

Capítulo II. Marco Teórico.

Hoy en día existen diversos métodos para la estimación de movimiento en escenas dinámicas. Cada uno de los métodos para dicha estimación se puede clasificar según su funcionamiento y las técnicas que utilizan para la construcción del campo de movimiento. Existen diversas clasificaciones entre las que destacan principalmente dos de ellas, las cuales son las basadas en correspondencia y las basadas en gradiente, sin embargo, también podemos encontrar las basadas en filtración espacio-temporal, basadas en fases y en redes neuronales.

En este capítulo se profundizará en el funcionamiento de las tres clasificaciones de mayor auge en la actualidad, estas son las basadas en correspondencia, basadas en gradiente y basadas en filtración espacio-temporal. Dentro de cada una de éstas clasificaciones se pueden encontrar varios métodos para resolver el problema de la estimación de características de movimiento.

A continuación se detalla tanto las bases del funcionamiento de cada clasificación como algunos de los métodos más populares dentro de ellas.

2.1 MÉTODO BASADO EN CORRESPONDENCIA.

En [Molina, 2002] se menciona que este tipo de métodos son muy parecidos a los métodos llamados stereo-point en visión por computadora. En este método como en todos los demás de análisis de movimiento trabaja sobre una disparidad entre escenas de diferente tiempo, es decir, cuando se toma una serie de imágenes en un transcurso de tiempo Δt con la misma

cámara. Al igual que los métodos basados en gradiente, los métodos de correspondencia utilizan un número pequeño de imágenes en una secuencia de tiempo.

En correspondencia, los puntos críticos, líneas y otros atributos de algún objeto de la primera imagen son puestos en comparación, esto es, en concordancia con atributos semejantes encontrados en imágenes posteriores. Tanto la equivalencia obtenida y representada por la correspondencia como el intervalo de tiempo en las imágenes determinan el campo de movimiento del objeto que posee las características estudiadas.

Existen dos técnicas comunes de tiempo real para disminuir el costo computacional necesario para hacer las correspondencias de características entre imágenes [Laplante, 1996]:

1. Reducir el número de características críticas.
2. Uso de métodos de multirestricciones para obtener una estimación del campo de movimiento.

La primera de las técnicas mencionadas con anterioridad es la más comúnmente utilizada, sin embargo, existen problemas cuando se necesita una gran exactitud de las características del objeto en movimiento. La segunda estrategia está siendo probada y estudiada por investigadores quienes buscan mejorar la robustez de las técnicas de correspondencia.

Hace algunos años varios investigadores trabajando sobre la tesis doctoral de Ullman, relacionada con el tema de correspondencia, determinaron que con tan solo comparar ocho puntos en imágenes dinámicas era suficiente para determinar el movimiento de un objeto y su estructura [].

El método más popular y de mejor desempeño dentro de correspondencia es el llamado Método de Correspondencia de Bloques, del cual se explica su funcionamiento a continuación.

2.1.1 ALGORITMO DE CORRESPONDENCIA DE BLOQUES.

Como se mencionó en el capítulo anterior, la estimación del campo de movimiento por medio de este método consiste en encontrar un punto de coordenadas (x', y') que corresponda al centro del segmento de una imagen en el tiempo t que minimice la correspondencia con el centro del mismo segmento de otra imagen con coordenadas (x, y) en el tiempo $t - \Delta t$. Esta estimación de la correspondencia se obtiene mediante la siguiente fórmula:

$$C(x, y, x', y') = \sum_{m=-M/2}^{M/2} \sum_{l=-L/2}^{L/2} E(x + m, y + l, t - \Delta t) - E(x' + m, y' + l, t)$$

En donde $E(x, y, t)$ representa el valor del brillo de un punto de la imagen en el tiempo t con coordenadas x, y . Los valores de M y L representan las dimensiones del segmento de la imagen a lo largo del eje X y del eje Y respectivamente. Generalmente $M = L$ por lo que obtenemos segmentos de comparación cuadrados [Molina, 2002].

La minimización de la correspondencia consiste en buscar el valor mínimo para la función $C(x, y, x', y')$ cercano al punto de coordenadas (x, y) para las funciones:

$$x' = x - F/2, \dots, x + F/2 \quad y' = y - F/2, \dots, y + F/2$$

Donde F es la dimensión del segmento de la imagen en donde se busca el valor mínimo. Una vez encontrado el valor de coordenadas x', y' en el cuál se minimizó el valor de $C(x, y, x', y')$, éste punto será el centro del nuevo segmento para la imagen en el tiempo

t. Cuando se obtuvieron los valores de (x, y) en $t - \Delta t$ y (x', y') en t , se pueden obtener las componentes del desplazamiento para el eje X y para el eje Y para cada píxel de coordenadas (x, y) en el tiempo $t - \Delta t$ por medio de las siguientes diferencias:

$$Sx = x' - x \quad Sy = y' - y$$

Los valores resultantes para Sx y Sy que corresponden al desplazamiento de un punto en los ejes X y Y respectivamente están dados por números enteros lo cual podría llevarnos a cometer errores [Laplante, 1996].

La siguiente gráfica ilustra el funcionamiento del método de correspondencia de bloques.

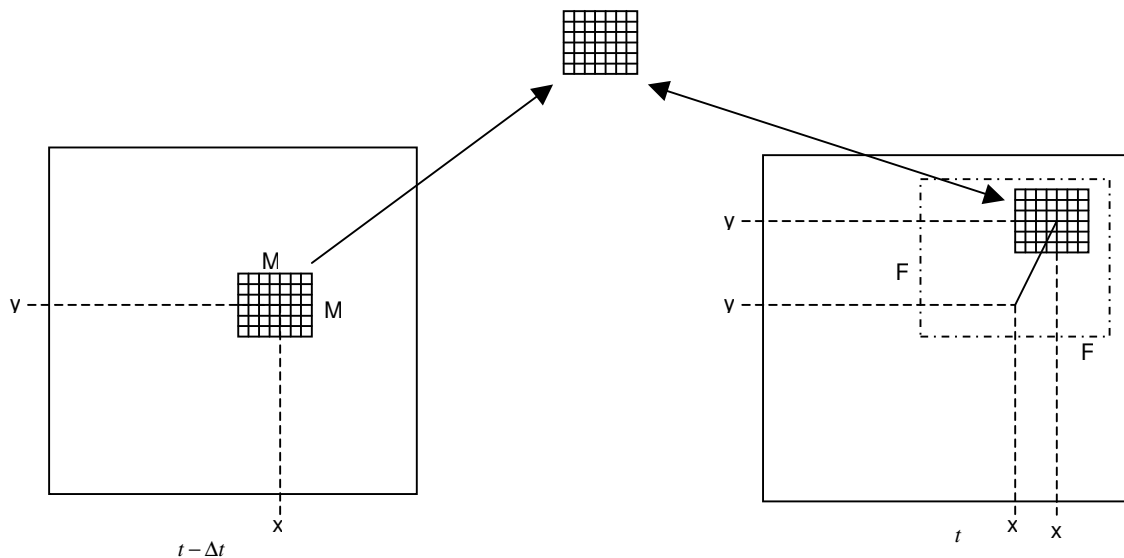


Figura 2.1 Funcionamiento de correspondencia de bloques.

El método de correspondencia de bloques es capaz de estimar movimientos rotacionales debido a que se va actualizando constantemente, esto es, que una vez

encontrados los puntos con coordenadas (x', y') , estos se convertirán en los puntos (x, y) a ser rastreados en la siguiente imagen.

Este método tiene algunas desventajas como es el hecho que es sensible al ruido, este problema se puede solucionar aumentando las dimensiones del segmento a ser comparado, esto es, aumentar los valores para M y L . Otra desventaja es que si el desplazamiento de un objeto es mayor a $F/2$, este no puede ser medido, por lo que se debe restringir la velocidad máxima de los objetos en movimiento, la forma más sencilla de solucionar este problema es aumentando el tamaño de L , sin embargo, el costo computacional aumentará. La complejidad para este algoritmo para segmentos de M^2 es $O(F^2M^2)$ [Laplante, 1996].

2.2 MÉTODOS BASADOS EN GRADIENTE.

Los métodos basados en gradiente proveen la solución al problema de estimación de movimiento observando los cambios en el brillo de la imagen. Estos cambios en el brillo son modelados por medio de ecuaciones parciales llamadas ecuaciones limitadoras (constraint equations) las que su solución nos provee el llamado campo de movimiento.

El gradiente es un vector, en donde sus componentes miden la rapidez en que los valores de los píxel cambian en la distancia y en las direcciones x e y . dx y dy son las distancias en las direcciones x e y respectivamente, en términos de número de pixels entre dos puntos [Penedo, 2000].

$$\frac{\partial f(x, y)}{\partial x} = \Delta_x = \frac{f(x + d_x, y) - f(x, y)}{d_x}$$

$$\frac{\partial f(x, y)}{\partial y} = \Delta_y = \frac{f(x, +d_y + y) - f(x, y)}{d_y}$$

En orden de detectar la presencia de una discontinuidad en el gradiente, debemos calcular el cambio en el gradiente en el punto (x, y) . Esto se puede hacer referenciando la medida aportada por la magnitud del gradiente y su dirección [Penedo, 2000].

$$M = \sqrt{\Delta^2 x + \Delta^2 y}$$

$$\theta = \tan^{-1}\left(\frac{\Delta y}{\Delta x}\right)$$

La ecuación diferencial parcial más importante para el modelado del campo de movimiento es obtenida al considerar los cambios en el brillo de la imagen con respecto al tiempo $E(x(t), y(t), t)$, dicha fórmula es la siguiente [Laplante, 1996]:

$$\frac{dE}{dt} = \frac{\delta E}{\delta x} \frac{dx}{dt} + \frac{\delta E}{\delta y} \frac{dy}{dt} + \frac{\delta E}{\delta t} = 0$$

Dada la ecuación anterior, si consideramos una u que corresponda a $\frac{dx}{dt}$ y una v que corresponda a $\frac{dy}{dt}$, las cuales son componentes del vector de velocidad \bar{V} sobre los ejes X y Y respectivamente se obtiene la función conocida como Optical Flow Constraint (OFC).

$$Exu + Eyv + Et = 0$$

Esta ecuación puede ser considerada como la ecuación de una línea recta dentro del plano (u, v) , es decir:

$$v = mu + c$$

Donde $m = -(E_x / E_y)$ es la pendiente y $c = -(E_t / E_y)$ es la intersección. Cualquier punto que pertenezca a esta línea es una posible solución al problema de la estimación del campo de movimiento, por lo que solamente la función OFC no proporciona un valor único y para poder obtener este valor único se necesitan de otras ecuaciones limitadoras. Estas funciones limitadoras extras pueden ser ya sea el uso de detección de bordes o el suavizado.

Teniendo la ecuación OFC podemos observar que la componente perpendicular de la velocidad que es paralela a ∇E se puede calcular por medio de la fórmula siguiente:

$$V_{\perp} = -\frac{E_t}{\|\nabla E\|} \frac{\nabla E}{\|\nabla E\|} \quad \text{Donde se asume que } \frac{dE}{dT} = 0 \text{ y } \|\nabla E\| \neq 0.$$

Dentro de los métodos basados en gradiente para la estimación del campo de movimiento podemos encontrar dos algoritmos principales que son: algoritmos basados en regularización (regularization-based approaches) y algoritmos basados en multirestricciones (multiconstraint-based approaches) [Laplante, 1996].

2.2.1 ALGORITMOS BASADOS EN REGULARISACIÓN.

En [Laplante, 1996] se explica que en este tipo de métodos, se define una función en la que una restricción de suavizado es usada para regularizar la solución de la ecuación diferencial parcial. La influencia de la restricción de suavizado depende directamente de una constante positiva la cual representa el peso de dicha restricción. La función es minimizada utilizando cálculo de variaciones o por relajación estocástica.

Dentro de las diversas soluciones basadas en regularización para la estimación del campo de movimiento, la más famosa fue propuesta por Horn y Schunk, la cual se basa en la minimización de la función:

$$\iint \left[(E_x u + E_y v + E_t)^2 + \alpha^2 (u_x^2 + u_y^2 + v_x^2 + v_y^2) \right] dx dy$$

Donde el término $(E_x u + E_y v + E_t)$ corresponde a la función OFC, el término $(u_x^2 + u_y^2 + v_x^2 + v_y^2)$ es una medida de la suavización del campo de movimiento, y α es el factor de peso que controla la influencia de la restricción de suavizado [Laplante, 1996]. Ésta función es minimizada utilizando el cálculo de variaciones, el cual produce una pareja de ecuaciones diferenciales parciales:

$$\nabla^2 u = \frac{E_x}{\alpha^2} (E_x u + E_y v + E_t) \quad \nabla^2 v = \frac{E_y}{\alpha^2} (E_x u + E_y v + E_t)$$

El siguiente par de ecuaciones es utilizado para estimar las componentes del campo de movimiento para cada instante de tiempo:

$$u_{(i,j,t)}^{n+1} = u_{(i,j,t)}^{-n} - \frac{E_{x(i,j,t)} [E_{x(i,j,t)} u_{(i,j,t)}^{-n} + E_{y(i,j,t)} v_{(i,j,t)}^{-n} + E_{t(i,j,t)}]}{(\alpha^2 + E_{x(i,j,t)}^2 + E_{y(i,j,t)}^2)}$$

$$v_{(i,j,t)}^{n+1} = v_{(i,j,t)}^{-n} - \frac{E_{y(i,j,t)} [E_{x(i,j,t)} u_{(i,j,t)}^n + E_{y(i,j,t)} v_{(i,j,t)}^n + E_{t(i,j,t)}]}{(\alpha^2 + E_{x(i,j,t)}^2 + E_{y(i,j,t)}^2)}$$

Donde n es el número de iteraciones. Se nota que la estimación de $E_{x(i,j,t)}$ y $E_{y(i,j,t)}$ requiere de la comunicación de los datos de los píxeles vecinos, así como la estimación de:

$$\bar{u}_{(i,j,t)} = (u_{(i-1,j-1,t)} + u_{(i-1,j+1,t)} + u_{(i+1,j-1,t)} + u_{(i+1,j+1,t)})/12 + (u_{(i-1,j,t)} + u_{(i+1,j,t)} + u_{(i,j-1,t)} + u_{(i,j+1,t)})/6$$

$$\bar{v}_{(i,j,t)} = (v_{(i-1,j-1,t)} + v_{(i-1,j+1,t)} + v_{(i+1,j-1,t)} + v_{(i+1,j+1,t)})/12 + (v_{(i-1,j,t)} + v_{(i+1,j,t)} + v_{(i,j-1,t)} + v_{(i,j+1,t)})/6$$

Debido a que el desempeño del algoritmo de Horn y Schunk es en función del número de iteraciones I_t que se necesitan para obtener la estimación final y las dimensiones de la imagen, la complejidad del algoritmo es $O(I_t^2 I^2)$ [Laplante, 1996].

La mayoría de algoritmos basados en regularización, tienden a producir campos de movimiento suavizados para los valores de la velocidad de propagación de los puntos estimados.

Aun cuando estos métodos convergen en un mínimo, esto no implica que converjan en la solución óptima, esto es debido a que el error mínimo de la estimación del campo de movimiento no corresponde a la iteración en la que el proceso obtuvo ese valor mínimo de la función.

2.2.2 ALGORITMOS BASADOS EN MULTIRESTRICCIÓN

En [Barron, et al., 1995] se explica como este tipo de algoritmos definen ecuaciones las cuales determinan un sistema de ecuaciones el cual es resuelto generalmente usando mínimos cuadrados o razonamiento probabilístico. Los métodos basados en multirestricciones son clasificados basándose en los métodos adoptados para la construcción del sistema de ecuaciones y en las técnicas matemáticas escogidas para su solución, estas técnicas matemáticas pueden ser tanto algebraicas como de “clustering”.

Dentro de este tipo de algoritmos podemos encontrar tres diferentes métodos los cuales toman una hipótesis inicial diferente. Estos métodos son los siguientes:

El primer método trata la hipótesis de lo estacionario, principalmente adoptado para el brillo de la imagen, esto es, toma la ecuación $dE/dt = 0$. Este método debe ser válido para cualquier función de movimiento invariante como son el contraste, entropía, curvatura, varianza del brillo, entre otras. Este método permite definir varias ecuaciones diferenciales

parciales de estructura semejante a la ecuación del campo de movimiento para el estudio de un mismo punto dentro de la imagen.

En el segundo método se obtienen nuevas ecuaciones de restricción tomando la derivada de la ecuación delimitadora fundamental (función del campo de movimiento o la función extendida del campo de movimiento) con respecto a x , y , t y se asume que estas derivadas son iguales a cero.

El tercer método supone que el cambio del campo de movimiento sigue un patrón aproximadamente lineal. Debido a esto se puede obtener una solución de la estimación del campo de movimiento suavizada, ésta se obtiene por medio de una aproximación lineal de la restricción aplicada a los puntos vecinos del punto en estudio. Este método sólo es válido cuando el campo de movimiento a ser estudiado es suavizado. Posteriormente se definen ecuaciones de vecinos para un punto de tamaño $N * N$, y por consecuencia, cuando el valor de N es muy grande, se llega a perder resolución en la imagen [Laplante, 1996].

• **SOLUCIONES ALGEBRAICAS.**

El proceso para la obtención de soluciones algebraicas puede ser considerado como un proceso a seguir que consta de cuatro pasos, los cuales son:

1. Repliegue de las imágenes por medio de un filtro Gausseano.
2. Estimación de las derivadas del brillo de las imágenes.
3. Determinar un sistema de ecuaciones equivalente por medio del método pseudo-inverso.
4. Estimación de las componentes de la velocidad al resolver el sistema de ecuaciones determinado.

La complejidad para este tipo de soluciones se obtiene por la siguiente expresión la cual consta de cuatro términos, donde cada uno de ellos corresponde a cada una de las fases mencionadas anteriormente:

$$C() = G^2 I^2 + QnI^2 + QN^2 \left[\frac{n}{2}(n-1) + n \right] I^2 + Qn^3 I^2$$

Donde I es la dimensión de la imagen, G es la dimensión del filtro Gausseano, Q es el número de ecuaciones de restricción definidas para cada píxel, n es el número de variables desconocidas del sistema de ecuaciones y N es el tamaño del área de estudio de los puntos vecinos. [Laplante, 1996]

- **SOLUCIONES DE CLUSTERING.**

Dado un sistema de ecuaciones determinado puede ser resuelto por medio de técnicas diferentes a las algebraicas. En este tipo de soluciones, cada ecuación de restricción representa una línea de restricción que puede ser dibujada en el plano (u, v) . Estas líneas de restricción que representan al mismo campo de movimiento deben interceptarse en un punto en común.

La función principal de las técnicas de clustering es solucionar los problemas de las soluciones de mínimos cuadrados, para los cuales la solución es un punto que se encuentra ubicado en la distancia mínima entre los puntos de intersección de todas las líneas de restricción. Dado que las técnicas de clustering mejoran la solución proporcionada por los mínimos cuadrados, las técnicas de clustering deben ser capaces de determinar la solución más probable.

Una de las primeras soluciones de clustering fue propuesta por Fennema y Thompson quienes basan su solución en la transformada de Hough y en la ecuación del campo de movimiento. La solución representa el hecho que por medio de la transformación de Hough a una línea en el plano (u, v) corresponde a un punto en el dominio transformado.

Las líneas que tienen una intersección común dentro del plano (u, v) producen en el espacio transformado puntos que son distribuidos a través de curvas bien definidas. La verificación de la presencia de dichas curvas es computacionalmente pesado y se deben de analizar una gran cantidad de líneas de restricción con el fin de obtener estimaciones robustas.

Otro método de solución por clustering fue propuesto por Schunk. En éste método la solución se obtiene agrupando las intersecciones de varias líneas de restricción con la línea de restricción del píxel central del segmento de píxeles a estudiar. Obtenido esto, se puede determinar el campo de movimiento por medio del grupo de intersecciones más dominante. Sin embargo, el resultado de la estimación del campo de movimiento es erróneo si la línea de restricción del píxel central es afectada por errores, los cuales pueden ser provocados por el ruido de las mismas imágenes.

2.3 METODO SUSAN DE DETECCION DE ESQUINAS PRINCIPALES.

Por sus siglas en inglés “Smallest Univalued Segment Assimilating Nucleus” se usan las esquinas principales de los objetos así como la filtración de ruido en las imágenes durante las búsquedas de los bordes.

Para su funcionamiento éste método utiliza una máscara circular la cual está formada por un centro al que se le denomina núcleo y un radio el cual puede ser variable. Esta máscara es barrida sobre las zonas de interés de la imagen, si el brillo de los píxeles que se encuentran bajo la máscara tienen el mismo valor que el núcleo de la máscara, entonces estos píxeles forman el área llamada USAN, por sus siglas en inglés “Univalued Segment Assimilating Nucleus”. [Starostenko, 2001] [Quevedo, 2000]

Los valores de los píxeles que se encuentran por debajo de la máscara, es decir en el área USAN se utiliza para determinar la presencia de bordes o esquinas y pueden dar tres posibles resultados:

1. Un máximo cuando el núcleo está en una región de la imagen donde todos los píxeles dentro de la imagen tienen el mismo brillo. Ver Figura 2.2.
2. En medio cuando el núcleo está muy cerca de un borde. Ver Figura 2.3.
3. Un mínimo cuando el núcleo está sobre una esquina. Ver figura 2.4

Dado que el método SUSAN no usa las derivadas de la imagen para la búsqueda de bordes y esquinas, se dice que es un buen algoritmo para la presencia de ruido en las imágenes. Además, SUSAN analiza diferentes regiones por separado y detecta la presencia de las fronteras por medio del valor mínimo de USAN.

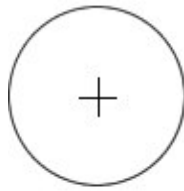


Figura 2.2 Valor máximo de un punto

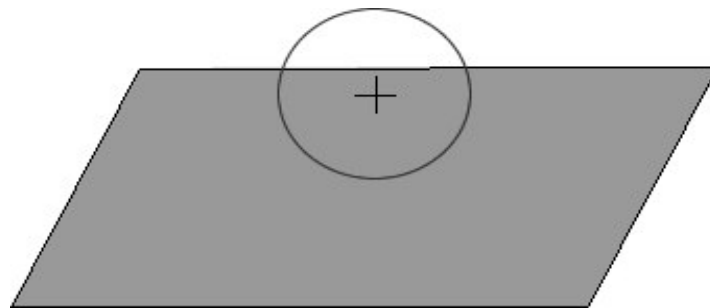


Figura 2.3 Valor medio de un punto



Figura 2.4 Valor mínimo de un punto

En [Starostenko, 2001] y [Smith, 1995] se presenta el procedimiento que sigue SUSAN para la detección de esquinas principales el cual es el siguiente:

1. Barrer la máscara circular sobre el núcleo o píxel de interés. Existen varias formas de calcular el valor de la máscara, sin embargo, el método más exacto es cuando se toma la función de distribución Gausseana que es:

$$g[i, j] = K * e^{-\frac{(i^2 + j^2)}{2\sigma^2}}$$

Donde K corresponde a una constante de normalización, σ es el parámetro de dispersión, i y j son dos dimensiones de la función.

2. Posteriormente se calcula el valor de similitud con el núcleo de cada píxel bajo la máscara por medio de la siguiente función:

$$C(r, r_0) = \exp\left[-\left[\frac{I(r) - I(r_0)}{t}\right]^6\right]$$

Donde r_0 es la posición del núcleo, r es la posición de cualquier otro punto bajo la máscara, $I(r)$ es el brillo del píxel y t es el contraste mínimo entre detalles que van a detectarse y el ruido a eliminar. Por lo que el número de esquinas detectadas depende directamente del valor de t .

3. Una vez obtenidos los valores bajo la máscara, se suman y el resultado se asigna a la posición del píxel central el cual fue estimado de la siguiente forma:

$$n(r_0) = \sum_r C(r, r_0)$$

4. Utilizando la ecuación:

$$R(r_0) = \begin{cases} g - n(r_0) \Leftrightarrow n(r_0) < g \\ 0 \Leftrightarrow n(r_0) \geq g \end{cases}$$

Donde g es el umbral establecido el cual es igual a $\frac{n_{\max}}{2}$, donde n_{\max} es el número máximo de todos los valores obtenidos por la máscara. El umbral g , sin embargo, puede tener otros valores que influyen sobre la calidad de la salida así como del número de esquinas encontradas.

5. Finalmente, se buscan todos los puntos falsos dentro del conjunto de esquinas encontradas. Estas esquinas falsas pueden ser producto de bordes entre esquinas borrosas. Este problema de esquinas falsas puede ser resuelto calculando el centro de gravedad del área USAN, la mediana de las coordenadas de x y y . Con éste valor se calcula la distancia entre el centro de gravedad y el núcleo de la máscara, si esta distancia es grande, se dice que estamos en presencia de esquinas falsas por lo que se eliminan del conjunto de esquinas encontrado.

2.4 Conclusiones.

Es muy importante tener conocimiento del funcionamiento de los métodos más modernos y a su vez más utilizados en el área de detección de movimiento para así poder atacar al problema de una forma más efectiva.

Todos los métodos mencionados anteriormente son solo algunos de los más utilizados en la actualidad para la detección de movimiento dentro de una secuencia de imágenes, sin embargo, como ya se mencionó en este capítulo, todos tienen ventajas y desventajas.

Comparando la capacidad, eficiencia y rapidez de los métodos basados en gradiente con los métodos basados en correspondencia y los de detección de esquinas principales pude darme cuenta que para la generación del campo de movimiento, son más eficientes los métodos de detección de esquinas principales sobre los otros dos.