

DESARROLLO DE UN SISTEMA DE VIGILANCIA PORTABLE

MANUEL ALEJANDRO JIMÉNEZ VIVAS

**FUNDACIÓN UNIVERSITARIA SAN MARTÍN
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA ELECTRÓNICA Y TELECOMUNICACIONES
BOGOTÁ
2009 II**

DISEÑO DE UN SISTEMA DE VIGILANCIA PORTABLE

MANUEL ALEJANDRO JIMÉNEZ VIVAS
032008
alejoman1@gmail.com

MONOGRAFÍA DE GRADO

ASESOR TÉCNICO
DIEGO DÍAZ MUÑOZ
DSP (DIGITAL SIGNAL PROCESSING)

FUNDACIÓN UNIVERSITARIA SAN MARTÍN
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA ELECTRÓNICA Y TELECOMUNICACIONES
BOGOTÁ
2009 II

Nota de aceptación

Diego Díaz Muñoz
Asesor

Jorge Arévalo
Jurado 1

Freud Romero
Jurado 2

Bogotá D.C. (Diciembre de 2009)

*A mi abuela y a mi tío.
"Aunque ya no estén a mi lado,
siempre los tendré presentes en mi corazón"*

AGRADECIMIENTOS

De antemano gracias a Dios por darme fortaleza y sabiduría para salir triunfante ante todas las adversidades y tropiezos que he encontrado en mi vida, sin dejarme desfallecer en la lucha por alcanzar mis sueños.

A la persona más valiosa, valiente y ejemplo de vida, mi Señora madre sin ti no tendría el don de la vida, gracias por estar presente en todos los momentos importantes de mi vida, gracias por tu apoyo y amor incondicional , gracias por exigirme a ser mejor cada día, gracias por ser la gran hacedora de este logro tan importante en vida personal y profesional, porque estoy seguro que de no ser por ti nada de esto hubiese sido posible, gracias.

Gracias a mis hermanos Camilo y Sebastián, por su apoyo, cariño y aguante durante este ciclo de mi vida, gracias por enseñarme el amor de hermanos y brindarme su amistad. Nunca los dejaré solos.

Gracias a mi Padre por sembrar en mi la semilla del conocimiento y apoyarme durante los primeros años carrera.

A mi Abuelita que me enseñó a ser perseverante, constante y recursivo para lograr las cosas que quiero, gracias por tu sabiduría para sortear las dificultades, gracias simplemente por enseñarme a sobrevivir, siempre estarás presente en mi vida.

Al mejor Tío, por mostrarme la otra cara de la vida, por siempre sonreír sin importar lo fuertes que fueran los problemas, gracias por enseñarme como levantarme después de cada tropiezo, gracias por ser mi Tío, mi Padre, mi Hermano mayor, mi Hermano menor y mi Amigo.

A mi mejor Amigo Harold Gil, por su apoyo en los momentos difíciles, por compartir en los mejores momentos de mi vida y por simplemente estar presente.

Al Ingeniero Diego Díaz, por toda la colaboración que me brindo para la realización de este proyecto, gracias por ayudarme a encontrar respuestas, sus conocimientos, su apoyo y animo en todo momento.

A Beto Cordoba, por su apoyo, consejos y recomendaciones durante la realización de este proyecto y los últimos semestres de mi carrera. Le deseo los mejores augurios en su vida profesional.

A Juancho y Eliana, por la colaboración brindada durante los últimos semestres de mi carrera.

A mi Alma Mater, por brindarme los conocimientos a través de sus ingenieros y profesores, durante este ciclo para poder llegar a la realización de este proyecto.

Y por último, todas las personas que directa ó indirectamente contribuyeron a culminación de este ciclo de mi vida.

CONTENIDO

	pág.
INTRODUCCIÓN	15
1. PROBLEMA	16
2. JUSTIFICACIÓN	18
3. OBJETIVOS	19
3.1 OBJETIVO GENERAL	19
3.2 OBJETIVOS ESPECÍFICOS	19
4. MARCO REFERENCIAL	20
4.1 ANTECEDENTES	20
4.2 MARCO CONCEPTUAL	21
4.2.1 Sistema de Vigilancia Basado en Video	21
4.2.2 Ejemplo Canal de Video	22
4.2.3 Fotograma	24
4.3 MARCO TEÓRICO	24
4.3.1 Modelo de Adquisición de Imagen	24
4.3.2 Súper Resolución	25
4.3.3 Detección de Movimiento	26
4.3.4 Principios físicos	30
4.3.5 Óptica del Ojo	31

4.3.6	Calidad Visual	33
4.4	ESTADO DEL ARTE	33
4.4.1	Componentes para la Tecnología de Vigilancia Inteligente	35
4.5	LIMITACIONES Y ALCANCES	36
5.	DISEÑO METODOLÓGICO	37
5.1	ETAPA 1 (Recopilación de Información)	37
5.2	ETAPA 2 (Determinar Resolución de los Fotogramas)	37
5.3	ETAPA 3 (Análisis Video)	37
5.4	ETAPA 4 (Determinar Frecuencia de Muestreo de los Fotogramas)	37
5.5	ETAPA 5 (Selección del Códec de Video)	38
5.6	ETAPA 6 (Evaluación de Desempeño y Determinación Requerimientos)	38
6.	DESARROLLO	39
6.1	FUNCIONAMIENTO DE LOS SISTEMAS DE VIGILANCIA USADOS EN LA ACTUALIDAD.	39
6.2	ALGORITMOS PARA EL RECONOCIMIENTO	39
6.2.1	Análisis de Componentes Principales	40
6.2.2	Análisis Lineal Discriminante (Linear Discriminant Analysis, LDA)	41
6.2.3	Estándares para el Reconocimiento	41
6.3	DETERMINAR LA VELOCIDAD DE LOS OBJETOS.	42
6.3.1	Proceso de Binarización de la Imagen.	43

6.3.2	Cálculo Centro de Masa	45
6.3.3	Cálculo Velocidad del Objeto	45
6.4	DETERMINAR FRECUENCIA DE MUESTREO DE LOS FOTOGRAMAS.	47
6.5	SELECCIÓN ALGORITMO PARA LA CAPTURA DE FOTOGRAMAS.	48
6.6	SELECCIÓN DEL CÓDEC DE REPRODUCCIÓN	50
6.7	REQUERIMIENTOS TÉCNICOS MÍNIMOS NECESARIOS	50
7.	PRUEBAS Y RESULTADOS	52
7.1	PRUEBA 1	52
7.2	PRUEBA 2	54
7.2.1	Binarización de los Fotogramas.	55
7.2.2	Cálculo de la Distancia Euclidiana y la Velocidad.	56
7.3	PRUEBA 3	57
7.3.1	Binarización de los Fotogramas.	58
7.3.2	Cálculo de la Distancia Euclidiana y la Velocidad.	59
7.4	PRUEBA 4	60
8.	CONCLUSIONES	62
9.	RECOMENDACIONES	64
	GLOSARIO	65
	BIBLIOGRAFÍA	66

LISTA DE TABLAS

	pág.
Tabla 1. Estándares de compresión de video.	48

LISTA DE FIGURAS

	pág.
Figura 1. Sistema de Video Cámara Simple.	21
Figura 2. Algoritmos de Compresión y Descompresión para JPEG.	23
Figura 3. Algoritmo de Compresión para MPEG.	23
Figura 5. Ilustración Modelo LR (Baja Resolución).	24
Figura 6. Ilustración de una imagen en HR.	26
Figura 7. Tareas principales para calcular las diferencias en la imagen.	27
Figura 8. Esquema general del modulo de actualización basado en Kalman.	28
Figura 9. Tareas principales del cálculo de diferencias entre imágenes.	30
Figura 10. El Ojo Humano.	32
Figura 11. Formación sobre la retina del tamaño y la resolución de la imagen.	34
Figura 12. Modelo de la arquitectura multiescala.	35
Figura 13. Análisis de video.	37
Figura 14. Proceso para determinar la frecuencia muestreo.	38
Figura 15. Eigenfaces estándar.	40
Figura 16. Ejemplo de seis clases usando LDA.	41
Figura 17. Diagrama del proceso para encontrar la velocidad de los objetos.	42
Figura 18. Proceso para binarización fotogramas.	43
Figura 19. (a) Fotograma capturado en t_0 , (b) Fotograma capturado en t .	43
Figura 20. Imagen resultado de la resta de los fotogramas t_0-t .	44
Figura 21. Imagen binarizada.	44
Figura 22. Desplazamiento centro de masa de un objeto.	45
Figura 23. Gráfica vector de la distancia euclidiana.	46
Figura 24. Espacio que capta la cámara.	46
Figura 25. Algoritmo para la captura de fotogramas.	49
Figura 26. Fotogramas extraídos para la Prueba 1.	52
Figura 27. Registro de errores en la Prueba 1.	53
Figura 28. Fotogramas extraídos para la Prueba 2.	54
Figura 29. Binarización de fotogramas.	55

Figura 30. Gráficas de los centros de masa.	56
Figura 31. Gráfica de la distancia recorrida por el objeto de interés.	57
Figura 32. Fotogramas extraídos de la secuencia de video para la Prueba 3.	58
Figura 33. Fotogramas binarizados para la Prueba 3.	59
Figura 34. Gráfica del centro de masa del objeto de interés.	59
Figura 35. Gráfica del vector de la distancia recorrida por el objeto de interés.	60
Figura 36. Zoom a un rostro ubicado a 10 metros de la cámara.	61

LISTA DE ANEXOS

Se anexan los códigos en matlab utilizados para el análisis de las imágenes documentadas en este proyecto.

RESUMEN

El presente proyecto estudia las principales características en la captura de video de los sistemas de vigilancia, de los cuales se selecciono e implemento un algoritmo de generación de video donde se modifica la frecuencia de muestreo, de tal manera que utilice la menor cantidad de recursos, posteriormente se evaluó el códec de video y se calcularon los requerimientos mínimos para su implementación en dispositivos portables.

Para la selección del algoritmo de video se realizaron las siguientes etapas:

Calculo de la resolución de los fotogramas. Una de las funciones principales del video de vigilancia es, aplicar software de reconocimiento de patrones y objetos incidentes en el escenario. La resolución se determinó a partir la cantidad de pixeles necesarios para el reconocimiento de objetos según lo proponen diferentes algoritmos de reconocimiento.

En la realización de este proyecto uno de los aspectos de mayor relevancia fue determinar la velocidad a la cual se desplazaban los objetos incidentes en el video, posteriormente con esta información se determinó la rata de muestreo adecuada de los fotogramas para obtener la mayor cantidad de información y no perder el registro de eventos importantes.

Selección de un algoritmo para mostrar el video. Para hacer una buena selección del algoritmo se tuvo en cuenta que los fotogramas no tuviesen ningún tipo de compresión en el momento de la generación del video y que este algoritmo fuera de fácil implementación en teléfonos celulares, debido a que estos son aspectos importantes dentro del desarrollo de este proyecto.

Evaluación del algoritmo seleccionado. Se realizaron pruebas al algoritmo seleccionado, para determinar los requerimientos técnicos mínimos de los dispositivos en los cuales pueda ser implementado el sistema propuesto. Estas pruebas se basaron en el tamaño de memoria requerido para el almacenamiento y la latencia del procesamiento.

INTRODUCCIÓN

La seguridad ha ganado gran importancia durante la última década dentro de la sociedad, debido a la constante amenaza de grupos terroristas y grupos de delincuencia común, por tal motivo los sistemas de vigilancia se hacen cada vez más indispensables dentro de los entes gubernamentales, la industria y las personas del común. Los sistemas de vigilancia modernos no solo deben cumplir con la función de registrar sucesos dentro de una secuencia entendible de video, sino también estar en la capacidad de contar con ciertas exigencias adicionales, como analizar las imágenes capturas, el uso de técnicas de compresión y transmisión eficiente de las imágenes, esto con el fin de incrementar la rapidez de alertar a las autoridades si llegase a existir algún peligro que atentase contra la seguridad y/o la integridad.

En los sistemas de vigilancia visual modernos es indispensable que los sucesos registrados sean grabados en imágenes de alta calidad, para que estas imágenes adquieran gran veracidad y credibilidad ante la sociedad y las cortes de justicia, en el momento de hacer una identificación ágil y certera de los objetos, rostros y movimientos de las personas incidentes en la secuencia de video grabada. Esto hará que la sociedad tenga más confianza en los entes gubernamentales y privados que prestan el servicio de brindar protección y con ello se aumentará el sentido de la seguridad para las personas que integran esta sociedad.

Este proyecto propone un sistema de vigilancia visual el cual comprime la secuencia de video capturada disminuyendo la rata de muestreo de los fotogramas y no haciendo compresión sobre estos últimos, conservando así la calidad visual de las imágenes y de esta manera pueda ser implementado en dispositivos portables, principalmente en teléfonos celulares.

1. PROBLEMA

En la actualidad los sistemas de seguridad se han vuelto indispensables no solo en lugares públicos si no como recurso en todas las empresas y hogares del mundo, y para cumplir con esta labor, es necesario que los sistemas de vigilancia sean cada vez más capaces de cumplir con las exigencias de la sociedad. Es por esta razón que los actuales sistemas de vigilancia basados en cámaras de video hacen énfasis en diversos parámetros entre los cuales se encuentran principalmente, el campo visual (el cual se enfoca en captar la mayor cantidad de área posible) y capturar la mayor cantidad de sucesos en tiempo, dejando en un segundo plano la calidad y resolución de la imagen, y tan solo los sistemas más modernos emplean técnicas para el análisis de la imagen y su posterior transmisión siendo esos sistemas los más costosos y difíciles de implementar.

En Colombia el no uso de sistemas de vigilancia que tengan como finalidad el análisis de la imagen ha permitido el incremento de la delincuencia común debido a que los videos tomados haciendo uso de tecnologías de video de baja resolución no permiten hacer un claro reconocimiento de los objetos y mucho menos identificación de rostros en las personas, ya que al hacer un zoom digital a la imagen esta carece de los pixeles necesarios para el reconocimiento de patrones.

Es por estas razones que se plantea la posibilidad de utilizar el teléfono celular, el cual en este momento no solamente cumple con la funcionalidad para hacer y recibir llamadas, sino que cada vez son más sofisticados en su sistema de multimedia, y vienen equipados con cámaras para fotografías con buena resolución, para ser utilizado como herramienta en un sistema de registro de video de alta calidad con el cual se pueda hacer análisis a la imagen para reconocimiento de patrones.

Expuesta la problemática y una posible solución, se hace indispensable saber lo necesario para llevar a cabo esta posible solución, y para ello se deben hacer evidentes ciertos cuestionamientos descritos a continuación:

¿Cuál es la tasa de fotogramas que se necesita para crear un video?

¿Son necesarias diferentes tasas de fotogramas para poder registrar ciertos eventos?

¿Cuál es la calidad necesaria para la imagen?

¿Cuál es la resolución que necesita la imagen?

¿Cómo hacer un archivo de video sin consumir muchos recursos de memoria?

¿Es necesario el audio para hacer el reconocimiento de una persona?

2. JUSTIFICACIÓN

La importancia de la seguridad ha tomado un valor crítico para la sociedad actual, la cual se siente bajo la constante amenaza de grupos terroristas y grupos de delincuencia común, es en este momento que surge la necesidad específica que sistemas de vigilancia cumplan con la función de poder monitorear actividades a distancia de lugares públicos y privados como lo son sistemas de transporte, bancos, centros comerciales, edificios gubernamentales, etc. Pero los sistemas de vigilancia modernos no solo tienen que tener la capacidad de cumplir con las funciones de capturar el campo visual y de registrar una secuencia entendible, sino que también deben tener la capacidad de emplear técnicas de análisis de imagen, técnicas para la compresión de la imagen y por último ser capaces de hacer una transmisión eficiente de la imagen. En tanto un sistema de vigilancia pueda cumplir con estos requerimientos, incrementará la facilidad y la rapidez de advertir a las autoridades si hay algún peligro que atente contra la seguridad o la integridad, y así mismo la veracidad de imágenes tomadas por los sistemas de vigilancia serán de mayor credibilidad ante la sociedad y las cortes de justicia ya que permitirán la identificación ágil y con certeza de los objetos y los rostros de las personas que están implicadas en las imágenes del video. Esto hará que la sociedad tenga más confianza en los entes gubernamentales y privados que prestan el servicio de brindar protección y con ello aumentará el sentido de la seguridad para las personas que integran esta sociedad.

El trabajar en escenarios con deficiencias en su estructura requiere el uso de robustos patrones de reconocimiento y métodos estadísticos. El uso de cámaras fijas las cuales pueden estar agrupadas en el área de interés ó dispersadas a través de todo el escenario requiere métodos automáticos para la compensación de las diferentes gamas cromáticas, la sincronización de las imágenes adquiridas y el registro del video con los marcos cartesianos de referencia locales.

Los avances tecnológicos también buscan la disminución en los costos de fabricación de los dispositivos y por supuesto los sistemas de vigilancia no son la excepción, por eso este proyecto plantea la posibilidad de usar el teléfono celular y toda la infraestructura ya implementada por las redes de telefonía móvil para poder brindar un servicio de sistemas de vigilancia basado en video con el cual se podrá hacer análisis de la imagen para la identificación de objetivos y se podrá transmitir de manera eficiente, sin tener la necesidad de crear ó utilizar redes privadas para la posterior transmisión a sitios remotos. Este proyecto hará un aprovechamiento máximo de los recursos que brinda la telefonía móvil para la construcción e implementación de un sistema de vigilancia con características similares a los sistemas robustos de vigilancia existentes los cuales no son fáciles de implementar y por el contrario presentan grandes costos en la infraestructura y en el tiempo de fabricación.

3. OBJETIVOS

3.1 OBJETIVO GENERAL

Diseñar y simular un sistema de vigilancia con una estructura basada en fotogramas.

3.2 OBJETIVOS ESPECÍFICOS

- Compilar y documentar información de los sistemas de vigilancia basados en video usados en la actualidad, su estructura y como se implementan.
- Determinar a qué tamaño y con cual resolución de los fotogramas, se puede realizar el reconocimiento de objetos.
- Determinar la velocidad máxima de los objetos dentro de la imagen con respecto a la frecuencia.
- Determinar la frecuencia más adecuada para capturar los fotogramas.
- Seleccionar un algoritmo para generación de video, que permita realizar las operaciones con un mínimo de recursos para hacerlo compatible con teléfonos celulares.
- Implementar un códec con el algoritmo seleccionado para convertir los fotogramas en video y realizar la evaluación de recursos y desempeño.
- Evaluar el algoritmo con una simulación, para determinar cuáles son los requerimientos técnicos mínimos de un teléfono celular donde pueda ser implementado el sistema diseñado.

4. MARCO REFERENCIAL

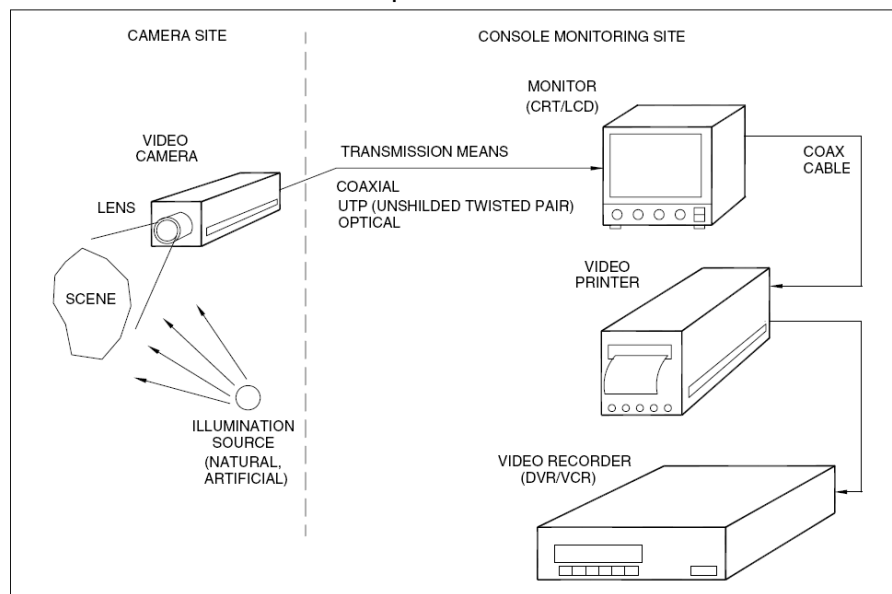
4.1 ANTECEDENTES

Durante toda la historia de la humanidad lo más importante para el ser humano ha sido su vida y la vida de sus seres queridos, al pasar el tiempo el valor de la propiedad también adquirió gran relevancia, es por esto que por muchos siglos se han desarrollado técnicas para proteger estos bienes de invasores y agresores que amenazan con quitarlos o destruirlos. Tanto en el pasado como en el presente las industrias y los gobiernos se han valido de contratar vigilantes para proteger sus instalaciones y prevenir posibles ataques. En la segunda guerra mundial el uso de personal para prestar guardia tuvo un considerable incremento y fueron empleados para resguardar de manera secreta industrias de manufactura de equipo militar e instalaciones gubernamentales y nacen varias empresas dedicadas a prestar este servicio especialmente en secreto (Herman, 2007).

Durante la década de los 60 con los avances en el campo de la electrónica permiten la incursión en el mercado de sistemas de alarma y sistemas de video, Radio Corporation of America (RCA), Motorola y General Electric fueron las compañías pioneras en construir cámaras de video con tubos al vacío para la industria de la seguridad. El uso de las cámaras de video creció rápidamente entre los años 60s y 70s debido al incremento de la confiabilidad, el bajo costo, y las mejoras tecnológicas en las cámaras. En la década de los 80s el continuo desarrollo conto con mejoras adicionales en las funciones y disponibilidad de accesorios para los sistemas seguridad de video, pero el mayor y más significativo avance en esa década fue la invención y posterior introducción de la cámara de video de estado sólido y para antes del inicio de los 90 la cámara de estado sólido (CCD) dispositivo acoplado por carga , con sensor de imagen era la mejor elección en las nuevas instalaciones de seguridad reemplazando rápidamente a las cámaras de tubo. En el pasado el sensor *Vidicon* de la cámara de tubo determinaba el rendimiento y la calidad visual alcanzable del sistema lo que lo convertía en el componente crítico del sistema de video. El tubo *Vidicon* era el enlace más débil en el sistema y presentaba desgaste con el uso. La complejidad y la variabilidad de la imagen en el tubo era de naturaleza análoga y daba menos seguridad para el análisis, que las imágenes obtenidas con tecnología de estado sólido. En comparación las cámaras de sensor CCD de estado sólido tienen más larga vida y son estables en todas las condiciones operativas. Otro de los factores de uso continuo y excesivo del video en los sistemas de seguridad fue el incremento de la capacidad de almacenamiento en los equipos, el bajo costo de fabricación de video cámaras de estado sólido, la disponibilidad de videograbadoras de casete económicas (VCR), la implementación de videograbadoras digitales (DVR) y el uso de computador personal (PC).

Ya para los comienzos de los 90s se aproximaba la integración de la tecnología computacional con la tecnología de video seguridad, ya que todos los componentes eran de estado sólido y la tecnología de video digital necesitaba memorias digitales a gran escala para manipular y almacenar grandes imágenes de video y la industria de la computación podía brindar lo necesario para este avance. Para conseguir una transmisión satisfactoria de la imagen de video la señal de video tuvo que ser comprimida para poder ser transmitida por la banda estrecha de las redes telefónicas existentes. A mitad de los 90s la industria de video por computadora ya tenía la compresión para la transmisión y contaba con los requerimientos para la industria y el gobierno, pero se necesitaba un medio de transmisión de mayor velocidad y de menor costo para transmitir las imágenes de video a locaciones mucho más lejanas, para entonces el Organismo de Proyectos de Investigación Avanzada de Defensa de los Estados Unidos (DARPA), la internet y la intranet proveen la habilidad de poder transmitir y recibir audio y video, comunicar y controlar datos de manera mundial en cualquier lugar, lo cual permitió la integración de los sistemas de seguridad basados en video con estos sistemas de comunicaciones (Kruegle, 2007).

Figura 1. Sistema de Video Cámara Simple.



(Kruegle, 2007)

4.2 MARCO CONCEPTUAL

4.2.1 Sistema de Vigilancia Basado en Video

Un sistema de vigilancia visual está constituido por una red de cámaras, este es el modelo típico usado por los CCTV (circuito cerrado de televisión), las cuales

emiten la imagen del campo visual continuamente casi durante las 24 horas del día los 7 días a la semana. Cada uno de los sistemas de vigilancia visual implementa un tipo de arquitectura específica, cuando estos sistemas se basan en arreglos de cámaras el tamaño de los datos a transmitir aumenta considerablemente, por esta razón los sistemas de vigilancia basados en video más sofisticados emplean cámaras móviles y esto permite el prestar más atención a otros parámetros de interés y no solo se limita a la transmisión de la imagen, entre estos parámetros se encuentran, proveer controles para adaptar al sistema a las condiciones del clima ó al tipo de iluminación y así lograr mejor registro de imagen en las escenas de interés. Las principales tareas de procesamiento que debe realizar un sistema de video son las que se nombran a continuación:

- Calibración de la cámara con respecto a un plano cartesiano de referencia local.
- Adquisición de la imagen de la escena.
- Adaptar un modelo del fondo de la escena.
- Detectar eventos en las regiones del primer plano.
- Registrar los sucesos.

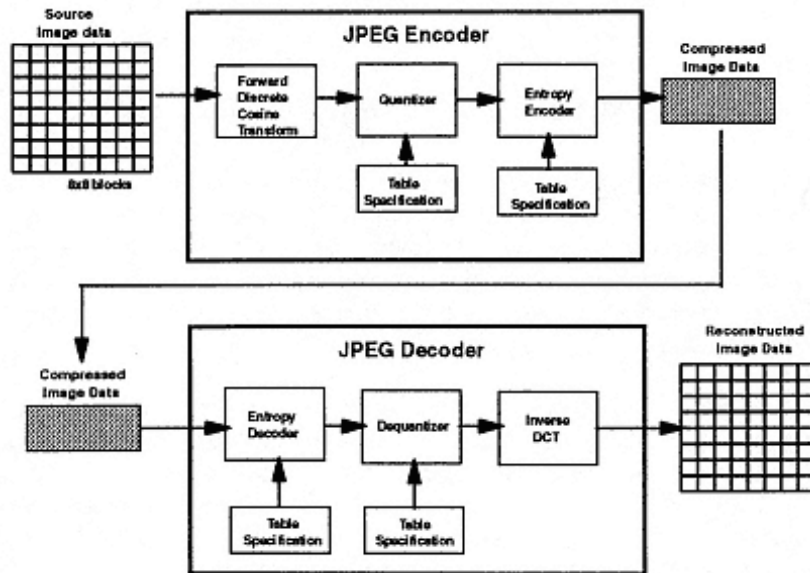
Todas las tareas anteriormente nombradas se relacionan entre sí, la calibración de la cámara y el registro de video pueden ser interpretadas como las herramientas para observación de los datos y para la tarea de conseguir automáticamente un procesamiento básico de la imagen para modelar el fondo y detectar eventos en primer plano (Chaudhuri & Taur, 2005).

4.2.2 Ejemplo Canal de Video

JPEG (Joint Photographic Experts Group), es un estándar de compresión de imagen, el cual se basa en la codificación de los coeficientes de información contenida en cada pixel de la imagen por medio de medio de la DCT (Transformada Discreta del Coseno), en bloques de 8x8 pixeles.

En la Figura 2, se muestra el diagrama de bloques de los algoritmos de compresión y descompresión del JPEG. Un simple fotograma es dividido en bloques de 8x8 los cuales se procesan independientemente. Cada bloque es transformado al espacio DCT, dando como resultado un bloque de 8x8 de los coeficientes de DCT. Estos coeficientes son cuantizados y convertidos en constantes. La constante de cuantificación para cada coeficiente DCT se elige para producir efectos de discontinuidades visuales mínimas, y el máximo reduce la entropía contenida en los coeficientes. Los coeficientes cifrados son entonces la entropía cifrada en una secuencia de datos (Westwater & Furht, 1997).

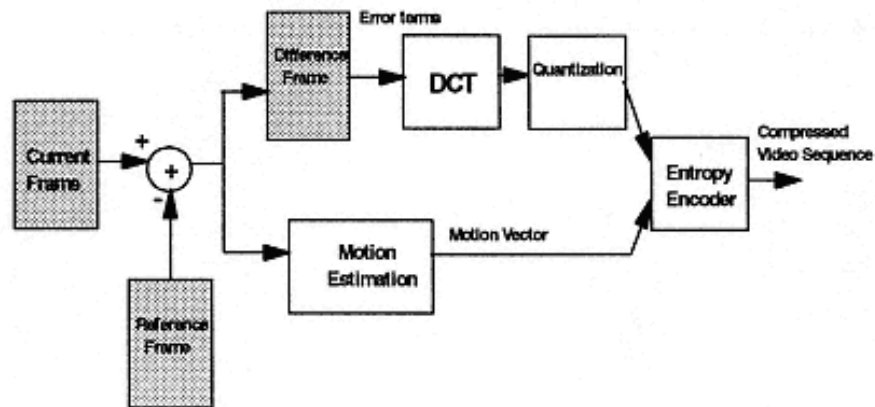
Figura 2. Algoritmos de Compresión y Descompresión para JPEG.



(Westwater & Furht, 1997)

Una secuencia de video se compone de varios fotogramas y al aplicar compresión JPEG para cada fotograma, se obtiene lo denominado M-PEG (JPEG en movimiento). Proporciona acceso aleatorio a cada fotograma individualmente. El estándar de compresión en movimiento MPEG es una extensión de la compresión DCT basada en imágenes móviles. El MPEG codifica los fotogramas haciendo un estimado de la diferencia entre del movimiento entre los fotogramas y codificando las diferencias en formato JPEG. La valoración del movimiento es de cálculo complejo, la cual requiere de equipo especializado para codificar lo cual agrega complejidad al algoritmo original del JPEG. En la Figura 3. Se ilustra el algoritmo de compresión MPEG para secuencia de fotogramas. (Westwater & Furht, 1997).

Figura 3. Algoritmo de Compresión para MPEG.



(Westwater & Furht, 1997)

4.2.3 Fotograma

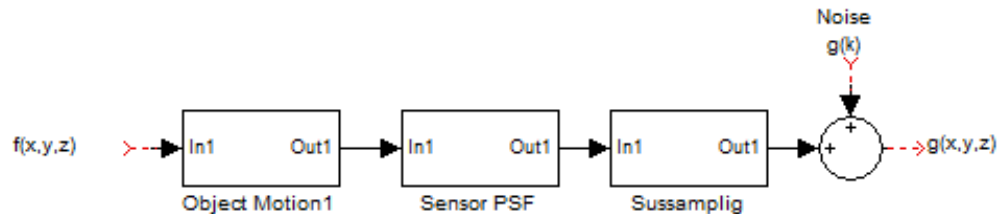
Fotograma son imágenes que al repetirse a una cierta frecuencia, generan en el cerebro la sensación de movimiento. Cada una de las que componentes una secuencia de video.

4.3 MARCO TEÓRICO

4.3.1 Modelo de Adquisición de Imagen

El modelo LR es sencillo y se puede aplicar para modelar secuencias de video incorporando información temporal.

Figura 4. Ilustración Modelo LR (Baja Resolución).



(Chaudhuri & Taur, 2005)

La señal que entra $f(x, y, k)$ denota que la imagen en HR se encuentra en plano focal del sistema de coordenadas (x, y) en un instante de tiempo kT , donde T hace relación a la duración de una muestra temporal del video, por ejemplo si un video contiene 10 fotogramas por segundo, entonces $T = 0.1seg$, asumiendo que la cámara de video esta fija. Un objeto se mueve dentro del campo visual de la cámara, está movimiento se puede representar por el desplazamiento de un vector y después el efecto de las dimensiones físicas del sensor LR es incorporado, este sensor es llamado función extendida de punto (PSF), en esta punto se asume que la cámara además de estar fija, tampoco esta desenfocada y no presenta manchas en movimiento, y también se asume que el obturador de la cámara tiene una duración el tiempo mucho menor que el tiempo entre cada fotograma, finalmente la imagen transformada pasa por una exploración LR seguido por un sensor de ruido. Por lo tanto de porción kth en HR (alta resolución) del fotograma f_{kt} se puede relacionar con la porción de kth en LR, se relacionan de la siguiente forma:

Ecuación 1

$$g_{kt} = AHf_{kt} + n_{kt}$$

Donde H es una matriz de filtros la cual contiene la dinámica del objeto. Teniendo ya definido el primer modelo para adquisición de la imagen se puede profundizar en el problema de la adquisición de la imagen en HR con cámara lenta. Al tener un número de muestras de LR $g_{kt}, k = 1, 2, \dots, n$ capturadas en baja resolución a razón de $1/T$, se quiere recuperar una porción en alta resolución de video HR a razón de $f_{m/s}^T, m = 1, 2, \dots, n$ donde s es un entero definido por factor de la muestra temporal, la muestra temporal es T (Chaudhuri & Taur, 2005).

4.3.2 Súper Resolución

Un algoritmo iterativo para poder construir fotogramas de HR (Alta Resolución) para una compresión de una secuencia video de LR (Baja Resolución), se propone para un fotograma kth teniendo en cuenta el estado anterior M_1 y para un estado siguiente M_2 , entonces el número de muestras será $g_i, i = k - M_1, \dots, k + M_2$. Usando el modelo de adquisición de imagen ya nombrado, el valor estimado de un fotograma en HR f_r ahora es denota por \hat{f}_k , y su demostración es la siguiente:

Ecuación 2

$$(\hat{f}_k, C)_{(k, k-i)} = \arg \min_{f_k, C_{k, k-i}} \left\{ \sum_{i=-M_2}^{M_1} \lambda_{k, k-i} \|AC_{k, k-i}f_k - g_{k-i}\|^2 + \lambda_1 \|Q_1 f_k\|^2 + \lambda_2 \|Q_2 f_k\|^2 \right\}$$

Donde A es la muestra de la parte baja de la matriz y $C_{k, k-i}$ representa el operador de compensación de movimiento entre un fotograma en HR f_k el fotograma anterior en HR f_{k-i} tal que

Ecuación 3

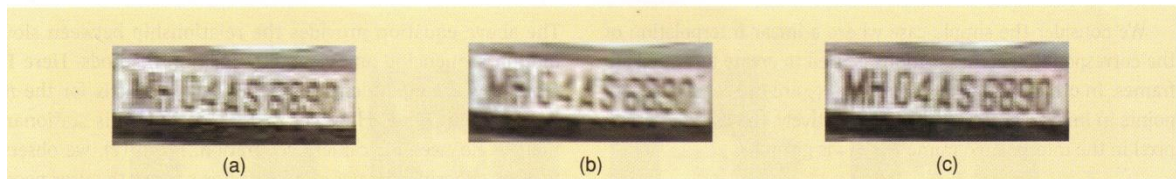
$$f_{k-i} = C_{k, k-i}f_k$$

Nótese que $C_{k, k-i}$ representa solo un fotograma HR y la densidad de movimiento de campo para cualquier fotograma kth . Por ejemplo el video grabado en MPEG solamente ofrece un vector de movimiento de un valor aproximado de de 16x16 bloques lo cual no sirve para propósitos de HR. Q_1 y Q_2 son operadores pasa altos que capturan entre bloque y bloque la suavidad de cada fotograma f_k . Aunque cuando se trabaja en descompresión de video el efecto de los bloques algunas veces es muy pronunciado en los fotogramas LR de las imágenes que son

capturadas a una baja tasa de bits. Ya que un fotograma de imagen en HR f_k y el campo de densidad de movimiento $C_{k,k-i}$ son de las observaciones de g_t , por lo tanto se plantean condiciones reguladas para poder resolver el problema. El operador Q_1 es el que exige una cierta cantidad de suavizado en la imagen en HR f_k , el operador Q_2 es definido entre los límites adyacentes de los bloques DCT. Por lo tanto este término es el que ayuda a la reconstrucción de los fotogramas en HR. Se considera que ambos operadores Q_1 y Q_2 son de estudio laplaciano. El término $\|AC_{k,k-i}f_k - g_{k-i}\|^2$ supervisa el cumplimiento de los datos, para los siguientes términos λ_1 , λ_2 y $\lambda_{k,k-i}$ son parámetros de regulación que controlan la suavización entre bloque y bloque y la exactitud de los datos. Obteniendo las condiciones iniciales para el campo de movimiento, se facilita resolver la ecuación para f_k , después se actualiza el vector de movimiento de HR $C_{k,k-i}$. Para obtener mejores resultados en la resolución de la imagen es necesario repetir este proceso y entre más veces se realice el proceso se obtendrán mejores valores para las estimaciones de f_k y de $C_{k,k-i}$.

Para aplicaciones en sistemas de vigilancia basados en video la expresión será resuelta solo para la región donde el objeto está en movimiento. Así se obtienen todos los fotogramas para una secuencia de video dada, que necesita ser interpolado temporalmente para correr un video en cámara lenta (Chaudhuri & Taur, 2005).

Figura 5. Ilustración de una imagen en HR.



(Chaudhuri & Taur, 2005)

4.3.3 Detección de Movimiento

En la actualidad han sido desarrollados gran variedad de métodos para la detección de cambios y se han incorporado en las cámaras, por ejemplo el método de diferenciación simple SD (Simple Difference), computa la diferencia total para cada instante de tiempo t $D_t(X, Y)$ entre la intensidad de los píxeles de dos imágenes, después se aproxima al umbral matemático TH para obtener la imagen de forma binaria $B(x, y)$. El método SD es de fácil y rápida implementación, pero también es muy sensible al ruido y a los cambios de iluminación, ya que podría afectar la escala de grises y prestarse para una mala interpretación de cambio en la estructura de los datos (Kapur, Sahoo, & Wong, 1985).

Para poder superar el problema de sensibilidad al ruido se crea el método de la derivada (DM), considerando una región de $n \times n$ píxeles en dos imágenes y se calcula la derivada del radio L_{ij} y se usa la definición de la varianza σ de las dos regiones R_i y R_j para obtener una imagen binaria como la que se describe a continuación:

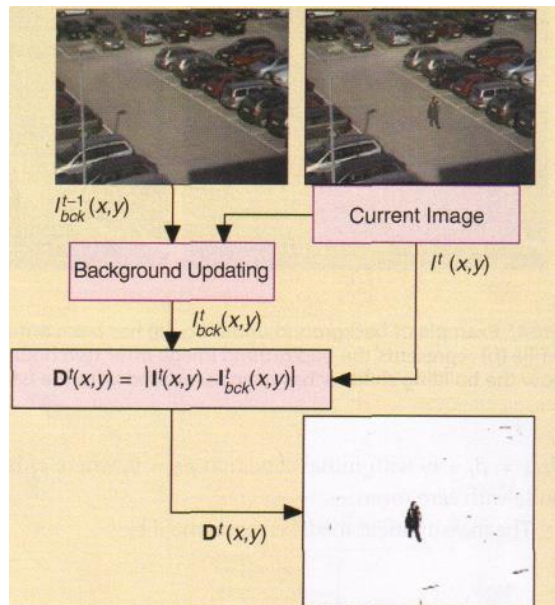
Ecuación 4

$$B(x,y) = \begin{cases} 0 & \text{si } L_{ij} < L_{Th} \\ 1 & \text{en otro caso} \end{cases} \quad \forall (x,y) \in R_i, R_j$$

Donde L_{Th} es el umbral.

El método de sombreado (SM) modela la intensidad que se da en un punto $I_p(x,y)$ como el producto de la incidencia de luz $I_i(x,y)$ y el coeficiente de sombreado S_p el cual es calculado para cualquier punto de acuerdo con el modelo de iluminación de Phong. Esto puede demostrar con facilidad si en una región a ocurrido algún cambio al hacer la comparación entre dos fotogramas consecutivos $I_{t-1}(x,y)$ y $I_t(x,y)$, esto es totalmente suficiente para poder calcular la varianza σ_i de la intensidad de los radios en cada región. Si la varianza σ_i se acerca a cero quiere decir que no ocurrieron cambios en el lugar (Foreti, et al., 2005).

Figura 6. Tareas principales para calcular las diferencias en la imagen de una cámara estática.



(Foreti, et al., 2005)

Los métodos DM y SM obtienen resultados similares. El nivel de ruido que afecta a la imagen generada por estos métodos es de menor intensidad que el nivel de

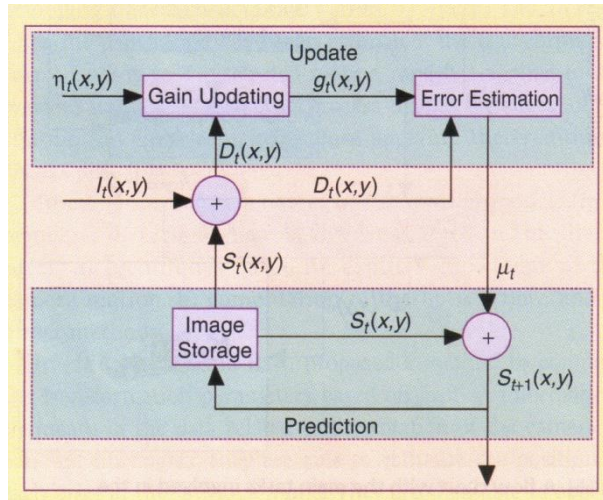
ruido que afecta a la imagen generada por el método SD, sin embargo lo detectado en relación a la forma, el lugar y el tamaño es de menor exactitud, por lo que el contorno del objeto es afectado significativamente y esto origina pérdidas parciales con respecto a la imagen original.

El método de LIG se basa en suponer que los píxeles que tienen una pendiente con varios niveles de grises que modelan el objeto y que todos los píxeles con niveles de grises similares hacen parte del mismo objeto.

La imagen es subdividida en varias imágenes de de $m \times m$, esto se hace para limitar los efectos de la iluminación en los cálculos. Cada sub-imagen de $m \times m$ se puede manipular como una imagen normal, y finalmente se aplica el proceso de umbral para aislar los píxeles del objeto del fondo. Este método presenta mejores resultados que los métodos anteriores (Parker, 2001).

Otro de los métodos que pretende minimizar errores en el proceso de detección de cambios es la de Actualización del Escenario el cual reduce los efectos del ruido, de la iluminación y de los cambios en el ambiente, este es el método usado por los sistemas avanzados de vigilancia basados en video, este modelo se basa en el filtro de Kalman, el proceso de Actualización del Fondo pretende determinar cambios importantes en la imagen usando cálculos aproximados en el escenario (Foreti, et al., 2005).

Figura 7. Esquema general del módulo de actualización basado en Kalman.



(Foreti, et al., 2005)

El módulo de actualización usa los niveles de grises $D_t(x, y)$ de la diferencia en las imágenes en un punto (x, y) y estima un valor de ruido $\eta_t(x, y)$ que afecta a la imagen entrante para poder actualizar la ganancia del filtro $g_t(x, y)$, luego estima un porcentaje error ε_t en base a la ganancia del filtro $g_t(x, y)$ y el valor de $\mu_t =$

$D_t(x, y) \cdot g_t(x, y)$. El modulo de predicción calcula el valor estimado del nivel de grises $S_t(x, y)$ en un punto (x, y) y lo compara con el valor de la siguiente imagen en el mismo punto (x, y) .

El modelo de multicapa ha sido usado en sistemas de vigilancia por que emplea la combinación de diferencias temporales con una plantilla agrupándolas para llevar a cabo un seguimiento del objeto, se modela para que cada vez que un objeto entra en el escenario quede estacionario por un predeterminado tiempo, esto se considera como una nueva capa la cual es añadida al escenario. La solución que plantea el modelo de multicapa se utiliza para resolver los problemas en cambio en el escenario como lo son huecos o fantasmas que se creaban en la imagen de fondo cuando el objeto comenzaba a moverse (Collins, Lipton, Fujiyoshi, & Kanade, 2001).

Este método también aborda el problema de los fondos multimodo que se generan en lugares al aire libre en donde los pixeles pueden cambiar de estado, para resolver esto se asume a cada pixel como una mezcla de distribuciones gaussianas y por tanto el valor probabilidad para cada pixel está dada por:

Ecuación 5

$$P(x_t) = \sum_{n=1}^k \omega_{i,t} * \eta \left(x_t, \mu_{i,t}, \sum_{i,t} \right)$$

Donde k es el numero de distribuciones, $\omega_{i,t}$, $\mu_{i,t}$ y $\sum_{i,t}$ son valores aproximados del tamaño, la media y la covarianza de la matriz de la mezcla gaussiana en un tiempo t y $*$ es el operador de la convolución, finalmente η es la función de densidad de probabilidad gaussiana que es igual a:

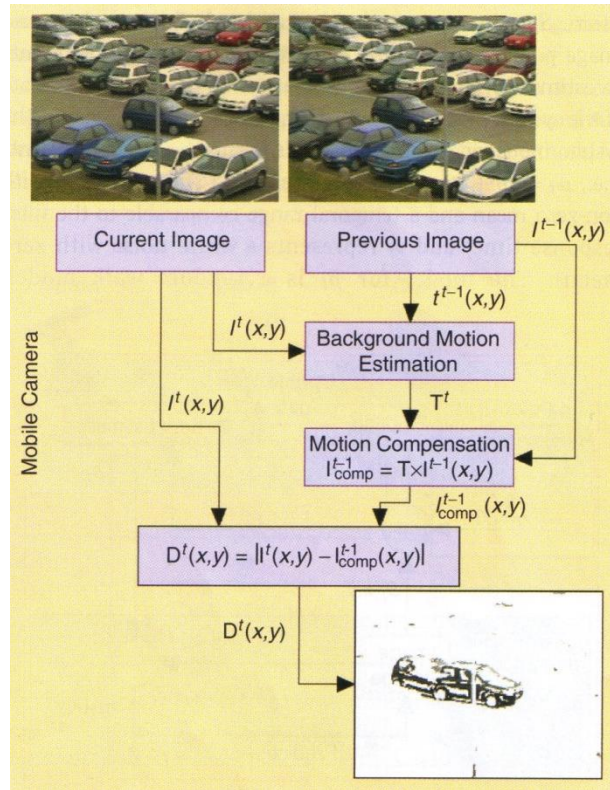
Ecuación 6

$$\eta(x_t, \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{n/2}} e^{-1/2(x_t - \mu)^T \Sigma^{-1} (x_t - \mu)}$$

El parámetro k es condicionado por el tamaño d la memoria disponible y el poder computacional de la computadora el valor esta en un rango entre 3-5.

Cada nuevo valor del pixel es comparado con las distribuciones de k existentes, y si el pixel esta dentro del 2.5 de la desviación estándar de una distribución entonces el pixel pertenece a esa distribución (Foreti, et al., 2005).

Figura 8. Tareas principales del cálculo de diferencias entre imágenes en el contexto de una cámara móvil.



(Foreti, et al., 2005)

4.3.4 Principios físicos

La óptica del ojo se basa en los principios físicos de refracción. Refracción es la flexión de los rayos de luz en la interfaz angulados de dos medios de comunicación transparentes con diferentes índices de refracción.

El índice de refracción n de un material es la relación entre la velocidad de la luz en el vacío C_0 a la velocidad de la luz en este material C ; $n = C_0/C$. El grado de refracción depende de la relación de los índices de refracción de los dos medios de comunicación, así como el ángulo entre el rayo de luz incidente y la interfaz normal: $n_1 \sin \phi_1 = n_2 \sin \phi_2$. Esto se conoce como la ley de Snell.

Lentes de refracción explotar a converger o divergir la luz, dependiendo de su forma. Paralelo los rayos de luz se doblan hacia el exterior al pasar a través de una lente cóncava hacia el interior y al pasar a través de una lente convexa. Estas propiedades se centra de una lente convexa se puede utilizar para la formación de imágenes. Debido a la naturaleza de la proyección, la imagen producida por la lente se invierte, es decir, girar 180° sobre el eje óptico.

Objetos a diferentes distancias de una lente convexa se centran en diferentes distancias detrás de la lente. En una primera aproximación, esto es descrito por la fórmula de Gauss para la lente:

Ecuación 7

$$\frac{1}{d_s} + \frac{1}{d_i} = \frac{1}{f}$$

Donde d_s es la distancia entre la fuente y el objetivo, d_i es la distancia entre la imagen y la lente, y f es la longitud focal de la lente. Un objeto infinitamente distante se centra en longitud focal, $d_i = f$. El recíproco de la longitud focal es una medida de la potencia óptica de una lente, es decir, qué tan fuertemente los rayos son doblados. La potencia óptica se define como $1m/f$, y se especifica en dioptrías.

Una variable de apertura, se añade a la mayoría de sistemas de imagen óptica, a fin de adaptarse a diferentes niveles de luz. Además de limitar la cantidad de luz que entra en el sistema, el tamaño de apertura también influye en la profundidad de campo, es decir, el rango de distancias de los objetos sobre los que aparecerá en el punto de mira en el plano de imagen. Una pequeña apertura produce imágenes con una gran profundidad de campo, y viceversa.

Otro efecto secundario de una apertura es la difracción. Difracción es la dispersión de la luz que se produce cuando la amplitud de una onda de luz es limitado. El resultado es una imagen borrosa. La cantidad de desenfoque depende de las dimensiones de la apertura en relación con la longitud de onda de la luz.

El tamaño se mide en términos de ángulo visual $\alpha = \text{atan}(s/2D)$ cubiertos por una imagen de tamaño s en la distancia D . En consecuencia, las frecuencias espaciales se miden en ciclos por grado (cpd) del ángulo visual.

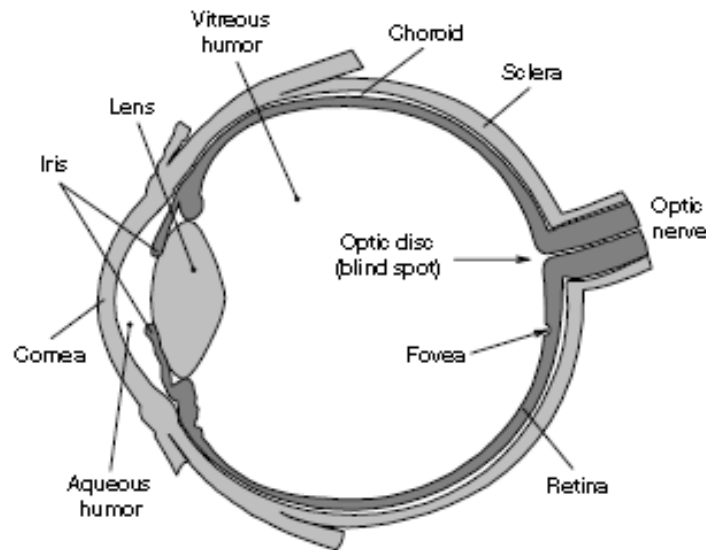
4.3.5 Óptica del Ojo

Hacer declaraciones generales acerca de la características ópticas del ojo es complicado ya que existen considerables variaciones entre cada individuo y además los componentes del ojo son sometidos a cambios continuos durante la vida, por esta razón las cifras aquí expuestas deben ser consideradas aproximadas.

El sistema óptico del ojo humano está formado por la córnea, el humor acuoso, el lente y el humor vítreo, como se ilustra en la siguiente figura. Los índices de refracción de estos cuatro componentes son 1,38, 1,33, 1,40 y 1,34,

respectivamente. El total de la potencia óptica del ojo es de aproximadamente 60 dioptrías. La mayor parte es suministrada por la transición del aire en la córnea, es aquí donde se produce la mayor diferencia en los índices de refracción (el índice de refracción del aire es cercano a 1). La lente ofrece sólo un tercio del poder total de refracción debido a que la óptica es de características similares en torno a los elementos.

Figura 9. El Ojo Humano.



(Winkler, 2005)

La importancia de la lente es que su curvatura y, por tanto, su potencia óptica puede ser aumentada voluntariamente por los músculos que se le atribuye. Este proceso se llama acomodación, el acomodación es fundamental para llevar los objetos a diferentes distancias en el foco en la retina. En los niños pequeños, la potencia óptica de la lente puede ser aumentado de 20 a 34 dioptrías.

Sin embargo, la capacidad de acomodación disminuye gradualmente con la edad hasta que se pierde casi por completo, una condición conocida como presbicia. Justo antes de entrar en la lente, la luz pasa por la pupila, el ojo de la apertura. La pupila es la apertura circular en el interior del iris, un conjunto de músculos que controlan su tamaño y, por tanto, la cantidad de luz que entra en el ojo en función de los niveles de iluminación exterior. La pigmentación del iris también es responsable por el color de nuestros ojos. El diámetro de la abertura pupilar se puede variar entre 1,5 y 8 mm, que corresponden a 30 veces el cambio de la cantidad de luz que entra en el ojo. La pupila es uno de los mecanismos del sistema visual humano que funciona para la adaptación de la luz (Winkler, 2005).

4.3.6 Calidad Visual

Es de suma importancia relacionar los parámetros de configuración con el sistema visual humano, ya que en la comunidad de video es muy popular relacionar la distancia de emisión con el tamaño de visualización, en múltiplos de la altura de la pantalla. La distancia de emisión preferida está entre 6 a 7 alturas de pantalla y para visualizaciones pequeñas está entre 3 a 4 alturas de pantalla del tamaño creciente de la visualización, otro de los factores es la suposición de resolución de la visualización o un cierto número de barrido de línea, el cual se usa actualmente en la televisión estándar. En el modelado de la visión las especificaciones de tamaño y la resolución de de la imagen que son proyectadas dentro de la retina son más que suficientes. Para una pantalla de altura H y una distancia de visión D , el tamaño es medido en grados desde el punto visual:

Ecuación 8

$$\alpha = 2 \operatorname{atan}(H/2D)$$

La resolución ó frecuencia espacial máxima f_{max} es medida en ciclos por grado del ángulo visual (cpd), esto se calcula con el número de barrido de líneas de acuerdo con el teorema de muestreo de Nyquist:

Ecuación 9

$$f_{max} = L/2 \alpha$$

En la Figura 11 se puede apreciar cómo se forman sobre la retina el tamaño y la resolución de la imagen de los formatos de video más populares, para un rango típico de distancias de visualización y alturas de pantalla.

En el contexto del modelado de la visión el tamaño y la resolución de la imagen proyectada sobre la retina para determinada altura H y distancia D , el tamaño se mide en grados del ángulo visual.

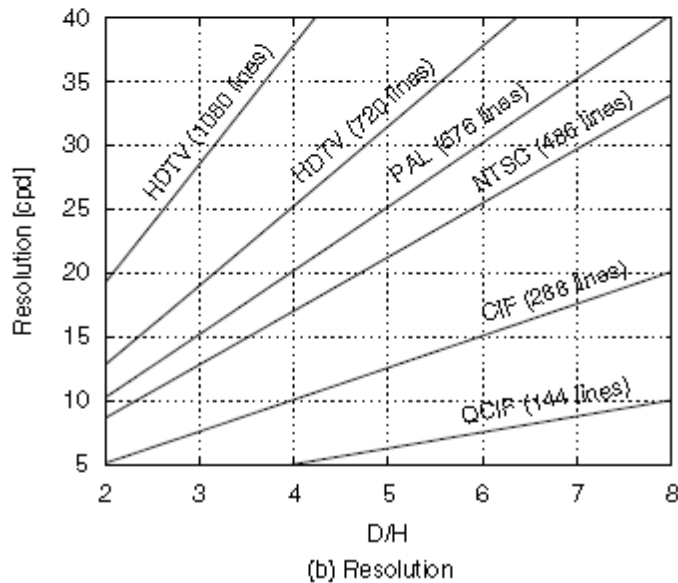
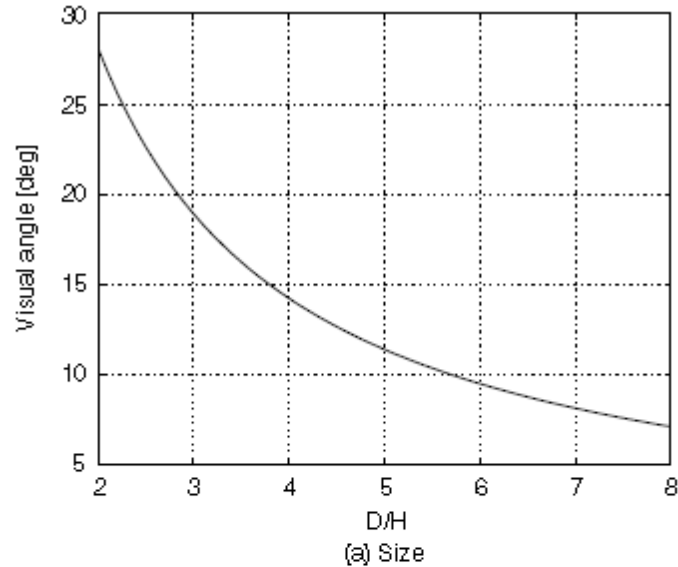
La distancia de emisión preferida esta alrededor de 6 - 7 alturas de pantalla, para visualizaciones más pequeñas, se acerca a 3 - 4 alturas de pantalla con un tamaño de visualización creciente.

4.4 ESTADO DEL ARTE

El análisis de video y el video de vigilancia son áreas de constante investigación, teniendo como puntos clave el video basado en detección y rendimiento, la identificación y seguimiento de personas y sistemas de vigilancia a gran escala. Un porcentaje significativo de tecnologías para detección basadas en video y

seguimiento fueron desarrolladas bajo la supervisión del gobierno de los Estados Unidos con un programa llamado Video Surveillance and Monitoring (VSAM).

Figura 10. Formación sobre la retina del tamaño y la resolución de la imagen.



(Winkler, 2005)

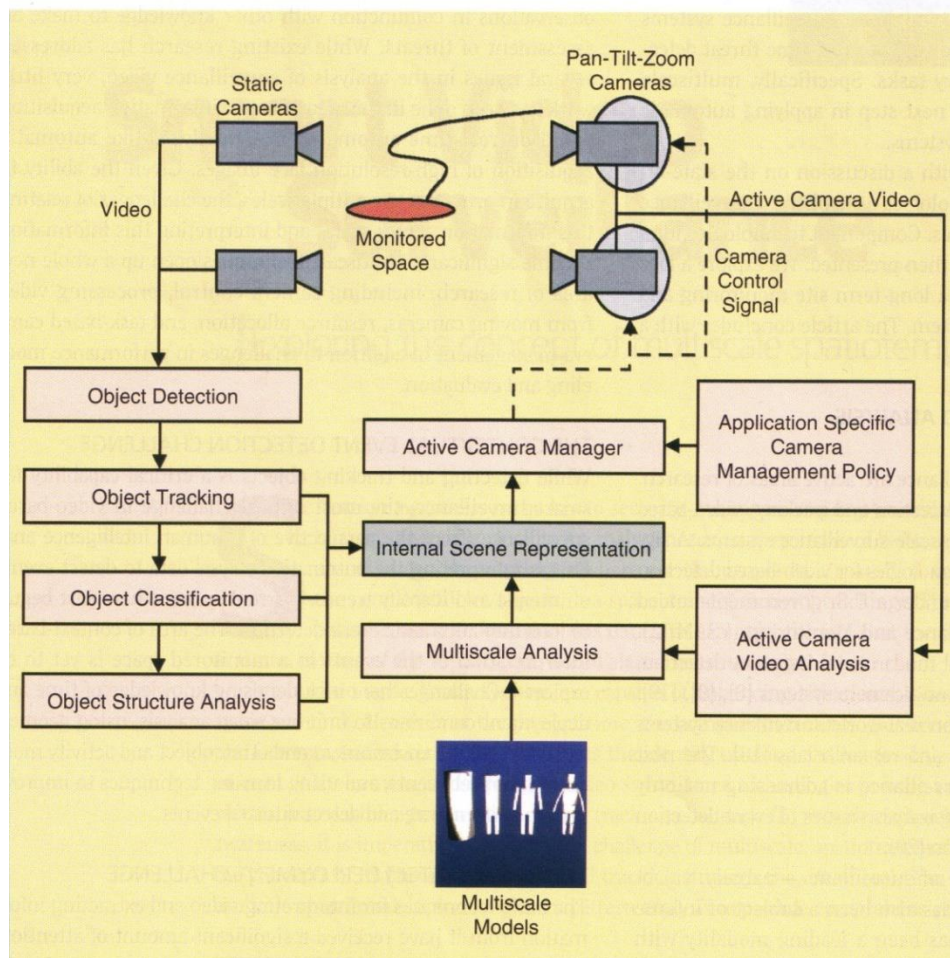
El reconocimiento de rostros ha venido siendo la modalidad líder de investigación en la industria de los sistemas y recientemente el gobierno de los estados unidos abrió un programa de investigación llamado Human ID el cual ha estimulado a la identificación de seres humanos desde grandes distancias usando técnicas basadas en la comparación de rostros para reconocimiento.

Las más avanzadas investigaciones concentran esfuerzos en los sistemas de vigilancia a gran escala también bajo supervisión del gobierno de los Estados Unidos, con el programa titulado Combats Zones That See. Este programa explora rápidamente el terreno desplegando sistemas de alto rendimiento con cámaras inteligentes que comunican sobre redes wireles, transmitiendo información a una estación central con el propósito de monitorear actividad de la zona y analizar patrones de movimiento térmicos.

4.4.1 Componentes para la Tecnología de Vigilancia Inteligente

El reto de la de la arquitectura multiescala incorpora varios desafíos técnicos, los cuales se presentan con una arquitectura general de un sistema de seguimiento multiescala. La meta del los sistemas de seguimiento multiescala es la de adquirir información acerca de objetos que se encuentren en el espacio monitoreado en variadas escalas de un marco unificado.

Figura 11. Modelo de la arquitectura multiescala.



(Hampampur, et al., 2005)

En la Figura 12 las cámaras estáticas ofrecen una visión global del escenario de interés, el zoom de la toma panorámica de las cámaras (Pam-Tilt-Zoom), tiene como finalidad obtener información detallada acerca de los objetos de interés en el escenario. El video obtenido de las cámaras estáticas es usado para detectar y hacer seguimiento de los objetos en dos y tres dimensiones. Adicionalmente, la cámara de imágenes fijas puede ser usada para extraer información adicional acerca de los objetos a un nivel más amplio, como el tipo de objeto (persona, carro) o los atributos del objeto (posición, velocidad), la escala de información es usada como una base para enfocar el arreglo de cámaras PTZ en un punto de atención. La información de las cámaras PTZ será utilizada para un análisis más específico. Por ejemplo si las cámaras fijas dan la información que el objeto es una persona el arreglo de cámaras PTZ se enfocara en el rostro de la persona para hacer un análisis de reconocimiento. Esta información que es mucho más amplia y tiene un análisis más específico es combinada para crear una representación propia del escenario, esto es lo que permite modelar en dos o tres dimensiones y así facilitar el registro en video de la detección de movimiento, reconocimiento, clasificación y análisis específico de objetos (Hampampur, et al., 2005).

4.5 LIMITACIONES Y ALCANCES

Como se expresó en el planteamiento de los objetivos, este proyecto permite determinar cuáles son los requerimientos técnicos mínimos necesarios en los dispositivos móviles, principalmente teléfonos celulares, para la posible implementación de un sistema de vigilancia visual, el cual no utilice compresión a los fotogramas capturados en la secuencia de video, si no una reducción significativa en la tasa de captura de los fotogramas.

Debido a que la grabación de un video de vigilancia se debe realiza bajo ciertas condiciones climáticas y de iluminación, ya que los algoritmos de reconocimiento requieren de escenarios controlados para un optimo funcionamiento, se presentan demasiados inconvenientes al hacerlo en zonas al aire libre. Por tanto la grabación de video propuesto en este proyecto se restringe a hacerla en recintos cerrados.

Los algoritmos de reconocimiento de objetos exigen un mínimo de pixeles en el área donde se encuentra ubicado el objeto a reconocer. Por esta razón, en el video capturado no debe existir una distancia considerable entre la cámara y el objeto de interés.

5. DISEÑO METODOLÓGICO

5.1 ETAPA 1 (RECOPIACIÓN DE INFORMACIÓN)

Se recopiló información referente a arquitecturas de los algoritmos basados en fotogramas, con el fin de adquirir las bases necesarias para poder diseñar un sistema de vigilancia basado en video. Esta etapa se cumplió durante el proceso de Seminario de Graduación.

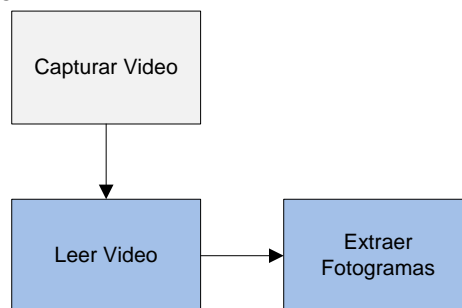
5.2 ETAPA 2 (DETERMINAR RESOLUCIÓN DE LOS FOTOGRAMAS)

En esta etapa se estableció la resolución adecuada de los fotogramas para hacer posible el reconocimiento de objetos, esta resolución se determinó a partir de estudiar diferentes algoritmos para el reconocimiento de objetos, los cuales hacen énfasis en el tamaño en píxeles que debe tener la zona en donde se encuentra el objeto de interés dentro del fotograma.

5.3 ETAPA 3 (ANÁLISIS VIDEO)

Para poder analizar las estructuras de video se utilizó el software Matlab® de la compañía The Mathworks, se investigó cuáles son los comandos que permiten capturar y leer video, descomponer el video en fotogramas, para su posterior manipulación fotograma a fotograma.

Figura 12. Análisis de video.

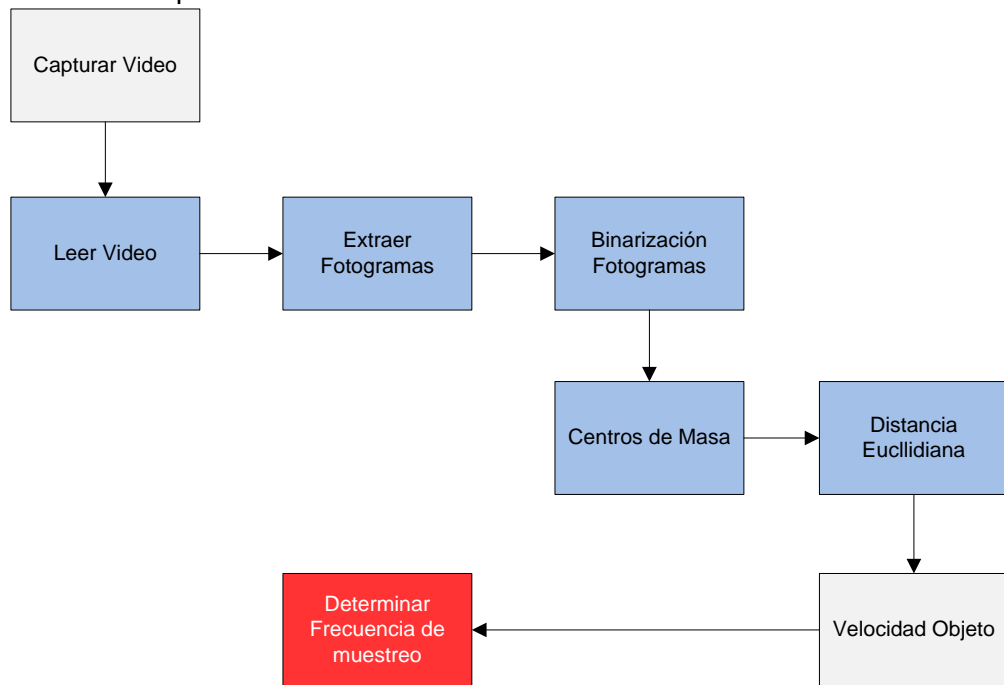


5.4 ETAPA 4 (DETERMINAR FRECUENCIA DE MUESTREO DE LOS FOTOGRAMAS)

Esta etapa se dividió en dos segmentos, el primero ellos fue el poder determinar la velocidad de los objetos que van a ser capturados por la cámara de grabación, y el segundo fue el poder decidir una frecuencia para la captura de los fotogramas la cual permita registrar la mayor cantidad de sucesos, este punto es de gran

importancia en el desarrollo del proyecto ya que la idea principal del proyecto es la de reducir el tamaño del video capturado reduciendo la tasa de captura de los fotogramas y no por compresión en los fotogramas capturados, por estas razones esta es la parte del proyecto que tomo la mayor cantidad de tiempo en el desarrollo.

Figura 13. Proceso para determinar la frecuencia muestreo.



5.5 ETAPA 5 (SELECCIÓN DEL CÓDEC DE VIDEO)

Teniendo la frecuencia de captura de los fotogramas se paso a buscar un códec de video de los ya existentes que en su arquitectura no haga ninguna compresión a los fotogramas al momento de generar la secuencia de video.

5.6 ETAPA 6 (EVALUACIÓN DE DESEMPEÑO Y DETERMINACIÓN REQUERIMIENTOS)

Ya teniendo claro el códec que se va implementaren la captura de video se hacen pruebas para calcular los tiempos de latencia y de ocupación de memoria en el procesamiento al capturar y generar video, para así poder definir los requerimientos técnicos mínimos de un teléfono celular para la posible implementación del sistema diseñado.

6. DESARROLLO

6.1 FUNCIONAMIENTO DE LOS SISTEMAS DE VIGILANCIA USADOS EN LA ACTUALIDAD.

Los sistemas de vigilancia que son utilizados en la actualidad se basan en arreglos de cámaras que deben estar en condiciones de contar con ciertos parámetros en su arquitectura para poder cumplir con la función de prestar vigilancia visual en entornos predeterminados para cada sistema. Los parámetros que deben contener las arquitecturas de los sistemas de vigilancia se evidenciaron en el desarrollo del marco referencial. En este aparte se hará referencia a los parámetros de mayor relevancia para la implementación de un sistema de vigilancia visual.

Un sistema de vigilancia moderno típico está basado en arreglos de cámaras móviles con las que se permite hacer control y énfasis en los eventos u objetivos de interés que ocurran en el escenario, estos sistemas deben también estar en las condiciones de poder transmitir el registro de sucesos del campo visual durante las 24 horas del día los 7 días a la semana, proveer controles para adaptar el sistema a las condiciones climáticas y de iluminación del entorno, con el fin de no perder el registro de eventos de importancia, para poder cumplir con estas exigencias, los sistemas de vigilancia modernos deben realizar ciertas tareas procesamiento fundamentales, como lo son.

- Calibración de la cámara con respecto a un plano cartesiano de referencia local.
- Adquisición de la imagen de la escena.
- Adaptar un modelo del fondo de la escena.
- Detectar eventos en las regiones del primer plano.
- Registrar los sucesos.

Teniendo implementadas estas tareas fundamentales en un sistema de vigilancia, se puede considerar este sistema de este tipo para hacerlo más robusto, al cual se le puedan adicionar tareas de procesamiento adicionales que puedan prestar servicios adicionales, como:

- Detección de movimiento.
- Mejoramiento de la resolución de la imagen.
- Compresión de los datos.

6.2 ALGORITMOS PARA EL RECONOCIMIENTO

El reconocimiento de objetos es una tarea fundamental para cualquier sistema de visión y en escenarios controlados se hace fácil su implementación y la

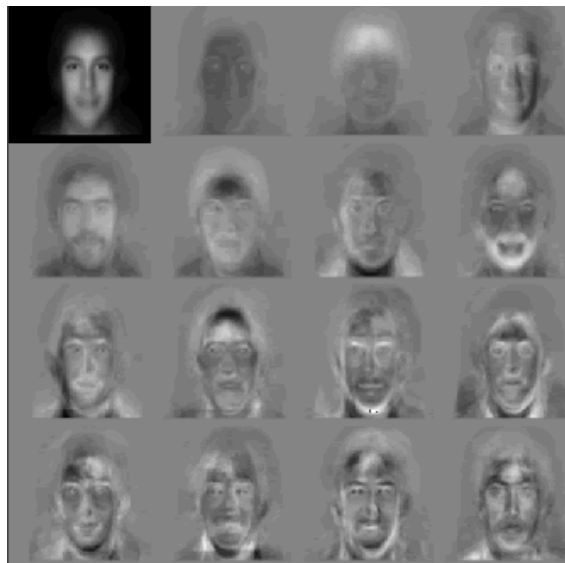
construcción de modelos en tres dimensiones de los objetos , pero en escenarios no estructurados se hace inviable aplicar técnicas de reconocimiento debido a las dificultades de segmentar adecuadamente el objeto incidente en el escenario.

El problema del reconocimiento se puede abordar de dos formas principalmente, la primera es la forma geométrica basándose en las características y la segunda forma es fotométrica basándose en su apariencia visual. Y las técnicas de mayor uso han sido el análisis de la componentes principales (PCA) y la técnica de análisis lineal discriminante (LDA).

6.2.1 Análisis de Componentes Principales (Principal Component Analysis, PCA)

PCA, comúnmente referida al uso de Eigenfaces, es la técnica impulsada por Kirby & Sirovich en 1988. Con PCA, el sondeo y la galería de imágenes deben ser del mismo tamaño y deben ser normalizadas previamente para alinear los ojos y bocas de los sujetos en las imágenes. La aproximación de PCA es luego utilizado para reducir la dimensión de los datos por medio de fundamentos de compresión de datos y revela la más efectiva estructura de baja dimensión de los patrones faciales.

Figura 14. Eigenfaces estándar.



((NSTC), et al., 2006)

Esta reducción en las dimensiones quita información que no es útil y descompone de manera precisa la estructura facial en componentes ortogonales (no correlativos) conocidos como Eigenfaces. Cada imagen facial puede ser

representada como una suma ponderada (vector de rasgo) de los eigenfaces, las cuales son almacenadas en un conjunto de una dimensión.

Una imagen de sondeo es comparada con una gallería de imágenes midiendo la distancia entre sus respectivos vectores de rasgos. La aproximación PCA típicamente requiere la cara completa de frente para ser presentada cada vez; de otra forma la imagen dará un resultado de bajo rendimiento. La ventaja primaria de esta técnica es que puede reducir los datos necesarios para identificar el individuo a 1/1000 de los datos presentados ((NSTC), Technology, & Biometrics, 2006).

6.2.2 Análisis Lineal Discriminante (Linear Discriminant Analysis, LDA)

LDA es una aproximación estadística para clasificar muestras de clases desconocidas basadas en ejemplos de entrenamiento con clases conocidas. Esta técnica tiene la intención de maximizar la varianza entre clases (ej. Entre usuarios) y minimizar la varianza de cada clase (Ej. De cada usuario). En la Figura 15 cada bloque representa una clase, hay grandes variaciones entre clases, pero pequeñas en cada clase. Cuando se trata con datos faciales de alta dimensión, esta técnica enfrenta el problema de muestras de tamaño pequeño que surge donde hay un número pequeño de ejemplos de entrenamiento comparados a la dimensionalidad del espacio de muestra.

Figura 15. Ejemplo de seis clases usando LDA.



((NSTC), et al., 2006)

6.2.3 Estándares para el Reconocimiento

La estandarización es una porción vital del avance del mercado y el estado del arte. Muchos trabajo se ha realizado en los estándares para facilitar la interoperabilidad y los formatos de intercambio de datos, lo que ayudará a facilitar el avance de la tecnología en una plataforma estandarizada. Las normas ANSI/INCITS (M1) 385-2004 e ISO 19794-5 de formato de intercambio de datos de reconocimiento facial son los mayores estándares en esta área y están

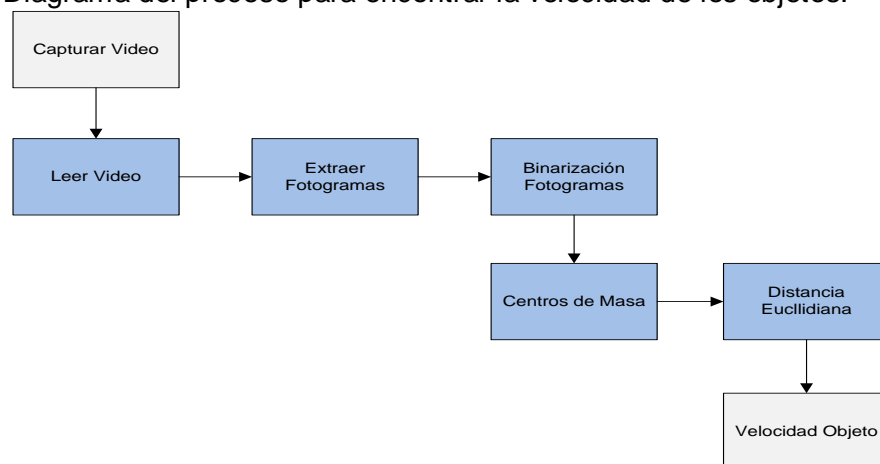
dirigidas al examen humano detallado de imágenes de rostros, verificación de identificación humana, e identificación y verificación facial automatizada.

Los estándares han establecido una imagen frontal definida y han irrumpido en sub-secciones dando tratamiento a imágenes frontales y no frontales (definiendo una imagen frontal a cinco grados del centro y una imagen no frontal está definida por la ubicación de los ojos). Estos estándares dejan otras imágenes tales como semiperfil indefinidas, pero aseguran que las imágenes enroladas alcanzarán el estándar de calidad necesario tanto para reconocimiento automatizado de rostros como para inspección humana de imágenes de rostros. Estos estándares también facilitan el uso de información de rostros en aplicaciones que tienen capacidad de almacenamiento limitada (Ej. Pasaportes, visas, licencias de conducir). Otros estándares como INCITS 398-2005 (Common Biometric Exchange Formats Framework, CBEFF), o Marco de trabajo de Formatos para Intercambios Comunes de Biometría, tratan específicamente con los elementos utilizados para describir los datos de biometría de forma común ((NSTC), et al., 2006).

La iluminación de interiores razonable controlada dada, la tecnología de punta de reconocimiento facial es de verificación del 90% a una tasa de falsa aceptación de 1%. El uso de modelos moldeables, los cuales mapean una imagen 2D sobre una grilla 3D en un intento de superar radiaciones posturales y de iluminación, puede mejorar significativamente el reconocimiento facial no frontal. El rendimiento de la lista de vigilancia decrece como función del tamaño de una galería el rendimiento utilizando listas de vigilancia más pequeñas es mejor que utilizando las más grandes.

6.3 DETERMINAR LA VELOCIDAD DE LOS OBJETOS DENTRO DEL VIDEO.

Figura 16. Diagrama del proceso para encontrar la velocidad de los objetos.



Para calcular este parámetro fue necesario primero conocer la posición del objeto de interés, para después calcular el centro de masa del objeto y finalmente trazar un vector velocidad entre dos fotogramas binarizados.

6.3.1 Proceso de Binarización de la Imagen.

En el proceso de binarización de la imagen se comparan dos fotogramas, el primero es una imagen fija en un instante de tiempo t_0 antes que el objeto entre en escena, y el segundo fotograma es tomado en un instante de tiempo t considerado el estado actual del objeto.

Figura 17. Proceso para binarización fotogramas.

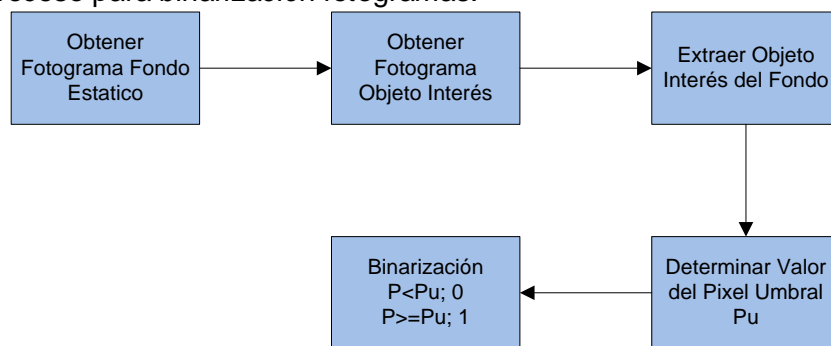


Figura 18. (a) Fotograma capturado en t_0 , (b) Fotograma capturado en t .



Al sustraer la imagen en t de la imagen t_0 se pueden identificar los pixeles que se modifican cuando el objeto entra en la escena.

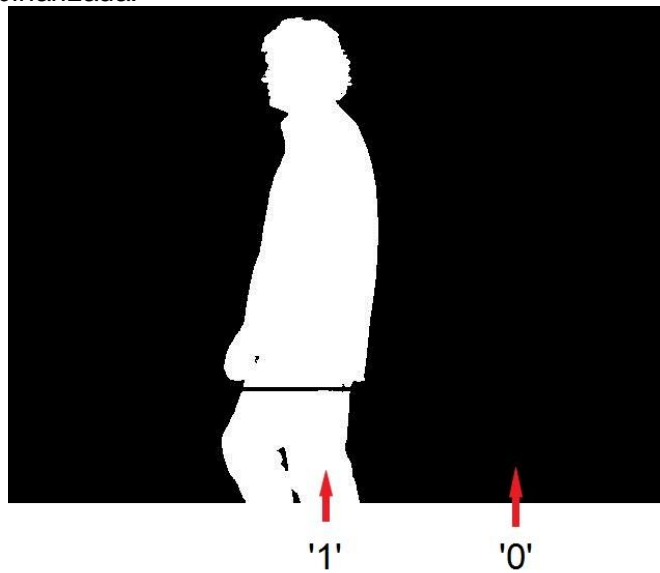
Figura 19. Imagen resultado de la resta de los fotogramas to-t.



Para el proceso de binarización se le asignara un valor de luminancia de '1' a los pixeles pertenecientes al objeto incidente en la escena, y a los demás pixeles de la imagen un valor de '0'. Para determinar los pixeles que componen el objeto y darles el valor de '1' se halla un valor de pixel umbral dentro de ese contorno.

Luminancia del pixel (0=negro → 1=blanco).

Figura 20. Imagen binarizada.



6.3.2 Cálculo Centro de Masa

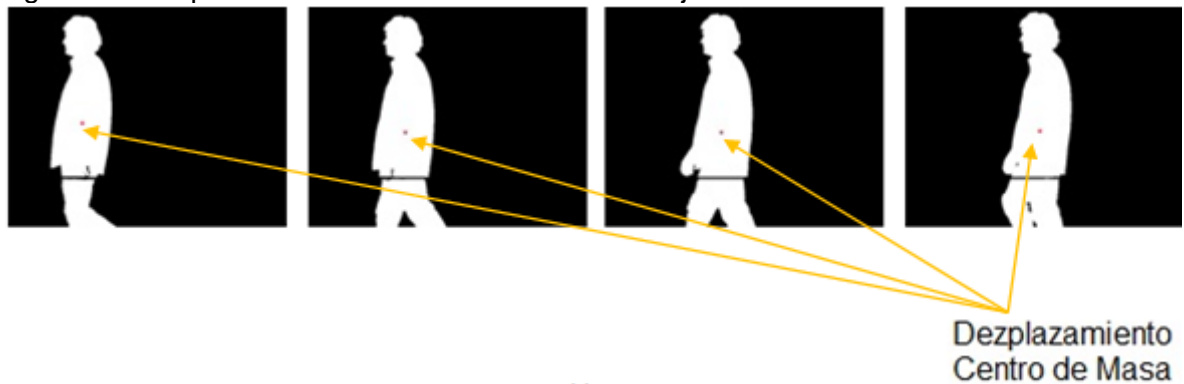
El centro de masa es hacia donde tiende la acumulación de la energía del objeto de interés, el hallar las coordenadas del centro de masa de un fotograma capturado en un instante de tiempo t y comparar esta información con la de otro fotograma capturado en un instante de tiempo $t+n$, esta comparación nos proporciona información, como la dirección de movimiento del objeto y la velocidad a la que se desplaza.

Para obtener las coordenadas del centro de masa se utiliza la siguiente ecuación.

Ecuación 10

$$x_c = \frac{\sum_{x=0}^{m-1} \sum_{y=0}^{n-1} x \times I_B(x, y)}{\sum_{x=0}^{m-1} \sum_{y=0}^{n-1} I_B(x, y)}; y_c = \frac{\sum_{x=0}^{m-1} \sum_{y=0}^{n-1} y \times I_B(x, y)}{\sum_{x=0}^{m-1} \sum_{y=0}^{n-1} I_B(x, y)}$$

Figura 21. Desplazamiento centro de masa de un objeto.



La secuencia de fotogramas mostrada en Figura 22, las imágenes fueron extraídas de una secuencia de video que se encuentra grabada a una tasa de 15fps (fotogramas por segundo), de derecha a izquierda se aprecian, el fotograma 60, el fotograma 65, el fotograma 70 y el fotograma 75, de la secuencia de video mencionada.

6.3.3 Cálculo Velocidad del Objeto

Para el calcular la velocidad instantánea de los objetos dentro del video se hace necesario conocer la distancia euclidiana (Ecuación 11) entre dos centros de masa del objeto de interés, lo cuales son calculados en dos fotogramas capturados consecutivamente y el otro parámetro que se debe conocer es el periodo de captura de los fotogramas. Con esta información se obtiene la velocidad del objeto en pixeles/segundo.

Ecuación 11

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

(Gonzalez, Woods, & Eddins, 2003)

Figura 22. Gráfica vector de la distancia euclidiana.

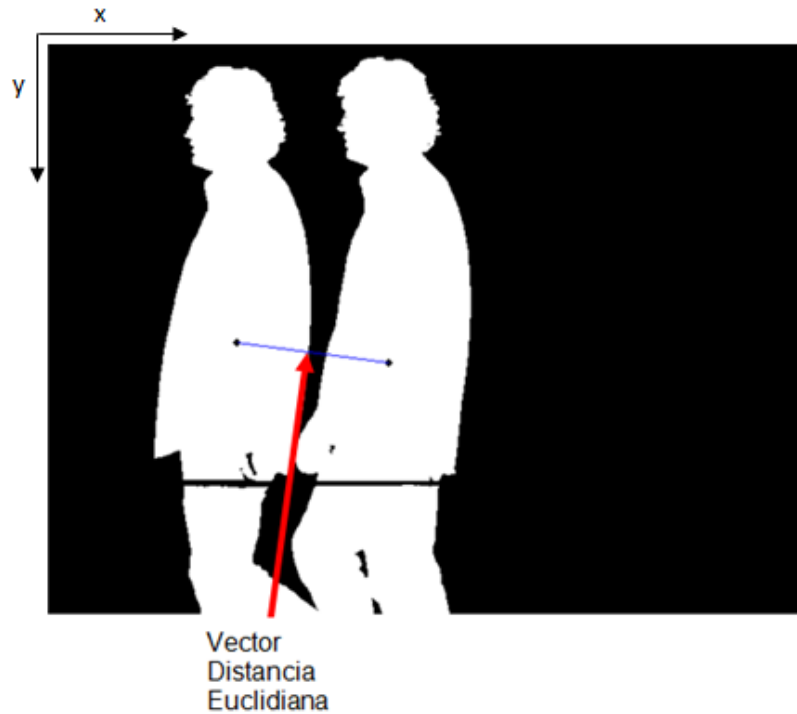
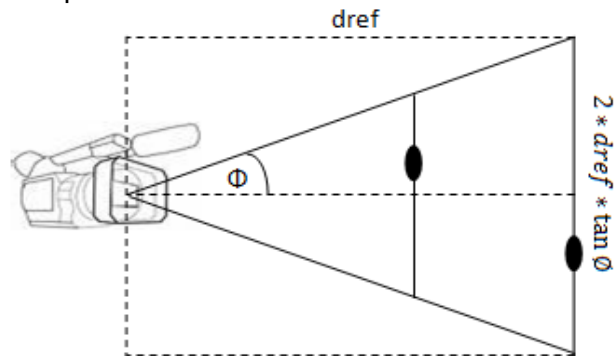


Figura 23. Espacio que capta la cámara.



En la Figura 23 se observa el grafico del vector de la distancia euclidiana entre fotograma 60 y el fotograma 75 referenciados en la Figura 22, este vector proporciona la información de la cantidad de desplazamiento del objeto de interés entre un fotograma y otro, en un tiempo determinado.

Como la distancia que se obtiene se encuentra en una magnitud de pixeles por segundo, se debe hallar su equivalente a metros por segundo, lo cual se obtiene a partir de la distancia horizontal que cubre la cámara la cual depende de la distancia referencial que existe entre la cámara y el objeto de interés.

De la Figura 23 se tiene que la distancia recorrida se puede obtener en función de la distancia referencial entre la cámara el objeto de y se puede expresar como:

Ecuación 12

$$n - \text{pixeles} = 2 * d_{ref} * \tan \phi$$

En donde n es la resolución horizontal del fotograma capturado.

Finalmente la velocidad de la distancia recorrida por el objeto de interés en metros es expresada como:

Ecuación 13

$$V = \frac{d}{T} = \frac{d}{T} \times \frac{2 * d_{ref} * \tan \phi}{\text{ancho fotograma}} \text{ metros/segundo}$$

En donde d es la distancia euclidiana, T es periodo de captura de los fotogramas y d_{ref} es la distancia referencial entre el objeto de interés y la cámara.

6.4 DETERMINAR FRECUENCIA DE MUESTREO DE LOS FOTOGRAMAS.

Con los resultados obtenidos de las pruebas realizadas, se dio paso a la determinación de de la rata de muestreo para la captura de fotogramas. Teniendo como parámetros para la selección los siguientes:

- Velocidad de los objetos incidentes en el escenario de grabación.
- Distancia referencial entre la cámara y el objeto de interés.
- La limitante de capturar los fotogramas en recintos cerrados, debido a las condiciones de iluminación inestables en espacios al aire libre.
- La resolución adecuada de los fotogramas para que se pueda implementar reconocimiento de objetos.
- Tamaño del fotograma capturado en el disco (ocupación de memoria).

En las pruebas de realizadas para calcular la velocidad de objetos de interés dentro de video, las cuales son las de mayor importancia, arrojaron los siguientes datos; la velocidad a la cual se desplazan las personas caminado se encuentra

alrededor de los 0.60 metros por segundo, el área aproximada de las locaciones que serán los futuros escenarios de grabación son en estimación de 10 metros² y el tamaño de ocupación de memoria de un fotograma de 640x480 pixeles en escala de grises sin ningún tipo de compresión es de aproximadamente 300Kbits, variando este valor con respecto a la incidencia de luz dentro del escenario.

Tomando como promedio el tiempo que le toma a una persona recorrer la mitad distancia máxima (5 metros), el cual es de 3 segundos y el tamaño promedio de cada fotograma. Se tomo la decisión de establecer la rata de muestreo de los fotogramas en la secuencia de video, en 2 fotogramas por segundo, logrando de esta manera capturar al menos 6 imágenes para hacer análisis al objeto de interés.

El decidir una rata de muestro como esta permite hacer la compresión planteada por el proyecto, ya que los estándares existentes plantean ratas de muestreo de mínimo 15 fotogramas por segundo y resoluciones de fotogramas más pequeñas, como se muestra en la siguiente tabla.

Tabla 1. Estándares de compresión de video.

Aplicación	Estándar	Tamaño	Rata de Fotogramas
Videoconferencia Tasa baja de bit	H.263	128x96 176x144	15-30
Video Básico Telefonía	H.261	176x144 352x288	10-20
Videoconferencia	H.261	352x288	15-30
Multimedia	MPEG-1	352x240	15-30
TV Digital – NTSC	MPEG-2	720x480	30
TV Alta Definición	MPEG-2	1200x800	30-60

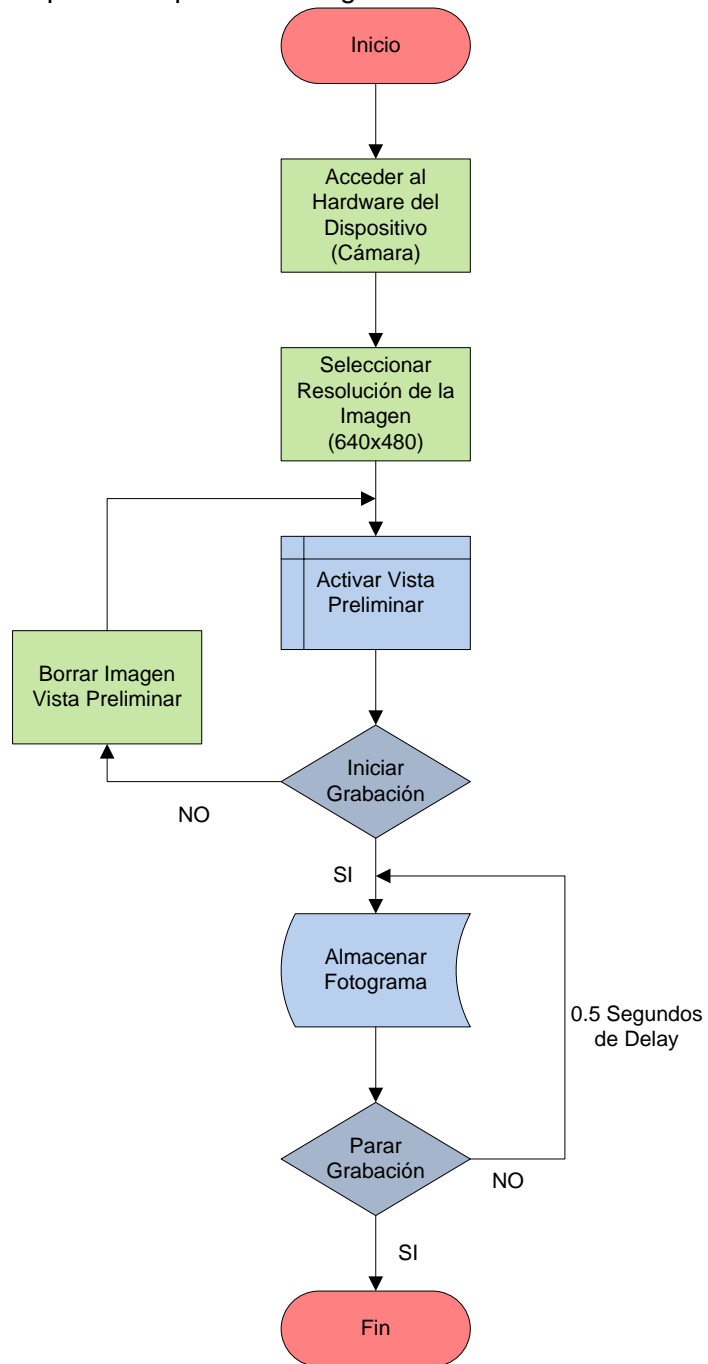
(Herman, 2007)

6.5 SELECCIÓN ALGORITMO PARA CAPTURA LA CAPTURA DE FOTOGRAMAS.

En la Figura 25 se describe el proceso propuesto para la captura de los fotogramas, en donde como primer paso es el acceder al hardware del dispositivo, es decir establecer una conexión con la cámara, cuando ya se establece esta conexión se adquiere una vista del escenario a grabar, mientras se encuentre activa esta vista preliminar solo se almacenará en la memoria volátil un solo fotograma el cual se borrara si no se activa la grabación, para darle espacio al siguiente fotograma de vista preliminar, de esta forma se aprovechan los recursos

de memoria del dispositivo, al momento de activar la grabación de la secuencia, cada fotograma capturado se almacena en la memoria a una frecuencia de dos por segundos y sin ningún tipo de compresión.

Figura 24. Algoritmo para la captura de fotogramas.



Los fotogramas son guardados sin formato, ya que las imágenes son para un propósito de análisis de la mismas posterior a la grabación, entonces no se hace necesario reproducir el video dentro del dispositivo.

6.6 SELECCIÓN DEL CÓDEC DE REPRODUCCIÓN DE LOS FOTOGRAMAS

Después de haber adquirido los fotogramas se determino la forma en reproducirlos en una secuencia de video, para este fin se estudiaron varias arquitecturas de códec para reproducir video, finalmente se selecciono el estándar de AVI (Audio Video Interleave), por razones expuestas a continuación.

AVI no es más que un formato de archivo desarrollado por la compañía Microsoft que puede guardar datos en su interior codificados de diversas formas y con ayuda de diversos codecs los cuales aplican diversos factores de compresión, pero también existe la posibilidad de almacenar los ficheros en un formato AVI "raw" o crudo, es decir, sin compresión.

Como uno de los propósitos específicos de este proyecto es el de no hacer ningún tipo de compresión a los fotogramas capturados no conviene crear una secuencia de video con un códec que haga compresión al reproducirlos. Con el reproductor AVI raw podemos montar los fotogramas en secuencia sin comprimirlos y es de uso gratuito, por estas razones se selecciono el reproductor AVI.

6.7 REQUERIMIENTOS TÉCNICOS MÍNIMOS NECESARIOS

Los requerimientos técnicos mínimos que deben contener los dispositivos portables en los cuales pueda ser implementado el sistema se determinaron a partir de la resolución de los fotogramas.

El tamaño de ocupación de memoria de un solo fotograma esta dado por:

$$\text{tamaño} = \text{mapa de color en bytes} \times \text{tamaño fotograma}$$

Este resultado se da en bytes haciéndose necesario convertir el mapa de color de bits a bytes.

Entonces un fotograma sin ningún tipo de compresión con características como las anteriormente mencionadas tiene un tamaño en disco de:

$$\text{tamaño} = 1 \text{ byte} \times 640 \times 480 = 307.2 \text{ Kbytes}$$

El resultado anterior fue para un solo fotograma, así mismo si se graba una secuencia de video de un minuto con una frecuencia 30 fotogramas por segundo sin compresión el video el tamaño del video es de:

$$\text{tamaño} = 1 \text{ byte} \times 640 \times 480 \times 30 \text{ fotogramas} \times 60 \text{ segundos} = 552 \text{ Mbytes}$$

Como se observa en el resultado el tamaño de un video sin compresión grabado con una rata de muestreo de fotogramas de las más usadas es muy grande por lo cual se hace necesario la compresión del video.

Como se determinó en este proyecto una compresión mediante la reducción de la rata de muestreo y esta se estableció en 2 fotogramas por segundo, se calcularon las requerimientos mínimos de memoria para una grabación continua de 30 minutos, de la siguiente forma:

En la memoria volátil del dispositivo siempre debe estar almacenado un fotograma correspondiente tiempo presente, por ende como mínimo esta memoria debe ser de 307.2 Kbytes pero se recomienda un tamaño mayor a este.

Para una grabación de 30 minutos se requiere una memoria de almacenamiento de:

$$\begin{aligned} \text{tamaño memoria} \\ = 1 \text{ byte} \times 640 \times 480 \times 2 \text{ fotogramas} \times 60 \text{ segundos} \times 30 \text{ minutos} \end{aligned}$$

$$\text{tamaño de memoria} = 1.10592 \text{ Giga byte}$$

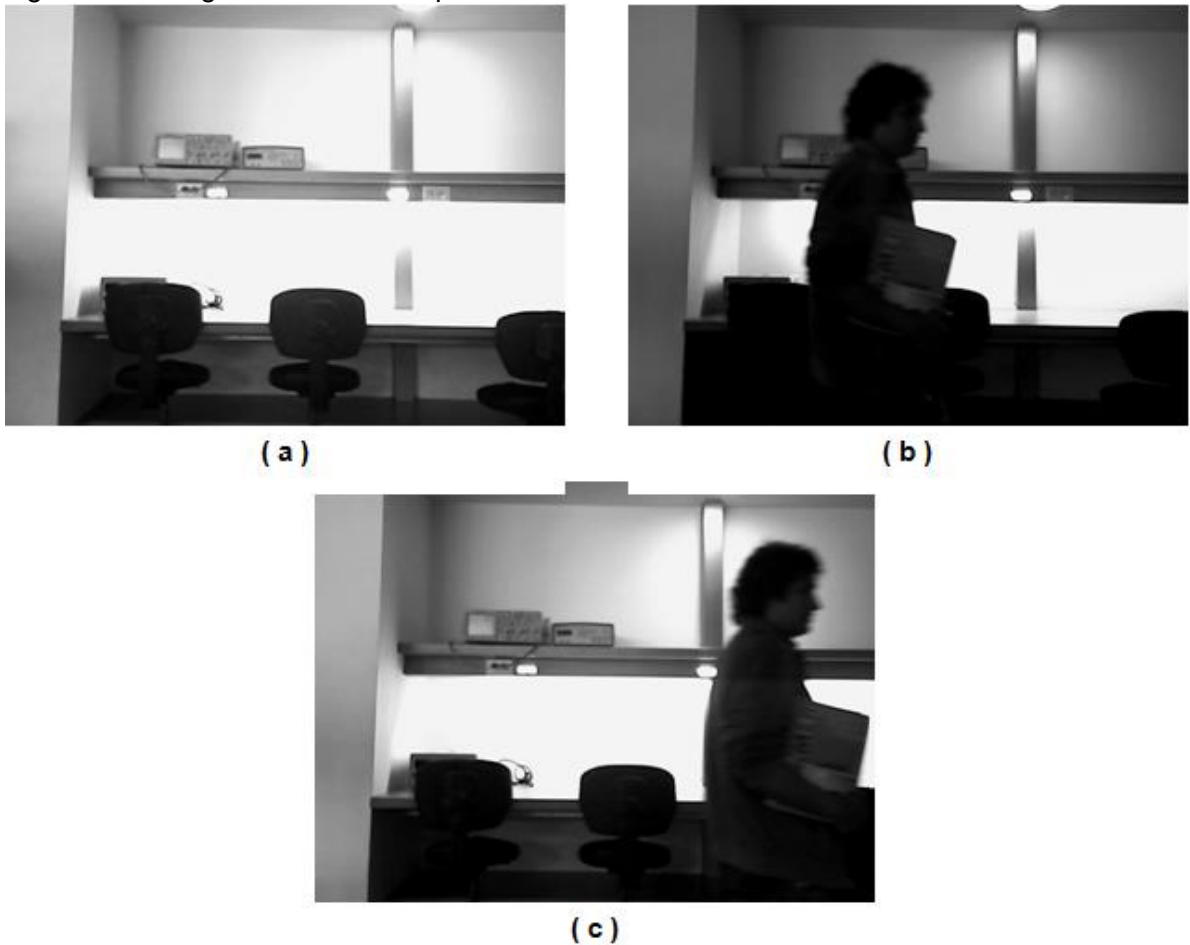
Requiriéndose como mínimo una capacidad de almacenamiento en el dispositivo portable de 1 Gigabyte.

7. PRUEBAS Y RESULTADOS

7.1 PRUEBA 1

Para esta prueba se utilizó el laboratorio de electrónica básica de la Fundación Universitaria San Martín (FUSM), en donde se grabó un video de 6 segundos con una tasa de muestreo de 15 fotogramas por segundo y resolución de 640x480 píxeles.

Figura 25. Fotogramas extraídos para la Prueba 1.

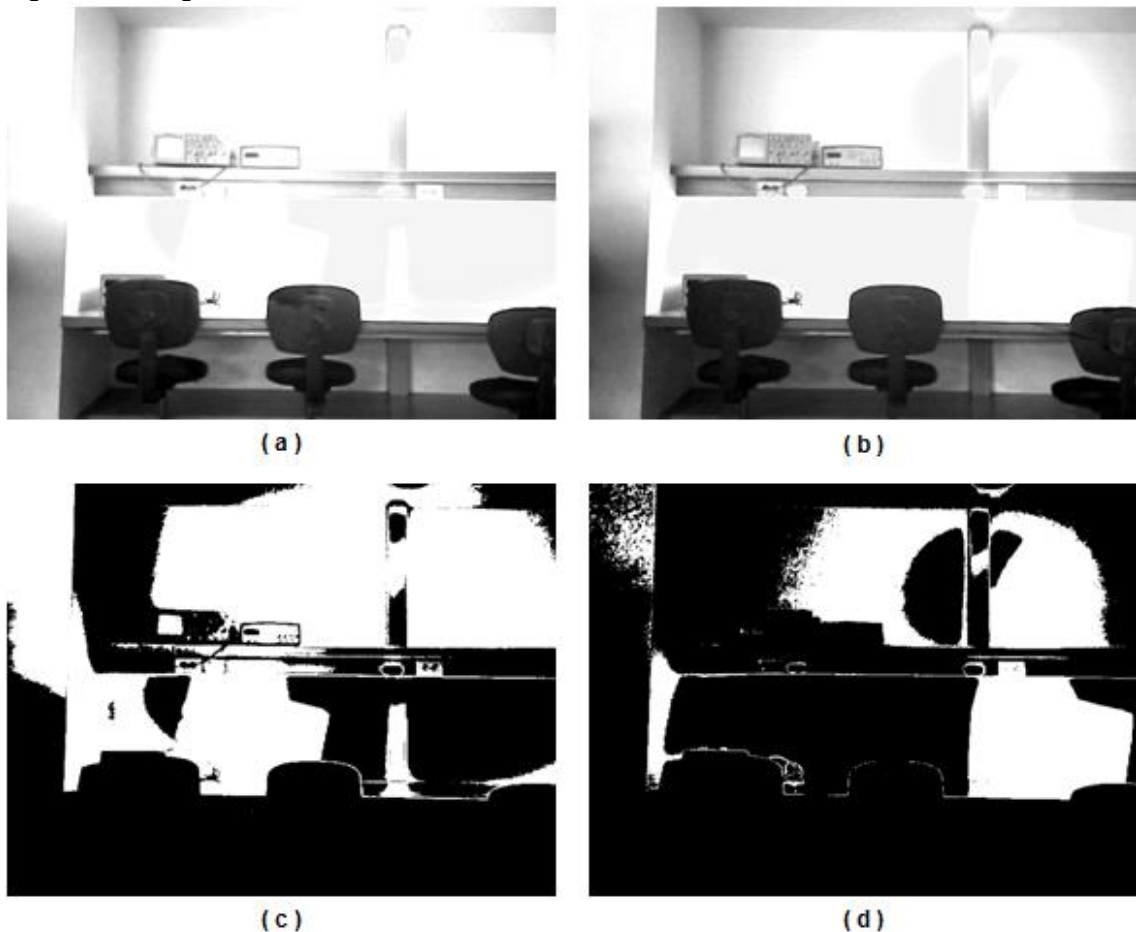


En la Figura 25. Fotogramas extraídos para la Prueba 1, se hace alusión a las siguientes imágenes (a) Fondo estático. (b) Fotograma 60 de la secuencia de video capturada. (c) Fotograma 75 de la secuencia de video capturada, en donde se puede apreciar el inconveniente de los cambios en la iluminación en los tres fotogramas de la figura anterior, esto se debe a que las cámaras tienen por defecto auto ajustar la apertura del diafragma cuando hay objetos incidentes en el

escenario, la apertura del diafragma permite controlar la incidencia de luz dentro de la grabación, por tal motivo en el fotograma 60 se nota que existe menos luz a comparación de la luz incidente en el fotograma 75, lo cual indica que existió un retardo en la apertura del diafragma cuando el objeto entro en el escenario.

Estas deficiencias de iluminación no permitieron hacer los cálculos con certeza en cuanto a la posición y distancia recorrida por el objeto, tal como se evidencia en la siguiente figura.

Figura 26. Registro de errores en la Prueba 1.



La imagen (a) de la Figura 26. Registro de errores en la Prueba 1. es el resultado de la resta entre el fotograma fondo estatico y el fotograma 60 y en la imagen (c) su equivalente binarizado, en la parte derecha de la figura la imagen (b) corresponde a la imagen resultante de la resta entre el fotograma 75 y el fondo estatico y en la parte inferior su equivalente binarizado. Como se puede apreciar en esta figura no es posible deducir con certeza cual es la ubicación del objeto en las imágenes debido a los cambios de iluminación que se dieron durante la

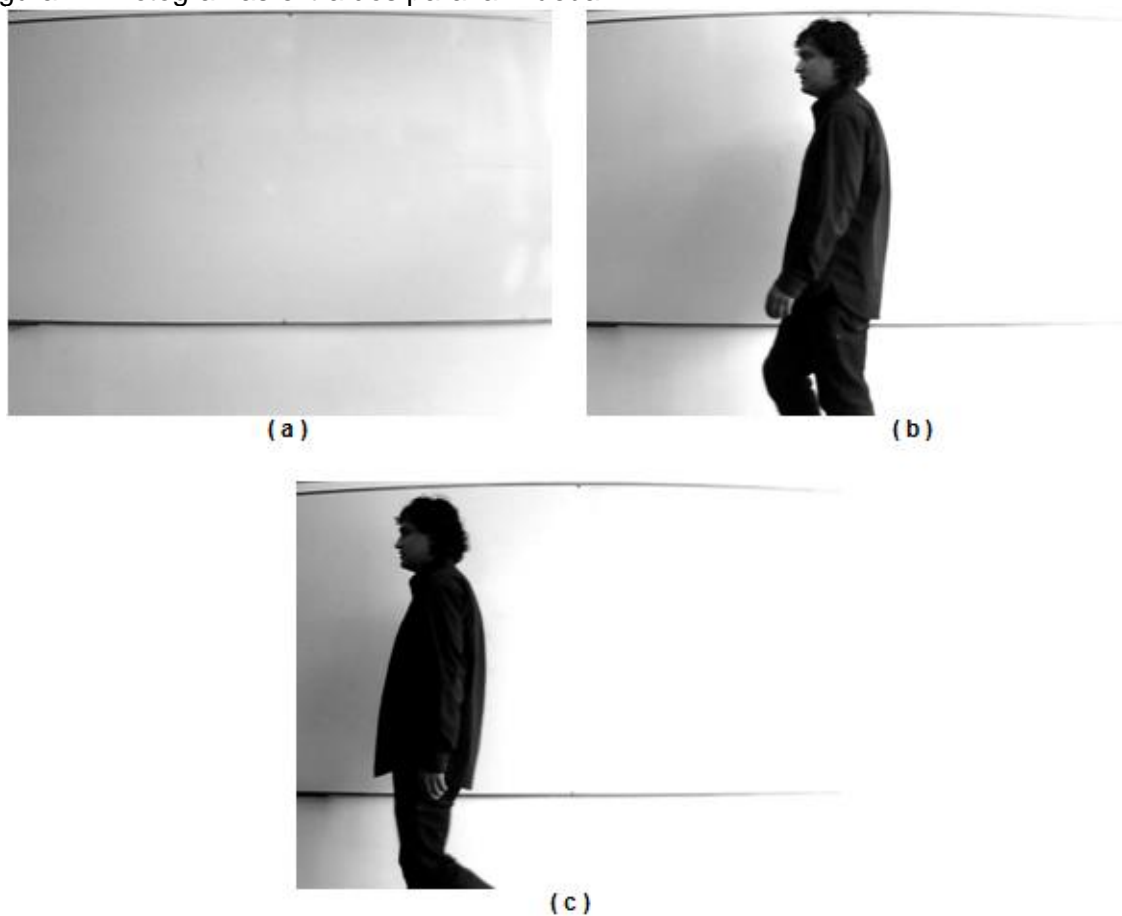
grabación del video, por lo cual surge el problema de valores diferentes en los pixeles que forman el fondo del escenario entre un fotograma y otro, por esta razón no se puede hacer una resta lineal entre las imágenes. Como se aclaró en el apartado de desarrollo se necesita binarizar el fotograma para hallar el centro de masa de los objetos y en esta prueba no es posible encontrar el valor de pixel umbral, ya que se presentan valores similares de pixeles dentro del objeto y en el escenario, por eso al hacer la binarización de los fotogramas brinda información errónea acerca de la ubicación del objeto de interés.

7.2 PRUEBA 2

Para esta prueba se mejoraron las condiciones de luz incidente en el escenario y se hizo una mejora en código de matlab para la manipulación de los fotogramas.

La locación para la grabación de esta prueba fue el laboratorio de comunicaciones de la FUSM. En donde se consiguió mejorar las condiciones de iluminación.

Figura 27. Fotogramas extraídos para la Prueba 2.

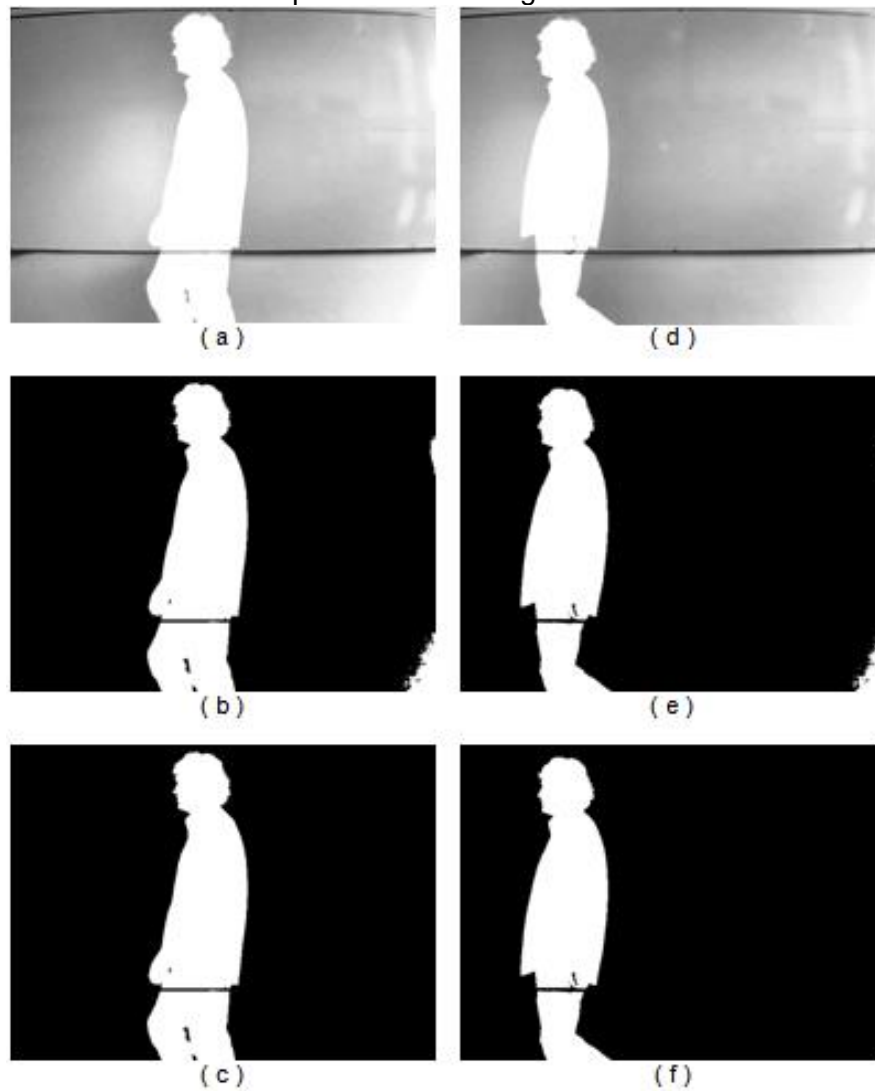


En la

Figura 278 se muestran las siguientes imágenes. (a) Fondo estático. (b) Fotograma 60 de la secuencia de video capturada. (c) Fotograma 75 de la secuencia de video capturada, en donde se hacen más evidente las diferencias existentes entre los fotogramas 60 y 75 con respecto al fondo del escenario, estas diferencias permiten facilidad al determinar la existencia de un objeto incidente en el escenario. El video fue grabado con una frecuencia de muestreo de 15 fotogramas por segundo y resolución de 640x480 pixeles y un tiempo de duración de 10 segundos.

7.2.1 Binarización de los Fotogramas.

Figura 28. (a) y (d) Fotogramas 60 y 75 base para la binarización, (d) y (e) resultado de la binarización de las imágenes con selección del pixel umbral arbitrario, (c) y (f) resultado de la binarización con el valor de pixel umbral escogido.



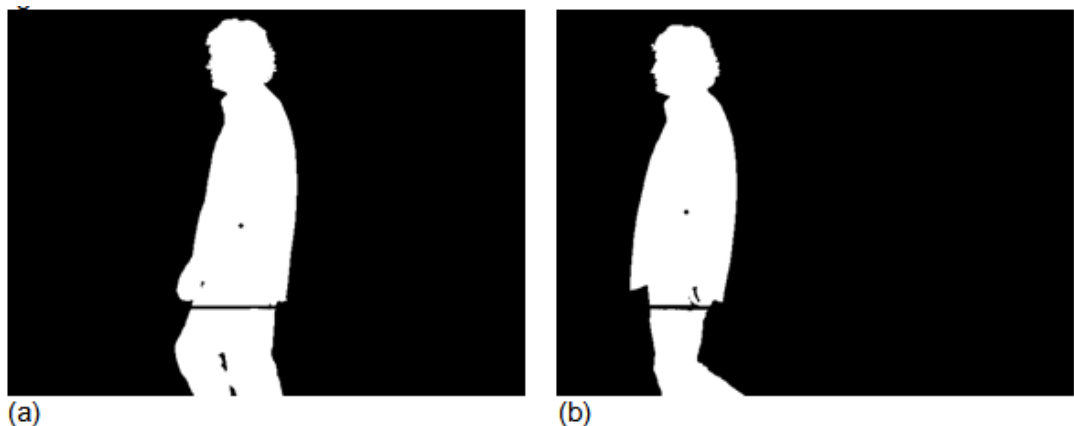
En el lado derecho de la Figura 28 hace alusión al proceso de binarización del fotograma 60 extraído de la secuencia video capturada para esta prueba, en la imagen (a) se observa como al restar el fotograma 60 del fondo estático se resalta la silueta del objeto incidente en el escenario en el segundo 4 de grabación, de igual manera se hace para el fotograma 75 de la secuencia y el resultado se muestra en la imagen (d). En las imágenes (b) y (e) se muestran las fotografías binarizadas correspondientes al número 60 y al número 75 respectivamente de secuencia de video, en estas imágenes se puede observar que aun se presentan problemas en la binarización ya que en la esquina inferior derecha de cada fotograma algunos pixeles quedaron con valor de '1' esto se registra como si existiera un objeto incidente en estas partes, lo cual no sucedió, esto se debe a los cambios en la incidencia de luz sobre la lente de la cámara, y si se hiciera el cálculo del centro de masa con estas imágenes se obtendría información inexacta.

El problema expuesto anteriormente se resolvió para las imágenes (c) y (f) de los fotogramas 60 y 75 respectivamente, escogiendo un valor del pixel umbral más cercano a los pixeles que forman la silueta del objeto de interés, de esa manera se eliminaron las formas que no pertenecían propiamente al objeto.

7.2.2 Cálculo de la Distancia Euclidiana y la Velocidad.

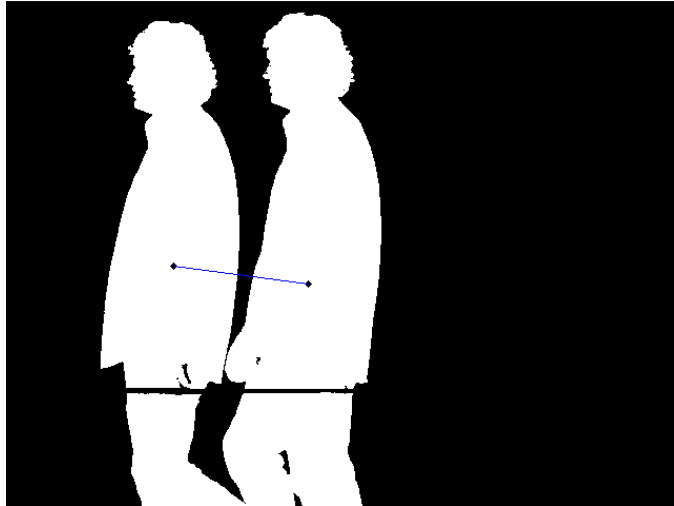
Para esta prueba el tamaño de las imágenes sobre las cuales se halló el centro de masa es de 640x480 pixeles. Para el fotograma 60 los cálculos arrojaron las coordenadas (288.033, 268.95) del centro de masa del objeto, puesto que los valores de las posiciones de los pixeles están dadas en valores enteros entonces se escogieron los valores de mayor proximidad a los valores del resultado, finalmente las coordenadas del centro de masa del objeto de interés en el fotograma 60 son (288, 269) y en el fotograma 75 son (159.99, 252.46) siguiendo el mismo criterio quedan (160, 252). Estas coordenadas se graficaron en la Figura 29.

Figura 29. Gráficas de los centros de masa.



Teniendo las coordenadas de los centros de masa del objeto de interés en dos fotogramas de una secuencia de video se emplea la ecuación de la distancia euclidiana para calcular la distancia recorrida por el objeto.

Figura 30. Gráfica de la distancia recorrida por el objeto de interés.



La distancia recorrida por el objeto interés la describe el comportamiento del vector graficado de color azul en la Figura 30 y tiene una magnitud de 129.0998 píxeles.

La distancia referencial desde la cámara hasta el objetivo es de 2.20 mts. y el ángulo de apertura del lente de la cámara con respecto a la distancia referencial es de 70.332° . Con estos datos se puede ya calcular el equivalente de la distancia recorrida en metros. La distancia recorrida por el objeto de interés fue de 0.625 mts. Como la frecuencia de muestreo de los fotogramas es de 15 fotogramas por segundo, entonces el tiempo transcurrido entre el fotograma 60 y el fotograma 75 fue de un segundo, finalmente se tiene que la velocidad del objeto de interés en esta prueba fue de 0.62 metros por segundo.

7.3 PRUEBA 3

Esta prueba el objeto de interés se enfoco en el movimiento de las manos de una persona incidente en escenario, utilizando un objeto en las manos de la persona de un color que resaltará del fondo estático y de los colores usados por la persona misma. La resolución en fotogramas capturados es de 640x480 píxeles, a una frecuencia de 15 fotogramas por segundo.

7.3.1 Binarización de los Fotogramas.

De la Figura 31 se observa el objeto de interés para esta prueba en particular en las imágenes (a) y (b) de los fotogramas 60 y 75 respectivamente, y se puede visualizar que el objeto es de un color diferente al de los demás colores de objetos incidentes en la grabación.

En la Figura 32, la imagen (a) es el Fotograma 60 de secuencia de video después del proceso de binarización y la imagen (b) corresponde al Fotograma 75. Esta prueba se facilitó a comparación de las pruebas anteriores, debido a que se suplieron rápidamente los inconvenientes presentados gracias las experiencias adquiridas en las otras pruebas.

Figura 31. Fotogramas extraídos de la secuencia de video para la Prueba 3.



(a)

(b)



(c)

Figura 32. Fotogramas binarizados para la Prueba 3. (a) Fotograma 60 binarizado, (b) Fotograma 75 binarizado.



7.3.2 Cálculo de la Distancia Euclidiana y la Velocidad.

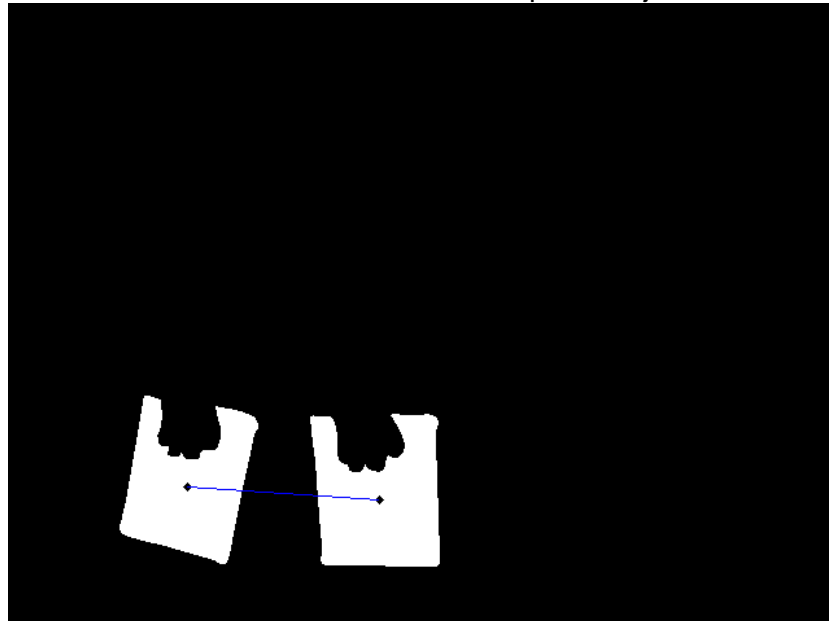
Al determinar el centro de masa del objeto de interés en los dos fotogramas a comparar, los resultados de las coordenadas en cada uno fueron las siguientes, para el Fotograma 60 el centro de masa se encuentra ubicado en (139,4,374) y para el Fotograma 75 se encuentra en (287,384) coordenadas en pixeles, teniendo en cuenta que el tamaño de los fotogramas es de 640x480 pixeles. La gráfica de estas coordenadas se evidencia en la Figura 33.

Figura 33. Gráfica del centro de masa del objeto de interés. (a) Centro de masa del objeto en el Fotograma 60, (b) centro de masa del objeto en el Fotograma 75.



La distancia recorrida por el objeto entre el segundo 4 y el segundo 5 de la secuencia de video capturada fue de 148.0429 pixeles. La distancia referencial desde la cámara hasta el objeto de interés es de 2.20 mts. y el ángulo de apertura del lente de la cámara con respecto a la distancia referencial es de 70.332° . Finalmente se obtuvo la velocidad del objeto de interés para esta prueba y fue de 0.71 metros por segundo. En la Figura 34 se muestra la grafica del vector de las distancia recorrida por el objeto de interés.

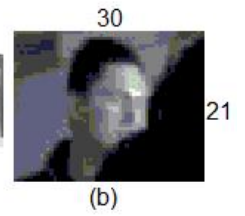
Figura 34. Gráfica del vector de la distancia recorrida por el objeto de interés.



7.4 PRUEBA 4

En esta prueba se hace alusión al tamaño de objetos capturados en el campo visual, a una distancia referencial de 10 metros de la cámara. El Figura 36 se observa la calidad visual de la imagen cuando se hace un zoom a un objeto de interés ubicado a una distancia de 10 metros, el tamaño del fotograma es de 640x480 pixeles, el rostro extraído presenta un tamaño de 10x11 pixeles, los cuales son suficientes para satisfacer los requerimientos mínimos de pixeles, para la aplicación de los algoritmos de reconocimiento de rostros PCA (Principal Component Analysis) y LDA (Linear Discriminant Analysis). Determinándose de esta manera el tamaño de 640x480 pixeles para cada fotograma ideales para la captura de la secuencia de video.

Figura 35. Zoom a un rostro ubicado a 10 metros de la cámara.



8. CONCLUSIONES

A través del desarrollo del proyecto se documentó como los sistemas de vigilancia usados en la actualidad basan el manejo de memoria bajando la resolución de los fotogramas utilizando codecs complejos como M-PEG, minimizando así la posibilidad de reconocer objetos dentro de la secuencia de video.

El algoritmo propuesto para la grabación de la secuencia de video saca provecho los recursos de la memoria RAM de los dispositivos ya que solamente tiene que almacenar un fotograma, debido a que el periodo para la toma de fotogramas al ser de 0.5 segundos durante el cual ese fotograma se guarda directamente en la memoria ROM y se elimina de RAM para dar paso al fotograma siguiente.

El proceso de binarización es muy confiable para determinar el área de objetos de interés incidentes en imágenes y para este proceso el valor del pixel umbral es fundamental para hacer cálculos erróneos acerca de la posición del objeto.

La incidencia de luz en el escenario de grabación es un factor crítico en la grabación de una secuencia de video y se hace incontrolable en espacios al aire libre debido a los constantes cambios, lo cual reduce la calidad de las imágenes adquiridas.

El calcular la velocidad de objetos dentro de una secuencia de video fue fundamental para determinar la tasa de muestreo de fotogramas.

En las pruebas realizadas del movimiento de una persona para una distancia referencial de 3.2 metros entre la cámara y el objeto de interés, con un ángulo de 70° de apertura del lente de cámara, arrojaron una velocidad promedio de 0.6 metros por segundo del movimiento de las personas.

Si es posible hacer compresión de una secuencia de video disminuyendo la tasa de muestreo, como se evidenció en el desarrollo del proyecto ya que un video de 1 minuto sin compresión en los fotogramas a una tasa normal de muestreo de 30 fotogramas por segundo, es de un tamaño de 500 Megabytes, y con la tasa de muestreo propuesta en este proyecto se puede grabar 30 minutos de grabación continua con una capacidad de almacenamiento de 1 Gigabyte, obteniendo una mejor calidad en los fotogramas adquiridos.

Para la reproducción de la secuencia de fotogramas adquirida se seleccionó el reproductor AVI raw, ya que en este brinda la posibilidad de almacenar ficheros sin compresión y es de uso gratuito.

No es necesario el color en una imagen para el reconocimiento de patrones ya que los algoritmos utilizados para este propósito utilizan imágenes en 2 dimensiones. El trabajar con una imagen en escala de grises sí facilita su análisis y velocidad de procesamiento.

Se seleccionó una resolución de fotogramas de 640x480 píxeles para un campo visual de captura de 10 metros, ya que las técnicas de reconocimiento estudiadas reducen los datos necesarios para identificar un individuo a 1/1000 de los datos presentados.

El algoritmo propuesto para la grabación de video permite almacenar cantidades finitas de fotogramas, no obstante debe estar supervisado de forma manual por el usuario para evitar, ya que se podrían saturar los recursos de memoria del dispositivo.

9. RECOMENDACIONES

En este proyecto se realizó el cálculo de la velocidad de objetos de interés dentro de un video, a partir de los centros de masa del objeto en fotogramas consecutivos, con base a esto se propone el estudio de detección de movimiento en secuencias de fotogramas con propósitos de predicción de movimiento, seguimiento y control para robótica.

El procesamiento digital de imágenes es un campo de acción muy extenso y poco explorado en el país se recomienda hacer estudios acerca de otras formas de compresión de imágenes y de video, como el uso de algoritmos basados en fractales, redes neuronales y algoritmos genéticos.

El campo de la biométrica específicamente el reconocimiento facial aun se encuentra aun en desarrollo lo cual es interesante para estudios posteriores.

GLOSARIO

CAMARA: el sensor de la cámara fotográfica convierte la escena visible por la lente en una señal eléctrica para transmitirla para su visualización. (Kruegle, 2007)

COMPRESIÓN: reducir el ancho de banda que ocupa una señal, quitando componentes poco significativas de la señal.

DCT (Discrete Cosine Transform) transformada discreta del coseno: es una de las transformaciones más importantes en el área de la compresión de imágenes digitales, la DCT que consiste en compactar la energía, los vectores de la base dependen del orden de la transformada, la DCT posee la capacidad de cuantificar los coeficientes utilizando valores de cuantificación que se eligen de forma visual. (E.U.I.T.Telecomunicación, 2008)

Entropía: f. Fís. Magnitud termodinámica que mide la parte no utilizable de la energía contenida en un sistema.

FRECUENCIA: Es el número de veces que en la unidad de tiempo se repiten las imágenes, en una secuencia de video.

MODELO DE ILUMINACIÓN DE PHONG: Es un modelo empírico simplificado para iluminar puntos de una escena, en este modelo, los objetos no emiten luz, sólo reflejan la luz que les llega de las fuentes de luz o reflejada de otros objetos. La luz que llega a un objeto puede ser de tres tipos, luz ambiental la cual proviene de todas las direcciones e ilumina todas las caras del objeto por igual, luz difusa que proviene de una dirección pero se refleja en todas direcciones y luz especular proviene de una dirección y se refleja sólo en una dirección. La iluminación de un punto se calcula como la suma de los tres tipos de iluminaciones: $I_{\text{Phong}} = I_{\text{amb}} + I_{\text{dif}} + I_{\text{esp}}$. (Kruegle, 2007)

RESOLUCIÓN: toda imagen digital está compuesta por pixeles tanto a lo largo como a lo ancho ej.: 800x600 pixeles (resolución utilizada en aplicaciones de internet), entre más cantidad de pixeles posea la imagen será de mayor resolución.

VSAM (VIDEO SURVEILLANCE AND MONITORING): Programa de los Estados Unidos para el desarrollo de tecnologías para vigilancia y seguimiento. (Hampampur, et al., 2005)

BIBLIOGRAFÍA

- Chaudhuri, S., & Taur, D. (2005). High-Resolution Slow-Motion Sequencing: How generate a slow-motion sequence from a bit stream. *IEEE Signal Processing Magazine: Surveillance Networks*, 22(2), 16-24.
- Collins, R., Lipton, A., Fujiyoshi, H., & Kanade, T. (2001). A system for video surveillance and monitoring. *Proc. IEEE*, 89(10), 1456-1477.
- E.U.I.T.Telecomunicación, D. d. I. A. y. C. (2008). Retrieved marzo de 2009, 2009, from www.diac.upm.es/acceso_profesores/asignaturas/tdi/tdi/transformadas/pdf/dct.pdf
- Foreti, G., Micheloni, C., Snidaro, L., Remagnino, P., & Ellis, T. (2005). Active video-Based Surveillance System: The low-level image and video processing techniques needed for implementation. *IEEE Signal Processing Magazine: Surveillance Networks*, 22(2), 26-36.
- Gonzalez, R., Woods, R., & Eddins, S. (2003). *Digital Image Processing Using Matlab* (Primera ed., pp. 624): Prentice Hall.
- Hampampur, A., Brown, L., Connel, J., Hass, N., Merlk, H., Pankanti, S., et al. (2005). Smart Video Surveillance: Exploring the concept of multiscale spatiotemporal tracking. *IEEE Signal Processing Magazine: Surveillance Networks*, 22(2), 38-45.
- Herman, K. (2007). *CCTV Surveillance:Video Practices and technology*. Burlington: Elsevier.
- Kapur, J. N., Sahoo, P. K., & Wong, A. K. C. (1985). A new method for gray-level picture thresholding using the entropy of the histogram, *Computer Vision, Graphics, Image Processing*. 29(3), 38-45.
- Kruegle, H. (2007). *CCTV Surveillance:Video Practices and technology*. Burlington: Elsevier.
- (NSTC), N. S. a. T. C., Technology, C. o., & Biometrics, S. o. (2006). *Face Recognition*: Executive Office of the President of the United States.
- Parker, J. R. (2001). Gray-level thresholding in badly illuminated images. *IEEE Trans. PattenAnal. Machine Inteli.*, 13(8), 813-819.
- Westwater, R., & Furht, B. (1997). *Real-Time Video Compression: Techniques and Algorithms*. Massachusetts: Academic Publishers.

Winkler, S. (2005). In W. Sussex (Ed.), *Digital Video Quality: Vision Models and Metrics* (Primera ed.). England: Willey.