SEGMENTACIÓN DE VIDEO PARTE 2

Francisco J. Hernández López

fcoj23@cimat.mx

MÉTODOS DE SEGMENTACIÓN BASADO EN ENERGÍAS

- Un tema de investigación en segmentación es tratar de agrupar a los pixeles que tiene una apariencia similar y que las fronteras de los grupos estén bien definidos
- Podemos formular el problema usando:
  - Métodos variacionales
  - Métodos basados en Markov Random Field (MRF)
- El problema de segmentación puede ser escrita como:

$$E(f) = \sum_{i,j} \underbrace{E_D(i,j)}_{} + \underbrace{E_R(i,j)}_{},$$

$$E_D(i,j) = E_S\left(I(i,j); \Omega\big(p(i,j)\big)\right)$$

$I(i,j)$ consistente con la estadística de la región $\Omega\big(p(i,j)\big)$ en la función $p(i,j)$ que deseamos estimar

$$E_R(i,j) = s_x(i,j)\delta(p(i,j) - p(i+1,j)) + s_y(i,j)\delta(p(i,j) - p(i,j+1))$$

Mide la inconsistencia entre los vecinos de (i,j) modulado con los términos de suavidad s_x y s_y

Alan C. Bovik. 2009. *The Essential Guide to Video Processing* (2nd ed.). Academic Press

SEGMENTACIÓN BINARIA BASADA EN MRF

- Segmentación binaria consiste en clasificar a cada pixel de la imagen en dos posibles clases (por ej. FG y BG)
- Dada una imagen, el objetivo es estimar un mapa de etiquetas p. Este problema se puede atacar usando métodos bayesianos [Marroquin et. al. 1987, Szeliski book 2011]
- Sea
  - $\mathcal{L}$ → Malla (grid o Lattice) 2D
  - $F = \{f(\vec{x}) | \vec{x} \in \mathcal{L}\}$ → Familia de variables aleatorias definidas en $\mathcal{L}$
  - $p(\vec{x})$ toma una etiqueta l en el conjunto de etiquetas $\{0,1\}$
- F es un MRF en $\mathcal{L}$ con respecto a un sistema de vecindad $N_{\vec{x}}$ si y solo si se cumplen las siguientes condiciones:
  - $P(f) > 0, \forall f \in \{0,1\}^n$ **(Positividad)**
  - $P(f(\vec{x}) | f(\vec{z}), \forall \vec{z} \in \mathcal{L} \backslash \vec{x}) = P(f(\vec{x}) | f(\vec{y}), \forall \vec{y} \in N_{\vec{x}})$ **(Markovianidad)**

SEGMENTACIÓN BINARIA BASADA EN MRF (C1)

- De acuerdo con la regla de Bayes

$$P(p|f) = \frac{P(f|p)P(p)}{P(f)},$$

- Tomando el $-log$ en ambos lados, tenemos

$$-\log[P(p|f)] = -\log[P(f|p)] - \log[P(p)] + C,$$

- Calculamos el MAP para p, minimizando

$$E(p, f) = E_D(p, f) + E_R(p),$$

- Para MRF, la probabilidad $P(p)$ es una distribución Gibbs que puede ser escrito como una suma de potenciales aplicados por parejas:

$$E_R(p) = \sum_{\vec{y} \in N_{\vec{x}}} \Phi\big(p(\vec{x}), p(\vec{y})\big)$$

GRAPH CUT

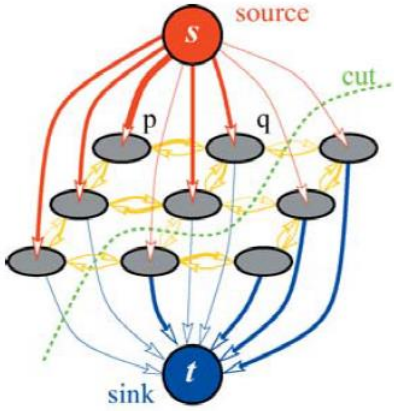
- Es una familia de algoritmos basados en MRF que resuelven el problema
  - Mínimo corte (min-cut)
  - Máximo flujo (max-flow)
- Un ejemplo de Graph Cut basado en MRF es el algoritmo de [Greig et. al.1989, Boykov et. al. 2001]
- Siguiendo el trabajo de [Greig et. al. 1989]

$$P(f|p) = \prod_{x=1}^{n} P(f(x)|p(x)) = \prod_{x=1}^{n} P(f(x)|1)^{p(x)} P(f(x)|0)^{1-p(x)}$$

$$P(p) \propto \exp\left[\frac{1}{2}\sum_{x=1}^{n}\sum_{y=1}^{n}\beta_{xy}\{p(x)p(y) + (1-p(x))(1-p(y))\}\right],$$

con $\beta_{xx} = 0$ y $\beta_{xy} = \beta_{yx} \geq 0$

GRAPH CUT (C1)



- Entonces el estimador MAP es aquel $\hat{p}$ que maximiza:

$$logP(p|f) = \sum_{x=1}^{n} \psi_x p(x) + \frac{1}{2} \sum_{x=1}^{n} \sum_{y=1}^{n} \beta_{xy} \{p(x)p(y) + (1-p(x))(1-p(y))\},$$

donde $\psi_x = \log\left\{\frac{P(f(x)|1)}{P(f(x)|0)}\right\}$ es una razón de verosimilitud entre las dos clases

- Note que para el caso binario $(p(x) \in \{0,1\})$ hay 2^n posibles valores del $logP(p|f)$ y n (numero de pixeles) puede ser muy grande, haciendo inviable la búsqueda de $\hat{p}$

- Graph Cut considera una red de capacidades con $n+2$ vértices
  - Un vértice s (source)
  - Un vértice t (sink)
  - (s,x) → Enlace dirigido de s a cada pixel x con capacidad $c_{sx} = \psi_x$ si $\psi_x > 0$
  - (x,t) → Enlace dirigido de x a t con capacidad $c_{xt} = -\psi_x$ si $\psi_x \leq 0$
  - (x,y) → Enlace no dirigido entre dos vértices internos con capacidad $c_{xy} = \beta_{xy}$

GRAPH CUT (C2)

- Para una imagen binaria f, dos particiones $B = \{s\} \cup \{x|f(x) = 1\}$, $W = \{t\} \cup \{x|f(x) = 0\}$ y sea $C(p) = \sum_{k \in B} \sum_{t \in W} c_{kl}$, entonces el conjunto de enlaces con un vértice en B y un vértice en W es llamado el corte (cut) y $C(p)$ su capacidad, el cual puede ser escrito como:

$$C(p) = \sum_{x=1}^{n} p(x) \max(0, -\psi_x) + \sum_{x=1}^{n} \big(1 - p(x)\big) \max(0, \psi_x) + \frac{1}{2} \sum_{x=1}^{n} \sum_{y=1}^{n} \beta_{xy} \big(p(x) - p(y)\big)^2,$$

GrabCut, Rother et. al. 2004

QMMF

- Sea $\hat{V}_l(\vec{x})$ la verosimilitud de un pixel x de pertenecer a cierta etiqueta $l(\vec{x}) \in \{FG, BG\}$
- Dicha verosimilitud es regularizada resolviendo el siguiente problema cuadrático:

$$\min_p \sum_{\vec{x}\in\Omega} \left\{ Q(p; V, \vec{x}) + \mu R_1(p; \vec{x}) + \lambda \sum_{\vec{y}\in N_{\vec{x}}} R_2(p; \vec{x}, \vec{y}) \right\},$$

Controla la entropía (medida de desorden)

Controla la suavidad de la solución

donde $N_{\vec{x}} = \{\vec{y}: \|\vec{x} - \vec{y}\|_2 = 1\}$

$$Q(p, V) = -\sum_{k=1}^{K} p_k^2 log V_k \,,$$

con K el número de clases

Mariano Rivera and Oscar Dalmau Cedeño. Variational Viewpoint of the Quadratic Markov Measure Field Models: Theory and Algorithms. IEEE Transactions on Image Processing, vol. 21, no. 3, pages 12461257, 2012.

QMMF (C1)

- Si consideramos $K = 2$ clases y $\mu = 0$ entonces el problema de segmentación lo podemos ver de la siguiente manera:

$$arg\min_{p} \frac{1}{2} \sum_{\vec{x} \in \mathcal{L}} U\big(p(\vec{x})\big) \quad sujeto\ a\ p(\vec{x}) \geq 0,$$

donde

$$U\big(p(\vec{x})\big) = p^2(\vec{x}) d_{FG}(\vec{x}) + [1 - p(\vec{x})]^2 d_{BG}(\vec{x}) + \lambda \sum_{\vec{y} \in N_{\vec{x}}} [p(\vec{x}) - p(\vec{y})]^2 W_\gamma(\vec{x}, \vec{y})$$

$$d_l(\vec{x}) = -log \hat{V}_l(\vec{x}),$$

$$W_\gamma(\vec{x}, \vec{y}) = \frac{\gamma}{\gamma + \|f(\vec{x}) - f(\vec{y})\|_2^2},$$

- La solución a este problema cuadrático puede ser encontrada iterando vía Gauss-Seidel:

$$p_{t+1}(\vec{x}) = \frac{d_{BG}(\vec{x}) + \lambda \sum_{\vec{y} \in N_{\vec{x}}} W_\gamma(\vec{x}, \vec{y}) p_t(\vec{y})}{d_{FG}(\vec{x}) + d_{BG}(\vec{x}) + \lambda \sum_{\vec{y} \in N_{\vec{x}}} W_\gamma(\vec{x}, \vec{y})}, \quad \text{con } p_0(\vec{x}) = \hat{V}_{FG}(\vec{x})$$

Francisco J. Hernandez-Lopez and Mariano Rivera. Change Detection by Probabilistic Segmentation from Monocular View. Machine Vision and Applications, pages 121, 2013.

SEGMENTACIÓN INTERACTIVA

- Sea $\mathcal{L} = I \cup U \cup N$ la matriz de pixeles con regiones: Interesting (I), Uninteresting (U) y No label (N)
- Sea el trimapa $T(\vec{x}) \in \{FG, BG, ne\}$
- Sea $\tilde{f}$ la imagen f normalizada entre $[0, nb-1]$, para cada canal RGB.
- Sea $\vec{k}_i = \left(k_i^R, k_i^G, k_i^B\right), i = 1, \dots, nc$, un vector de intensidades de $\tilde{f}$, con $nc = nb \times nb \times nb$ y nb el número de bins en cada canal de color RGB.

1. Calculamos histogramas para cada clase $l \in \{FG, BG\}$

$$h_l(\vec{k}_i) = \frac{\sum_{\vec{x}\in\mathcal{L}} \delta(T(\vec{x}) - l)\delta(\left\|\tilde{f}(\vec{x}) - \vec{k}_i\right\|_2^2)}{\sum_{\vec{x}\in\mathcal{L}} \delta(T(\vec{x}) - l)}$$

2. Calculamos la verosimilitud para cada clase

- $\hat{V}_l(\vec{x}) = \frac{V_l(\vec{x})+\epsilon}{V_{FG}(\vec{x})+V_{BG}(\vec{x})+2\epsilon}$, con $V_l(\vec{x}) = h_l\left(\tilde{f}(\vec{x})\right)$

SEGMENTACIÓN INTERACTIVA

3. Calculamos el mapa de segmentación p con QMMF

4. Actualizamos el Trimapa:

$$T(\vec{x}) \leftarrow \begin{cases} FG & p(\vec{x}) > 1/2 \\ BG & p(\vec{x}) < 0.05 \\ T(\vec{x}) & otro\ caso \end{cases}$$

5. Si se alcanza un máximo número de iteraciones o los cambios en T son menores que el 2% de la imagen, entonces parar con solución p, en otro caso, ir al paso 1.

Francisco J. Hernandez-Lopez and M. Rivera. AVScreen: a Real-Time video augmentation method. J. of Real-Time Image Process., pages 113, 2013.

SEGMENTACIÓN DE MOVIMIENTO

- También conocido como segmentación de flujo óptico
- Está relacionado con dos problemas:
  - Detección de cambios
  - Estimación de movimiento
- El objetivo de los algoritmos de segmentación de movimiento es determinar el número de modelos de movimiento que mejor se ajustan a la escena y el soporte espacial de cada modelo de movimiento
- Tipos de modelos paramétricos comúnmente usados:
  - Afine
  - Perspectiva
  - Mapeo cuadrático

Alan C. Bovik. 2009. *The Essential Guide to Video Processing* (2nd ed.). Academic Press

SEGMENTACIÓN USANDO DOS FRAMES

- Irani et. al.1994, proponen un modelo paramétrico multietapa de movimiento dominante. El procedimiento es el siguiente:

1. Calcular el vector de traslación 2D (d_x, d_y) dominante sobre toda la imagen:

$$\begin{bmatrix} \sum \frac{\partial I}{\partial_x}\frac{\partial I}{\partial_x} & \sum \frac{\partial I}{\partial_x}\frac{\partial I}{\partial_y} \\ \sum \frac{\partial I}{\partial_x}\frac{\partial I}{\partial_y} & \sum \frac{\partial I}{\partial_y}\frac{\partial I}{\partial_y} \end{bmatrix} \begin{bmatrix} d_x \\ d_y \end{bmatrix} = \begin{bmatrix} -\sum \frac{\partial I}{\partial x}\frac{\partial I}{\partial t} \\ -\sum \frac{\partial I}{\partial y}\frac{\partial I}{\partial t} \end{bmatrix},$$

En caso de que el movimiento dominante no sea una traslación, entonces la traslación estimada llega a ser una aproximación de primer orden del movimiento dominante

Alan C. Bovik. 2009. *The Essential Guide to Video Processing* (2nd ed.). Academic Press

SEGMENTACIÓN USANDO DOS FRAMES (C1)

2. Etiquetar todos los pixeles que corresponden al movimiento dominante:

   a) Registrar las dos imágenes usando el movimiento dominante estimado

   b) El problema se reduce a etiquetar regiones estacionarios entre las imágenes registradas

   Calculamos

$$FDN_{k,r}(\vec{x}) = \frac{\sum_{\vec{y}\in\mathcal{N}_{\vec{x}}} |I(\vec{y},k) - I(\vec{y},r)||\nabla I(\vec{y},r)|}{\sum_{\vec{y}\in\mathcal{N}_{\vec{x}}} |\nabla I(\vec{y},r)|^2 + c},$$

   Esto se puede implementar usando Multiresolución (Una pirámide Gausiana)

   donde:

   $FDN_{k,r}$ → Diferencias normalizadas entre el frame al tiempo k y el frame al tiempo r

   $\mathcal{N}$ → vecindario local del pixel $\vec{x}$

   c → constante para evitar inconsistencias (división entre cero)

Alan C. Bovik. 2009. *The Essential Guide to Video Processing* (2nd ed.). Academic Press

SEGMENTACIÓN USANDO DOS FRAMES (C2)

2. Etiquetar todos los pixeles que corresponden al movimiento dominante:

   c) Calculamos una medida de confiabilidad del movimiento

$$R(\vec{x}, k) = \frac{\lambda_{min}}{\lambda_{max}},$$

donde:

$\lambda_{min}, \lambda_{max}$ → los eigenvalores más pequeño y más grande de la matriz de coeficientes A:

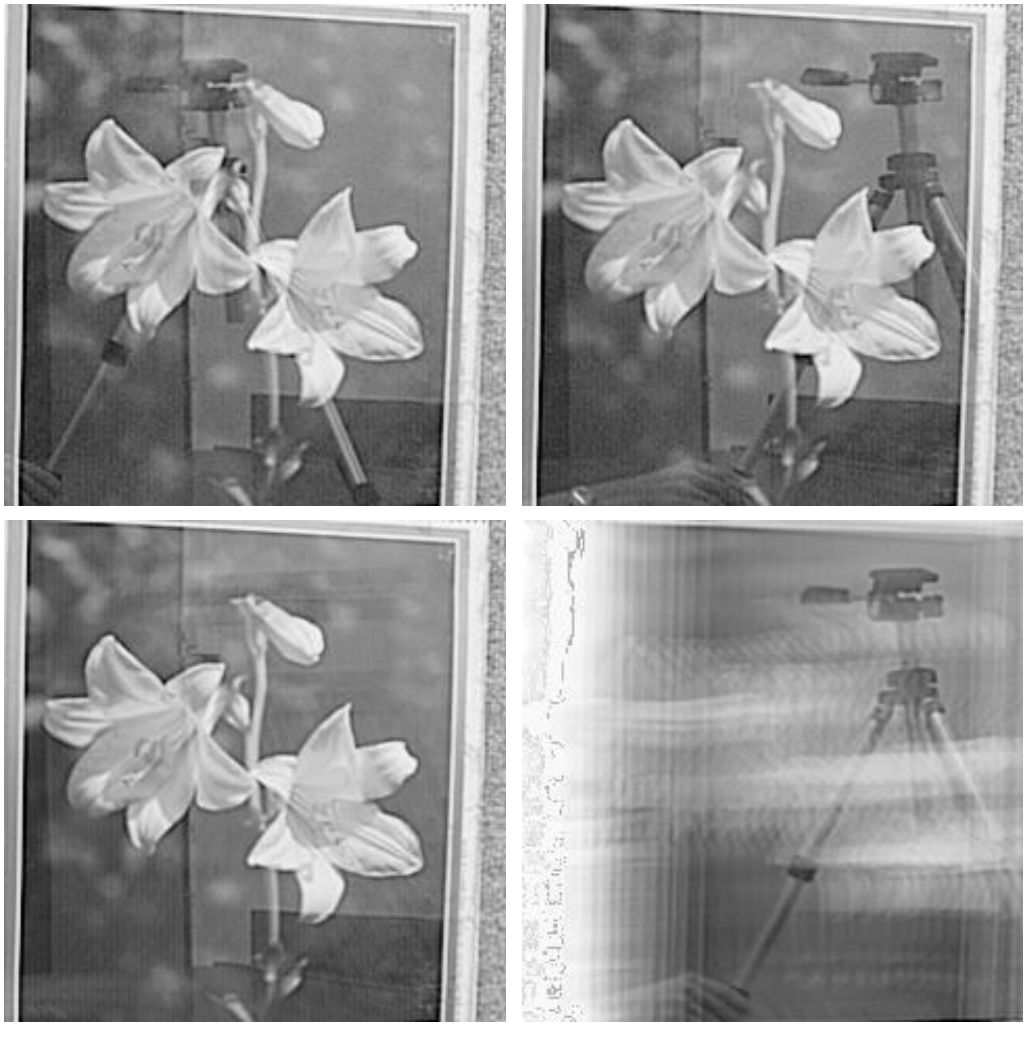
$$A = \begin{bmatrix} \sum \frac{\partial I}{\partial_x}\frac{\partial I}{\partial_x} & \sum \frac{\partial I}{\partial_x}\frac{\partial I}{\partial_y} \\ \sum \frac{\partial I}{\partial_x}\frac{\partial I}{\partial_y} & \sum \frac{\partial I}{\partial_y}\frac{\partial I}{\partial_y} \end{bmatrix}$$

Entonces un pixel es clasificado como estacionario si su FDN es bajo y su R es alto

3. Estimar los parámetros con un modelo de movimiento de mayor orden (afine, perspectiva, cuadrático, etc.) sobre la nueva región de análisis. Iterar pasos 2 y 3 hasta obtener una segmentación satisfactoria

Alan C. Bovik. 2009. *The Essential Guide to Video Processing* (2nd ed.). Academic Press

RECONSTRUCCIÓN DE OBJETOS "TRANSPARENTES"



Primera y última imagen de una secuencia

Reconstrucción de los dos movimientos independientes segmentados (las flores y el trípode)

Peleg, Michal Irani Benny Rousso Shmuel. "Computing occluding and transparent motions." Int. J. of Computer Vision 12.1 (1993): 5-16.

SEGMENTACIÓN DE MOVIMIENTO USANDO K-MEANS

- Wang y Adelson 1994, emplearon K-Means para segmentar
- Partir la imagen en bloques no traslapados y uniformemente distribuidos, usando un modelo *affine* para estimar el campo de movimiento dentro de cada bloque
- Determinar la confiabilidad de los parámetros estimados en cada bloque:

$$\bar{\eta}^2 = \sum_{\vec{x} \in \mathcal{B}} \|\vec{v}(\vec{x}) - \hat{v}(\vec{x})\|^2 ,$$

A partir de los parámetros estimados del modelo affine

A partir del Flujo óptico

donde $\mathcal{B}$ es un bloque de pixeles

- Los parámetros de movimiento de cada bloque con residual pequeño son seleccionados como modelos semilla

Alan C. Bovik. 2009. *The Essential Guide to Video Processing* (2nd ed.). Academic Press

SEGMENTACIÓN DE MOVIMIENTO USANDO K-MEANS (C1)

- Dados N vectores de parámetros semilla $\vec{a}_1, \vec{a}_2, \dots, \vec{a}_N$, donde $\vec{a}_n = [a_{n,1} \quad a_{n,2} \quad a_{n,3} \quad a_{n,4} \quad a_{n,5} \quad a_{n,6}]^T, n = 1, \dots, N$

- El problema ahora es encontrar K centros de agrupación $\vec{c}_1, \vec{c}_2, \dots, \vec{c}_K$, con $K \ll N$ y las etiquetas $k = 1, \dots, K$, asignadas a cada vector de parámetros $\vec{a}_n$, el cual minimiza:

$$\sum_{n=1}^{N} D(\vec{a}_n, \vec{c}_k)$$

con $D(\vec{a}, \vec{b}) = \left((\vec{a} - \vec{b})^T M (\vec{a} - \vec{b}) \right)^{\frac{1}{2}}$,

$M = diag(1, dim^2, dim^2, 1, dim^2, dim^2)$ es una matriz de escala de 6×6 y dim es el tamaño en pixeles de la imagen.

M.I.T. Media Laboratory Vision and Modeling Group, Technical Report No. 262, February 1994.
Appears in *Proceedings of the SPIE: Image and Video Processing II, vol. 2182, San Jose, February 1994.*

SEGMENTACIÓN DE MOVIMIENTO USANDO K-MEANS (C2)

- Entonces podemos usar K-means:
  1. Inicializar $\vec{c}_1, \vec{c}_2, \dots, \vec{c}_K$ de forma arbitraria
  2. Para cada bloque semilla $n, n = 1, \dots, N$, encontrar k:
  $$k = arg \min_s D(\vec{a}_n, \vec{c}_s),$$
  Donde s toma valores en el conjunto $\{1,2, \dots, K\}$, nota: si la mínima distancia excede un cierto umbral, entonces el bloque no se etiqueta
  3. Definir S_k como el conjunto de bloques semillas cuyo vector de parámetros afine es cercano a $\vec{c}_k, k = 1, \dots, K$, entonces, actualizar:
  $$\vec{c}_k = \frac{\sum_{n \in S_k} \vec{a}_n}{\sum_{n \in S_k} 1}$$
  4. Repetir los pasos 2 y 3 hasta que la diferencia entre los k-mean $\vec{c}_k$ de la iteración previa y actual sean menores a cierto umbral

Alan C. Bovik. 2009. *The Essential Guide to Video Processing* (2nd ed.). Academic Press

SEGMENTACIÓN DE MOVIMIENTO USANDO K-MEANS (C3)

- Ya que tenemos los centros k-means, entonces etiquetamos a cada pixel $\vec{x}$:

$$z(\vec{x}) = arg\min_{k} \|\vec{v}(\vec{x}) - \mathcal{T}(\vec{c}_k; \vec{x})\|^2 ,$$

donde

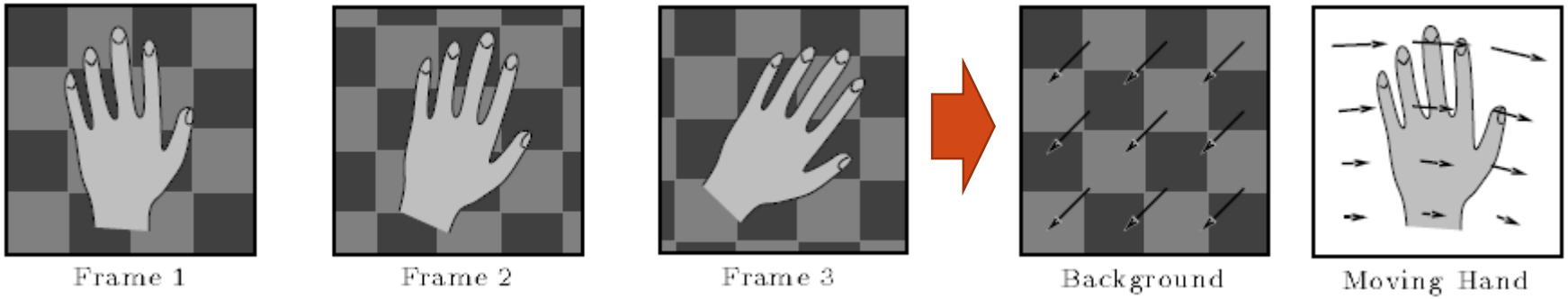
$$\mathcal{T}(\vec{c}_k; \vec{x}) = \begin{bmatrix} (c_{k,1} - 1)x_1 + c_{k,2}x_2 + c_{k,3} \\ c_{k,4}x_1 + (c_{k,5} - 1)x_2 + c_{k,6} \end{bmatrix},$$

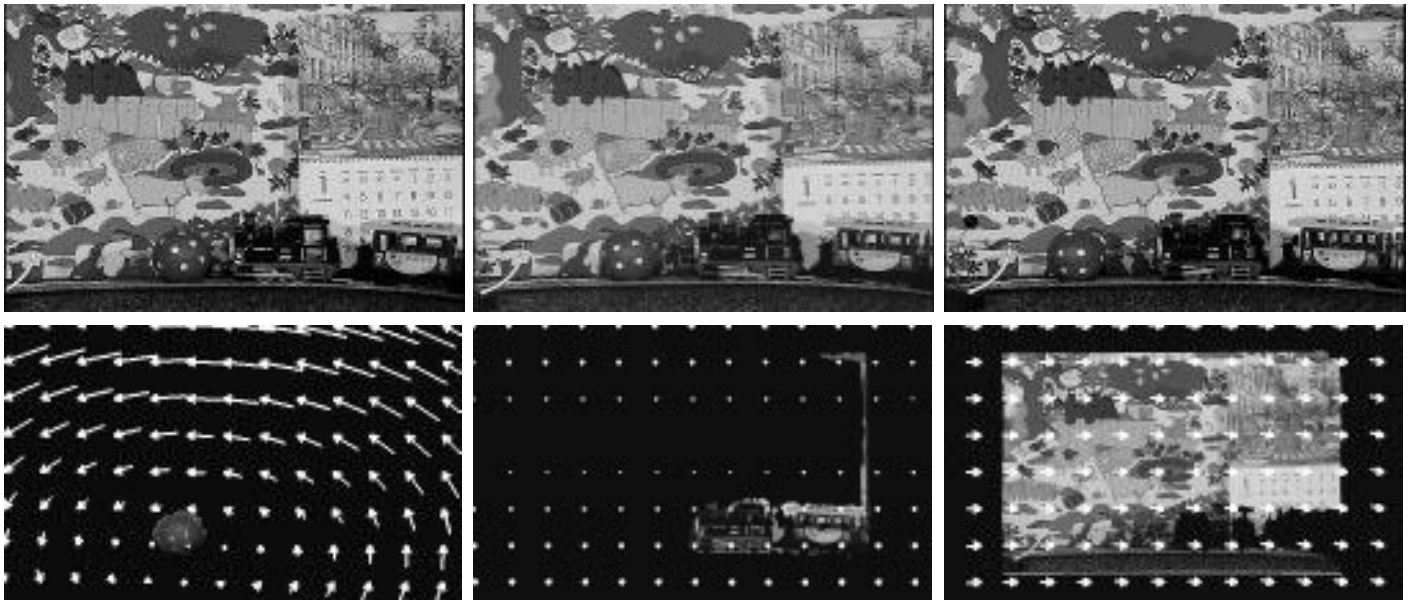
$$\vec{v}(\vec{x}) = \begin{bmatrix} v_1(\vec{x}) \\ v_2(\vec{x}) \end{bmatrix}$$

- Todo el procedimiento se puede repetir estimando nuevos parámetros semillas sobre las regiones estimadas en la iteración anterior y además los clusters se pueden dividir o mezclar entre las iteraciones

Alan C. Bovik. 2009. *The Essential Guide to Video Processing* (2nd ed.). Academic Press

REPRESENTACIÓN DE MOVIMIENTO CON ETIQUETAS





Segmentación correspondiente con el movimiento de: bola, trenecito y el fondo

Wang, John YA, and Edward H. Adelson. "Representing moving images with layers." Image Processing, IEEE Transactions on 3.5 (1994): 625-638.