

Capítulo II.

Modelos de aproximación.

Existe la gran variedad de técnicas de aproximación para estimar el movimiento. Cada una de ellas tiene su enfoque definido e ideas importantes. Entre todas las aproximaciones existentes se destacan dos grupos principales: basados en gradiente y correspondencias. Los otros grupos se basan en las fases aplicando diferentes filtros y en redes neuronales artificiales. Normalmente todos estos grupos se dedican a construir el campo de flujo óptico. Pero el grupo basado en gradiente y de los filtros se puede usar para encontrar los bordes de objetos en movimiento de la escena.

En este capítulo describiré los métodos de los grupos mencionados anotando sus ventajas y desventajas en general.

2.1. Algoritmos basados en gradiente.

Los algoritmos basados en gradiente proveen la solución al problema de estimación de movimiento desde la observación en el tiempo de los cambios en el brillo de la secuencia de imágenes. Estos cambios son modelados por ecuaciones diferenciales parciales llamadas ecuaciones de limitadores. El campo de los vectores de velocidad obtenido al resolver las ecuaciones se llama el flujo óptico o flujo de imagen [Laplante, 1996].

La ecuación más importante para modelar el campo de flujo óptico se obtiene al considerar los cambios en el brillo de la imagen $E(x, y, t)$:

$$\frac{dE}{dt} = \frac{\partial E}{\partial x} \frac{dx}{dt} + \frac{\partial E}{\partial y} \frac{dy}{dt} + \frac{\partial E}{\partial t} = 0 \quad (2.1)$$

En la forma abreviada es: $E_x u + E_y v + E_t = 0$, donde $u = dx/dt$ y $v = dy/dt$ que representan los componentes del vector por x y y . También la ecuación puede ser representada en forma de línea: $v = mu + c$, donde $m = -(E_x/E_y)$ y $c = -(E_t/E_y)$. Cualquier punto en esta línea puede ser la solución posible. Entonces una ecuación no es suficiente para determinar la solución única. Otros limitadores como los bordes y suavizado se aplican para obtener la solución computacional. El problema de apertura aquí también se relaciona con el flujo óptico. De hecho al considerar la ecuación del flujo óptico, se observa que sólo el componente perpendicular V_{\perp} de la velocidad aparente, que es paralelo al gradiente ∇E puede ser percibido y estimado:

$$V_{\perp} = - \frac{E_t}{\|\nabla E\|} \frac{\nabla E}{\|\nabla E\|}, \text{ donde } \|\nabla E\| \neq 0 \quad (2.2)$$

El problema de apertura no estará presente en el caso si es posible determinar al menos dos ecuaciones distintas como en el caso de las esquinas.

La versión extendida de la ecuación de flujo óptico es:

$$E_x u + E_y v + E(u_x + v_y) + E_t = 0 \quad (2.3)$$

que incluye la divergencia del campo de velocidad $(u_x + v_y)$. Al analizar estas dos ecuaciones se obtiene que la primera es mejor en la presencia de calibración en la mayoría de las condiciones de movimiento.

En la estimación del campo de movimiento por medio de gradiente existen diferentes problemas como el ruido, oclusión, discontinuidades [Laplante, 1996]. El primer problema se relaciona con el de apertura por la ecuación que puede tener varias soluciones. En el segundo caso el problema se refiere a la aplicabilidad de las ecuaciones, condiciones

bajo las cuáles el campo puede ser modelado. El tercer problema se debe a la sensibilidad del método a las discontinuidades en el brillo de la imagen.

Existen dos tipos principales de algoritmos basados en gradiente: de regularidades y multilimitadores.

2.1.1. Algoritmos basados en regularidades.

En estos algoritmos se usa el suavizado para la regularización de solución de la ecuación diferencial parcial. La función es minimizada por los cálculos de variaciones o el relajamiento estocástico. La aproximación más famosa propuesta por Horn y Shunck se basa en la minimización de función [Laplante, 1996]:

$$\iint [(E_x u + E_y v + E_t)^2 + \alpha^2 (u_x^2 + u_y^2 + v_x^2 + v_y^2)] dx dy \quad (2.4)$$

donde el primer término es la ecuación de flujo óptico y el segundo es el suavizado. α es el factor de peso que controla la influencia de suavizado. La minimización se realiza por medio de los cálculos de variaciones que se basan en el sistema de dos ecuaciones diferenciales:

$$\begin{aligned} V^2 u &= \frac{E_x}{\alpha^2} (E_x u + E_y v + E_t) \\ V^2 v &= \frac{E_y}{\alpha^2} (E_x u + E_y v + E_t) \end{aligned} \quad (2.5)$$

La aproximación de Horn y Schunck, como las otras basadas en la regularización, tiende a producir los campos de flujo óptico suavizado para la propagación de valores de vectores desde la estimación de puntos. La profundidad de propagación depende del factor de peso y el número de iteraciones en calcular los Laplacianos. Aunque el proceso de estimación converge a un mínimo, no se da la solución óptima, porque el error más bajo en la estimación del flujo óptico no corresponde a la iteración en la cuál el proceso obtuvo el mínimo.

La otra aproximación que entra a esta categoría es de Fennema [Fennema, 1979] llamado "Determinación de velocidad en las escenas con varios objetos en movimiento". Se basa en la transformación de intensidad de gradiente. Se aplica para las translaciones rígidas sin rotación. La variación de intensidad con el tiempo y el gradiente espacial aplicado a cada punto dan la velocidad y dirección del movimiento de objeto en este punto. Estas características no son suficientes para determinar el vector único de desplazamiento, por eso se aplica la técnica de agrupamiento por la velocidad común. Para mejorar el resultado se aplica el filtro de paso bajo (efecto de suavizado) a la imagen, que reduce el ruido y elimina algunas discontinuidades. La técnica trabaja muy bien con las escenas que contienen diferentes objetos en movimiento lateral de diversas direcciones.

2.1.2. Algoritmos basados en multilimitadores.

En estos algoritmos se definen las diferentes ecuaciones que determinan el sistema que se resuelve por medio de las técnicas de cuadrado más bajo o probabilísticamente [Laplante, 1996]. Estos algoritmos pueden ser clasificados en base a la aproximación adoptada para construir el sistema de ecuaciones y a las técnicas matemáticas para la solución del sistema. Existen tres aproximaciones principales. En la primera se supone que $dE/dt = 0$ en el caso estacionario y se definen varias ecuaciones diferenciales parciales de flujo óptico. En la segunda se obtienen nuevas ecuaciones al tomar las derivadas de las ecuaciones fundamentales del flujo óptico con respecto a los ejes x , y , t . En la tercera se asume que el cambio de flujo óptico es lineal. Esto es válido sólo en caso de que el campo es suavizado. Se definen las ecuaciones de vecinos del cuadrado $N*N$ de un punto. Un número grande de N va a conducir a la pérdida de resolución.

Se puede anotar que la calidad de estimación de flujo óptico depende de las ecuaciones escogidas. Por ejemplo, los métodos basados en las segundas derivadas del brillo de imagen son muy sensibles al ruido. Es bueno tener las soluciones de suavizado para usar el flujo óptico en la segmentación de objetos en movimiento, seguimiento de objetos y estimación de movimiento. Pero por otro lado la filtración de la imagen puede destruir información importante en la reconstrucción de objetos 3D y en la estimación de

movimiento 3D. El funcionamiento de estos algoritmos depende del número de ecuaciones y de la aproximación adoptada para resolver el sistema. Las aproximaciones pueden dividirse en dos clases: algebraica, donde el sistema de ecuaciones se resuelve con los métodos numéricos y de agrupamiento, donde cada solución posible se considera como el punto en el dominio multidimensional. Estos puntos son agrupados en orden de seleccionar la solución más probable.

2.1.2.1. Soluciones algebraicas.

Las soluciones algebraicas pueden ser vistas como el proceso de cuatro fases [Laplante, 1996]:

- Convolución de imágenes con el filtro de Gauss.
- Estimación de derivadas del brillo de la imagen.
- Estimación del sistema de ecuaciones determinada utilizando el método pseudo-inverso.
- Estimación de los componentes de velocidad al resolver el sistema.

Dos aproximaciones que entran a esta clase son de Liu [Liu, 1998] "Extracción de los bordes en movimiento e implementación en tiempo real". Y la segunda que entra es de Boutheymy [Boutheymy, 1989] "Algoritmo de la probabilidad máxima en el marco para determinar los bordes en movimiento".

Extracción de los bordes en movimiento e implementación en tiempo real.

La extracción de los bordes en movimiento y la computación del flujo óptico son dos subproblemas que no pueden ser resueltos independiente uno de otro. Existen varios métodos basados en esquema iterativo que resuelven este problema. Esta aproximación usa el método no iterativo que extrae simultáneamente los bordes en movimiento y calcula el flujo óptico. Se basa en la descomposición de secuencia de imágenes por medio de los polinomios hermitos 3D representando la proyección en perspectiva del movimiento de 3D. La ventaja principal de este algoritmo es su eficiencia en ejecución.

Hablando en general, el problema de detección de movimiento incluye dos subproblemas: computación de flujo óptico y segmentación de movimiento. El flujo óptico mide cuantitativamente el movimiento asociado con los objetos percibidos. La segmentación de movimiento distingue cualitativamente diferentes objetos.

La idea básica de esta aproximación es hallar la correspondencia entre propiedades locales de la imagen y el modelo de movimiento 3D. Elementos necesarios del esquema son el modelo de movimiento que se basa en el movimiento 3D arbitrario, el método para estimar las propiedades de la imagen para lo cual se usa la descomposición de imagen basada en polinomios hermitos, y el procedimiento para extraer los detalles del modelo como los bordes en movimiento. El uso del polinomio hermito es eficiente por su característica de separación.

Algoritmo de la probabilidad máxima en el marco para determinar los bordes en movimiento.

Este algoritmo se concentra en la determinación de los bordes en movimiento en la secuencia de imágenes. Más precisamente, el método simultáneamente pertenece a un procesamiento local de lo siguiente: posición, dirección espacial y componente de velocidad que es perpendicular al borde de la imagen.

Para determinar los bordes en movimiento el algoritmo usa tres etapas básicas: modelación, diseño de hipótesis competitivas y definición del coeficiente de probabilidad. La primera etapa consiste en determinar las áreas de la superficie donde se encuentra el borde en movimiento. El interés principal de este método es cubrir dos problemas: contraste de intensidad y aspecto de movimiento desde los puntos de vista analítico y estructural. El paso de modelado fue motivado por tres factores principales. Primero, el proceso tiene que ser local para las mediciones de densidad y la determinación del amplio rango de los campos de movimiento. El segundo, los cálculos computacionales tienen que realizarse en un conjunto de intensidades para poder implementar el algoritmo en el tiempo real. El tercero, la estrategia de bordes tiene que ser integrado en el proceso de algún modo.

En la segunda etapa se definen dos hipótesis. Si la estimación de movimiento está modelada por un área S que se define por el vector de parámetros Φ entonces hay un volumen π en 3D (x, y, t) que puede incluir al menos un borde en movimiento o incluir nada. Respectivamente las hipótesis son:

H_0 : No existe la estimación de movimiento. Entonces la distribución de intensidad dentro de π es modelada por un constante C_0 más el ruido, según la función de distribución de Gauss con la varianza σ^2 .

H_1 : Existe una estimación del movimiento. Entonces el área $S(\Phi)$ divide π en dos π_1 y π_2 y la intensidad se modela otra vez según los constantes C_1 dentro de π_1 y C_2 dentro de π_2 , donde $C_1 \neq C_2$.

El problema ahora es seleccionar una de las dos hipótesis. Para esto se implementa la tercera etapa. Se usan las funciones de probabilidad asociadas con cada hipótesis. Estas funciones se construyen por las probabilidades de la función de intensidades dentro de π . Según el coeficiente que se obtiene después de las funciones se escoge una u otra hipótesis.

Como el resultado el algoritmo da el flujo óptico de contornos de objetos. Para estimar los detalles espacio temporales con este algoritmo es suficiente tener dos imágenes de secuencia. Pero a veces esta determinación puede ser mejorada tomando en consideración más de dos imágenes. El algoritmo es sólo una parte de la estimación completa de movimiento que consiste en tres etapas: procesamiento local, estructura intermedia, estimación del campo de flujo óptico. En la primera etapa se determinan los bordes espacio temporales y/o la segmentación en regiones según los criterios hechos jerárquicamente, que toman en cuenta la información de movimiento parcial. En la segunda etapa se obtiene la estructura intermedia de la imagen como el resultado de juntar los bordes y/o regiones. En la etapa final se reconstruyen los vectores de desplazamiento combinando las observaciones locales. Este algoritmo da las discontinuidades de velocidad en la imagen y produce un conjunto de primitivas espacio temporales del nivel intermedio.

2.1.2.2. Soluciones basadas en agrupamiento.

En las soluciones basadas en agrupamiento [Laplante, 1996] el punto principal es resolver el problema de mínimos cuadrados de las soluciones, donde la misma se localiza en el punto ubicado en la distancia mínima entre los puntos de intersección en el conjunto de líneas limitadoras. Al utilizar este método de mínimos cuadrados se aplica el otro de agrupamiento que mejora la solución final.

Una de las aproximaciones propuestas por Fennema y Thompson [Fennema, 1979] se basa en la transformada de Hough y limitadores del flujo óptico. Se toma en cuenta que la transformada de Hough de línea en el punto (u, v) corresponde al punto en el dominio transformado. Las líneas que tienen la intersección común en el plano producen los puntos que se distribuyen por las curvas bien definidas. Esta aproximación es computacionalmente pesada debido a las verificaciones de la gran cantidad de las curvas generadas.

La otra aproximación que fue propuesta por Schunck [Laplante, 1996] obtiene la solución por el agrupamiento de las intersecciones de varias líneas limitadoras con una línea de pixel central entre sus vecinos. El agrupamiento se hace alrededor de la línea central, desde entonces el grupo dominante de las intersecciones identifica la solución al problema de flujo. Pero el punto débil que puede presentar esta aproximación es la precisión de los resultados en las imágenes ruidosas.

El otro algoritmo más reciente fue propuesto por Frigui [Frigui, 1999] que se llama "El algoritmo robusto competitivo del agrupamiento (RCA)". Este algoritmo combina el agrupamiento jerárquico y de particiones. En el primero no se considera la prioridad de los grupos, se toman sólo los vecinos locales en cada paso sin considerar los conocimientos globales de los grupos. El segundo algoritmo de particiones es dinámico, donde los puntos pueden moverse de un grupo al otro. Incorporan los conocimientos de los grupos. El algoritmo RCA determina el número óptimo de los grupos y considera los conocimientos de la forma global de los grupos que se obtienen usando los prototipos. Para detectar los detalles se usan los conceptos de estadística robusta. Los grupos sobrepuestos se regularizan por el algoritmo de lógica difusa.

En el primer paso el algoritmo divide los datos en un número grande de grupos pequeños que reduce su sensibilidad al inicio. Conforme el algoritmo avanza los grupos

adyacentes compiten por los puntos y los que pierden, desaparecen gradualmente. RCA usa dos diferentes conjuntos de pesos para cada punto. El primero es de limitadores probabilísticos que representa los grados de compartimento entre grupos. El segundo es de posibilidades de los grados de puntos con respecto a los grupos. Estos dos conjuntos se usan para obtener la estimación robusta de los prototipos de grupos.

2.1.3. El método basado en gradiente para detectar los bordes de objetos en movimiento.

A parte de construir el campo de flujo óptico también se puede detectar los bordes de objetos en movimiento utilizando el método de multiplicación del gradiente espacial y temporal [Jain, 1995].

En una escena pueden estar los objetos móviles y estáticos y los segundos podrían ser un obstáculo para detectar el movimiento de objetos dinámicos. Por esa razón se propone el algoritmo de detección de los bordes en movimiento usando la combinación de dos gradientes: temporal y espacial haciendo el operador lógico AND entre ellos:

$$E_t(x, y, t) = dF(x, y, t)/dS * dF(x, y, t)/dt = E(x, y, t) * D(x, y) \quad (2.6)$$

donde $dF(x,y,t)/dS$ y $dF(x,y,t)/dt$ son dos gradientes espacial y temporal respectivamente de la función de intensidad F . Para calcular la primera parte diferentes detectores de bordes de objetos estáticos pueden ser usados y para la segunda parte es la diferencia entre valores de gris en primera y segunda escenas de la secuencia.

Después de obtener la secuencia de imágenes con los bordes de objetos móviles se puede detectar la dirección de desplazamiento y la velocidad, teniendo un intervalo de tiempo entre las imágenes. Para detectar la dirección de desplazamiento del objeto cada marco de la secuencia de imágenes con bordes de objetos móviles se divide en regiones y después se calcula el desplazamiento de la región que pertenece al objeto en movimiento.

Este método basado en el cálculo de gradiente es bueno para los movimientos de la velocidad que puede detectar el ojo humano. El tamaño de desplazamiento de los objetos entre las escenas no tiene gran importancia para este algoritmo.

Existen diferentes soluciones basadas en gradiente para estimar el campo de flujo óptico que usan ecuaciones de complejidad diferente. Con el tiempo han aparecido nuevas aproximaciones cuya complejidad es más grande debido al uso de análisis numérico más sofisticado. Según Laplante [Laplante, 1996] las que se basan en multilimitadores dan la estimación del campo de movimiento más robusta, pero hay unas aproximaciones que se usan más frecuente debido a su simplicidad de aplicación y la posibilidad de remover el ruido usando los filtros. Y las de multilimitadores no requieren alguna filtración pero son computacionalmente pesados.

2.2. Algoritmos de correspondencias.

Se basan en la técnica de correspondencias de detalles destacados del objeto en movimiento [Laplante, 1996]. Al pasar estos detalles por el marco interno se buscan los que son parecidos estimando así el movimiento. Los detalles se clasifican en nivel alto (líneas, formas) y nivel bajo (esquinas, patrones, cambio de colores, curvaturas). Los resultados obtenidos con detalles de nivel alto son mejores que de otro tipo pero se requiere el preprocesamiento de imágenes (filtración, extracción de bordes) con el costo computacional correspondiente. Por otro lado las técnicas de correspondencia no deben depender de rotación y cambio de escala. En este caso es mejor usar los detalles de nivel bajo. Entre estas técnicas la más famosa es correspondencia de bloques que se basa en el brillo de imagen. Este algoritmo es afectado por el error acumulativo.

2.2.1. Técnica de correspondencia de bloques.

La técnica de correspondencia de bloques [Laplante, 1996] toma los segmentos de imagen como detalles para buscar sus correspondientes. En este caso la estimación del campo de movimiento consiste en encontrar las coordenadas (x', y') del centro de

segmento en el tiempo t que minimizan la correspondencia con el segmento en el tiempo $(t-\Delta t)$ con coordenadas (x, y) . La fórmula para obtener la estimación es:

$$C(x, y, x', y') = \sum_{m=-M/2}^{M/2} \sum_{l=-L/2}^{L/2} (E(x+m, y+l, t-\Delta t) - E(x'+m, y'+l, t))^2 \quad (2.7)$$

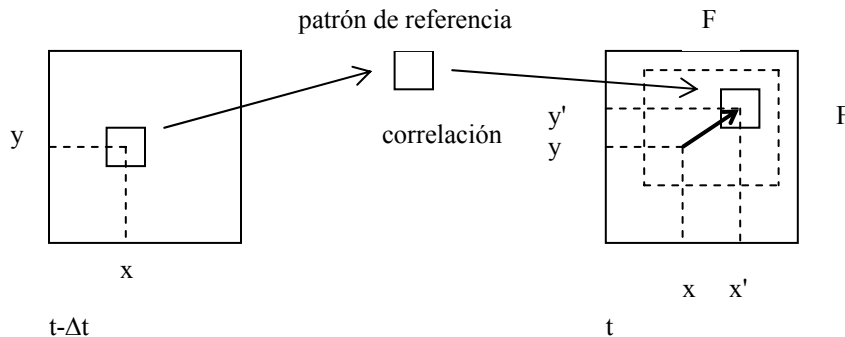
donde $E(x, y, t)$ es el valor de brillo de la imagen en el tiempo t , M y L son dos dimensiones del patrón por eje x y y respectivamente. El proceso de minimización consiste en encontrar el valor mínimo de la función $C(x, y, x', y')$ cerca del punto (x, y) para

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

donde F es la dimensión del segmento de imagen en el cual se busca el valor mínimo. Las coordenadas (x', y') identifican la posición del segmento en el tiempo t . Las diferencias

$$S_x = x' - x \quad y \quad S_y = y' - y$$

son los componentes de desplazamiento del segmento, o sea, su velocidad.



Con esta técnica el patrón del segmento de imagen se usa como el patrón de referencia para buscar el desplazamiento en el tiempo t . El mismo patrón se usa después en las siguientes imágenes, lo que ayuda a resolver el problema de crecimiento progresivo de patrones y evitar los desplazamientos incorrectos. Usando el patrón de referencia se puede calcular el error acumulativo que tiende a crecer con el tiempo. El otro problema es que los componentes de desplazamiento se calculan usando números enteros, entonces si el desplazamiento es de un pixel, el error es muy alto. Para reducir el crecimiento de error se usan las esquinas y bordes de la imagen.

La técnica de correspondencia de bloques es sensible al ruido, pero este problema puede ser resuelto al aumentar el tamaño del patrón de referencia. Otros mecanismos para mejorar la robustez son [Laplante, 1996]:

- Evitar el seguimiento de los patrones que no caracterizan suficientemente a la imagen porque la probabilidad que el dicho patrón sea más bien el ruido es muy alta.
- No tomar en cuenta los desplazamientos que salen de los límites de umbralización predefinidos.

La técnica de correspondencia de bloques se divide en:

- La búsqueda completa del bloque [Lagendijk, 1999].

El patrón de referencia se busca pasando por cada pixel del área. Es computacionalmente compleja, sensible al ruido.

- La búsqueda reducida en complejidad.
- La búsqueda logarítmica de 2D.

Fue propuesta por Jain en 1981 [Cheung, 1998]. La búsqueda usa el cruzado griego (+) en cada paso. El algoritmo se divide en tres etapas [Manning]. En la primera etapa se determina el paso s por la fórmula $2^{\lfloor \log_2 d \rfloor - 1}$ donde d es el desplazamiento posible. Después entre los 5 bloques $[0,0][0,+s][0,-s][-s,0][+s,0]$ se busca el que más parece al patrón con qué se compara. En la etapa dos se checa si el bloque parecido está en el centro del área de búsqueda entonces el paso s se reduce en mitad, si no s se queda el mismo y el centro de búsqueda se mueve al nuevo. En la etapa tres se checa si s es igual a un pixel entonces se examinan todos los 8 bloques cercanos y se escoge el mejor. Si s es mayor que uno entonces se asignan los nuevos bloques a comparar y el procedimiento se repite desde la etapa 2.

- La búsqueda de tres pasos.

Fue introducida por Koga en 1981 [Cheung, 1998]. Se basa en el principio de refinamiento de búsqueda. El paso inicial se toma como $\lceil d/2 \rceil$, donde d es el desplazamiento máximo posible. A partir del centro $[cx,cy]$ se examinan los bloques $[cx-s,cy-s]$, $[cx-s,cy]$, $[cx-s,cy+s]$, $[cx, cy-s]$, $[cx,cy]$, $[cx,cy+s]$, $[cx+s,cy-s]$, $[cx+s,cy]$, $[cx+s,cy+s]$. Se escoge el más parecido al patrón que después será el centro.

Después de cada etapa se reduce el paso en mitad hasta llegar a s igual a un pixel y de allí se determina el bloque más parecido [Manning].

- La búsqueda ortogonal.

Este algoritmo fue propuesto por Puri en 1987 [Cheung, 1998]. Es un híbrido de dos búsquedas [Manning]: de tres pasos y 2D logarítmica. La idea principal consiste en búsqueda entre par horizontal y vertical donde el paso s se reduce en mitad empezando con $s = \lfloor d/2 \rfloor$. Se examinan tres puntos horizontales con el centro en medio, después con el mismo paso s - los tres puntos verticales donde el centro es el bloque más parecido. En siguiente etapa el paso s se reduce en mitad y el procedimiento se repite hasta llegar a un pixel.

- La búsqueda una a la vez.

Consiste principalmente en dos etapas [Manning]: horizontal y vertical. El bloque se empieza a buscar en el centro del área de búsqueda por línea horizontal. Una vez encontrado el bloque más parecido la búsqueda entra a segunda etapa donde el desplazamiento va por vertical encontrando al final el bloque más parecido al patrón.

- La búsqueda cruzada.

Fue propuesta por Ghanbari en 1990 [Cheung, 1998]. Los cuatro puntos en donde se busca son de letra X. Se toma el paso $s = \lfloor d/2 \rfloor$ al principio y después s se divide en mitad en cada etapa escogiendo el centro que más parece al bloque buscado. La búsqueda termina cuando s será igual a un pixel.

- La búsqueda basada en bloque de gradiente descendente.

Fue propuesta por L.K.Liu y E.Feig en 1996 [Cheung, 1998]. Este algoritmo usa 9 puntos en cada paso con s igual a un pixel para checar la correlación. El bloque más parecido se escoge como el centro. La condición de paro es ya se encontró el bloque deseado o el algoritmo llegó al borde. Como s igual a un pixel hay unos puntos que se superponen de otros pasos. Este método es bueno para los movimientos pequeños.

Todos estos algoritmos descritos arriba en cada paso determinan la dirección y el centro de la búsqueda. Pero existen algunos casos donde no se puede determinar la dirección correcta:

- No hay patrones distinguidos (bloques planos).

- Presencia de los bordes rectos (bloques de líneas).

En estos dos casos el verdadero vector de movimiento es similar a otros vectores que lo rodean, la estimación de movimiento resulta incoherente. Cheung [Cheung, 1998] propone mejorar el método de correspondencia de bloques agregando el vector de movimiento coherente que se obtiene en tomar la mediana entre los vectores de movimiento que rodean el vector estimado.

Siguiendo con los tipos de algoritmos de correspondencia de bloques:

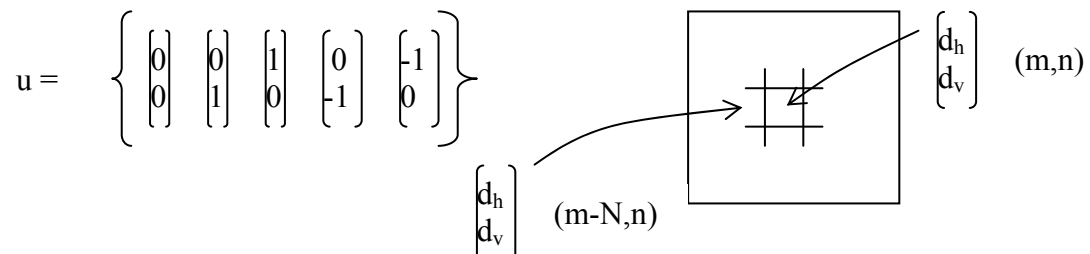
- Correspondencia de bloques en jerarquía [Lagendijk, 1999].

Empieza con la estimación en nivel bajo de resolución. Después el desplazamiento encontrado se pasa a nivel más alto de resolución para su refinamiento.

- Correspondencia de bloques recursiva [Lagendijk, 1999].

Desplazamiento del bloque es dado por el desplazamiento de sus vecinos mas la actualización.

$$\begin{bmatrix} d_h \\ d_v \end{bmatrix} (m,n) = \begin{bmatrix} d_h \\ d_v \end{bmatrix} (m-N,n) + u$$



(2.8)

- Algoritmos de reproducción para la correspondencia de bloques [Manning].

Estos algoritmos reducen el número de correspondencias por desplazar el patrón inmediatamente al encontrar el vector con distorsión menor que del centro actual. El refinamiento de los vectores de movimiento procede en la dirección de distorsión mínima. El tamaño de paso s se reduce al encontrar la distorsión mínima. Los algoritmos de reproducción buscan la mejor posición en diferencia de los algoritmos reducidos en complejidad que solamente tratan de mejorar la posición.

El método de correspondencia generalmente usa pocos marcos de la secuencia. Puntos críticos, líneas u otros detalles de objetos en primer imagen de la secuencia se ponen en correspondencia con los detalles similares en siguientes imágenes. El desplazamiento y el intervalo de tiempo entre una y otra imagen determinan el flujo óptico de objetos procesados. El proceso de correspondencia puede sufrir de explosión combinatoria en sistemas de tiempo real. Para resolver este problema se usan dos técnicas [Laplante, 1996]:

- Reducción de número de detalles críticos dependiendo de la aplicación.
- Uso de los métodos de multiresolución como jerárquico para estimar el flujo óptico.

Normalmente los algoritmos de correspondencia usan una estrategia de minimización.

Muchos métodos de correspondencia usan la técnica de jerarquía y la filtración previa de paso de banda o de paso bajo para disminuir las posibilidades de correspondencias falsas y después hacer la pirámide de refinamiento de soluciones para la mejor correlación del brillo de imágenes. Esta técnica de filtración con la búsqueda jerárquica puede caer en el problema de apertura. Se puede construir el flujo óptico en las esquinas donde no está presente este problema pero en las regiones que contienen las líneas rectas la determinación de flujo se vuelve más imprecisa.

2.2.2. Estimación robusta del movimiento para las cámaras calibradas para la secuencia de imágenes monoculares.

Existen diferentes métodos basados en correspondencia que son más avanzados en la estimación de flujo óptico y incluyen el movimiento transitorio y circular. Uno de ellos está descrito en el artículo de Robert Wagner [Wagner, 1999]. Su idea principal es la estimación de movimiento de la cámara basada en el conjunto de puntos de correspondencia tomados de la secuencia de imágenes monoculares. El método se basa en la descomposición de conjunto de parámetros en subconjuntos que son optimizados por separado y en la detección de diferentes movimientos. El modelo de error cónico permite distinguir entre correspondencias factibles y no relacionados con los puntos de objetos de 3D. La estabilidad y precisión de los resultados obtenidos son bastante sensibles a la

exactitud del modelo de la cámara, los parámetros de la cámara, los resultados de la estimación de movimiento y las correspondencias obtenidas de la imagen. La estimación de subconjuntos en vez de uno grande evita los errores de compensación relacionados con distintos parámetros. En general la calidad y precisión de estimación de movimiento depende del tratamiento correcto de todos errores posibles. Los errores producidos por distorsiones aleatoria y sistemática como el ruido, por distorsiones de proyección, por correspondencias escasas relacionadas con los puntos de objeto 3D en diferentes movimientos tienen que tomarse en cuenta. La estimación de error usando el cono es buena solución para este problema. En la calibración de cámara se usa el mapeo

$\Pi: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ entre los puntos dados en coordenadas de 3D y sus posiciones respectivas en el plano de la imagen. Π puede ser interpretado como la composición $\Pi_3 \circ \Pi_2 \circ \Pi_1$ de transformaciones entre dos sistemas de coordenadas euclidianas de 3D $\Pi_1: \mathbb{R}^3 \rightarrow \mathbb{R}^3$, proyección $\Pi_2: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ y el mapeo $\Pi_3: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ que convierte las posiciones de pixeles de la cámara a las coordenadas de pixeles en el sistema físico. Los experimentos de este método muestran la alta precisión y estabilidad, su uso práctico se aplica muy bien a las aplicaciones automovilísticas por su capacidad de computarizar el movimiento de objetos en alta velocidad con buena precisión.

En esta área los algoritmos que se usan más son de correspondencia de bloques. Unos de ellos escogen mejor las correspondencias que otros y entonces el resultado es más preciso. Para encontrar los bloques parecidos unos evalúan más candidatos, otros menos, entonces la complejidad se varía dependiendo de aproximación. Debido al uso relativamente fácil de estos algoritmos, su aplicación es muy extensa en diferentes áreas. Pero si se trata del procesamiento del nivel bajo los algoritmos de correspondencia pueden caer en los problemas como de apertura.

2.3. Métodos de filtración espacio temporal.

Son los métodos más recientes que consideran la secuencia de imágenes de 2D como las imágenes espacio temporales de 3D. Después de aplicar los filtros de tres dimensiones al volumen de la imagen en espacio tiempo. Si los filtros son orientados, localizados y sensibles al rango de frecuencias espacio temporales entonces es posible realizar la detección robusta de movimiento. Existen dos estrategias básicas para usar el volumen 3D espacio temporal [Laplante,1996]:

- Detección de movimiento en dos dimensiones espaciales se convierte al problema de detección de orientación del plano en volumen 3D espacio temporal.
- Detección de movimiento de objetos en 2D se convierte al problema de encontrar texturas orientadas en volumen 3D espacio temporal.

El propósito principal en usar estas estrategias es evitar el problema de apertura en los bordes y procesar las imágenes en el nivel bajo de operaciones con píxeles.

2.3.1. Método basado en fases.

Uno de los métodos que se propone es de Fleet [Fleet, 1992] caracterizado por dos pasos principales:

- Aplicar la colección de filtros de velocidad sintonizados.
- Para cada respuesta de filtro:
 - medir los componentes de velocidad de imagen;
 - remover las mediciones irrelevantes.

En el primer paso se usan los filtros de Gabor que proveen la descomposición de un paso de banda espacio temporal de 0.8 octavos medido en $\sigma_k = 1/\sigma$, donde σ es la desviación estándar que determina la frecuencia de ancho de banda. Es importante tomar este ancho de banda pequeño porque se reduce la sensibilidad a las iluminaciones significativas y frecuencias bajas. Después a cada salida de los filtros se aplica el gradiente de fase para obtener las estimaciones de componentes de velocidad.

Existen varias propiedades que se aplican a este método [Fleet, 1992]:

- El uso de la información de fase es independiente de la imagen con las características específicas. Entonces los problemas como detección, localización, descripción y correspondencia están resueltos.
- La estructura de la imagen es separada según la velocidad y escala, entonces la estimación de velocidades múltiples está permitida con los vecinos locales. Esto puede ayudar en los problemas de transparencia, reflexión y oclusión parcial.
- El método es eficiente y cómodo para el procesamiento paralelo. Cada canal de velocidad puede trabajar independientemente y varias operaciones son locales, insensibles al desplazamiento y separables en espacio tiempo.

Con respecto a los filtros iniciales hay las siguientes propiedades:

- La técnica no es limitada al uso de filtro Gabor. El uso de la información de fase no depende estrictamente de los filtros de velocidad sintonizados.
- Los filtros tienen que remover las iluminaciones significativas y atenuar las frecuencias bajas para evitar el efecto de sobreposición en la representación de submuestras.
- La representación inicial basada en las salidas de los filtros es eficiente porque es submuestreada en el coeficiente razonable y cuantificada.

El problema que puede surgir de las fases es su estabilidad con respecto a la perturbación de escala en la entrada. Fleet [Fleet, 1992] mostró que la fase es estable bajo las perturbaciones de escala pequeñas y es cuasi-lineal como la función de la posición espacial. Estas dos características dependen de la correlación que existe entre los filtros sintonizados a las escalas de localizaciones espaciales.

El método de Fleet [Fleet, 1992] puede ser aplicado a las mediciones de disparidad binocular. Si la configuración geométrica de dos cámaras se conoce, entonces la disparidad del punto de la escena (distancia relativa entre proyecciones de un punto singular de la escena en dos superficies de imagen) codifica la profundidad 3D de un punto. Se asume que las disparidades cruzadas correspondientes a los detalles de la escena que están más cerca que el punto fijo, son negativas. Las disparidades no cruzadas correspondientes a los detalles de la escena que están más lejos del punto fijo, son positivas. La computación de disparidades del punto de imagen implica la correspondencia de detalles de imagen de un punto de vista con el otro para obtener la aproximación a disparidad de un punto de la escena. En el algoritmo de Fleet se hace la correspondencia de las fases por medio de

determinación de desplazamiento requerido para igualar las fases de señales de izquierda y derecha.

El algoritmo de Fleet fue analizado por el autor en las entradas libres del ruido. El siguiente análisis de estabilidad y singularidad de fase se requiere con los modelos más realistas. También la mayor parte de la atención tiene que ser prestada a la precisión de mediciones de velocidad por distorsiones fotométricas y velocidades múltiples locales.

2.3.2. Método basado en algoritmo SUSAN.

El otro tipo de filtro se podría usar para detectar las esquinas de los objetos de la escena. Después al poner una imagen sobre la otra de secuencia con sus esquinas correspondientes se podría ver si hubo el movimiento de objetos o no. El filtro se llama SUSAN que fue propuesto por Smith [Smith, 1997].

SUSAN se traduce como "Smallest Univalued Segment Assimilating Nucleus". Los algoritmos de SUSAN cubren los tópicos como filtración del ruido en las imágenes, búsqueda de los bordes y búsqueda de las esquinas.

El principio de algoritmos de SUSAN para detectar los detalles de la imagen es la máscara circular con el centro llamado núcleo. Si el brillo de cada pixel bajo la máscara tiene el mismo valor que el núcleo de la máscara entonces el área se llama USAN (Univalued Segment Assimilating Nucleus). Esa área contiene la información sobre la estructura de imagen en la región que cubre la máscara. El área de USAN está en máximo cuando el núcleo está en la región plana de superficie de la imagen, está en medio cuando el núcleo está muy cerca del borde y está en mínimo cuando el núcleo está en la esquina. Esta propiedad del área de USAN se usa para determinar la presencia de bordes y las figuras de dos dimensiones (esquinas). Esa máscara se pasa por cada punto de la imagen que será el núcleo con su USAN área correspondiente. Como SUSAN no usa las derivadas de imagen para la búsqueda de los bordes y las esquinas se explica por qué sus algoritmos son buenos en la presencia del ruido en la imagen. El ruido es suficientemente pequeño para ser considerado en la función USAN. La integración de valores individuales en el cálculo de las áreas reduce el efecto del ruido.

SUSAN analiza diferentes regiones por separado usando las mediciones directas locales y encuentra los lugares donde las fronteras de región particular están en el mínimo de la función, o sea, encuentra las esquinas que forman las regiones. Para el algoritmo de SUSAN no hay ninguna condición sobre la estructura de región.

El algoritmo de SUSAN para buscar las esquinas es el siguiente:

- Poner la máscara circular cerca del núcleo o pixel de interés.

Los valores de la máscara se toman de la función de distribución de Gauss:

$$g[i,j] = C * e^{-\frac{(i^2+j^2)}{2\sigma^2}} \quad (2.9)$$

donde C es la constante de normalización, σ es el parámetro de dispersión, i, j son dos dimensiones de la función. Después de la función se hace la aproximación digital para obtener los círculos con el radio 3.4 pixeles que da 37 pixeles en total.

- Usando la ecuación

$$C(r, r_0) = \exp - \left[\frac{I(r) - I(r_0)}{t} \right]^6 \quad (2.10)$$

se calcula el valor de similitud con el núcleo de cada pixel bajo la máscara. En la ecuación escrita arriba r_0 es la posición del núcleo, r es la posición de cualquier otro punto dentro de la máscara, $I(r)$ es el brillo de pixel, t determina el contraste mínimo entre detalles que se van a detectarse y el ruido a eliminar (umbral de luminosidad). De cómo escoger t depende el número de esquinas reportadas. Este umbral controla la cantidad de salida.

- Después los valores obtenidos bajo la máscara se suman

$$n(r_0) = \sum_r C(r, r_0) \quad (2.11)$$

El resultado se asigna a la posición de pixel central que fue estimado.

➤ Usando la ecuación

$$R(r_0) = \begin{cases} g-n(r_0), & \text{si } n(r_0) < g \\ 0, & \text{en otros casos} \end{cases} \quad (2.12)$$

se obtienen los valores de USAN área que producen las esquinas de imagen más destacadas. En la ecuación g es el umbral establecido igual a $n_{\max}/2$, donde n_{\max} es el número máximo de todos los valores obtenidos. Umbral g puede tener otros valores que afectarán a la calidad de salida. Aunque ese también afecta al número de esquinas encontradas lo más importante es su efecto a la forma de las esquinas detectadas.

➤ Examinar USAN área para detectar los puntos que son falsos en conjunto de las esquinas.

El problema de esquinas falsas puede ocurrir cuando los bordes entre dos regiones son borrosos. Para resolverlo se busca el centro de gravedad de USAN (mediana de las coordenadas de valores por eje x y y). Después se obtiene la distancia entre el centro de gravedad y el núcleo de la máscara. Si la distancia es grande entonces el núcleo es un punto falso en USAN áreas y por esa razón será rechazado.

2.4. Métodos basados en redes neuronales artificiales.

Existen dos algoritmos descritos por Zhou y Chellappa [Zhou, 1992] basados en el conjunto de métodos de redes neuronales artificiales (ANN): por lote y recursivo para calcular el campo de disparidad usando la secuencia de movimientos laterales estereos. El algoritmo de lote primero determina las tendencias de entradas usando todas las imágenes de secuencia. Después se calcula el campo de disparidad por redes neuronales. El segundo algoritmo usa el método recursivo de los cuadrados mínimos para obtener las tendencias de entradas en vez de valores de disparidad. Cuando el siguiente marco de la secuencia llega a ser disponible, las tendencias de entradas se reactualizan primero y después se aplica la red neuronal para estimar los valores de disparidad bajo limitantes epipolares (disparidad

binocular), fotométricos (calibración de cámara) y de suavizado original (pertenece a la imagen). El método no requiere el procedimiento de suavizado antes de procesar los datos. En la estructura paralela este algoritmo es muy rápido y eficiente para las aplicaciones de la visión de robot en el tiempo real. Los dos algoritmos usan las derivadas como mediciones primitivas para la correspondencia. En el método recursivo las tendencias de entradas de neuronas son reactualizadas recursivamente. Si los resultados intermedios no son requeridos, el procedimiento de correspondencia es implementado una sólo vez. En realidad los métodos recursivos existentes obtienen el campo de disparidad muy suave y denso. Los resultados experimentales muestran que los métodos recursivos necesitan menos iteraciones que los de lote. Este hecho se debe al uso del esquema de reactualización de entradas (especialmente para los píxeles ocultos). La buena estimación de las entradas hace que la red neuronal converja más rápido. Zhou y Chellappa [Zhou, 1992] suponen que si la disparidad máxima es amplia y la secuencia de imágenes es grande, entonces será posible detectar de los píxeles ocultos y reducir bastante el número de iteraciones.

En el cálculo de flujo óptico, basado en valores de intensidad y en las curvaturas principales bajo el movimiento rígido local y los limitadores de suavizado original, también se pueden usar las redes neuronales con la función máxima de evolución. Para detectar las discontinuidades de movimiento, se usa el proceso de línea. Para detectarlas correctamente primero se encuentran los elementos ocultos de mediciones iniciales de movimiento. Después con el proceso de línea se localizan las discontinuidades usando la información de elementos ocultos. Los valores de intensidad y sus curvaturas principales son estimados con la buena precisión usando una técnica de polinomios. Para asegurar la convergencia de la red, se usan las reglas determinísticas de decisión. Para mejorar el algoritmo se usan los marcos múltiples. Aquí también son aplicables los algoritmos de lote y recursivo. Ambos métodos se necesitan calcular el flujo óptico no más de dos veces.

Wright [Wright, 1992] analiza la habilidad del algoritmo de una capa de red neuronal para aprender los aspectos de la estructura de imagen local de la secuencia. Y después prueba las redes entrenadas para su procesamiento de imagen y las capacidades de suprimir el ruido.

El punto inicial de análisis fue el problema de remover el ruido de Gauss de la secuencia de imágenes usando la técnica de redes neuronales. Su estrategia es empezar con la arquitectura más simple de la red y ver sus capacidades, después moverse a un estado más complejo. El estado inicial es una capa de la red y el entrenamiento por la regla δ . En el segundo estado se usa la red de una capa y la regla de entrenamiento es no supervisada. Los resultados que obtuvo el autor son que la red basada en la regla δ aprenderá el paso de banda o el paso bajo central de la función de pesos cuando está entrenada por el escaneo de la imagen natural pasando la ventana.

Para la filtración espacio temporal la red neuronal demuestra que existe una alta correlación entre los niveles de gris de pixeles con las mismas coordenadas en los marcos adyacentes. Aplicando el filtro espacio temporal de paso bajo del tipo aprendido por la regla de δ reduce el ruido de pixel.

El valor posible que puede ser asociado con las redes neuronales aplicados a los métodos espacio temporales es la habilidad de escoger el filtro adecuado para la secuencia de imágenes en particular.

En general las aproximaciones de gradiente y correspondencia son sensibles a variaciones de contraste y diferentes deformaciones de imagen. También es muy común la restricción a la velocidad única en todas las posiciones de la imagen [Fleet, 1992]. En la aproximación de las fases el enfoque principal es el análisis de estabilidad de la fase que dependerá de la estructura de imagen [Fleet, 1992]. El problema del uso de redes neuronales artificiales es escoger los parámetros de entrada en la manera óptima que abarque la gran cantidad de diversos imágenes [Zhou, 1992].

A pesar de varios problemas que tienen los algoritmos descritos en este capítulo, es posible aplicarlos a las situaciones reales adoptándolos a cada una. De las aplicaciones que existen en el área de estimación de movimiento se hablará en el siguiente capítulo.