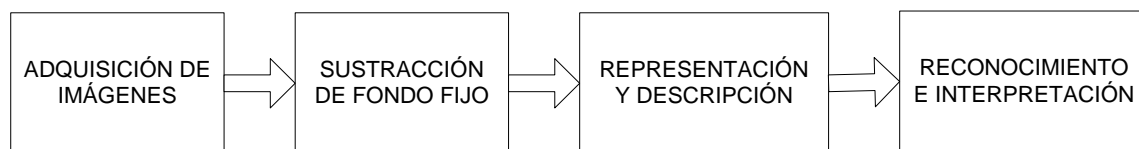
El análisis de una secuencia de *frames* de video o análisis de imágenes dinámicas es una aplicación de la visión computacional que permite realizar el seguimiento y localización de los objetos en movimiento presentes en una secuencia de *frames* de vídeo.

Al ser parte de la visión computacional, el análisis de imágenes dinámicas se realiza siguiendo las mismas etapas de la visión computacional, en las que aplicando procesamiento digital de imágenes se extrae la información necesaria para cumplir con la tarea de localización y monitoreo.

4.1 ETAPAS DE UN SISTEMA DE VISIÓN COMPUTACIONAL

Los sistemas de visión computacional para el análisis de secuencias de imágenes, aplicados a la detección de movimiento se componen de las etapas de adquisición, sustracción de fondo, representación y descripción, reconocimiento e interpretación como se observa en la Figura 3.

Figura 3. Diagrama de bloques de un sistema de visión artificial



La segmentación tiene como objetivo desmembrar del fondo (*background*) de la escena, los objetos en movimiento o regiones de interés que constituyen el primer plano (*foreground*) o frente de la escena. En el análisis de secuencias de video se presentan cuatro casos, como los clasifica Hernández⁴, los cuales se observan en la tabla 1, para este proyecto el problema que se desea analizar corresponde al caso 1, lo que implica que las técnicas de segmentación más eficaces son las basadas en el modelado y posterior sustracción de fondo.

Tabla 1. Análisis de la escena

Escena	Caso 1	Caso 2	Caso 3	Caso 4
Primer plano	Móvil	Móvil	Estático	Estático
Fondo	Estático	Móvil	Móvil	Estático

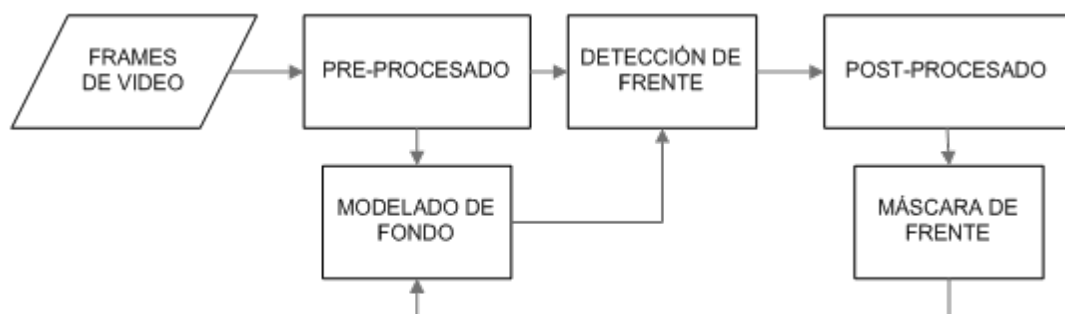
⁴ HERNÁNDEZ LÓPEZ, Francisco Javier. Segmentación de secuencias de video en tiempo real utilizando una WebCam. Centro de investigación en matemáticas. CIMAT. México. 2008

Según Sonsoles⁵, las secuencias de video no son ideales; es decir, el fondo de la escena no es completamente estático, por esto la segmentación automática de objetos es bastante compleja y constituye el proceso más importante en el procesamiento de video. También en su trabajo plantea cuales son los problemas más frecuentes que se deben resolver para obtener un algoritmo de detección de objetos eficiente computacionalmente y en tiempo real; estos inconvenientes son

- Cambios de iluminación de la escena.
- Sombras ocasionadas por los objetos en movimiento.
- Captura y actualización del fondo inicial de la escena.
- Configuración de los parámetros de funcionamiento del algoritmo de estimación de fondo y detección de objetos.
- Ruido presente en la secuencia de imágenes.
- Fondos multimodales.
- Camuflaje.

Con el fin de solucionar dichos problemas se utiliza un algoritmo de sustracción de fondo fijo generalizado, representado en el siguiente diagrama de bloques en el que se identifican las etapas de pre-procesado, modelado de fondo, detección de frente y post-procesado.

Figura 4. Etapas de un algoritmo de sustracción de fondo fijo



Fuente: Robust techniques for background subtraction in urban traffic video de Ching, Cheung y Kamath. Center for applied Scientific Computing

⁵ HERRERO MARTÍN, Sonsoles. Análisis comparativo de técnicas de segmentación de secuencias de video basadas en modelado de fondo. Universidad Autónoma de Madrid. 2009

4.1.1 Adquisición del *frame* de vídeo. Para realizar la adquisición de imágenes de la secuencia de video, es necesario contar con un dispositivo físico (Videocámara) sensible a una determinada banda del espectro electromagnético (espectro visible). Actualmente existen en el mercado diferentes tipos de dispositivos que pueden ser utilizados para este fin, desde los más básicos, como por ejemplo los que responden únicamente a la presencia o no presencia de luz hasta las videocámaras CCD que se utilizan con frecuencia en los sistemas de procesamiento de imágenes, debido a que brindan una mejor resolución que las videocámaras normales. La señal de salida de estos dispositivos puede ser digital o analógica.

Para efectos del procesamiento digital debe usarse un conversor analógico/digital (tarjeta digitalizadora) para las señales que no se encuentren en formato digital. Una vez obtenida la imagen digital el siguiente paso consiste en el procesamiento de la imagen.

4.1.2 Pre-procesado de video. El pre-procesado de video es la aplicación de técnicas de tratamiento a la señal de video digital para efectuar operaciones de filtrado espacial y operaciones morfológicas, con el fin de mejorar o suprimir detalles de la imagen y obtener una mejor interpretación visual de la escena.

La tarea principal de esta etapa es identificar y mejorar las características de las imágenes de la secuencia de video que se desean segmentar y suprimir los cambios insignificantes en la escena. Esta operación se realiza con el objetivo de aumentar el éxito de detección de los objetos de interés.

Para los sistemas de segmentación por movimiento es imprescindible una etapa de pre-procesado, ya que se requiere eliminar impurezas y variaciones en el fondo de la escena como los efectos de la lluvia o la nieve. Se requiere de filtros espaciales o temporales que realicen esta operación; sin embargo, esta solución implica que se deba reducir el número de *frame* analizados para cumplir con los requisitos de ejecución en tiempo real.

Las técnicas de pre-procesado de video que más se utilizan son: ajustes geométricos, ajustes radiométricos o de intensidad, derivada de la imagen y profundidad de la escena.

4.1.2.1 Ajustes geométricos. Son técnicas que permiten restaurar los valores de los píxeles que han sufrido alguna variación en el canal de la luminancia por efecto de los cambios de iluminación o movimientos de la cámara. Las técnicas más conocidas son la dilatación, la erosión, la rotación y las transformaciones proyectivas.

4.1.2.2 Ajustes radiométricos o de intensidad. Son técnicas que se encargan de corregir las variaciones de luminancia ocasionadas por la utilización de fuentes de iluminación inapropiadas. La solución consiste en normalizar el valor del pixel que corresponde al canal de luminancia en la imagen actual a la media y varianza de la imagen de fondo, donde para cada pixel de la imagen se obtiene dos umbrales de decisión referentes a los valores de luminancia iniciales. Estas técnicas se aplican en conjunto con la etapa de actualización de fondo; sin embargo, implican un aumento en el tiempo de estabilización de fondo de la escena⁶.

4.1.2.3 Derivada de la imagen. Son técnicas que calculan la variación de un punto en la imagen o la variación total de la imagen, con el fin de identificar las regiones que son insensibles a los cambios de iluminación o para la detección de características puntuales de los objetos que se mueven a alta velocidad en la escena.

4.1.2.4 Profundidad de la escena. Los métodos de sustracción de fondo apoyados en las técnicas de segmentación por profundidad, son ineficientes en la extracción de objetos en movimiento, mientras los objetos de interés posean características similares con el fondo; por ejemplo, comparten la misma textura, el mismo color y se encuentran mínimamente distanciados. Sin embargo, esta técnica es robusta en aplicaciones con iluminación variable, según Gordon⁷ esta técnica es un complemento importante en la segmentación de objetos por sustracción de fondo.

4.1.3 Modelado de fondo. Es la etapa principal de los sistemas que realizan la segmentación de objetos por movimiento. Se encarga de realizar la captura inicial del fondo, su actualización y representación de acuerdo al análisis de la secuencia de video.

Para obtener un buen modelo del fondo de la escena, inicialmente se supone que los valores de los pixeles que corresponden al objeto en movimiento varían de forma regular, mientras que el valor de los pixeles del fondo no cambia. Esta premisa solo se cumple en entornos controlados, donde se tiene un fondo estático o unimodal. En ambientes con objetos que pertenecen al fondo y presentan movimientos repetitivos y cambios de iluminación bruscos se conocen como fondos multimodales.

⁶ COSTANTINI R., RAMPONI G., BRACAMONTE J. Countering illumination variations in a video Surveillance environment. In Proc. of SPIE. Electronic Imaging Conference. San Jose, USA. 2001

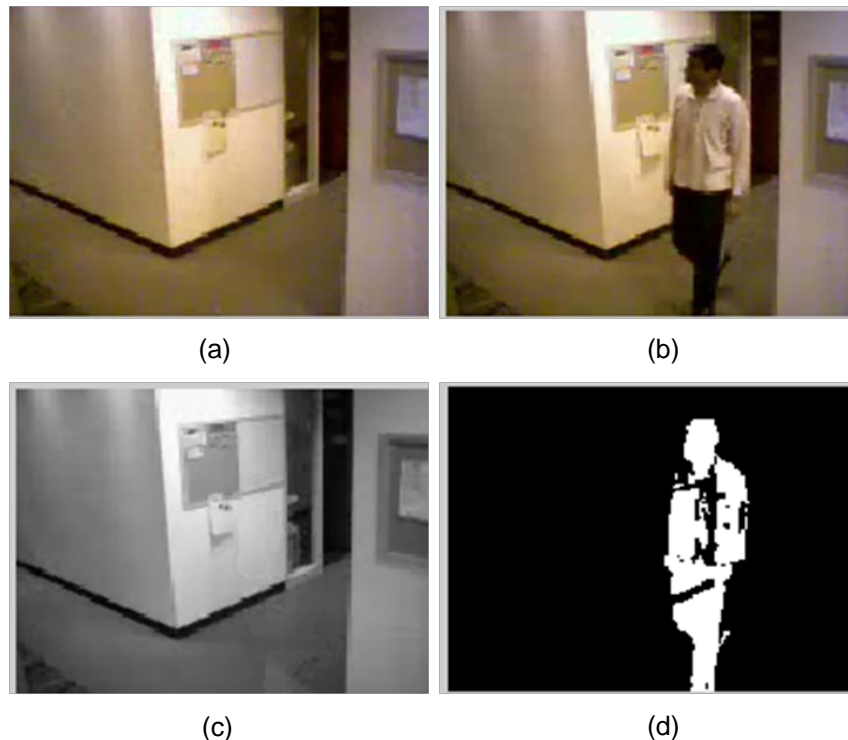
⁷ GORDON G., DARRELL T., HARVILLE M., WOODFILL J. Background Estimation and Removal Based on Range and Color. CVPR. 1999. IEEE Company Society Conference on Comp Vision and Patterns Recognitions. 1999-2-2459

La velocidad de cambio de los píxeles en la escena, es un parámetro muy importante para determinar el factor de actualización del fondo⁸, el cual controla la adaptación del modelo a cambios de iluminación leves y movimientos sutiles de objetos que pertenecen al fondo.

El proceso de modelado consiste en capturar una primera imagen como imagen de fondo, luego esta se toma como referencia inicial para realizar una actualización de acuerdo a los cambios previstos en las imágenes siguientes de la secuencia de video.

El proceso de inicialización de fondo se debe realizar bajo un estricto control de los objetos en movimiento, ya que si se presenta esta condición el fondo inicial capturaría objetos ajenos al fondo propio de la escena y produciría una detección incorrecta de los objetos de interés. La estimación inicial del fondo se debe realizar cuando no exista movimiento en la escena, como se observa en la Figura 5. La Figura 6 representa una estimación inicial incorrecta del fondo de la escena.

Figura 5. (a) Inicialización de fondo sin objetos en movimiento, (b) *Frame* actual, (c) Adaptación de fondo y (d) Detección correcta de un objeto en movimiento.



⁸ SAN MIGUEL AVEDILLO, Juan Carlos. Transmisión de secuencias de video a tasa binaria muy baja y adaptable basada en generación y transmisión de descripciones. Universidad Autónoma de Madrid. Escuela Politécnica Superior. Septiembre de 2006

Figura 6. (a) Inicialización de fondo con objetos en movimiento, (b) *Frame* actual, (c) Adaptación de fondo con inicialización incorrecta y (d) Detección incorrecta de objetos en movimiento.



Según Cristani⁹, el modelado de fondo tiene tres características definidas. La primera es el modelo de representación, es la descripción matemática con la cual se interpreta el fondo de la escena. La segunda es el modelo de inicialización, describe el método que se utiliza para obtener una estimación inicial de la escena como la captura de un *frame* inicial o el promedio de una secuencia de imágenes sin objetos en movimiento. La tercera es el modelo de adaptación o actualización del fondo de la escena, corresponde a las técnicas de ajuste de cambios en el entorno.

4.1.3.1 Modelo de representación del fondo. En este proyecto, los modelos de representación de fondo que se investigaron se clasifican como métodos básicos, paramétricos y no paramétricos.

Según el estado del arte, la técnica de diferencia de *frame*, el filtro promedio temporal y el filtro de mediana aproximado se agrupan como métodos básicos.

⁹ CRISTANI M., BICEGO M. y MURINO V. Multi-level background initialization using Hidden Markov Models. workshop on Video surveillance 2003 p.11-20

La técnica de diferencia de *frames*¹⁰, consiste en realizar una detección de los objetos de frente por medio de una sustracción de *frames*. En la Ecuación 1, el modelo de fondo en el instante t es (B_t) que corresponde a una imagen I en el instante t-1 de la secuencia de video.

$$B_t = I_{t-1} \quad Ec.1$$

La principal ventaja de esta técnica es su rápida respuesta a cambios repentinos en el primer plano de la escena y simplicidad computacional. Sin embargo, su capacidad para asimilar las variaciones de luminancia y ruido en la imagen es mínima. Se requiere de una etapa de post-procesado para el relleno de los objetos detectados, ya que es una técnica de detección de contornos activos.

El filtro promedio temporal¹¹, es una técnica cuyo modelo de fondo (B_t) se calcula como el promedio de una colección de imágenes consecutivas, las cuales no poseen objetos en movimiento como se describe en las Ecuaciones 2 y 3.

$$|I_t - B_t| > \tau \quad \rightarrow \quad B_t = \frac{1}{N} \sum_{n=t}^{n=t-N} I_n \quad Ec.2$$

$$|I_t - B_t| \leq \tau \quad \rightarrow \quad B_t = I_t \quad Ec.3$$

Esta técnica presenta grandes inconvenientes para el modelado del fondo, cuando se tienen objetos con movimientos muy suaves y entornos multimodales, además limita la posibilidad de detección de objetos a la escogencia de un solo umbral para toda la imagen.

El filtro de mediana aproximado¹², es una técnica que consiste en almacenar en un buffer los valores de los pixeles de una secuencia de imágenes, de acuerdo a un patrón de clasificación y comportamiento. Su principal ventaja radica en la rapidez de adaptación a cambios leves de luminancia y sencillez computacional. La formación del modelo de fondo radica en estimar el valor medio de la imagen, operando de manera acumulativa si la diferencia entre el fondo actual y el *frame* actual es mayor o diminutiva en caso contrario. Las Ecuaciones 3, 4, 5 describen matemáticamente este proceso.

$$I_t > B_t \quad \rightarrow \quad B_t = B_{t-1} + 1 \quad Ec.4$$

$$I_t < B_t \quad \rightarrow \quad B_t = B_{t-1} - 1 \quad Ec.5$$

$$I_t = B_t \quad \rightarrow \quad B_t = B_{t-1} \quad Ec.6$$

¹⁰ EWERTH Ralph, FREISLEBEN Bernd. Frame difference normalization: an approach to reduce error rates of cut detection algorithms for MPEG videos. 2003

¹¹ HERRERO MARTÍN., Op. cit

¹² Ibid.

Su principal desventaja radica en la lenta adaptación del fondo, cuando se produce un cambio en los objetos del fondo como lo menciona Cheung y Kamath¹³, debido a que el tiempo de actualización se limita a la velocidad de procesamiento de los *frames*.

Las técnicas paramétricas describen el modelo de fondo a partir de distribuciones de probabilidad estándar, comúnmente gaussiana, modelando características específicas del fondo lo que genera una mayor confiabilidad en la estimación de entornos con cambios bruscos de iluminación y vibraciones de la cámara.

El método de la gaussiana Simple (SG) interpreta los cambios que ocurren en el *frame* de fondo (B_t), asignado a cada pixel del fondo una distribución unimodal definida por los parámetros de la media (μ_t) y la varianza (σ_t^2) como se describe en las Ecuaciones 7 y 8.

$$\mu_t(x, y) = \sum_{i=1}^{i=t} \frac{B_i(x, y)}{t} \quad Ec. 7$$

$$\sigma_t^2(x, y) = \sum_{i=1}^{i=t} \frac{B_i^2(x, y)}{t} - \mu_t^2(x, y) \quad Ec. 8$$

Para cada instante de tiempo se evalúa el valor del pixel y se determina si pertenece al fondo o no. El proceso de selección consiste en determinar si el valor del pixel está dentro o fuera de la gaussiana definida para ese pixel.

El filtro de Kalman es una técnica recursiva para determinar los parámetros correctos de un sistema que evoluciona con el tiempo. Dados unos estimadores iniciales y los parámetros propios del sistema dinámico, el filtro va prediciendo y auto ajustándose con cada nueva medida.

Los fondos multimodales son aquellos que contienen objetos en movimiento periódico, como las hojas de arboles o las olas del mar, estas situaciones plantean un reto en el procesamiento de imágenes, ya que los pixeles que componen estas imágenes no tienen un valor discreto de luminancia, si no que corresponde a un conjunto finito de valores característicos; para resolver este tipo de problema se conoce una técnica llamada mezcla de gaussianas (*MoG*) que propone una solución a este problema¹⁴ que consiste en modelar la intensidad de los píxeles con una mezcla de k distribuciones gaussianas definidas por los siguientes

¹³ CHEUNG S. C, KAMATH C. Robust techniques for background subtraction in urban traffic video. Proc Elect Imaging: Visual Comm Image Proce 2004

¹⁴ STAUFFER C, GRIMSON W. Learning Patterns of activity using real-time Tracking. IEEE. 2000

parámetros, la media $\mu_{k,t}$, la varianza $\sigma_{k,t}^2$, y el peso $w_{k,t}$ de acuerdo a las Ecuaciones 9 y 10.

$$\mu_{k,t}(x, y) = \sum_{i=1}^{i=t} \frac{B_i(x, y)}{t} \quad \text{Ec. 9}$$

$$\sigma_{k,t}^2(x, y) = \sum_{i=1}^{i=t} \frac{B_i^2(x, y)}{t} - \mu_t^2(x, y) \quad \text{Ec. 10}$$

La *MoG* tiene la desventaja de requerir un tiempo elevado de procesamiento y es bastante sensible a cambios repentinos de iluminación.

Las técnicas no paramétricos son bastantes complejas. Estas se basan en el modelado de fondo a través del almacenamiento de los últimos valores del píxel, cálculo de rangos de valores del píxel y ajustes de funciones de predicción entre otros. Algunos modelos no paramétricos son densidad de núcleo¹⁵, modelos ocultos de Markov¹⁶ y métodos basados en código¹⁷ (*Codebooks*).

4.1.3.2 Modelo de inicialización de fondo. El proceso de inicialización de fondo consiste en generar un fondo inicial de la escena a partir de un conjunto de características obtenidas de un grupo de imágenes pertenecientes a la secuencia de video donde no existe objetos en movimiento. Esta condición difícilmente se puede controlar en entornos abiertos, por ello es necesario capacitar al modelo con un fondo inicial robusto. El método más conocido para inicializar el fondo es la mediana de una secuencia previa de la escena sin objetos.

4.1.3.3 Modelo de adaptación del fondo. El proceso de adaptación de fondo constituye el pilar fundamental para la segmentación de objetos por movimiento. Proporciona al fondo la capacidad de adaptarse a medida que la secuencia de video evoluciona. Las técnicas más utilizadas son la media móvil y la mediana aproximada, en ambas técnicas el proceso consiste en asignar un buffer a cada píxel y la imagen de fondo se adapta de acuerdo al comportamiento del píxel en *frames* anteriores. La técnica *running average* se adapta fácilmente a las variaciones leves de iluminación y el fondo se actualiza de acuerdo a la Ecuación 11.

¹⁵ ELGAMMAL A, DURAISWAMI R, HARWOOD D, DAVIS LS. Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. IEEE Jul. 2002

¹⁶ STENGER B., RAMESH V., PARAGIOS N., COETZEE F., BUHMANN J. M. Topology Free Hidden Markov Models: Application to Background Modeling. Conference on Computer Vision 2001

¹⁷ BUTLER D., SRIDHARAN S., BOVE. Real-time Adaptive Background Segmentation. IEEE Conference on April 2003.

$$B_t(x, y) = \alpha I_t(x, y) + (1 - \alpha)B_{t-1}(x, y) \quad Ec. 11$$

Los algoritmos de actualización selectiva, permite renovar solo aquellos píxeles seleccionados como fondo, si algún objeto permanece un periodo de tiempo mayor al de la actualización, este pasara a ser parte del modelo de fondo de la escena.

Los algoritmos de actualización no selectiva renuevan todos los píxeles de la imagen de fondo sin tener ningún criterio de selección o discriminación, causando que objetos en movimiento pasen a ser parte del fondo.

4.1.4 Detección de objetos de primer plano. Es la máscara de discriminación de regiones, permite enmascarar los objetos de primer plano e identificar cuáles son los píxeles que conforman dichos objetos en movimiento; para obtener esta máscara de selección, es necesario realizar la comparación del *frame* de video actual con el modelo de fondo en ese preciso instante, este proceso está en función de umbral de clasificación y las condiciones de iluminación de la escena.

El método más conocido para realizar la detección de frente por sustracción de fondo es realizando la diferencia entre la imagen actual y el fondo, este proceso concluye en dos estados donde los píxeles cuya diferencia es alta corresponde a objetos en movimiento, y en caso opuesto pertenecen al fondo como se observa en las Ecuaciones 12 y 13.

$$(I_t(x, y) - B_t(x, y)) \leq \tau \rightarrow F_t(x, y) = 0 \quad Ec. 12$$

$$(I_t(x, y) - B_t(x, y)) > \tau \rightarrow F_t(x, y) = 1 \quad Ec. 13$$

La comparación entre el *frame* de video y el fondo se realizo por medio de la diferencia absoluta, ya que esta operación de sustracción proporciona solo resultados positivos, los cuales son posibles referenciar con el umbral de clasificación mencionado anteriormente, esta operación permite obtener una máscara de discriminación de regiones con bordes gruesos, lo que facilita el post-procesamiento de la imagen para la extracción o segmentación del objeto.

4.1.4.1 Umbral de clasificación. La escogencia adecuada del umbral de clasificación permite obtener una máscara de discriminación que seleccione los píxeles cuyo resultado de la diferencia absoluta es un valor superior al umbral fijado, por consiguiente la segmentación de los objetos en movimiento se limita a seleccionar un umbral óptimo de clasificación.

La máscara de discriminación es un matriz binaria compuesta por pixeles de valor 0, si pertenece al fondo y pixeles de valor 1, si se considera como un objeto en movimiento como se muestra en la Figura 7.

Figura 7. Máscara de discriminación de frente



El umbral se estableció de forma empírica¹⁸, a partir de un proceso experimental de ensayo y error sobre una base de datos compuesta por videos tomados en la escena de análisis, en segundo caso el umbral se establece por medio de un cálculo adaptable¹⁹, donde se realizan una serie de operaciones dependiendo de parámetros como el número de falsas alarmas, la cantidad características locales especificadas, el modelo de fondo empleado y la información espacial de la escena.

4.1.5 Técnicas de post-procesado. las técnicas de sustracción de fondo y detección de objetos en movimiento, en la mayoría de los casos, no realizan un trabajo completo; por consiguiente es necesario aplicar una etapa de post-procesado o mejoramiento de la máscara de discriminación, con el fin aumentar las posibilidades de una correcta detección de objetos y el éxito de las etapas de descripción, reconocimiento e interpretación de la escena.

Existen diversas técnicas de post-procesado que proponen algoritmos basados en el análisis del modelo de color de la imagen, en el estudio de contornos, crecimiento de regiones entre otras²⁰.

¹⁸ SKIFSTAD K, JAIN R. Illumination independent change detection for real world image sequences. Graphics Image Process June 1989

¹⁹ TOTH D., AACH T., METZLER V., Bayesian spatial-temporal motion detection under varying illumination illumination-invariant change detection. EUSIPCO, Tampere, Finland, 2000

²⁰ CUCCHIARA R, GRANA C, PICCARDI M, PRATI A. Detecting moving objects, ghosts and shadows in video streams. IEEE 2003

4.1.6 Representación y descripción. Consiste en la extracción de los rasgos o información cuantitativa de las características básicas de los objetos por los cuales pueden ser diferenciados. También se realiza la clasificación de los objetos teniendo en cuenta parámetros como áreas, longitud del eje mayor, longitud del eje menor, entre otros; de tal manera se tienen en cuenta solo los objetos que cumplen con el perfil deseado, ya sean personas, autos u otros objetos de interés.

Después de realizado el proceso de detección, segmentación de movimiento y clasificación del objeto, se realiza la localización del centroide del objeto con respecto al origen de coordenadas de la imagen. Aunque el centroide está relacionado con una propiedad física de los objetos como es su masa, tiene una aplicación particular en el procesamiento digital de imágenes donde se considera cada pixel como una masa puntual ficticia.

4.1.7 Reconocimiento e interpretación. El reconocimiento es el proceso de etiquetado o asignación de un nombre a un objeto, basándose en la información que proveen sus descriptores. La interpretación involucra la asignación de significado a un conjunto de objetos de interés reconocidos.

5. DESCRIPCIÓN GENERAL DEL SISTEMA

Este proyecto consiste en el diseño e implementación de un sistema que permite monitorear y registrar los peatones que cruzan por un entorno determinado, comprendiéndose entorno como un pasillo, una calle, una puerta de acceso, o cualquier otro ambiente que facilite el flujo de peatones en dos direcciones. El sistema está en capacidad de clasificar los peatones como objetivos prioritarios de la escena, determinar su orientación y en función de esta, realizar el registro del peatón en tiempo real.

Un sistema de tiempo real es aquel en el que para que las operaciones computacionales estén correctas no depende sólo de que la lógica e implementación de los programas computacionales sean correctos, sino también del tiempo en el que dicha operación entrega su resultado. Si las restricciones de tiempo no son respetadas el sistema se dice que ha fallado²¹.

5.1 PROCESO DE ADQUISICIÓN DE VIDEO

Este proceso permite obtener una matriz multidimensional con los *frames* que componen la escena. Se realizó utilizando el *toolbox* de adquisición de imágenes de Matlab[®], el cual provee al programador de una colección de funciones y herramientas específicas para la captura, almacenamiento y procesamiento de video en tiempo real, además de ser compatible con diferentes tipos de dispositivos de captura de video.

5.2 ESTIMACIÓN Y SUSTRACCIÓN DEL FONDO

En términos básicos, se define el fondo como la parte estacionaria de la escena, luego para una estimación óptima del fondo, se precisa de una etapa de inicialización, donde la secuencia de video solo contenga elementos de *background*; de esta forma se obtiene un modelo de referencia inicial.

Se diseñaron diferentes algoritmos con el lenguaje técnico de programación de Matlab[®], en los cuales se aplican las técnicas más conocidas y empleadas para la estimación y sustracción de fondo en tiempo real, las cuales se enumeran a continuación:

²¹ DONALD Gillies.YUXIN Zheng, Dynamic Interactions with the Philosophy of Mathematics, Theory. 2001. p. 437-459.

1. Diferencia de *frames* adyacentes.
2. Método de múltiples gaussianas
3. Filtro de mediana aproximado
4. Filtro de mediana aproximado selectivo

Se diseñó un método para normalizar la luminancia, y aumentar la expectativa de éxito en la segmentación de objetos en movimiento; reducir la detección de sombras y mitigar el problema de semejanza de color entre los objetos del primer plano y el fondo de la escena. Sin embargo, este método no se aplica en la propuesta final del proyecto, debido a la incompatibilidad con los métodos de actualización de fondo progresiva, cómo se explica en la sección 6.3.5.

5.3 SEGMENTACIÓN DE OBJETOS EN MOVIMIENTO

Para realizar la segmentación de objetos, se utilizó la técnica de detección de frente por diferencia absoluta, como se explica en la sección 4.1.4.1, la cual consiste en comparar la imagen actual con la imagen de fondo y de acuerdo a un umbral de clasificación, asignar a cada pixel un valor binario, basado en el siguiente criterio:

$$\text{Pixel } I_{(x,y)} = \begin{cases} 1, & \text{si } I_{(x,y)} \in \text{objeto en movimiento} \\ 0, & \text{si } I_{(x,y)} \in \text{fondo estático} \end{cases}$$

Al final, se obtiene una matriz que contiene los pixeles en valor binario, la cual asocia los objetos de primer plano que se encuentran en movimiento y se le conoce como máscara de primer plano.

5.4 CLASIFICACIÓN AUTOMÁTICA DE OBJETOS

La selección de los objetos con perfil de peatón, es una tarea compleja debido a la no rigidez de los movimientos del cuerpo humano, el vestuario, los problemas de iluminación en los *frames*, los problemas de oclusión; además la apariencia de los humanos cambia, no solo con el cambio del punto de vista sino que también existe un margen de variación relacionada con las partes visibles del cuerpo humano.

De acuerdo a estos antecedentes, se desarrolló un algoritmo que clasifica los objetos de acuerdo al área que ocupa en la escena, este patrón de selección está en función solamente de la altura a la cual se coloca la cámara de video y no constituye un parámetro configurable.

5.5 SEGUIMIENTO DE OBJETOS EN MOVIMIENTO

Inicialmente se aplicó el algoritmo de seguimiento por medio del filtro de Kalman, sin embargo, las restricciones al seguimiento de un solo objeto, los criterios de inicialización del filtro, y su característica de función de densidad de probabilidad unimodal, implica que su aplicación para el seguimiento de varios objetos sea un proceso mucho más complejo y costoso computacionalmente; por lo tanto, se diseñó un algoritmo que organiza los peatones por orden de llegada en la escena, realiza el emparejamiento y determina la correspondencia de los objetos nuevos con los objetos previos en la escena, por medio de la comparación de la posición de sus centros de masa.

Se utilizó el método de la distancia euclidiana, como técnica de comparación entre los objetos presentes y los objetos futuros, o visto desde otra perspectiva como la correspondencia de un objeto entre el $frame_{t-1}$ y el $frame_t$.

5.6 REGISTRO Y CONTEO DE OBJETOS

El algoritmo para el conteo y registro de peatones se realizó en base al cambio de etiquetas, para tal fin la escena se dividió en dos áreas previamente definidas, son áreas principales de conteo, en las cuales el peatón toma la etiqueta correspondiente al área donde está ubicado, cuando se produce el cambio de etiqueta para el peatón, el sistema reconoce que este cruzó la línea de conteo e inmediatamente modifica la variable de conteo, en la sección 6.7 se explica con detalle el conteo de objetos.

5.7 ESPECIFICACIÓN DE REQUISITOS

Este proyecto de desarrollo e implementación de software debe cumplir con ciertas condiciones de funcionalidad; es importante que el sistema funcione de forma correcta y en unidades de tiempo razonable, debe ser flexible, robusto y adaptable en la tarea para el cual se diseñó.

5.7.1 Especificaciones funcionales. Los requisitos funcionales son aquellos que indican que debe realizar la aplicación, que servicios adicionales ofrece, los estados que el sistema resuelve y las situaciones en las cuales el sistema puede fallar; a continuación los requisitos funcionales:

- La aplicación debe ejecutarse en tiempo real, cumpliendo con los objetivos de detección, seguimiento, registro y conteo de peatones en un ambiente controlado y para un máximo de 4 transeúntes en la escena.
- El algoritmo de adquisición de video debe estar en la capacidad de procesar y suministrar los *frames* que componen la secuencia de video de forma inmediata.
- El algoritmo de estimación de fondo debe capturar y modelar una imagen de fondo inicial, que sea insensible a los cambios leves de luz, pero que permita percibir los movimientos de objetos ajenos a la escena.
- El algoritmo de clasificación debe filtrar las áreas que no corresponden con el perfil de peatón en la escena y etiquetar aquellas que cumplen con el perfil.
- El algoritmo de seguimiento debe capturar al peatón desde el momento mismo que ingresa a la escena y no perder su posición hasta que abandone la región de monitoreo.
- El sistema debe estar en la capacidad de recuperarse en un tiempo determinado, en caso de perder al peatón.

5.7.2 Especificaciones no funcionales. A continuación los requisitos que detallan como debe responder el sistema y lo que se espera de él.

- La visualización de los resultados debe ser intuitiva, se debe tener la certeza de que el sistema está realizando el seguimiento y conteo de los peatones en tiempo real.
- El sistema debe responder de forma consecuente con el movimiento de los peatones en la escena, mientras se cumplan las restricciones de velocidad de los peatones.
- Una de las restricciones más importantes de la aplicación está relacionada con la posición de los peatones en la escena, si estos se encuentran muy próximos entre sí, el sistema no es capaz de diferenciar uno del otro, y el resultado será tomado como un nuevo transeúnte.
- El sistema funcionara correctamente mientras el número de peatones sea inferior o igual a cuatro personas en un mismo instante en la escena, ya que el sistema se limita al diseño del software y los recursos de hardware utilizados.

6. DESARROLLO FUNCIONAL DEL SISTEMA

Para el desarrollo e implementación del sistema de monitoreo de perfiles de objetos se utilizó un computador con un procesador pentium® Dual-Core E5200 @ 2.50GHz y 2G de memoria RAM, con el sistema operativo Windows XP Versión 5.1 Service Pack 3 V.5857 de Microsoft®.

La aplicación se diseñó en el entorno de programación de Matlab® Versión 7.6.0.324 (R2008a), con la ayuda de los paquetes de herramientas *Image Acquisition Toolbox™* Versión 3.1 (R2008a), *Image Processing Toolbox™* Versión 6.1 (R2008a) y el entorno de desarrollo grafico *GUIDE* de Matlab®.

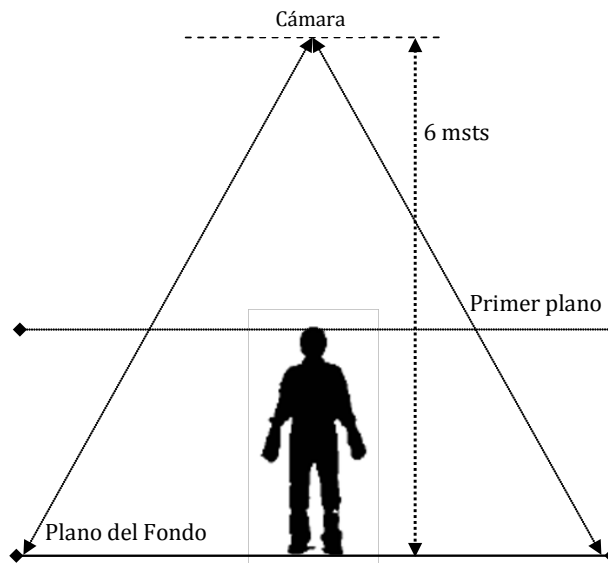
Se utilizó dos tipos de cámaras *WEB* convencionales para la transferencia de los *frames* de video al computador por medio de la interfaz *USB* (*Genius eye 320* y *Genius Look 110*), son cámaras de reducidas prestaciones y bajo costo, pero que cumplen con los requisitos de video para la implementación del prototipo del sistema que se implementó.

6.1 CONSIDERACIONES INICIALES DEL SISTEMA

El sistema implementado es libre de parámetros para el usuario y usa una cámara de video fija para localizar y monitorear en tiempo real a los transeúntes en la zona de interés, determinar su dirección y mantener un registro discriminado. A continuación se relacionan ciertas consideraciones que se deben tener en cuenta para el óptimo funcionamiento del sistema.

6.1.1 Consideraciones en la adquisición de video. Para una buena captura de la escena la cámara debe estar totalmente fija, ubicada a una altura de 6 metros del suelo, cuyo eje focal debe estar de forma perpendicular al plano del fondo como se observa en la Figura 8.

Figura 8. Esquema de ubicación de la cámara de video



6.1.2 Consideraciones del entorno de la escena. El escenario se controla parcialmente; el sistema logra adaptarse a cambios de iluminación suaves, sin embargo, es necesario que el entorno no esté expuesto directamente a rayos de luz solar, ya que la cámara web utilizada es muy sensible a la reflexión de la luz producida por el fondo y los objetos. En cuanto al color del fondo no requiere un control exhaustivo.

6.1.3 Consideración de movimiento de los peatones. La velocidad del peatón está restringida a la respuesta de captura de la cámara y la velocidad de procesamiento del software, para este proyecto se considera que el peatón se desplaza en condiciones normales.

6.2 PROCESO DE ADQUISICIÓN DE VIDEO DIGITAL

Para el tratamiento de la señal de video digital en aplicaciones de tiempo real es necesario efectuar las operaciones lo más rápido posible, para lograr tomar decisiones al instante en que se analiza la escena. Por consiguiente, se desarrolló un algoritmo de adquisición de *frames* en tiempo real, compatible con cámaras de video comerciales.

El mercado de dispositivos de captura de imagen y video es extenso, luego dependiendo del hardware de captura que se escoja, el primer paso para trabajar con dichos dispositivos es realizar la instalación de los controladores.

Una vez instalado el dispositivo, es necesario identificar sus propiedades claves de acceso a través del software de programación, las cuales son: el nombre del adaptador, el id de identificación y los formatos de video soportados.

Un adaptador, es un software que permite la comunicación con el hardware de captura utilizando el controlador del dispositivo. El *toolbox* de adquisición de imágenes incluye en su registro diferentes adaptadores de acuerdo a las marcas y tipos de dispositivos reconocidos en el mercado.

La función `imaqhwinfo` retorna información sobre los adaptadores disponibles.

```
>>imaqhwinfo
ans =
    InstalledAdaptors: {'coreco' 'winvideo'}
    MATLABVersion: '7.6 (R2008a)'
    ToolboxName: 'Image Acquisition Toolbox'
    ToolboxVersion: '3.1 (R2008a)'
```

El adaptador (`winvideo`), soporta los dispositivos de video conectados, luego se obtiene los identificadores para el hardware especificado.

```
>> Device=imaqhwinfo('winvideo')
Device =
    AdaptorDllName: [1x81 char]
    AdaptorDllVersion: '3.1 (R2008a)'
    AdaptorName: 'winvideo'
    DeviceIDs: {[1] [2]}
    DeviceInfo: [1x2 struct]
```

Con los argumentos anteriores se obtiene el resto de información sobre el dispositivo y los formatos de video que soporta.

```
>> Camara2=imaqhwinfo('winvideo',2)
Camara2 =
    DefaultFormat: 'RGB24_352x288'
    DeviceFileSupported: 0
    DeviceName: 'Look 110'
    DeviceID: 2
    ObjectConstructor: 'videoinput('winvideo', 2)'
    SupportedFormats: {1x10 cell}
```

Con la información sobre la cámara web instalada, se establece la conexión virtual con el dispositivo, por medio del objeto de video; se crea con la función `videoinput`, y las propiedades claves de acceso. La función `videoinput` es una estructura de construcción, cuyos argumentos de entrada son las propiedades anteriormente definidas.

```
>> camvid2=videoinput('winvideo',2) %se crea el objeto
Summary of Video Input Object Using 'Look 110'.
Acquisition Source(s): input1 is available.
Acquisition Parameters: 'input1' is the current selected source.
                        10 frames per trigger using the selected
                        source.
                        'RGB24_352x288' video data to be logged upon
                        START.
                        Grabbing first of every 1 frame(s).
                        Log data to 'memory' on trigger.
Trigger Parameters: 1 'immediate' trigger(s) on START.
Status: Waiting for START.
0 frames acquired since starting.
0 frames available for GETDATA.
```

El objeto de video permite acceder completamente a las características principales del hardware de captura instalado.

```
>> imaqhwinfo(camvid1)
ans =
        AdaptorName: 'winvideo'
        DeviceName: 'USB Video Device'
        MaxHeight: 120
        MaxWidth: 160
        NativeDataType: 'uint8'
        TotalSources: 1
        VendorDriverDescription: 'Windows WDM Compatible Driver'
        VendorDriverVersion: 'DirectX 9.0'
```

La óptima adquisición del *frame* de video se obtiene realizando una configuración correcta del objeto de video.

Para identificar y controlar las propiedades del objeto se utiliza la función `get` y `set`; la primera permite obtener las características habilitadas y no habilitadas del dispositivo y la segunda controla las características habilitadas del hardware.

Finalizada la configuración del objeto, se debe inicializar para realizar la respectiva adquisición de datos, La función `start` prepara el objeto de video para dicha tarea.

```
start (camvid1) %se inicializa el objeto de video
```

El proceso de *data logging* se configura de forma manual, con el fin de sincronizar la ejecución del sistema con la operación de obturación de la cámara; es decir; una vez se captura el *frame*, el sistema lo procesa antes de la siguiente captura.

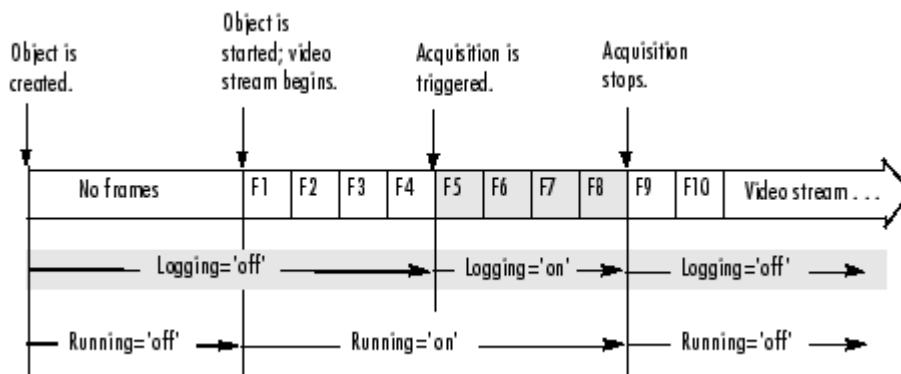
```
triggerconfig(camvid2,'manual');
```

Los *frames* capturados se almacenan en la memoria caché del sistema o en el disco duro, luego son llamados por medio de la función `getdata`, que los retorna como un arreglo matricial $m \times n$, donde m es el numero de muestras adquiridas y n el numero del canal de la fuente de video.

```
trigger(camvid2); %Orden de disparo para adquirir frame
frame=getdata(camvid2,1); %Adquisición de una muestra
```

La siguiente figura representa el procedimiento de adquisición de datos de video.

Figura 9. Fases de la adquisición de datos de video



Fuente: Manual del usuario del *toolbox* de adquisición de imágenes de Matlab®

6.3 ESTIMACIÓN Y SUSTRACCIÓN DE FONDO

La estimación y sustracción del fondo, consiste en separar el fondo o *background* del primer plano o *foreground*. Conceptualmente, se distingue el fondo del primer plano, si se define el fondo como la parte estacionaria de la escena, mientras el primer plano está formado por los objetos en movimiento o temporalmente estáticos.

Para separar los objetos en movimiento presentes en la escena, se realiza una sustracción de fondo, donde se calcula la diferencia entre los píxeles del *background* y los píxeles de la imagen adquirida.

Se requiere una representación precisa del fondo para detectar con claridad todos los objetos no estacionarios de la escena; una vez que objetos tales como automóviles, personas, entre otros; se identifican como *foreground*, constituyen la información de entrada a módulos de procesamiento de nivel superior que realizan tareas sobre dichos objetos como seguimiento, reconocimiento y clasificación.

Los resultados de las siguientes técnicas implementadas en este proyecto serán analizados y validados en el capítulo de pruebas y análisis de resultados.

6.3.1 Diferenciación de *frames* adyacentes. La técnica más simple de sustracción de fondo es la diferencia de *frames* adyacentes, consiste en:

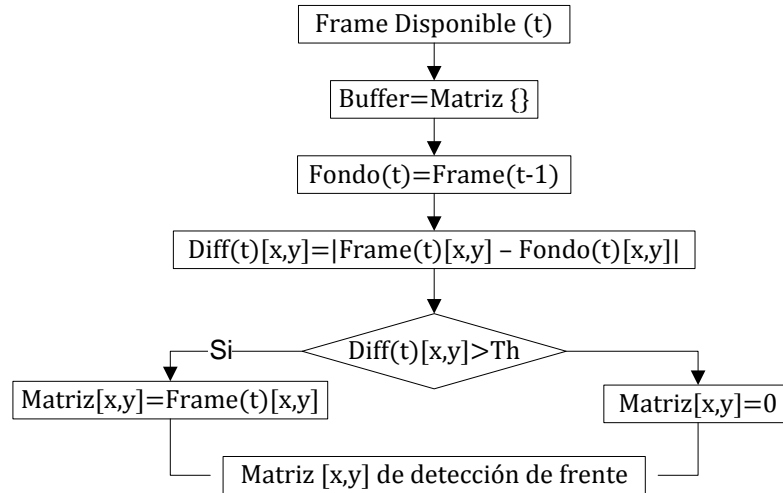
$$|imagen_i - imagen_{i-1}| > umbral \quad Ec. 14$$

El modelo de fondo estimado es el *frame* anterior consecutivo en la secuencia de video. Es un algoritmo que omite objetos de primer plano con movimientos casi o totalmente nulos y cambios de iluminación sutiles, además de que identifica como objetos de primer plano a aquellos objetos que pertenecen al fondo, que debido a los cambios repentinos de luz son detectados como objetos extraños en la escena y es vulnerable a los problemas de camuflaje de objetos de primer plano.

Este método solo funciona bajo determinadas condiciones de velocidad de los objetos y velocidad de captura (*fps*). La detección de objetos de primer plano es muy dependiente del umbral que se seleccione.

El algoritmo desarrollado es el siguiente:

Figura 10. Diagrama de flujo del método de *frames* adyacentes



El *frame* disponible en el instante $t=1$, se captura en formato *YCbCr*, luego se realiza la conversión de color a escala de grises y se toma como fondo inicial de la escena. Para el instante $t+1$ se captura la imagen de video y se realiza la diferencia absoluta entre los píxeles del fondo y el *frame* $t+1$; luego se efectúa la comparación de píxeles con un umbral de decisión (Th) y se extraen los objetos de primer plano cuyo valor de diferencia es superior a este umbral, y se clasifican como píxeles de primer plano, los cuales asocian intrínsecamente un objeto en movimiento, en caso contrario a dicha posición se le asocia el valor 0, el cual indica que es un píxel del fondo.

El valor de estos píxeles se almacena en una matriz buffer, en sus respectivas coordenadas, de tal forma que los objetos en movimiento son segmentados por la matriz de detección de frente. Para los instantes $t+n$ se continúa el mismo ciclo.

6.3.2 Método de múltiples gaussianas. La técnica de Stauffer y Grimson²² se reconoce como un método recursivo para la estimación de fondo y detección de objetos de primer plano mediante la suma ponderada de gaussianas. A diferencia del filtro de Kalman que rastrea la evolución de una sola función gaussiana, el método de mezcla de gaussianas rastrea múltiples funciones gaussianas simultáneamente.

El algoritmo se implemento de la siguiente manera:

²² STAUFFER C. y GRIMSON Wel. Adaptive Background Mixture Models for Real-Time Tracking. IEEE. 1999

Se realiza una comparación entre el pixel de entrada y el pixel del fondo medio, si el valor del pixel de entrada es cercano al valor del pixel del fondo, se considera un emparejamiento entre los pixeles; es decir, la diferencia absoluta entre los pixeles es menor que la desviación estándar multiplicada por un umbral positivo de desviación.

$$|I_t - \mu(I)_{t-1}| \leq \text{Umbral} * \sigma \quad \text{Ec. 15}$$

Se actualizan las variables asociadas al pixel como el peso asociado, la media del valor estimado y la varianza con respecto al nuevo pixel. Para pixeles emparejados, el conjunto de ecuaciones aumenta la confianza sobre el valor medido y aumenta la confiabilidad del valor estimado. Para pixeles no emparejados, el peso asociado al pixel disminuye de forma exponencial, mientras el valor del pixel medio estimado y la varianza se mantiene. Para los casos positivos aumentan el grado de confianza reflejado en un factor de aprendizaje del modelo.

El criterio de emparejamiento de pixeles, determina cuales pixeles hacen parte del modelo de fondo y cuales al primer plano, las condiciones evaluadas son:

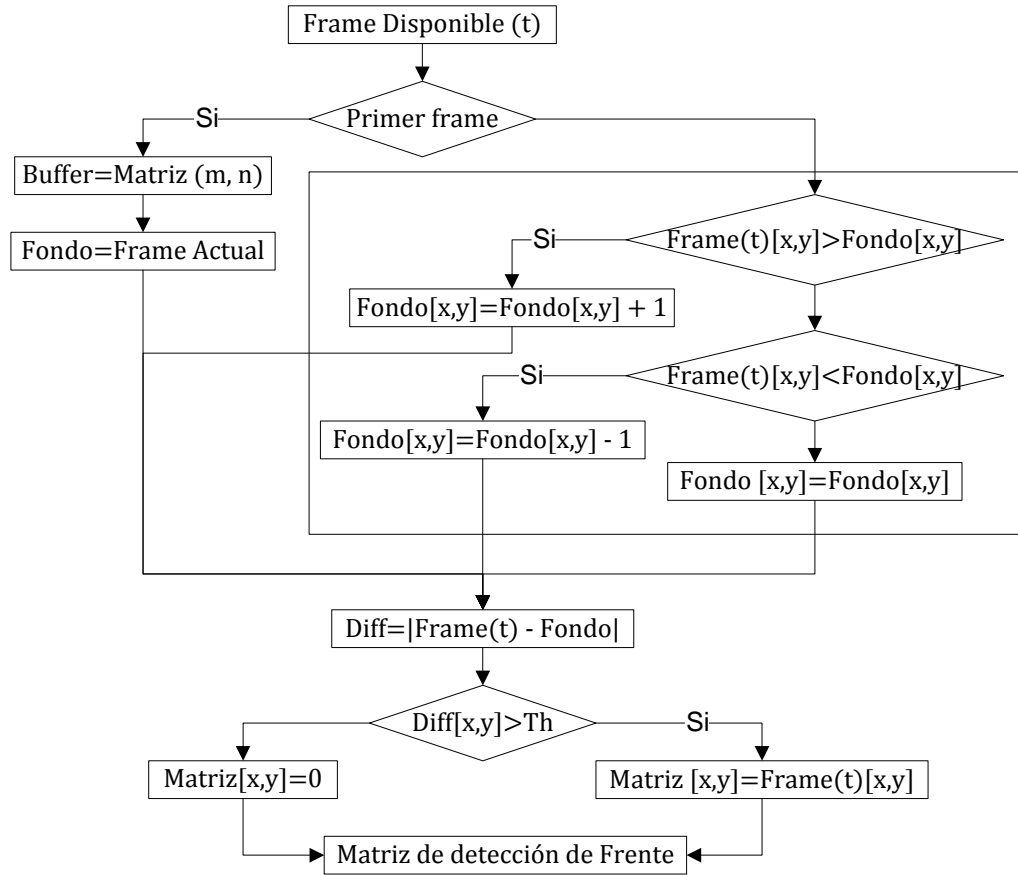
- Se seleccionan los pixeles con mayor grado de confianza y se ordenan de forma descendente.
- Se aplica un umbral para escoger los pixeles con mayor peso asociado.
- Se obtiene un modelo de fondo compuesto por los pixeles con mayor peso asociado.

Al final los pixeles que no están emparejados corresponden a objetos de primer plano de la escena.

6.3.3 Filtro de mediana aproximado. Este método es uno de los más básicos y robustos para la estimación de fondo y sustracción de objetos de primer plano.

El algoritmo que se implemento es el siguiente:

Figura 11. Diagrama de flujo del algoritmo de filtro de mediana aproximado



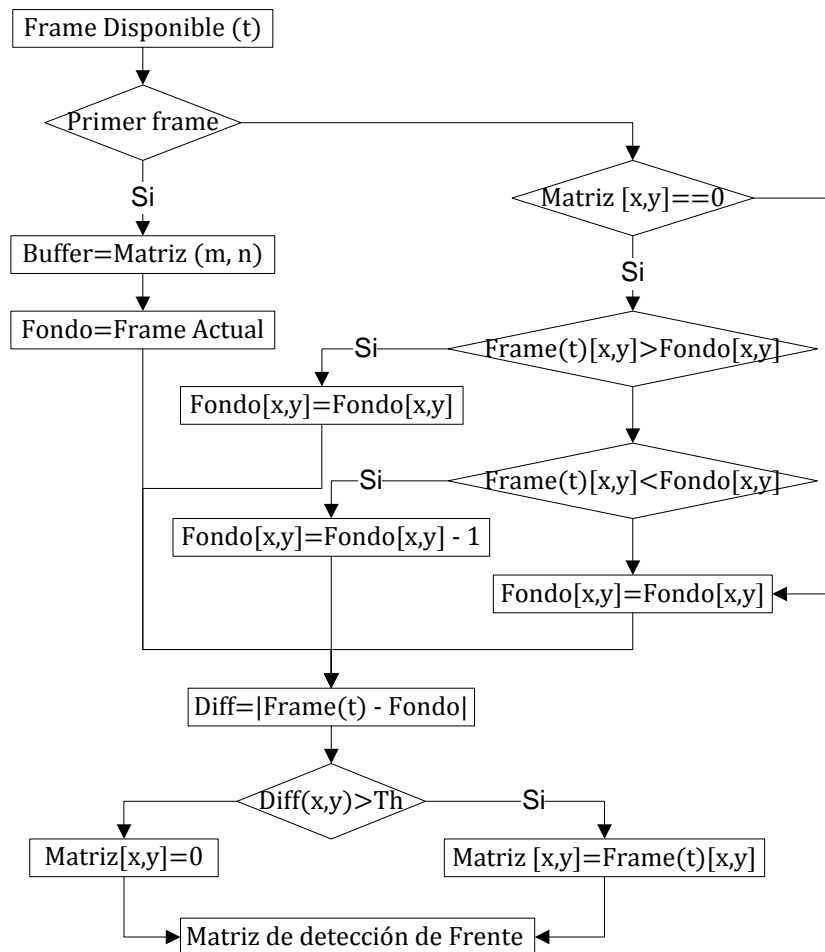
El sistema de monitoreo se especifica para el caso específico de una cámara estática; en primer lugar, se realiza una captura previa del fondo de la escena, luego se define una matriz dinámica de píxeles o buffer. El pre-procesamiento de conversión de color, comparación de píxeles y extracción de objetos en movimiento se realiza de igual forma que en la técnica de diferencia de *frames* adyacentes.

En el filtro de mediana aproximado, se introduce un nuevo proceso en la etapa de estimación y sustracción de fondo; esta etapa se encarga de actualizar el fondo total de la escena. El modelo de fondo se obtiene de la comparación entre el fondo actual y el fondo previo, determinando la posición de los píxeles que deben ser actualizados y aquellos que se mantendrán intactos.

6.3.4 Filtro de mediana aproximado selectivo. El filtro de mediana aproximado selectivo es una mejora al método convencional, el cual corrige el problema de inclusión de objetos de primer plano al fondo; es decir, solo se actualizan los píxeles que pertenecen al fondo de la escena ($\text{Matriz}(x,y)=0$).

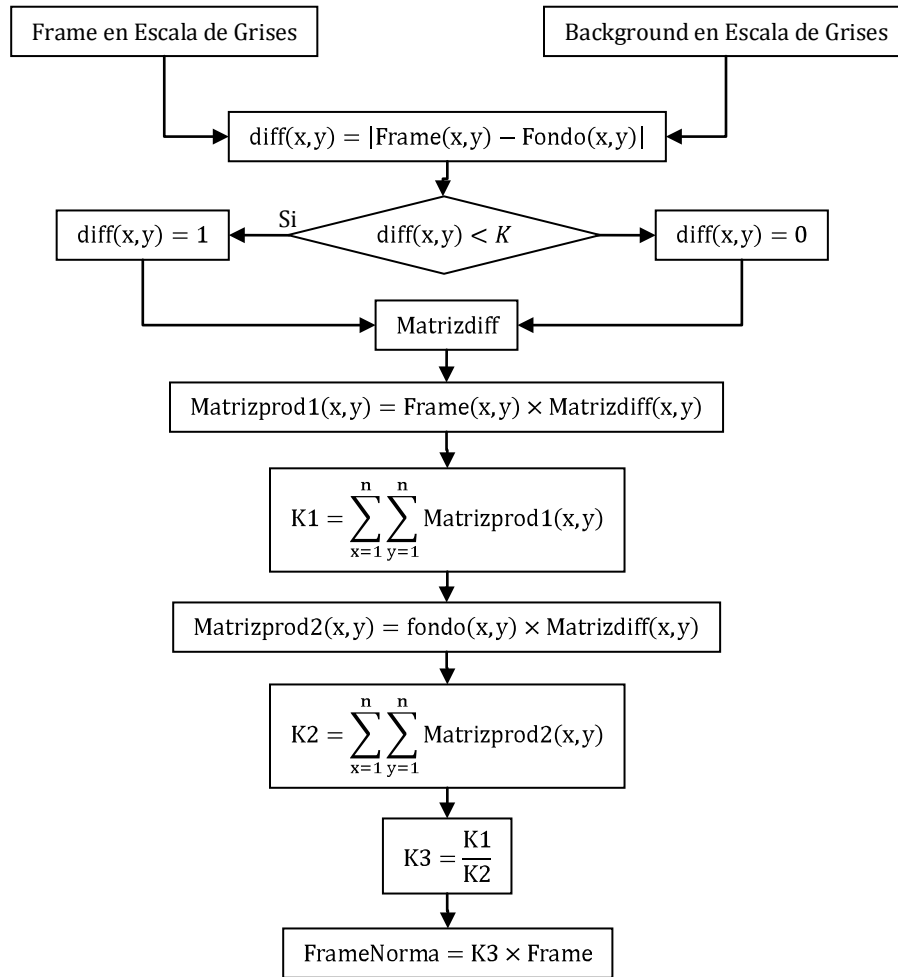
El proceso de selectividad consiste en identificar para cada nuevo *frame* los píxeles que corresponden al fondo o al primer plano; los píxeles que son clasificados como primer plano no se tienen en cuenta para la actualización, de esta forma se evita que objetos que no pertenecen al fondo pasen a formar parte de éste, cuando se detienen dentro de la región de interés.

Figura 12. Diagrama de flujo del algoritmo de filtro de mediana selectivo



6.3.5 Normalización de la luminancia. Para aumentar la confiabilidad del sistema de estimación y sustracción de fondo, se realizó un algoritmo para normalizar los valores de luminancia en los *frames*, como se observa en la siguiente figura. Esta técnica permite controlar los cambios bruscos progresivos incrementales de luminancia que se presentan en los *frames* de video respecto al fondo, con el fin de evitar las falsas detecciones de frente.

Figura 13. Diagrama de Flujo del bloque de normalización de luminancia



El factor de luminancia k , es un parámetro de referencia que permite identificar los píxeles que presentan un cambio brusco entre el fondo y el *frame* de video actual, con el objetivo de normalizar los niveles de luminancia en la imagen, este parámetro se escoge de manera experimental por ensayo y error.

6.4 SEGMENTACIÓN DE OBJETOS EN MOVIMIENTO

La segmentación es un proceso que divide una imagen en regiones u objetos cuyos píxeles poseen atributos similares, dicho proceso permite extraer los objetos de interés de la imagen para su posterior descripción y reconocimiento. Se

constituye como uno de los procesos más importantes en un sistema automatizado de visión artificial²³.

El método de segmentación que se empleó en este proyecto se basa en el modelado o estimación de fondo, con un proceso posterior de sustracción o comparación de *frames*. El proceso de segmentación consiste en un algoritmo que mantiene un modelo matemático de apariencia del fondo, como referencia de comparación para extraer los objetos en movimiento (primer plano) de la secuencia de video. La segmentación automática de los objetos es bastante compleja, siendo una de las tareas más complicadas dentro del procesamiento de video, como se señala Harville²⁴.

El algoritmo de segmentación que se desarrolla en este proyecto consiste en comparar los píxeles del *frame* de entrada con los píxeles del modelo de fondo estimado y seleccionar los píxeles pertenecientes a un objeto en movimiento. Esta discriminación se realiza por la diferencia absoluta entre los píxeles del fondo y el *frame* de video, en función de un umbral que depende de la escena, el ruido de la cámara y las condiciones de iluminación.

La elección del umbral es una operación decisiva para lograr segmentar los objetos de la secuencia de video. En el resultado del cálculo de la imagen diferencia, los niveles de gris pertenecientes a los píxeles de primer plano, tendrán valores superiores a los píxeles en ausencia de movimiento, por lo tanto el problema de localización y segmentación del objeto en movimiento se resuelve al seleccionar un valor de umbral óptimo.

La matriz resultante representa una imagen binaria con valor lógico 0, si el píxel pertenece al fondo y valor lógico 1, si el píxel pertenece a un objeto en movimiento.

El umbral se estableció de manera empírica como afirma Kasturi²⁵, se fijó un valor específico para todos los píxeles que componen la escena, de acuerdo a pruebas realizadas sobre videos de muestra tomados en un entorno controlado.

$$Umbral = \frac{Medida\ Experimental}{255} \quad Ec. 16$$

²³ JIMÉNEZ, J. G. Visión por Computador. Ed. Paraninfo. 2000

²⁴ HARVILLE M. Foreground Segmentation Using Adaptive Mixture Models in Color and Depth. Event. Detection and Recognition of Events in Video. IEEE. 2001

²⁵ KASTURI R. Machine Vision. McGRAWHILL. Editions 1995

6.5 CLASIFICACIÓN AUTOMÁTICA DE OBJETOS

La detección automática de personas y actividades en secuencias de video es un área de investigación en creciente desarrollo; la necesidad de garantizar la seguridad de personas y bienes en entornos cada vez mas masificados explica la enorme demanda de sistemas de video inteligente.

La gran complejidad del problema radica en la dificultad de modelar una persona, debido a la gran variedad de apariencias, poses, movimientos e interacciones existentes entre ellas. La clasificación de los objetos en la escena con perfil de persona, puede basarse en técnicas como el modelado de forma, dimensiones del cuerpo, agrupamiento de trayectorias y extracción de rasgos. De acuerdo al objetivo planteado se decidió realizar la clasificación de objetos por medio de la técnica de extracción de rasgos.

Una vez finalizado el proceso de segmentación de la región de interés, se aplica la etapa de extracción de características de los objetos de interés, dicha operación es un proceso muy importante para la operación de clasificación, ya que se identifican las características que distinguen a cada uno de los objetos que se desean reconocer en la secuencia de video.

A partir de cada objeto segmentado, se procede a extraer las características que guarden la información más relevante y que permitan tener un patrón común de diferencia entre ellos, ya que este patrón facilita la identificación del objeto en los siguientes *frames*, mientras este se desplace dentro del área de interés de la escena.

De acuerdo al objetivo del proyecto, de manera heurística se evaluaron las diferentes características de cada objeto y se determinó que el método más flexible pero no el más robusto para realizar el conteo y registro de los objetos, es realizar una discriminación por área; es decir, los perfiles de objetos cuya área de pixeles conectados con valor 1 mayores a 2000 pixeles son considerados transeúntes; este rango fue determinado de forma empírica y únicamente está en función de la altura de ubicación de la cámara, es un parámetro estático por lo tanto no representa una variable configurable.

A continuación se relacionan las características que se evaluaron para la clasificación de objetos en movimiento.

- *Area*. Es un escalar que representa el número de pixeles que pertenecen al objeto de interés.

- **BoundingBox.** Es un vector que contiene las coordenadas del rectángulo más pequeño que encierra al objeto de interés.
- **Centroid.** Es un vector que contiene las coordenadas del centro de masa del objeto de interés, donde el primer elemento del vector es la coordenada horizontal y el segundo elemento es la coordenada vertical.

6.6 SEGUIMIENTO DE OBJETOS EN MOVIMIENTO

El seguimiento de objetos en secuencias de video es un tema de investigación bastante amplio y complejo, debido al gran número de aplicaciones reales en el campo de la vigilancia inteligente, además de incursionar en otras áreas como la domótica, visión robótica y la visión robótica espacial.

El seguimiento de objetos pertenece al nivel intermedio de la visión, e involucra el encuentro de relaciones coherentes entre características de los objetos de interés en una serie de *frames* consecutivos con respecto al color, textura, posición y velocidad.

El análisis de la posición y la trayectoria de un peatón pueden determinar eventos, que impliquen un nivel de alerta, en términos de prevención o de asistencia. Por ejemplo, una anciana sola en su vivienda sufrió una fuerte caída, el sistema que realiza el seguimiento de la persona, dispara inmediatamente una señal de asistencia al centro de atención médica o le informa a un familiar para que este solicite la ayuda correspondiente.

Para la etapa de seguimiento de perfiles de objetos, específicamente personas se desarrolló un algoritmo recursivo y recurrente; donde se aprovechó el poder de cálculo y las bondades del procesamiento matricial que posee Matlab[®].

Actualmente existen dos líneas de investigación²⁶ para el seguimiento de personas en secuencias de video. La primera se basa en el seguimiento a partir de modelos y la segunda son técnicas recursivas donde la complejidad se reduce usando otro método.

La primera línea se subdivide en diferentes técnicas que a continuación se mencionan:

- Figura esquelética.

²⁶ IZQUIERDO GUERRA Walter y GARCIA REYES Edel. Seguimiento y conteo de personas en ambientes exteriores con una cámara fija. Reporte técnico Centro de aplicaciones de tecnología de avanzada. Diciembre 2009

- Contornos de dos dimensiones.
- Modelos Volumétricos.
- Modelos híbridos.
- Modelos de movimiento.

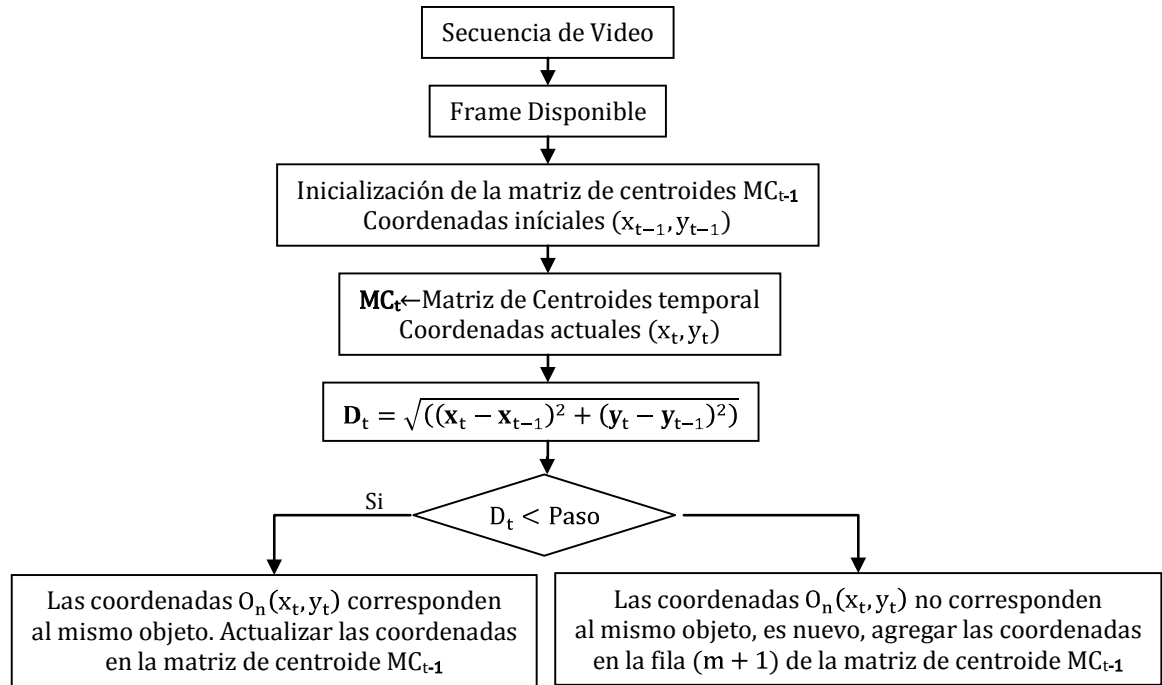
Las técnicas recursivas consideran que para el seguimiento no es necesaria la información estructural del objeto. Entre las técnicas más conocidas se tienen:

- Seguimiento basado en regiones.
- Seguimiento basado en contornos activos.
- Seguimiento basado en rasgos.

6.6.1 Método de la distancia euclidiana. El método que se utilizó para realizar el seguimiento de perfiles de objetos correspondientes a personas, se clasifica como una técnica recursiva con seguimiento basado en rasgos o patrones.

El centroide es una característica de los objetos que retorna las coordenadas del centro de masa o el momento de inercia del objeto. El algoritmo se basa en la teoría de la distancia euclidiana para realizar el seguimiento a través de la secuencia de imágenes, utilizando las coordenadas del objeto en el *frame* anterior (x_{t-1}, y_{t-1}) y el *frame* presente (x_t, y_t) .

Figura 14. Diagrama de flujo para seguimiento de objetos



Se tiene una matriz MC_{t-1} de tamaño $m \times 2$, donde m corresponde al número de objetos que se encuentran en la escena y 2 columnas correspondientes a las coordenadas (x, y) respectivamente.

Se realiza el cálculo de la distancia euclidiana para cada objeto de la matriz MC_t con respecto a la matriz inicializada MC_{t-1} , donde MC_t es una matriz de centroides temporal. El resultado se almacena en la matriz de distancias D_t , luego se compara con un parámetro, llamado *Paso de cambio*, en unidades de pixel, si la distancia entre un punto y otro supera este cambio, se considera que las coordenadas corresponden a otro objeto o un objeto nuevo en la escena, entonces el nuevo objeto tendrá la posición $m + 1$ en la matriz MC_{t-1} , en caso contrario en la matriz MC_{t-1} se actualiza el objeto respectivo con las nuevas coordenadas del centroide.

La matriz MC_{t-1} se especifica de la siguiente manera:

$$MC_{t-1} = \begin{bmatrix} x_{01} & y_{01} \\ x_{02} & y_{02} \\ x_{03} & y_{03} \\ x_{04} & y_{04} \end{bmatrix}$$

posición de los objetos

La matriz MC_t es una matriz de posiciones aleatorias que contiene los centroides mientras son emparejados con la matriz MC_{t-1} , cuya distribución de objetos siempre será la misma.

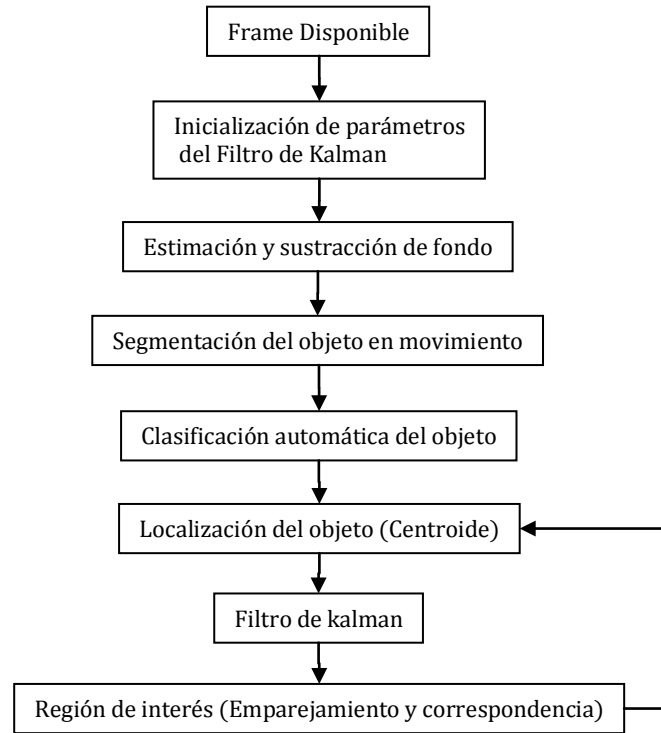
De esta forma, se realiza el seguimiento del objeto a través de la secuencia de video y se obtiene una matriz ordenada de posiciones para que en la etapa siguiente se realice el etiquetado y conteo de objetos.

6.6.2 Método del filtro de Kalman. Durante el proceso de investigación y desarrollo de este proyecto, se implementó el algoritmo del filtro de Kalman para el seguimiento de objetos, el cual se utiliza para estimar la posición de un punto o característica cuantitativa en movimiento, respecto a la incertidumbre de la medida.

Esta técnica es bastante eficiente para el seguimiento de un objeto, no obstante, cuando el número de objetos aumenta, la aplicación del filtro de Kalman se torna en un proceso mucho más complejo, ya que las características de la aplicación y los problemas que se presenta para el seguimiento, hacen que este no se modele como una función unimodal, sino multimodal, donde la complejidad de entrada y salida de peatones, originan un gran inconveniente en la inicialización y depuración del filtro para el posterior seguimiento de dichos objetos.

Para aplicar el filtro de Kalman al seguimiento de un objeto, es necesario proporcionar una característica representativa del mismo, que se establece como la observación del objeto. Para el cálculo de esta característica (centroide), la cual determina la posición del objeto, es necesario realizar un procesamiento previo a la imagen como se observa en el diagrama de flujo.

Figura 15. Diagrama de flujo para el filtro de Kalman



6.7 REGISTRO Y CONTEO DE OBJETOS

Para realizar la etapa final del sistema de monitoreo de perfiles de objetos se empleó una metodología basada en el cambio de etiqueta; El objetivo de esta etapa es discriminar la orientación del objeto de acuerdo al seguimiento del mismo.

El área de captura de la escena se divide en dos áreas iguales con etiquetas T1 y T2 respectivamente, y de acuerdo al seguimiento del objeto se discrimina su posición y se determina el área en la que se encuentra dicho objeto; una vez ubicado e identificado, se le asigna el valor correspondiente de etiqueta, en un vector columna de etiquetas (VCT_{t-1}).

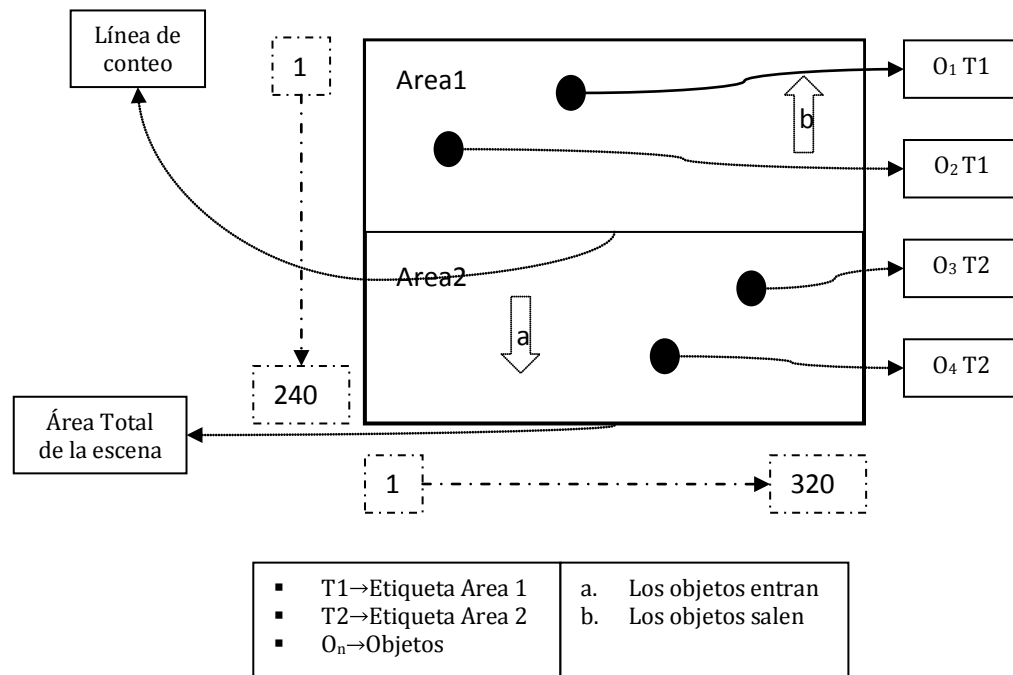
$$VCT_{t-1} = \begin{bmatrix} T1 \\ T2 \\ T1 \\ T2 \end{bmatrix}$$

En el análisis del siguiente *frame*, la etiqueta se corrige o se mantiene de acuerdo a las nuevas coordenadas de posición del objeto y se verifica si el valor de la etiqueta cambia o se mantiene.

Si el valor de la etiqueta cambia, significa que el objeto a cruzado la línea de conteo, entonces se produce el conteo del evento correspondiente de acuerdo a la interpretación de la orientación del movimiento de dicho objeto.

Si el valor de la etiqueta no cambia, significa que el objeto se mantiene en la misma región; el sistema descarta el conteo para este objeto y continúa con el proceso para el siguiente objeto rastreado, como se ilustra en la Figura 16.

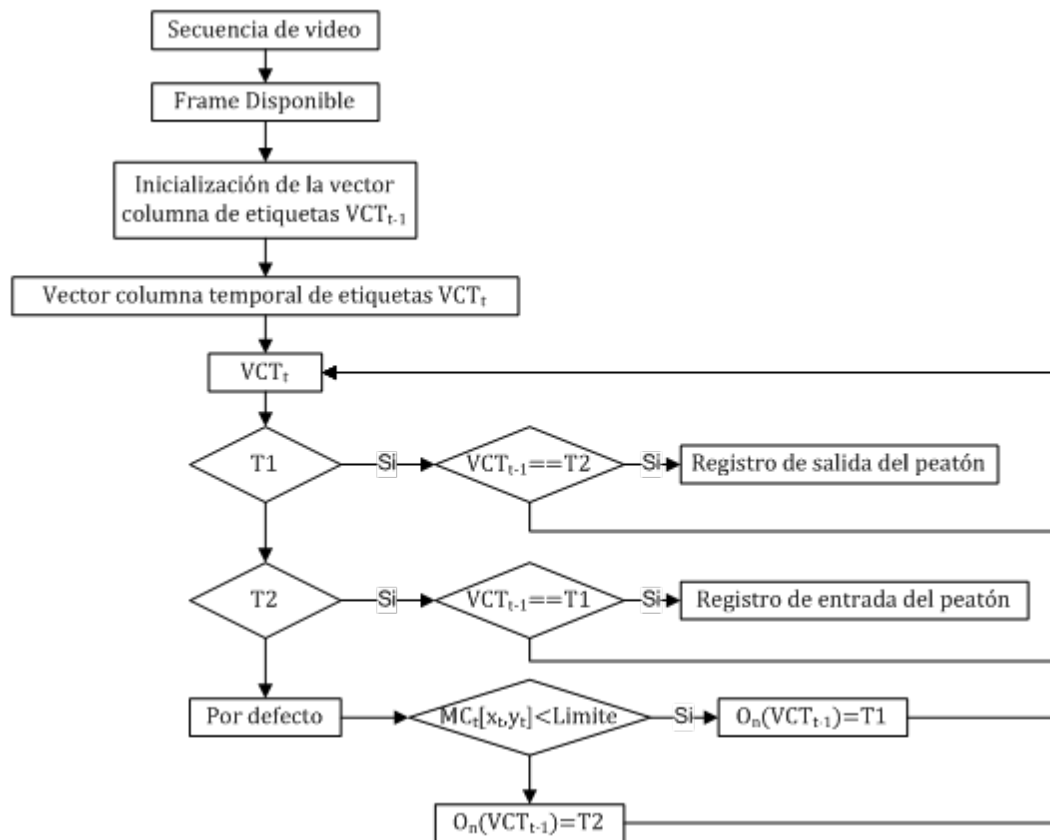
Figura 16. Registro y conteo de objetos



La lógica para el conteo del objeto de acuerdo al cambio de etiqueta es la siguiente:

- Si la etiqueta asociada al objeto O_n cambia de T1 a T2, el objeto se cuenta como evento de entrada.
- Si la etiqueta asociada al objeto O_n cambia de T2 a T1, el objeto se cuenta como evento de salida.

Figura 17. Diagrama de flujo de la etapa de conteo



7. PRUEBAS Y ANÁLISIS DE RESULTADOS

Las pruebas para los métodos de estimación y sustracción de fondo se realizaron utilizando el video de entrenamiento de Matlab[®]. Para la implementación final del sistema de monitoreo de perfiles de objetos; se escogió como escenario el pasillo del bloque de administración de la Universidad Surcolombiana, La cámara se ubicó a 6 metros de altura, con un área de captura de fondo de 30 metros².

Las aplicaciones en el campo de la visión artificial son complejas y puntuales; las bases de datos existentes sobre dicho campo, como aplicaciones de conteo; donde el área es una restricción muy importante que varía según la implementación, implica que no exista una base de datos estándar para el conteo de peatones, por ende, se optó por realizar videos al sistema trabajando en tiempo real para su análisis posterior.

7.1 ESTIMACIÓN Y SUSTRACCIÓN DE FONDO

En el desarrollo de este proyecto se implementaron y desarrollaron diferentes algoritmos en el lenguaje técnico de programación de Matlab[®], cuyos resultados se analizan a continuación, para dicho análisis se utilizó el video de entrenamiento de Matlab[®] (`vipmen.avi`), el cual simula los cambios más importantes en el procesamiento de video, como cambios bruscos de luz, camuflaje de objetos, mezcla de peatones. Este video representa la escena de dos peatones que presentan trayectorias opuestas y similitud de color con el fondo.

7.1.1 Método de diferencia de *frames* adyacentes. Las imágenes que se observan en las siguientes Figuras (18, 19, 20) se capturaron en diferentes instantes de tiempo en el que se detecta las bondades y desventajas de este método de modelado y sustracción de fondo.

Figura 18. Detección de objetos en movimiento utilizando la diferencia de *frame* adyacente (*Frame 36, 38,40*)



De acuerdo a los resultados se deduce que este método es bastante eficiente para la detección de objetos en movimiento; sin embargo, la segmentación de dicho objeto se condiciona al evento transitorio; es decir, si el objeto se detiene en algún lugar del área de captura, se produce el ocultamiento o pérdida del mismo dentro de la escena.

Figura 19. Ocultamiento del objeto en estado estacionario (*Frame 57 59 60*)



Es un método que funciona correctamente mientras se cumplan las condiciones de movimiento del objeto, condiciones normales de iluminación y tiempo de muestreo, sin embargo una gran desventaja es la sensibilidad del umbral de detección de movimiento, ya que si se escoge muy alto puede no detectar movimientos sutiles del objeto y si se escoge muy bajo, el ruido de la señal de

video produce falsas detecciones de movimiento. Es un método que no se adapta fácilmente a los cambios de iluminación y requiere de un post-procesado de la señal de video, como filtrado, suavizado y realce de los *frames* de video.

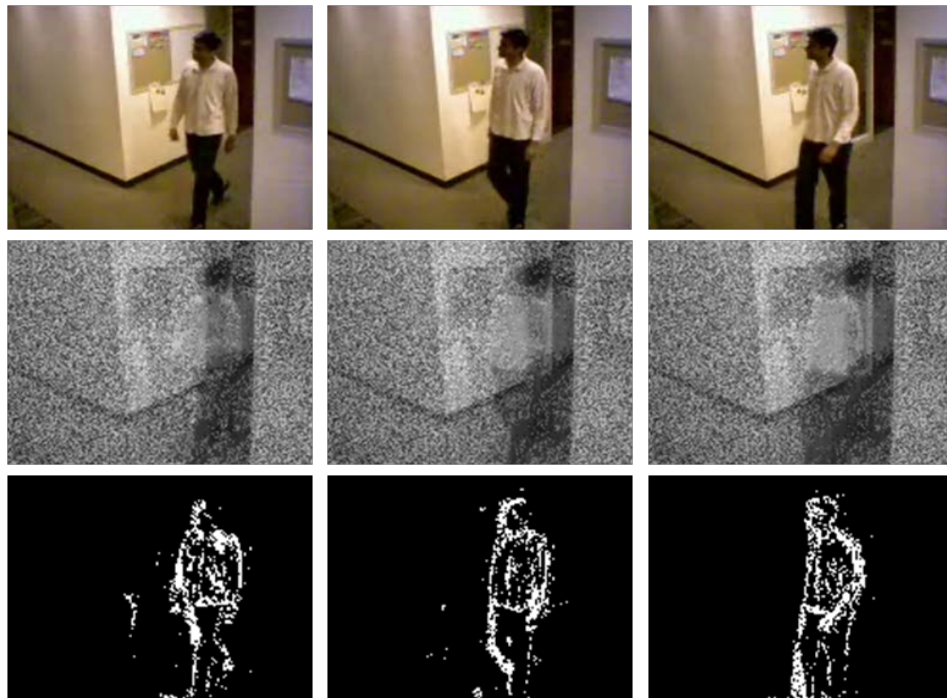
Figura 20. Detección inadecuada debido a cambios bruscos de luz y camuflaje (*Frame* 103, 105, 107)



Se consideró que este método es bastante eficiente en el procesamiento de imágenes en tiempo real, pero no es preciso y robusto para la aplicación que se desea implementar.

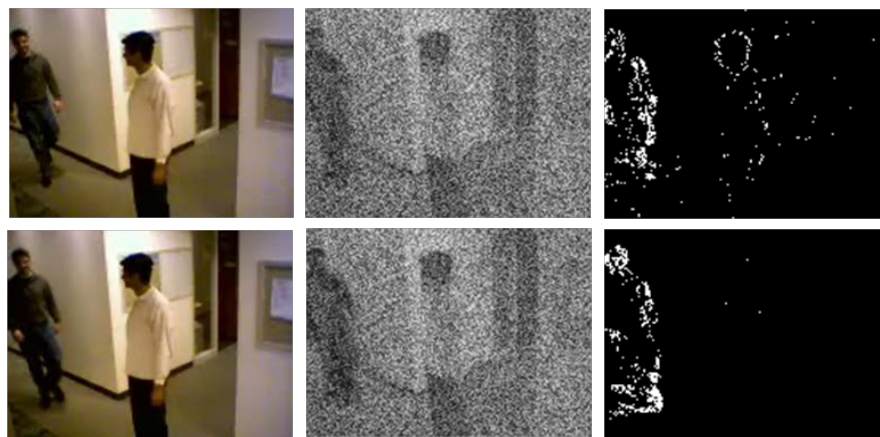
7.1.2 Método de múltiples gaussianas. Las imágenes que se observan en la Figura 21, representan una secuencia de video, donde se aprecia la secuencia de tres *frames* (fila 1), la estimación de fondo para cada instante de tiempo (Fila 2), detección y segmentación del perfil del objeto (Fila 3).

Figura 21. Estimación del modelo de fondo y extracción de los objetos en movimiento del primer plano (*Frame 33 36 39*)



Este método obtiene una estimación de modelo multimodal del fondo, sin embargo, requiere de una etapa de inicialización del número de componentes gaussianos por lo general de 3 a 5. Se trabajó con tres componentes gaussianos; además de otros parámetros como: el número de componentes de fondo, el umbral de desviación positivo, el coeficiente de aprendizaje, entre otros.

Figura 22. Actualización inadecuada del modelo de fondo (*Frame 101,103*)



El problema generado por la similitud de los objetos con el fondo y el estado estacionario del objeto, hacen que esta técnica no realice una estimación adecuada del modelo de fondo de la escena, si se observa en la Figura 22, el objeto de la derecha desaparece totalmente de la escena; es posible evitar este inconveniente con un módulo de post-procesado o con una reconfiguración de los parámetros de actualización de fondo con un porcentaje bajo de cada *frame*, respecto al fondo anterior y desviaciones altas, para no incorporar el objeto al fondo.

La Figura 23 representa la alta sensibilidad que tiene este método ante los cambios bruscos de luz, un inconveniente presente en la desarrollo de este trabajo de grado; además se requiere de una técnica computacionalmente sencilla, que es una desventaja de este algoritmo, ya que requiere de un alto costo computacional.

Figura 23. Alta sensibilidad del modelo de fondo debido a cambios bruscos de luz (*Frame 118*)



7.1.3 Filtro de mediana aproximado. La Figura 24 representa el método de filtro de mediana aproximado. El filtro de mediana aproximado es una técnica bastante conocida y utilizada para el procesamiento de video digital en aplicaciones como video-vigilancia y medicina; sin embargo, este método presenta ciertas desventajas como la actualización de fondo no selectiva, es decir, que todo objeto en la escena que se detenga por un tiempo superior al tiempo de actualización del fondo, el algoritmo lo reconoce como parte de éste, ocasionando el problema de ocultamiento o pérdida del objeto de interés en la máscara de detección de frente. Este efecto se observa en la Figura 24, el objeto va desapareciendo de la máscara de detección de frente a medida que se perpetúa en el fondo.

Figura 24. Estimación del modelo de fondo por el método de mediana aproximada (*Frame 95, 99*)



Este algoritmo es sensible a los cambios bruscos de luz y debido a ello se producen falsas detecciones y pérdida de los objetos; los resultados se observan en la Figura 25.

Figura 25. Estimación errónea del modelo de fondo debido a cambios bruscos de luz (*Frame 118, 120*)



En la siguiente figura se observan los inconvenientes que se producen cuando no existe una actualización selectiva del fondo, un objeto que pasa a ser parte de éste, cuando retoma su trayectoria, produce un cambio drástico en el modelo de fondo, originando en la máscara de detección de frente dos representaciones del mismo objeto, la primera perteneciente a la sombra dejada por el objeto

parcialmente capturado como fondo y la otra producida por el movimiento del objeto real.

Figura 26. Actualización inadecuada del modelo de fondo (*Frame 249, 253*)



La técnica que se analiza a continuación corrige este problema, fue una mejora realizada por los autores del proyecto, utilizando las ventajas que proporciona el cálculo matricial y la herramienta de procesamiento de imágenes de Matlab[®], se logra realizar una ventana de no actualización de fondo, que encierra al objeto detectado y esta evita que el objeto en condiciones estacionarias pase a ser parte del fondo. Fue una mejora excelente, ya que se logra realizar una segmentación óptima del objeto en condiciones de movimiento o viceversa.

7.1.4 Filtro de mediana aproximado selectivo. En la Figura 27 se observa la corrección del problema de ocultamiento del objeto, debido a la actualización selectiva del fondo, se determina que este método soluciona el inconveniente planteado. La estimación óptima del modelo de fondo y la detección de los objetos de frente, en condiciones de movimiento y no movimiento, sin embargo, aún persiste el problema de adaptación del fondo a cambios bruscos de luz, como se observa en la Figura 28; este inconveniente se resuelve parcialmente con la técnica de normalización de luminancia, pero no es aplicable para el sistema en tiempo real.

Figura 27. Estimación del modelo de fondo con actualización selectiva de regiones (*Frame 95, 99*)



En la Figura 28, se observan los efectos producidos por cambios bruscos de luz, no obstante, este método responde bastante rápido a los cambios leves de iluminación del ambiente. Por esta razón se decidió restringir el entorno de la aplicación, es decir, el sistema se implementa en un entorno controlado.

Figura 28. Inconvenientes de adaptación del modelo de fondo debido a cambios bruscos de iluminación (*Frame 118, 122*).



Esta técnica se seleccionó en la propuesta final del proyecto, debido a su bajo costo computacional y gran robustez en el procesamiento de imágenes en tiempo real.

7.1.5 Normalización de la luminancia para el filtro de mediana aproximado selectivo. En la Figura 29, se observa los resultados de la mejora para la corrección de los cambios bruscos de luminancia. El cambio en la detección y segmentación de los objetos es evidente si se compara dichos resultados con los mostrados en las Figuras 23, 25, 28. No obstante, existe una desventaja al adicionar el bloque normalizador de luminancia, ya que este restringe los niveles de intensidad de los píxeles del *frame* de video con respecto a los niveles de intensidad de los píxeles del fondo, la diferencia absoluta entre los píxeles del objeto y los píxeles del fondo es muy pequeña que en ocasiones no supera el umbral para la detección de frente. Además, este método solo es aplicable cuando el nivel de luz se incrementa, en caso contrario el algoritmo no responde de forma satisfactoria, debido a que se debe reinicializar los parámetros iniciales, por ende, se menciona como informe de investigación y no fue aplicado en la propuesta final del trabajo de grado.

Figura 29. Corrección de los efectos generados por cambios bruscos de luz (*Frame* 118, 122)



7.2 SEGUIMIENTO DE OBJETOS EN MOVIMIENTO

Se implementaron dos métodos de seguimiento, el primero se basa en la correspondencia de centroides aplicando el criterio de distancia euclidiana y el segundo, el método de seguimiento de filtro de Kalman.

7.2.1 Método de distancia euclidiana. Este método no demanda un alto costo computacional y no es muy robusto. Su funcionamiento se basa en las distancias entre las posiciones de los centroides de aquellos objetos que tienen área mayor a

2000 píxeles para ser considerados como transeúntes. De forma tal que el bloque detector compara la distancia de la posición de un centroide en el *frame t* a la posición de los centroides del *frame t-1*, si hay alguna distancia que sea menor a un umbral de distancia de 40 píxeles es considerado como el mismo objeto, sino se reconoce como un objeto nuevo. El proceso se realiza para todos los centroides presentes en el *frame t*, una vez el bloque detector encuentra correspondencia de centroides asigna una etiqueta a dicho transeúnte, que es identificada mediante una marca de color.

El método no es muy robusto debido a su limitación en velocidad. Si un transeúnte pasa de forma rápida y su centroide actual no cumple la condición de emparejamiento al centroide en el *frame t-1*, no cumple el criterio de distancia, no es seguido y pierde su etiqueta. La técnica de seguimiento por distancia euclidiana se observa en la Sección 7.1, Figura 33.

7.2.2 Método del filtro de Kalman. En la Figura 30, se observa los resultados sobre el seguimiento del perfil de una persona a través de un pasillo; el identificador verde en forma de diamante corresponde a la observación (centroide calculado) y el identificador rojo en forma de cuadrado corresponde a la (estimación del centroide) realizado por el filtro de Kalman.

Figura 30. Seguimiento de un peatón utilizando el filtro de Kalman (*Frame 34, 36, 38*)



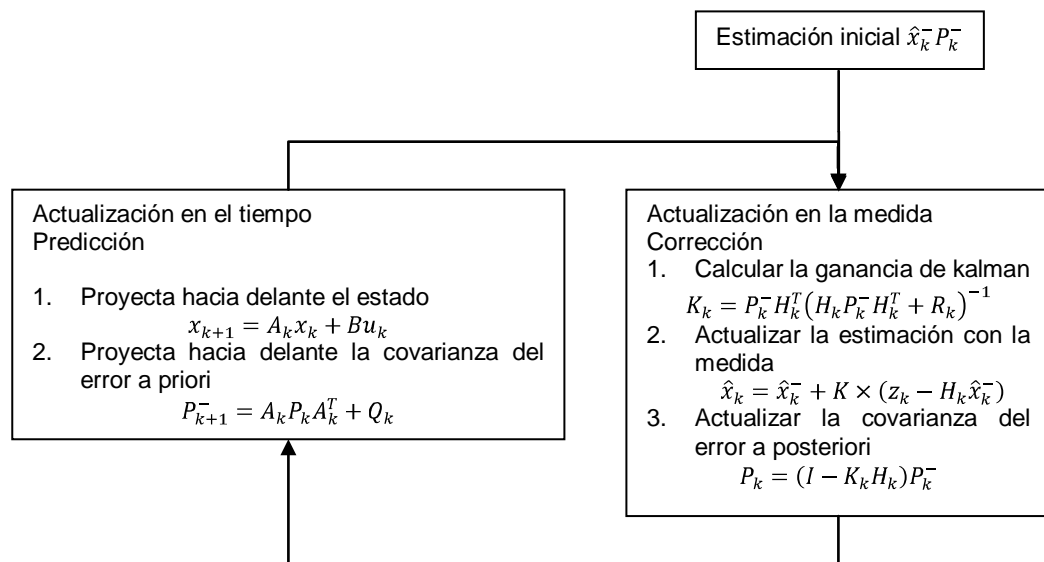
El objetivo de este método es predecir la aparición futura del objeto en la secuencia de video, en función de la medida anterior; es decir; de acuerdo a los estimadores iniciales, configurados con parámetros previamente definidos, debe predecir la estimación y la covarianza del error, auto ajustándose con cada nueva observación (centroide calculado). Esta técnica es bastante robusta para el

seguimiento de un solo objeto en imágenes 2D; en el desarrollo de esta propuesta logramos realizar el seguimiento en forma general de un solo objeto en 2D; pero no fue posible aplicar esta técnica a la propuesta final, porque el filtro de Kalman permite obtener la solución que minimiza el error de estimación, a través de la formalización matemática del estimador de estado clásico recursivo. La formulación que da lugar al filtro de Kalman pasa por la presunción de ciertas restricciones en lo que al modelo se refiere, como linealidad e incertidumbre al modelo, asociando ruido al estado y las medidas²⁷.

El filtro de Kalman, se considera como un estimador recursivo, el cual presenta las dos etapas propias de un estimador clásico:

- El proceso de predicción se desarrolla en el instante t-1, en el que se realiza una estimación a priori del vector de estado x_{k+1} y de la matriz de covarianza del error de estimación P_{k+1}^- .
- El proceso de corrección se desarrollada en el instante t, en el que a partir del nuevo vector de medida \hat{x}_k^- , de la ecuación de salida del modelo y de P_k^- , se calcula la matriz de estimación K_k que permite actualizar la estimación del vector de estado \hat{x}_k y la covarianza del error del error de estimación para ese instante P_k .

Figura 31. Diagrama de flujo del filtro de Kalman



²⁷ ROMERA MARRON, Martha. Seguimiento de Múltiples Objetos en Entornos Interiores muy Poblados Basado en la Combinación de Métodos Probabilísticos y Determinísticos. Tesis Doctoral, Universidad de Alcalá, Escuela Politécnica Superior, Departamento de Electrónica, España, 2008.

De acuerdo a la formulación del filtro de Kalman, la estimación proporcionada por el filtro \vec{x}_k presenta un error, que es una función de densidad de probabilidad normal y parametrizada; por tanto, el filtro de Kalman es un estimador unimodal y por ende, el filtro no permite resolver el problema de seguimiento de múltiples objetos con un único estimador, sino que es necesario utilizar un filtro por cada objeto a seguir; entonces se concluye que el filtro no es el algoritmo más adecuado para resolver el problema de seguimiento de peatones planteado en este trabajo de grado.

7.3 ANÁLISIS DE RENDIMIENTO Y CONFIABILIDAD DEL SISTEMA

Para cuantificar el rendimiento y confiabilidad de la propuesta del sistema final implementado se utilizó una matriz de confusión donde se definen las siguientes salidas del sistema.

- Falso positivo (FP): grupo de pixeles conectados que son detectados como el perfil de un objeto en movimiento, pero que dicho perfil no corresponde al de un peatón y son contados por el sistema.
- Falso negativo (FN): grupo de pixeles conectados que son detectados como el perfil de un objeto en movimiento, que corresponde al perfil de un peatón, y no es contado por el sistema.
- Verdadero positivo (VP): grupo de pixeles conectados que son detectados como el perfil de un objeto en movimiento, que corresponde al perfil de un peatón, y es contado por el sistema.
- Verdadero negativo (VN): grupo de pixeles no conectados que son detectados como el perfil de un objeto en movimiento por el sistema.

Tabla 2. Matriz de confusión para el conteo de peatones

	FP		FN		VP		VN		Real	
	Suben	Bajan	Suben	Bajan	Suben	Bajan	suben	Bajan	Suben	Bajan
Muestra 1		1	1		19	25			20	24
Muestra 2		4	3		12	26			15	22
Muestra 3			1	1	4	9			5	8
Muestra 4			2		16	17			18	17
Muestra 5				1	5	9			5	10
Muestra 6		2			13	13			11	13
Muestra 7			1		10	14			11	14
Muestra 8				1	12	19			12	20
Muestra 9		3	1		21	19			22	16
Muestra 10			4	2	17	15			21	17

Se analizaron 10 muestras de video, con un promedio de 31 peatones por cada muestra, las pruebas de monitoreo de perfiles se realizaron a una velocidad de muestreo de 30 FPS.

De acuerdo a la tabla de resultados se concluye que el sistema diseñado e implementado cumple en un 95%, con el objetivo planteado en este proyecto, si se compara los VP con el conteo real, se obtiene un error promedio del 5%. Se obtuvo un total de 10 FP, esto se debe a que el sistema se implementó en un entorno parcialmente controlado, donde no se tuvo en cuenta los focos oblicuos de luz artificial o ambiente, debido a ello, las sombras de los peatones ocasionan falsas detecciones, originando un doble conteo cuando los peatones cruzan la línea de conteo.

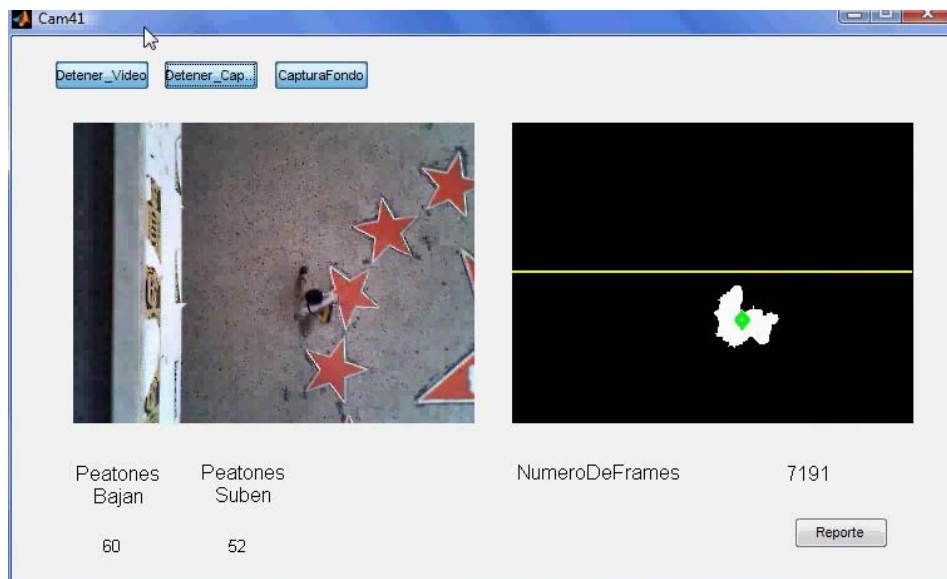
De acuerdo al total de FN, se observa que el sistema de monitoreo de perfiles falló en el conteo de ciertos peatones, debido a las siguientes causas:

- El peatón cruzó el área de captura a una velocidad no detectable para la etapa de seguimiento, ya que no cumplió con el umbral de distancia permitida para realizar un seguimiento y emparejamiento con el perfil de dicho peatón en el *frame* anterior.
- El número de peatones presentes en el área de captura es superior al número de peatones contables por el sistema, es decir, en un instante dado cruzaron 5 o más peatones y el sistema solamente registró 4 de ellos, ya que por defecto es el número máximo de conteo por *frame*.
- Los peatones se traslaparon antes de cruzar la línea de conteo y el sistema detectó el grupo de peatones como un solo peatón.

En las siguientes figuras, se relaciona la ventana de presentación del software, ejecutándose en tiempo real y particularidades del sistema de monitoreo de perfiles de objetos.

En la Figura 32 se observa el *frame* 7191, en donde la línea de color amarillo representa el límite de conteo. Los transeúntes son identificados por medio de etiquetas de color. La marca de color verde indica que a dicho transeúnte se le está realizando seguimiento. En la figura se evidencia una óptima segmentación del peatón a pesar de que tiene texturas similares a las del fondo, una condición importante del objeto es que mantiene un movimiento constante y permite que se cumpla los tiempos de procesamiento en el sistema para cada uno de los *frames* que componen la escena.

Figura 32. Ejecución del sistema de conteo con segmentación de un transeúnte



En la Figura 33, se observa el seguimiento que se realiza a los dos transeúntes y su etiquetado correspondiente, observando una buena segmentación de los objetos en la escena. Los peatones carecen de formas definidas en el transcurso de la escena, debido a la compleja motricidad el cuerpo humano, por esta razón se optó por clasificar a los peatones por medio del criterio de área.

El criterio de área es un parámetro general que para esta aplicación en particular no es bastante robusto, debido a que el proceso de clasificación se centra en un solo patrón de referencia. En un futuro se podría modificar este criterio de clasificación por uno mucho más complejo como la clasificación de peatones por un modelo 3D, reconocimiento de rostro y parámetros laterales (brazos-piernas-cabeza).

Figura 33. Software de conteo monitoreando dos transeúntes



En la Figura 34, se aprecia el inconveniente funcional más característico de esta aplicación, el traslape de transeúntes. Debido a la técnica de seguimiento el sistema no tiene la capacidad de discernir cuando se mezclan dos transeúntes y reconoce al grupo de peatones como uno solo, luego el sistema realiza el seguimiento del grupo de peatones y registra el conteo.

Figura 34. Ejecución del sistema de conteo con traslape de transeúntes



8. CONCLUSIONES Y RECOMENDACIONES

Para realizar el rastreo de perfiles de objetos en tiempo real, fue necesario diseñar un sistema con la capacidad de adaptación a los cambios de iluminación del entorno, con este fin, se investigó y se desarrolló diferentes algoritmos, cuyas características de flexibilidad, robustez y eficiencia se determinan de acuerdo a la aplicación que se desee diseñar.

En el transcurso de la investigación y desarrollo del proyecto, se realizaron pruebas para determinar cuál de las técnicas cumplía cabalmente con los requisitos de estimación y sustracción del modelo de fondo, para el análisis posterior de seguimiento, interpretación y reconocimiento como lo exige la aplicación. El método con el cual se obtuvo los mejores resultados fue el filtro de mediana aproximado; sin embargo, fue necesario modificar la etapa de actualización de fondo, por una etapa de actualización selectiva; consiste en actualizar sólo los píxeles clasificados como fondo (píxeles de valor 0 en la máscara de extracción de frente).

El método de diferencia de *frame* adyacentes, es la técnica más rápida para la detección de movimiento, se adapta rápidamente a entornos con cambios de iluminación leves, pero presenta ciertos inconvenientes respecto a las condiciones de movimiento del objeto; por ejemplo, si el objeto se detiene dentro del área de captura de la cámara, este desaparece de la máscara de extracción; además se obtiene una máscara de detección de frente basada en contornos, lo cual implica que se debe realizar un post-procesamiento para realizar la unión y relleno de contornos y así, obtener una buena segmentación.

El filtro de mediana aproximado es un método de bajo costo computacional, con la ventaja de que se adapta a ciertos cambios de iluminación del entorno como los que causa la luz ambiente. Los problemas de la adaptación de fondo con este algoritmo, son la inclusión de objetos al fondo (*background*), cuando pierde su movilidad y las falsas detecciones, debido a las sombras que se producen mientras el objeto se desplaza, bajo un foco de luz oblicuo al plano del fondo. Este problema se soluciona en entornos parcialmente controlados modificando la sub-etapa de actualización del filtro, otorgando un criterio de actualización selectivo.

Se desarrollaron otros algoritmos con mayor costo computacional, como se observa en el capítulo de descripción funcional del sistema, pero los resultados no fueron los mejores, el método de múltiples gaussianas presenta inconvenientes para la detección de objetos de frente, cuando la textura y luminancia de este tiene similitud con el fondo o cuando éste, inicialmente en movimiento se detiene; es muy poco robusto a cambios repentinos de iluminación. Esta técnica es más utilizada para estimar el modelo de fondo con características multimodales, donde existe un compromiso entre objetos en movimiento constante y objetos cuyas

partes tienen movimientos transitorios dentro de la escena como las hojas de los árboles.

Se diseñó un algoritmo que solucionara el problema de cambios de iluminación que fuera común a todas las técnicas investigadas, se ideó una forma de normalizar la luminancia en los valores de los píxeles de cada *frame* de video; consiste en sumar los píxeles en el *frame* t ; si son mayores a una constante de normalización k , se obtiene una segunda constante k_1 y se suma a los píxeles en el fondo t que son mayores a la constante k , donde se obtiene una segunda constante k_2 , la división de k_1 entre k_2 permite normalizar el nivel de intensidad del *frame* t , filtrando los valores que son menores al cociente.

El bloque de normalización de luminancia tiene la desventaja de hacer cero aquellos píxeles que tienen un alto nivel de intensidad en el *frame* t que pueden pertenecer a un objeto en movimiento; si este bloque se aplica antes del bloque de estimación y sustracción de fondo, ocasiona que disminuya el número de píxeles detectados que pertenecen a objetos en movimiento.

El método de seguimiento de Kalman, es bastante eficiente para el seguimiento de objetos en condiciones de oclusión, debido a su propiedad de estimador recursivo, el cual permite predecir la posición futura del objeto, reduciendo al máximo el área de búsqueda del objeto dentro de la escena; sin embargo, es una técnica con una gran desventaja debido a su formulación matemática de estimador clásico, lo hace aplicable solamente a funciones unimodales.

La propuesta del sistema final realiza el seguimiento de hasta cuatro objetos presentes en la escena, pero puede modificarse para ampliar el conteo a un número mayor de objetos y la utilización de técnicas más robustas. El método de seguimiento basado en la técnica de distancia euclidiana se limita por la velocidad del peatón. Si este se desplaza de forma rápida y la distancia de la posición del centroide del *frame* t al *frame* $t-1$ es mayor al umbral de distancia permitido, se considera como un objeto nuevo y el sistema no realiza el seguimiento.

El sistema obtiene los centroides de los objetos que cumplen con el requisito de área, para ser consideradas como transeúntes, Las áreas que no cumplen con el perfil de peatón son descartadas por la etapa de clasificación. El proceso de conteo consiste en una línea que indica la división que debe sobrepasar el peatón para ser contado.

El sistema de clasificación de peatones requiere ciertos ajustes. De acuerdo al estado del arte, los peatones carecen de un perfil definido y el criterio de área aplicado en este proyecto no es el más robusto, pero sí el más generalizado en cuanto al monitoreo de perfiles de objetos en movimiento.

Se propone para estudios futuros técnicas de corrección de sombras, técnicas que corrijan el efecto del camuflaje, ya que el proceso de segmentación por sustracción de fondo es afectado en gran medida por estos dos factores. Hay que evidenciar también que este trabajo de grado se constituye como recurso bibliográfico para aplicaciones mucho más específicas en el campo de videovigilancia y de videoseguridad. También se propone como estudio inicial para aplicaciones relacionadas a la visión robótica, enfocada en el desarrollo de robots de servicio, robots guía para museos, robots contadores de mercancías, entre otros. Es importante destacar que se abordaron temas de segundo nivel de complejidad en el campo de la visión computacional, lográndose obtener resultados satisfactorios de acuerdo a los objetivos de la propuesta inicial del proyecto.

BIBLIOGRAFÍA

ABRIL CAÑAS, Mauricio, VALENCIA, Mauricio, RESTREPO, Bonnie Johana y HOLGUÍN Germán Andrés. Sistema de visión artificial para el registro de densidad peatonal en tiempo real. Universidad Tecnológica de Pereira. 2007.

CHEUNG, Sen-Ching y KAMATH, Chandrika. Robust techniques for background subtraction in urban traffic video. Electronic Images Visual Communications 2004.

COLLAZOS CARRERA, Antonio. Detección de objetos abandonados en estancias controladas. Universidad Rey Juan Carlos 2009.

DUQUE QUINTANA, Juan Carlos y MENDIETA BORJA, Milton Giovanni. Síntesis animada de voz visual a partir del video de un rostro. Universidad Javeriana. 2005.

DURUCAN, Emrullah y EBRAHIMI, Touradj. Robust and illumination invariant change detection based on linear dependence for surveillance applications. EUSIPCO. 2000.

GONZALES JIMÉNEZ, Javier. Visión por Computador. 2000.

GONZALES, Rafael. WOODS, Richard y EDDINS, Steven. Digital Image Processing Using Matlab®. New jersey. Pearson Practice-Hall. 2004.

HERRERO MARTIN, Sonsoles. Análisis comparativo de técnicas de segmentación de secuencias de video basadas en el modelado de fondo. Universidad autónoma de Madrid. 2009.

JAIN, Ramesh. KASTURI, Rangachar y SCHUNK, Brian. Machine Vision. McGRAW-HILL. 1995.

KENTARO, Toyama y BARRY, Brumitt. Wallflower: Principles and Practice of Background Maintenance. Conference on computer vision. 1999.

MARRÓN ROMERO, Martha. Seguimiento de Múltiples Objetos en Entornos Interiores muy Poblados Basado en la Combinación de Métodos Probabilísticos y Determinísticos. Tesis Doctoral, Universidad de Alcalá. Escuela Politécnica Superior. Departamento de Electrónica. 2008.

RAIGAL, Eva Burgueño. Aplicación del filtro de partículas al seguimiento de objetos en secuencias de imágenes. Universidad Rey Juan Carlos. 2003.

RODRÍGUEZ MUÑOZ, patricia. Aplicación al filtro de Kalman al seguimiento de objetos en secuencias de imágenes. Universidad Rey Juan Carlos. 2003.

SIEBEL Nils. Design and implementation of people tracking algorithms for visual surveillance applications. The University of reading. 2003.

The MathWorks, Inc. Image acquisition Toolbox™ V3.1, User's Guide, Image processing Toolbox™ V6.1, User's Guide, Matlab® Started Guide, Creating Graphical User Interfaces, User's Guide.

VÉLEZ José. Francisco, MORENO Ana Belén, SÁNCHEZ Ángel y SÁNCHEZ MARÍN José Luis Esteban. Visión por computador. Universidad Rey Juan Carlos. 2º Edición. 2003.

ANEXO A. Especificaciones técnicas de las cámaras web utilizadas

Genius EYE 320

Tipo de dispositivo: Cámara web

Anchura: 5.1 cm

Profundidad: 3.9 cm

Altura: 4.3 cm

Peso: 69 g

Formato vídeo digital: WMV

Captura de vídeo: 640 x 480 a 30 frames por segundo (fps)

Imagen fija: JPEG - 640 x 480, 320 x 240, 352 x 288, 176 x 144, 160 x 120

Sensor de imagen tipo: CMOS

Ajuste de foco: Manual

Interfaz de ordenador: Hi-Speed USB

Interfaces: 1 x Hi-Speed USB - 4 PIN USB tipo A

Requerimientos del sistema

Software incluido: Controladores y utilidades, Genius CrazyTalk Cam Suite

Sistema operativo requerido: Microsoft Windows XP SP2, Microsoft Windows Vista, Apple MacOS X 10.4.8 ó posterior

Dispositivos periféricos / interfaz: CD-ROM

Windows XP SP2 - Pentium 4 - 1.6 GHz - RAM 256 MB - HD 260 MB

Windows Vista - Pentium 4 - 1.6 GHz - RAM 256 MB - HD 260 MB

MacOS X 10.4.8 o posterior - Core 2 Duo - 2.33 GHz - RAM 256 MB - HD 260 MB

Genius LOOK 110

Sensor de imagen tipo: CMOS de 100kpixeles

Ajuste de foco: manual

Interface: compatible con USB1.1/1.0

Formato video digital: JPEG/WMV

Tamaño de imagen: 640 x 480 (con software de interpolacion), 352 x 288, 320 x 240, 176 x 144, 160 x 120 pixeles

Video resolucion: 352 x 288 pixeles a 30fps

640 x 480 pixeles a 15fps

Requerimientos del sistema

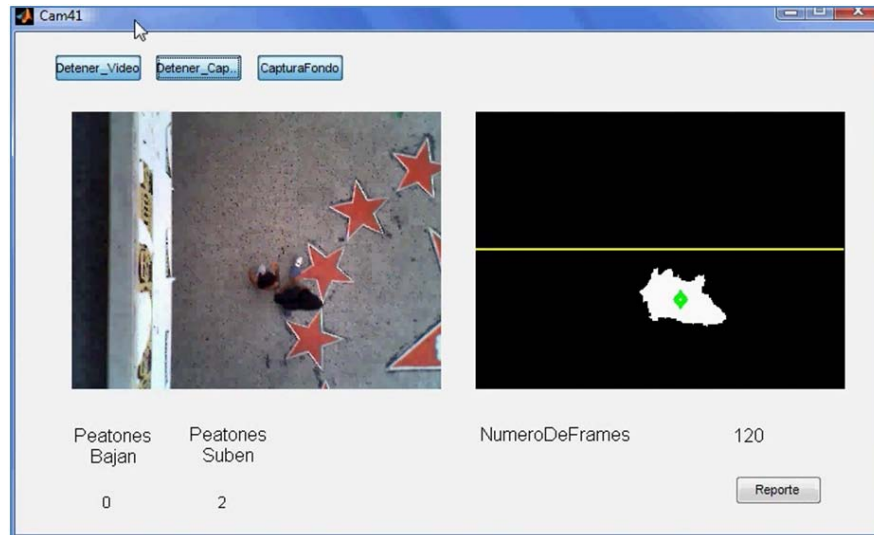
Procesador Pentium 4 1Ghz o superior

Memoria RAM de 256MB o superior

Tarjeta de video con mínimo 16MB de RAM

Windows Vista, XP, 2000, Me, 98SE

ANEXO B. Descripción de la interfaz gráfica de usuario del software de monitoreo de perfiles de objetos utilizando visión artificial.



- Botón (Iniciar_Video/Detener_Video): inicia y detiene la visualización previa de video a través del objeto de video de Matlab[®], como se observa en la ventana izquierda del guide.
- Botón (Iniciar_Captura/Detener_Captura): Controla los tiempos de captura de *frames*, para el análisis de conteo de perfiles de objetos, permite inicializar la captura del *frame* para una estimación inicial del fondo de la escena.
- Botón (CapturaFondo): Es un botón habilitado para renovar la captura de fondo a voluntad del usuario del sistema, cuyo fin es reducir el tiempo de muestreo y estabilización del fondo inicial de la escena.
- Botón (Reporte): este botón genera un informe en formato pdf del conteo que se está realizando en ese instante, con la fecha y hora en la que se genera.
- Las etiquetas (Peatones Bajan, Peatones Suben y NumeroDeframes) indican la información sobre el conteo de peatones y la cantidad de *frames* de video muestreados y procesados.
- La ventana de la derecha visualiza los resultados de segmentación de los perfiles de objetos que se están monitoreando y muestra la identificación y seguimiento de los perfiles de objetos que ingresan y salen de la escena.