

INSTITUTO POLITECNICO NACIONAL

Sección de Estudios de Posgrado e Investigación

Unidad Culhuacán

**“Detección Y Seguimiento de Personas Basado
en Vectores de Movimiento”**

T E S I S

**QUE PARA OBTENER EL GRADO DE DOCTOR EN CIENCIAS DE
INGENIERIA EN COMUNICACIONES PRESENTA:**

JOSUE AARON HERNANDEZ GARCIA
MAESTRO EN CIENCIAS DE INGENIERIA EN
MICROELECTRONICA

ASESOR DE TESIS:
Dr. HECTOR MANUEL PEREZ MEANA

Mayo 2011





INSTITUTO POLITÉCNICO NACIONAL

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

ACTA DE REVISIÓN DE TESIS

En la Ciudad de México D. F., siendo las 12:00 horas del día 24 del mes de mayo del 2011 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Estudios de Posgrado e Investigación de SEPI-ESIME-CULH. para examinar la tesis titulada:

"Detección y Rastreo de Movimiento Relevante Usando Secuencias de Video"

Presentada por el alumno:

Hernández

García

Josué Aarón

Apellido paterno

Apellido materno

Nombre(s)

Con registro:

A	0	7	0	2	0	8
---	---	---	---	---	---	---

aspirante de:

Doctorado en Comunicaciones y Electrónica

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

LA COMISIÓN REVISORA

Director de tesis

Dr. Héctor Manuel Pérez Meana

Dr. Volodymyr Ponomaryov

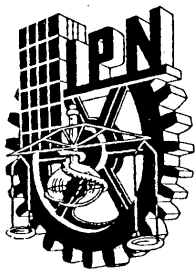
Dra. Mariko Nakano Miyatake

Dr. Juan Carlos Sánchez García

Dr. Gonzalo Isaac Duchén Sánchez

PRESIDENTE DEL COLEGIO DE PROFESORES

Dr. Gonzalo Isaac Duchén Sánchez



INSTITUTO POLITÉCNICO NACIONAL
SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

CARTA CESIÓN DE DERECHOS

En la Ciudad de México D. F. el día 30 del mes de mayo del año 2011, el que suscribe Josué Aarón Hernández García alumno del Programa de Doctorado en Comunicaciones y Electrónica con número de registro A070208, adscrito a La Escuela Superior de Ingeniería Mecánica y Eléctrica, manifiesta que es autor intelectual del presente trabajo de Tesis bajo la dirección del Dr. Héctor Manuel Pérez Meana y cede los derechos del trabajo intitulado Detección y Rastreo de Movimiento Relevante Usando Secuencias de Video, al Instituto Politécnico Nacional para su difusión, con fines académicos y de investigación.

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

M en C Josué Aarón Hernández García

Nombre y firma

RESUMEN

Los requerimientos de seguridad actuales demandan sistemas de detección de movimiento más confiables, precisos, fáciles de manejar y que sean capaces de adaptarse a cualquier escenario o área de vigilancia. Los grandes avances que se tienen en el área de procesamiento de imágenes hace posible la creación de algoritmos y sistemas que tratan de satisfacer estas necesidades, sin embargo la complejidad, el alto costo de adquisición ó las limitaciones en su funcionamiento hacen que estos sistemas originen un gran número de falsas alarmas y adicionalmente sean costosos. Sin embargo, el avance que ha experimentado la tecnología digital durante los últimos años ha permitido mejorar de manera significativa los sistemas de seguridad y control de acceso mediante la manipulación de imágenes y video digital. Esto ha hecho posible el desarrollo de sistemas eficientes de detección y rastreo de movimiento de personas y vehículos etc. con baja complejidad computacional, contribuyendo de esta manera a incrementar el uso de estos esquemas para el control de acceso y monitoreo de lugares de acceso restringido. Debido a que la demanda de esquemas de seguridad confiables y de bajo costo se ha incrementado durante las últimas dos décadas, varios sistemas han sido propuestos durante los últimos años, los cuales básicamente emplean la información contenida en secuencias de video, para detección y rastreo del movimiento de seres humanos, vehículos y otros objetos. Este artículo presenta el desarrollo de un algoritmo para la detección y rastreo del movimiento de seres humanos con baja complejidad computacional, basado en la estimación de vectores de movimiento. El sistema propuesto es capaz de diferenciar entre movimientos permitidos y no permitidos, sin la intervención de ningún operador humano. Los resultados experimentales proporcionados muestran un funcionamiento adecuado del sistema propuesto.

ABSTRACT

Nowadays the security requirements demands motion detection systems more reliable, precise, easy to handle and to able to adapt in any scene or monitoring area. Image processing area has been great advances and it makes possible the creation of algorithms and systems which can satisfy these requirements. However, complexity, high cost of acquisition or the limitation in their operation make these systems to have great number of false alarms. However the advances of the digital technology during the last several years have allowed important improvements in the development of security and access control systems by means of images and video processing. This fact has done it possible the development of low cost and efficient motion detection and tracking systems for applications that require access control and monitoring of access restricted places. Because the necessity of low cost and high performance motion detection and tracking systems has increased during the last years, several systems have been in the literature, based on the analysis on the information contained in image sequences to detect the movement of persons and vehicles, etc. This paper proposes a human and vehicle motion detection and tracking system, based on motion vectors, which is able to discriminate between allowed and not allowed movements without the intervention of any human operator. Evaluation results show the desirable features of proposed method.



INDICE

Capítulo 1. Introducción

1.1. Motivación	1
1.2. Objetivo General	2
1.3. Metas Particulares	2
1.4 Organización de la Tesis	3

Capítulo 2. Estado del Arte

Introducción.	5
2.1. Sistemas de Video vigilancia Actual	5
2.2. Reconocimiento Basado en Movimiento	6
2.3. Problema de Apertura	8
2.4. Problema de Correspondencia	9
2.5. Movimiento en Imágenes Espacio-Tiempo	10
2.6. Movimiento Común y Movimiento Relativo	11
2.7. Importancia del Movimiento.	12
2.8. Técnicas de Extracción de la Información de Movimiento	13
2.9. Correspondencia de Movimiento	13
2.9.1. Parametrización de Trayectorias	14
2.9.2. Flujo Óptico	15
2.9.2.1 Mecanismos de Rango Corto	15
2.9.2.2. Mecanismos de Rango Amplio	16
2.9.2.3. Aproximación Basada en el Gradiente	16
2.9.2.4. Aproximación Basada en la Correlación	17
2.9.2.5. Aproximación Basada en la energía Espacio-Temporal	18
2.9.3. Correlación	19
2.9.3.1. Segmentación	21
2.9.3.2. Evaluación y Comparación	21
2.10. Aplicaciones del Reconocimiento Basado en Movimiento	22
2.11. Sistemas Actuales	22
2.11.1. Sistema Basado en Aproximación Probabilística	22
2.11.2. Modelo Basado en la Silueta Humana	25
2.11.3. Detección de Personas Usando Reglas Difusas	28
2.12. Conclusiones	31

Capítulo 3. Sistema Propuesto

Introducción	32
3.1. Bosquejo General del Sistema	32
3.2. Codificador y Decodificador Genérico de Video	34
3.2.1. Estimación del Movimiento	35
3.2.2. Criterio de Mapeo	36



3.2.3. Algoritmos para la Estimación de Movimiento	37
3.2.4. Método de Búsqueda Exhaustiva	37
3.2.5. Búsqueda Logarítmica Bidimensional	38
3.2.6. Búsqueda Unidimensional en Paralelo Jerárquico	39
3.2.7. Estimación Jerárquica de Movimiento	40
3.3. Estimación de Vectores de Movimiento	42
3.4. Codificación Espacial	44
3.5. Cuantización	45
3.6. Velocidad y Dirección de un Vector	49
3.7. Filtrado de Vectores de Movimiento	50
3.8. Estimación de la Dirección del Movimiento	52
3.9. Criterios de Análisis	54
3.10. Magnitud	56
3.11. Conclusiones	57

Capítulo 4. Resultados Experimentales

<i>Introducción</i>	58
4.1. Sistema Propuesto Doctorado	58
4.1.1. Detección de Movimiento	58
4.1.2. Seguimiento del Movimiento	64
4.1.3. Pruebas de Detección Erróneas	65
4.2. Comparación de Sistemas	68
4.3. Conclusiones	69

Capítulo 5. Conclusiones Generales.

<i>Introducción</i>	70
5.1. Conclusiones	70
5.2. Investigación Futura	71

Capítulo 6. Reconocimiento de Personal.

<i>Introducción</i>	72
6.1. Biometría	72
6.2. Basado en Modelos	74
6.3. Basado en Siluetas	75
6.4. Revistas y Eventos que Publican el Tema	78
6.4.1. Revistas	78
6.4.2. Eventos	78
6.5. Investigadores Destacados	78
6.6. Conclusiones	79

Publicaciones	80
----------------------	----

Referencias Bibliográficas	I
-----------------------------------	---



CAPITULO 1

Introducción

1.1. Motivación.

Un área de investigación que ha sido muy extensa durante los últimos años es la detección y rastreo de objetos en movimiento utilizando la información contenida en secuencias de video y el conocimiento adquirido previamente acerca de la forma que poseen los diferentes tipos de objetos a nuestro alrededor. Las investigaciones referentes a detección y reconocimiento basadas en los sistemas de vídeo vigilancia están en auge; la seguridad en interiores así como en los alrededores de los edificios se ha incrementado gracias a la evolución tecnológica y a la disminución del costo de este tipo de tecnología, siendo la meta principal en esta área de investigación el lograr crear un método capaz de detectar y clasificar todo tipo de movimiento con un alto porcentaje de precisión para poder determinar la acción a realizar de acuerdo al tipo de objeto detectado.

Esta evolución se observa fácilmente si se piensa en la interpretación de videos de vigilancia; anteriormente se llevaba a cabo por un operario al que se le mostraba una gran cantidad de información en forma de video en uno o varios monitores; sin embargo, en la actualidad y cada vez presentes en más entornos se utilizan herramientas automáticas de procesamiento.

Se pueden señalar dos ventajas obvias derivadas de este último escenario: Incremento de la robustez de la seguridad y decremento de los costos económicos y del esfuerzo humano al permanecer por varias horas mirando monitores de seguridad.

Los sistemas de video inteligentes se desarrollan mediante la implementación de herramientas de análisis capaces de obtener información significativa de una secuencia de video. Para ello, es necesario procesar un flujo continuo de imágenes. Así, el ser capaz de detectar movimiento de algún objeto y poder llevar a cabo su seguimiento abre un abanico amplio de herramientas posibles para solventar situaciones frecuentes en esta área: objetos abandonados, movimiento en zonas en las que no debería producirse, dirección del movimiento, comportamientos humanos sospechosos, etc.

Por mencionar algunas de las limitaciones que se asocian a estos sistemas de video inteligentes son:

- i. **Fiabilidad:** se refiere a las restricciones de iluminación, cambios en la forma de caminar debido a factores extraordinarios, velocidad de procesamiento del algoritmo implementado, movimientos dentro de la escena de objetos secundarios, entorno y calidad de la cámara.
- ii. **Costo:** se refiere, sobre todo, al costo computacional. Éste presenta un umbral, más allá



del cual la eficiencia del sistema se ve seriamente perjudicada. Si se rebasa este umbral holgadamente podría incluso verse puesta en evidencia la eficacia de este mismo sistema. Por lo tanto, fiabilidad y costo plantean un conflicto que ha de resolverse con base a una serie de prioridades impuestas intrínsecamente por el propio entorno.

- iii. **Operatividad:** un sistema de vigilancia debe estar operativo durante periodos largos de tiempo en entornos cambiantes y debe ser capaz de almacenar solamente información previa y después del evento y no todo el tiempo de operación.

Con la realización de esta investigación se pretende disminuir las limitaciones referentes a la fiabilidad y al costo, seleccionando y comparando diferentes métodos de detección y seguimiento de objetos, una de las motivaciones principales del proyecto.

En lo referente a la aplicación de vigilancia, se propone la monitorización de la detección y seguimiento de personas, el cual pretende mostrar una secuencia de video y un mensaje de texto indicando el número de personas que se encuentran dentro del escenario en caso de detectarse movimiento relevante.

En segundo lugar, el proyecto se centra en una de las principales limitaciones de este tipo de métodos “*adaptabilidad*”, actualmente existen sistemas de detección robustos capaces de reconocer personas específicas con un gran grado de precisión, sin embargo su campo de aplicación es limitado debido a que fueron creados para una acción específica, trabajando en robustecer el método de detección en vez de la aplicación es posible lograr una mayor adaptabilidad, ampliando su robustez y campos de acción.

1.2. Objetivo General

El objetivo de esta investigación es proponer un método de detección de movimiento de baja complejidad computacional y con gran posibilidad de adaptabilidad, presentando un bajo índice de error durante la detección, robustez en su funcionamiento bajo diferentes escenarios controlados; principales problemas que presentan gran parte de los sistemas de detección de movimiento actuales.

1.3. Metas Particulares

- i. Analizar y comprender diferentes técnicas de detección y reconocimiento de objetos en movimiento.
- ii. Investigación acerca de los sistemas actuales y sus métodos de detección con el fin de entender el estado actual del área.
- iii. Desarrollo y simulación de un método capaz de superar las limitantes de precisión y



confiabilidad presentadas en otros sistemas propuestos para la detección de movimiento, logrando desarrollar un sistema capaz de clasificar los movimientos dentro de la escena como significativo o no significativo.

- iv. Investigación del funcionamiento y confiabilidad del cálculo de vectores de movimiento a través del uso de macrobloques, basándose en el estándar de compresión MPEG2.
- v. Desarrollo y simulación de un método basado en la dirección de los vectores de movimiento calculados a través de la técnica de macrobloques sobre una secuencia de video para determinar la magnitud y dirección del objeto dentro de la escena y clasificarlo.
- vi. Verificación del funcionamiento del sistema bajo escenarios controlados para probar la confiabilidad y precisión en la detección.

1.4. Organización de la Tesis.

La tesis se realiza en 6 capítulos y dos anexos, que nos muestran los avances actuales en sistemas de detección de movimiento, así como una nueva propuesta que surge a partir de las necesidades que presentan algunos sistemas de detección de movimiento. Al final de la tesis se encuentran las publicaciones realizadas durante esta investigación y referencias bibliográficas.

Capítulo 1. En este capítulo se da una breve introducción acerca de la motivación y objetivos generales de la investigación así como un bosquejo actual de los sistemas de detección de movimiento y sus limitantes actuales.

Capítulo 2. En este capítulo se realiza una revisión de los conceptos básicos de detección de movimiento, el estado del arte actual teniendo como base principal una comparación directa del funcionamiento del sistema visual humano y el flujo óptico generado a través de vectores de movimiento.

Capítulo 3. En este capítulo, discutiremos a fondo cada una de las etapas del método propuesto, sus principales ventajas y limitantes, la importancia del uso de macrobloques para el cálculo de los vectores de movimiento, la adaptabilidad contenida en el método propuesto y su aplicación en el área de video vigilancia.

Capítulo 4. En este capítulo se evalúa el sistema propuesto haciendo una primera



comparación con otros sistemas desde el punto de vista de detección de movimiento relevante en la escena.

Capítulo 5. Conclusiones Generales y Trabajos Futuros.

Capítulo 6. Introducción teórica al estado del arte actual para el reconocimiento de personas con la finalidad de proporcionar una cuadro actual, para los posibles trabajos futuros en el tema.

Publicaciones

Referencias Bibliográficas



CAPITULO 2

Estado del Arte

Introducción

En este capítulo se realiza una revisión de los avances en sistemas de video vigilancia, principales campos de acción, así como conceptos básicos de detección de movimiento, teniendo como base principal una comparación directa del funcionamiento del sistema visual humano con la visión por computador. Posteriormente se explican los principales problemas a los que se enfrentan las diferentes técnicas de extracción de información de movimiento como es el problema de apertura y el problema de correspondencia. Puntos importantes como el movimiento relativo, el movimiento común y su nivel de importancia para la extracción de información son también abarcados en este capítulo, seguido de una clara explicación de las principales técnicas de extracción de movimiento; finalizando con una breve introducción al campo de aplicaciones de los métodos de reconocimiento basados en el movimiento.

2.1. Sistemas de Video vigilancia actual

El monitoreo a través de video es hoy en día el sistema más utilizado para la vigilancia de lugares como edificios, bancos, oficinas gubernamentales, centros comerciales, prisiones, aeropuertos etc. Actualmente gran parte de estos sistemas llegan a tener un alto costo de mantenimiento además de requerir una intervención humana las 24 horas al día, haciendo a este tipo de sistemas en gran parte inefectivos. Existen sistemas como los arreglos de CCTV que requieren de guardias de seguridad que estén mirando el arreglo todo el tiempo [1], además de utilizar gran cantidad de cintas de video para poder grabar toda la información que se registra durante el día, representado así un alto costo de almacenamiento, un nivel bajo de eficiencia, ya que para poder detectar algún tipo de irregularidad es necesario revisar toda la cinta para poder obtener las imágenes más importantes del evento detectado, esto representa una pérdida de tiempo y baja eficiencia ya que este tipo de sistemas de vigilancia no garantizan que todos los eventos irregulares serán detectados debido a que en algún momento los guardias de seguridad pueden tener algún tipo de distracción ocasionando que giren su vista hacia otro objeto, dando como resultado, un abandono momentáneo del monitoreo el cual podría ser importante.

Actualmente se cuenta con herramientas como la visión artificial que consiste en enseñar a las computadoras a “ver”, permitiendo que ciertos procesos interpreten lo que ellas “ven” y así ser capaces de tomar decisiones para dar una solución a diversos problemas [3], dando la posibilidad de crear modernos sistemas de monitoreo que disminuyan la intervención humana e incrementen la precisión en la detección de movimiento dentro del área de vigilancia.



Fig. 2.1. Cambio de una imagen de video a una imagen digital.

2.2. Reconocimiento Basado en Movimiento

El reconocimiento de un objeto o su movimiento está basado en el movimiento que este genera en una secuencia de imágenes. El movimiento juega un papel muy importante en el sistema visual humano.

Nosotros tenemos la habilidad de reconocer a distancia a una persona caminando por su modo de andar, gestos particulares en las manos, pasos de baile, aves volando y todo lo que esté constituido por una secuencia compleja de movimiento.

La percepción del movimiento nos ayuda a reconocer diferentes objetos y sus movimientos en la escena deduciendo su profundidad relativa, rigidez, etc.

Nuestro **HVS** es muy sensitivo al movimiento y tendemos siempre a enfocar nuestra atención a objetos móviles [4]. Tomado en cuenta los sistemas de cómputo presentes, podemos establecer una analogía entre la visión humana y la visión por computador, debido a que ambos cuentan con un elemento sensorial (el ojo y la cámara) y un procesador de la información (el cerebro y la computadora) [3]. En psicología, la percepción del movimiento ha sido estudiada extensivamente usando las pantallas móviles de luz Johansson (**MLD**). Los **MLDs** consisten de puntos brillantes adheridos a las articulaciones de un actor con vestimenta negra y moviéndose en frente de un fondo negro.

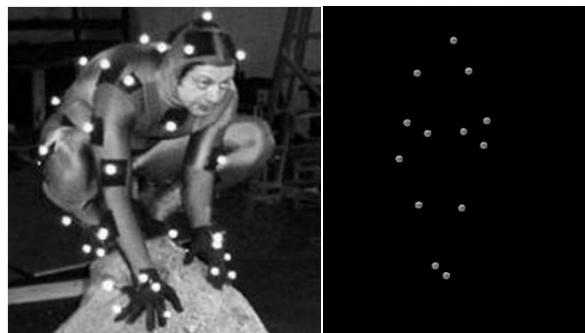


Fig. 2.2. Pantallas de Luz Móviles Johansson.



La colección de puntos entrega solo información **2D** y no información estructural, debido a que los puntos no están conectados. Aquí el grupo de puntos estáticos permanecen sin sentido para los observadores mientras que su movimiento relativo crea una vívida impresión de una persona caminando, corriendo, danzando, etc. El género de una persona e incluso el modo de andar de un amigo pueden ser reconocidos basándose solamente en estos puntos. También ha sido probado que invirtiendo los **MLDs** (de cabeza) no son usualmente reconocidos incluso para los movimientos más simples. Esto sugiere que la familiaridad de un observador con un movimiento particular juega un papel muy importante en la facilidad con la cual uno puede reconocer dicho movimiento. Así un movimiento invertido no es, natural ni familiar, por lo que es más difícil de reconocer [4].

Hay dos teorías acerca de la interpretación de los **MLD**. En la primera, la gente usa la información de movimiento en el **MLD** para recuperar la estructura **3D** y subsecuentemente usar esa estructura para reconocer el objeto. Mientras que con la segunda teoría la información de movimiento es usada directamente para reconocer un movimiento, sin recuperar la estructura.



Fig. 2.3. Secuencia de Video mostrando los puntos de luz y la imagen real.

Ha sido de gran interés en la visión por computador, la determinación de estructuras del movimiento. Con esta finalidad las coordenadas **3D** de puntos sobre el objeto móvil y su movimiento **3D** son recuperadas, desde una secuencia de cuadros. Este problema es formulado en términos de sistemas de ecuaciones lineales y no lineales dando posiciones **2D** de puntos móviles entre unos pocos cuadros. En esta aproximación, la estructura **3D** no es suficiente para un reconocimiento robusto y exacto ya que la reconstrucción es sensible al ruido y múltiples pistas como movimiento, textura, profundidad, etc. son necesarios.

Otra aproximación para el análisis de movimiento trata con el uso directo de la información de movimiento para su reconocimiento. En esta aproximación el énfasis no es necesariamente sobre la estructura estática y la información del movimiento no es extraída de un cuadro en un tiempo. Más bien, se usa una secuencia que contenga un gran número de cuadros para extraer la información de movimiento continuamente. El uso de secuencias largas permite reconocer movimientos de niveles altos, como caminar o correr. Estos movimientos consisten de una serie coordinada y compleja de eventos que no pueden ser entendidos mirando solo pocos cuadros. Además, movimientos más complejos pueden ser examinados a un nivel más apropiado [4] [5].



El reconocimiento basado en el movimiento es una aproximación que favorece el uso directo de la información de movimiento extraída de una secuencia de imágenes para el propósito de reconocimiento. El conocimiento acerca del objeto o movimiento es usado para construir modelos que servirán en el proceso de reconocimiento. Existe una distinción entre *reconocimiento basado en el movimiento* y *reconocimiento de movimiento*: el reconocimiento basado en el movimiento es una *aproximación* general que favorece el uso de la información de movimiento para el propósito de reconocimiento, mientras que, reconocimiento de movimiento es un logro que puede ser alcanzado con esta aproximación. En esta aproximación incluimos también el rastreo usando modelos, es decir el rastreo de un movimiento nos da no solo la forma sino también un modelo de movimiento. Y con esto se entiende que si un rastreo apropiado es logrado entonces el reconocimiento también se logra [4] [3].

Hay dos pasos principales en el reconocimiento basado en el movimiento, el primer paso consiste en encontrar una apropiada representación para los objetos o movimientos que queremos modelar, desde el movimiento en la secuencia de imágenes. Estas representaciones pueden ser relativamente de bajo nivel. De entrada la trayectoria de un punto particular sobre un objeto moviéndose, una velocidad o dirección a través de la secuencia, la cual puede, si es necesario, ser organizada dentro de representaciones de alto-nivel. El segundo paso consiste de la igualación de alguna entrada desconocida con un modelo. Los métodos aquí son más estándar, y son a veces técnicas comunes de clasificación de patrones. El término “reconocimiento” es a veces usado como un equivalente de clasificación. Vectores característicos son clasificados, es decir, asociados con un grupo representativo de acuerdo a una distancia medida para ese grupo. Los grupos pueden ser también agrupados o divididos dependiendo de una distancia predefinida medida o de acuerdo a algún parámetro introducido por el usuario [4].

2.3. Problema de Apertura

Hasta ahora hemos aprendido que la estimación de movimiento esta cercanamente relacionada con los cambios espacio-temporales de los valores de una imagen, los cuales pueden ser fácilmente derivadas con operadores locales que calculen la derivada espacial y temporal. Tal operador solo “ve” en un pequeño sector (igual al tamaño de su máscara) del objeto observado. Nosotros podríamos observar este efecto colocando una máscara o apertura sobre la imagen [5]. La figura 2.4 muestra la base física del problema de apertura, mostrando el movimiento de una barra observada a través de una apertura que es pequeña comparada con la dimensión de la barra. La figura 2.4a muestra la posición de la barra en tiempo t . La figura 2.4b muestra la posición de la barra en tiempo $t + \Delta t$, asumiendo que la barra se está moviendo hacia el lado derecho con una velocidad de $\Delta x / \Delta t$, y la figura 2.4c muestra la posición de la barra en tiempo $t + \Delta t$ asumiendo que la barra se mueve hacia abajo con una velocidad $\sqrt{2} \times \Delta x / \Delta t$. Cuando la barra se mueve solo a través de la pequeña apertura, no podemos diferenciar entre la situación en la figura 2.4b y 2.4c.

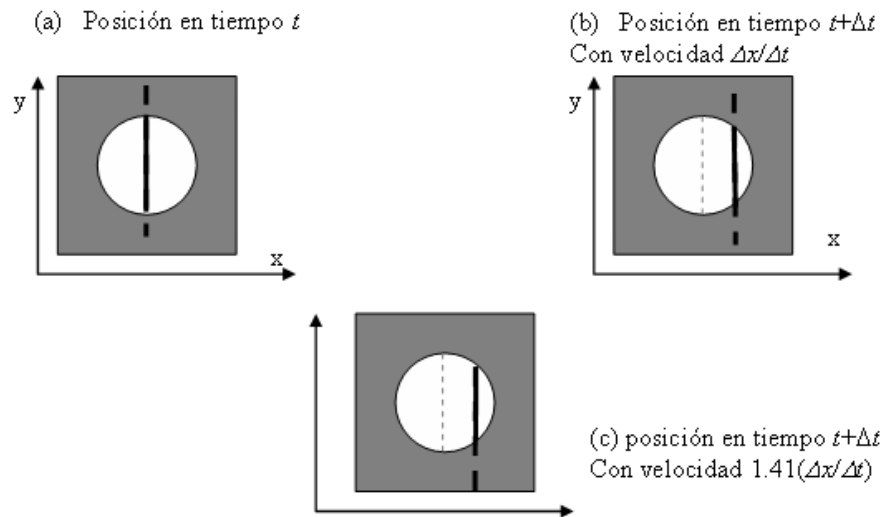


Fig. 2.4. Base Física Del Problema De Apertura.

De manera más general no es posible calcular la verdadera velocidad observando una pequeña vecindad de un punto que está experimentando un movimiento. La única información directamente aplicable de la medición local es la componente de velocidad normal para el contorno fundamental.

Generalizando el problema de apertura existe en regiones de la imagen que tienen gradientes de intensidad fuertemente orientados por ejemplo en los bordes. Por el otro lado, regiones de la imagen que tienen gradientes de intensidad más fuertes de un nivel más alto, tales como esquinas o regiones de texturas no sufren del problema de apertura [7].

2.4. Problema de Correspondencia

El problema de correspondencia es causado por el hecho de que no podamos encontrar el punto correspondiente en un borde en la imagen subsiguiente de una secuencia de imágenes, porque no tenemos ningún medio para distinguir los diferentes puntos en un borde [5]. En otras palabras, dado n cuadros tomados a diferentes instantes de tiempo y dado m puntos in cada cuadro, proyectar un punto en un cuadro a otro punto en un cuadro siguiente tal que ninguno de los dos puntos se proyecten dentro del mismo punto.

La ocurrencia de oclusión y desocclusión también suma la dificultad al problema. Es necesario introducir restricciones para limitar el espacio de búsqueda. Las restricciones incluyen uniformidad próxima, velocidad máxima, cambios de velocidad, pequeña suavidad de movimiento, movimiento común, equivalente consistente, etc. Una importante característica en el problema de correspondencia es el convertir las restricciones cualitativas arriba mencionadas en expresiones cuantitativas [4] [5].

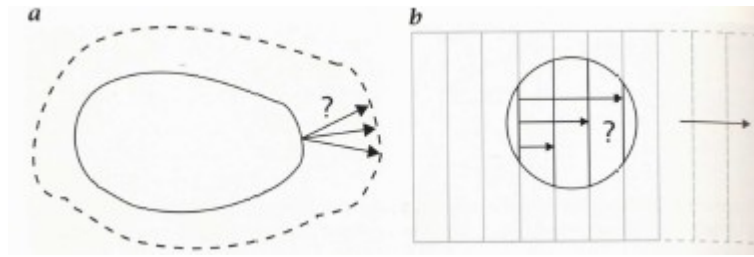


Fig. 2.5. Ilustración Del Problema De Correspondencia: (a) Un Objeto 2D Deformable; (b) Rejilla Regular.

2.5. Movimiento en Imágenes Espacio-Tiempo

Podemos pensar en una imagen **3D** espacio-temporal como una pila de imágenes consecutivas que podrían ser representadas como una imagen cúbica como se muestra en la figura 2.6. En cada cara visible del cubo proyectamos una sección cruzada en la dirección correspondiente, así una diapositiva xt es mostrada en la cara superior y una diapositiva yt en la cara derecha del cubo. Las diapositivas fueron tomadas a profundidad, marcadas por la línea blanca en la cara frontal, como muestra la última imagen de la secuencia [5].

En una imagen espacio-temporal un píxel se extiende a un volumen, es decir representa un valor de gris en un elemento pequeño de volumen con las extensiones Δx , Δy y Δt . Aquí se enfrentan los límites de nuestra imaginación visual, cuando tratamos de capturar el verdadero dato **3D**. Además, necesitamos una apropiada representación de los datos para hacer a las características esenciales de interés visible [8].

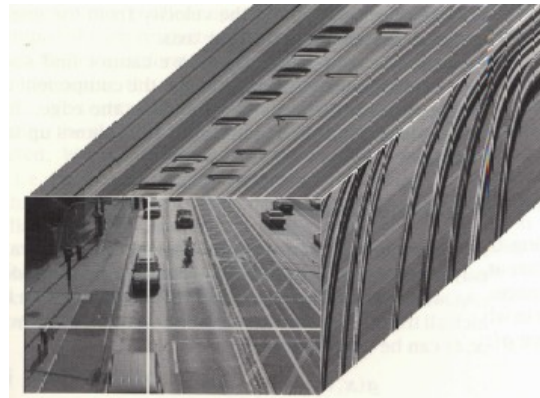


Fig. 2.6. Secuencia De Imágenes 3D, Representada Como Una Imagen Cúbica. El Eje Del Tiempo Corre Dentro De La Profundidad, Apuntando Hacia El Observador.

Para analizar el movimiento en imágenes espacio-temporales, primero consideramos un ejemplo simple con una coordenada en el espacio y una en el tiempo, figura 2.7a. Una ausencia de movimiento de un objeto **1D** muestra una estructura vertical de valores de gris. Si el objeto se mueve entonces este es desplazado de imagen a imagen formando una estructura inclinada de valores de gris. La velocidad está directamente vinculada a la orientación en imágenes espacio-temporales [8]. En un simple caso de imágenes espacio-temporales **3D** la velocidad está dada por:



$$v = -\tan \varphi. \quad (2.1)$$

Donde φ es el ángulo entre el eje t y la dirección en la cual los valores de gris son constantes. El signo menos en la ecuación 2.1 es porque los ángulos son positivos a contra reloj, la extensión a dos dimensiones espaciales es directo e ilustrado en la figura 2.7b.

$$v = \begin{bmatrix} \tan \varphi_x \\ \tan \varphi_y \end{bmatrix}. \quad (2.2)$$

Los ángulos φ_x y φ_y son definidos análogamente por el ángulo entre los componentes x y y de un vector en dirección de los valores constantes de gris en el eje t .

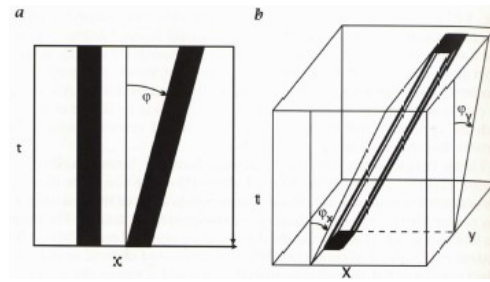


Fig. 2.7. Imágenes Espaciotemporales: (a) Imagen 2D Con Una Coordenada Espacial y Temporal; (b) Imagen 3D.

2.6. Movimiento Común y Movimiento Relativo

Cutting y Proffitt [11] mostraron que el movimiento relativo es un importante aspecto en la percepción visual humana. La percepción del movimiento absoluto de un objeto puede ser dividido en dos componentes de movimiento: movimiento común y movimiento relativo.

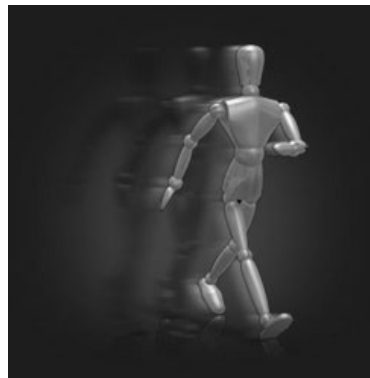


Fig. 2.8. Diseño Representado un Movimiento Relativo.



El movimiento común es la percepción global de movimiento de un objeto relativo al observador, donde cada elemento del objeto comparte el movimiento, mientras que el movimiento relativo es el movimiento de un elemento con respecto a otros elementos. El movimiento absoluto de un punto, es decir su trayectoria es definida como la suma de los movimientos comunes y relativos. Cutting y Proffitt [11] crearon la hipótesis de que una reducción del proceso es aplicada para ambos movimientos, relativo y común y que el que se resuelve primero, explican, es el que perciben los observadores. Encontraron que el movimiento relativo entre los elementos de un objeto, primero deben ser extraídos, es decir, el movimiento relativo es más relevante para el entendimiento del movimiento y reconocimiento de un objeto que el movimiento común [11].

Múltiples trayectorias pueden ser usadas para calcular el movimiento relativo. Por ejemplo ángulos relativos pueden ser calculados entre pares de puntos relativos para un eje dado o como la unión de ángulos entre tres puntos en cada cuadro. La diferencia entre los ángulos es entonces determinada entre cuadros sucesivos, y puede ser considerada como velocidades angulares si suponemos que el tiempo entre cada cuadro es constante. Un problema con trayectorias múltiples, es que, el cálculo del movimiento relativo para cada elemento es combinatoriamente inestable. También, el movimiento relativo entre dos elementos no relacionados podría no ser muy relevante. Métodos heurísticos pueden ser desarrollados para seleccionar apropiadamente movimientos relativos entre todos los posibles movimientos [4].

2.7. Importancia del Movimiento

Los casos de movimiento son definidos como cambios significantes o discontinuidades en el movimiento. Un cambio brusco de dirección, un alto o una pausa, pueden proporcionar pistas importantes acerca del tipo de objeto y su movimiento. El estudio de un sub-muestreo temporal en algunas secuencias mostró que eligiendo imágenes donde ocurrieran pocas actividades entre eventos de movimiento, era mejor para la comprensión en lugar de tomar imágenes en periodos constantes de tiempo. Los casos de movimiento son usualmente detectados por la presencia de discontinuidades, que podrían ser encontradas, por ejemplo calculando la escala-espacio de una curva de velocidad [4].

Basándose en consideraciones psicológicas, Engel y Rubin [23] describen los cambios significativos de movimiento como movimientos de fronteras, de los cuales ellos encuentran cinco tipos: inicios suaves, términos suaves, pausas, impulsos iniciales e impulsos terminales. Todos ellos considerados como casos de movimiento que particionaron un movimiento global dentro de sus partes psicológicas [4].

Goddard en su trabajo relativo a movimiento humano, uso cambios en la velocidad rotacional, y la segmentación de cuerpos junto con cambios en la dirección, como casos de movimiento, por ejemplo, seis rangos de valores de velocidad angular fueron usados, junto con cuatro cuantizaciones de la orientación (≤ -200 ; $[-200, -100]$; $[-100, 0]$; $[0, 100]$; $[100, 200]$; ≥ 200 grados/segundo), tomando para los cuatro cuadrantes el valor de 90° . Cualquier cambio en la orientación o velocidad angular constituye un caso de movimiento, el cual dispara alguna acción en su sistema de conexión [14].



2.8. Técnicas de Extracción de la Información de Movimiento

Hay dos pasos principales en el reconocimiento basado en el movimiento. Primero se debe definir una representación del movimiento, para que la información apropiada del movimiento sea extraída de la secuencia de imágenes y organizada dentro de modelos de movimiento. Segundo, la representación de un modelo desconocido deberá ser comparada con un modelo almacenado para igualación o reconocimiento [4].

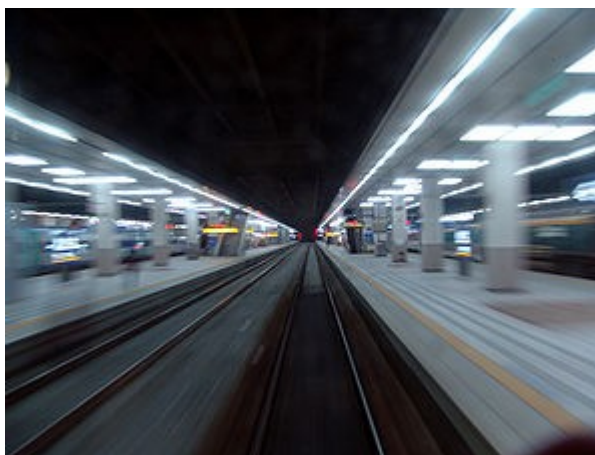


Fig. 2.9. Información de Movimiento.

Hay dos métodos generales para extraer movimiento bidimensional: correspondencia de movimiento y el flujo óptico. La correspondencia de movimiento es vinculada a la igualación de características muestreadas a través del tiempo, mientras que el flujo óptico consiste en el cálculo del desplazamiento de cada píxel entre cuadros [5].

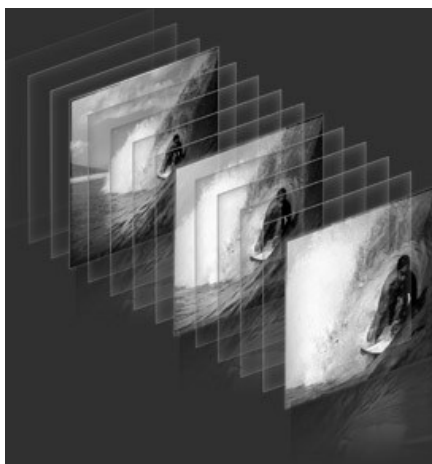


Fig. 2.10. Extracción de Movimiento Bidimensional.

2.9. Correspondencia de Movimiento

La correspondencia de movimiento trata con la extracción de puntos importantes, rasgos característicos que pueden ser rastreados en tiempo. Esta correspondencia para



múltiples cuadros resulta en lo que es llamado trayectoria del movimiento, es decir, una secuencia de localidades (x_i, y_i) para $i=1, \dots, n$ donde n es el número de cuadros en la secuencia. Esta trayectoria de movimiento puede ser considerada un valor de vector, es decir, en cada tiempo tendremos dos valores x y y , sin embargo un valor simple de vector es mejor para cálculos y por consiguiente la parametrización de las trayectorias es necesaria. Las trayectorias pueden ser parametrizadas de varias maneras, por ejemplo, velocidad y dirección, las velocidades v_x y v_y y la curvatura espacio-temporal. Representaciones parametrizadas pueden ser analizadas para identificar importantes eventos de movimiento [4]. Jenkins [9] propuso un algoritmo para rastrear una localización **3D** de puntos desde una vista estero de la localización **2D** de las imágenes, el cual está basado en una suposición general de suavidad, afirmando que la localización, velocidad escalar y dirección de movimiento de un punto dado son relativamente iguales de un cuadro a otro.

2.9.1.1. Parametrización de Trayectorias

Trayectorias simples usualmente no proporcionan suficiente información para el reconocimiento y como se mencionó antes estos son valores de vectores, los cuales no son tan fáciles de manipular como con valores simples. La parametrización de trayectorias es entonces muy útil. Una representación es la velocidad relativa de x y y en el tiempo, estos son calculados como sigue:

$$v_x = \frac{x_i - x_{i-1}}{\Delta t}, \quad v_y = \frac{y_i - y_{i-1}}{\Delta t}. \quad (2.3)$$

Donde (x_{i-1}, y_{i-1}) y (x_i, y_i) son la localización de un punto en los cuadros $i-1$ y i . La velocidad es un desplazamiento invariante, pero no de rotación invariante o escala invariante. Velocidad y dirección son otras parametrizaciones útiles, las cuales están definidas como:

$$s_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \quad (2.4)$$

$$d_i = \arctan\left(\frac{y_{i+1} - y_i}{x_{i+1} - x_i}\right)$$

Donde s_i es la velocidad y d_i es la dirección de un punto en el cuadro i ; la velocidad y dirección son invariantes en desplazamiento y rotación. Además la dirección es de escala invariante, pero no de aceleración. La dirección sin embargo es más sensitiva al ruido, debido a la operación no lineal del $\arctan$. Parametrización de la velocidad y parametrización de la aceleración y dirección son bastante fáciles de calcular desde las trayectorias y generan curvas que son fáciles de interpretar. Sin embargo no es muy obvio como combinar estas trayectorias para un alto nivel de interpretación [4].

La curvatura espacio-temporal k de una trayectoria es otra representación común. Esta es determinada como sigue:



$$k = \frac{\sqrt{A^2 + B^2 + C^2}}{((x')^2 + (y')^2 + (t')^2)^{3/2}}. \quad (2.5)$$

Donde:

$$A = \begin{vmatrix} y' & t' \\ y'' & t'' \end{vmatrix}, \quad B = \begin{vmatrix} t' & x' \\ t'' & x'' \end{vmatrix}, \quad C = \begin{vmatrix} x' & y' \\ x'' & y'' \end{vmatrix}. \quad (2.6)$$

La notación $|\cdot|$ denota el determinante. Una aproximación discreta es usada para calcular las derivadas, por ejemplo $x'_i = x_i - x_{i-1}$ y $x''_i = x'_i - x'_{i-1}$. Se asume que Δt es constante por lo que $t'=1$ y $t''=0$. Se ha mostrado en geometría diferencial que cualquier espacio curvo es completamente definido por su curvatura y torsión, hasta una rotación y translación rígida. La curvatura tiene ventaja sobre otras parametrizaciones de trayectorias, ya que una función simple es capaz de capturar información, como opción de las dos funciones con v_x y v_y o aceleración y dirección. La curvatura de una trayectoria fue usada por Allemn y Dyer [10] para detección de movimientos cíclicos.

Cualquier tipo de trayectoria puede ser además procesada para ser usada en reconocimiento. Por ejemplo la escala-espacio puede ser calculada suavizando progresivamente una curva usando mascarar gaussianas y detectando cruces por cero. Trayectoria de escala-espacio fueron usadas en severos estudios para describir el reconocimiento del movimiento humano [4] [10].

2.9.2. Flujo Óptico

La percepción del movimiento podría ser tratada como un proceso de dos pasos: Medición del movimiento 2D proyectado en la superficie de la imagen y la Interpretación del movimiento 2D para dibujar las conclusiones acerca de la estructura 3D y movimiento de la escena. Tomando estas dos consideraciones podemos decir que podría haber (y posiblemente hay) *dos distintos mecanismos para medir el movimiento 2D* proyectado en la mayoría de los animales avanzados incluyendo el humano [7] [15].

2.9.2.1. Mecanismos de Rango-Corto

Estos mecanismos resultan dentro de una caracterización inmediata de la imagen en movimiento como un campo de velocidades 2D y han sido referidos como *la aproximación al flujo de la imagen*. En esta aproximación la etapa de medición es responsable de construir un campo de flujo óptico 2D de imágenes variantes en el tiempo de la escena y, la etapa de interpretación toma el campo de velocidades como su entrada para extraer información acerca de la profundidad y la velocidad en cada punto del campo visual [7].



2.9.2.2. Mecanismos de Rango-Amplio

Es el principio de la aproximación “*discrete snapshots*”. Las características salientes descubiertas en dos discrete snapshots son igualadas para recuperar el desplazamiento espacial de cada característica. *La base de la aproximación de Ullman* es también conocida como la aproximación “discrete snapshots”. En esta aproximación la etapa de medición es responsable de identificar características distintivas en dos o más “*discrete snapshots*” de la escena dinámica igualándolos a través de los snapshots, Esto se conoce como igualación de muestra. La salida de la etapa de medición, que se puede visualizar como un mapa que da la posición de varias características en un grupo de imágenes tomadas en instantes discretos de tiempo, al final la etapa de interpretación, usa esta información para derivar las posiciones **3D** de todos los puntos que corresponden a las muestras y velocidades de los cuerpos rígidos que contienen estos puntos [7].

Tomando en cuenta la información que nos proporcionan estos dos mecanismos de medición de movimiento, podemos concluir que *El flujo de la imagen* describe en cada punto de la superficie de la imagen una proyección **2D** de la velocidad instantánea **3D** del punto correspondiente en la escena y el *flujo óptico* es una representación que es asociada con cada punto del campo retinal es decir una velocidad de vector **2D** que representa la proyección de la velocidad **3D** de un punto correspondiente en la escena. concluyendo que el *flujo de la imagen* y el *flujo óptico* generalmente no son iguales, para poder tener una idea más clara al respecto tomemos en cuenta el siguiente ejemplo: una imagen con escenario estacionario con respecto al plano de la imagen e iluminado por una fuente de luz con movimiento no uniforme. Debido a que la escena es *estacionaria* el flujo de la imagen será cero en toda la imagen, pero debido a que la fuente de luz se mueve, sin embargo, el brillo en los patrones de la imagen varía resultando dentro de un campo de flujo óptico diferente de cero. En otras palabras, el *campo de flujo óptico* es la distribución de velocidades aparentes que pueden ser asociadas con la variación de patrones de brillo de la imagen [7]. Existen diferentes técnicas para detectar el campo de flujo óptico en una secuencia de imágenes, pero, todas ellas comprenden solo tres diferentes tipos de aproximaciones:

2.9.2.3. Aproximación Basada en el Gradiente

Las técnicas de detección que se encuentran dentro de esta aproximación típicamente trabajan sobre la suposición de la conservación de la intensidad de la imagen. Algunas técnicas usan una función de intensidad de imagen mejor que la intensidad de la imagen por si misma. Esta técnica supone que para un punto dado en la escena la intensidad I en el punto de la imagen correspondiente permanece constante sobre el tiempo. Esto es, si un punto P en la escena (figura 2.6) se proyecta en el punto (x,y) de la imagen en tiempo t y dentro del punto $(x+\delta x, y+\delta y)$ de la imagen en tiempo $(t+\delta t)$ podemos escribir:

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t). \quad (2.7)$$

Expandiendo el lado derecho usando las series de Taylor sobre (x,y,t) e ignorando los términos de segundo y más alto orden, obtenemos:



$$I(x + \delta x, y + \delta y, t + \delta t) = I(x, y, t) + \delta x \frac{\partial I}{\partial x} + \delta y \frac{\partial I}{\partial y} + \delta t \frac{\partial I}{\partial t}. \quad (2.8)$$

Combinando las dos ecuaciones (2.7) y (2.8) resulta en la siguiente expresión:

$$\delta x \frac{\partial I}{\partial x} + \delta y \frac{\partial I}{\partial y} + \delta t \frac{\partial I}{\partial t} = 0. \quad (2.9)$$

Dividiendo todo por δt ; denotando las derivadas parciales de I por I_x , I_y e I_t y denotando el vector de velocidad local por (u, v) obtenemos:

$$I_x u + I_y v + I_t = 0. \quad (2.10)$$

Así, la restricción local proporciona una ecuación lineal en las variables u y v . Como consecuencia, la velocidad del vector (u, v) no puede ser determinada localmente sin aplicar restricciones adicionales. La ecuación (2.10) también es conocida como la línea de restricción de movimiento.

Esta técnica sufre de problemas de apertura como un sistema bajo restricciones compuesto de una ecuación simple en dos variables, particularmente los dos componentes escalares del vector de velocidad local. Sin embargo esta técnica puede ser utilizada en regiones donde el problema de apertura no existe [7] [15].

2.9.2.4. Aproximación Basada en la Correlación.

Estas técnicas usan la suposición de la conservación de la distribución de intensidad local. Usualmente trabajan con dos imágenes sucesivas de la escena variante en el tiempo. Para cada píxel de la primera imagen, buscan un píxel equivalente en la 2ª imagen, la salida es un vector de desplazamiento para cada píxel en la primera imagen. La búsqueda consiste, encontrar el mejor equivalente para el píxel bajo consideración entre algunos píxeles candidatos en la segunda imagen, los píxeles candidatos son determinados usando una búsqueda estratégica burdo-fino o imponiendo una limitación física en el desplazamiento máximo que puede tomar lugar entre dos imágenes consecutivas en una secuencia de imagen, La región cubierta por los píxeles candidatos es llamada la *ventana de búsqueda*. El mejor equivalente se basa en maximizar una media equivalente. Esta medida es usualmente la correlación entre una pequeña ventana de correlación alrededor del píxel bajo consideración y una ventana correspondiente alrededor del equivalente candidato.

En esta forma más ingeniosa el problema de apertura podría hacer que un escenario donde varios píxeles en W_s tengan la misma medida equivalente mostrada sobre una línea recta. Lo mejor que puede ser estimado en tales casos es la componente del vector de desplazamiento a lo largo del gradiente fundamental de intensidad. Esto es, normal al borde fundamental de intensidad, descrito por d_n en la figura 2.11 [7].

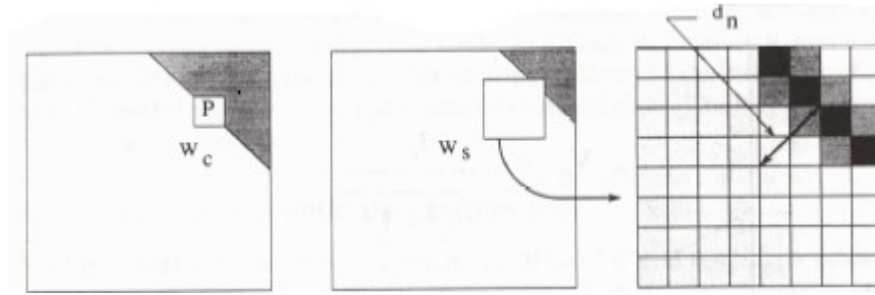


Fig. 2.11. El Problema De Apertura En Los Métodos Basados En La Correlación: Imagen.

2.9.2.5. Aproximación Basada en la Energía Espacio-Temporal

Es bien entendido que las frecuencias espacial y temporal están relacionadas con la velocidad de un estímulo móvil por:

$$\omega_t = \omega_x v + \omega_y u. \quad (2.11)$$

Las variables u y v se refieren a las 2 componentes ortogonales del vector de velocidad 2D, ω_x y ω_y se refieren a las 2 componentes ortogonales de la frecuencia espacial del estímulo visual y ω_t se refiere a la frecuencia espacial correspondiente.

De acuerdo con la ecuación (2.11), si las frecuencias espacial y temporal de una señal visual son graficadas en el espacio ω_x , ω_y , ω_t estarán situados sobre un plano. En otras palabras, la energía de un estímulo móvil es contenida en frecuencias espacio-temporales que se sitúan sobre un plano en el espacio ω_x , ω_y , ω_t . La velocidad de la imagen (u, v) puede ser encontrada si determinamos este plano [7].

Varios métodos han sido desarrollados, sin embargo exactitud y medidas densas son difíciles de alcanzar. Un problema con el flujo óptico en general es que este es susceptible al problema de apertura como se vio en las aproximaciones mencionadas anteriormente, ya que, en algunas condiciones, solo permite el cálculo preciso del flujo normal, es decir, la componente paralela al gradiente. Este es también propenso al suavizado excesivo de la frontera, es decir, píxeles del fondo alrededor de la frontera de los objetos podrían tener un valor de flujo diferente de cero. Presenta también problemas con múltiples objetos móviles, donde la segmentación puede ser difícil de alcanzar y múltiples velocidades legítimas en una pequeña vecindad que podrían ocurrir como un resultado de transparencia. A pesar de todo, el flujo óptico ha sido satisfactoriamente usado como una fuente de información [4] [15].

Existen también otras técnicas para extraer información de movimiento de las imágenes las cuales no se consideran casos tan particulares como lo las dos técnicas anteriores (método de correspondencia, flujo óptico) por estar relacionadas dentro de la aplicación de dichas técnicas, sin embargo es importante mencionarlas para tener una idea más clara acerca de ellas.



2.9.3. Correlación.

La técnica de correlación es una aproximación la cual se origina de analizar el desplazamiento entre dos imágenes consecutivas. Para encontrar un rasgo característico de la primera imagen con la segunda, tomando la imagen uno y comparando esta con la imagen dos dentro de un cierto rango. Con este rango buscamos la posición de la mayor similitud entre las dos imágenes [5].

La correlación $g(x)$ entre dos funciones es:

$$g(x, y) = h(x, y) \circ f(x, y) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} f^*(i, j) h(x+i, y+j). \quad (2.12)$$

Donde f^* es el complejo conjugado el cual en el caso de imágenes, al ser números reales f^* es igual a f . en la figura 2.7 puede verse la correlación entre dos señales. El funcionamiento es parecido al de la convolución. Así, se tienen las siguientes funciones:

$$f(x) = \begin{cases} 1 & 0 < x < 1 \\ 2 & 1 < x < 3 \\ 1 & 3 < x < 4 \\ 0 & \text{en cualquier otro caso} \end{cases} \quad (2.13)$$

$$h(x) = \begin{cases} 2 & -1 < x < 1 \\ 0 & \text{en cualquier otro caso} \end{cases}$$

Observando los resultados mostrados en la figura 2.12 se puede deducir que una posible aplicación de la correlación es la localización de un patrón dentro de una imagen, ya que el valor donde la correlación es máxima corresponde a las coordenadas donde ese patrón se encuentra. Sin embargo si se toma el segundo caso de la misma figura se ve que si un objeto engloba a otro dará la misma respuesta. Otro caso en el que la correlación fallaría sería si buscamos un objeto grande y oscuro si también está presente otro objeto más pequeño, pero mucho más claro. Para solucionar esto se tiene la *correlación normalizada*, que evita además posible cambios en los niveles de gris entre el modelo que se busca y el objeto presente en la imagen [3].

$$r(m, n) = \frac{\sum_x \sum_y [f(x, y) - \bar{f}(x, y)][W(m+x, n+y) - \bar{W}]}{\sqrt{\sum_x \sum_y [f(x, y) - \bar{f}(x, y)]^2 \sum_x \sum_y [W(m+x, n+y) - \bar{W}]^2}} \quad (2.14)$$

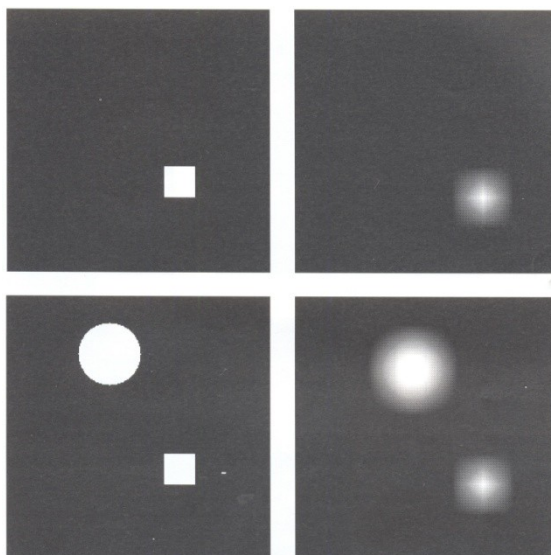


Fig. 2.12. Gráfica Del Concepto De Correlación. Detección de un Cuadro Mediante Correlación.

En la figura 2.13 puede verse el uso de la correlación normalizada. En este caso se trata de buscar un componente electrónico en la imagen. Puede observarse que la correlación no es inmune a los cambios en la escala, ni a la rotación del objeto.

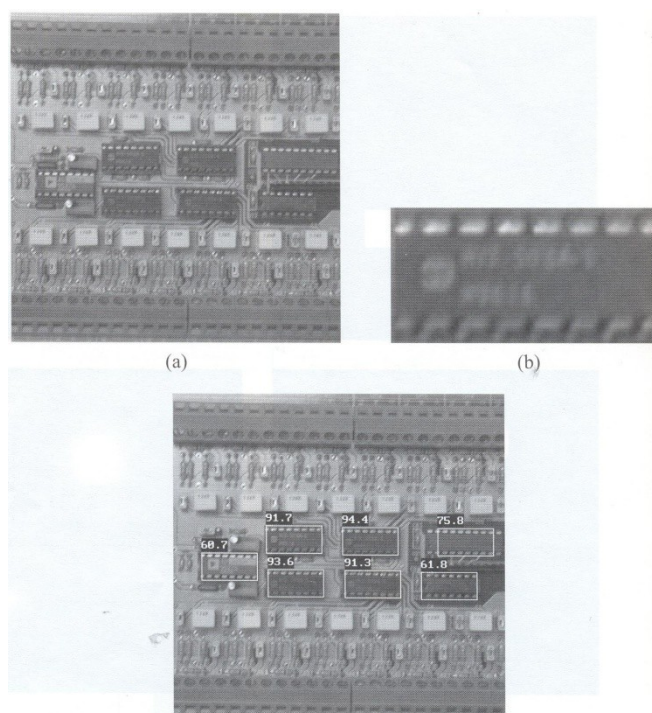


Fig. 2.13. Correlación Normalizada: (a) Imagen; (b) Modelo Ampliado; (c) Resultados de la Correlación Normalizada.



2.9.3.1. Segmentación

Una imagen digital contiene una gran variedad de patrones relacionados directamente con los objetos de la escena y, de acuerdo con las definiciones arriba planteadas, habrá otros patrones que no tengan este significado. Para estudiar con mayor facilidad aquellos patrones de interés, es necesario separarlos del resto de la imagen; este proceso puede ser **real** o **virtual**. Es *real* si los píxeles que se refieren a un patrón son asignados a una imagen intermedia en donde éstos ocupan sus mismas posiciones relativas, resultando tantas imágenes figura 2.14 como patrones se deseen separar. Es *virtual* si los píxeles del patrón únicamente se etiquetan como pertenecientes a una clase dada o si sus valores se cambian todos a uno predefinido, pero sin separarse de la imagen original.

Al reconocer los objetos de una escena, se le segmenta también, ya que la segmentación es el agrupamiento de puntos de ésta en regiones conectadas que poseen una correspondencia significativa con los objetos presentes en la escena; sin embargo, la segmentación no necesariamente conduce a un reconocimiento de patrones. Al transformar y operar sobre una imagen, el fin último es siempre el reconocimiento de los patrones que la componen, en el contexto de un problema, con el objetivo de auxiliar en el modelado del comportamiento del sistema físico que presenta [24].

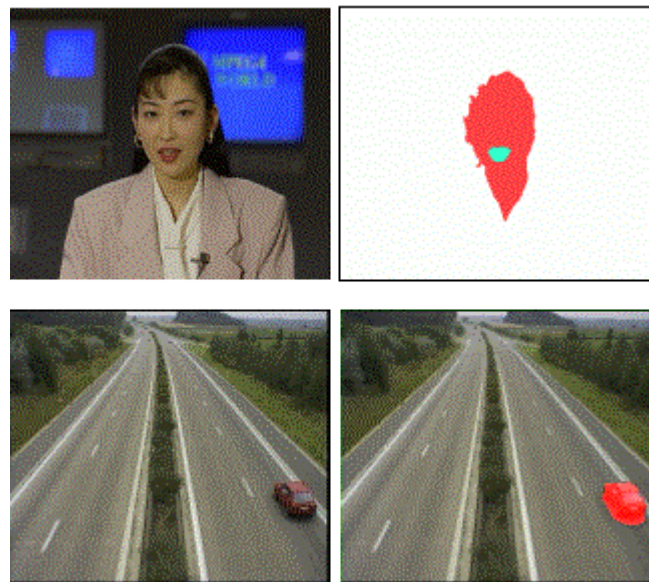


Fig. 2.14. Segmentación de Una Imagen de Video.

2.9.3.2. Evaluación y Comparación

En contraste con los métodos diferenciales, los cuales están basados en la continuidad del flujo óptico, la aproximación de correlación es insensible a los cambios de intensidad entre dos imágenes. Esto hace a las técnicas basadas en la correlación; muy útiles, para el procesamiento de imágenes estéreo donde siempre ocurre una ligera variación de intensidad entre la imagen izquierda y derecha, debido al uso de dos cámaras diferentes. Incluso si nosotros extendemos la técnica de correlación para múltiples correlaciones en más de dos cuadros, esta continua siendo una aproximación discreta en tiempo, aunque esta carece de



la elegancia de los otros métodos los cuales son formulados en espacios continuos antes de ser implementados para imágenes discretas [5].

2.10. Aplicaciones del Reconocimiento Basado en Movimiento

Las aplicaciones para el reconocimiento basado en movimiento son muy amplias, este ya ha sido utilizado en el campo de la medicina para el estudio del movimiento ventricular izquierdo, para revelar daño, deterioro o anormalidades, en el análisis de comportamiento clínico (la localización de varias articulaciones del cuerpo de un paciente son rastreadas y analizadas para anormalidades usando una computadora), el estudio de la unión relativa de ángulos a través de un círculo. Similarmente en los deportes y entrenamientos atléticos existen algunos sistemas nuevos, por ejemplo, digitalizar el patrón de una jabalina y calcular su velocidad y ángulo de ataque y comparar los valores contra el modelo de un perfecto lanzamiento de jabalina. Sistemas de reconocimiento basados en movimiento pueden también ser usados en el análisis de movimientos de danza y como una herramienta instructiva para bailarines. En métodos que crean alguna compresión de imágenes pudiendo ser usados para señales en tiempo real. Sistemas de vigilancia, para los cuales una detección de movimiento simple podría no ser suficiente, técnicas de reconocimiento de movimiento podrían ayudar a aclarar entre posible o permisibles tipos de movimiento desde un tipo de movimiento no deseable en una escena particular. En sistemas de monitoreo automático esas técnicas podrían ayudar a localizar problemas destacando posibles características en una línea automatizada, dando las restricciones de los movimientos permitidos [4] [6].

2.11. Sistemas Actuales.

2.11.1. Sistema basado en aproximación probabilística.

Una de las investigaciones que comprende el reconocimiento de personas a través de secuencias de video es la realizada por un investigador de nombre Song, La aproximación tomada por Song para el reconocimiento de personas es Probabilística, en la cual las características y variaciones comunes del movimiento humano son encapsuladas en un modelo estadístico simple, este modelo es almacenado como una gráfica que contiene vértices y bordes. Cada vértice constituye un punto característico del movimiento de alguna parte del cuerpo humano. El vértice es representado por una posición euclidiana (x,y) indicando la posición media de la característica en el cuerpo humano. Así como la velocidad media (V_x, V_y) . En adición, hay una matriz de covarianza que relaciona cada uno de estos cuatro parámetros a aquellos de cada uno de los vértices. Cada dos vértices conectados por un borde son considerados por ser probabilísticamente dependientes uno de otro. Y aquellos no conectados son considerados independientes [15].

La reducción de puntos característicos para cada modelo a evaluar requiere que el investigador sea cada vez más inteligente a la hora de elegirlos en primera instancia. Sería impráctico que simplemente encontramos puntos en movimiento en la imagen completa y



alimentamos estas características en el modelo en una evaluación simple. En primera instancia, el número de características sería muy largo dando un funcionamiento pobre. En segunda estaríamos proveyendo un gran número de puntos con características extrañas, tales como alguna parte del fondo de la imagen. Por este motivo el uso de la segmentación es usado para separar la imagen principal del fondo de la imagen. La segmentación de una imagen es la acción de un algoritmo que separa regiones de una imagen de manera que resaltan como el humano las percibiría de manera natural. Cuando hablamos de movimiento una aproximación natural sería la de segmentar aquellas parte de la imagen que están generando movimiento relativo del fondo. Este proceso es llamado extracción de fondo. Una vez que la imagen es segmentada de acuerdo al movimiento de fondo, podemos rastrear los puntos característicos de manera separa en cada región. Esto permite correr un modelo de evaluación para región de manera individual de mejor manera que una que basado en la imagen completa [15].



Fig. 2.15. Ejemplo del algoritmo de rastreo de puntos diseñado por Kanade-Lucas_Tomasi.

El algoritmo diseñado por Kanade-Lucas-Tomasi está basado en una implementación que toma como referencia el algoritmo KLT diseñado por Stan Birchfield [16] a pesar de la amplia modificación que se realizó para mejorar la velocidad de acuerdo a los detalles necesarios para este tipo de aplicaciones teniendo como finalidad la robustez ante cambios bruscos de características. El algoritmo es ejecutado primero en una versión sub muestreada de la imagen. Esta imagen sub muestreada es calculada filtrando como primera etapa la imagen original por medio de un filtro gaussiano y después removiendo las columnas y filas impares. El rastreo es procesado primero en una imagen sub muestreada dos veces para obtener una aproximación del desplazamiento.

La fórmula de rastreo es usada para calcular el desplazamiento por cada iteración del algoritmo. Una vez que la conversión del desplazamiento es cercano a cero el rastreo es



completado. Podemos decir que el algoritmo ha fallado si no converge después de algunas iteraciones. El algoritmo KLT tiene varios parámetros numéricos que fueron seleccionados para su implementación. El tamaño de la ventana deberá de ser de 7×7 . Este parámetro es el tamaño de la región sobre la cual la sumatoria de la ecuación es evaluada. Las iteraciones para el rastreo son realizados hasta que una iteración simple tiene un desplazamiento menor a 0.1 píxeles. Si se completan 10 iteraciones sin un desplazamiento menor a 0.1 píxeles la característica es descartada.

Una consideración a este modelo es que un modelo simple puede solamente detectar a una persona caminando de derecha a izquierda o viceversa pero no en ambos sentidos. Esta restricción existe porque el modelo contiene un valor medio para la velocidad x en cada punto característico.

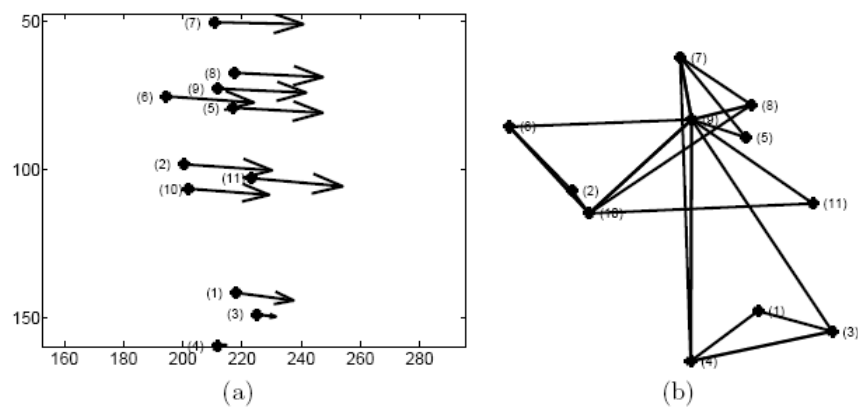


Fig. 2.16. a) Posición media y velocidad media de los puntos característicos b) Grafica triangular del modelo.



Fig. 2.17. En la figura (a) se muestra un ejemplo de un error de detección, la persona a la derecha no es detectado por el algoritmo, la figura (b) es incorrectamente clasificada como una persona caminando.



2.11.2. Modelo Basado en la Silueta Humana.

El modelo inicial es obtener tanto el movimiento de objetos en el primero plano del campo de visión, tales como gente caminando, coches en movimiento, etc. La idea básica de este modelo depende de la intensidad de los pixeles estacionarios su valor de brillo el cual tiene una alta redundancia sobre valores de intensidad previamente entrenada.

Para obtener el modelo de fondo inicial, primero el número de los radios de redundancia de intensidades para un pixel de locación x en todas las imágenes es calculada para varios segundos de video para poder distinguir los pixeles en movimiento de los estacionarios. Una secuencia de entrenamiento es llamada para esta secuencia de video. Cada valor de intensidad tomado para la locación x del pixel en la secuencia de video es comparado con los valores de intensidad tomados hasta el momento. De esta manera la variación de intensidad de ese pixel es almacenado dentro de una cadena de celdas asignadas para la locación x del pixel en un mapa histórico. En el mapa histórico la primera celda de la cadena almacena el número de diferentes valores de intensidad tomada para la locación x del pixel durante la secuencia de entrenamiento, las siguientes dos celdas Fig. (2.18) para la primera etiquetada como 1a almacena el valor de intensidad, la segunda celda se etiquetada como 1b almacena la redundancia del valor de intensidad.

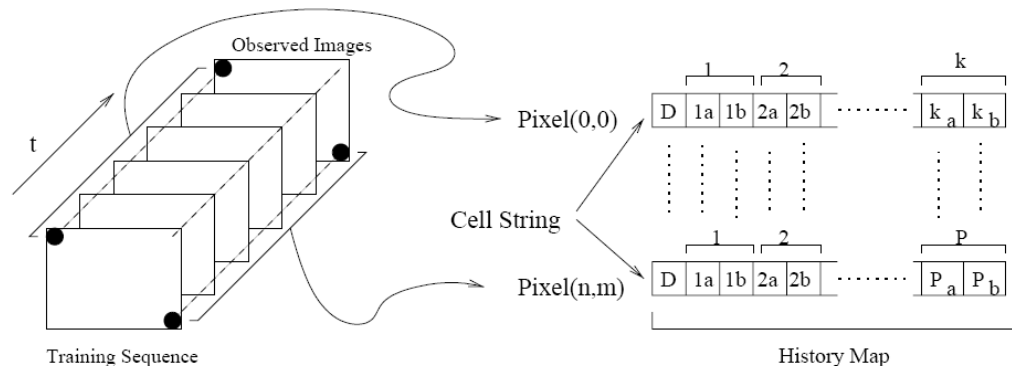


Fig. 2.18. (Izquierda) Secuencia de Entrenamiento, (Derecha) Estimación del Mapa Histórico.

El resultado de este algoritmo para la secuencias de entrenamiento se muestra en la Fig. (2.19a) los resultados del algoritmo HRR que produce el modelo inicial de fondo se muestra en la Fig. (2.19b) y en la Fig. (2.19c) se muestra el mapa de pixeles erróneos entre el fondo ideal y fondo estimado obtenido por el algoritmo HRR.

La idea para el desarrollo del algoritmo HRR es inspirada en el filtro de mediana. La ventaja del algoritmo HRR comparado con otros algoritmos depende del uso del filtro de mediana.



- El uso de la media se basa en asumir que el fondo de cada pixel será visible más del 50 por ciento del tiempo de duración de la secuencia de entrenamiento. Pero en el algoritmo presentado, el asumir que la visibilidad del fondo de cada pixel llegar a ser el radio de mayor redundancia de los valores de intensidad durante la secuencia de entrenamiento. Esto es más flexible y aplicable para eventos reales que el filtro de mediana y además agrega todas las ventajas de los algoritmos dependientes de un filtrado de mediana.
- El requerimiento de memoria para poder aplicar el algoritmo HRR es menor comparado con otros algoritmos basados en el filtro de mediana.

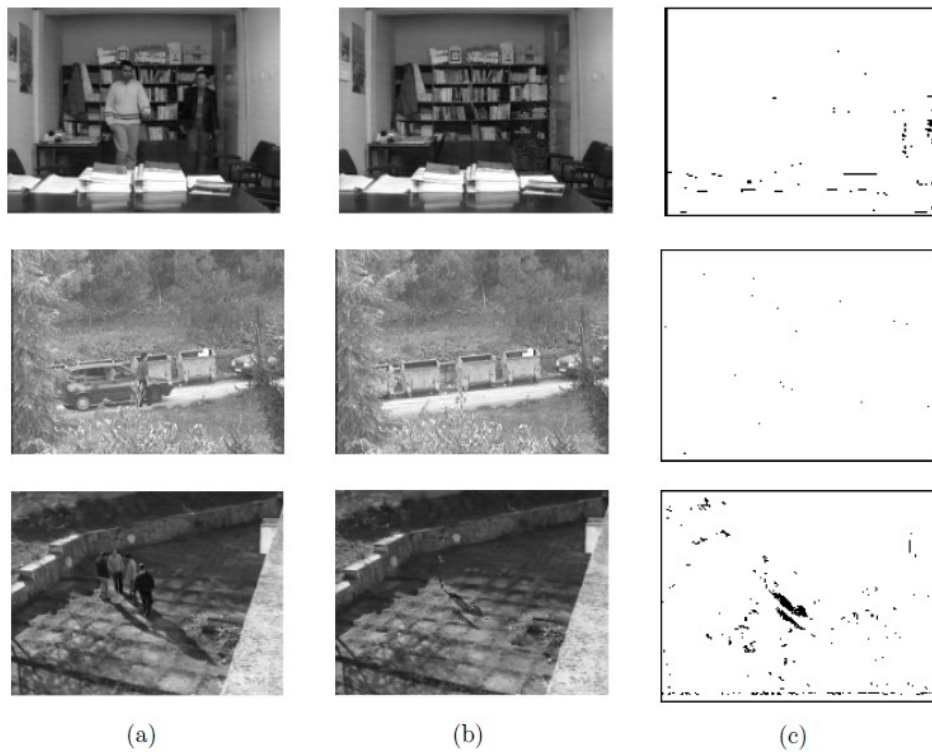


Fig. 2.19. Resultados de la ejecución del algoritmo, a) imagen de la secuencia entrenada, b) cálculo del modelo de fondo, c) Mapa de los pixeles errores.

Para probar la confiabilidad del modelo de fondo basado en el método HRR se creó un escenario Fig. (2.20). Dentro de este escenario hay una persona generando movimiento en una dirección específica y esperando un periodo de tiempo para después alejarse de ese punto. Después del seguimiento por un periodo de tiempo y cuando nadie está más en se punto de la escena una nueva persona comienza a generar movimiento pero generando una intensidad distinta.

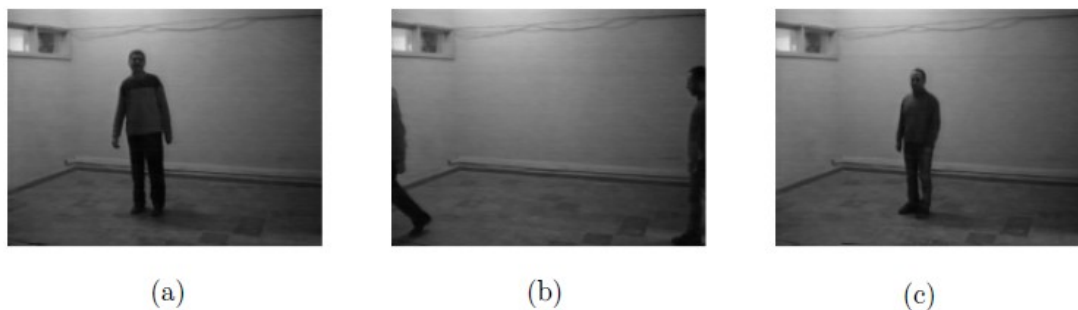


Fig. 2.20. Ejemplo de los cuadros de video utilizados para la prueba del algoritmo.

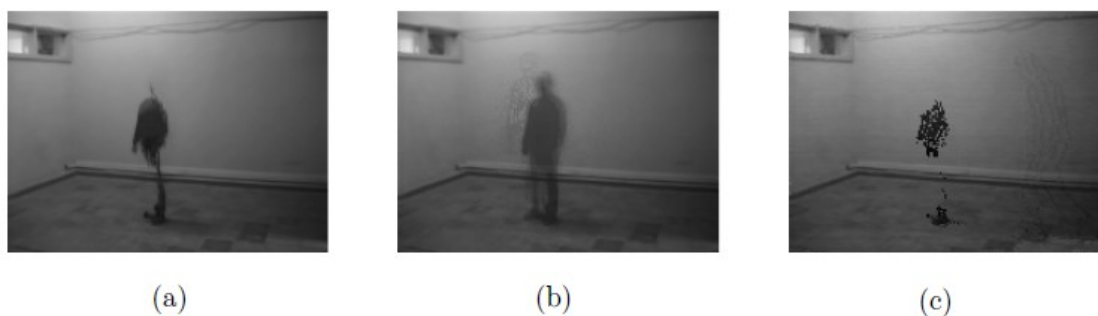


Fig. 2.21. Resultados de la inicialización del modelo de fondo bajo diferentes algoritmos sobre una secuencia de entrenamiento, a) algoritmo basado en la mediana b) Algoritmo W4, Algoritmo basado en HRR.

Los resultados experimentales en la secuencia de video se muestran en la Fig. (2.22) los tres algoritmos son aplicados sobre una secuencia de video de 5 segundos para obtener el modelo de fondo, el periodo de aprendizaje ha sido muy corto para poder tener una mayor visibilidad al momento de comparar los algoritmos. El algoritmo HRR produce resultados más robustos y confiables comparados con los otros algoritmos.

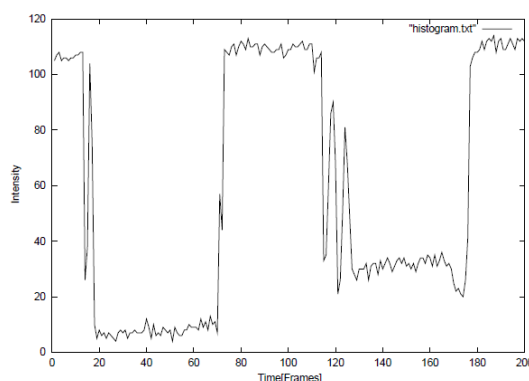


Fig. 2.22. Análisis de los cambios de intensidad calculada a diferente tiempo.



La tabla que se muestra a continuación muestra la comparación de este método de detección con otros algoritmos, una de las ventajas que presenta el algoritmo es la operación a bajo nivel y la capacidad de detectar más de una persona en la escena aun cuando presenten diferentes direcciones al mismo tiempo. El algoritmo depende de etapas de mantenimiento para el correcto funcionamiento de cada una de las etapas que lo envuelven.

Environment	Objects	Motion Type	Sequence
Indoor 1	2 people	walking complex	1500
Indoor 1	2 people	walking complex	1500
Indoor 3	4 people	walking complex	2750
Outdoor 1	multi people, cars	walking complex, slow	2500
Outdoor 2	multi people, car	walking complex, slow	3250
Outdoor 3	5 people	walking complex in shadow	1500
Outdoor 4	5 people	walking complex, in shadow	1500
Outdoor 5	multi people, cars	walking, running complex	6500
Outdoor 6	single person	walking, running	1300

Tabla. 2.1. Resultados obtenidos de las pruebas experimentales bajo ambientes controlados y no controlados.

A pesar de que los resultados presentados para este algoritmo presentan mejores resultados que el anterior tiene ciertas desventajas debido al método de extracción de información que maneja, el cual es basada en valores espaciales de la imagen. Si la imagen tuviera un alto contenido de sombras la efectividad podría verse disminuida al momento de querer discriminar el movimiento creado por la sombra al movimiento creado por el objeto que la produce. Otro punto importante es que las imágenes deben ser previamente entrenadas para que los resultados sean óptimos lo cual delimita su uso cuando los escenarios serán variables y las secuencias hasta cierto punto no podrán ser entrenadas.

2.11.3. Detección de Personas Usando Clasificación Basada en Reglas Difusas

El algoritmo propuesto consiste de 5 etapas, adquisición de la imagen, segmentación, pre-procesamiento, identificación de la burbuja mayor, extracción de las características de la burbuja y clasificación del movimiento.

Después de adquirir la secuencia de imágenes para la detección de movimiento la etapa de segmentación usa dos diferentes métodos, la substracción del cuadro adyacente y la sustracción del fondo dependiendo de la tasa de cuadros en la secuencia de video.

Para una secuencia con una tasa alta el método de substracción del cuadro adyacente es usado dado que el cambio por el movimiento entre cuadros adyacentes es muy pequeño. Por lo tanto la diferencia resultante $D(x,y,t)$ entre un cuadro de entrada $F(x,y,t)$ y el siguiente



cuadro adquirido $F(x,y,t+c)$ después de un intervalo de tiempo definido C está dado por:

$$D(x,y,t) = \text{abs}\{F(x,y,t) - F(x,y,c+t)\} \quad (2.15)$$

De esta manera se elimina el fondo estacionario dejando solamente las regiones de movimiento deseadas. En algunas secuencias de video el proceso de adquisición podría tener cambios erráticos debido a iluminación, reflexión o ruido. Para reducir los cambios de intensidad o sensibilidad causados por estos factores. Se aplica el filtro Gaussiano espacial antes de la diferenciación.

Después se obtiene la imagen del movimiento la cual es calculada con el uso de un umbral para reducir información innecesaria.

La etapa de pre-procesamiento consiste de dos etapas, operaciones morfológicas y etiquetado burbuja lo cual ayuda a preparar la imagen de movimiento para la etapa de identificación de burbuja.

Operaciones morfológicas, las siguiente operaciones morfológicas son realizadas en la imagen en movimiento.

- Cierre, morfología en las zonas suavizadas de los contornos.
- Completar huecos, una operación flood-fill es utilizado para completar pequeños huecos en la imagen.
- Eliminación de movimiento en los bordes, pixeles que se localizan lejos de los bordes son eliminados para evitar ambigüedad en la región que posee el posible movimiento relevante.

La etapa final para la detección del movimiento es el etiquetado burbuja o regiones de movimiento detectado. Estas regiones son contadas y etiquetadas de manera individual. De esta manera se calcula el tamaño de cada una de ellas.

En esta etapa la identificación de la mayor burbuja para determinar el movimiento humano. La burbuja mayor se define como la burbuja en movimiento que es exhibida por el movimiento de una persona. Este método utiliza los vectores de movimiento el cual fue utilizado por primera vez por Shio y Sklansky [18].

El vector de movimiento es dado por.

$$\vec{\delta} = (\delta_x, \delta_y) = \{(m,n) | \min_{m,n \in \{0, \dots, \pm l\}} \{C(m,n)\}\} \quad (2.16)$$



Donde:

$$C(m, n) = \sum_{x=-1}^1 \sum_{y=-1}^1 |F(x, y, t+1) - F(x+m, y+n, t)| \quad (2.17)$$

La dirección del movimiento estimado es restringida a un valor entero entre -1 y 1 para cada locación (m, n) . El vector de movimiento promedio es calculado después para cada burbuja. En algunas ocasiones cuando se generan varias o algunas burbujas menores cerca de una burbuja mayor se realiza en proceso de centroide para combinar esas burbujas con la burbuja mayor en caso de que exista correspondencia.

En la etapa final, la base difusa utilizada para extraer la mayor burbuja tiene tres variables los cuales representan el criterio de detección distancia del vector de movimiento, cambio en las texturas y porcentaje del área de la elipse de covarianza son formados por las tres burbujas características, calculadas anteriormente.

Un conjunto de 5 reglas difusas IF-THEN Para un sistema de inferencia Mamdani con funciones de membresía trapezoidal.

Rule R_1 : If X_1 is NEAR and X_2 is SMALL and X_3 is UNFIT then Y is REJECT	ELSE
Rule R_2 : If X_1 is NEAR and X_2 is BIG and X_3 is UNFIT then Y is REJECT	ELSE
Rule R_3 : If X_1 is FAR and X_2 is SMALL and X_3 is UNFIT then Y is REJECT	ELSE
Rule R_4 : If X_1 is FAR and X_2 is BIG and X_3 is UNFIT then Y is REJECT	ELSE
Rule R_5 : If X_1 is NEAR and X_2 is SMALL and X_3 is FIT then Y is ACCEPT	

Las reglas son agregadas con un operador de implicación mínima y difusa, utilizando un centroide difuso para obtener un valor de salida confiable.

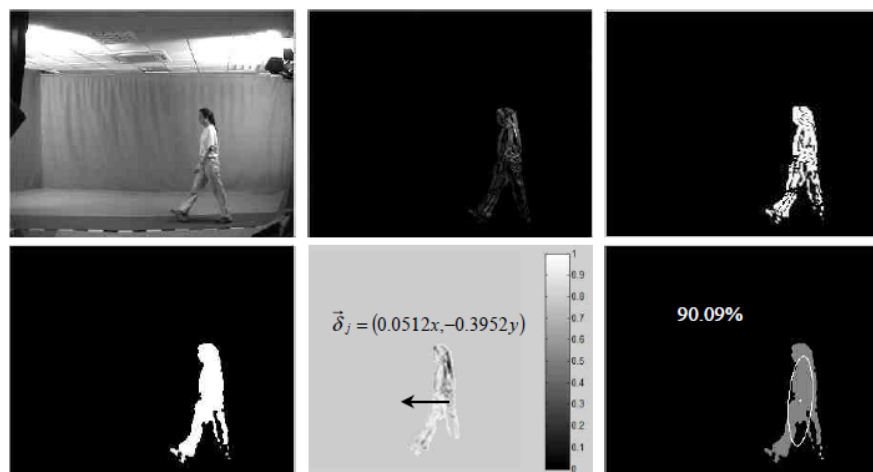


Fig. 2.23. Muestra de una secuencia de video analizada con el Algoritmo propuesto.



Algorithm	Detection accuracy (%)
Multi-feature classification metric (dispersedness, area of image blob) [7]	82.80%
Fuzzy rule-based reasoning for classification from the vertical projection of motion cues [6]	92.60%
Fuzzy rule-based classification of moving blob regions	93.75%

Tabla. 2.2. Resultados obtenidos de las pruebas experimentales bajo ambientes controlados y no controlados.

Los resultados experimentales muestran logros en las detecciones así como en la velocidad de detección, comparándolo con aproximación previas que tienen las mismas bases. De igual manera, la flexibilidad y adaptabilidad que muestra el algoritmo lo hace ideal para aplicaciones en tiempo real [18].

2.12. Conclusiones

Los principales métodos de extracción de información de movimiento así como los problemas a los que se enfrentan fueron presentados en este capítulo, mostrando al final que aún estamos lejos de lograr crear un sistema, capaz de replicar el sistema visual humano en su totalidad. Sin embargo, en su forma más avanzada, la detección de movimiento posee la capacidad de tres o dos dimensiones, puede determinar el tamaño y dirección del movimiento de un objeto, puede identificar y eliminar una acción repetitiva como el movimiento de un objeto, puede eliminar parte o gran parte del ruido contenido en la escena y puede compensar la vibración de la cámara. El flujo óptico es una técnica de compensación de movimiento para secuencias de imágenes que nos permite realizar este tipo de detecciones y clasificaciones de una manera precisa, esta técnica ha sido la base teórica para muchas investigaciones en el área y es la base principal de esta investigación.



CAPITULO 3

Sistema Propuesto

Introducción

En este capítulo, discutiremos a fondo cada una de las etapas del sistema propuesto. En este método la eliminación de información irrelevante y la extracción de la información de movimiento se realizan por medio vectores de movimiento, calculados a partir del uso de macrobloques, estos vectores de movimiento nos proporcionan magnitud, velocidad y dirección para poder determinar la importancia del movimiento dentro de la escena y de esta manera lograr aislar el objeto de interés del resto de la escena disminuyendo así el tiempo de cálculo y procesamiento durante la detección y rastreo.

3.1. Bosquejo General del Sistema

El flujo óptico es un proceso dentro de la visión por computador el cual es usado en una amplia variedad de aplicaciones, desde el análisis de una imagen 3D, la compresión de video y experimentos físicos.

El termino flujo óptico fue primeramente usado por el psicólogo James Jerome Gibson en su estudio de la visión humana en 1980-1981. A este respecto Horn y Schunk [HS80, HS81] desarrollan una simple manera de calcular el flujo óptico basado en una regularización. Este primer trabajo fue seguido posteriormente por un gran número de contribuciones con métodos alternativos propuestos. Por mencionar algunos, podemos mencionar los métodos de filtrado espacio-temporal iniciados por Adelson y Bergen [AB85], aquellos basados en la energía [Hee88], los métodos basados en fase [FJ90] y los métodos basados en igualación de regiones [BA83, BYX83, Ana89] [15].

Los componentes normales del flujo de la imagen y el flujo óptico son iguales si la imagen variante en el tiempo corresponde a una superficie Lambertian experimentando un movimiento puramente de translación bajo una iluminación uniforme espacial y temporalmente [7]. En el procesamiento de imágenes es posible estimar la trayectoria de cada píxel entre cuadros sucesivos de video, produciendo un campo de trayectorias del píxel conocido como flujo óptico [19]. El cálculo del flujo óptico es una herramienta muy útil en aplicaciones como el rastreo o detección de movimiento ya que a través de este se obtiene un vector de flujo el cual indica la posición anterior y actual del píxel, a través de este vector es posible calcular tanto la velocidad como la dirección real del objeto por cuadro. Sin embargo este no es un método práctico por varias razones.

Un cálculo exacto del flujo óptico es computacionalmente muy complejo ya que se calcula un vector de flujo por cada píxel en la escena generando un gran consumo de



tiempo, lo cual no puede ser aceptable para un algoritmo de aplicación en tiempo real.

Tomando el flujo óptico como base, podemos observar que el cálculo de vectores de movimiento nos proporciona información relevante dentro de la escena, de los objetos que presentan movimiento. Velocidad, trayectoria, y magnitud, son algunos puntos principales que se pueden obtener a partir de un vector de movimiento. El uso de la técnica de macrobloque es muy confiable para realizar dicho cálculo de vector, un ejemplo puntual sería el estándar de compresión MPEG basado en macrobloques y además cuenta con una etapa adicional de predicción y compensación de movimiento.

Tomado esto como referencia el sistema propuesto está basado en el cálculo de vectores por macrobloques tomando el principio aplicado al estándar de compresión MPEG2, creando así un método robusto y confiable para la extracción de movimiento en la escena.

Es importante mencionar que la función principal de un vector de movimiento es el indicar la posición actual de un objeto dentro del cuadro de video con respecto a su posición anterior, para que de esta manera un sistema adicional sea capaz de recrear el cuadro sin necesidad de tener toda la información contenida en el. Sin embargo, para nuestra aplicación solo es necesario extraer las coordenadas del vector dentro del cuadro de video para delimitar el área de búsqueda y poder estimar que objeto está produciendo estos cambios de posición. El diagrama del algoritmo propuesto se muestra en la figura 3.1.

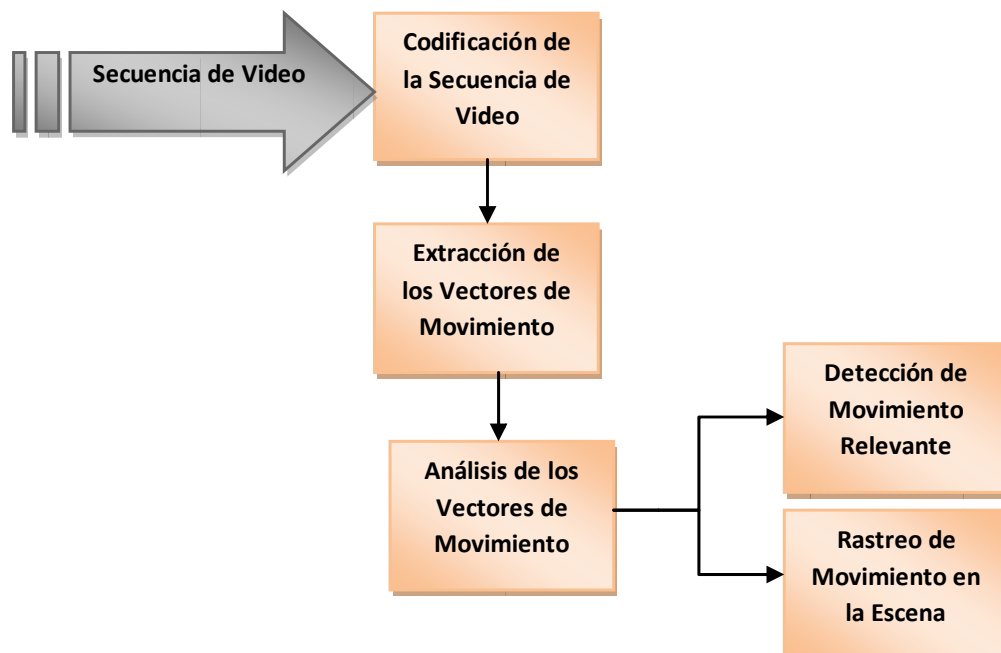


Fig. 3.1. Diagrama A Bloques Del Algoritmo Propuesto.



3.2. Codificador y Decodificador Genérico de Video

Generalmente un codificador de video consiste de dos etapas que son: la reducción de la redundancia temporal, como primera etapa, y la reducción de la redundancia espacial. La reducción de la redundancia temporal trata de estimar o generar una representación de la trama en el tiempo t a partir de la trama en el tiempo $t-1$. El método de compresión de video que emplea solamente la reducción de redundancia temporal se llama codificador “interframe”.

En el codificador estándar de video se usa un codificador basado en la DCT para reducir la redundancia espacial. El método de la reducción espacial es igual al codificador DCT del JPEG. El codificador de video que emplea solamente la reducción espacial se conoce como codificador “intraframe” y la combinación de codificadores “interframe” e “intraframe” se conoce como codificador híbrido, el cual se usa generalmente para codificación de video.

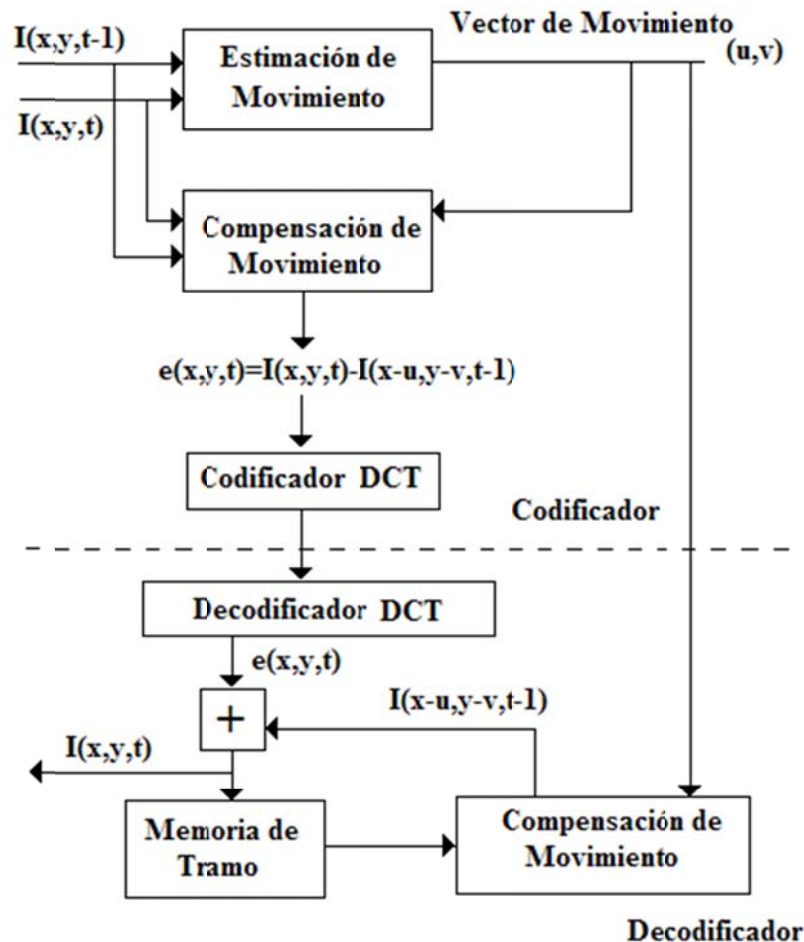


Fig. 3.2. Codificador/Decodificador de Video.



Podemos observar de la Fig. (3.2), que usando dos tramas consecutivas en tiempo (trama en el tiempo t y trama en el tiempo $t-1$), se estima el movimiento de cada bloque desde la trama $t-1$ a la trama t . Como resultado de la estimación del movimiento, para cada bloque se calcula un vector de movimiento (u,v) . Usando estos vectores de movimiento, se compensa el movimiento de los bloques de la tramas en tiempo $t-1$ y finalmente se calcula la diferencia entre dos tramas (uno compensado). Esta diferencia es considera como error de predicción y se calcula como se muestra en la ecu. (3.1)

$$e(x, y, t) = I(x, y, t) - I(x - u, y - v, t - 1) \quad (3.1)$$

Donde:

$e(x, y, t)$

Es error de predicción

$I(x, y, t)$

Es trama en el tiempo t

$I(x - u, y - v, t - 1)$

Es la trama compensada por el vector de movimiento (u,v) en el tiempo $t-1$.

Generalmente el error de predicción es muy pequeño, por lo tanto la señal de error tiene una gran cantidad de redundancia espacial, la cual se reduce usando el codificador de DCT como se realiza en el JPEG.

En el decodificador, se reciben los coeficientes de DCT de la señal de error y los vectores de movimiento. Seguidamente a partir de los coeficientes recibidos se reconstruye la señal de error. Posteriormente aplicando el vector de movimiento a la imagen de la trama construida anteriormente y guardada en la memoria, se calcula la trama compensada $I(x - u, y - v, t - 1)$. Finalmente usando la señal de error y la imagen de la trama compensada en tiempo $t-1$ se estima la imagen de trama en tiempo t $I(x, y, t)$, la cual se almacena temporalmente en la memoria para poder estimar la trama de tiempo $t+1$.

3.2.1 Estimación de movimiento

En el codificador y decodificador de video, la estimación de movimiento es un proceso muy importante. El proceso de estimación de movimiento se describe en la Fig.(3.3). Primero la trama actual se divide en varios bloques no traslapados de tamaño $N \times M$, los cuales se llaman “macrobloques”, como se muestra en la Fig.(3.3a). El punto superior-izquierdo es un punto representativo de un macrobloque, el cual se representa como (x,y) .

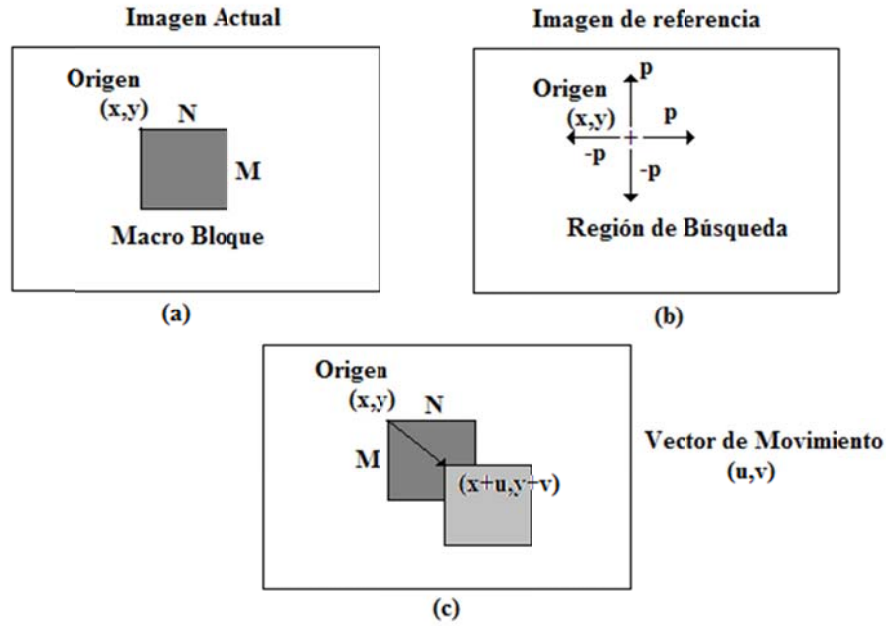


Fig. 3.3. Proceso de estimación de movimiento.

El movimiento del bloque se estima tomando (x,y) como referencia. En la trama de referencia (la trama anterior o posterior) de la trama actual, se realiza la búsqueda de mejor acoplamiento. La región de búsqueda no es la región completa de la trama, sino una región dentro de la trama de referencia, debido a una suposición que un objeto no se mueve tanto de una trama a otra trama consecutiva. La región se representa como $[-p, p]$ en ambos ejes, tomando (x,y) como centro de la región; la Fig. (3.3b) muestra la región $[-p, p]$. Dentro de la región bajo análisis se realiza una búsqueda y se obtiene una posición $(x+u, y+v)$ que se la que proporciona el mejor acoplamiento, donde el vector de (u,v) se conoce como vector de movimiento. Diversos estudios han mostrado que en la codificación de video el tamaño de macrobloque mas adecuado es $N=M=16$.

3.2.2 Criterio de mapeo

El criterio más usado para determinar grado de acoplamiento es MAE (Mean Absolute Error), el cual esta dado por la ecu. (3.2).

$$MAE(i, j) = \frac{1}{MN} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} |C(x+k, y+l) - R(x+i+k, y+j+l)| \quad (3.2)$$

Donde:

$C(x+k, y+l)$

Es (k,l) -ésimo píxel del macrobloque (x,y) de la trama actual



$$R(x+i+k, y+j+l)$$

Es bloque de posición $(x+i, y+j)$ en la trama de referencia.

Aquí, $-p \leq i \leq p, -p \leq j \leq p$. Cuando el criterio $MAE(i, j)$ es mínimo, la posición (i, j) se determina como vector de movimiento.

3.2.3 Algoritmos para la estimación de movimiento

Se han desarrollado varios algoritmos para realizar esta tarea. Los principales algoritmos para la estimación de movimiento son los siguientes

1. Método de búsqueda exhaustiva (Full-Search Method)
2. Búsqueda logarítmica bidimensional (Two-Dimensional Logarithmic Search)
3. Búsqueda jerárquica paralela unidimensional (PHODS)
4. Búsqueda jerárquica (Hierarchical Motion Estimation)

3.2.4 Método de búsqueda exhaustiva (Full-Search Method)

Un método más simple para encontrar vector de movimiento es una búsqueda completo dentro de región $[-p, p]$. Quizá el método de búsqueda completo es mejor método desde el punto de vista su funcionamiento, sin embargo este método no es práctica, debido a que su complejidad es demasiada alta.

Complejidad de Algoritmo

El cálculo del criterio MAE para un punto (i, j) requiere NM operaciones, cada una de ellas consiste de tres operaciones básicas, las cuales son una substracción, un cálculo de valor absoluto y una suma, lo cual implica 3NM operaciones.

El rango de búsqueda es $-p \leq i \leq p, -p \leq j \leq p$, por lo tanto en $MAE(i, j)$ se calcularía $(2p+1)^2$ veces. Considerando que el tamaño de una trama es $I \times J$ y el número de

macrobloques dentro de una trama es $\frac{IJ}{NM}$ y un típico valor del número de tramas para la Transmisión de TV es 30 por segundo, tomando en cuenta los datos anteriores se obtiene que el número de operaciones que se deben realizar por segundo es de $\frac{IJF}{NM}(2p+1)^2 3NM = 3IJF(2p+1)^2$. En un caso típico, $I=720, J=480, F=30, p=15$, se requieren 29.89 GOPS (Giga operations per second), mientras que para $p=7$, se requieren 6.99 GOPS. Este número de operaciones es excesivo para un uso práctico.

3.2.5 Búsqueda logarítmica bidimensional (Two-Dimensional Logarithmic Search)

El método de búsqueda logarítmica es similar a una búsqueda binaria y está dada por

1. Calcular los valores de MAE del origen (0,0) y ocho puntos alrededor del origen (total nueve puntos).

Los nueve puntos son $(0,0), (0,d_1), (0,-d_1), (-d_1,0), (d_1,0), (d_1,d_1), (d_1,-d_1), (-d_1,d_1)$ y $(-d_1,-d_1)$, la distancia d_1 se define como $d_1 = 2^{k-1}$ donde $k = \lceil \log_2 p \rceil$.

2. Escoger el punto que tenga el mínimo valor MAE de los nueve puntos, el cual se considera como el punto del nuevo origen. Seguidamente se seleccionan ocho puntos con una distancia $d_2 = \frac{d_1}{2}$ desde el origen.

3. Repetir operación hasta que la distancia sea 1 pixel.

Ejemplo: $p=7$, $k = \lceil \log_2 7 \rceil = 3$

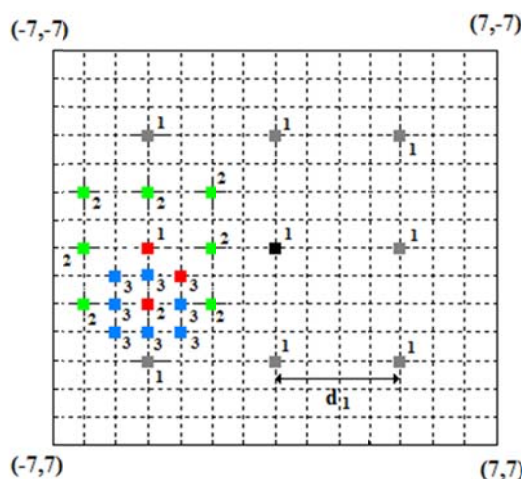


Fig. 3.4. Búsqueda Logarítmica.

Complejidad de Algoritmo

En este algoritmo, $8k+1$ puntos se calcula su valor de MAE. Por lo tanto número de operaciones son:

$$\frac{IJF}{NM} (8k+1) \times 3 \times NM = 3IJF(8k+1)$$

En caso típico, $I=720$, $J=480$, $F=30$, $p=15$, $k=4$ se requiere aproximadamente 1.0GOP.

3.2.6. Búsqueda Unidimensional en Paralelo Jerárquico (PHODS)

Diferente de la búsqueda logarítmica, la estrategia de búsqueda de este método se realiza independientemente en cada eje.

El algoritmo se realiza las siguientes etapas.

- (1) Definir la distancia $S = 2^{\lfloor \log_2 p \rfloor}$ y el punto de origen inicial el cual es (d_i, d_j) .
- (2) En paralelo, se calculan independientemente los puntos con el mínimo MAE en cada eje.
 - Eje-x mínimo local: Dentro de los tres puntos $(d_i - S, d_j), (d_i, d_j), (d_i + S, d_j)$, encontrar el punto con el menor valor MAE. Aquí el nuevo origen en eje-x d_i se actualiza con relación al punto mínimo encontrado.
 - Eje-y mínimo local: Dentro de los tres puntos $(d_i, d_j - S), (d_i, d_j), (d_i, d_j + S)$, encontrar el punto con el menor valor MAE. Aquí el nuevo origen en eje-y d_j se actualiza con relación al punto mínimo encontrado.
 - El valor de S se actualiza como $S = \frac{S}{2}$

Ejemplo: $p=7$, $S = 2^{\lfloor \log_2 7 \rfloor} = 4$

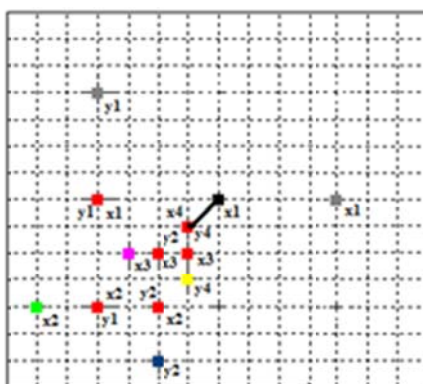


Fig. 3.4. Resultados del algoritmo PHODS.

Complejidad de Algoritmo

El número de puntos que calcula MAE es $(\lfloor \log_2 S \rfloor + 1) * 4 + 1$. Por ejemplo cuando $p=7$, entonces $S=4$ y el número de puntos de búsqueda es 13, cuando $p=15$, $S=8$ y el número de puntos de búsqueda es igual a 17.

Por lo tanto el número total de operación es $\frac{IJF}{NM} 13 \times 3 \times NM = 39IJF$, cuando $p=7$ y $\frac{IJF}{NM} 17 \times 3 \times NM = 51IJF$

En caso típico, $I=720$, $J=480$, $F=30$, $p=15$, necesita 528.76 MOP y para $I=720$, $J=480$, $F=30$, $p=7$, se necesita 404.35 MOP (Millones de operaciones por segundo).

3.2.7. Estimación Jerárquica de Movimiento

Este método está basado en reducir la resolución de la trama actual y la trama de referencia usando filtros pasa bajas. Aquí en cada aplicación del filtro pasa baja se realiza una decimación, por lo tanto el número de puntos en las que se realiza búsqueda se disminuye.

1. Aplicar dos veces el filtrado pasa baja y su respectiva decimación. La posición (x, y) de la imagen original es equivalente a la posición $\left(\frac{x}{2}, \frac{y}{2}\right)$ de la imagen de nivel 1 de descomposición y a la posición $\left(\frac{x}{4}, \frac{y}{4}\right)$ de la imagen de nivel 2 de descomposición.
2. Empieza la búsqueda del vector de movimiento en el nivel 2. En el nivel 2, el tamaño de la imagen es $\frac{1}{16}$ y el tamaño de macrobloque es 4×4 . Suponiendo que el punto (u_2, v_2) es tiene el mínimo MAE del macrobloque actual.
3. En el nivel 1, la búsqueda de vector de movimiento se realiza en los macrobloques de tamaño 8×8 con el origen de búsqueda $\left(\frac{x}{2} 2u_1, \frac{y}{2} + 2v_2\right)$. El rango de búsqueda es solamente $[-1, 1]$ alrededor de píxel del origen.
4. En el nivel 0, la búsqueda de vector de movimiento se realiza en los macrobloques de tamaño 16×16 con el origen de búsqueda en $(x + 2u_1, y + 2v_1)$. El rango de búsqueda es solamente $[-1, 1]$ alrededor de píxel del origen.

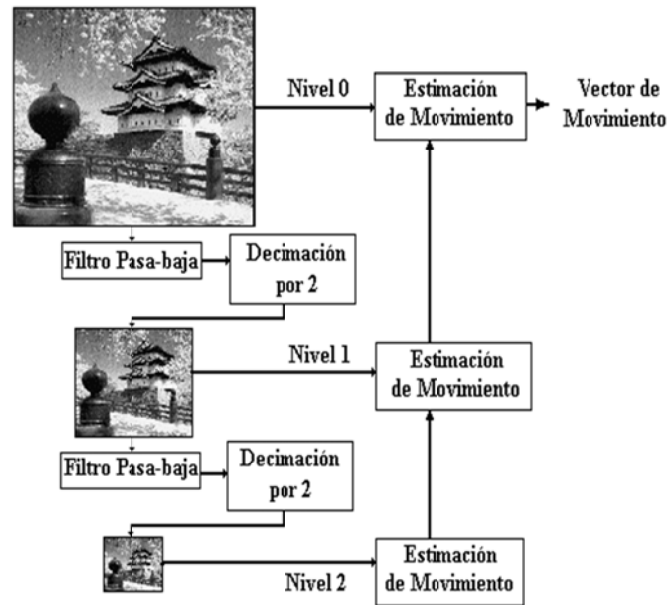


Fig. 3.5. Algoritmo de Estimación de movimiento Jerárquico.

Complejidad de Algoritmo

Tamaño original de tramo: 720 x 480

Tamaño de macrobloque: 16x16

Numero de tramas/segundo: 30

Nivel 2

Tamaño de imagen: 180x120

Tamaño de macrobloque: 4x4

Numero de macrobloque: $\frac{IJF}{NM} = \frac{180 \times 120 \times 30}{4 \times 4} = 40500$

Parámetros de búsqueda = $\left\lceil \frac{p}{4} \right\rceil = 4$ p=15

Número de búsqueda = $(2 \times 4 + 1)^2 = 81$

Número de operación por punto de búsqueda = tamaño de macrobloque x 3 = 4x4x3=48

Finalmente complejidad de la búsqueda en el nivel 2 es $40500 \times 81 \times 48 = 157.46 MOP$

Nivel 1

Tamaño de imagen: 360x240

Tamaño de macrobloque: 8x8

Numero de macrobloque: $\frac{IJF}{NM} = \frac{360 \times 240 \times 30}{8 \times 8} = 40500$

Parámetros de búsqueda=1

Número de búsqueda = $(2 \times 1 + 1)^2 = 9$

Número de operación por punto de búsqueda = tamaño de macrobloque $\times 3 = 8 \times 8 \times 3 = 192$

Finalmente complejidad de la búsqueda en el nivel 2 es $40500 \times 9 \times 192 = 69.98MOPs$

Nivel 0

Tamaño de imagen: 720x480

Tamaño de macrobloque: 16x16

Numero de macrobloque: $\frac{IJF}{NM} = \frac{720 \times 480 \times 30}{16 \times 16} = 40500$

Parámetros de búsqueda=1

Número de búsqueda = $(2 \times 1 + 1)^2 = 9$

Número de operación por punto de búsqueda = tamaño de macrobloque $\times 3 = 16 \times 16 \times 3 = 768$.

Finalmente la complejidad de la búsqueda en el nivel 2 es $40500 \times 9 \times 768 = 279.94MOPs$

La complejidad total es suma de la complejidad de los tres niveles, el cual es

$157.46 + 69.98 + 279.94 = 507.38MOPs$

3.3. Estimación de los Vectores de Movimiento

Existen diferentes métodos para el cálculo de vectores, como se muestra a continuación:

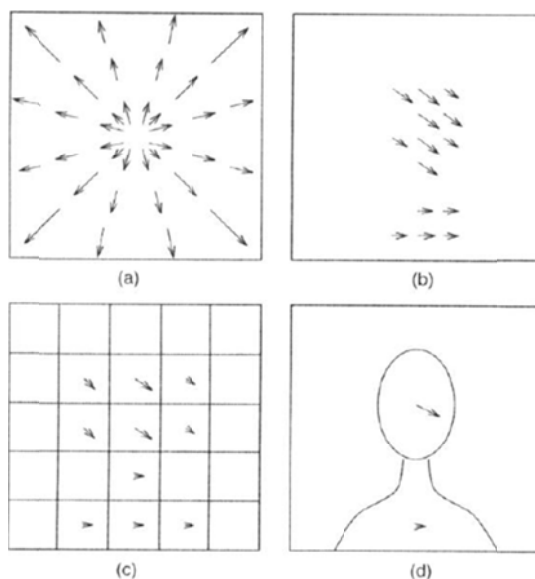


Fig. 3.6. (a) Imagen Completa (b) Pixel a Pixel (c) Bloques (d) Regiones.

La estimación de movimiento se basa en buscar un área en el cuadro de referencia para encontrar una región muestreada 'equivalente' $M \times N$. Esto se realiza comparando el bloque



en el cuadro actual con algunas o todas las regiones $M \times N$ posibles en el área de búsqueda y encontrando la región que dé el mejor ‘equivalente’, tal que, la región aspirante que minimice la energía residual sea elegida como el mejor ‘equivalente’ [19].

El movimiento reduce las similitudes entre las imágenes e incrementa la cantidad de datos necesarios para crear una imagen diferencia. La compensación de movimiento se utiliza para incrementar la similitud de las imágenes. La figura 3.7 muestra el principio.

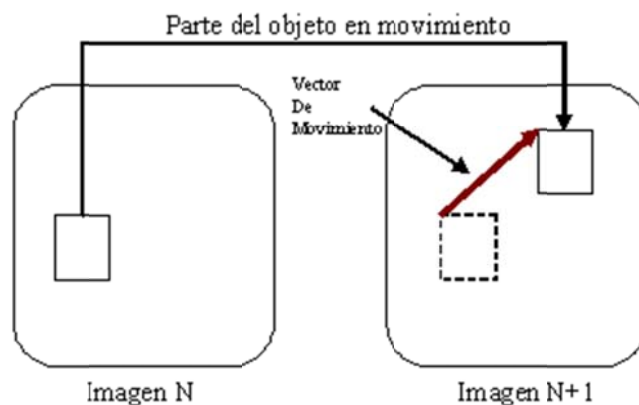


Fig. 3.7. Principio De La Compensación De Movimiento.

Basándose en la correlación espacial y temporal entre los vectores de movimiento, se hace una predicción del movimiento. Esto permite reducir la búsqueda: se consideran sólo los movimientos parecidos al predicho, reducir la información a enviar: solo se transmite la diferencia entre el movimiento predicho y el “real”.

Cuando un objeto se mueve en la pantalla puede aparecer en otra posición en la imagen siguiente pero generalmente no cambiará su apariencia. La diferencia de imagen puede ser reducida, midiendo el movimiento en el codificador. Este movimiento es enviado al decodificador como un vector. El decodificador usa este vector para correr parte de la imagen previa a un lugar más adecuado en la nueva imagen. Un vector controla el movimiento de imagen entera de la imagen conocida como macrobloque. El tamaño del macrobloque depende de la codificación DCT y la estructura de muestreo del color. La figura 3.8 muestra que en un sistema 4:2:0, el espaciamiento vertical de las muestras de croma es exactamente el doble que las muestras de luminancia. Un simple bloque 8×8 de muestras de crominancia se extiende sobre la misma área que 4 bloques 8×8 de luminancia. Por lo tanto, esta es la mínima área que puede ser desplazada por un vector. El estimador de movimiento trabaja comparando los macrobloques de luminancia de dos imágenes sucesivas [16].

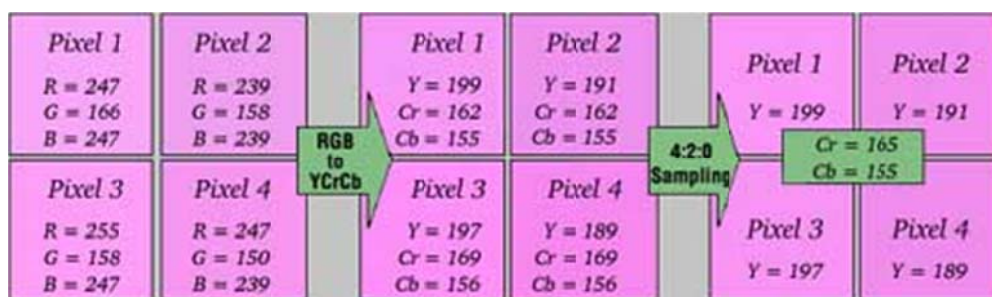


Fig. 3.8. Formato de Compresión 4:2:0.

Un macrobloque en la primera imagen es usado como referencia. Cuando se encuentra la correlación más grande, esta es asumida como la correlación que representa al movimiento. Este vector de movimiento tiene una componente vertical y una horizontal. En material típico, el movimiento es continuo a través de las imágenes. Una mejora en la compresión se logra si estos vectores se transmiten diferencialmente. Consecuentemente, si un objeto se mueve a velocidad constante, los vectores diferenciales serán cero. Los vectores de movimiento están asociados a macrobloques, no a objetos concretos [19]. Puede haber ocasiones en que parte del macrobloque se movió y parte no. En este caso es imposible compensar apropiadamente.

Si el movimiento de la parte móvil es compensado trasmitiendo un vector de desplazamiento, la parte estacionaria estará mal compensada y habrá la necesidad de corregir datos de diferencia. Si no se envía un vector, la parte estacionaria estará bien compensada pero habrá que corregir la parte móvil. Un compresor inteligente podría comparar ambas técnicas y quedarse con la que requiere menos datos diferenciales [16].

Esta etapa es la más importante para el funcionamiento del sistema de reconocimiento basado en el movimiento, ya que es aquí en donde se calculan los vectores de movimiento de cada imagen, los cuales son extraídos y utilizados para poder determinar cuándo un movimiento llega a ser importante o no.

La precisión en el cálculo de los vectores de movimiento estará basada en el tamaño del macrobloque utilizado para calcular el mejor equivalente es decir aquella zona que disminuya al máximo la energía.

3.4. Codificación espacial

El primer paso en la codificación espacial, es desarrollar un análisis de frecuencias espaciales mediante una transformada. Una transformada es una manera de expresar una forma de onda en un dominio diferente, en este caso, el de frecuencia. La salida de la transformada es un conjunto de coeficientes que indican cuanto de una determinada frecuencia está presente. La transformada más conocida es la de Fourier. Esta transformada encuentra cada componente de frecuencia multiplicando muestra a muestra la señal de



entrada por su respectiva función base e integrando el producto. La transformada de Fourier tiene la desventaja de requerir de coeficientes para los términos seno y coseno de cada frecuencia. En la transformada coseno, la señal de entrada es reflejada con respecto al eje de las ordenadas antes de multiplicarla por las funciones base. Este reflejo cancela todas las componentes seno y duplica las componentes coseno de la señal. Ahora los coeficientes seno de la transformada son innecesarios y solo se necesita un coeficiente para cada frecuencia.

La Transformada Discreta Coseno (DCT) es la versión muestreada de la transformada coseno, y es usada ampliamente en 2D. Un bloque de 8x8 píxeles es transformado en un bloque de 8x8 coeficientes. Ya que la operación requiere la multiplicación por fracciones, algunos coeficientes tendrán longitud de palabra más larga que los valores de los píxeles. Típicamente, un bloque de píxeles de 8 bits, incurrirá en un bloque de coeficientes de 11 bits. Por lo tanto, la DCT no produce una compresión, sino lo contrario. No obstante, la DCT convierte la fuente de píxeles en una forma en donde es más fácil la compresión [16]. Los coeficientes significativos de la DCT de un bloque de imagen o muestra residual son típicamente las bajas frecuencias posicionadas alrededor del coeficiente DC (0,0).

Los coeficientes DC diferentes a 0 son almacenados alrededor de la parte superior izquierda del coeficiente de DC y la distribución es simétricamente áspera en las direcciones horizontales y verticales. La Fig. (3.9b) grafica la probabilidad de los coeficientes diferentes a 0, pero esta vez los coeficientes son sesgados esto es porque el campo de la imagen tiene un componente de alta frecuencia muy fuerte en el eje vertical [19].

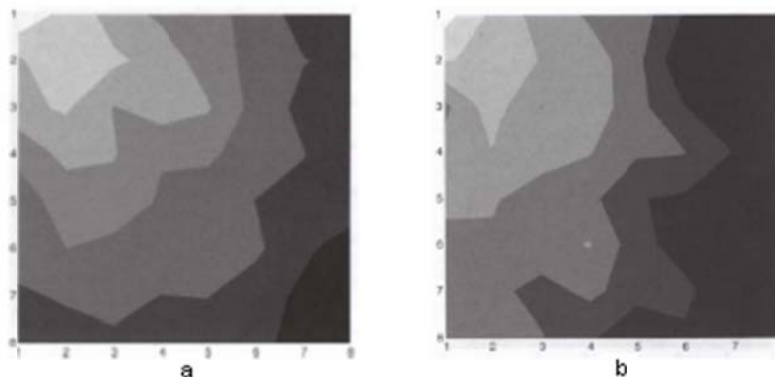


Fig. 3.9. Coeficientes de Distribución DCT, (a) Coeficientes Normales, (b) Coeficientes Sesgados.

3.5. Cuantización

Un cuantizador mapea una señal con un rango de valores de “x” para una señal cuantizada con un rango reducido de valores “y”. Un cuantizador escalar mapea un señal muestreada de entrada para un valor de salida cuantizado. Un cuantizador vectorial mapea un grupo de señales muestreadas de entrada para un grupo de valores cuantizados. Un

ejemplo simple de cuantización escalar es el proceso de redondear un numero fraccional al entero más cercano. El proceso es con pérdidas debido a que no es posible determinar el número fraccional original del entero redondeado.

$$FQ = \text{round}\left(\frac{X}{QP}\right) \quad Y = FQ \cdot QP. \quad (3.3)$$

En la compresión de video e imágenes la operación de cuantización se efectúa en 2 partes, primero, un cuantizador adelantado FQ en el codificador y un ‘cuantizador inverso’ IQ en el decodificador. Un parámetro critico es el step size QP entre los valores re-escalados sucesivamente, si el step size es grande el rango de valores cuantizados es pequeño y puede ser representado eficientemente durante la transmisión, pero los valores re-escalados suelen ser una cruda aproximación de la señal original. Si el step size es pequeño los valores re-escalados igualan a la señal original muy cercanamente pero un rango mayor de valores cuantizados reduce la eficiencia de compresión. Un cuantizador vectorial mapea un grupo de datos de entrada a un valor simple (codeword) y en el decodificador, cada codeword mapea una aproximación del grupo original de datos de entrada (un vector). El grupo de vectores son almacenados en el codificador y decodificador en un codebook (figura 3.10). La llave principal en el diseño del cuantizador vectorial incluye el diseño del codebook y una búsqueda eficiente en el codebook para encontrar el vector óptimo [19] [16].

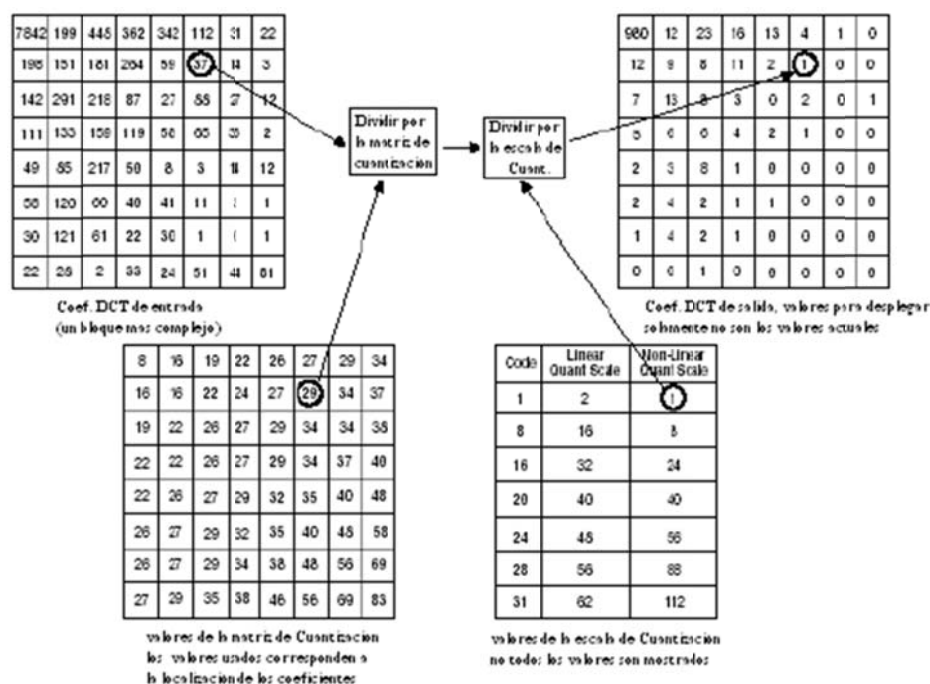


Fig. 3.10. Proceso de Cuantización.



Como se mencionó anteriormente los vectores de movimiento indican la posición anterior y actual de un macrobloque. En el estándar MPEG2 estos vectores de movimiento son generados en imágenes tipo P y B, los vectores de movimiento para las imágenes tipo P son adelantados los cuales también podríamos llamar reales y para las imágenes tipo B son adelantados, atrasados o ambos (figura 3.11), de acuerdo a la posición del cuadro y a la cantidad de información contenida en ella.

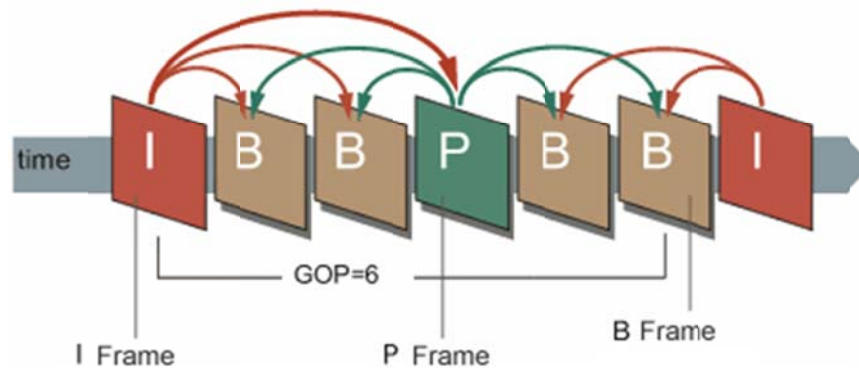


Fig. 3.11. Grupo De Imágenes.

Como se puede observar en la figura 3.8 el cálculo del vector de movimiento en el codificador consiste en encontrar el “Best Matching” del cuadro anterior con respecto al cuadro actual, durante el proceso de compresión, nosotros extraemos la información generada en los cuadros de video, tipo P, ya que, como se mencionó anteriormente, los vectores de movimiento generados en estos cuadros, representan la información real del movimiento, es decir, que a través de ellos podemos nosotros calcular la dirección y velocidad real del o de los objetos dentro de la escena.

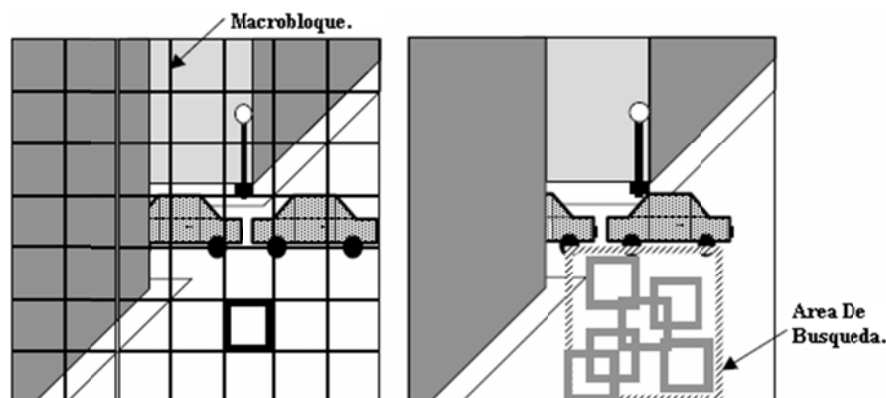


Fig. 3.12. Calculo Del Vector De Movimiento.

Al extraer la información de los vectores de movimiento del compresor obtenemos los componentes ‘x’ y ‘y’ del vector, como se muestra en la figura 3.13. A través de estos

componentes nosotros calculamos tanto la velocidad como la dirección del movimiento, como se muestra en la figura 3.14. El tiempo de proceso es entre 10 y 15 cuadros por segundo, el estándar es de 30 cuadros por segundo en video digital.

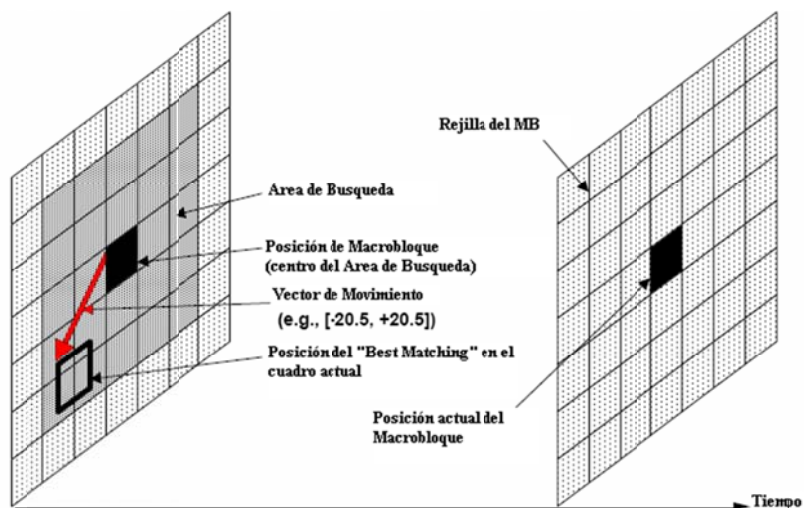


Fig. 3.13. Representación Gráfica De Un Vector De Movimiento.

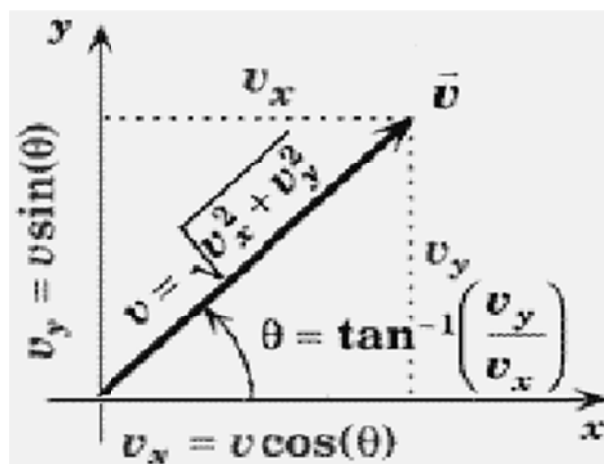


Fig. 3.14. Calculo de la Velocidad de un Vector.

De esta manera obtenemos la velocidad y dirección de cada uno de los vectores dentro de la escena, sin embargo, no podemos utilizar esta información de manera directa ya que debemos recordar que el algoritmo de macrobloques va a calcular vectores de movimiento de todos los objetos o píxeles que produzcan cambios dentro de la escena. Por lo tanto los vectores de movimiento no solo van a registrar movimiento provocado por los objetos (movimientos reales) también van a registrar cambios en la iluminación los cuales



comúnmente son representados por una cantidad muy pequeña de vectores y por lo general son vectores aislados, con dirección y velocidad irregular. Este tipo de vectores en aplicaciones de rastreo o detección de movimiento son conocidos como vectores de ruido ya que no representan al movimiento real de los objetos en la escena, por lo tanto debemos de eliminarlos para evitar posible errores de detección.

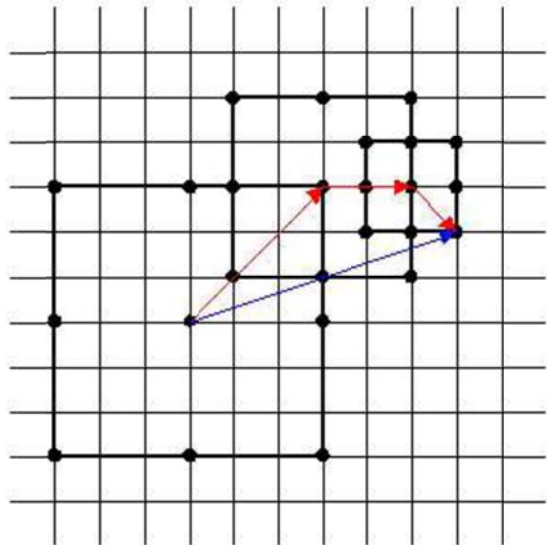


Fig. 3.15. Desplazamiento de Vector entre escenas.

3.6. Velocidad y Dirección de un Vector

La velocidad de un cuerpo en movimiento puede cambiar tanto su magnitud como su dirección. En un plano bidimensional podemos describir la velocidad del vector de la siguiente manera:

- Especificar la velocidad del cuerpo en el eje x (componente X) y la velocidad en el cuerpo y (componente Y).
- Alternativamente un vector bidimensional puede ser especificado completamente afirmando su velocidad (es decir la magnitud del vector) y su dirección como un ángulo.

La velocidad del vector y sus componentes forman un ángulo recto como se muestra en la figura 3.15. La cantidad de la componente X nos dice que tan rápido se está moviendo el cuerpo en la dirección x independientemente de que tan rápido este pueda moverse en otras direcciones.

Para una gran cantidad de situaciones las partes más importantes de un vector son sus componentes X y Y (componente X, componente Y). La dirección de un vector se entiende como el ángulo de la velocidad del vector como se muestra en la figura 3.16.

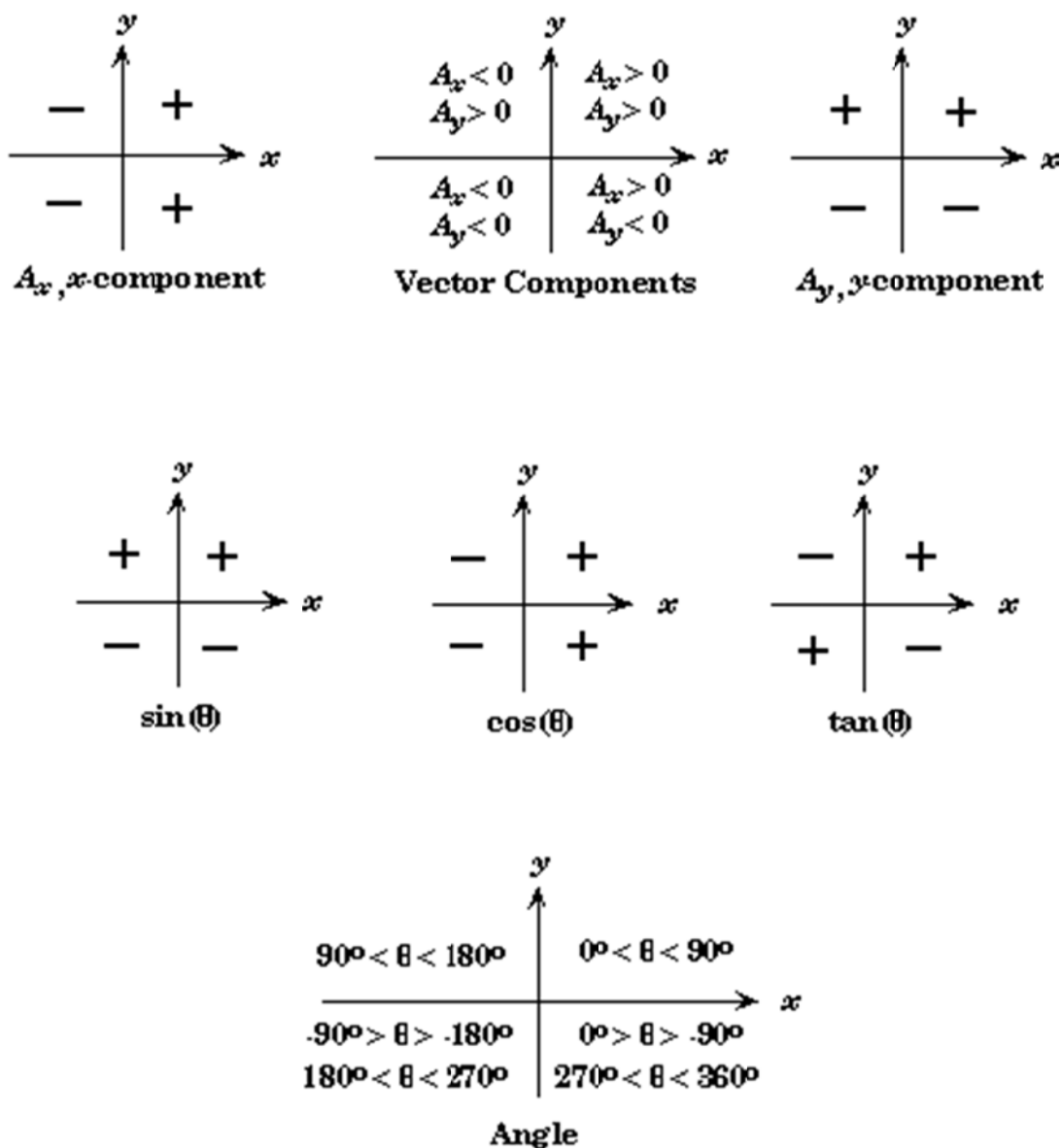


Fig. 3.16. Calculo de la Dirección de un Vector.

3.7. Filtrado de Vectores de Movimiento

La primera etapa del pre-procesamiento es el filtrado de vectores. Esta etapa es muy importante para la detección de movimiento, desde que la principal herramienta que tenemos para decir si el movimiento detectado es relevante o no, serán los datos de magnitud, velocidad y dirección calculados con los vectores de movimiento. Mientras más precisa sea la información obtenida en esta etapa, más precisa será la detección de movimiento.

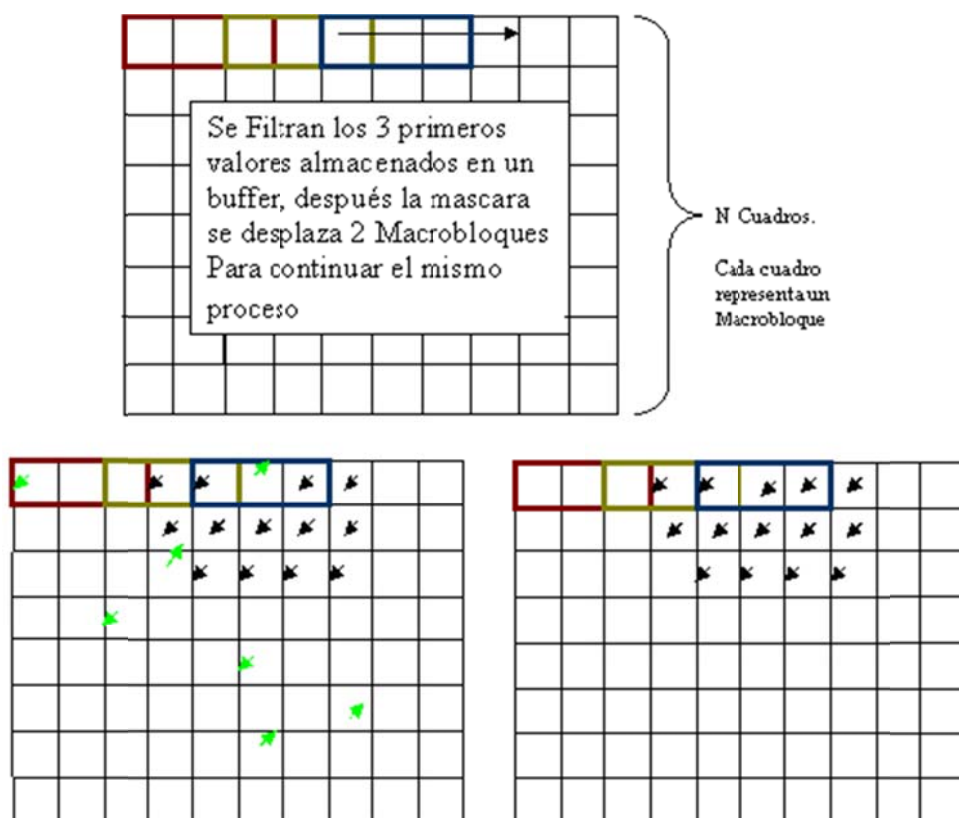


Fig. 3.17. Proceso De Filtrado Para Eliminar Los Vectores De Ruido.

Esta etapa está basada en dos pasos muy importantes. El primer paso consiste en el filtrado de vectores, en donde la tarea principal es eliminar los vectores de ruido y el segundo paso es de discriminación el cual se basa en las características que presenta el comportamiento humano ante un evento determinado.

Tomando en cuenta la forma de operar de un filtro de mediana el filtrado de vectores se realiza de manera similar, la principal diferencia es que en esta etapa utilizamos una máscara de 3x1 y esta se mueve cada 2 macrobloques como se muestra en la figura 3.17.

El sistema tiene la facilidad de ir registrando la dirección real del movimiento, esto tiene como finalidad corregir o cambiar la dirección de posibles vectores de movimiento que tengan una dirección diferente al movimiento real.

En la figura 3.18 podemos observar lo que se conoce como una raw extraction de los vectores de movimiento, en esta figura podemos observar varios vectores de ruido los cuales no representan el movimiento real de los objetos en la escena. Si nosotros usáramos la información de los vectores, de esta manera, el algoritmo tendría problemas de detección.



Fig. 3.18 Vectores De Movimiento Sin Etapa De Filtrado.

3.8. Estimación de la Dirección del Movimiento

Parte importante para el correcto funcionamiento de los sistemas de detección de movimiento es el análisis del comportamiento humano en cierto tipo de eventos, cumple ciertas características, las cuales pueden ser útiles durante la etapa de refinamiento de vectores.

Si nosotros fijamos la posición de la cámara, cerca de la entrada al área restringida, es decir, que el punto de vigilancia sea la parte superior de la escena, entonces por lógica, el movimiento relevante ocurrirá de arriba hacia abajo como se muestra en la figura 3.19, de esta manera podemos tomar como información relevante solo aquellos vectores de movimiento que se generen de arriba hacia abajo, en otras palabras vectores de movimiento que se encuentren dentro de un rango de 200° y 340° grados y como información irrelevante aquellos que se encuentren fuera de ese rango.

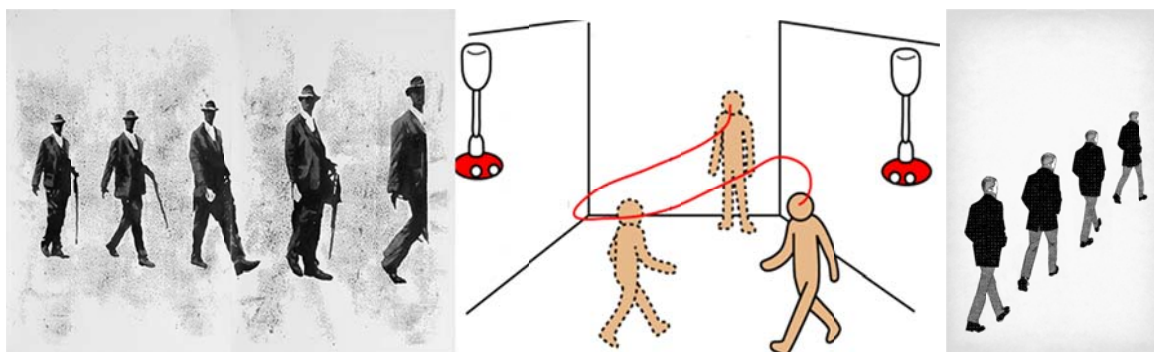


Fig. 3.19. Análisis Del Comportamiento Humano.

De esta manera podemos decir, no importa como sea o que dirección tenga el objeto en movimiento, si este es relevante siempre generara vectores de movimiento en la dirección establecida y de igual manera, el número de vectores de movimiento generados dependerá directamente del tamaño del objeto.



Fig. 3.20. Objetos distintos generando movimiento en la misma dirección.

A partir de este paso la información de movimiento obtenida por medio de los vectores de movimiento es suficiente para determinar la presencia de un objeto relevante en el cuadro.

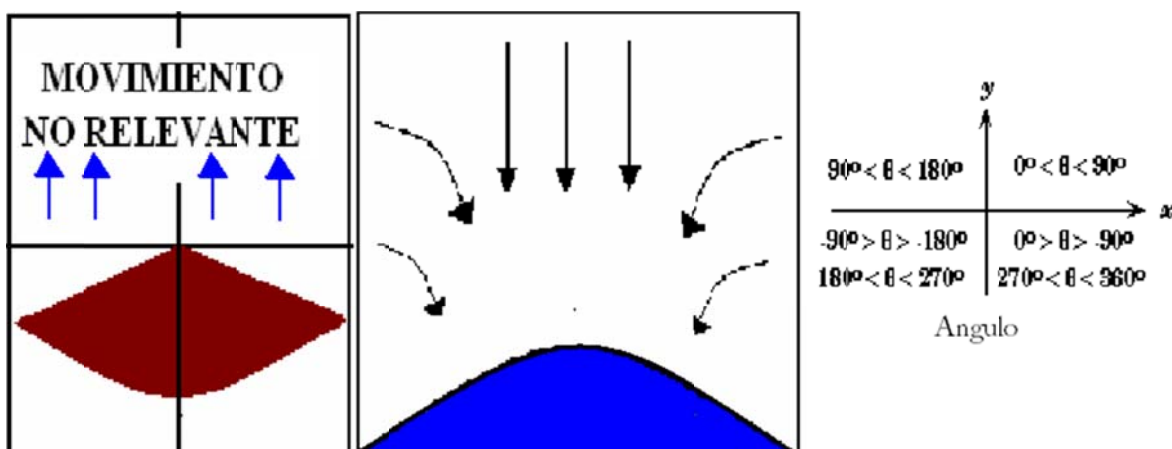


Fig. 3.21. Representación De La Dirección De Un Movimiento Relevante.

Otro factor importante en el algoritmo es, la distancia entre el objeto y la cámara, mientras mayor sea la distancia entre ambos mayor será el área a vigilar pero, menor será el número de vectores de movimiento a analizar, y mientras más cerca este la cámara del objeto menor será el área a vigilar y mayor el número de vectores a analizar.

Por lo tanto, la distancia debe ser tal que, permita un área de vigilancia amplia, permitiendo clasificar de esta manera el nivel de relevancia del movimiento y que genere a su vez una cantidad determinada de vectores de movimiento que faciliten la clasificación y análisis de cada uno de los movimientos en la escena.

En la figura 3.22 podemos ver un ejemplo de la distancia incorrecta entre la cámara y el objetivo a analizar, como se mencionó anteriormente. De no conservar una distancia

adecuada entre ambos la precisión del sistema se vería seriamente afectada.

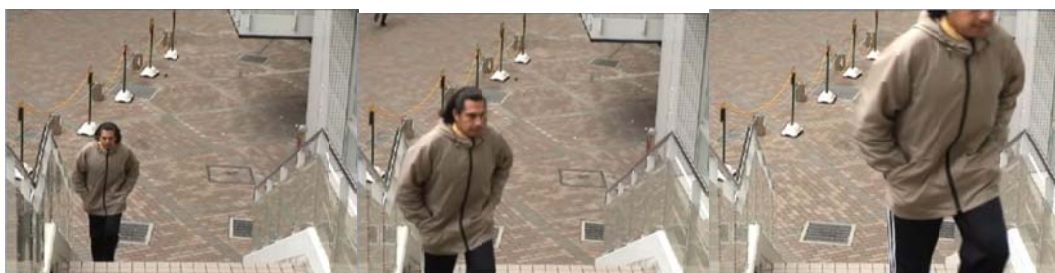


Fig. 3.22. Distancia Incorrecta.

En la figura 3.23 podemos ver una correcta distancia entre el objetivo y la cámara, podemos ver que durante toda la trayectoria se conserva la misma magnitud en el objetivo, por lo tanto la confiabilidad en el sistema es mayor comparado con la escena mostrada en la figura 3.22.

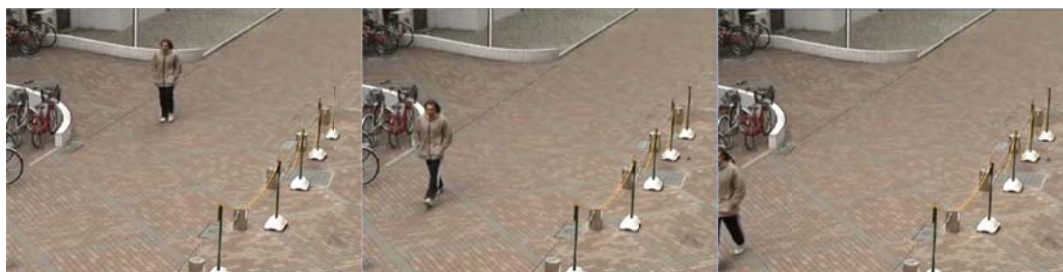


Fig. 3.23. Importancia De Una Correcta Distancia Entre La Cámara Y El Objeto.

3.9. Criterios de análisis

Al tener nosotros una correcta distancia podemos agregar divisiones para medir la importancia del movimiento y poder rastrear la posición del objeto en un tiempo determinado como se muestra en la figura 3.24. De igual forma se pueden usar varios contadores (figura 3.21) con la finalidad, no solo de saber cuántos objetos están generando movimiento relevante, si no también, de que dirección o parte de la escena proviene el movimiento y cuántos de estos movimientos entran al área restringida. De esta manera será posible rastrear y analizar el movimiento relevante generado.

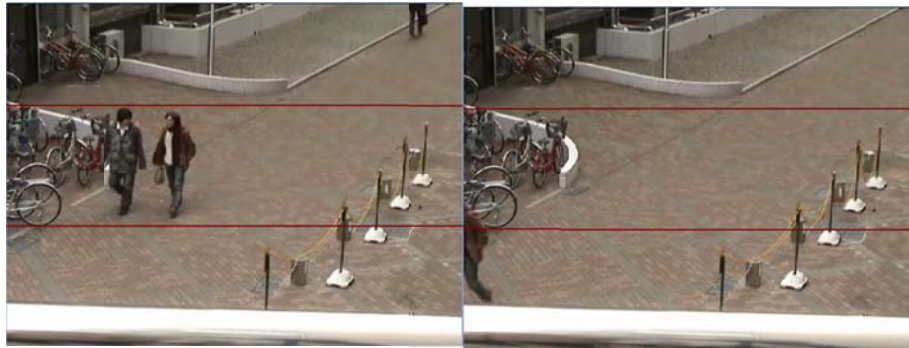


Fig. 3.24. División Del Área Para Determinar Un Nivel De Importancia.

Los contadores se basan en la magnitud y dirección de los vectores de movimiento de los objetos dentro de la escena para su funcionamiento. A través de estos contadores no solo podemos saber cuántos objetos acceden al área de monitoreo, también, nos da una clara idea de cuántos de estos objetos entran o cruzan el contador inferior el cual indica acceso al área restringida. De esta manera podemos tener un mejor control en la zona y un mejor registro del movimiento para su futuro análisis.



Fig. 3.25. Contadores De Movimiento Relevante.

La magnitud de los objetos nos ayuda a determinar cuando el movimiento podría ser relevante y cuando podría tratarse de un movimiento irrelevante con trayectoria relevante, como se muestra en la figura 3.25. Donde el objetivo está generando un movimiento irrelevante pero a su vez podemos invertir el ángulo de búsqueda para poder realizar un seguimiento durante el monitoreo.



Fig. 3.26. Seguimiento de Movimiento.

En la figura anterior podemos ver dos resultados distintos, el resultado de la izquierda es obtenido sin la etapa de filtrado, solamente estableciendo el ángulo de búsqueda aplicado a los vectores de movimiento, incluso es posible notar que el seguimiento se acorta en la parte superior de la escena. En cuanto al segundo resultado es posible observar que la etapa de filtrado nos da un rastreo más suave y con mayor precisión dentro de la escena, podemos notar en la parte superior de la escena que el rastreo es más exacto comparado con el resultado anterior.

3.10. Magnitud

El número mínimo de vectores de movimiento que deberá existir para decir que se trata de un objeto importante es de 9 a 12, y a través de estas magnitudes es cómo podemos saber el número de objetos relevantes en la escena.

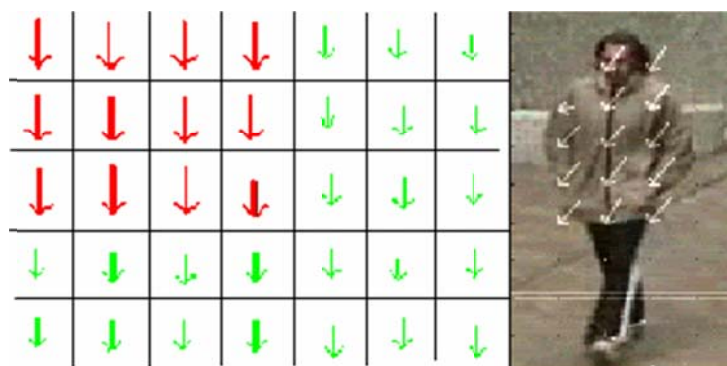


Fig. 3.27. Magnitud Generada Por Un Objeto Relevante.

Una vez establecidas las restricciones en el sistema, este es capaz de detectar, el número, trayectoria y nivel de importancia de movimiento(s) relevante(s) dentro de la escena, con un alto índice de exactitud en la detección y en diferentes ambientes a comparación del primer sistema propuesto en la maestría.



3.11. Conclusiones

Los sistemas de detección de movimiento se componen principalmente de restricciones en la escena, para una mejor detección. Actualmente la demanda en este tipo de sistemas ha ido creciendo, pero, poder desarrollar un algoritmo capaz de detectar cualquier movimiento en cualquier tipo de escenario utilizando una sola cámara de video es prácticamente imposible, debido a las limitantes de la tecnología actual.

El uso de macrobloques para el cálculo de los vectores de movimiento, ayuda a disminuir la complejidad del sistema, ayudando a enfocarnos solo en el procesamiento de estos, obteniendo un mejor resultado en la detección. La etapa de filtrado de vectores de movimiento es la más importante, ya que de ella dependen las restricciones del algoritmo. Esta etapa nos ayuda a eliminar los vectores de ruido en la escena dejando solamente los vectores que representan el movimiento real, sin embargo esta etapa depende de una correcta distancia entre la cámara y el objeto para un buen funcionamiento, en otras palabras, cada una de las restricciones en el algoritmo dependerán unos de otros para un correcto funcionamiento del algoritmo.



CAPITULO 4

Resultados Experimentales

Introducción

En este capítulo se compara el sistema propuesto con sistemas u algoritmos actuales o que toman las mismas bases para realizar la detección del movimiento en la escena. Las primeras evaluaciones realizadas con el sistema basado en macrobloques tanto en sitios abiertos como cerrados, mostrando las ventajas que este representa el usar los vectores de movimiento. Al final del capítulo, se muestra una evaluación global del sistema comparando con sistemas de detección propuestos en la actualidad, mostrando las ventajas que muestra la etapa de procesamiento aplicada.

4.1. Sistema Propuesto Doctorado

En los resultados obtenidos con el sistema anterior, puede observarse una correcta detección de movimiento bajo un gran número de restricciones, las cuales hacen que su aplicación para fines de seguimiento sea mínima a pesar de que la etapa de procesamiento logra eliminar gran parte de la información irrelevante.

Tomando estos puntos en cuenta, decidimos trabajar sobre la teoría que maneja el flujo óptico (sección 2.9.2) pero utilizando el método de macrobloques para el cálculo de los vectores de movimiento. Logrando así obtener un mejor sistema que logra cubrir las deficiencias del primero, manteniendo un buen índice de detección en las pruebas bajo ciertas restricciones.

4.2. Detección de Movimiento

Las secuencias de video utilizadas tienen una resolución media, por lo que se conoce con anterioridad el resultado que deberá presentar el algoritmo durante la detección. Las escenas muestran áreas de libre tránsito, la elección de estas áreas, es, debido a que en ellas podemos encontrar una gran cantidad de movimiento tanto relevante como no relevante, esto con la finalidad de probar la eficiencia del algoritmo en ambos aspectos.



Fig. 0.1. Secuencias De Video de Movimiento Simple y Complejo.

La primera secuencia de video mostrada en la figura 4.6 muestra la primera prueba realizada con el sistema, en esta secuencia de video se muestra una persona generando movimiento de afuera hacia adentro de acuerdo a la posición de la cámara, la detección de movimiento es simple desde que no se encuentra algún otro objeto generando movimiento durante la secuencia. La figura 4.7 muestra el resultado obtenido por el sistema durante el proceso de detección.



```

      READY TO MONITOR THIS AREA
*****
*TOP side 1 person(s) Detected*
*****
ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
>>>>      1 Person(s)      <<<<
ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
  
```

Fig. 0.2. Resultado De La Primera Prueba De Detección.



Como se puede apreciar en el resultado, el sistema es capaz de detectar el movimiento relevante en la escena, además de indicarnos en donde se originó el movimiento, el número de objetos que lo generaron y el número de objetos que accedieron.

En la figura 4.8 se muestra una secuencia de video con un movimiento irrelevante. Este movimiento se considera irrelevante ya que se genera de abajo hacia arriba, y de acuerdo a las restricciones y a los criterios que se tomaron para el filtrado de los vectores de movimiento, se considera movimiento relevante todo aquel movimiento generado de afuera hacia adentro, tomando en cuenta la posición de la cámara.



Fig. 0.3. Secuencia de Video, Movimiento Irrelevante.

El resultado generado en esta secuencia es, la ausencia de generación de vectores de movimiento durante el análisis de la secuencia de video y al ser movimiento irrelevante el sistema no lo toma en cuenta, por lo que el sistema solo despliega en pantalla el mensaje de monitoreo como se muestra en la figura 4.9.

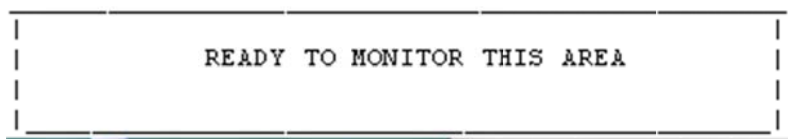


Fig. 0.4. Resultado De La Segunda Prueba De Detección.



De igual manera se realizan las detecciones de movimiento en las siguientes secuencias de video, pero con la diferencia de que el número de objetos relevantes e irrelevantes cambia al igual que la trayectoria de cada uno de los objetos en la escena.

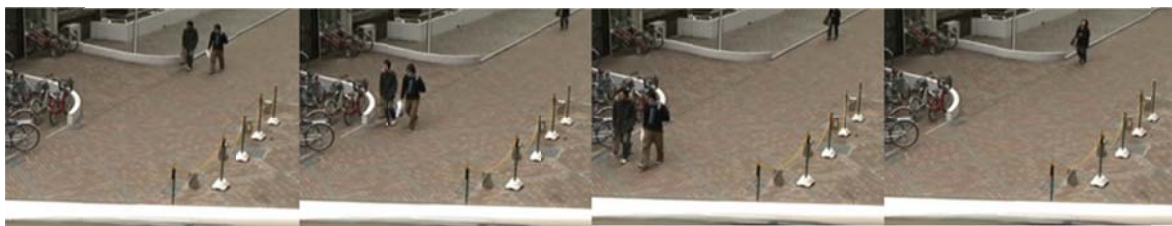


Fig. 0.5. Tercera Secuencia De Video, Tres Movimientos Relevantes En Escena.

En la figura 4.10 se muestra la tercera secuencia de prueba, donde podemos encontrar 3 movimientos relevantes, el resultado de esta prueba se muestra en la figura 4.11.



```

|-----|
|               |
|   READY TO MONITOR THIS AREA   |
|               |
|-----|
*****
*TOP side 2 person(s) Detected*
*****
ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
>>>> 2 Person(s) <<<<
ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
*****
*TOP side 1 person(s) Detected*
*****

```

Fig. 0.6. Resultado De La Tercera Prueba De Detección.

La secuencia de video de la cuarta prueba es mostrada en la figura 4.12. En esta secuencia podemos observar una persona generando movimiento relevante y varias personas generando movimiento irrelevante al mismo tiempo, el resultado de la prueba de detección se muestra en la figura 4.13.



Fig. 0.7. Cuarta Secuencia De Video, Un Movimiento Relevante En Escena.

En esta secuencia (figura 4.12) se observa que la detección de movimiento es compleja desde que existen cuatro objetos generando movimiento irrelevante al mismo tiempo que el movimiento relevante aparece en la escena, sin embargo, el sistema de detección muestra un correcto resultado (figura 4.18) durante la detección, mostrando, de esta manera que es capaz de discriminar información irrelevante de una manera correcta.





```

|-----|
|               READY TO MONITOR THIS AREA               |
|-----|
|*****|
|*TOP side 1 person(s) Detected*|
|*****|
|ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!|
|>>>>      1 Person(s)      <<<<|
|ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!|

```

Fig. 0.8. Resultado De La Cuarta Prueba De Detección.

La figura 4.14 muestra la quinta secuencia de prueba. Los resultados pueden observarse en la figura 4.15.

En el resultado de esta prueba puede observarse que el sistema marca tres movimientos relevantes cuando existen cuatro, esto se debe a que la secuencia de video termina antes de que puedan apreciarse completamente los dos últimos objetos, por lo que el sistema no es capaz de reconocer el número de objetos en su totalidad, registrando de esta manera los dos movimientos relevantes completos y tomando a los dos últimos objetos como un solo movimiento relevante.



Fig. 0.9. Quinta Secuencia De Video, Dos Movimientos Relevantes En Escena.

```

|-----|
|               READY TO MONITOR THIS AREA               |
|-----|
|*****|
|*TOP side 1 person(s) Detected*|
|*****|
|*****|
|*TOP side 1 person(s) Detected*|
|*****|
|*****|
|*TOP side 1 person(s) Detected*|
|*****|
|*****|

```

Fig. 0.10. Resultado De La Quinta Prueba De Detección.

En la secuencia de video de la figura 4.16 podemos observar dos movimientos relevantes los cuales aparecen de ambos lados de la escena, al final la trayectoria de ambos objetos se convierte en irrelevante por lo que el sistema solo genera vectores de movimiento cuando los objetos generan movimiento dentro del rango establecido para después dejar de generarlos al convertirse en movimiento irrelevante. El resultado de esta prueba se muestra en la figura 4.17.



Fig. 0.11. Sexta Secuencia De Video, Dos Movimientos Relevantes.

```

|-----|
|               READY TO MONITOR THIS AREA               |
|-----|
*****
*RIGHT side 1 person(s) Detected*
*****
*****
*LEFT side 1 person(s) Detected*
*****

```

Fig. 0.12. Resultado De La Sexta Prueba De Detección.

4.2.1. Seguimiento del Movimiento

Es importante mostrar que la etapa de filtrado y las restricciones implementadas al momento de tomar la muestra necesario de vectores de movimiento de acuerdo al desplazamiento de la persona, es en gran medida una parte medular del correcto



funcionamiento del sistema.

En la figura 4.18 podemos observar cómo se vería el desplazamiento de un objetivo sin la etapa de filtrado y la mejora que se puede observar al implementar el filtro de mediana entre el cálculo del vector y la restricción es el ángulo de su trayectoria.

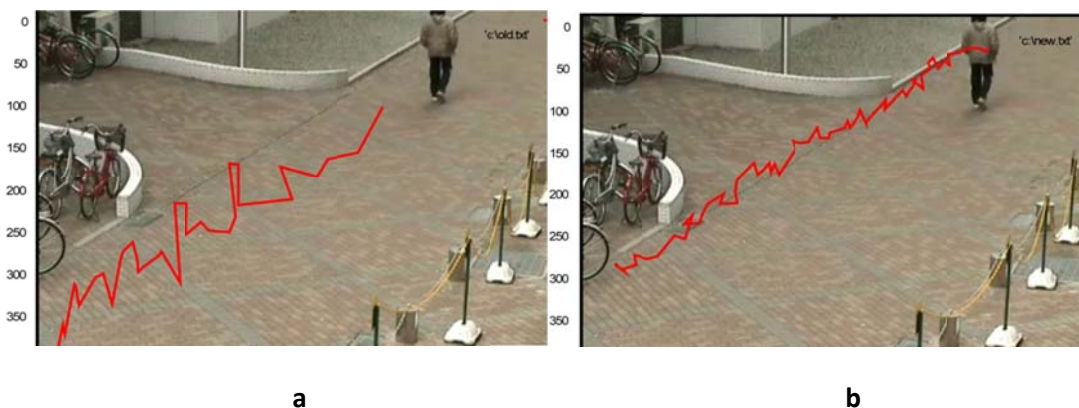


Fig. 0.13. Seguimiento del movimiento a) sin etapa de filtrado b) Con etapa de filtrado.

Se puede observar en los resultados anteriores que el sistema tiene buenos resultados durante la detección de movimiento, cuando es utilizado tomando en cuenta las restricciones que se mencionan en el capítulo 3. La distancia que exista entre el objeto a analizar y la cámara será de gran importancia para una correcta detección, si nosotros no tomamos en cuenta estas restricciones el sistema simplemente no podría funcionar de una manera correcta, como se muestra a continuación.

4.2.2. Pruebas De Detección Erróneas.

Si no tomamos en cuenta la distancia que deberá de haber entre el objeto y la cámara durante el proceso de detección, el resultado final será una lectura errónea de la secuencia de video, debemos recordar que una herramienta principal para poder determinar que el movimiento detectado es generado por un humano es la magnitud de vectores de movimiento, si la distancia no es la correcta o la cámara se encuentra muy cerca del objeto esta magnitud cambiaría por lo que la lectura o resultado final también como se muestra en las siguientes figuras.



Fig. 0.14. Secuencia de Video, Primera Prueba Errónea.

```

|-----|
|          READY TO MONITOR THIS AREA          |
|-----|
*****
*TOP side 1 person(s) Detected*
*****
*****
*TOP side 1 person(s) Detected*
*****
ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
>>>>      1 Person(s)      <<<<
ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
*****
*TOP side 1 person(s) Detected*
*****
ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
>>>>      3 Person(s)      <<<<
ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!

```

Fig. 0.15. Resultado De La Primera Secuencia Errónea.

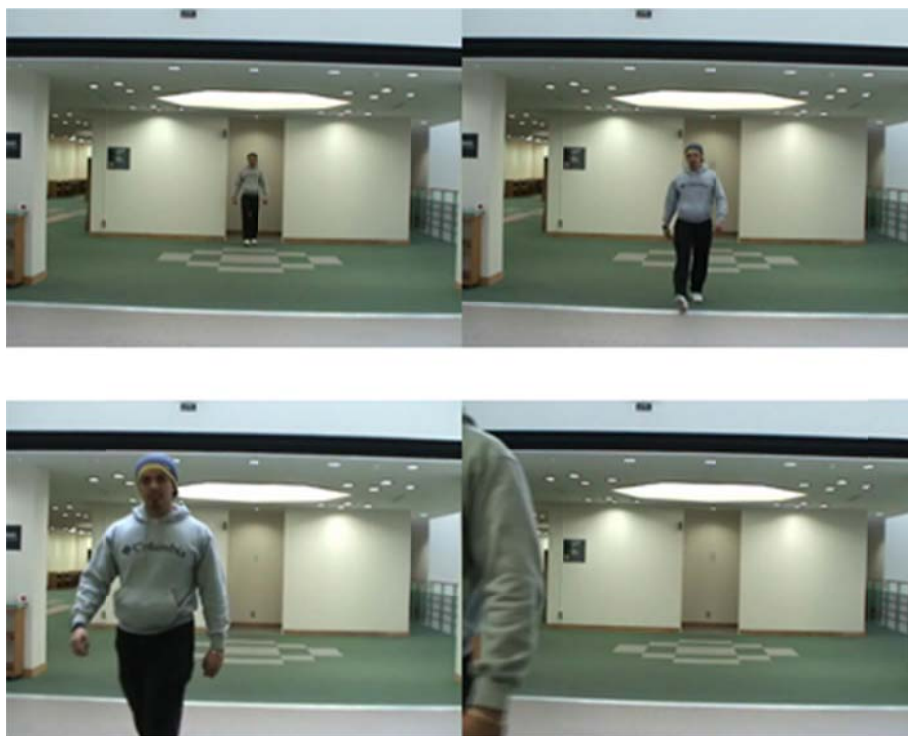


Fig. 0.16. Secuencia De Video, Segunda Prueba Errónea.

```

|-----|
|               READY TO MONITOR THIS AREA               |
|-----|
ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
>>>>      1 Person(s)      <<<<
ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
>>>>      1 Person(s)      <<<<
ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
*****
*TOP side 2 person(s) Detected*
*****

```

Fig. 0.17. Resultado De La Segunda Secuencia Errónea.



Fig. 0.18. Secuencia De Video, Tercera Secuencia Errónea.

```

|-----|
|               READY TO MONITOR THIS AREA               |
|-----|
ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
>>>>      1 Person(s)      <<<<
ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
>>>>      2 Person(s)      <<<<
ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
*****+*****
*LEFT side 1 person(s) Detected*
*****
ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
>>>>      4 Person(s)      <<<<
ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
*****
*TOP side 3 person(s) Detected*
*****

```

Fig. 0.19. Resultado De La Tercera Secuencia Errónea

3.2. Comparación de Sistemas.

Analizando los resultados obtenidos el sistema propuesto, podemos observar claras diferencias entre ellos, como son las ventajas que presenta el nuestro sistema frente a los demás (contadores, detección de movimiento humano basado en la magnitud, uso en diferentes escenarios, análisis de trayectorias, etc.).

Una de las principales ventajas que muestra el sistema es que es capaz de detectar más de un movimiento relevante dentro de la escena, mientras que el sistema 1 está diseñado para detectar solo movimientos simples, es decir, un solo movimiento relevante en la escena.



3.3. Conclusiones.

Se han comparado diferentes sistemas para realizar la detección de movimiento relevante. Los mejores resultados presentados se dan en el sistema propuesto, a pesar de que los otros son capaces de realizar una correcta detección de movimiento bajo ciertas restricciones, no brindan la estabilidad adecuada para su uso en situaciones diversas.

Nuestro sistema fue estudiado y simulado con la idea de crear un sistema de detección, capaz de detectar movimiento relevante bajo cualquier circunstancia y con la mínima intervención humana durante su funcionamiento. Los resultados obtenidos presentan un sistema capaz de detectar y rastrear movimiento relevantes bajo un gran número de circunstancias, permitiendo también, un análisis de la trayectoria de los objetos en movimiento. Los resultados obtenidos pueden ser comparados con los resultados de sistemas actuales que se encuentran hoy en el mercado, dando como resultado final, una aplicación real del sistema en sistemas de detección.



CAPITULO 5

Conclusiones Generales

Introducción

Durante la realización de este trabajo se propuso un sistema capaz de superar las limitantes presentadas en el sistema anterior, el uso de macrobloques como técnica para el cálculo de vectores nos ayuda a reconocer de manera más precisa los objetos en movimiento dentro del escenario y de esta manera es posible para nosotros basarnos en la dirección y magnitud de cada uno de ellos para el reconocimiento de movimiento. Este capítulo menciona las conclusiones generales de ambos sistemas, dejando notar los trabajos futuros que se pueden realizar combinando los resultados de nuestro segundo sistema para que sea capaz no solo de detectar si no también de reconocer.

5.1. Conclusiones

A pesar de los grandes avances tecnológicos y del surgimiento de diferentes investigaciones en el área de procesamiento de imágenes, que tienen como objetivo principal, crear un sistema capaz de imitar al Sistema Visual Humano, con el fin de lograr un sistema de detección de movimiento capaz de trabajar de forma automática, actualmente, no existe un sistema de detección de movimiento capaz de detectar y clasificar cualquier objeto de acuerdo a las características que presentan en su movimiento y que a su vez sea capaz de trabajar en todas las áreas posibles.

Los sistemas propuestos plantean una solución a las desventajas o necesidades que presentan algunos sistemas de detección de movimiento actuales, como es el alto costo de los equipos o la alta complejidad computacional en su funcionamiento y teniendo como meta final el poder crear un sistema capaz de acercarse al objetivo principal de esta área.

En estos sistemas se elimina la información innecesaria dentro del cuadro para poder realizar una correcta detección de movimiento, el primero lo logra trabajando sobre la información espacial contenida en cada cuadro logrando eliminar casi en su totalidad las variaciones de luz y sombra presentes en la secuencia de video. Por el contrario, el segundo sistema basa esta eliminación tanto en la información espacial como en la temporal de la secuencia de video, por medio de un proceso de detección de vectores de movimiento por medio de macrobloques y un filtrado de vectores de movimiento.

La detección de movimiento en ambos casos, es basado en la magnitud que presenta el objeto en el cuadro, la principal diferencia es que, el primer sistema se basa en la información total contenida en cada cuadro (píxeles) para calcular la magnitud y poder determinar el nivel de importancia del movimiento, mientras que el segundo sistema utiliza la cantidad de vectores de movimiento generados después de la etapa de preprocesamiento para poder determinar la magnitud del o los objetos contenidos en la secuencia. Ambos sistemas presentan una baja complejidad computacional y un correcto funcionamiento como muestran los resultados, siempre y cuando apliquemos las restricciones mencionadas.



De esta manera logramos cubrir parte del objetivo de nuestra propuesta, crear un método de detección de movimiento más robusto, confiable y menos complejo capaz de adaptarse a otros métodos con el fin de acercarse en cierta medida al sistema visual humano.

5.2. Investigación Futura.

Hoy en día el reconocimiento de personas es una realidad, contamos con algoritmos de reconocimiento de rostro, reconocimiento por medio de huella digital, reconocimiento por medio de retina, y la “huella digital” actual, reconocimiento a través de la forma de caminar.

La idea global de esta investigación es realizar un sistema no solo capaz de detectar un movimiento dentro de la escena, también deberá de ser capaz de reconocer a la persona o personas que están realizando ese movimiento.

La combinación de técnicas para lograr sistemas híbridos es ya una necesidad, los algoritmos existen de manera individual pero la combinación de técnicas para lograr eficientar su funcionamiento aun no es muy clara. En estos momentos tenemos una propuesta teórica sólida, basada en la teoría actual.

En el siguiente capítulo discutiremos con mayor detalle los puntos antes mencionados, dando un argumento teórico sólido de nuestra propuesta futura.



CAPITULO 6

Reconocimiento de Personas (Investigación Futura).

Introducción

La detección, autenticación o reconocimiento de humanos siempre ha sido un tema constante de investigación. La revisión de la literatura contemporánea (2009) revela interés en obtener resultados computacionalmente confiables, donde las personas estén alejadas y no tengan que interactuar directamente con la fuente de captación, con el objetivo de realizar labores de video-vigilancia. La modalidad biométrica que más se ajusta a estas condiciones, es la forma de caminar (en inglés Gait).

6.1. Biometría

La biometría es la ciencia que se encarga de identificar, reconocer o autenticar personas, basándose en sus características fisiológicas o de comportamiento (Fig. 6.1). El autor da su apreciación en alto, medio y bajo (tabla. 6.1), atendiendo a los requerimientos que siguen:

Universalidad: cuan común es encontrarlo en los individuos.

Individualidad: que tan único o diferente es la huella biométrica.

Permanencia: que tan invariante es durante un periodo de tiempo.

Colección: que tan fácil es la adquisición, medición y almacenamiento.

Atendiendo a cuestiones prácticas también se puede tener en cuenta:

Ejecución: que tan preciso, veloz y robusto es el sistema en el manejo de la huella biométrica.

Aceptación: Que tanta aprobación tiene la tecnología entre el público.

Evasivo: Que tan fácil es engañar al sistema de autenticación.

En la (Fig. 6.1) se ubica la forma de caminar como un comportamiento de la persona, sin embargo, en [2] la trata mediando entre fisiológica y de comportamiento, debido a que aspectos físicos, como el peso de la persona, longitud del paso, longitud del pie, etc., incluyen en la forma de caminar.

Cada una de estas modalidades biométricas (tabla 6.1), se ajustan para su uso según sus Fig. 1. Clasificación de las modalidades biométricas limitantes, aunque en la mayoría se precisa de cierta resolución, calidad (imagen, video etc.) y cercana o trato directo con las personas para su uso (rostro, ADN, huellas, oreja etc.).

Un área donde se requiere del reconocimiento de individuos habiendo baja resolución, oscurecimiento, movimiento a distancia, etc., es en el tema de video-vigilancia. Además, en cuestiones de delitos o prófugos, los individuos pueden ocultar su rostro sus manos, etc. La modalidad biométrica que reúne estas circunstancias es la forma de caminar (en inglés Gait). La necesidad de su uso está presente en una considerable cantidad de trabajos



referenciados en [3,4] y más reciente [5,6,7,8,9,10,11,12,13,14,15,16,17,18,19], en busca de lograr el reconocimiento de las personas, usando su forma de caminar y aumentar su contabilidad. Esta modalidad biométrica está muy justificada, primeramente por estudios realizados por los pioneros de ella (M. Nixon, T. Tan, R. Chellappa), que referencian evidencias en la literatura, ciencias de la psicología y medicina [20], pero en este momento (2009), hay suficientes trabajos con resultados positivos sobre las bases de datos ya creadas (Ej. sección (5), UMD Maryland, CMU shilouette, CASIA, Soton aurdor, etc.), que demuestran su potencial futuro. Aunque, aun no se han alcanzado resultados para considerar la forma de caminar, un sistema de reconocimiento computacionalmente contable, esto se verifica en una nota emitida (2009) [21], referida al Centro de Excelencia Biométrica del FBI, donde están considerando los avances de las técnicas en esta área (Gait), esperando que se convierta en una tecnología lo suficientemente contable, para poder incluirla en su sistema de Identificación de Última Generación (NGI, siglas en ingles), para próximas versiones de su sistema de identificación multimodal. Considerando que contamos con una secuencia de video en la que se ha extraído satisfactoriamente el recorrido de una persona al caminar, obviando algunos pasos como modelado del fondo, segmentación del movimiento etc., o como se encuentra en la mayoría de las bases de datos, entonces, podemos comenzar, como en toda tarea de reconocimiento, en lograr un conjunto mínimo de características que discriminen, al menos, a los entes presentes en la base de datos. Estas características se ajustan a la forma en que se ha modelado el cuerpo humano.

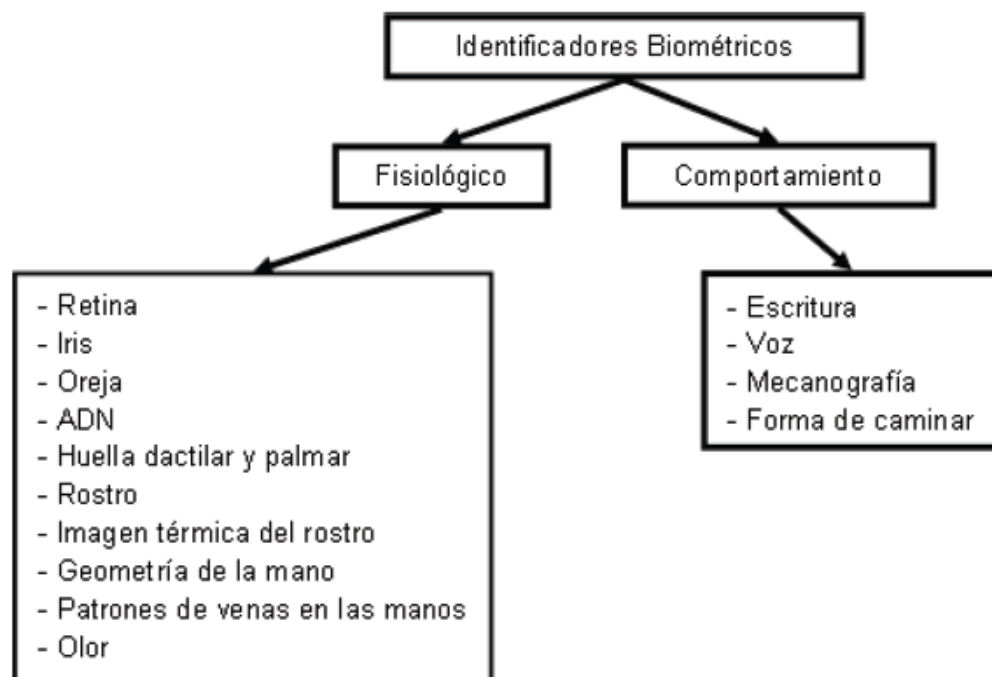


Fig. 6.1. Clasificación de las modalidades biométricas.



Identificador Biométrico	Universalidad	Individualidad	Permanencia	Colección	Ejecución	Aceptación	Evasivo
Retina	H	H	M	L	H	L	L
Iris	H	H	H	M	H	L	L
Iris	H	H	H	M	H	L	L
Oreja	M	M	H	M	M	H	M
ADN	H	H	H	L	H	L	L
Huella dactilar y palmar	M	H	H	M	H	M	M
Rostro	H	L	M	H	L	H	H
Geometría de la mano	M	M	M	H	M	M	M
Venas en las manos	M	M	M	M	M	M	L
Olor	H	H	H	L	L	M	L
Escritura	L	L	L	H	L	H	H
Escritura	L	L	L	H	L	H	H
Voz	M	L	L	M	L	H	H
Mecanografía	L	L	L	M	L	M	M
Forma de caminar	M	L	L	H	L	H	M

Tabla. 6.1. Comparación de modalidades biométricas en H (alta), M (Medio), L (Baja).

6.2. Basado en Modelos

La forma de caminar de las personas es descrita por un modelo del cuerpo, que se ajusta a este en cada cuadro de la secuencia del video a lugares característicos del cuerpo (puntos de unión entre extremidades) (Fig. 6.2). La creación de la base de datos o el estudio de la selección de estos puntos es lograda, en muchos casos, usando marcas (puntos de luz, sensores, etc., ajustados al cuerpo) o más reciente [5,8] usando la silueta del contorno (Fig. 6.6b). Cuando se requiere de su detección automática, esto suele ser de cierta dificultad [5,8].

Obtenida la secuencia de modelos se pueden extraer parámetros relacionados con la forma del cuerpo (ancho, alto, longitud de las extremidades, etc.) o en función del movimiento (longitud del paso, tiempo del ciclo, velocidad, ángulos de unión entre las extremidades, etc.) (Fig. 6.4). Se plantea en [22,4,5,8], que la extracción de los parámetros angulares en estos modelos, provenientes de videos 2D, resultan ser de gran dificultad, debido a oclusiones, tipo de vestuarios, objetos que puede portar el hombre, error de segmentación, diferentes ángulos de vista de la cámara, etc. , además, el costo computacional de los algoritmos basados en modelos es relativamente alto. Aunque, se exploran variantes recientes por reconocidos investigadores del tema [8,5], usando descriptores elípticos de Fourier para localizar estos puntos característicos. Algunos trabajos que han usado modelado pueden ser encontrados en [4,22,3].

Recientemente [5,8,9,11,10] para extracción de las características y en [23] en estudios de factores que dificultan la detección, reconocimiento y/o identificación. La representación basada en modelos no es robusta a cambios de escala, posición, necesita buena calidad en la secuencia que se capta de "Gait 2 demanda alto costo computacional [19].



Fig. 6.2. Poses del cuerpo al caminar, modelado usando segmentos.

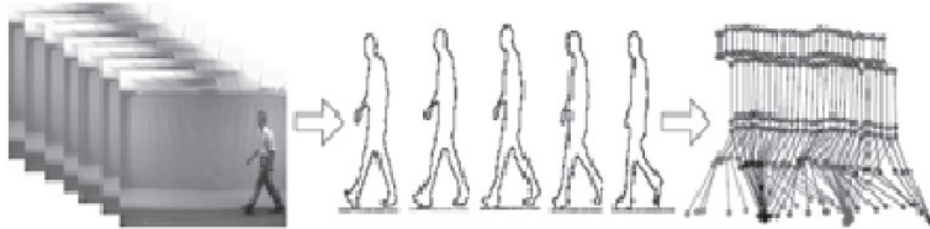


Fig. 6.3. Extracción de modelos usando contorno de siluetas.

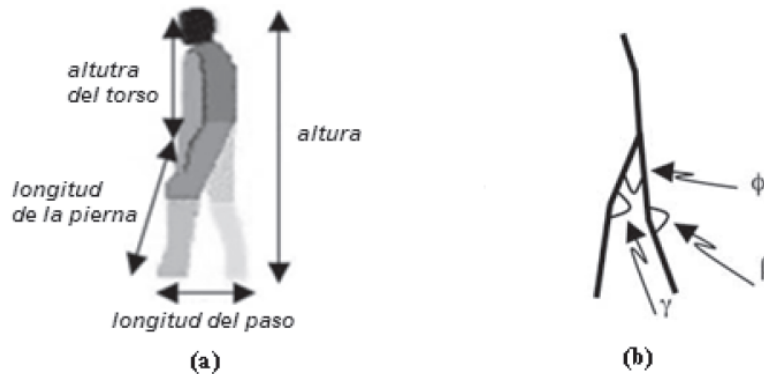


Fig. 6.4. (a) Parámetros de la Forma, (b) parámetros del movimiento.

6.3. Basado en Silueta

La silueta es determinada para cada imagen de la secuencia del video, como la región que enmarca la presencia (si existe) de una o varias personas. Es muy común ver pasos como: extracción del fondo y segmentación en dos clases (fondo y personas) [22], (Fig. 6.5). Otros disminuyen información quedándose solo con el borde de la región en blanco, nombrándola silueta del contorno (Fig. 6.6b). También es logrado algo muy parecido a una silueta, con información proveniente de aplicar el flujo óptico a la secuencia de las imágenes de videos [25] (Fig. 6.7), este método es sensible a cambios de brillo [25]. La información del tiempo, como antes se dijo, puede ser usada para lograr un volumen 3D, es decir, $(2D + t)$ de la secuencia de siluetas de poses en cada cuadro del video (Fig. 8), el uso de esta representación es vista en [7].



Fig. 6.5. (a) Fondo, (b) Fondo más Persona, (c) Imagen Segmentada en dos Clases.



Fig. 6.6. (a) Imagen Segmentada (Silueta), (b) Detección de Borde (Contorno de la silueta).

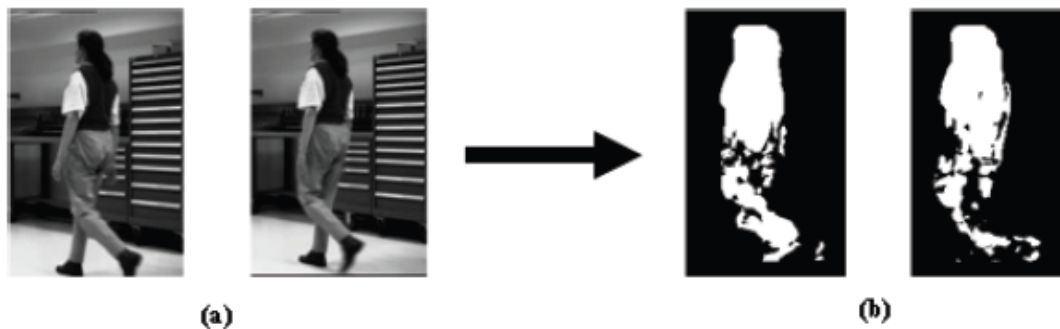


Fig. 6.7. Imágenes Consecutivas de un video, (b) Región en Blanca los puntos que se han movido.

La mayoría de los investigadores han usado las siluetas, obteniendo el vector de características basado en la forma o en el movimiento del sujeto, es decir, consideran el movimiento del cuerpo como una secuencia estática de poses o la distribución generada en un espacio (XYT) por el cuerpo al caminar. Una relación de artículos referenciados, usando estos tipos de representación es vista en [22,3]. Aportes más actuales usando silueta son observados en [12,13,9,7,14,15,16,17,6,18,19]. A partir de las siluetas también es posible obtener cierta información del movimiento o de la estructura, independientemente; en un inicio se tiene el ancho, centro y alto de la persona y de estos, según estudios anatómicos



[26], se logra la posición vertical del pecho y tobillos el ancho del pecho, longitud del paso, etc., y de esta manera se refuerza el poder discriminador del vector que describe la forma de caminar (Fig. 6.9) [4].



Fig. 6.8. Volumen $(2D+t)$ de la secuencia de Siluetas.

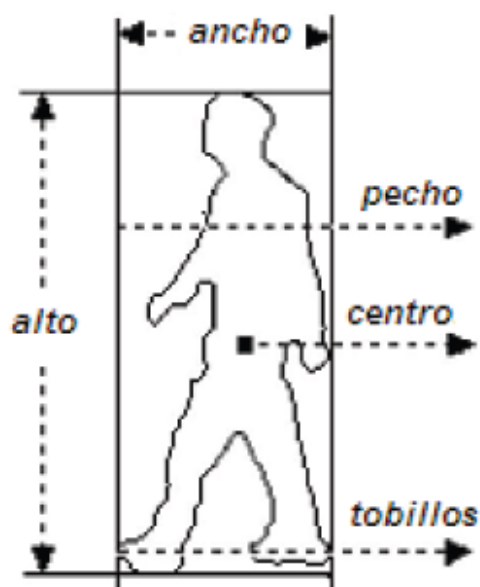


Fig. 6.9. Características extraídas del contorno de la silueta.



6.4.Revistas y Eventos que Publican del Tema

6.4.1. Revistas

1. IEEE Trans. Systems, Man and Cybernetics.
2. IEEE Trans. Pattern Analysis and Machine.
3. Proceedings of IEEE.
4. IEEE Trans. Image Processing.
5. EURASIP Journal on Advances in Signal Processing.
6. AI in Engineering.
7. IEEE Trans. Circuits Syst. Video Techn.
8. IEEE Trans. Pattern Anal. Mach. Intell (PAMI).
9. The Journal of Bone and Joint Surgery.
10. International Journal of Biometrics (IJBM).

6.4.2. Eventos

1. International Conference on Pattern Recognition.
2. IEEE Workshops on Application of Computer Vision.
3. IEEE Transactions on Information Forensics and Security.
4. International Conference on Patter Recognition (ICPR).
5. IEEE International Conference on Biometrics.
6. Technical Committee on Biometrics (IAPR TC4).
7. International Summer School on Biometrics, every year in Alghero, Italy.
8. International Conference on Audio and Video-Based Biometric Person Authentication.
9. IEEE Conference on Advanced Video and Signal Based Surveillance.
10. IEEE International Conference on Biometrics, Identity and Security (BIIdS), organized
11. and sponsored by the IEEE Biometric Council.
12. ASEAN Forum on Biometric Interoperability.

6.5.Investigadores Destacados

- Marck Nixon, University of Southampton, UK.
- Teniu Tan. Chinese Academy of Science Institute.
- Rama Chellappa (College Park, MD, USA).
- Massimo Tistarelli, University Sassari, Alghero, Italy.
- Stan Li, Chinese Academy of Science.
- Sudeep Sarkar, University of South Florida.



6.6.Conclusiones

Sin duda, el reconocimiento de personas por la forma de caminar (gait), está siendo un área de gran atención por los investigadores de la biometría en estos momentos (2009), con intereses marcados en la video-vigilancia, debido a que ofrece ventajas, que pocas modalidades biométricas poseen, al conseguir las muestras a distancia con baja resolución y sin necesidad de una atención directa de las personas a la fuente de captación (cámara). Matemáticamente se han obtenido buenos resultados en alcanzar un espacio de características invariante a cambios de posición, escala y rotación, pero los esfuerzos en la actualidad, van dirigidos a lograr un conjunto de características en el andar del hombre, las cuales sean invariantes a factores como: diferentes _ángulos de vista, tipo de ropa, calzado, velocidad, terreno, edad, objetos llevados etc. (sección 1.2). Hay aportes que tratan de fusionar la forma de caminar con otras modalidades, como rostro, lo cual no es ventajoso, porque se prescindirá de cierta resolución y la atención de la persona a la cámara, eliminando con esto la mayor ventaja de esta modalidad. Otra variante con muy buen futuro, está en el enriquecimiento usando la biometría suave.



PUBLICACIONES

➤ **Relevant Motion Detection From Motion Vectors in MPEG Stream.**

Josue A. Hernandez Garcia, Hiroyoshi Morita, Hua Yu, I. G. B. Baskara Nugraha, Hector Perez Meana.

IPS Information Processing Society of Japan, Forum on Information Technology FIT 2006, 5th Information Science and Technology Forum, Pages 1-5, September 2006.

➤ **Direction-Based Motion Detection Using the MPEG2 Encoder.**

Josué A. Hernández-García, Hiroyoshi, Morita, Héctor Pérez-Meana, M. Nakano-Miyatake.

2nd WSEAS Int. Conf on COMPUTER ENGINEERING and APPLICATIONS (CEA'08) Acapulco, Mexico, January 25-27, 2008 ISSN: 1790-5117 ISBN: 978-960-6766-33-6 Page 199

➤ **Movement Detection and Tracking Using Video Frames.**

Josue Hernandez¹, Hiroshi Morita², Mariko Nakano-Miyatake¹, and Hector Perez-Meana¹.

14th Iberoamerican Conference on Pattern Recognition, CIARP 2009. Guadalajara ,Jalisco, Mexico, November 2009. LNCS 5856, ISSN 0302-9743 pages 1054-1061.

➤ **Detection and Tracking of Human Movement Based on the Motion Vectors Estimation.**

Josué Hernández García¹, Hiroshi Morita², Mariko Nakano Miyatake¹, Héctor Pérez Meana¹

Rev. Fac. Ing. Univ. Antioquia N.º 56. Diciembre 2010 56 pp. 142-151. Diciembre, 2010.



Referencias Bibliográficas.

- [1] Arturo de la Escalera, "Computation Vision", Pearson Education S.A., Madrid 2001, Ed. Prentice Hall, pp. 1-3, 60-80.
- [2] Claudette Cedras and Mubarak Shah, "Motion-Based Recognition", Computer Vision lab, Department of Science, University of central Florida Orlando, pp. 129-144.
- [3] Bernand Jähne, "Digital Image Processing 5th edition", Ed. Springer, pp. 376-400.
- [4] Klette Zamperoni, "Handbook of Image Processing Operators", Ed. Wiley, pp. 114-150.
- [5] Ajit Singh, "Optic Flow Computation", IEEE Computer Society Press, pp. 20-43.
- [6] B. Jähne, "Motion Determination in Space-Time Image", Springer-Verlag, Antibes France, pp. 161-173.
- [7] Jenkis, S.F.; D.J. Morton; P.D. Dukes, "Comparison of techniques for detection of reverse motion in human movement". 1967, pp. 25-27
- [8] Steven M. Seitz , Charles R. Dyer, View-Invariant "Analysis of Cyclic Motion, International Journal of Computer Vision", v.25 n.3, pp.231-251, Dec. 1997.
- [9] J.E. Cutting D.R. Proffitt and L.T. Kozlowski, "A Biomechanical Invariant for Gait Perception," . Experimental Psychology: Human Perception and Performance, vol. 4, no. 3, pp. 357-372.
- [10] Johansson, "Visual Perception of Biological Motion and a Model for Its Analysis," Perception & Psychophysics, vol. 14, no. 2, pp. 201-211, 1973.
- [11] David Stavens, "Optical Flow Estimation", Stanford Artificial Intelligence Lab., pp 33-43.
- [12] Víctor Paladino, "Introducción a la Compresión de Video Bajo el Estándar MPEG-2", Instituto de Ingeniería Eléctrica de la Facultad de Ingeniería, Diciembre de 2005, pp. 3-24.
- [13] P.N. Tudor, "MPEG-2 Video Compression, Tutorial", Electronics & Communication Engineering Journal, December 1995, pp 1-14.
- [14] Meter Symes, "Digital Video Compression Copyrighted Material", Ed. McGraw-Hill, 2003, pp. 1-47.
- [15] J. R. Pierce, "An Introduction to Information Theory, Symbols Signals and



Noise”, Second Edition, pp. 78-100.

- [16] Engel and J. M. Rubin, “Detecting Visual Motion Boundaries”, In Proc. Workshop on Motion: Representation and Analysis, Charleston, S.C., May 1986, pp. 107—111.
- [17] Jorge Lira Chavez, “Introducción al Tratamiento Digital de Imágenes”, Ed. Ciencia de la Computación, 2002, pp. 16-36, 219-223.
- [18] Rafael C. Gonzalez, Richard E. Woods, “Digital Image Processing”, Ed. Addison-Wesley, pp. 21-25.
- [19] John C. Russ, “The Image Processing Handbook, Fourth Edition”, Ed. CRC Press 2002, pp. 1-8, 20-24, 265-269.
- [20] Karthik Hariharakrishnan, Dan Schonfeld, “Video Tracking Using Block Matching”, IEEE 2003 Volume III, pp. 945-948.
- [21] Guy Gilboa, Nir Sochen, “Image Enhancement and Denoising by Complex Diffusion Processes”, IEE Transactions on Pattern Analysis and Machine Intelligence vol. 26, 2004, pp. 1020-1035.
- [22] Longin Jan Latecki, Roland Mieziánko, “Motion detection based on local variation of spatiotemporal texture”, IEEE Computer Society Conference 2004, pp. 1-6.
- [23] C. Stauffer, W. E. L. Grimson, “Learning patterns of activity using real-time tracking”, IEEE 2000, pp. 747-754.
- [24] C. Wren, “Pfinder: Real-time tracking of human body”, IEEE 1997, pp. 780-785.
- [25] David Moore, “A real-world system for human motion detection and tracking”, California Institute of Technology 2003, pp. 3-36.
- [26] Arun Hampapur, Lisa Brown, “Smart Video Surveillance”, IEEE SPM, 2005, pp. 38-51.
- [27] Gian Luca Foresti, Christian Micheloni, “Smart Video-based Surveillance System”, IEEE Signal Processing Magazine, 2005, pp. 25-37.
- [28] Yang Song, Xiaolin Feng, “Towards Detection of Human Motion”, proceedings of CVPR 2000, vol I, pp.810-817.
- [29] Shinya Yamamoto, Yasushi Mae, “Realtime Multiple Object Tracking Based on Optical Flow”, IEEE International Conference on Robotics and Automation, 1995, pp. 2328-2333.



- [30] Yu Zhong, Anil K. Jain, “Object tracking using deformable templates”, IEEE TPAMI, 2000, pp. 544-549.
- [31] Rafael Molina Soriano, “Del Procesamiento a la Vision Artificial”, Depto Ciencias de la Computación e I.A., Universidad de Granada, pp. 3-29.

ANEXO A

PRINCIPALES PUBLICACIONES

Detección y rastreo de movimiento humano basado en la estimación de vectores de movimiento

Detection and tracking of human movement based on the motion vectors estimation

*Josué Hernández García¹, Hiroshi Morita², Mariko Nakano Miyatake¹,
Héctor Pérez Meana^{1*}*

¹ESIME Culhuacan, Instituto Politécnico Nacional, Av. Santa Ana 1000, Col.
San Francisco Culhuacan, 04430 México D. F. México.

²Universidad de Electro-Comunicaciones, 1-5-1, Chofugaoka, Chofu-shi.
Tokio. 182-8585. Japón.

(Recibido el 4 de septiembre de 2009. Aceptado el 6 de septiembre de 2010)

Resumen

Este artículo presenta el desarrollo de un algoritmo para la detección y rastreo del movimiento de seres humanos con baja complejidad computacional, basado en la estimación de vectores de movimiento. El sistema propuesto es capaz de diferenciar entre movimientos permitidos y no permitidos, sin la intervención de ningún operador humano. Los resultados experimentales proporcionados muestran un funcionamiento adecuado del sistema propuesto.

----- *Palabras clave:* Detección de movimiento, vectores de movimiento, seguimiento de trayectoria, movimiento relevante, estimación de vectores movimiento

Abstract

This paper proposes a human and vehicle motion detection and tracking system, based on motion vectors, which is able to discriminate between allowed and not allowed movements without the intervention of any human operator. Evaluation results show the desirable features of proposed method.

----- *Keywords:* Motion detection, movement detection, movement tracking, relevant motion, motion vectors estimation

* Autor de correspondencia: teléfono/fax: + 52 + 55 + 565 620 58, correo electrónico: hmperezm@ipn.mx. (H. Pérez)

Introducción

La detección automática de movimiento y el rastreo de objetos ha sido un tópico de activa investigación durante los últimos años, debido a su potencial impacto en el desarrollo de sistemas de seguridad confiables usando la información contenida en secuencias de video; junto con alguna otra información previamente adquirida relacionada con la forma, velocidad, dirección del movimiento y textura, etc. de los objetos bajo análisis contenidos en la secuencia de video. Estos esquemas han sido empleados en el desarrollo de diversos sistemas de monitoreo en exteriores, casetas de cobro, en edificios, estacionamientos, túneles, calles y puentes, así como en el desarrollo de sistemas de seguridad en bancos, edificios de gobierno, centros comerciales, estadios, centros penitenciarios y aeropuertos, etc. [1]. Todas estas aplicaciones requieren el desarrollo de sistemas de vigilancia inteligentes capaces de clasificar la mayoría de los movimientos con la mayor precisión posible. Adicionalmente estos sistemas deben ser robustos contra falsas alarmas, así como ser capaces de sincronizarse y operar junto con otros algoritmos.

Para llevar a cabo adecuadamente las aplicaciones antes mencionadas, cualquier sistema de vigilancia debe satisfacer varios requerimientos como son: confiabilidad, calidad de imagen, un número reducido de falsas alarmas, interoperabilidad con otros sistemas, etc. Existen varios sistemas propuestos previamente que presentan estas características, sin embargo su adquisición, en la mayoría de los casos, es costosa, así como los son sus costos de operación y mantenimiento. Adicionalmente, en algunos casos se requiere una gran capacidad de almacenamiento, ya que se debe efectuar la grabación durante las 24 horas del día, así como la intervención humana para revisar el video grabado. Por lo anterior el desarrollo de sistemas de vigilancia de bajo costo, los cuales no requieran la intervención humana, es un tópico de activa investigación. Como resultado, durante los últimos años varios sistemas que cumplen con estos requerimientos han sido propuestos en la literatura [1, 2]. Algunos de estos esquemas

requieren supervisión humana como es el caso del sistema propuesto por Jones [1] el cual es un esquema de video para la detección de movimiento de bajo costo, capaz de monitorear con muchas cámaras de video y con la ayuda de un operador, un espacio determinado las 24 horas del día desde un sitio remoto. Este sistema es capaz de grabar y enviar secuencias de video solamente cuando algún evento relevante está teniendo lugar, situación que es detectada usando la diferencia entre imágenes consecutivas de una secuencia de video. Con esta finalidad el sistema analiza las ventanas de vigilancia para determinar si la diferencia entre cuadros consecutivos es mayor que un umbral determinado de antemano. Para mejorar la detección, se puede usar el histograma de la distribución de luz con el fin de corregir las variaciones de iluminación dentro de la ventana. Zhang [2] propuso el sistema QLS basado en el análisis de consistencia lineal, el cual se enfoca solamente en aquellos elementos que son estrictamente necesarios para el cálculo de la solución. Debido a esto el QLS reduce los cálculos al mínimo, maximizando de esa manera su eficiencia [2]. Adicionalmente el algoritmo QLS no requiere calibración de la cámara y es numéricamente estable. Lorenzo Favalli y col [3] propusieron un sistema construido alrededor de los rasgos del esquema de codificación MPEG2 [4-7] y como consecuencia comparte también las limitaciones conocidas en el esquema de estimación de movimiento tales como el algoritmo de acoplamiento de bloques, movimientos lineales rígidos, oclusiones, fallas locales y aproximaciones de bloque. Este sistema se comporta muy bien cuando es usado para llevar a cabo procesos de inserción y recuperación de video, ya que no requiere el uso de técnicas de segmentación para la extracción y seguimiento de objetos. Dado que este sistema depende exclusivamente de la información que, sobre el movimiento, proporciona el codificador MPEG [4-7], éste produce una carga computacional adicional mínima al decodificador ya que emplea, para el 75% del trabajo de seguimiento de movimiento, la información ya presente en el flujo de bits. Hariharakrishnan y col. [8] proponen un

algoritmo de detección de movimiento en el cual el seguimiento se obtiene prediciendo las fronteras de los objetos de interés, usando los vectores de movimiento, seguida por una actualización de bordes por medio de un algoritmo de detección de oclusiones para resolver las mismas [8]. Para mejorar la estimación de movimiento entre tramas una aproximación basada en bloques adaptables es usada, junto con esquemas de modulación eficientes para controlar los huecos existentes entre las tramas empleadas para el seguimiento de un objeto. Las aplicaciones más comunes del algoritmo de seguimiento propuesto en [8] son la compresión de video usando el estándar MPEG4 [5, 6] y la recuperación de contenido basado en estándares tales como H.264L [7]. Yoneyama y col. [9] proponen un algoritmo para detección y clasificación simultánea de movimiento basado en mapas direccionales los cuales capturan la distribución espacio-temporal de las direcciones locales de movimiento a través del video; enfatizando la información tanto local como global del movimiento a través, tanto del tiempo como del espacio. Hernández y col [10, 11] separadamente proponen esquemas para la detección de movimiento basado en información obtenida del codificador MPEG. Finalmente en [12-16] se proponen otros eficientes algoritmos para la detección de movimiento. Todos estos sistemas presentan muchas características deseables, sin embargo aún existen algunas limitaciones que deben ser resueltas.

Este artículo propone un esquema para la detección de movimiento capaz de operar tanto en lugares abiertos como cerrados y bajo diversas condiciones ambientales; siendo además capaz de detectar los movimientos relevantes y eliminar movimientos irrelevantes en el área monitoreada. Con esta finalidad se estiman los vectores de movimiento [4-6] que permitan obtener información confiable acerca de los objetos que están generando movimiento dentro de la escena siendo monitoreada [4]. Para determinar si el movimiento es relevante o no, los vectores de movimiento son divididos en sus componentes vertical y horizontal, los cuales pueden ser positivos o negativos, con el fin de

obtener un vector de movimiento resultante que indica la dirección del movimiento del objeto bajo análisis; a partir del cual se puede determinar si el movimiento es relevante o no. Aquí un movimiento relevante es aquel que indica que una o varias personas están ingresando o aproximándose a una zona restringida. Para reducir las falsas alarmas la detección de movimiento se lleva a cabo introduciendo varias restricciones a la información del movimiento disponible en las tramas de video como es la distancia y posición de la cámara con respecto al objeto o persona siendo monitoreada.

Algoritmo propuesto

El sistema propuesto, mostrado en la figura 1, inicialmente recibe una secuencia de imágenes las cuales se dividen en bloques no traslapados de $N \times N$ a partir de los cuales se estiman los vectores de movimiento, MV. Una vez estimados los MV son filtrados para eliminar la distorsión debida al ruido, iluminación deficiente, etc. Seguidamente se estima la trayectoria del objeto o persona que esta generando movimiento a partir del vector resultante obtenido usando los MV de cada bloque, determinándose así si éste es un movimiento relevante o no lo es. Finalmente se obtiene el seguimiento de la trayectoria de la persona o personas. A continuación se describe con detalle cada uno de estos procesos.

Estimación de los vectores de movimiento

La estimación del vector de movimiento, MV, de un macro-bloque dado consiste en encontrar una región de 16×16 píxeles en el cuadro correspondiente al instante t , que aproxime lo más cercanamente posible al correspondiente macro-bloque en el cuadro de referencia, donde el cuadro de referencia corresponde a la imagen de la secuencia de entrada en el instante $t-1$. Tomando en cuenta que el intervalo de tiempo entre el cuadro t y el cuadro $t-1$ es relativamente pequeña, solamente es necesario realizar la búsqueda dentro de un área cercana al macro-bloque a analizar, cuyo tamaño es ligeramente

mayor. Aquí la distancia y dirección entre el macro-bloque bajo análisis y la región del cuadro de referencia que minimice un criterio dado de antemano, usualmente en MAE dado por (1), corresponde al vector de movimiento correspondiente a ese macro-bloque [4].

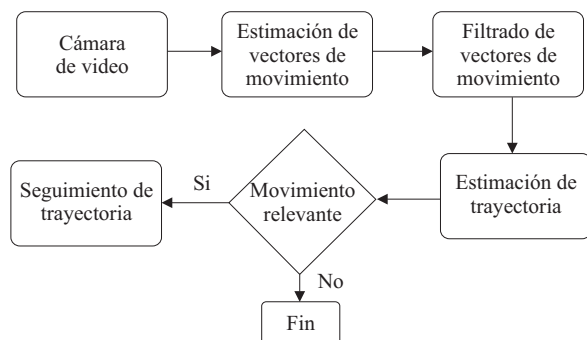


Figura 1 Sistema propuesto para la detección de movimiento relevante

$$MAE(i, j) = \frac{1}{MN} \sum_{k=0}^{M-1N-1} \sum_{l=0}^{N-1} |C(x+k, y+l) - R(x+i+k, y+j+l)| \quad (1)$$

donde, en la mayoría de los casos $M=N=16$, $C(x+k, y+l)$ es (k, l) -ésimo píxel y (x, y) son las coordenadas del extremo superior derecho del macro-bloque bajo análisis, respectivamente, en la trama actual y $R(x+i+k, y+j+l)$ es el macro-bloque en la posición $(x+i, y+j)$ en la trama de referencia con $-p \leq i, j \leq p$. Así, la región de búsqueda no es la trama de referencia completa, sino una región dentro de la misma que engloba la posición del macro-bloque bajo análisis en la trama de referencia, debido a que en general se puede suponer que un objeto no se mueve excesivamente rápido de una trama a otra en relación al número de tramas/s. Aquí la región de búsqueda se representa como $[-p, p]$ en ambos ejes, tomando (x, y) como centro de la región. Existen diversos mecanismos para la obtención de los MV, siendo el método de Estimación Jerárquica de Movimiento el que presenta la menor complejidad computacional menor [4]

Filtrado de los vectores de movimiento

Varios aspectos deben ser tomados en cuenta para realizar la estimación de un movimiento relevante a partir de los vectores de movimiento estimados anteriormente, siendo un movimiento relevante aquel realizado de afuera hacia adentro de la zona restringida por una persona o vehículo. Esto sugiere la necesidad de filtrar o eliminar los MV producidos ya sea por cambios abruptos en el brillo de la imagen o debidos a pequeñas vibraciones de la cámara; así como los MV producidos por el movimiento de objetos pequeños; lo mismo que MV que indiquen desplazamientos dentro de la zona de seguridad o de adentro hacia afuera de ésta. Así para eliminar los MV que no proporcionan información relativa a un movimiento relevante es importante considerar varios aspectos tales como la distancia entre la cámara y el escenario; ya que esto permite tener una mejor perspectiva del objeto, así como MV con una magnitud similar durante toda la trayectoria [11]. Esto además permite el usar un número constante de MV para representar el movimiento de una persona desplazándose hacia la zona restringida. Resultados obtenidos experimentalmente muestran que el movimiento de una persona se puede representar usando entre 9 y 13 MV como se muestra en la figura 2.

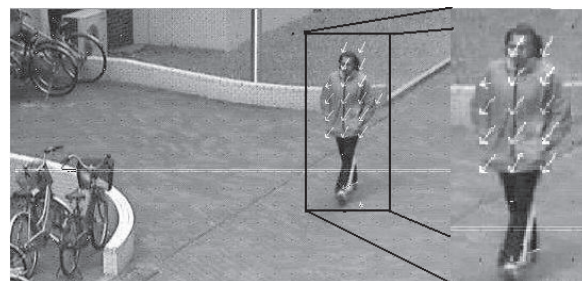


Figura 2 Posición de la cámara que permite detener un número determinado de MV para detectar el movimiento de una persona

Algunos de los MV estimados podrían presentar distorsión debido a cambios abruptos en el brillo de la imagen o pequeñas vibraciones de la cámara lo que se puede interpretar como la presencia de

ruido en los MV, lo que sugiere la necesidad de filtrar estos MV para reducir la distorsión. Así mismo, debido a que la cantidad de MV ruidosos es pequeña en comparación con los no ruidosos, así como al hecho de que los MV ruidosos no son continuos o similares entre sí como sucede con aquellos creados por movimientos reales, los MV ruidosos pueden ser fácilmente distinguidos del resto. Por otro lado, debido a que los MV son calculados uno por uno es necesario crear un registro temporal de almacenamiento, a partir del cual se llevará a cabo el proceso de filtrado de mediana mostrado en la figura 3 [16].

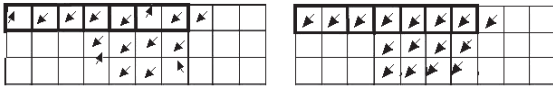


Figura 3 Filtrado de mediana de los vectores de movimiento para eliminar distorsión [16]

Una vez que los vectores de movimiento de cada macro-bloque son estimados y filtrados como se describió anteriormente, se procede a estimar la magnitud y dirección del movimiento de aquellas regiones que presenten más de 9 macro-bloques contiguos con vectores de movimiento diferentes de cero. De esta manera se eliminan todos aquellos MV que no proporcionan información útil para la estimación un movimiento real dentro de la escena de interés. Así dados N vectores de movimiento, MV , $N \geq 9$, pertenecientes a N macro-bloques continuos, se lleva a cabo la estimación de la magnitud y dirección del movimiento de las regiones de interés las cuales estarán dadas por:

$$|MV| = \left(\left(\sum_{k=1}^N MV_x(k) \right)^2 + \left(\sum_{k=1}^N MV_y(k) \right)^2 \right)^{1/2} \quad (2)$$

$$\theta = \tan^{-1} \left(\frac{\sum_{k=1}^N MV_y(k)}{\sum_{k=1}^N MV_x(k)} \right) \quad (3)$$

donde $MV_x(k)$ y $MV_y(k)$ son las componentes del vector de movimiento del k -ésimo macro-bloque y θ es el ángulo que indica la dirección del movimiento. Finalmente usando la información proporcionada por θ , se procede a eliminar el vector indicando el movimiento resultante, dado por (2) y (3), si su dirección es opuesta a la zona de acceso restringido como se muestra en la figure 4. Así el método propuesto es capaz de medir el nivel de relevancia de cada movimiento de acuerdo con su posición en la escena.



Figura 4 Detección de movimiento significativo en dirección a la zona restringida [16]

Estimación de la trayectoria

Una vez estimados los MV se calcula el ángulo del vector resultante, dado por (3), el cual permite determinar la dirección del movimiento con fin de clasificar el evento como un movimiento relevante o irrelevante. Para esto se toma en cuenta que un movimiento relevante será aquel en el cual una persona o vehículo ingresa o trata de ingresar a un lugar restringido, lo que significa que solo las trayectorias cuyos ángulos se encuentren entre los 220 y 340 grados serán consideradas como relevantes. Dicho rango fue calculado tomando en cuenta dos consideraciones importantes: la primera es relativa a la forma de caminar de un ser humano; mientras que la segunda toma en cuenta el hecho que no importa la trayectoria del desplazamiento de la persona si esta se realiza de adentro hacia afuera de la zona restringida, o si el movimiento se realiza dentro de ésta [16].

Otro factor importante que debe considerarse es la distancia entre la cámara y el escenario, ya que, además de obtenerse vectores constantes durante toda la trayectoria, como se mencionó anteriormente, una distancia correcta, como se muestra en la figura 2, permite introducir divisiones dentro de la escena para medir la importancia del movimiento de acuerdo a la posición de la persona en un tiempo determinado como se muestra en la como se aprecia en la figura 5, descartando los movimientos realizados dentro de la zona de seguridad. Estas divisiones además permiten usar varios contadores con el propósito de determinar no solo cuántas personas generan movimientos relevantes, si no también registrar de que dirección y cuántos accesos se producen a la zona restringida, lo cual permite rastrear y analizar los movimientos relevantes generados.



Figura 5 Zonas de seguridad y alerta establecidas con una adecuada posición de la cámara

Rastreo del movimiento generado

Otro factor importante es el rastreo del movimiento relevante generado que permita, además de determinar si se produjo éste, seguir la trayectoria de la persona que lo generó hasta que ésta abandone la zona restringida, con la posibilidad de analizar la evolución del movimiento de la misma para determinar el origen de este movimiento [14, 15]. Así es deseable que el sistema propuesto permita calcular, además del número de personas que están generando

movimientos relevantes en un momento dado, la dirección y el punto desde el cual se originó dicho movimiento dentro del escenario de interés [15]. Tomando en cuenta los requerimientos de los sistemas de seguridad, este tipo de información puede ser muy importante. El seguimiento de objetos o personas dentro de las secuencias de video podría presentar serias dificultades cuando los objetos se mueven relativamente más rápido que la tasa de cuadros/s, sin embargo existen muchas aplicaciones, como es el seguimiento de personas en sistemas de seguridad en las cuales los objetos se muevan a velocidades menores a 60 cuadros/s. En este tipo de aplicaciones se pueden emplear técnicas de compresión de video en las cuales los MV obtenidos de la secuencia de imágenes se pueden usar para estimar la magnitud y dirección del movimiento usando las ecuaciones (2) y (3). Con esta finalidad la imagen de los objetos cuyo movimiento se desea seguir se divide en macro-bloques, donde el movimiento de cada uno de ellos se representa por medio de un MV. Así usando la estimación de los MV cuya magnitud y orientación determina el desplazamiento de los nodos de la malla, mediante el uso de (2) y (3) se puede estimar la posición del objeto en cuadro correspondiente al instante t con relación al cuadro de la imagen en el instante $t-1$. Así repitiendo el proceso de manera sucesiva se puede estimar la trayectoria del objeto dentro la escena. Tal estimación permite determinar el nivel de importancia del movimiento que cada objeto genera, pudiendo así decidir cuando una trayectoria debe generar una alerta.

Experimental

El sistema propuesto se evaluó desde el punto de vista de su capacidad para detectar y rastrear movimientos relevantes dentro de una escena donde normalmente no debería de haber movimientos significativos en cierta dirección. El algoritmo fue evaluado en una PowerMac G4 con una velocidad de BUS de 167MHz y una velocidad del CPU de 1,25GHz con 1,25GB en RAM. Las secuencias de video utilizadas para probar el algoritmo fueron previamente grabadas con una

cámara de video digital con una resolución de 640x480 píxeles por cuadro, por lo que se conoce con anterioridad el resultado que deber presentar el algoritmo durante la detección. Las escenas muestran tomas al aire libre, debido a que en ellas podemos encontrar una gran cantidad de movimiento tanto relevante como no relevante, esto con la finalidad de probar la eficiencia del algoritmo en ambos aspectos.

La figura 6 muestra una secuencia de 4 imágenes en la cual se presenta un movimiento simple de afuera hacia adentro de la zona restringida, de acuerdo a la posición de la cámara [16]. Aquí la detección del movimiento es simple ya que no se encuentra ningún otro objeto generando movimiento significativo durante la duración de la secuencia. En esta secuencia aparece una persona generando movimiento de afuera hacia adentro de la zona restringida, estando presentes otras personas que generan movimientos desde adentro hacia fuera de la zona restringida, es decir movimientos no significativos. Así mismo la figura 6(b) muestra el resultado desplegado por el programa desarrollado para llevar a cabo la realización del algoritmo propuesto. Así el método propuesto es capaz de contar el número de movimientos que aparecen dentro de la escena, analizando la trayectoria de cada uno para determinar el número de ellos que ingresan al área de seguridad. Seguidamente en la figura 7(a) se observa una secuencia de 6 imágenes las cuales registran 3 accesos ilegales a una zona restringida. Por su parte la figura 7(b) muestra el resultado desplegado por el algoritmo desarrollado para llevar a cabo la realización del algoritmo propuesto [16]. Así los resultados experimentales obtenidos muestran que el sistema propuesto es capaz de realizar inicialmente la detección de dos movimientos relevantes en forma simultánea, los cuales representan accesos ilegales a la zona restringida. Posteriormente el sistema detecta un tercer acceso a la zona restringida.

Si no tomamos en cuenta la distancia mínima que deberá de haber entre el objeto y la cámara durante el proceso de detección, el resultado final será una lectura errónea de la secuencia de video, ya que

se debe recordar que la herramienta principal para poder determinar que el movimiento detectado es generado por un ser humano es el número de MV. Así, si la distancia no es la correcta, esto es si la cámara se encuentra muy cerca del objeto, el número de MV cambiaría. Esto origina que el resultado final, como se muestra en las figuras 8 y 9, sea erróneo. Aquí al inicio la detección es correcta, sin embargo conforme la distancia se reduce el número de accesos ilegales detectados llega a ser erróneo, ya que una persona se caracteriza con un determinado número de vectores de movimiento cuando la distancia entre la cámara y la persona es adecuada, en otro caso, al incrementarse el tamaño relativo de la persona, el número de vectores de movimiento llega a ser mayor dando como resultado que el sistema detecte hasta cuatro accesos ilegales en lugar de solamente uno.

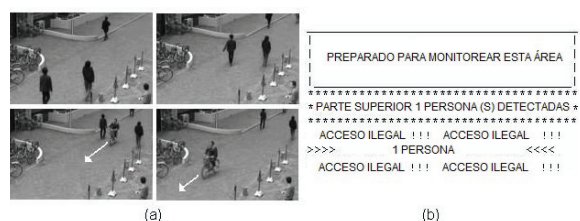


Figura 6 (a) Secuencia de imágenes mostrando un acceso simple a la zona restringida [16]. (b) Resultado desplegado por el programa que realiza el algoritmo propuesto

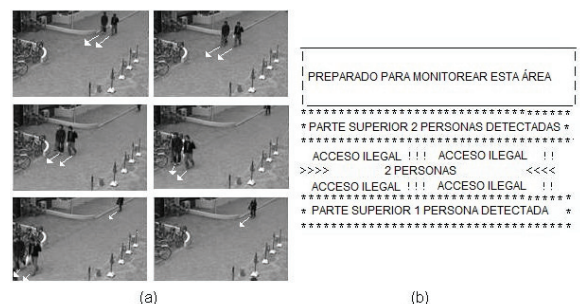


Figura 7 (a) Secuencia de imágenes mostrando 3 accesos a la zona restringida. Dos de ellos de manera simultánea [16]. (b) Resultado desplegado por el programa usado para la realización del algoritmo propuesto

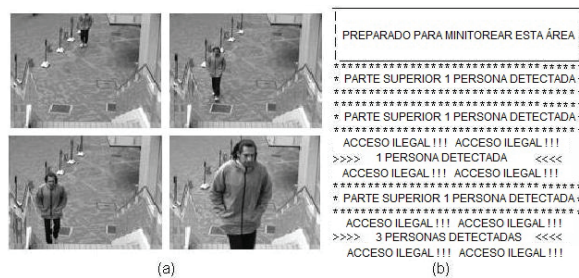


Figura 8 (a) Secuencia de imágenes obtenida cuando la distancia entre la cámara y la persona no es la correcta. (b) Resultado del programa desarrollado

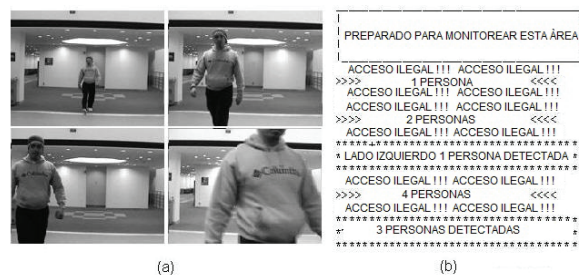


Figura 9 (a) Secuencia de imágenes obtenida cuando la distancia entre la cámara y la persona generando movimiento es menor que la recomendada. (b) Resultado producido por el programa desarrollado

Seguidamente se evalúa la capacidad de seguimiento del sistema propuesto. La figura 10 muestra el funcionamiento del algoritmo propuesto cuando es requerido a seguir la trayectoria de una persona, donde para obtener una trayectoria más suave o menos irregular, la trayectoria estimada mediante los vectores de movimiento es filtrada con un filtro pasa-bajas [16]. Las versiones la trayectoria estimada y su versión filtrada se muestran en las figuras 10(a) y 10(b).

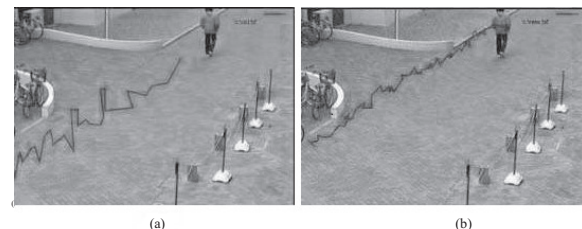


Figura 10 a) Seguimiento de la trayectoria a partir de los vectores de movimiento. (b) Trayectoria suavizada usando filtrado pasa bajas

Finalmente la tabla 1 presenta una comparación del algoritmo propuesto con otros previamente reportados en la literatura. Los resultados muestran que el algoritmo propuesto es competitivo con otros algoritmos similares.

Tabla 1 Comparación del comportamiento del algoritmo propuesto con otros dos algoritmos previamente reportados en la literatura

	<i>Detección de 2 o más movimientos</i>	<i>Confiabilidad</i>	<i>Lugares abiertos y cerrados</i>	<i>Depende de la distancia al objeto</i>	<i>Análisis de trayectoria</i>	<i>Costo</i>	<i>Detección de movimientos complejos</i>	<i>Robustez</i>	<i>Tiempo de detección</i>
Propuesto	Si	Alta	Amplio	Si	Si	Bajo	Amplia	Alto	2 a 4 cuadros/s
Ref. [10]	No	Media	Limitado	Si	No	Medio	Limitada	Medio	10 a 15 cuadros/s
Ref. [9]	Si	Alta	Amplio	Si	Limitada	Alto	Media	Alto	10 a 20 cuadros/s

Conclusiones

Este artículo propone un sistema automático para la detección de movimiento basado en la estimación de los vectores de movimiento, obtenidos a partir de una secuencia de video. Inicialmente los MV se filtran para reducir ruido presente para mejorar la estimación del movimiento. Adicionalmente, la posición de la cámara, es fundamental ya que si la distancia que separa a la cámara de la frontera entre la zona de seguridad y una zona de alerta es adecuada, el sistema es capaz de determinar si el movimiento es relevante, esto es de afuera hacia adentro de la zona de seguridad; o si éste es no relevante, es decir de dentro hacia fuera de la zona restringida. Finalmente, dividiendo la imagen en bloques no traslapados en forma de malla y estimando el movimiento de los vértices de cada bloque, es posible llevar a cabo el seguimiento de la trayectoria del objeto o persona de interés.

Los resultados obtenidos en la evaluación muestran que el sistema propuesto es capaz de detectar y clasificar el movimiento de cualquier objeto de acuerdo a las características y dirección que presenta el mismo, siendo capaz de trabajar tanto en interiores como es exteriores. El sistema de detección y seguimiento de movimiento propuesto está intentado para usarse como parte de un sistema de vigilancia, donde monitorear y detectar entradas relevantes es el objetivo principal. La técnica propuesta muestra ser confiable con respecto a cambios abruptos en el brillo, medio ambiente y parámetros de movimientos discontinuos, así como la dirección de los objetos. La principal ventaja de esta técnica es que usa solamente la información proporcionada por los vectores de movimiento, para detectar y llevar a cabo el seguimiento de los movimientos relevantes dentro de la escena de interés, teniendo además la habilidad de monitorear mientras lleva a cabo la detección de uno o más objetos dentro de la zona de alerta. El sistema propuesto es competitivo con otros previamente reportados en la literatura.

Agradecimientos

Un especial agradecimiento al Consejo Nacional de Ciencia y Tecnología (CONACYT) de México, Al Instituto de Ciencia y Tecnología del Distrito Federal (ICYTDF) de México, a la Universidad de Electro-Comunicaciones de Japón, así como al Instituto Politécnico Nacional por el apoyo proporcionado durante el desarrollo de esta investigación.

Referencias

1. B. T. Jones. "Low-Cost Outdoor Video Motion and Non-Motion Detection". *Proceedings of Int. Carnahan Conference on Security Technology*. San Jose (CA). 1995. pp. 376-380.
2. Z. Zhang. "Mining Surveillance Video for Independent Motion Detection". *IEEE International Conference on Data Mining*. Houston (TX). 2005. pp. 741-744.
3. L. Favalli, A. Mecocci, F. Moschetti. "Object Tracking For Retrieval in MPEG2". *IEEE. Trans. on Circuit and Syst. for Video Technology*. Vol. 10. 2000. pp. 427-432.
4. P. Symes. "Digital video compression". Ed. McGraw Hill. New York. 2001. pp. 137-149.
5. I. E. Richardson. *H.264 and MPEG-4 Video Compression, Video Coding for the Next-Generation Multimedia*. Ed. Wiley. Londres. 2004. pp. 27-41.
6. J. Watkinson. *The MPEG Handbook: MPEG-1, MPEG-2, MPEG-4*. Ed. Focal. Burlington (MA). 2001. pp. 257-254.
7. P. N. Tudor. "MPEG-2 Video Compression: Tutorial". *Journal of Electronics and Communication Engineering*. Vol. 7. 1995. pp. 257-264.
8. K. Hariharakrishnan, D. Schonfeld, P. Raffy, F. Yassa. "Video Tracking Using Block Matching". *IEEE. International Conference on Image Processing*. Barcelona. 2003. pp. 945-948.
9. A. Yoneyama, Y. Nakajima, H. Yanagihara, M. Sugano. "Moving Object Detection from MPEG Video Stream". *Systems and Computers in Japan*. Vol. 30. 1999. pp. 1-11.
10. J. Hernández García, H. Pérez Meana, M. Nakano Miyatake. "Video Motion Detection using the Algorithm of Discrimination and Hamming Distance". *Lectures Notes in Computer Science*. Vol. 3563. 2005. pp. 321-330.

11. J. M. Gryn, R. Wlides, J. K Tsotsos. "Detecting Motion Patterns via Directional Maps with Applications to Surveillance". *Computer Vision and Image Understanding*. Vol. 113. 2009. pp. 291-307.
12. S. Avidan. "Support Vector Tracking". *IEEE Trans. on Image Processing*. Vol. 26. 2004. pp. 1064-1071.
13. H. Nguyen, A. Smeulders. "Fast Occluded Object Tracking by a Robust Appearance Filter". *IEEE Trans. on Image Processing*. Vol. 26. 2004. pp. 1099-1103.
14. M. Lin, C. Tomasi. "Surfaces with Occlusions from Layered Stereo". *IEEE*. Vol. 26. 2004. pp. 1073-1098.
15. T. Sebastian, P. Klein, B. Kimia. "Recognition of Shapes by Editing Their Sock Graphs". *IEEE. Trans. on Pattern analysis and machine intelligence*. Vol. 26. 2004. pp. 550-571.
16. J. Hernández, H. Morita, M. Nakano-Miyatake, H. Pérez-Meana. "Movement Detection and Tracking Using Video Frames". *Lecture Notes in Computer Science*. Vol. 5856. 2009. 1054-1061.

Eduardo Bayro-Corrochano
Jan-Olof Eklundh (Eds.)

LNCS 5856

Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications

14th Iberoamerican Conference on Pattern Recognition, CIARP 2009
Guadalajara, Jalisco, Mexico, November 2009
Proceedings

 Springer

Volume Editors

Eduardo Bayro-Corrochano
CINVESTAV, Unidad Guadalajara
Department of Electrical Engineering and Computer Science
Jalisco, México
E-mail: edb@gdl.cinvestav.mx

Jan-Olof Eklundh
KTH - Royal Institute of Technology
Centre for Autonomous Systems
School of Computer Science and Communication
Stockholm, Sweden
E-mail: joe@nada.kth.se

Library of Congress Control Number: 2009937938

CR Subject Classification (1998): I.5, I.4, I.2.10, I.2.7, F.2.2

LNCS Sublibrary: SL 6 – Image Processing, Computer Vision, Pattern Recognition,
and Graphics

ISSN 0302-9743
ISBN-10 3-642-10267-0 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-10267-7 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

springer.com

© Springer-Verlag Berlin Heidelberg 2009
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12787630 06/3180 5 4 3 2 1 0

XVIII CASI 2009 Workshop I: Intelligent Computing for Remote Sensing Imagery

Spectral Estimation of Digital Signals by the Orthogonal Kravchenko Wavelets $\{\widehat{h_a(t)}\}$	989
<i>Victor Kravchenko, Hector Perez Meana, Volodymyr Ponomaryov, and Dmitry Churikov</i>	
Video Denoising by Fuzzy Directional Filter Using the DSP EVM DM642	997
<i>Francisco J. Gallegos-Funes, Victor Kravchenko, Volodymyr Ponomaryov, and Alberto Rosales-Silva</i>	
Image Authentication Scheme Based on Self-embedding Watermarking	1005
<i>Clara Cruz-Ramos, Rogelio Reyes-Reyes, Mariko Nakano-Miyatake, and Héctor Pérez-Meana</i>	
Unified Experiment Design, Bayesian Minimum Risk and Convex Projection Regularization Method for Enhanced Remote Sensing Imaging	1013
<i>Yuriy Shkvarko, Jose Tuxpan, and Stewart Santos</i>	
Intelligent Experiment Design-Based Virtual Remote Sensing Laboratory	1021
<i>Yuriy Shkvarko, Stewart Santos, and Jose Tuxpan</i>	

XIX CASI 2009 Workshop II: Intelligent Fusion and Classification Techniques

Optimizing Classification Accuracy of Remotely Sensed Imagery with DT-CWT Fused Images	1031
<i>Diego Renza, Estibaliz Martinez, and Agueda Arquero</i>	
Filter Banks for Hyperspectral Pixel Classification of Satellite Images...	1039
<i>Olga Rajadell, Pedro García-Sevilla, and Filiberto Pla</i>	
Minimum Variance Gain Nonuniformity Estimation in Infrared Focal Plane Array Sensors	1047
<i>César San-Martin and Gabriel Hermosilla</i>	
Movement Detection and Tracking Using Video Frames	1054
<i>Josue Hernandez, Hiroshi Morita, Mariko Nakano-Miyatake, and Hector Perez-Meana</i>	

Movement Detection and Tracking Using Video Frames

Josue Hernandez¹, Hiroshi Morita², Mariko Nakano-Miyake¹,
and Hector Perez-Meana¹

¹National Polytechnic Institute, Av. Santa 1000, 04430 Mexico D. F. Mexico

²The University of Electro-Communications, 182-8585, Tokyo, Japan
mariko@infinitum.com.mx, hmperezm@ipn.mx

Abstract. The use of image processing schemes as part of the security systems have been increasing, to detect, classify as well as to track object and human motion with a high precision. To this end several approaches have been proposed during the last decades using image processing techniques, because computer vision let us to manipulate digital image sequences to extract useful information contained in a video stream. In this paper we present a motion detection algorithm using the movement vectors estimation which are subsequently filtered to obtain better information about real motion into a given scenes. Experimental results show that the accuracy of proposed system.

Keywords: Motion Vectors, Movement Detection, Surveillance System, Surveillance system development.

1 Introduction

The advance of electronic and computer technologies increase the ability to perform video analysis to extract useful information from a video frames to carry out [1] motion detection and characterization [2], remote sensing, and pattern recognition, among others [3]-[5]. Pattern recognition is a research area that has been amply studied during the last years, it use information contained in video sequences; In many applications of pattern recognition, approaches with recognition capability are usually based on a corpus of data which is treated either in a holistic manner or which is partitioned by application of prior knowledge [5] like shape, velocity, direction, texture, magnitude, behavior and so on from different kind of objects around the area of interest.

Several approaches have been proposed to solve the problem of detecting and tracking motion during the last several years [6]-[16]. Some of them are describe in the following paragraphs [6]. Reference [7] proposes a motion detection approach based on the MPEG image compression algorithm in which the estimation of detection motion and the moving object direction only use the information contained in the MPEG motion vectors and the DC coefficients of the DCT directly extracted from the MPEG bit stream of the processed video. Evaluation results of this method shows that it can handle several situations where moving objects are present in the scene observed with a mobile camera. However, the

efficiency of the moving object detection depends on the quality of MPEG motion vectors [8-12]. This method also has a constraint in the context of content-based video browsing and indexing that should be taken into account when the MPEG encoder is selected. Reference [9] proposed an algorithm that intends to extend the capabilities to MPEG2 streams, allowing tracking objects selected by a user throughout the sequence. The way in which the tracking has been realized, is through exploitation of the motion information already present in the bit stream and generated by the decoder, another important feature of this scheme is that it is not necessary to decode the full sequence, if the user wants to access only part of the video stream [10]. This proposed scheme performs well in assisting the information insertion/retrieval process. However, no segmentation or filtering techniques are used for the extraction and tracking of the objects, since it relies exclusively on the motion information already provided by the MPEG encoder. Evaluation results show that the algorithm performs well though it is slightly dependent on the object shape. Hariharakrishnan et al [8] proposes an algorithm in which the tracking is achieved by predicting the object boundary using motion vectors, followed by contour update, using occlusion/disocclusion detection. An adaptive block-based approach has been used for estimating motion between consecutive frames. Here an efficient modulation scheme is used to control the gap between frames used for object tracking. The algorithm for detecting occlusion proceeds in two steps. First, covered regions are estimated from the displaced frame difference. Next these covered regions are classified into actual occlusions and false alarms using the motion characteristics. Disocclusion detection is also performed in a similar manner [8].

This paper proposes an algorithm for detection and tracking of movement of objects and persons based on a video sequence processing. Evaluation results show that proposed scheme provides a fairly good performance when required to detect relevant movements and tracking the motion of objects under analysis.

2 Proposed Movement and Tracking Detection System

The proposed system, firstly estimates the motion vectors, using an input video frames, which are then filtered to reduce distortion due to noise and deficient illumination. Next using the estimated motion vectors the movement trace is estimated to determine if it is a relevant or irrelevant motion. Finally if the movement is relevant, its tracking is carried out until the object left the restricted zone.

2.1 Motion Vector Estimation

The motion estimation is carried out dividing, firstly, the actual image at time, $t-1$, into a non-overlapping macro-blocks of 16×16 pixels. Next in the image frame at time t , the algorithm looks for the region that closely matches the macro-block under analysis [2]. Taking in account that the time difference in time between two consecutive images in a given frame is relatively small, only is necessary to carry out the

analysis in a region slightly larger than the given macro-block. Here the distance and direction that minimizes the criterion given by [5]

$$MAE(i, j) = \frac{1}{MN} \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} |C(x+k, y+l) - R(x+i+k, y+j+l)| \quad (1)$$

Where $C(x+k, y+l)$ is the (k, l) th pixel of macro-block (x, y) of the actual frame and $R(x+i+k, y+j+l)$ is the macro-block in the position $(x+i, y+j)$ in the reference frame, where $-p \leq i \leq p$ $y-p \leq j \leq p$. Then the vector motion magnitude and direction is obtained from the actual position (x, y) in the actual frame, to the position $(x+i, y+j)$ in which the MAE (i, j) is minimum. This process is illustrated in Fig. 1.

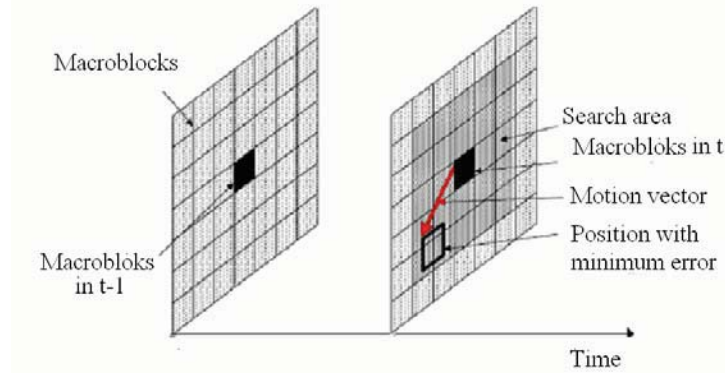


Fig. 1. Motion vector estimation process

Several methods have been proposed for motion vector estimation, which are based on the MAE (Mean Absolute error) minimization (1), all of them providing similar results although their computational complexity presents important differences. Among the Hierarchical estimation method presents the less computational complexity and then is one of the most widely used motion vector estimation method. Here to reduce the computational complexity, the image is decimated by 2 using low pass filters, reducing in such way the image size. Thus the position (x, y) in the original image becomes the position $(x/2, y/2)$ in the image corresponding to the first decomposition level. Subsequently a second decomposition level is applied using a low pass filter, such that the original (x, y) point becomes the point $(x/4, y/4)$ in the second decomposition level image [5].

After the second decomposition level is performed, the motion vector estimation starts in the second decomposition level, in which the image size is $1/16$ of the original image size, with a macro-block of size 4×4 . Assuming that the point (u_2, v_2) corresponds to the minimum MAE of the actual macro-block, the search of the motion vector in the decomposition level 1 is carried out in the macro-blocks of 8×8 whose search start in $(x/2 + 2u_2, y/2 + 2v_2)$, with a search range of $[-1, 1]$ around the origin pixel. Finally the motion vector estimation in the level decomposition 0 is carried out with macro-blocks of size 16×16 , with search starting at point $(x + 2u_1, y + 2v_1)$ with a search range equal to $[-1, 1]$, around the origin pixel. Here

(u_1, v_1) corresponds to the minimum MAE at the level decomposition 1. These motion vectors consist of horizontal and vertical components which are used to estimate the motion vector magnitude and angle. Finally using these data it is possible to determine if the movement is relevant or not. Figure 2 illustrates this process.

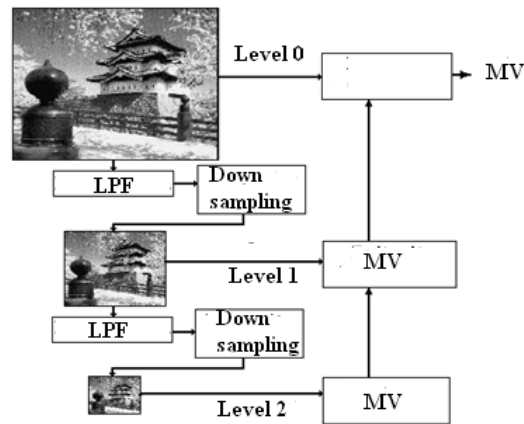


Fig. 2. Hierarchical motion vector estimation algorithm

Several times the estimated motion vectors only represent abrupt illumination changes in the image or small vibrations of the camera used to capture the image sequence, resulting in wrong movement estimation if these distortions are not cancel or at least reduced. To this end, the estimated motion vectors are filtered using a one dimensional median filter as shown in Fig. 3, before the overall movement estimation. This allows to cancel most motion vectors that no provides useful information, such as those due to illumination changes, background movement, etc. This process is shown in Fig. 3.

Because the amount of noisy motion vectors is relatively small in comparison with the correctly estimated motion vectors, as well due to the fact that the noisy vectors are continuous ore similar among them, as happen with the actual motion vectors, the noisy motion vectors can be easily distinguish form the remaining ones. On the other hand, because the motion vectors are estimated one by one it is not necessary to create a temporal register, to be use during the filtering process [5]. Thus after the noisy vectors have been eliminated, the information generated by objects moving in no relevant directions is cancelled, based on the fact that movement direction is different to that of the security zone. This fact allows the proposed algorithm to measure the relevance level, according to their position in the scene. Thus a motion vector is eliminated if its position does not change after several video sequences or if the motion direction is opposite the restricted zone, as shown in Fig. 4.

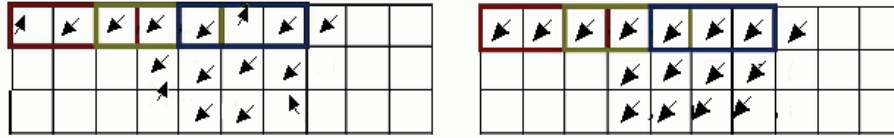


Fig. 3. Motion vectors filtering process

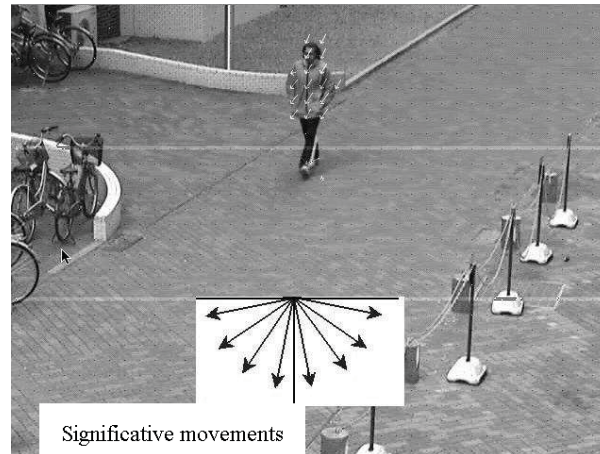


Fig. 4. Relevant movement detection in the direction of the restricted zone

2.2 Trajectory Estimation

Once the motion vectors have been estimated, it is necessary to estimate other important parameters such as: the speed, direction and movement tracking of the moving object present in the video sequence. To this end firstly it can be used the information provided by the motion vector to estimate the movement angle to classify the movement in, either, relevant or not relevant. To this end, we can take in account that a relevant movement is that in which the people enter or intend to enter into the restricted zone. This fact implies that, according with the video camera position, the motion vectors angles must be between 200 and 340 degrees. This fact takes in account that a movement is not relevant if it takes place inside the restricted zone, or if its direction is from inside to outside the restricted zone.

Other important factor that must be considered to obtain an accurate estimation of the motion vectors is the camera position, because it allows to obtain constant motion vectors, with similar magnitude during all trajectory, avoiding detection errors. This fact also allows that a person's magnitude may be represented using among 9 to 13 motion vectors. A correct distance also allows adding some divisions to measure the movement importance and be able to track the position of a given object during a determined time; as well as to add several counters to determine the number of



Fig. 5. Tracking ability of proposed algorithm

persons that are generating relevant movements, as well as the direction of these movements.

2.3 Movement Tracking

Other important issue is the tracking of the generated relevant movement, which means that not only is important to determine if the movement a relevant movement has been produced, but also to have the ability to track the movement of the person generating it until the go out of the restricted zone. The main difficulty to track the object movement using a sequence of successive images is the object localization in the image sequences, especially when the movement of them is relatively faster than the images rate. For these reason, the tracking systems usually uses moving models that intend to predict the variations of the image due to possible movements present in the object trajectory. Although these methods provide fairly good results, they may be computationally complex. Thus to reduce the computational complexity, the object is divided in $N \times M$ non overlap blocks. Next using the motion vector estimation described above, the movement of each corner is estimated. Once the motion vectors are obtained, the tracking of object motion in each frame is obtained as the resultant vector of all individual motion vectors in such frame. Finally to obtain a smoother trajectory the resulting vector in the previous step is filtered using a low pass filter.

3 Evaluation Results

The proposed system was evaluated using computer simulation using a Power Mac G4, with a data bus speed of 167MHz, a CPU of 1.25GHz and 1.25GB of RAM memory. The video sequences used to evaluate the proposed algorithm were previously recorded with a resolution of 640x480 pixels per image. The obtained results show that the proposed algorithm is able to accurately detect the relevant movements and correctly track the person or object motion.



Fig. 6. Detection performance of proposed algorithm, here in the fourth and fifth image there are a relevant movement

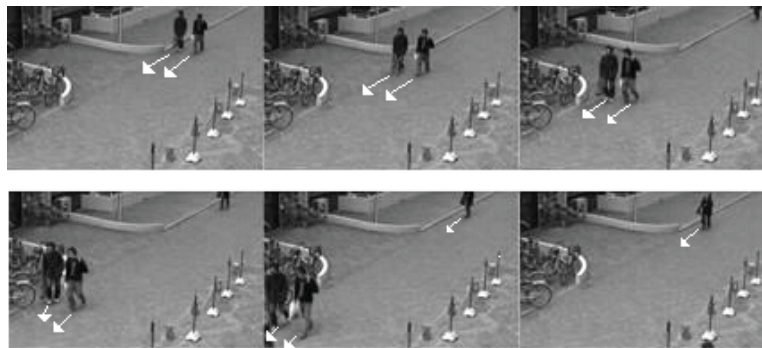


Fig. 7. Two persons moving to the restricted zone

4 Conclusions

This paper proposed an automatic movement detection system, using a video sequence, based on the motion vector estimation, which are filtered to eliminate the noisy and distorted vectors, due to illumination variations and background movement. Proposed algorithm is also to discriminate between relevant and non relevant movements which allows only take in account the movements whose direction is from outside to inside the restricted zone. Using also a motion vector estimation the algorithm is also able to tract the trajectory of a given person whose movement is from inside to outside of the restricted zone.

Acknowledgements

The authors thanks the National Science and Technology Council, to The Science and Technology Institute of the D. F. Mexico, and to the National Polytechnic Institute of Mexico for the financial support provide during the realization of this research.

References

1. Jones, B.T.: Low-Cost Outdoor Video Motion and Non-Motion Detection. *Processing of Security Technology*, 376–380 (1995)
2. Paladino, V.: Introduction of Video Compression Under the Standard MPEG-2. *The Institute of Electronic Engineer, Spain*, 3–24 (2005)
3. Richardson, I.E.: H.264 and MPEG-4 Video Compression, Video Coding for the Next-Generation Multimedia, pp. 27–41. Wiley, UK (2004)
4. Watkinson, J.: *The MPEG Handbook: MPEG-1, MPEG-2, MPEG-4*. Focal Press (2001)
5. Tudor, P.N.: MPEG-2 Video Compression: Tutorial. *Journal of Electronics and Communication Engineering*, 1-5 (December 1995)
6. Zhang, Z.: Mining Surveillance Video for Independent Motion Detection. In: *IEEE International Conference on Data Mining*, pp. 741–744 (2005)
7. Favalli, L., Mecocci, A., Moschetti, F.: Object Tracking For Retrieval in MPEG2. *IEEE, Trans. on Circuit and Syst. for Video Technology*, 427–432 (2000)
8. Hariharakrishnan, K., Schonfeld, D., Raffy, P., Yassa, F.: Video Tracking Using Block Matching. In: *IEEE, International Conference on Image Processing*, pp. 945–948 (2003)
9. Yoneyama, A., Nakajima, Y., Yanagihara, H., Sugano, M.: Moving Object Detection from MPEG Video Stream. *Systems and Computers in Japan* 30(13), 1–11 (1999)
10. Avidan, S.: Support Vector Tracking. *IEEE* 26(8), 1064–1071 (2004)
11. Nguyen, H., Smeulders, A.: Fast Occluded Object Tracking by a Robust Appearance Filter. *IEEE Trans. on Image Processing* 26(8), 1099–1103 (2004)
12. Lin, M., Tomasi, C.: Surfaces with Occlusions from Layered Stereo. *IEEE* 26(8), 1073–1098 (2004)
13. Sebastian, T., Klein, P., Kimia, B.: Recognition of Shapes by Editing Their Sock Graphs. *IEEE Trans. on Image Processing* 26(5), 550–554 (2004)
14. Coding Audiovisual Objects Part 2, International Standard Organization / Int. Electronics Communications (ISO/IEC), 14496
15. Hernández García, J., Pérez-Meana, H., Nakano Miyatake, M.: Video Motion Detection using the Algorithm of Discrimination and Hamming Distance. In: Ramos, F.F., Larios Rosillo, V., Unger, H. (eds.) *ISSADS 2005. LNCS*, vol. 3563, pp. 321–330. Springer, Heidelberg (2005)
16. Gryn, J.M., Wlides, R., Tsotsos, J.K.: Detecting Motion Patterns via Directional Maps with Applications to Surveillance. *Computer Vision and Image Understanding* 113, 291–307 (2009)

**GESTS International Transactions on
Computer Science and Engineering**

Volume 43, Number 1, November 2007



Global Engineering, Science, and Technology Society
<http://www.gests.org>

GESTS International Transactions

Editorial Board

Pedro Arezes, Portugal	Gu-min Jeong, Korea	Yasuaki Okada, Japan
Karunesh Arora, India	Marcel Jirina, Czech	Raymond Panneton, Canada
Bruce M. Bae, Korea	Misurec Jiri, Czech	Sung-Kyo Park, Korea
Isabel Lopez Barrio, Spain	Ho-Youl Jung, Korea	Sungyong Park, Korea
N. Tamer Bayazit, Turkey	Sun Mee Kang, Korea	Dariusz Pleban, Poland
Michael Beddoes, Canada	Faouzi Kamoun, ARAB Emi.	Laurent Poinot, France
Aicha Bouzid, Tunisia	P. Kazienko, Poland	Ilyas Potamitis, Greece
Guoyue Chen, Japan	Jin Young Kim, Korea	Hae Chang Rim, Korea
Junga Choi, USA	Jong Kuk Kim, Korea	Tracey L. Rogers, Australia
Seong Young Choi, Korea	Seungcheon Kim, Korea	Antonio Samagaio, Portugal
Seung Ho Choi, Korea	Sung-A Kim, Korea	Koji Sasaki, Japan
Kyung Ryul Chung, Korea	Heuy-Dong Kim, Korea	Tetsumi Sato, Japan
Yung Chang Chung, Korea	Pyung-Soo Kim, Korea	JongSoo Seo, Korea
Michael Cohen, Japan	Ioannis Krikidis, France	Qinzhong Shi, Japan
Tony Collings, Australia	Mira Oh, Korea	Tae-Bo Shim, Korea
Franco Cotana, Italy	Stjepan Lakusic, Croatia	Seonghan Shin, Japan
D. Cvetkovic, Yugoslavia	Dong Chun Lee, Korea	Gequn Shu, China
Pawel Czarnul, Poland	InSeop Lee, USA	Jung-Ho Sohn, Korea
Yuki Denda, Japan	Kyung Suk Lee, Korea	Mathias Stani, Austria
Jichang Dong, China	Joo-Hun Lee, Korea	Chuan-Ching Sue, Taiwan
Vesna Dragcevic, Croatia	Jung Hyun Lee, USA	Soo-Young Suk, Japan
Nikos Fakotakis, Greece	Liang-Teh Lee, Taiwan	Chakib Tadj, Canada
Yinzhang GUO, China	Won Chul Lee, Korea	S. k. Tang, Hong Kong
Daeho Ha, Thailand	Jemni B. A. Leila, Tunisia	Carlos Ventura, Canada
Sami Harari, France	Valeri Lenchine, Korea	Wei-Hui Wang, Taiwan
Kiyoshi Hayakawa, Japan	Zishun Liu, Singapore	Seon-Mi Woo, Korea
HIdechito Hayashi, Japan	D. Shing-Min Liu, Taiwan	Ja-Ling Wu, Taiwan
Hiromitsu Hhta, Japan	Feng Liu, USA	Don-lin Yang, Taiwan
Chan-hsiao Ho, Taiwan	Giuseppe Miccoli, Italy	Sung-il Yang, Korea
Kwang Seok Hong, Korea	Soyeon Min, Korea	Koichi Yoshihisa, Japan
Naohiro Ishii, Japan	Yasuo Mitani, Japan	Fernando Zacarias, Mexico
A. Jagniatinskis, Lithuania	Noboru Nakasako, Japan	Jianyuan Zhu, China
J. S. Roger Jang, Taiwan	Takanobu Nishiura, Japan	
Wan-Ho Jeon, Korea	Hiromitsu Ohta, Japan	

Table of Contents

Performance of Learning based Navigation in Obstacle Avoidance Problem	1
SeungGwan Lee	
Circuit Transformation Method from BOTA Circuits into UCC-based Circuits	9
Norbert Herencsar and Kamil Vrba	
Development of a Museum Exhibition System Using Podcast Based on PUR	17
Kumiko Iwazaki, Hidehiko Agata, and Takami Yasuda	
Influence of the Amplifier Non-Idealities on the Current-Mode MFB Low-Pass Filter	25
David Kubanek and Kamil Vrba	
Particle Swarm Optimisation for Segmenttton of Microcalcifications in Mam- mograms	33
K.Geetha and K.Thanushkodi	
Development of a 3DBlog System and its Applications	43
Yoshio Nishio, Shigeki Yokoi, and Takami Yasuda	
Motion Detection Algorithm Based on MPEG Video Coding	55
Josue Hernandez-Garcia, Hiroyoshi Morita, Wu Katsu, Baskara Nugraha, Mariko Nakano-Miyatake, and Hector Pérez-Meana	
Real-time Disk Scheduling for Non-hiccupped Continuous Media Playbacks	67
Sungchae Lim	
Retrieval-oriented Summarization	79
Weijiang Li, Tiejun Zhao, and Wenmao Zang	
Quantum-inspired Huffman Coding	91
A. S. Tolba, M. Z. Rashad, and M. A. El-Dosuky	
Simulation and Analysis of OFDM Modulation	105
Anna Shklyeva and David Kubanek	

Motion Detection Algorithm Based on MPEG Video Coding

Josue Hernandez-Garcia¹, Hiroyoshi Morita², Wu Katsu², Baskara Nugraha²
Mariko Nakano-Miyatake¹, and Hector Pérez-Meana¹

¹ National Polytechnic Institute, Mexico, Av. Santa Ana 1000, San Francisco
Culhuacan, 04430 Mexico D.F. Mexico
hmpm@prodigy.net.mx

² The University of Electro-Communications, Chofu-shi,
Tokyo 182-8585, Japan
hiroyoshi-m@par.odn.ne.jp

Abstract. Security systems based on automatic motion detection must satisfy several requirements such as: accurate objects tracking and elimination of irrelevant motion information in the security area to avoid false alarms; sensitivity reduction color, shades and illumination changes, etc. To meet these requirements, several reliable security systems have been proposed during the last several years, based on pixel as well in coded data domains, although several problems still must be solved. This paper proposes a motion detection algorithm that takes advantage of the motion vectors information generated by the MPEG-2 encoder to detect relevant motion in a video stream, together with a tuning step that is used to eliminate irrelevant information. Experimental results show that the proposed algorithm performs fairly well under different conditions eliminating irrelevant information and proving fairly good movement detection under different cases.

1 Introduction

Automatic motion detection and object tracking has been a topic of active research during the last several years; because their potential application in the developing of reliable security systems using the information contained in video sequences, together with previously acquired information such as: shape, speed, moving direction and texture etc. of the objects contained in the video sequence. They have found many applications in outdoors monitoring, doorways, buildings, car parks, tunnels, streets and bridges, as well as in the developing of security systems in banks, government buildings, shopping centers, stadiums, prisons and airports etc. [1]. All of these applications require the development of intelligent surveillance systems able of classifying most kind of motion with a precision as highest as possible. Additionally these systems also must be able to work under several circumstances such as: synchronization among different algorithms, robustness against false alarms, etc.

To properly perform in such applications, any surveillance systems have to satisfy several requirements such as: trustworthiness, image quality, a reduced number of false alarms, interoperability with other systems etc. However several systems satisfying these requirements are expensive, with high maintenance cost or require large storage capacity, in addition to require recording and human intervention during 24 hours per day. Thus the developing inexpensive surveillance systems without requiring the human intervention are an active research topic. As a result,

during the last several years several approaches for tracking and motion detection have been proposed. One of these approaches is based on a simple technique without the use of any classification or optimization stage. Other systems require human supervision such as that proposed by Jones [1], a low cost video motion detection system which is able to supervise 24 Hrs per day; in which one person can monitor many cameras from remote sites. This system is able to record and send a video sequence only when a relevant event is occurring; situation that is detected using the difference between consecutive pictures. To this end firstly the system looks at the surveillance windows to determine if there is any difference between consecutive pictures larger than a given threshold. To improve the detection, it can also be used the histogram of the light distribution within the window and then correct for overall ambient light variations. Zhongfei Zhang [6] proposed the QLS system that is based on a linear consistency analysis. Since the QLS approach only focuses on those elements that are strictly necessary to compute a solution, it reduces the computation to a minimum and achieves a maximum efficacy. The QLS has the following desirable features: no camera calibration is required; the statistics computed in the algorithm are stable because the low condition numbers of matrices resulting avoids the unstable matrix computation problem of high condition numbers. Lorenzo Favalli, Alessandro Mecocci and Moschetti [7] proposed a system built around the features of the MPEG2 coding scheme, and as a consequence, it shares its known limitations in the motion-determination scheme with those of the block-matching algorithm, linear rigid motion, occlusions, local failures and block approximation. This system performs very well in assisting for information insertion/retrieval processes. No segmentation techniques are used for the extraction and tracking objects. This system relies exclusively on the motion information already provided by the MPEG encoder, producing a minimal overhead at the decoder since uses, for 75% of the tracking work, the information already present in the bit stream. Reference [8] presents a moving detection algorithm in which the tracking is achieved by predicting the objects boundary using the motion vectors, followed by a border update, by mean of an occlusion/disocclusion detection algorithm. To improve the motion estimation between frames an adaptive block-based approach is used. In addition, efficient modulation schemes are used to control the gap between the frames used for object tracking. A very important part of this algorithm is the occlusions detection, which consists of two steps: Firstly, the covered regions are estimated using the differences between the displaced frames; and then, using the motion characteristics, the covered regions are classified into actual occlusions and false alarms. The disocclusion detection is performed in a similar manner. The most common applications of the tracking algorithm proposed in [8] are the video compression using MPEG4 and the content retrieval based on standards like H.264L. All of these systems present many desirable features, however they still have several limitations that must be solved.

This paper proposes a moving detection algorithm, intended to be used in several places and under different conditions; that be able of detecting not only the relevant human motion, but also of eliminating irrelevant motions inside the monitored area. In general every change between of frames can be expressed as a simple displacement of pixels, although sometimes it is due to vibrations or abrupt changes between frames and not due to a real human movement. To reduce this problem, the proposed

algorithm used MPEG-2 encoder, available from MPlayer [13], to estimate the movement vectors because, the estimation of the movement vectors (MV) does not require to transmit and decode the coded video stream; besides that the MPEG-2 contains a matrix used to compensate the prediction errors; and this compensation requires less bytes than the whole video frame; allowing also to obtain reliable information about all the objects which are generating movement into the scene; enabling to the system to classify them as relevant or non-relevant motion [2].

To determine if the movement is relevant or not, the movement vectors are divided in horizontal and vertical movement vectors which can be either positive or negative; where a movement vector with a positive value means that the motion is to the right or downwards in the image, while a negative value means that the motion is to the left or upwards, respectively. To reduce the false alarms, relevant motion detection is carried out using several constraints in the motion information already available in the video stream, generated by the MPEG-2 encoder. All involved constraints such as: distance, filtering, counter and tracking have a very important role during the detection process, doing the movement vector the main tool in the proposed motion detection algorithm.

The paper is organized as follows. The structure and basic principles supporting the proposed motion detection algorithm together with an analysis of the advantages and disadvantages of proposed algorithm, in comparison with other previously proposed motion detection methods, are provided in section 2. Experimental results showing the performance of proposed algorithm are given in section 3, and finally the conclusions of this research and future works are provided in section 4.

2 Proposed Motion Detection Algorithm

To detect relevant motion in a scene where normally there would not be movement vectors of MPEG-2 video coding scheme can provide important information about the movement of all objects into a given scene such as: speed, direction and magnitude; which are the main information used by the proposed algorithm, together with some previous knowledge available about the human behavior, to classify and detect the relevant motion, under certain constrains.

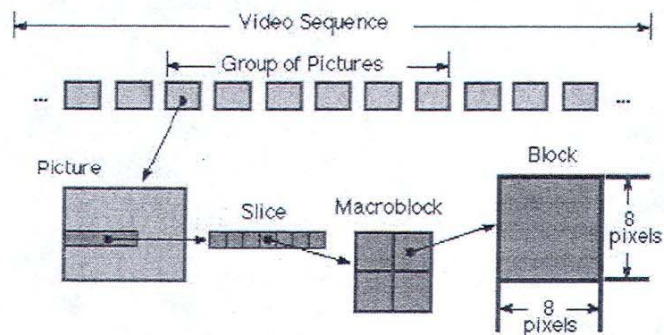


Fig.1. MPEG video sequence

The group of pictures (GOP) of the MPEG, show in Fig. 2, have three different kinds of frames which are needed to support bidirectional and differential codification while keeping the error propagation as small as possible [4], [5]. These picture frames are: The Intra frames (I-frames) are encoded independently of any other pictures in the frame. This achieves moderate compression by reducing spatial redundancy, although does not reduce the temporal redundancy. These frames can be used periodically to provide access points in the bit stream where the decoding can begin. The second type is called Predictive frames (P-frames) which can use the previous I- or P-frame for motion compensation and may be used as a reference picture for further prediction. These frames are divided into non-overlapping frames called macroblock, which can be either predicted or intra-coded. By reducing spatial and temporal redundancy, P-frames offer increased compression compared with the I-frames. These kinds of frames contain one forward movement vector per each macroblock, which indicate the previous position of each macroblock in the reference frame. The third kind of pictures is called Bidirectionally-predictive frames (B-frames), which use the previous and subsequent I- or P-frames for motion-compensation, offer the highest compression. Each macroblock in a B-frame can be forward, backward or bidirectionally predicted or intra-coded. To enable backward prediction from a future frame, the coder reorders the frames from their natural display order to bit stream order so that the B-frame is transmitted after the previous and next reference frames. This process introduces a reordering delay dependent on the number of consecutive B-frames. The B-frames contain forward, backwards or both kinds of movement vectors per each macroblock. Finally a regular GOP is defined by two parameters N and M , which denote the number of pictures in the GOP and the spacing between consecutive P-frames, respectively [5], [4].

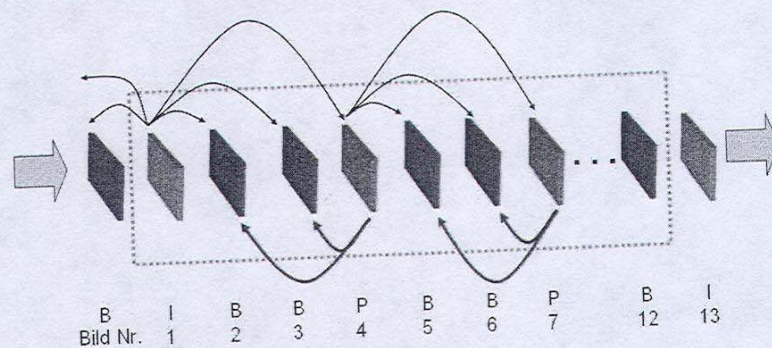


Fig.2. MPEG group of pictures

2.1 Movement Vector Estimation

This technique exploits temporal redundancy among consecutive frame by attempting to predict the $(n+1)$ th frame, from the n th frame using a motion estimation algorithm; in which a macroblock is involved in finding a 8×8 (16×16) sampled region inside the n th frame that closely matches the current macroblock at the $(n+1)$ th frame. Thus an area in the n th frame centered on the current macroblock position, called the search

area, is analyzed and the 8×8 (16×16) region within the search area that minimizes a matching criterion is chosen as the 'best match' [3]. The motion compensation then, consists on selecting the 'best matching' region in the $(n+1)$ th frame, subtracting it from the current macroblock to produce a residual, luminance and chrominance macroblock, that is encoded and transmitted together with a movement vector, which describes the position of the best matching region. Within the encoder, the residual is encoded and decoded and then added to the matching region to form a reconstructed macroblock which is stored as a reference for further motion prediction [3].

2.2 Tuning the Movement Vectors

Consider the P frames and the fact that MPEG encoder estimates the motion vectors in all directions, according with the object's movement, producing the estimation of motion vectors from 0 to 360 degrees. However some of these movement vectors may not represent a relevant movement for our application into the frame and then they can be considered as noisy movement vectors. Taking in account the particular application, security in this case, we can determine that motion is relevant if it occurs from bottom-to-top according with the camera position, as shown in Fig. 3. For these reasons the refinement of movement vectors is an important step, because it will help us to determinate which objects are generating relevant information and which are not doing it.

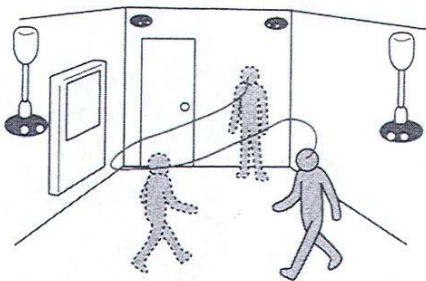


Fig.3. Human Behavior

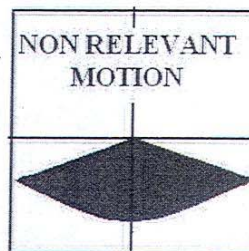


Fig.4. Relevant Movement

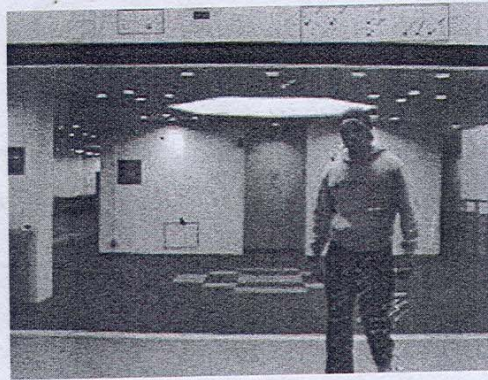


Fig.5. Noise movement vectors

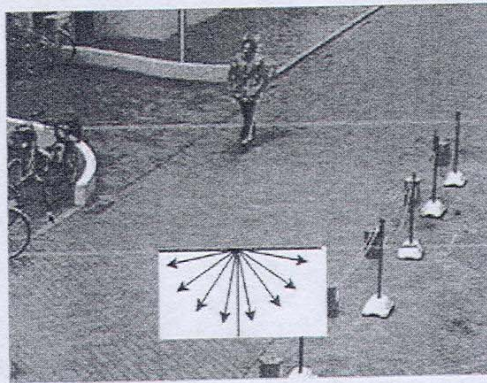


Fig.6. Refinement Movement Vectors

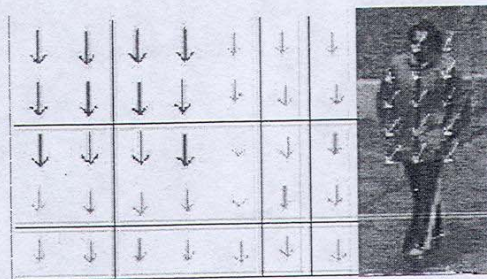


Fig.7. Magnitude of motion vectors for a human images

According to the theory described above we consider as relevant information all movement vectors between 220 and 340 degrees as shown in Fig. 4, considering the camera position, human behavior and security area. Refinements works as a median

filter used to eliminate noisy movement vectors as well as movement vector out of range, as shown in Fig. 5. Eliminate noisy movements vector is an important step to get the correct information about the real motion into the video stream. With this information we can not only determine how many peoples are in the scene but also determine if this motion is generated by a person (Fig. 6), because the magnitude of a person is between 9 and 13 macroblocks, Fig. 7. The reliance of this measure increase using a refinement step; although this magnitude relies on the distance between the camera and the object.

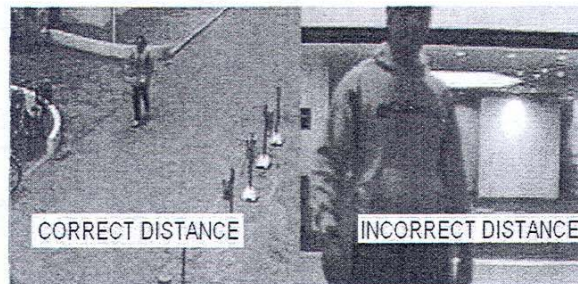


Fig.8. Correct and incorrect distance between the camera and the security area



Fig.9. Counters to determine the number of illigal access

2.3 Distance and counters.

Another important constraint is the distance between the camera and the security area, as shown in Fig. 8, in order to have better perspective of the objects into this area. Counting the number of peoples that are generating relevant motion as well as counting the number of persons who access into the security area is an important step; because it allows us to know how many peoples is generating important motion and

how many of this people access into the security area. This is important because we must take into account that in certain cases some persons change their direction before decide to access or not into the security area (Fig. 9).

3 Evaluation Results

The motion detection algorithm was proved using a Power Mac G5 with 1.25GBytes in RAM; Bus Speed of 167MHz and a CPU Speed of 1.25GHz. The system was evaluated in open area because in such places we can found many movements in many directions making this situation excellent to prove if the propose algorithm is able to work according with the performance mention above. The video frames used to test the performance of proposed algorithm show that it is able to eliminate irrelevant information taking into account just the relevant movement generated by an object into the security area. In the next frame, Fig.10, a person generated motion from top to bottom, there is just one person in the video sequence; this is the simplest test because there is no other motion in any different direction. The result of this test, Fig.11, shows the correct detection using this algorithm.

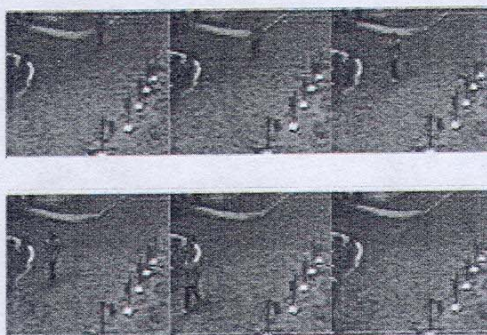


Fig.10. Detection of the movement generated by one person

```
[mpeg2video @ 0x484640]
[mpeg2video @ 0x484640] |
[mpeg2video @ 0x484640] | READY TO MONITOR THIS AREA |
[mpeg2video @ 0x484640] |
[mpeg2video @ 0x484640] |
[mpeg2video @ 0x484640] *****
[mpeg2video @ 0x484640] *TOP side 1 person(s) detected*
[mpeg2video @ 0x484640] *****
[mpeg2video @ 0x484640] ILLEGAL ACCESS!!! ILLEGAL ACCESS!!!
[mpeg2video @ 0x484640] >>>> 1 Person(s) <<<<
[mpeg2video @ 0x484640] ILLEGAL ACCESS!!! ILLEGAL ACCESS!!!
```

Fig.11. Results obtained when only are one illegal access.

In the frame shown in Fig. 12 a person is generating motion from top to bottom and some persons are generating motion from bottom to top, according with this, the algorithm must be able to generate motion vectors of this person, count how many relevant persons appears in the security area and how many access happened.

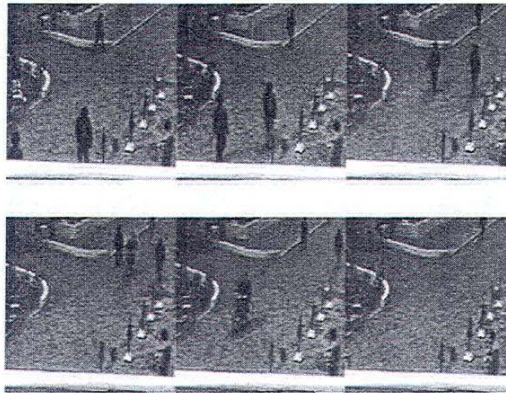


Fig.12. A person generating relevant movement

```
[mpeg2video @ 0x484640]
[mpeg2video @ 0x484640] |
[mpeg2video @ 0x484640] | READY TO MONITOR THIS AREA
[mpeg2video @ 0x484640] |
[mpeg2video @ 0x484640] |
[mpeg2video @ 0x484640] | *****
[mpeg2video @ 0x484640] | *TOP side 1 person(s) detected*
[mpeg2video @ 0x484640] | *****
[mpeg2video @ 0x484640] | ILLEGAL ACCESS!!! ILLEGAL ACCESS!!!
[mpeg2video @ 0x484640] | >>> 1 Person(s) <<<
[mpeg2video @ 0x484640] | ILLEGAL ACCESS!!! ILLEGAL ACCESS!!!
```

Fig.13. Detection of the relevant movement produced by one person

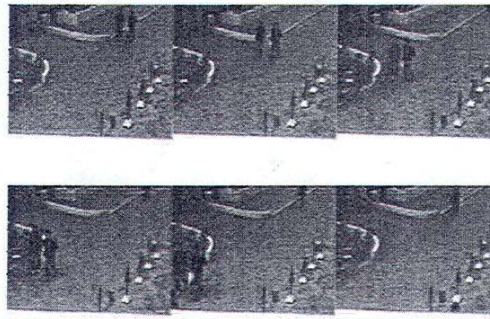


Fig.14. Three peoples generating relevant movement.

In the frame shown in Fig. 14 we can see two persons into the security area after that some persons appear generating motion from bottom to top and one person may appear generating motion from top to bottom while two persons access into the security area, according to this, the algorithm must be able to count 2 persons at the beginning and after that trigger and alarm when these two persons access into the security area and at the same time the system has to count one person more who appears in the scene while the illegal access is occurring. Figure 15 shows the detection results.

```
[mpeg2video @ 0x484640]
[mpeg2video @ 0x484640] |
[mpeg2video @ 0x484640] | READY TO MONITOR THIS AREA
[mpeg2video @ 0x484640] |
[mpeg2video @ 0x484640] |
[mpeg2video @ 0x484640] *****
[mpeg2video @ 0x484640] *TOP side 2 person(s) detected*
[mpeg2video @ 0x484640] *****
[mpeg2video @ 0x484640] ILLEGAL ACCESS!!! ILLEGAL ACCESS!!!
[mpeg2video @ 0x484640] >>>> 2 Person(s) <<<<
[mpeg2video @ 0x484640] ILLEGAL ACCESS!!! ILLEGAL ACCESS!!!
[mpeg2video @ 0x484640] *****
[mpeg2video @ 0x484640] *TOP side 1 person(s) detected*
[mpeg2video @ 0x484640] *****
```

Fig.15. Detection of the relevant movement produced by three persons

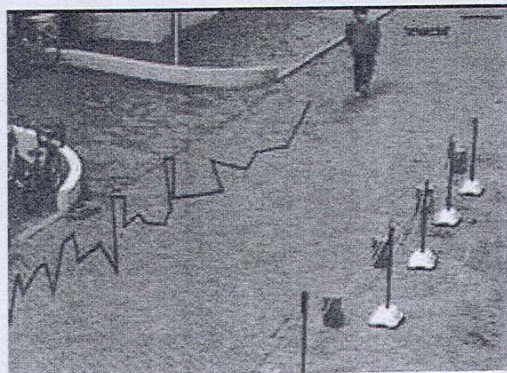


Fig.16. Tracking with refinement of the motion vector

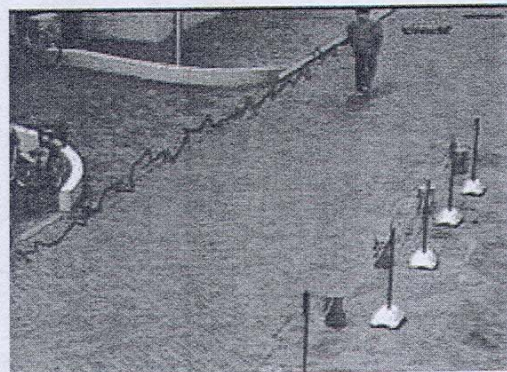


Fig.17. Tracking with refinement of the motion vectors

The refinement step provides us reliable information about the real motion. Figure 16 shows the result of a tracking algorithm using the MPEG2 movement vector

information in a raw way, in this figure we can see that the tracking suffer of many noisy motion vectors which can give wrong information about the real position of the object, giving as a result a irregular and non-reliable information about the real trajectory of each object. Finally in Fig. 17 shows the person tracking with vector refinement, which provides a more reliable information as compare with that provide by Fig. 18.

4 Conclusions

Evaluation results show that the proposed algorithm provides good results working using the motion vectors information to help us to decrease the computation al algorithm. Taking on account only the relevant object information into the security area, it is possible to obtain clear information about how many objects are really important, triggering not only an alarm when a illegal access occurs, providing also information how many illegal access occurs.

As we mention before the purpose of proposed algorithm is establishing a security area, thus it must be able to be used in many circumstances and any places. Evaluation results show that, using the MPEG encoder, the proposed algorithm is able to accurately detect the movement, determining if it relevant or not. The tracking ability of propose algorithm can also be improved using a refining of the movement vectors.

Acknowledgment

I would like to thank to the Ministry of Education Science and Culture of Japan, The University of Electro-Communications of Tokyo Japan, The National Science and Technology Council (CONACYT) and The National Polytechnic Institute Mexico city for the support during the realization of this research.

References

- [1] B. T. Jones, "Low-Cost Outdoor Video Motion and Non-Motion Detection", *Processing of Security Technology*, pp. 376-380, 1995.
- [2] V. Paladino, "Introduction of Video Compression Under the Standard MPEG-2", *The Institute of Electronic Engineer, Spain*, pp.3-24, 2005.
- [3] I. E. Richardson, "H.264 and MPEG-4 Video Compression, Video Coding for the Next-Generation Multimedia", Wiley, UK, pp. 27-41, 2004.
- [4] J. Watkinson, "The MPEG Handbook: MPEG-1, MPEG-2, MPEG-4," Focal Press, 2001.
- [5] P. N. Tudor, "MPEG-2 Video Compression: Tutorial," *Journal of Electronics and Communication Engineering*, December, pp. 1-5, 1995.
- [6] Z. Zhang, *Mining Surveillance Video for Independent Motion Detection*, (ed.), IEEE Internacional Conference on Data Mining, pp. 741-744, 2005.
- [7] L. Favalli, A. Mecocci and F. Moschetti, "Object Tracking For Retrieval in MPEG2," *IEEE, Trans. on Circuit and Syst. for Video Technology*, 427-432, 2000.
- [8] K. Hariharakrishnan, D. Schonfeld, P. Raffy and F. Yassa, "Video Tracking Using Block Matching", *IEEE, International Conference on Image Processing*, pp. 945-948, 2003.

Direction-Based Motion Detection Using the MPEG2 Encoder

Josue A, Hernandez

National Polytechnic Institute
Department of Postgraduate Studies
Av. Santa Ana #1000
Mexico
jahdezg@calmecac.esimecu.ipn.mx

Hiroyoshi, Morita

University of Electro-Communications
Tokyo #182-8585
Japan
hiroyoshi-m@par.odn.ne.jp

Hector M, Perez

National Polytechnic Institute
Department of Postgraduate Studies
Avenida Santa Ana #1000
Mexico
hmppm@calmecac.esimecu.ipn.mx

Abstract: Now a day, the used of biometrics as part of the security systems have been increasing; to detect, classify and to track human motion with a high precision are part of this demand, for that reason many approaches have been proposed and, since computer vision let us to manipulated digital image sequences many of these techniques use the information content in a video stream has the main tool to do this task. In this paper we present a simple but robust technique to detect and to track human motion using MPEG2. the proposed method take advantage of the MPEG2 encoder extracting the movement vector from the P frames and filter them in order to obtain better information about real motion into the scene. Our method implies very low computational cost and low complexity compare with other techniques. Experimental results have shown good accuracy, demonstrating that the proposed technique could be reliable.

Key-Words: MPEG encoder, Movement Vectors, Motion Detection, Tracking, Surveillance System.

1 Introduction

The ability to perform video analysis in the compressed domain has become more attractive regarding the large amount of video data available [1]. the use of the standard MPEG to compress digital video has been used for many years and, until now it represent a reliable compressor concerning detection of shot changes and camera motion characterization [2].

Pattern recognition is a research area that has been amply studied during the last years, its use information contained in video sequences; In many applications of pattern recognition, approaches with recognition capability are usually based on a corpus of data which is treated either in a holistic manner or which is partitioned by application of prior knowledge [3] like shape, velocity, direction, texture, magnitude, behavior and so on from different kind of objects around us.

In this paper, we proposed a relevant technique working in MPEG encoder domain to detect and to track persons, based on their walk's direction. Every change between frames should be expressed as a simple displacement of pixels, but sometimes this is not true due to camera's vibrations or abrupt changes on brightness or shadow displacements between frames. Trying to decrease these sort of problems *MPEG-2 encoder* available from *Mplayer*¹ to calculate the movement vectors is used in this technique.

Related work is described in section 2, a short

MPEG review are shown in section 3, a deep explanation of our method is written down in section 4. Experimental results of the proposed technique are shown in section 5 ending with conclusions in section 6 followed by the reference .

2 Related Work

The problem of detecting and tracking motion has been investigated in the past, with different approaches, some of these approaches are based in simple techniques having not any aid from a classifier or refiner, we give a brief background of similar techniques on motivation to use the propose technique.

Abdeljabar Benzougar, Patrick Bouthemmy and Roman Fabelt [1] this approach is based on the MPEG coded domain. the detection is stated as a Markovian labeling issue in terms of macroblocks conforming or not to the estimated dominant image motion assumed to be due to the camera motion. The dominant motion estimation and the moving object detection stages only utilize MPEG motion vectors and the DC coefficient of DCT directly extracted from the MPEG bit stream of the processed video.

This method demonstrate that can handle various situations where moving objects are present in the scene observed with a mobile camera. However, the efficiency of the moving object detection depend on the quality of MPEG motion vectors, since they do not used a filtering step to avoid this problem. this method

¹<http://www.mplayerhq.hu/design7/news.html>

also has a constrain in the context of contend-based video browsing and indexing that should be taken into account when the MPEG encoder is selected.

Lorenzo Favalli, Alessandro Mecocci and Fulvio Moschetti [8] the aim of this work is the realization of a tool that could extend these capabilities to MPEG2 streams, allowing to track objects selected by a user throughout the sequence. the way in which the tracking has been realized, is through exploitation of the motion information already present in the bit stream and generated by the decoder, another important feature of the proposed scheme is that it is not necessary to decode the full sequence if the user wants to access only part of it. the proposed scheme on this paper performs very well in assisting the information insertion/retrieval process. However, no segmentation or filtering techniques are used for the extraction and tracking of the objects, since it relies exclusively on the motion information already provided by the MPEG encoder. Results show that the algorithm performs well though it is slightly dependent on the object shape.

Karthik Hariharakrishnan, Dan Schonfeld and Fathy Yassa [9] tracking is achieved by predicting the object boundary using motion vectors, followed by contour update, using occlusion/disocclusion detection. An adaptive block-based approach has been used for estimating motion between frames.

An efficient modulation scheme is used to control the gap between frames used for object tracking. The algorithm for detecting occlusion proceeds in two steps. First, covered regions are estimated from the displaced frame difference. these covered regions are classified into actual occlusions and false alarms using motion characteristics. Disocclusion detection is also performed in a similar manner.

3 MPEG Overview

Since MPEG2 is well know standard of compression, we will not to go deeply into this subject. However, we consider important to explain, how the movement vectors are calculated into the encoder, such process involves two important steps: *Motion Estimation*, and *Motion Compensation*.

Motion Estimation for a macroblock involves find a 16x16-sample region in a reference frame that closely matches the current macroblock. The reference frame is a previously-encoded frame from the sequence and may be before or after the current frame in display order. An area in the reference frame centered on the current macroblock position (the search area) is searched and the 16x16 region within the search area that minimizes a matching criterion is chosen as the

'best match'[3]. In the *Motion Compensation* the selected 'best' matching region in the reference frame is subtracted from the current macroblock to produce a residual macroblock (luminance and chrominance) that is encoded and transmitted together with a *movement vector* describing the position of the best matching region. Within the encoder, the residual is encoded and decoded and added to the matching region to form a reconstructed macroblock which is stored as a reference for further motion-compensated prediction [3].

MPEG have three different sort of frames Intra frames *I frames*, Predictive frames *P frames* and Bidirectionally-Predictive frames *B frames* which are needed to support bidirectional and differential codification while the error propagation is diminished and a N number of this frames creating the **GOP** as shown in *fig.1*. Since *P frame* is forward predicted from a previous frame which can be *I frame* or *P frame* its contains movement vectors that indicates the previous position of each macroblock in the reference frame, on the other hand, *B frame* is built using the previous and the next I or P frames for motion compensation, and offer the highest degree of compression. each macroblock in a *B frame* can has forward movement vectors, backward movement vectors or bidirectionally movement vectors predicted or intra-coded [3].

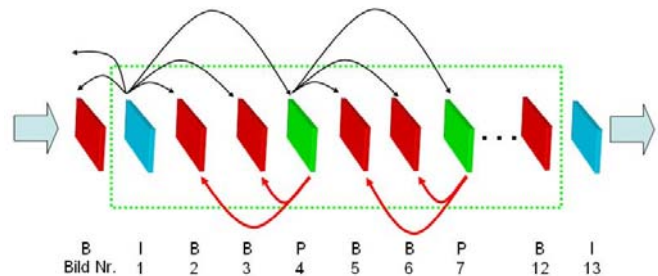


Figure 1: MPEG Group of Pictures (GOP)

4 Movement Vectors Extraction

We already know that MPEG estimate movement vectors in *P frames* and *B frames*. In this technique we use only movement vectors from *P frames*. Since, they represent the real direction from each object. These movement vectors are divided in a *horizontal* and a *vertical* part. These parts can be positive or negative. A positive value means motion to the right or motion downwards and a negative value means motion to the left or motion upwards, respectively. However, the data extracted from the MPEG encoder content in *MPlayer* is not represented in that way, instead

of that, it only shows the (x,y) position of each macroblock (40 macroblocks in the x axes and 30 macroblocks in the y axes using a 640×480 frame) and its old and new pixel position. For that reason, we should convert this data from pixel position to distances between coordinates (x,y) and to $(0-360)$ degrees.

Some of the movement vectors extracted from the compressor just indicate abrupt light variation or small vibration on the videocamera. On this cases, we call them noisy vector since they do not represent a real motion. also the movement vectors calculated because of the motion from fixed elements like trees or shrubs are analyzed to be eliminated.

4.1 Filtering Movement Vectors

As we mention above, there are two kind of movement vectors that we should eliminate in order to get only reliable information about all the real motion into the scene. the first step is to eliminate the noisy vectors, for that, we focused on how the median filter works and the behavior of this vectors.



Figure 2: Noisy Movement Vectors

Since the quantity of these vectors is small because of the runtime process of the decoder and they are not continuous or similar between them like those ones created by a real motion could be easy to distinguish from the rest. However, since the movement vectors are calculated one by one a buffer was necessary to retain certain number of them as well as determinate the size of our filter's mask that give us the best filtering. After the noisy vectors are eliminated from our data, we continue with the next step, eliminate those movement vectors generated by the motion from fixed objects, some of these vector are eliminated by the filter and the others are eliminated based on their position and nonuniform direction. Our method is able to measure the level of importance from each real motion according with their position in the scene dur-

ing the process, according with that, a movement vector is eliminated if its position does not change after some frames and considering also that its direction is not going to be uniform as well as its continuity. We can see in *fig.2* and *fig.3* the results from this process.

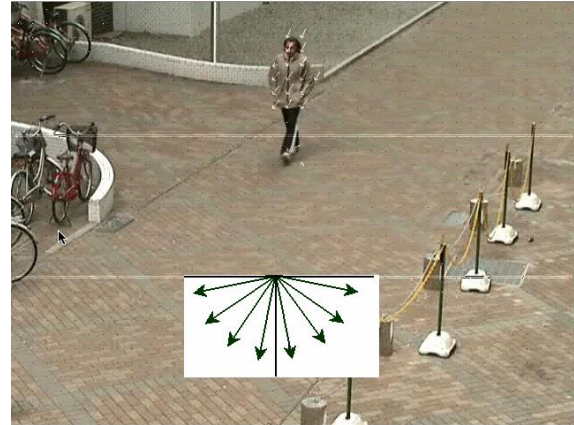


Figure 3: Filtered Movement Vectors

5 Movement Parameters

In order to provide information for higher level computer vision systems, it is necessary to estimate movement parameters such as velocity, tracking and direction of moving objects from the video stream. A general description of each one is achieved and some considerations on issues are given:

5.1 Motion Direction

Utilizing the information given by the movement vectors in order to determine the motion direction to classify the event as relevant or irrelevant the degrees previous calculated are used.

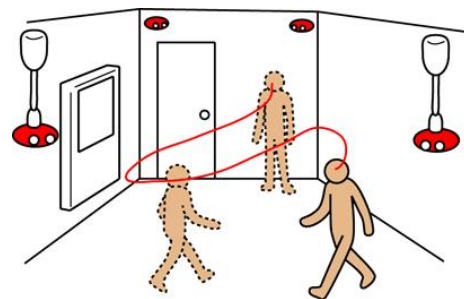


Figure 4: Relevant Human Displacement

Since the idea is try to detect an entrance motion in a place where there would not be motion in

that direction only the movement vectors between 220 and 340 degrees are consider as relevant. Such range was calculate using two considerations the first one is based on the human behavior about the way they walk and the second one is taken into account that it does not matter the object's trajectory if it is made from top to bottom will generate movement vector into this range as is shown in *fig.4*.

The distance between the videocamera an the scenario (*fig.5*) is as well, an important constrain in order to have a better perspective of the object, it also help us to conserve a constant magnitude in each element in it. the magnitude of a person after filter the vectors is between 9 to 13 as shown in *fig.6*.



Figure 5: Videocamera-Scenario Correct Distance

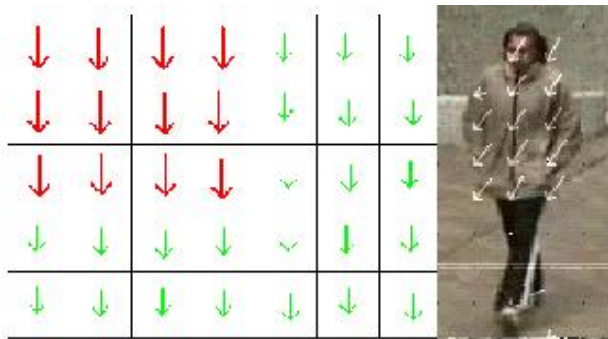


Figure 6: Magnitude Calculated From a Person

5.2 Motion Tracking

Motion tracking means, not only followed an object during the process but also analyze this trajectory to determinate the origin of it. for that reason, the use of counters (*fig.7*) let us to calculate the number of persons that are generating relevant motion according with their direction and determinate from which place they appears into the scenario (*talking about surveillance systems these kind of information could be very important*). This process used the direction of each movement vector and three imaginary divisions (*fig.8*) to estimate the position from each element that generates real and relevant motion, such estimation help us

to determinate the importance level from a each object to decide if it can be consider as an alerting trajectory or not.



Figure 7: Counters



Figure 8: Measure the Object's Position

6 Experimental Results

The basic idea is simple, *to detect and to track* relevant motion in a scene where there would normally be no motion in certain direction.

This method was tested over a *Power Mac G5* with 1.25GB in RAM, Bus Speed of 167MHz and a CPU Speed of 1.25GHz. We decide to monitor free areas since we can find many objects generating motion in different directions doing this situation an excellent scenario to test if the propose technique is able to work according with the performance mention above.

As we can see below, our technique is able to detect relevant motion into the scene, in the frame shown in (*fig.9*) a person generated motion from top to bottom, this is a simplest test because there is no other motion in a different direction; The result of this test shown in (*fig.10*) is presented.



Figure 9: Detection of a Simple Relevant Motion

```
[mpeg2video @ 0x484640]
[mpeg2video @ 0x484640] |
[mpeg2video @ 0x484640] |          READY TO MONITOR THIS AREA
[mpeg2video @ 0x484640] |
[mpeg2video @ 0x484640] |
[mpeg2video @ 0x484640] |*****
[mpeg2video @ 0x484640] *TOP side 1 person(s) detected*
[mpeg2video @ 0x484640] |*****
[mpeg2video @ 0x484640] ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
[mpeg2video @ 0x484640] >>>> 1 Person(s) <<<<
[mpeg2video @ 0x484640] ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
```

Figure 10: Result of the First Test



Figure 11: A Person Generating Relevant Motion

```
[mpeg2video @ 0x484640]
[mpeg2video @ 0x484640] |
[mpeg2video @ 0x484640] |          READY TO MONITOR THIS AREA
[mpeg2video @ 0x484640] |
[mpeg2video @ 0x484640] |*****
[mpeg2video @ 0x484640] *TOP side 1 person(s) detected*
[mpeg2video @ 0x484640] |*****
[mpeg2video @ 0x484640] ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
[mpeg2video @ 0x484640] >>>> 1 Person(s) <<<<
[mpeg2video @ 0x484640] ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
```

Figure 12: Result of the Second Test

In the case of (fig.11) a person is generating motion from top to bottom and some persons are generating motion from bottom to top, according with this, our technique must be able to, count how many relevant objects in motion appears and analyze their or its trajectory to determinate how many of them access in the security area. the result is shown in fig.12

Next test (fig.13) and (fig.14) shown reliable results from a more complex monitoring.



Figure 13: Three People Generating Relevant Motion

```
[mpeg2video @ 0x484640]
[mpeg2video @ 0x484640] |
[mpeg2video @ 0x484640] |          READY TO MONITOR THIS AREA
[mpeg2video @ 0x484640] |
[mpeg2video @ 0x484640] |*****
[mpeg2video @ 0x484640] *TOP side 2 person(s) detected*
[mpeg2video @ 0x484640] |*****
[mpeg2video @ 0x484640] ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
[mpeg2video @ 0x484640] >>>> 2 Person(s) <<<<
[mpeg2video @ 0x484640] ILLEGAL ACCESS!!!ILLEGAL ACCESS!!!
[mpeg2video @ 0x484640] |*****
[mpeg2video @ 0x484640] *TOP side 1 person(s) detected*
[mpeg2video @ 0x484640] |*****
```

Figure 14: Result of the Third Test

7 Conclusion

The motion detection and tracking described in this paper was applied in a surveillance scenario where to monitor and detect relevant entrance is the main goal. the proposed technique shown to be reliable with respect abrupt changes of brightness, changes in the environment, discontinuous motion parameters and objects direction. The main advantage of this technique is the use-only of movement vectors' information to detect and to track all relevant motion into the scenario having the ability to continue monitoring while a 1 or more objects have been detected.the run time process was decreased comparing with others techniques as well as the complexity.

Counting only the relevant object into the security area give a clear information about how many objects are really important, filtering the noisy movement vectors is an important part on this method because in that way we can eliminate irrelevant information and get a low error rate during the detection.

Acknowledgements: The research was supported by the National Science and Technology Council (CONACYT) and in the case of the first author, it was also supported by PIFI-IPN scholarship. I would like to thank the University of Electro-Communication for all the support given during this research.

References:

- [1] A. Benzougar, P. Bouthemy and R. Fablet, MRF-Based Moving Object Detection From MPEG Coded Video, *IEEE* 0-7803, 2001, pp. 402–404.
- [2] V. Paladino, , *Introduction of Video Compression Under the Standard MPEG-2*, Institute of Electronic Engineer, Spain, pp. 3–24.
- [3] Marks S. Nixon, Tieniu N. Tan and R. Chellappa, Human Identification Based on Gait, Springer –Verlag, UK–Beijing, China–Maryland USA 2006, pp. 1–15.
- [4] Bernard J. T. Jones, ASTRA Developments Ltd, 1995, *Low-Cost Outdoor Video Motion and Non-Motion Detection*, *IEEE* Denmark, pp. 376–380.
- [5] Iain E. G. Richardson, *H.264 and MPEG-4 Video Compression, Video Coding for the Next-Generation Multimedia*, Wiley, UK, 2004, pp. 27–41.
- [6] Watkinson J., *The MPEG Handbook: MPEG-1, MPEG-2, MPEG-4*, Focal Press, 2001, pp. 10–15, 150–200.
- [7] Shai Avidan, *Support Vector Tracking*, *IEEE*, Vol. 26 No. 8, 2004, pp. 1064–1071.
- [8] L. Favalli, A. Mecocci and F. Moschetti, *Object Tracking For Retrieval in MPEG2*, *IEEE*, 2000, pp. 427–432.
- [9] K. Hariharakrishnan, D. Schonfeld, P. Raffy and F. Yassa, *Video Tracking Using Block Matching*, *IEEE*, 2003, pp. 945–948.