

Visión por Computadora I

Ing. Andrés F. Brumovsky

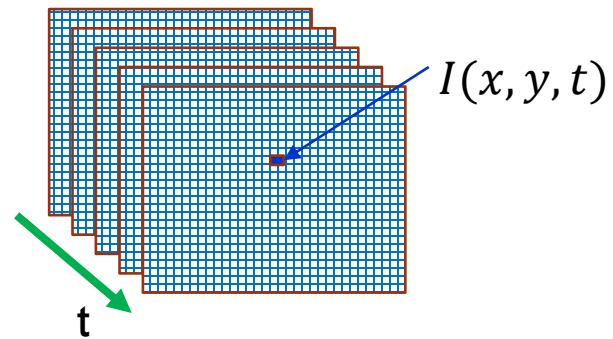
(abrumov@fi.uba.ar)

Ing. Maxim Dorogov

(mdorogov@fi.uba.ar)

Laboratorio de Sistemas Embebidos -FIUBA





INTRODUCCIÓN A MOVIMIENTO

Video

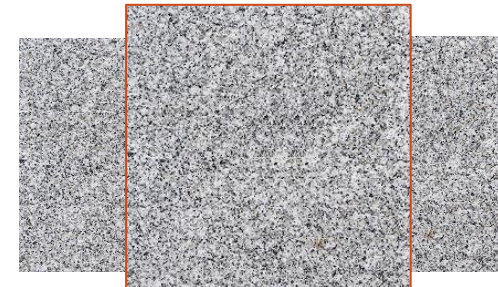
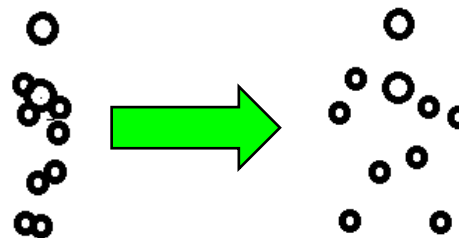
- Secuencia de imágenes capturadas a lo largo del tiempo
- Usualmente rápido (pocos cambios de imagen a imagen)
- A tiempos regulares (60Hz, 30Hz, 24Hz, etc.)
- Nuestra señal suma entonces una nueva variable, el tiempo.

Algunas aplicaciones de seguimiento

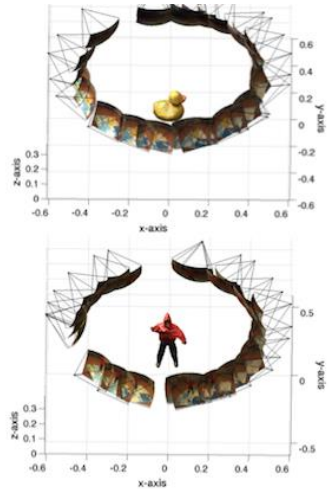
1. Separación de fondo (background subtraction)
 - Cámara estática
 - Separar el fondo estático del/los objeto/s en movimiento
2. Detección de borde de toma
 - Para hacer el corte preciso entre tomas de distintas cámaras
3. Segmentación por movimiento
 - Segmentar el video en objetos “coherentemente” en movimiento

Movimiento empobrecido

- Inclusive el movimiento empobrecido puede evocar una alta percepción de acción.



INTRODUCCIÓN A MOVIMIENTO



(a) Camera positions of the selected video frames for Duck and Action Man.



(b) Rendered views of the estimated 3D geometry.

Otras aplicaciones de seguimiento

1. Segmentación de objetos en espacio/tiempo
2. Estimación de estructura 3D
3. Aprendizaje de modelos dinámicos
4. Reconocimiento de eventos y actividades
5. Mejora de la calidad de video (estabilización)
6. Hardware específico: Cámaras por eventos (event/neuromorphic cameras)

Estimación de movimiento

1. **Métodos basados en características**
 - Extraer características visuales (esquinas, texturas, etc.) y seguirlas a través de los cuadros.
 - Esto da lugar a los campos de movimiento escasos (basados solo en las características que son buenas para el seguimiento). Aún así es un seguimiento robusto.
 - Adecuados cuando los movimientos son grandes (decenas de píxeles)
2. **Métodos directos o “densos” (Dense Flow)**
 - Buscan recuperar el movimiento de cada píxel a partir de variaciones espacio- temporales de los niveles de brillo
 - Da lugar a campos densos, pero sensibles a variaciones de apariencia
 - Adecuados cuando los movimientos son pequeños. Flujo óptico.



DENSE FLOW

- **Flujo óptico**

- Movimiento aparente de objetos o superficies
- Si no hay puntos de referencia no se puede predecir (parte blanca)

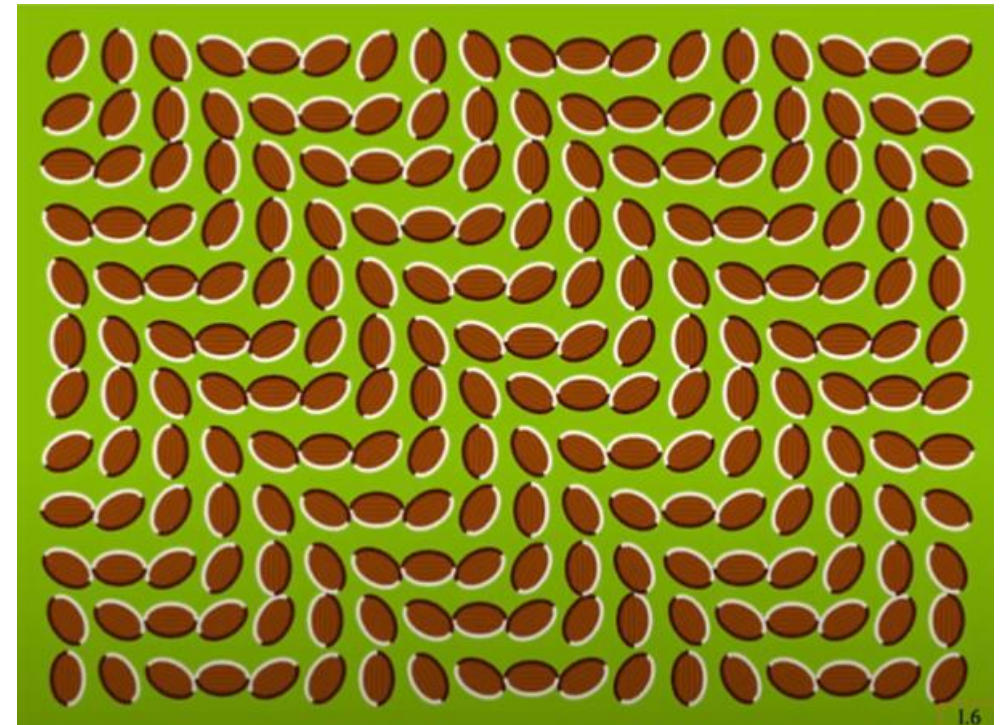
- ¿Cómo estimamos el movimiento de $I(x, y, t)$ a $I(x, y, t + 1)$?

- Buscamos resolver el problema de esta correspondencia.
- Dado un pixel en $I(x, y, t)$ buscar por píxeles cercanos del mismo “color”
- Este es el problema de “flujo óptico”

- **Constancia de color:** Un punto en $I(x, y, t)$ se parece a un punto en $I(x', y', t + 1)$

- Para imágenes en tonos de gris hablamos de “constancia de brillo”

- **Pequeños movimientos:** Se asume que los puntos no se mueven muy lejos entre t y $t + 1$



RESTRICCIONES DEL FLUJO ÓPTICO

- Restricción de constancia de brillo

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) \quad (1)$$

- Restricción de movimiento (pequeños desplazamientos)
 - 1 píxel o menos (las cosas cambian “suavemente”)
 - Podemos escribir una aproximación de Taylor

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) + \frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t + \dots$$

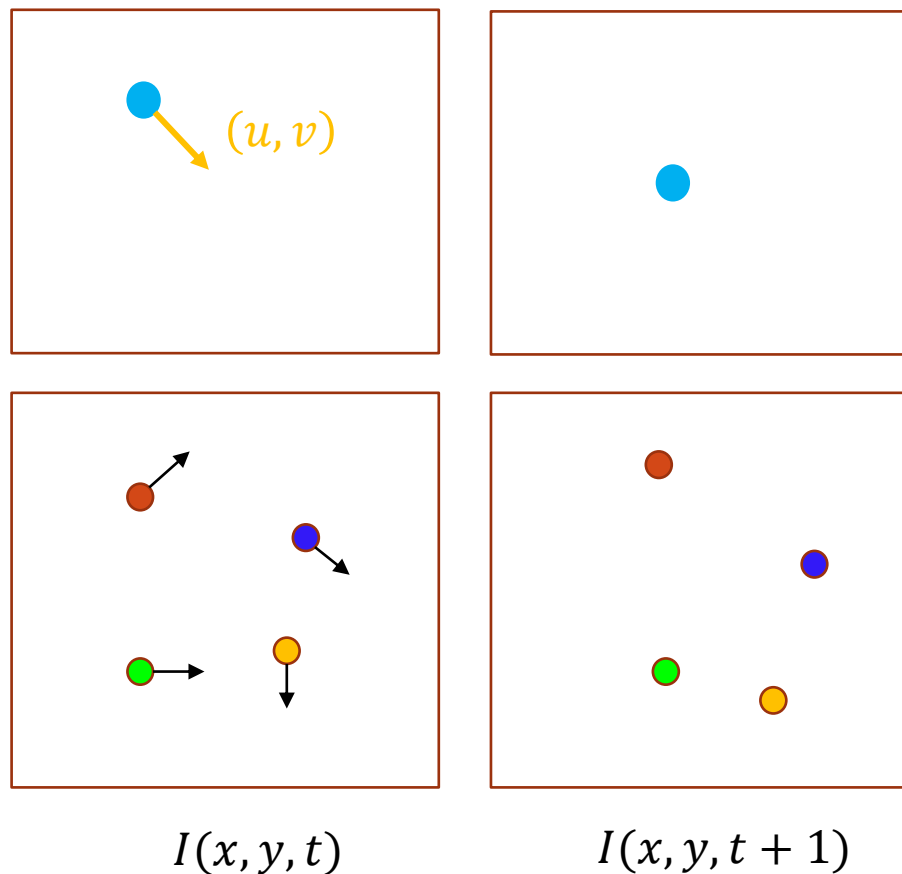
$$I(x + \Delta x, y + \Delta y, t + \Delta t) \approx I(x, y, t) + \frac{\partial I}{\partial x} \Delta x + \frac{\partial I}{\partial y} \Delta y + \frac{\partial I}{\partial t} \Delta t \quad (2)$$

- Restamos ambas ecuaciones y dividimos por Δt

$$I_t + I_x u + I_y v = 0$$

Donde:

$$I_x = \frac{\partial I}{\partial x}; I_y = \frac{\partial I}{\partial y}; I_t = \frac{\partial I}{\partial t}; u = \frac{\Delta x}{\Delta t}; v = \frac{\Delta y}{\Delta t}$$

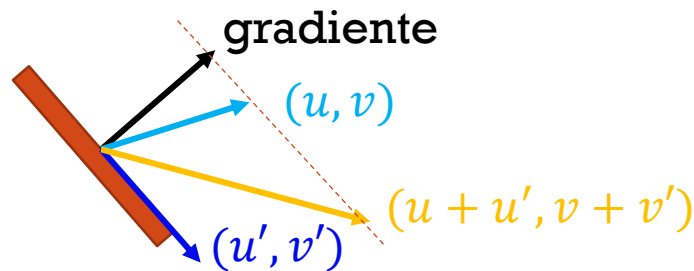


RESTRICCIONES DEL FLUJO ÓPTICO

- Esta última se conoce como ecuación de restricción de brillo constante

$$I_t + I_x u + I_y v = 0$$

- ¿Cuántas incógnitas y ecuaciones tenemos por píxel? → **2 incógnitas** (u, v) ...pero **una sola ecuación!**
- La componente de movimiento perpendicular al gradiente (paralelo al borde) no puede determinarse. Es decir, si (u, v) satisface la ecuación, también lo hace $(u + u', v + v')$.



PROBLEMA DE APERTURA

- Restricciones de flujo adicionales
 - Píxeles cercanos se mueven juntos (Restricción local)
 - El movimiento debe ser consistente a lo largo de toda la imagen (Restricción global)
 - Considerar únicamente regiones distintivas (como esquinas)
- **Enfoque global: Smooth Optical Flow (Horn y Schunk)**
 - Plantean el error en el flujo óptico

$$E = \iint [(I_x u + I_y v + I_t)^2 + \alpha^2 (\|\nabla u\|^2 + \|\nabla v\|^2)] dx dy$$

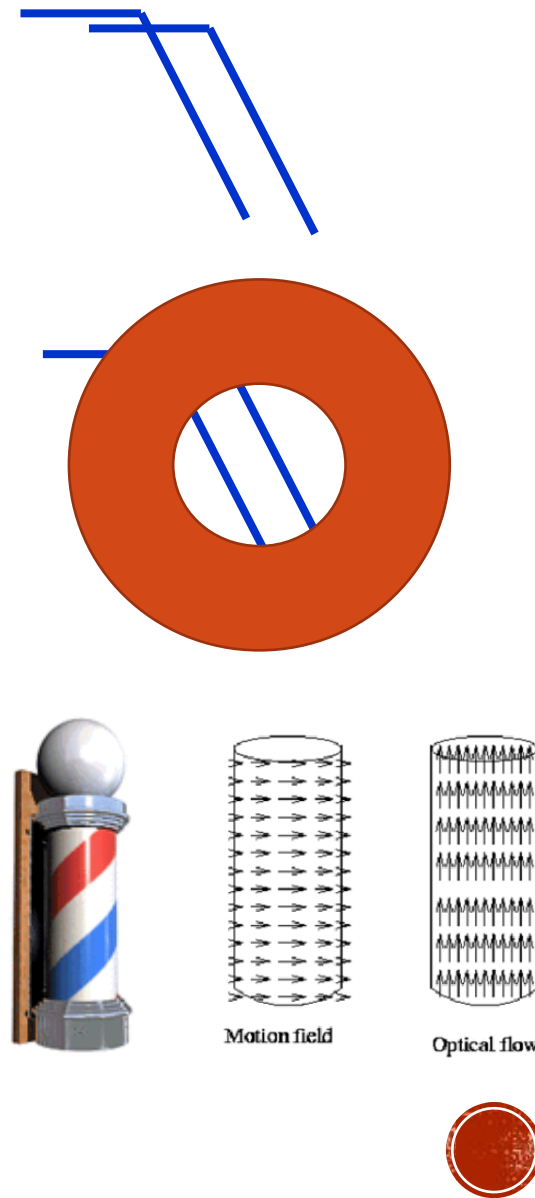
Es decir, se asume que el campo de velocidad tiende a variar lentamente a lo largo de la imagen. Se penalizan cambios bruscos de u y v

¿Cuándo este último error vale cero?

- Se buscan los (u,v) en cada punto (píxel) que minimicen:

$$e = e_s + \lambda e_c$$

El parámetro λ permite darle mayor o menor peso a la restricción de consistencia de brillo que a la de transiciones suaves y viceversa. Su definición será en base a cuánto le confiemos a los datos en cada caso.



LUCAS-KANADE

- **Enfoque local:** La idea es imponer más restricciones *locales* a cada píxel
- Se asume que el flujo de velocidades es suave localmente (en píxeles vecinos)
- Tanto que se asume que píxeles vecinos (por ejemplo en una ventana de 5x5) tienen el mismo (u,v)

$$0 = I_t(p_i) + \nabla I(p_i) \cdot [u, v]$$

$$\begin{bmatrix} I_x(p_1) & I_y(p_1) \\ I_x(p_2) & I_y(p_2) \\ \vdots & \vdots \\ I_x(p_{25}) & I_y(p_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(p_1) \\ I_t(p_2) \\ \vdots \\ I_t(p_{25}) \end{bmatrix}$$

$$\begin{matrix} A & d = b \\ (25 \times 2) & (2 \times 1) & (25 \times 1) \end{matrix}$$

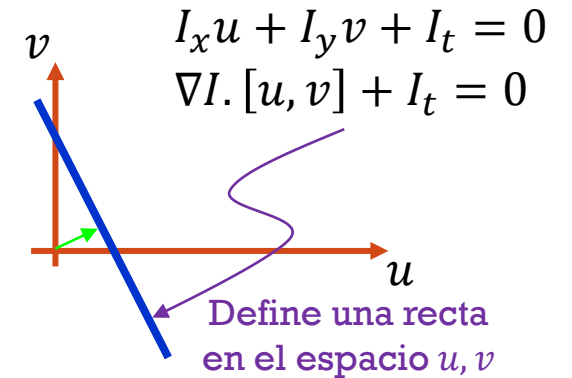
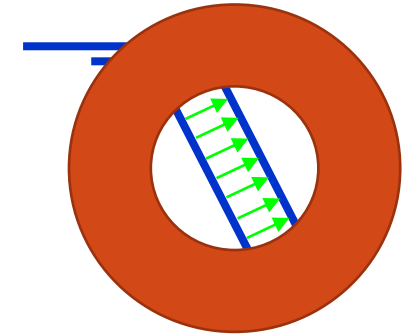
Es decir, pasamos a tener 25 ecuaciones por píxel!

- ¿Cómo lo resolvemos? → por mínimos cuadrados, minimizando $\|Ad - b\|^2$

$$(A^t A)d = A^t b$$

$$\begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix}$$

Estas sumatorias son sobre todos los píxeles en la ventana de $K \times K$ evaluada



LUCAS-KANADE

- ¿Cuándo es este sistema resoluble?
 - Cuando $A^t A$ es inversible
- Entonces $A^t A$ debe estar bien condicionada. Esto puede verse a través de la relación entre sus autovalores
 - λ_1/λ_2 no debe ser muy grande (considerando a λ_1 como el autovalor más grande)

- ¿A qué nos recuerda esto?

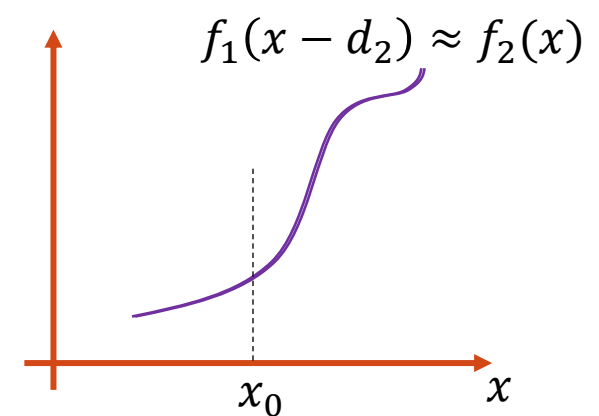
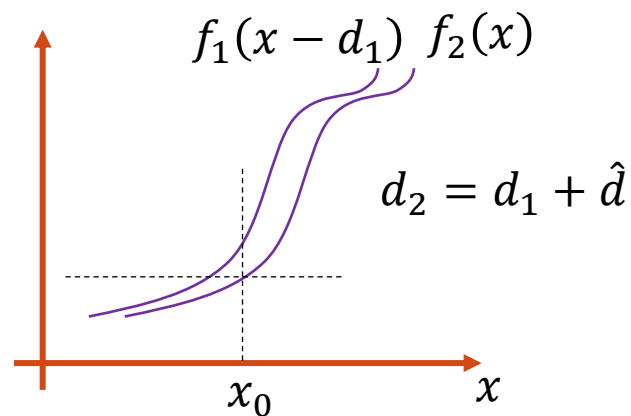
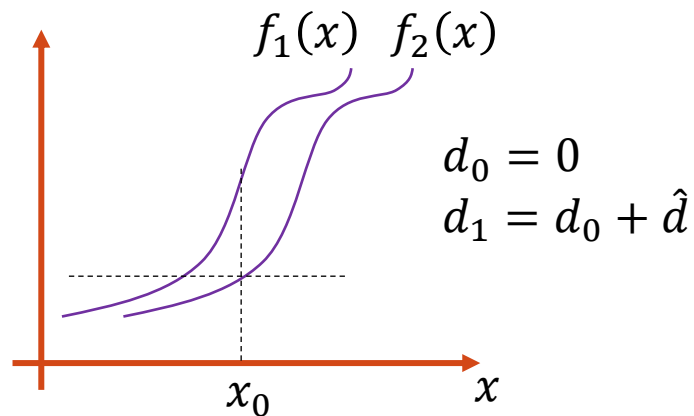
$$A^t A = \begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix} = \sum \begin{bmatrix} I_x \\ I_y \end{bmatrix} [I_x \quad I_y] = \sum \nabla I (\nabla I)^t$$

- Mismo criterio que el detector de esquinas de Harris. $M = A^t A$ es la matriz de momentos de orden 2.
 - Los autovectores y autovalores de M se relacionan con la dirección y magnitud del borde.
 - Recordando, algo era una buena esquina cuando λ_1 y λ_2 eran razonablemente grandes y comparables.
- ¿Qué pasa en imágenes color RGB?
 - En la misma ventana de 5×5 tendríamos $25 \times 3 = 75$ ecuaciones por píxel!
 - ¿Y si tuviésemos una “ventana de un solo píxel”? ¿No podríamos resolver el sistema de dos incógnitas (u, v) pero ahora tres ecuaciones?



ERRORES POSIBLES DE LUCAS-KANADE

- Movimientos grandes (más de un píxel) – La aproximación por Taylor no es buena
 - Transiciones no lineales (aún siendo suaves) → *Refinamiento iterativo*
 - Saltos de intensidad (ya no locales) → Estimación gruesa a fina (*coarse-to-fine flow*)
- Algoritmo iterativo de Lukas-Kanade
 1. Estimar la velocidad resolviendo las ecuaciones de Lucas-Kanade para cada píxel
 2. Deformamos la imagen del instante I_t al instante I_{t+1} usando el resultado del paso anterior. (Usando las técnicas de interpolación convencionales)
 3. Comparamos contra la verdadera imagen en $t + 1$. Volvemos a calcular el vector de desplazamiento y repetimos hasta converger.

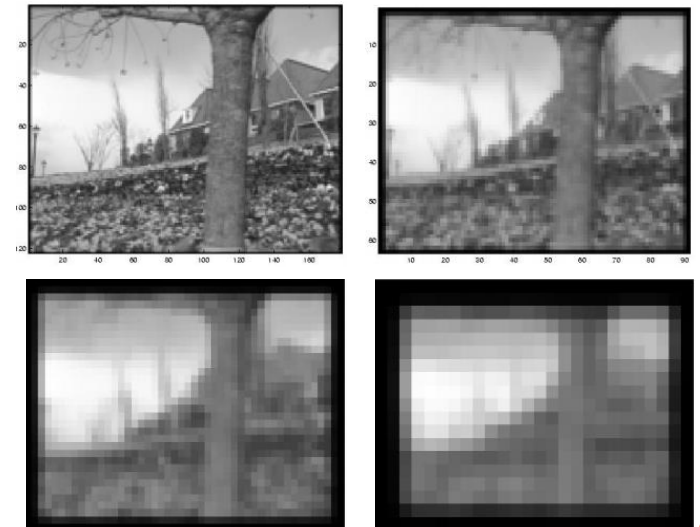
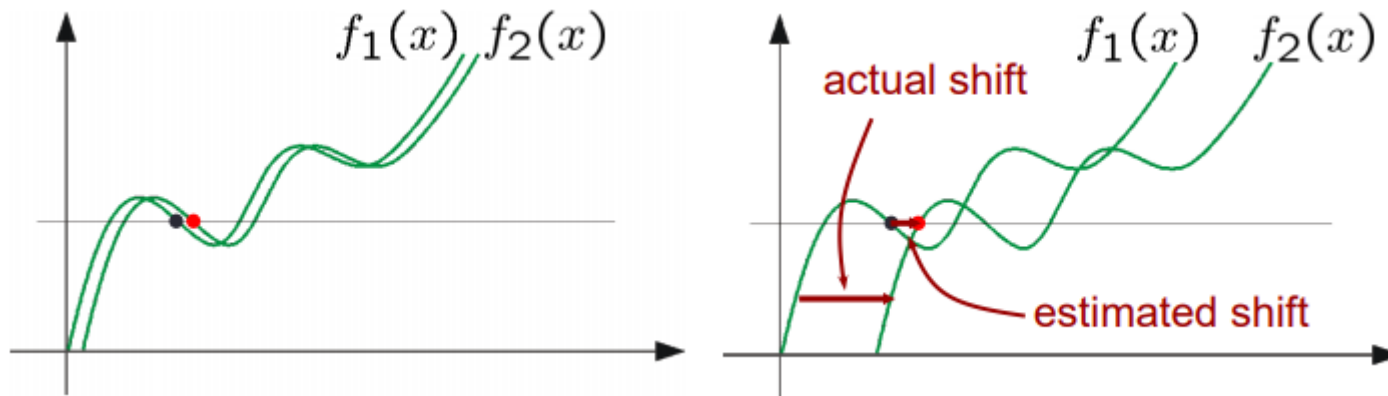


- Al implementarlo suele ser útil filtrar previamente con un pasa-bajos (para una mejor estimación de las derivadas y aproximaciones lineales en la intensidad de la imagen)



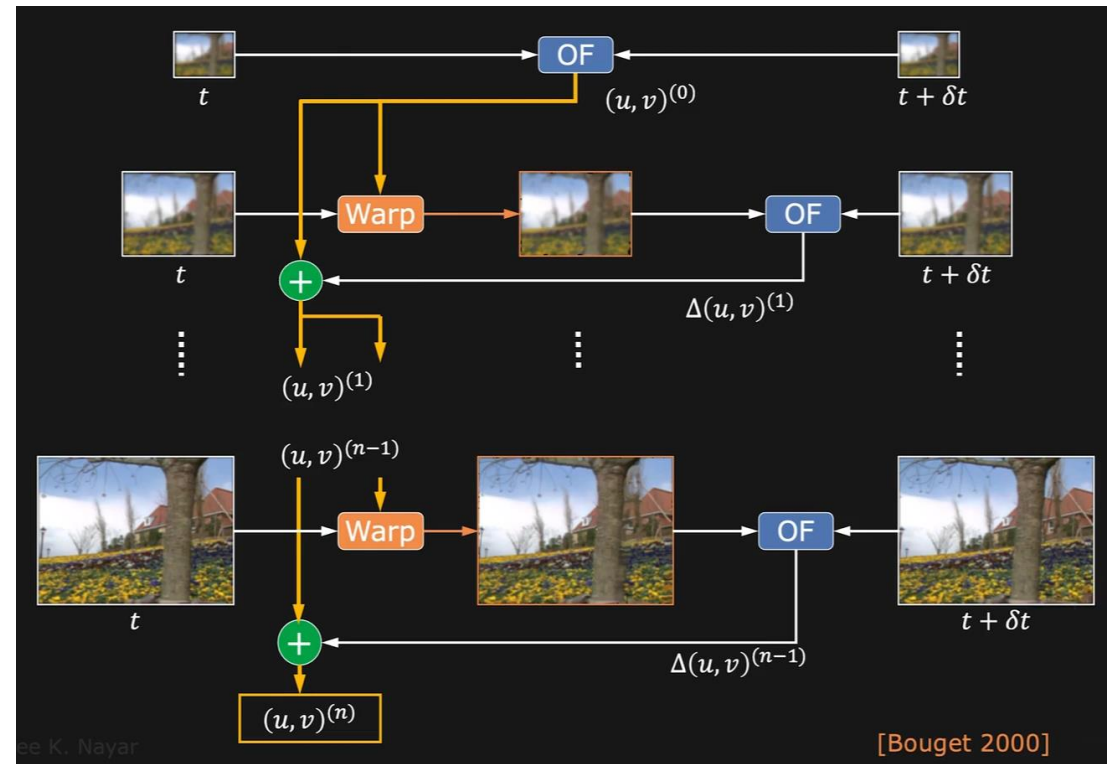
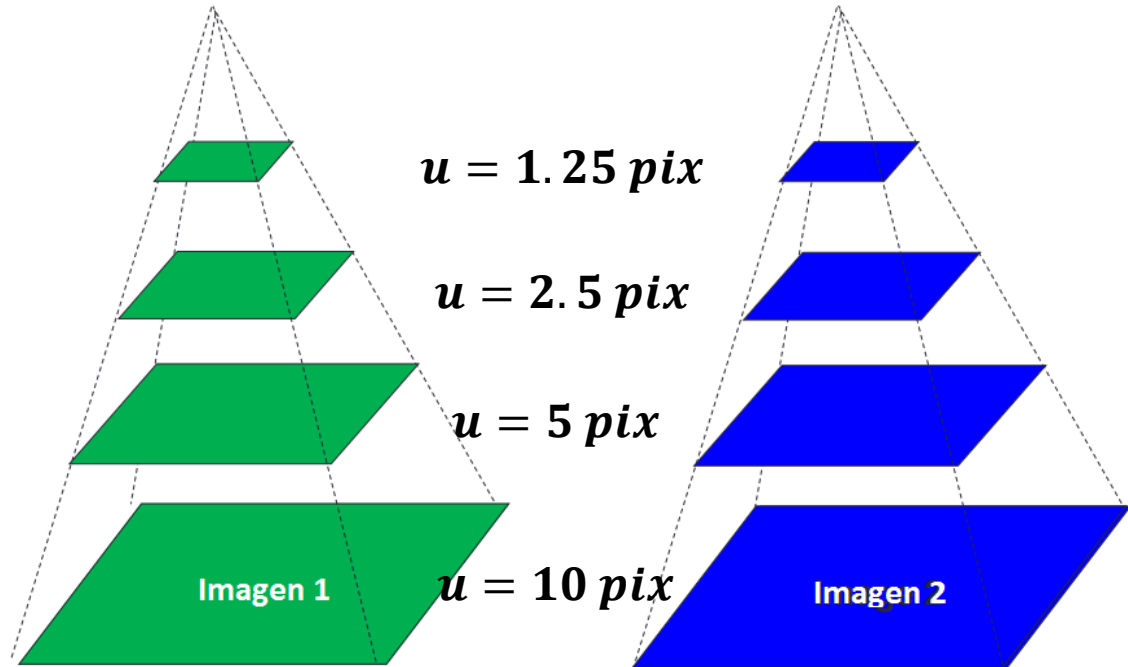
LUCAS-KANADE POR JERARQUÍAS

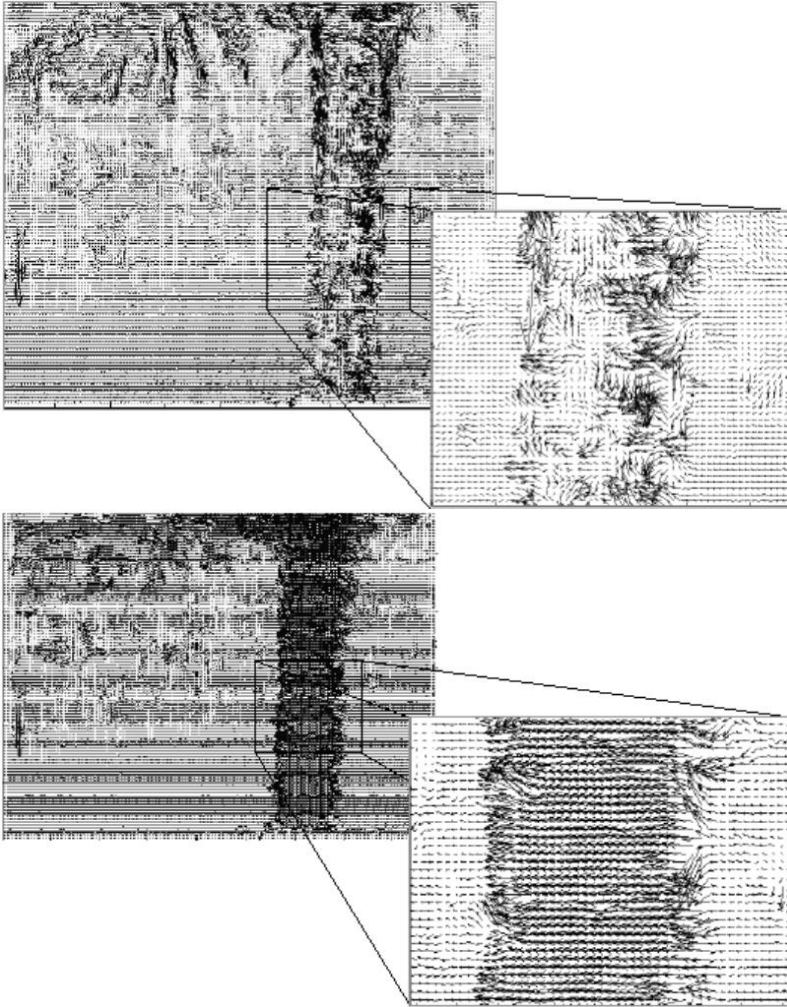
- Nos habíamos planteado qué pasaba si teníamos movimientos de varios píxeles (pérdida de variaciones locales)
- Tenemos una ambigüedad debida al aliasing temporal de la imagen donde muchos píxeles pueden tener el mismo nivel de intensidad
- Para superar esta situación podemos hacer una estimación de grueso a fino. Reduciendo la resolución!



LUCAS-KANADE POR JERARQUÍAS

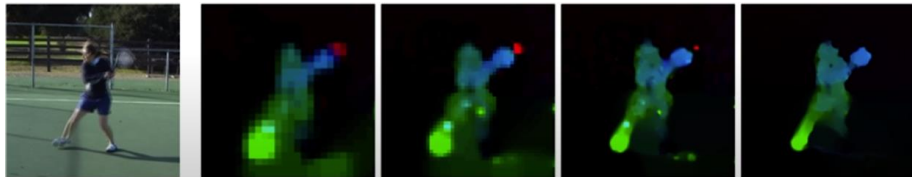
- La manera de realizar esto es a través de pirámides gaussianas (deben ser gaussianas para evitar problemas de aliasing espacial)



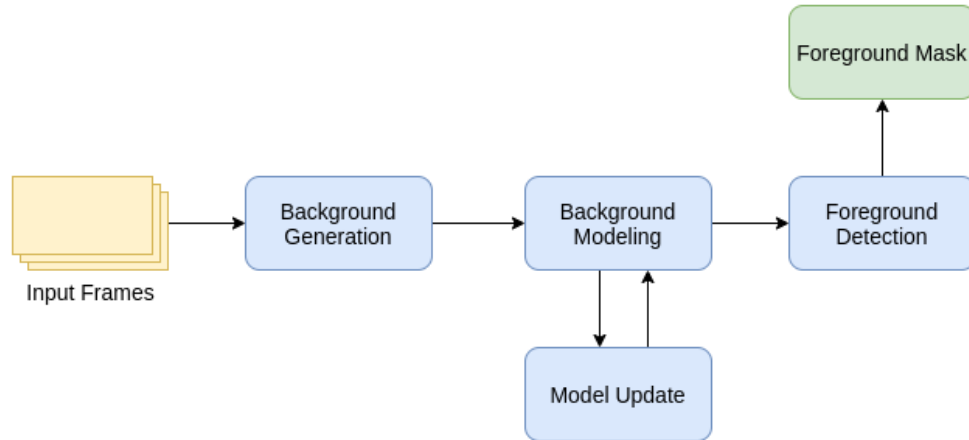


LK POR JERARQUIAS - RESULTADOS

- **Sin pirámides:** Falla en las regiones de grandes movimientos
- **Con pirámides:** Mejores resultados – Problemas en los bordes (hay píxeles que aparecen y desaparecen)
- **LK-escaso (sparse):** Consiste en aplicar LK por jerarquías solo a los lugares donde hay buenas características para seguir (esquinas). OpenCV usaba LK-denso y actualmente usa LK-escaso.
- Aproximándonos a la actualidad (Brox et al, CVPR 2009) se utiliza el concepto de Lucas-Kanade pero con algunos agregados:
 - + Constancia de gradiente
 - + Region matching
 - + Minimización de energía con término de suavidad
 - + Keypoint matching (para grandes desplazamientos)
- Más en la actualidad todavía (Fisher et al, 2015)
 - [FlowNet: Learning Optical Flow with Convolutional Networks](#)



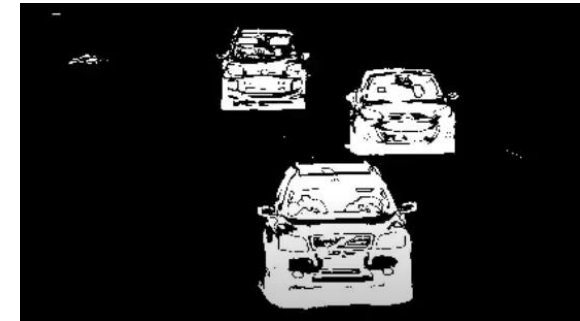
SUSTRACCIÓN DE FONDO



Naive background subtraction:

La mediana como estimador

- Se eligen N frames aleatorios y se calcula la mediana (background).
- Se resta el frame actual con la mediana y se binariza para obtener la mascara del objeto (*foreground*)
- Cada cierto intervalo se actualiza el modelo de *background* recalculando la mediana.



- Detección de objetos en movimiento con cámaras estáticas.

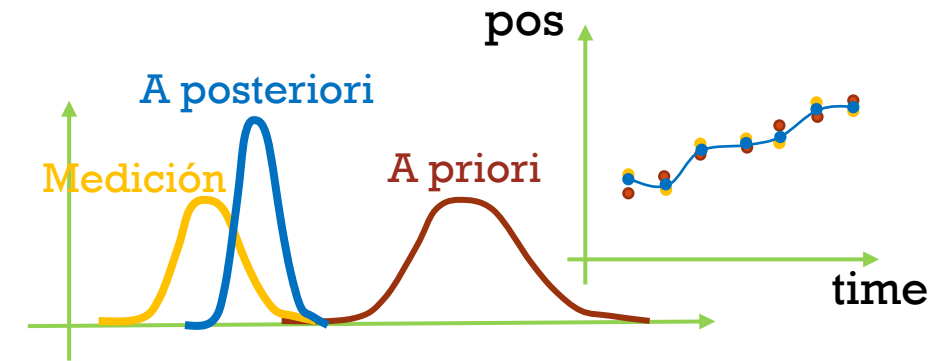
Enfoque “actual”:

- La distribución de intensidad de brillo de los pixeles del fondo se modela como una mezcla de gaussianas
- Se “aprende” el fondo y se generan mascaras de segmentación para los objetos en movimiento
- [“An Improved Adaptive Background Mixture Model for Realtime Tracking with Shadow Detection”](#). P. KaewTraKulPong and R. Bowden, 2001
- [“Visual Tracking of Human Visitors under Variable-Lighting Conditions for a Responsive Audio Art Installation”](#). B. Godbehre, A. Matsukawa, K. Goldberg , 2012

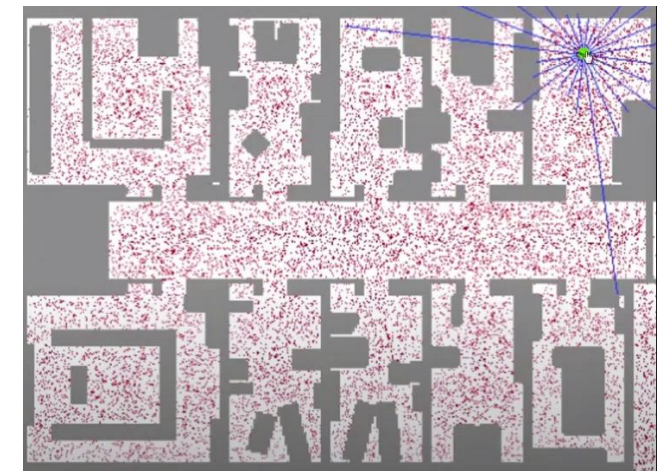


SEGUIMIENTO: RESUMEN

- **Desafíos:**
 - Puede ser difícil computar el flujo óptico en todos los píxeles.
 - Puede haber movimientos grandes (cosas que se mueven muy rápido) entre frames
 - Quizás haya que tener en cuenta la dinámica en consideración
 - Los errores de seguimiento se pueden acumular (y perder eventualmente al objeto)
 - Puede haber oclusiones del objeto a seguir.
- **Seguimiento de “buenas características”** – Shi-Tomasi (también SIFT)
- **Seguimiento basado en dinámica**
 - Dado un modelo de movimiento esperado, predecir dónde estarían los objetos en el frame siguiente, inclusive antes de ver la imagen.
 - Esta es la diferencia entre “seguimiento” y “detección”
 - **Objetivo 1:** Restringir la búsqueda
 - **Objetivo 2:** Obtener mejor estimación ya que el ruido de medición está atenuada por la dinámica a priori.
- **Seguimiento como una “inferencia”**
 - Estado desconocido (X): Parámetros reales que queremos detectar
 - Mediciones (Y): Observaciones ruidosas del estado subyacente
- Modelos lineales, con distribuciones gaussianas (predicciones y mediciones) → **Filtro de Kalman**
- Modelos no lineales, con distribuciones arbitrarias → **Filtros de partículas**



X: 2D – pos+vel med: 1D – pos



TP:7

- **Objetivo:**

- Implementar el detector de fondo naive usando la mediana como estimador. El algoritmo debe recibir el parámetro N (cantidad de frames utilizados para la estimación) y el intervalo de tiempo para recalcular el fondo.
- Se deben generar las mascaras de foreground y aplicarlas a los frames para segmentar los objetos en movimiento.
- Comparar con alguno de los métodos vistos en la practica basados en mezcla de gaussianas

