

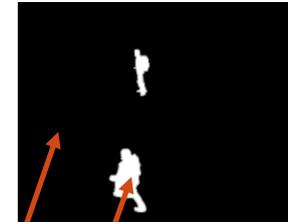
SUBSTRACCIÓN DE FONDO

Francisco J. Hernández López

fcoj23@cimat.mx



SUBSTRACCIÓN DE FONDO



<http://changedetection.net/>

- El camino más simple para substrair fondo es:
 - Asumir que tenemos una imagen del fondo en la cual no existen objetos moviéndose
 - Entonces las diferencias entre la imagen del fondo y la imagen actual son causadas por la presencia de objetos moviéndose
- Background (BG) → Pixeles que no cambian
- Foreground (FG) → Pixeles que han cambiado
- Pasos en la substracción del fondo
 - a) Modelo del fondo
 - b) Inicialización del fondo
 - c) Mantenimiento del fondo
 - d) Detección del FG

MODELO DEL FONDO

- Describe el tipo de modelo usado para representar el fondo
- Determina la habilidad que tiene el modelo con fondos que pueden ser unimodales o multimodales



baseline/office



dynamicBackground/fall

INICIALIZACIÓN DEL FONDO

- Generalmente el modelo del fondo es inicializado con:
 - El primer frame de la secuencia de video
 - Algún procesamiento sobre un conjunto de frames de entrenamiento
- Reto:
 - Obtener el modelo del fondo de entre un conjunto de frames donde más de la mitad de ese conjunto contiene objetos en movimiento
- Técnicas:
 - Mediana
 - Histogramas
 - SVM (Support Vector Machine)

MANTENIMIENTO DEL FONDO

- Mecanismo usado para adaptar el modelo a los cambios que se pueden presentar en la escena a través del tiempo
- Esquemas del mantenimiento del fondo:

- Mantenimiento del fondo a ciegas

$$B_{t+1}(\vec{x}) = (1 - \alpha)B_t(\vec{x}) + \alpha I_t(\vec{x}),$$

con $\alpha \in [0,1]$ un parámetro de aprendizaje

- Mantenimiento del fondo selectivo

$$B_{t+1}(\vec{x}) = (1 - \alpha)B_t(\vec{x}) + \alpha I_t(\vec{x}), \quad \text{si } \vec{x} \text{ es BG}$$

$$B_{t+1}(\vec{x}) = (1 - \beta)B_t(\vec{x}) + \beta I_t(\vec{x}), \quad \text{si } \vec{x} \text{ es FG}$$

con $\beta \ll \alpha$

- Esquemas Fuzzy

Toma en cuenta la incertidumbre de la clasificación, por ejemplo con el resultado de la detección del FG ir graduando la regla de actualización

MANTENIMIENTO DEL FONDO (C1)

- **Velocidad de aprendizaje**
Determina la velocidad de adaptación del modelo del BG a los cambios en la escena, estos pueden ser:
 - Fijos
 - Ajustados dinámicamente por algún método estadístico
 - Fuzzy
- **Mecanismos de mantenimiento del fondo**
 - Contadores → Número de veces que un pixel es clasificado como FG, si este número es más grande que un umbral, entonces se puede considerar al pixel como BG
 - CUSUM, Lindstrom et al.
- **Frecuencia de actualización**
Actualizar solo cuando sea necesario, el mantenimiento se puede realizar en cada frame, pero si no hay cambios significativos, entonces los pixeles no requieren de la actualización

DETECCIÓN DEL FG

- Consiste en comparar la imagen del modelo del BG con la imagen actual para etiquetar los píxeles ya sea como FG o como BG
- Preprocesamiento
 - Evita la detección de cambios debido al movimiento de la cámara o los cambios de iluminación
 - Ajustes geométricos:
 - Similitud
 - Affine
 - Proyectivo
- Prueba
 - Usualmente se aplica un umbral a la diferencia entre el modelo del BG y el frame actual

DETECCIÓN DEL FG (C1)

- Umbral

Existen diferentes tipos de umbral:

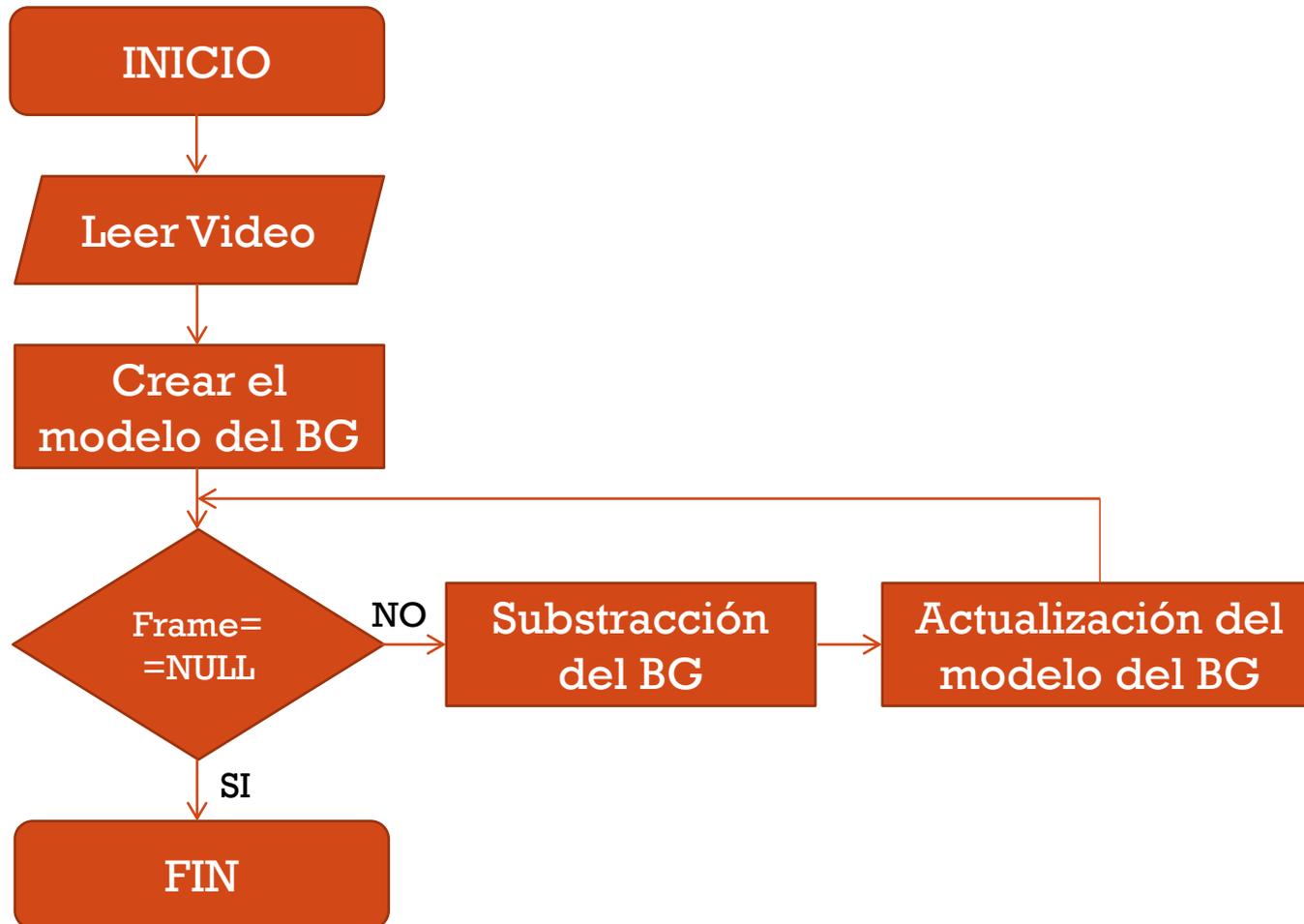
- Fijo
- Adaptable (Varianza temporal entre las intensidades del modelo del BG y el frame actual)

- Posprocesamiento

La idea es resaltar la consistencia de la máscara del FG:

- Eliminando pixeles aislados (morfología matemática)
- Inferencia Fuzzy entre las máscaras anterior y actual, etc.

DIAGRAMA GENERAL DE LA SUBSTRACCIÓN DE FONDO



PRINCIPALES RETOS

- Ruido en la imagen
Pobre calidad en la imagen, ya sea por ser adquiridas de WebCams o después de alguna compresión de video
- Movimientos irregulares de la cámara (Jitter)
El viento puede causar que la cámara se balancee, causando cierto movimiento irregular en la secuencia de video



cameraJitter/badminton

PRINCIPALES RETOS (C1)

- Ajustes automáticos de la cámara
 - Autoenfoco
 - Control de ganancia fotométrica
 - Balance de blancos
 - Control de brillo
- Cambios de iluminación
 - Graduales
 - Repentinos

<http://research.microsoft.com/en-us/um/people/jckrumm/wallflower/testimages.htm>



Wallflower → Time of Day



Wallflower → Light Switch

PRINCIPALES RETOS (C2)

- **Bootstrapping**

Durante el periodo de entrenamiento, el BG no está disponible, no hay un momento en el que se pueda tomar una sola imagen del BG sin objetos moviéndose

- **Camuflaje**

Regiones o pixeles en donde el FG y el BG no pueden distinguirse uno del otro.



Wallflower → Bootstrapping



Situaciones de camuflaje

PRINCIPALES RETOS (C3)

- **FG aperture**

Cuando un objeto en movimiento tiene regiones lisas (color uniforme), entonces los cambios dentro de esas regiones pueden no ser detectados.

La máscara del FG puede contener falsos negativos

- **Objetos del BG moviéndose**

Estos objetos no deberían considerarse como parte del FG

- **Objetos del BG insertados**

Un nuevo objeto del BG puede ser insertado

Estos objetos no deberían considerarse como parte del FG

PRINCIPALES RETOS (C4)

- **BG dinámico**

El BG puede vacilar o ser dinámico y entonces esto requiere modelos que puedan representar conjuntos disjuntos de valores de pixel

- **Beginning moving object**

Cuando un objeto que estaba inicialmente en el fondo se mueve, entonces ambos, dicho objeto y la nueva parte del BG revelada (llamada “fantasma”) son detectados

- **Sleeping FG object**

Algunos objetos del FG que llegan a tener poco movimiento pueden llegar a no ser distinguidos a partir del BG y entonces este sería incorporado al BG

Manejar esta situación depende del contexto

PRINCIPALES RETOS (C5)

- Sombras

Las sombras pueden ser detectadas como FG y pueden venir desde objetos del BG o desde los objetos que se están moviendo

Es considerado un campo de investigación por si mismo



shadow/peopleInShade



shadow/bungalows

MODELOS DEL BG TRADICIONALES

- Modelos básicos

Ya que tenemos el modelo del fondo B , entonces los pixeles del frame actual son clasificados como FG aplicando un umbral:

$$d(I_t(\vec{x}), B_{t-1}(\vec{x})) > T$$

con $d(\cdot, \cdot)$ una medida de distancia

- Modelos estadísticos

Estos ofrecen mayor robustez a los cambios de iluminación y fondo dinámico, pueden ser clasificados en las siguientes categorías:

- ✓ Modelos Gaussianos (Mixture Of Gaussians **MOG**, Kernel Density Estimation **KDE**)
- ✓ Support Vector Machine (SVM)
- ✓ Modelos de aprendizaje de subespacios

MODELOS DEL BG TRADICIONALES (C1)

- Modelos de cluster

Supongamos que cada pixel en el frame puede ser representado temporalmente por clusters:

- ✓ K-means
- ✓ Codebook
- ✓ Basic sequential clustering

- Modelos de redes neuronales (NN)

En este caso el fondo es representado por medio de los pesos de una red neuronal entrenada con N frames

La red aprende a clasificar cada pixel como BG o FG, algunos trabajos son:

- ✓ General Regression NN
- ✓ Multivalued NN
- ✓ Competitive NN
- ✓ Dipolar competitive NN
- ✓ Self Organizing NN, etc.

MODELOS DEL BG TRADICIONALES (C2)

- Modelos de estimación

El BG es estimado utilizando un filtro, este filtro puede ser:

- ✓ Wiener filter
- ✓ Kalman filter
- ✓ Chebychev filter

ALGORITMO: MANTENIMIENTO DEL BG USANDO EL FILTRO DE LA MEDIANA

1. Inicialización: Adquirir K frames
 - Calcular la mediana para los valores de intensidad de cada pixel en los K frames (la mediana representa el modelo del BG actual)
2. Leer el frame $K + 1$
 - Calcular las diferencias entre este frame y el modelo del BG actual
3. Aplicar un umbral a las diferencias
4. Utilizar alguna combinación de filtros de suavizado u operadores morfológicos para remover regiones muy pequeñas y llenar huecos de regiones grandes
5. Actualizar la mediana (modelo del BG) incorporando el frame actual
6. Regresar al paso 2

SUBSTRACCIÓN DEL BG POR MEZCLA DE GAUSSIANAS (MOG)

- Propuesto por Stauffer y Grimson en 1999
- La idea es modelar a cada pixel de forma independiente con una mezcla de Gaussianas
- Considerando un pixel con intensidad en escala de gris g_t
- La historia reciente de g_t puede ser modelada con K Gaussianas

$$\mathcal{N}_{k,t}(\mu_{k,t}, \sigma_{k,t}) \text{ con } k = 1, \dots, K$$

- ¿Cuántas K ? → Depende de la eficiencia computacional que se requiera

- Generalmente $K \in [3,7]$

$$P(g_t) = \sum_{k=1}^K w_{k,t} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(g_t - \mu_{k,t})^2}{\sigma_{k,t}^2}\right)$$

- A cada Gaussiana se le asocia un peso $w_{k,t}$

ALGORITMO: MANTENIMIENTO DEL BG VÍA UNA MOG

1. Inicializar:

- a) Elegir un número K de Gaussianas
- b) Probar $\alpha \in [0.01, 0.1]$
- c) Para cada pixel inicializar las K Gaussianas con μ_k, σ_k, w_k

2. Leer el siguiente frame t con intensidad g_t para cada pixel

- a) Comparar g_t con todas la Gaussianas

Si esta dentro de los 2.5 de la desviación estándar de la media, entonces considerarlo como un “match”

- b) De las Gaussianas que hacen “match”, seleccionar la mejor: l

3. Si hay un “match” con la Gaussiana l

- a) Actualizar los pesos $w_{k,t}$:

$$w_{k,t} = (1 - \alpha)w_{k,t-1} + \alpha M(k, t)$$

$$\text{con } M(k, t) = \begin{cases} 1 & \text{si } k = l \\ 0 & \text{otro caso} \end{cases}$$

Otra opción:

$$w_{k,t} = \begin{cases} (1 - \alpha)w_{k,t-1} & \text{si } k \neq l \\ w_{k,t-1} & \text{si } k = l \end{cases}$$

ALGORITMO: MANTENIMIENTO DEL BG VÍA UNA MOG (C1)

3. Si hay un “match” con la Gaussiana l

b) Calcular

$$\begin{aligned}\rho &= \alpha \mathcal{N}(g_t | \mu_l, \sigma_l) \\ \mu_{l,t} &= (1 - \rho)\mu_{l,t-1} + \rho g_t \\ \sigma_{l,t}^2 &= (1 - \rho)\sigma_{l,t-1}^2 + \rho(g_t - \mu_{l,t})^2\end{aligned}$$

4. Si no hay un “match”

a) Calcular $l = \arg \min_k (w_k)$

b) Eliminar la Gaussiana l y poner una nueva Gaussiana con:

$$\begin{aligned}\mu_{l,t} &= g_t \\ \sigma_{l,t}^2 &= 2 \max_k \sigma_{k,t-1}^2 \\ w_{l,t} &= 0.5 \min_k w_{k,t-1}\end{aligned}$$

ALGORITMO: MANTENIMIENTO DEL BG VÍA UNA MOG (C2)

5. Calcular

$$\#G = \arg \min_m \left(\sum_{k=1}^m w_{k,t} > T \right)$$

Nota: Ordenar los pesos de las Gaussianas de mayor a menor

Substracción de fondo:

Si un pixel es mayor a 2.5 veces la desviación estándar de cualquiera de las $\#G$ Gaussianas, entonces es clasificado como FG.

6. Aplicar algún pos-procesamiento (suavizado, dilataciones, erosiones, etc.)

7. Regresar al paso 2

MODELO DEL BG NO PARAMÉTRICO

- En escenas al aire libre (outdoor) \exists un rango amplio de variaciones, los cuales pueden ser muy rápidos:
 - Waving trees
 - Rippling water
 - Ocean waves
- Modelar estas variaciones requiere una representación flexible de la distribución de probabilidad del BG en cada pixel
- Esto motiva el uso de estimadores de densidad no paramétricos para modelar el BG:
 - Una técnica en particular es el Kernel Density Estimation (**KDE**)

KDE

- Dada una muestra $S = \{x_i\}_{i=1,2,\dots,N}$ a partir de una distribución con función de densidad $p(x)$, entonces un estimador de la densidad en x puede ser:

$$\hat{p}(x) = \frac{1}{N} \sum_{i=1}^N K_{\sigma}(x - x_i)$$

donde K_{σ} es una función kernel (función de ventana) con un ancho de banda (escala) σ

- El kernel K satisface que

$$\begin{aligned} K(t) &\geq 0 \\ \int K(t) dt &= 1 \end{aligned}$$

KDE (C1)

- Elgammal et. al. introducen un modelo del BG estimando la densidad del kernel
- Sean x_1, x_2, \dots, x_N una muestra de intensidades de un pixel
- Supongamos que en cada pixel tenemos ciertos valores característicos (features) en dimensión d
- Entonces la estimación de la función de densidad puede ser generalizada de la siguiente manera:

$$p(x_t|BG) = \frac{1}{N} \sum_{i=1}^N \prod_{j=1}^d K_{\sigma_j}(x_t^j - x_i^j)$$

donde x_t^j es el feature en la j -ésima dimensión al tiempo t

- El pixel es considerado *FG* si

$$p(x_t|BG) < th, \text{ con } th \text{ un umbral}$$

EVALUACIÓN DEL RENDIMIENTO

- Puede ser en términos de tiempo, memoria o de que tan bien el algoritmo clasifica los píxeles con el menor número de falsos posibles
- En términos de detección y dado un Ground Truth (GT), podemos usar:
 - Evaluación de curvas Receiver Operating Characteristics (ROC)
 - Evaluación de similitud

EVALUACIÓN DE CURVAS ROC

- Para las curvas ROC es necesario tener métricas basadas en el Ground Truth (GT)
- True Positives (TP) → Píxeles bien clasificados como FG
- True Negatives (TN) → Píxeles bien clasificados como BG
- False Positives (FP) → Píxeles mal clasificados como FG
- False Negatives (FN) → Píxeles mal clasificados como BG

- Positive Detection Rate (PDR):

$$PDR = \frac{TP}{TP + FN}$$

- Negative Detection Rate (NDR):

$$NDR = \frac{TN}{TN + FP}$$

- Presición:

$$P = \frac{TP}{TP + FP}$$

- Medida de efectividad:

$$F = \frac{2 * PDR * P}{PDR + P}$$

EVALUACIÓN DE SIMILITUD

- Sea A una región detectada
- Sea B el correspondiente GT
- Entonces la similitud entre A y B se define como:

$$S(A, B) = \frac{A \cap B}{A \cup B}$$