

Strategies for Object Segmentation, Detection and Tracking in Complex
Environments for Event Detection in Video Surveillance and Monitoring
TEC2011-25995 EventVideo (2012-2014)

eVi Workshop

23rd May 2014

*Escuela Politécnica Superior
Universidad Autónoma de Madrid*



*e
v
i
d
e
o*



EventVideo



TEC2011-25995 EventVideo (2012-2014)

Strategies for Object Segmentation, Detection and Tracking in
Complex Environments for Event Detection in Video Surveillance
and Monitoring

<http://www-vpu.eps.uam.es/eventvideo/>

Proyecto financiado por el Ministerio de Economía y Hacienda



Desarrollado por el *Video Processing and Understanding Lab* de la
Escuela Politécnica Superior de la Universidad Autónoma de
Madrid

- WP1: Framework establishment and maintenance
 - Sesión 1
- WP2: Video object segmentation
 - Sesión 2
- WP3: People and events modelling and detection
 - Sesiones 3 y 5
- WP4: Video object tracking
 - Sesión 4
- WP5: Integration, orchestration and exploration
- WP6: Testing and evaluation, and dissemination

- 10:15 Sesión 1: Infraestructura
- 11:00 Sesión 2: Segmentación
 - Segregación frente-fondo en entornos multimodales
 - Segmentación con cámaras PTZ
- 11:30 Sesión 3: Detección de personas
 - Detección de personas en tiempo real
 - Segmentación persona-fondo
- 12:00 Pausa
- 12:15 Sesión 4: Seguimiento
 - Seguimiento de objetos a largo plazo
 - Seguimiento basado en atención visual
- 12:45 Sesión 5: Detección de eventos
 - Actividades e interacciones de personas
 - Detección de regiones estáticas: robo y abandono
 - Detección de anomalías

Strategies for Object Segmentation, Detection and Tracking in Complex Environments for Event Detection in Video Surveillance and Monitoring
TEC2011-25995 EventVideo (2012-2014)

Sesión 1: INFRAESTRUCTURA

Jesús Bescós y Carlos Sánchez
j.bescos@uam.es .carlos.sanchezb@estudiante.uam.es



UNIVERSIDAD AUTÓNOMA DE MADRID

Introducción

VPU

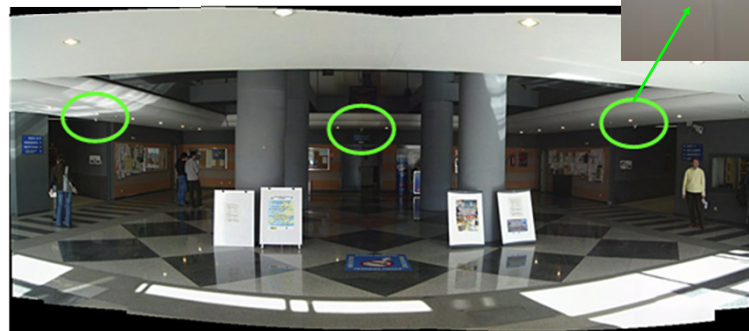
The diagram illustrates the infrastructure for event video. It starts with a central 'event video' logo. To the left is a photo of a group of people celebrating. To the right is a photo of a group of people with their arms raised, overlaid with a white skeleton overlay. Below these are two boxes: 'Infraestructura de captura' (Capture Infrastructure) and 'Infraestructura de proceso' (Processing Infrastructure), connected by a double-headed arrow. Below each box is a faint, mirrored version of the box's text.

TEC2011-25995 EventVideo (2012-2014) 2/7 Infraestructura

Cámaras fijas



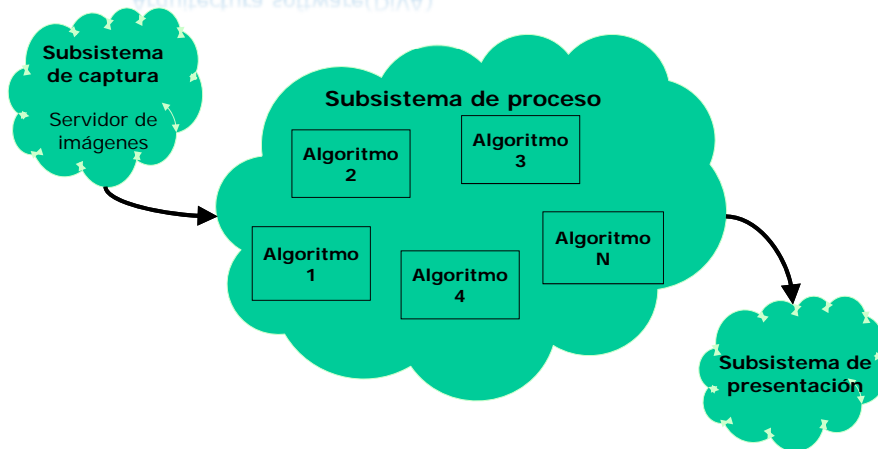
Cámaras PTZ

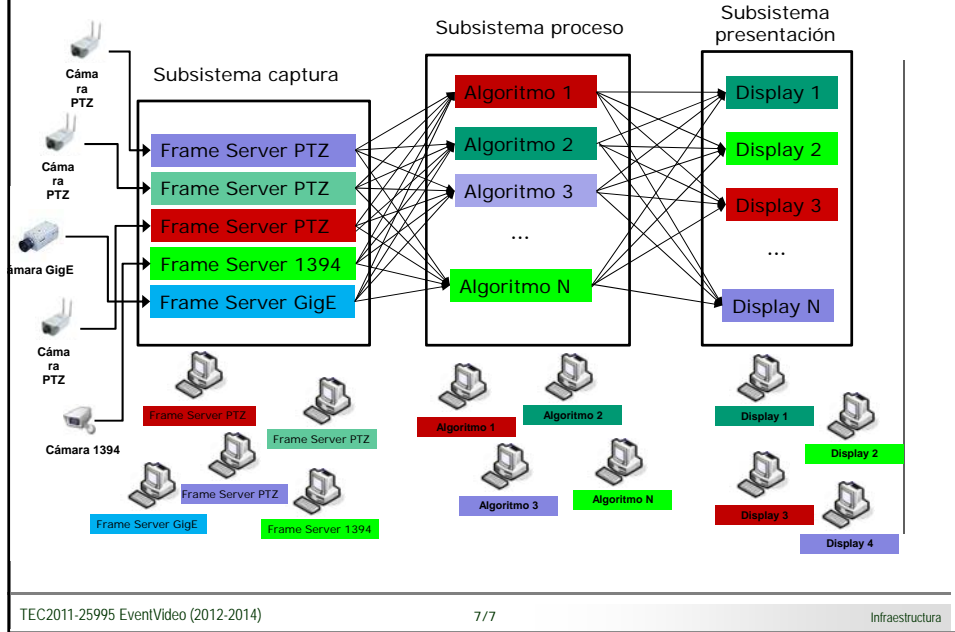


Equipamiento



Arquitectura software(DiVA)





Strategies for Object Segmentation, Detection and Tracking in Complex Environments for Event Detection in Video Surveillance and Monitoring
TEC2011-25995 EventVideo (2012-2014)

Sesión 2: Segmentación

Segregación frente-fondo en entornos multimodales.

Alejandro Blanco Carrasco
Marcos Escudero-Viñolo
{Alejandro.blancoc,Marcos.Escudero}@uam.es



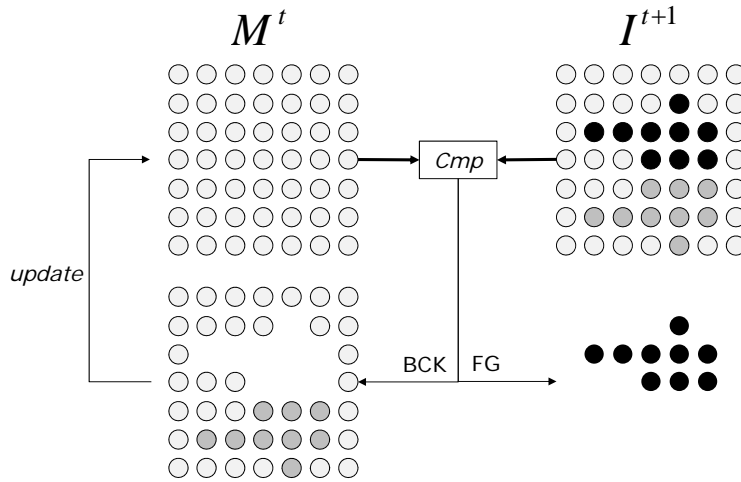
Modelando fondo en entornos dinámicos



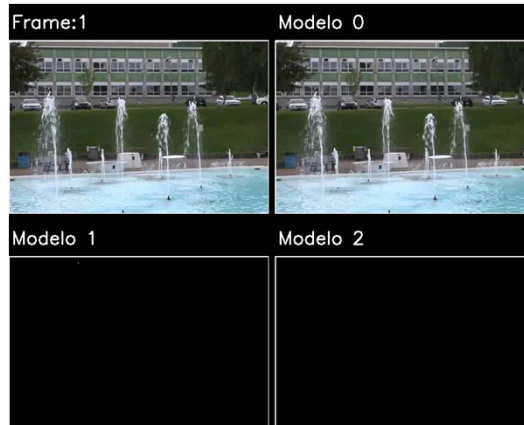
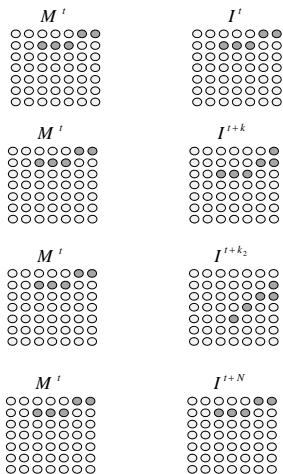
¿Qué es la sustracción de fondo?



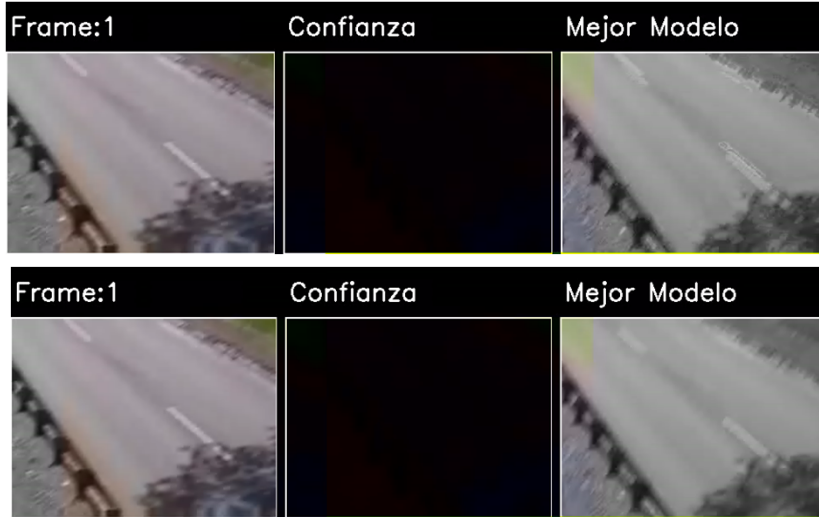
¿Qué es la sustracción de fondo?



Principales problemas
Multimodalidad



Principales problemas
Ruido en bordes / vibración

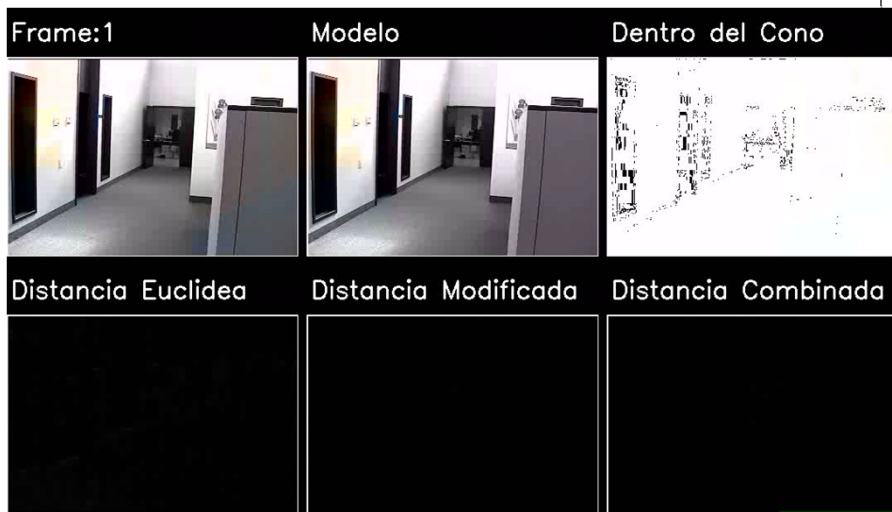


TEC2011-25995 EventVideo (2012-2014)

5/10

Segregación frente-fondo en entornos multimodales.

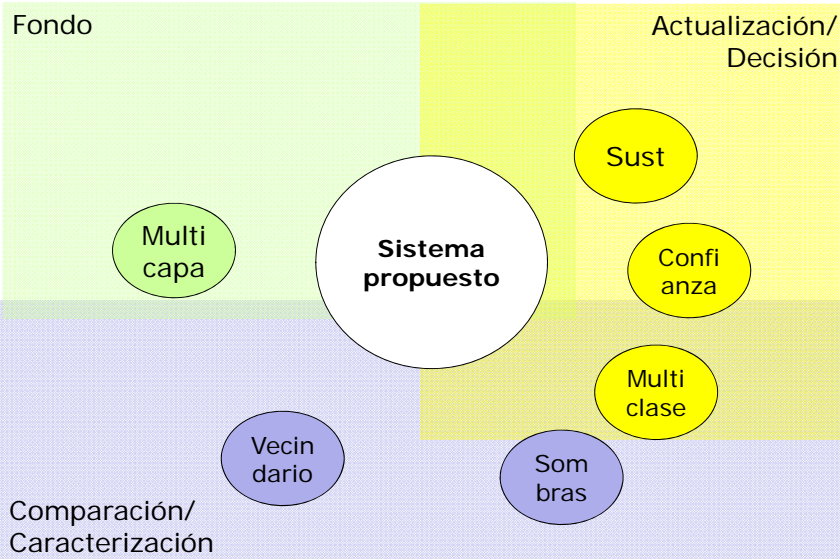
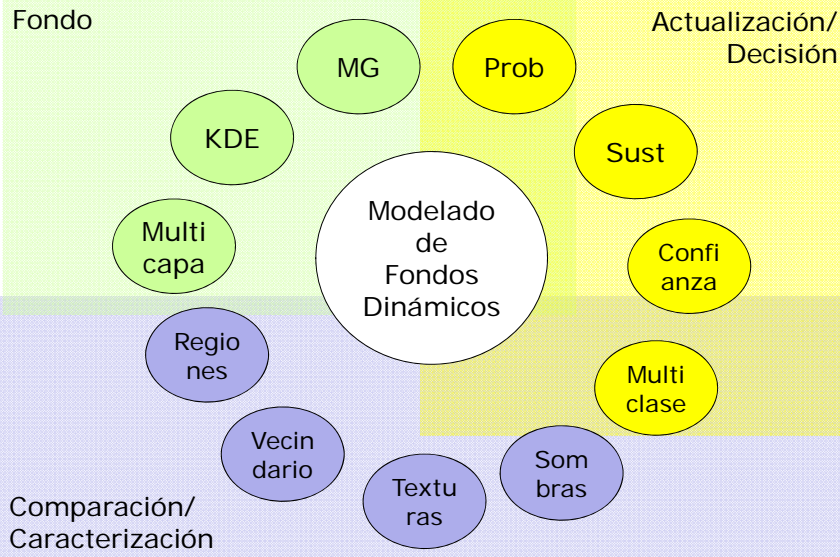
Principales problemas
Sombras



TEC2011-25995 EventVideo (2012-2014)

6/10

Segregación frente-fondo en entornos multimodales.



Interfaz

Factor k
(por pixel / por capa)

- Define la máxima y mínima permisividad del sistema
- La permisividad controla los procesos de actualización

Número de capas

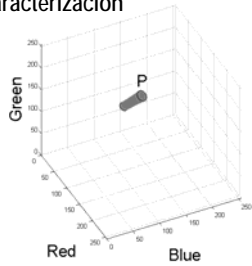
- Decide el grado de multi-modalidad soportado
- Almacena cambios temporales del fondo

The screenshot shows a software window titled 'multi-modal environments'. It contains several control panels:

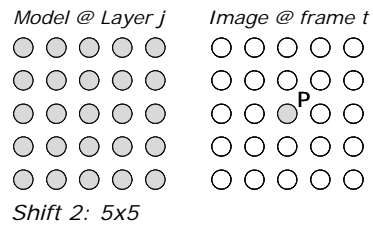
- Factor k:** A panel with 'Minimum Confidence' and 'Maximum Confidence Factor (k)' sliders. A red box highlights these sliders, with a red dashed box containing text about confidence limits and their effects on updates and background selection.
- Número de capas:** A panel with 'Number of Models' and 'k factor' sliders. A green box highlights these, with a green dashed box containing text about modal support and temporal storage.
- Factor shift:** A panel with 'Shift Number' and 'threshold' sliders. A blue box highlights these, with a blue dashed box containing text about neighborhood definition and robustness to noise.
- Altura del tronco de cono:** A panel with 'Lower' and 'Upper' sliders. A yellow box highlights these, with a yellow dashed box containing text about distance correction space and robustness to shadows.

- [1] Stauffer, C. & Grimson, W. Adaptive Background Mixture Models for Real-Time Tracking Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, IEEE Computer Society, 1999, 2, 2246.
- [2] Elgammal, A.; Duraiswami, R.; Harwood, D. & Davis, L. Background and foreground modelling using nonparametric kernel density estimation for visual surveillance Proceedings of the IEEE, 2002, 90, 1151 – 1163.
- [3] Goyette, N.; Jodoin, P.; Porikli, F.; Konrad, J. & Ishwar, P. Changedetection.net: A new change detection benchmark dataset. Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on, 2012, 1, 1-8.
- [4] Piccardi, M. Background subtraction techniques: a review. Systems, Man and Cybernetics, 2004 IEEE International Conference on, 2004, 4, 3099 – 3104.
- [5] Evangelio, R.; Patzold, M. & Sikora, T. Splitting Gaussians in Mixture Models Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on, 2012, 300 -305.
- [6] Evangelio, R. & Sikora, T. Complementary background models for the detection of static and moving objects in crowded environments Advanced Video and Signal Based Surveillance (AVSS), 2011 8th IEEE International Conference on, 2011, 71 -76.
- [7] Morde, A.; Ma, X. & Guler, S. Learning a background model for change detection Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on, 2012, 15 -20.
- [8] Cavallaro, A. & Ebrahimi, T. Video object extraction based on adaptive background and statistical change detection 2000, 465-475.
- [9] Colmenarejo, A.; Escudero-Viñolo, M. & Bescós, J. Class-driven Bayesian background modelling for video object segmentation Electronics Letters, 2011, 47, 1023 -1024.
- [10] Hofmann, M.; Tiefenbacher, P. & Rigoll, G. Background segmentation with feedback: The Pixel-Based Adaptive Segmenter Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on, 2012, 38 -43.
- [11] Schick, A.; Bauml, M. & Stiefelwagen, R. Improving foreground segmentations with probabilistic superpixel Markov random fields Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on, 2012, 27 -31.
- [12] Maddalena, L. & Petrosino, A. The SOBS algorithm: What are the limits? Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on, 2012, 21 -26.
- [13] Van Droogenbroeck, M. & Paquot, O. Background subtraction: Experiments and improvements for ViBe Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on, 2012, 32 -37.
- [14] M Escudero-Viñolo, J Bescós. A robust framework for region based video object segmentation. Image Processing (ICIP 2010), 17th IEEE International Conference on, 2010. 3461-3464.

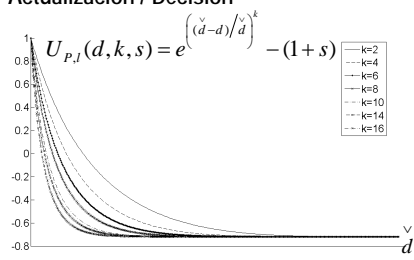
Caracterización



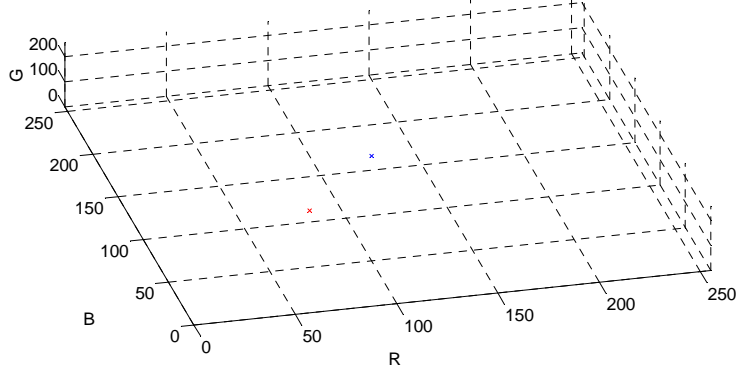
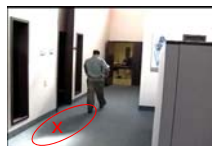
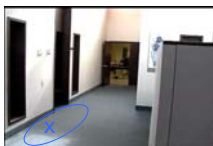
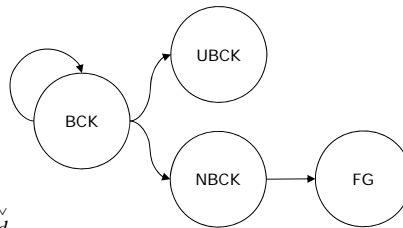
Comparación



Actualización / Decisión



$C_{p,l}^{t+1} = C_{p,l}^t + U_{p,l}(d, k, s)$



Strategies for Object Segmentation, Detection and Tracking in Complex Environments for Event Detection in Video Surveillance and Monitoring
TEC2011-25995 EventVideo (2012-2014)

Sesión 2: SEGMENTACIÓN Segmentación con cámaras PTZ

Alberto Palero Almazán
alberto.palero@uam.es
Jesús Bescós Cano
j.bescos@uam.es



UNIVERSIDAD AUTÓNOMA DE MADRID

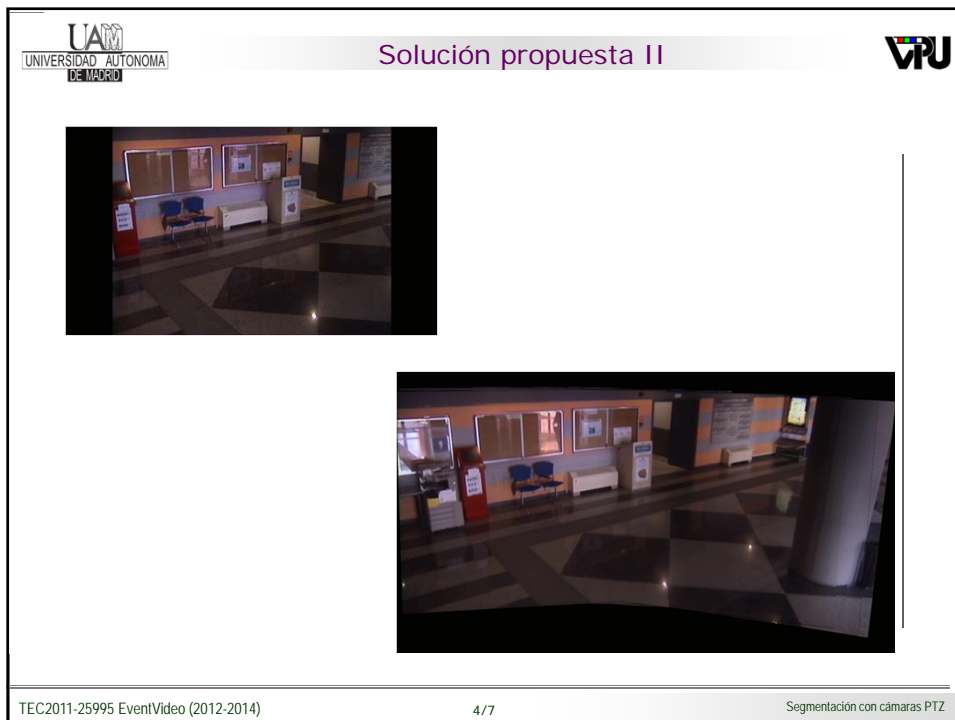
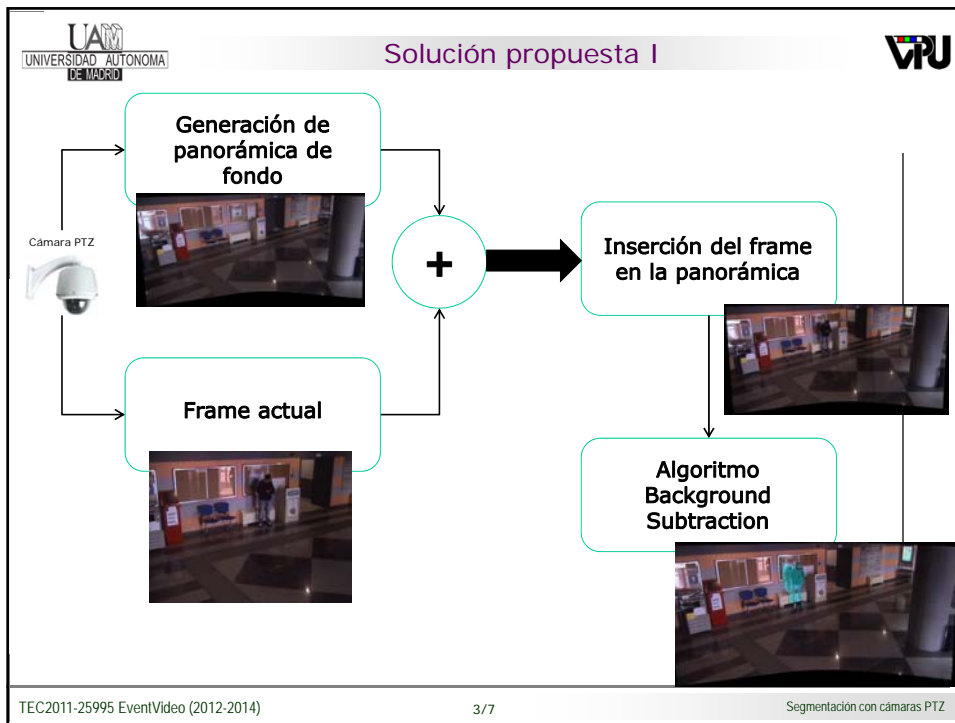
Cámaras PTZ vs Segmentación

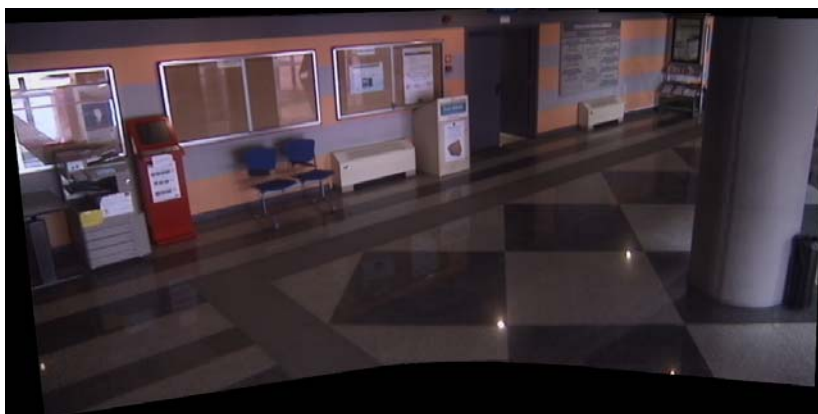
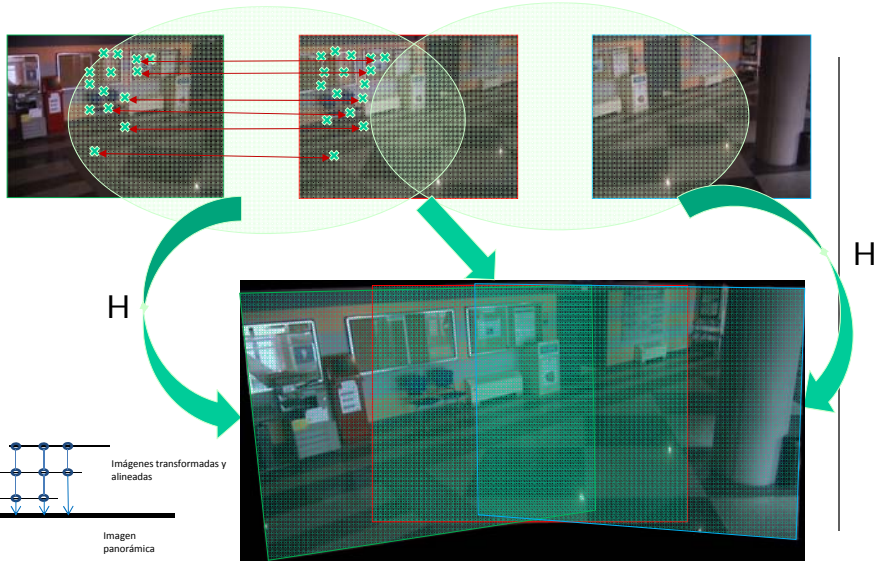
VPU

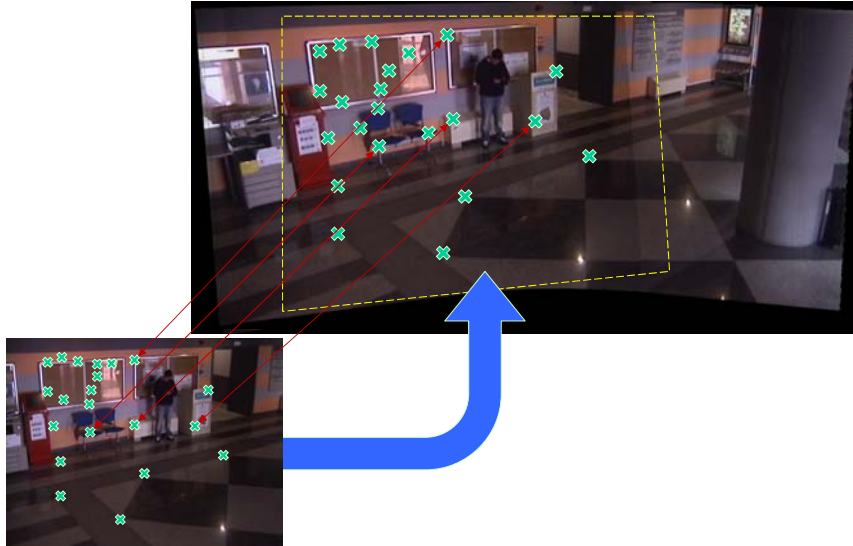
Cámara PTZ

The diagram illustrates the concept of PTZ vs Segmentation. It shows a PTZ camera (Pan-Tilt-Zoom) pointing to three different views of a hallway. The top view is a wide shot of the hallway. The middle view is a closer shot of the hallway. The bottom view is a closer shot of the hallway, showing a person sitting on a bench. This illustrates how a PTZ camera can capture different views of the same scene, which is then used for segmentation.

TEC2011-25995 EventVideo (2012-2014) 2/7 Segmentación con cámaras PTZ







Strategies for Object Segmentation, Detection and Tracking in Complex Environments for Event Detection in Video Surveillance and Monitoring
TEC2011-25995 EventVideo (2012-2014)

Sesión 3: DETECCIÓN DE PERSONAS

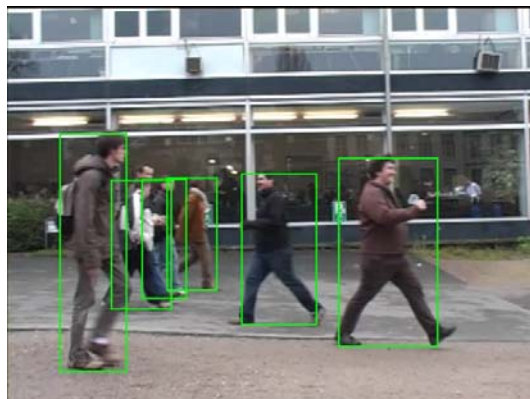
Patricia Marín
Ricardo Sánchez
Álvaro García
alvaro.garcia@uam.es



Detección de personas

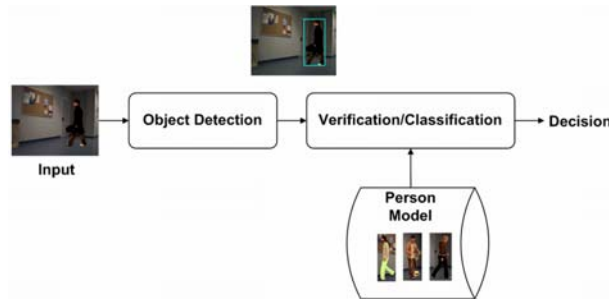


- ¿Qué es la detección de personas?



■ Detección personas

1. Definir/entrenar modelo persona
2. Detección de objetos
3. Calsificación objetos (persona/no persona)



■ Detección/Extracción objetos

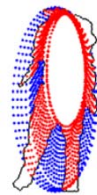
- Segmentación
- Búsqueda exhaustiva

■ Modelo persona

- Apariencia
- Movimiento
- Combinación

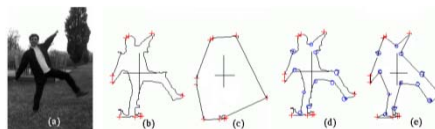
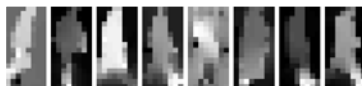
Ellipse fitting

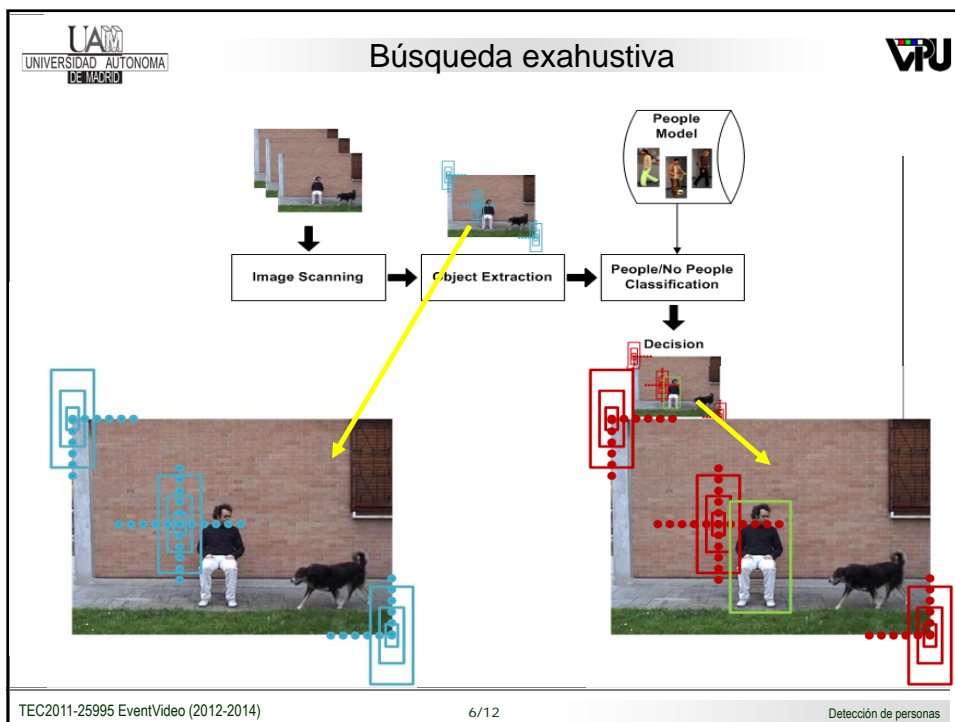
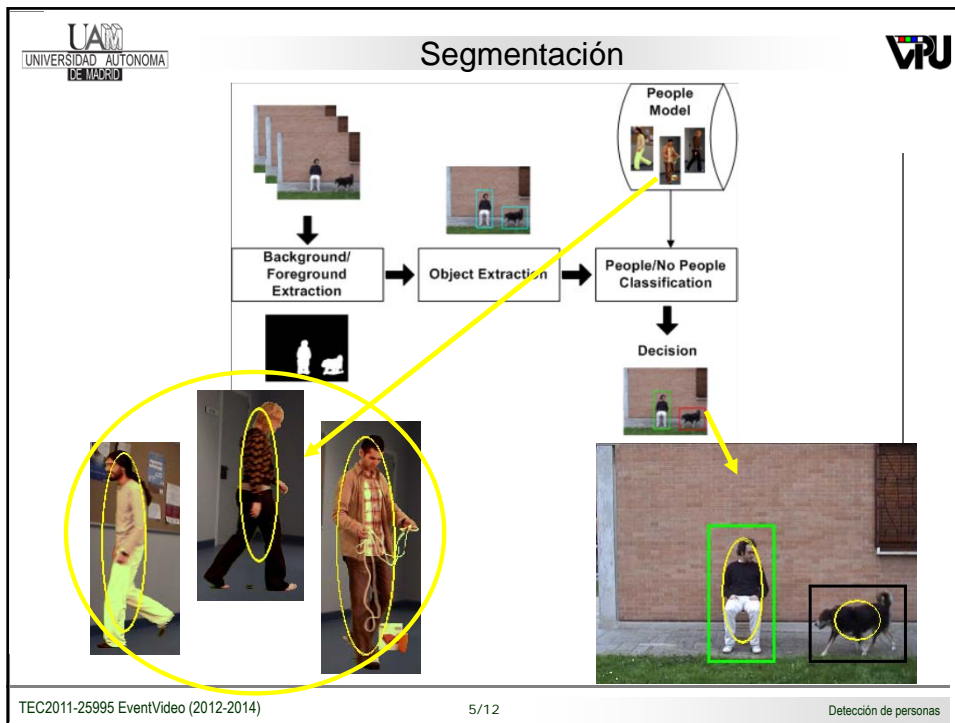
Silhouette matching

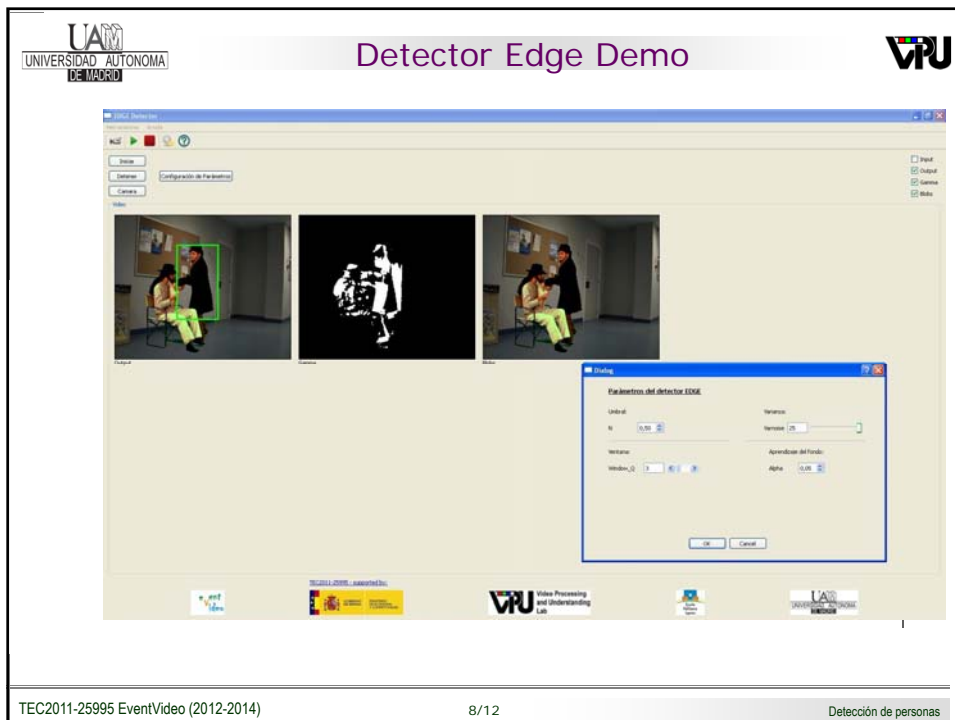
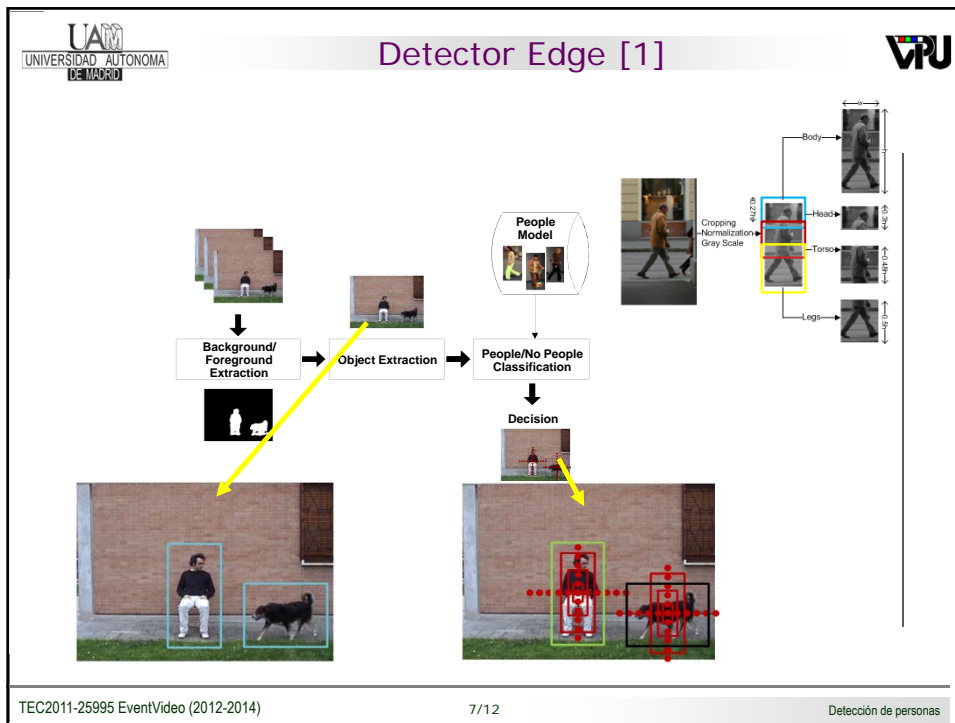


Motion patterns

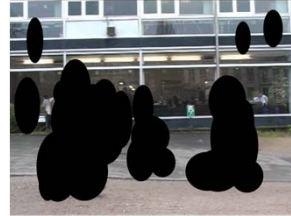
Ghost representation







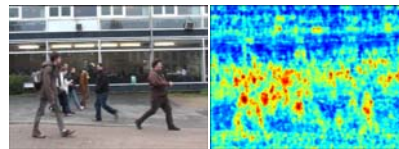
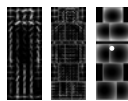
- Objetivo: Ninguna persona (o parte del cuerpo) es clasificada como fondo



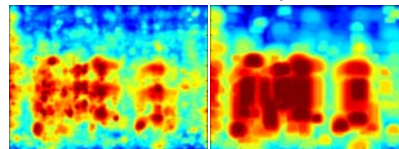
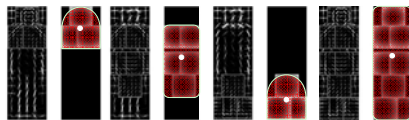
- Aplicaciones:

- Pre/Post-procesado:
 - Detección de personas
 - Tracking
 - Segmentación
- Estimación densidad de personas

Algoritmo original [3]

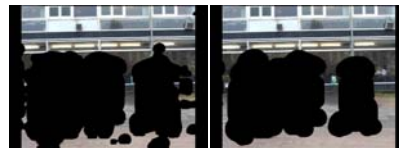


Mapas confianza DBP/DEBP (Kernel de extensión)



Segmentación DEBP-P

- Binarización and Post-procesado





- [1] A. Garcia-Martin; J.M. Martinez, "Robust Real Time Moving People Detection in Surveillance Scenarios," *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, vol., no., pp.241,247, Aug. 29 2010-Sept. 1 2010.
- [2] A. Garcia-Martin; A. Cavallaro; J:M. Martinez, "People-background segmentation with unequal error cost," *Image Processing (ICIP), 2012 19th IEEE International Conference on*, vol., no., pp.157,160, Sept. 30 2012-Oct. 3 2012.
- [3] P. Felzenszwalb; R. Girshick; D. McAllester; D. Ramanan, "Object Detection with Discriminatively Trained Part Based Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 9, Sep. 2010.

Strategies for Object Segmentation, Detection and Tracking in Complex Environments for Event Detection in Video Surveillance and Monitoring
TEC2011-25995 EventVideo (2012-2014)

Sesión 4: SEGUIMIENTO

Seguimiento de objetos a largo plazo

Jorge Sanjuan Garcia
Rafael Martín Nieto
{jorge.sanjuan, rafael.martinn}@uam.es



Introducción



■ Video Object Tracking

- Localizar un objeto a lo largo del tiempo usando una cámara (o varias)



■ Video Object Tracking

- Localizar un objeto a lo largo del tiempo usando una cámara (o varias)

- Aplicaciones:

- Seguridad y vigilancia
- Control de tráfico
- Edición de video



- Problemas:

- Movimientos complejos
- Cambios de iluminación
 - Local y global
- Ruido
- Oclusiones
- Cambios de escala
- Objetos similares



■ Originalmente diseñado para seguimiento de profesores durante una clase

- Ejecución en tiempo real
- Seguimiento a largo plazo (long-term)
- Mono objetivo
- Capacidad de recuperar el objeto (persona)
 - En caso de salir y entrar en la escena

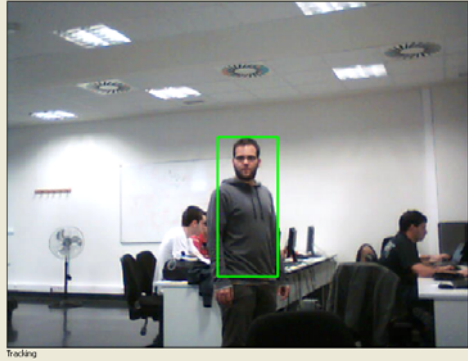
■ Técnica utilizada:

- Seguimiento basado en puntos (KLT)
- Refinamiento de la posición mediante búsqueda de color (Mean Shift)

UNIVERSIDAD AUTÓNOMA DE MADRID Demo VPU

GUI_PKLIFilter_Tracker

Área



Tracking

Iniciar




Detener

Camara

Configuración

TEC2011-25995 - supported by:



TEC2011-25995 EventVideo (2012-2014)
5/7
Seguimiento de objetos a largo plazo

UNIVERSIDAD AUTÓNOMA DE MADRID Demo VPU

ventanaConfiguracion

Show Features

Object Recovering:

Activated

Frames without features:

Threshold: [Slider]

Feature detection filtering (Fetaure Motion):

FM Threshold Down:

FM Threshold Up:

Color model configuration for Gradient Ascent:

Min Samples:

Mean-Shift configuration:

Iterations:

Stadistical filtering:

Stadistical distance filtering respect object: [Slider]

TEC2011-25995 EventVideo (2012-2014)
6/7
Seguimiento de objetos a largo plazo

- Antonio González Huete, Jesús Bescos Cano: Seguimiento y producción automática mediante cámaras PTZ en entornos de red. Trabajo de Fin de Máster, Septiembre 2013.

Strategies for Object Segmentation, Detection and Tracking in Complex Environments for Event Detection in Video Surveillance and Monitoring
TEC2011-25995 EventVideo (2012-2014)

Sesión 4: SEGUIMIENTO

Seguimiento basado en atención visual

Victor Fernández
victor.fernandez@uam.es

Miguel Angel García
miguelangel.garcia@uam.es

José M. Martínez
josem.martinez@uam.es



UAM UNIVERSIDAD AUTÓNOMA DE MADRID

Introducción

VPU

■ Vídeo-seguimiento

- Determinar automáticamente la trayectoria de objetos en movimiento a lo largo de las imágenes (cuadros) de una secuencia de vídeo.
- Asociar objetos identificados en una imagen con objetos identificados en imágenes previas.

t $t + N$

TEC2011-25995 EventVideo (2012-2014) 2/24 Seguimiento basado en atención visual

■ Vídeo-seguimiento

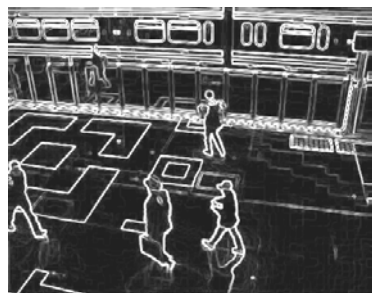
- Objetos = blobs → Segmentación frente / fondo.



- Blob:
 - Geometría: área, centroide, ancho, alto.
 - Imagen original.

■ Vídeo-seguimiento basado en atención visual

- Encontrar características visuales distintivas (**puntos de atención de foco**) que ayuden a asociar blobs entre distintas imágenes.



Ej.: características de orientación (bordes)

■ Atención visual basada en prominencia (*saliency*)

- Obtener una **imagen de prominencia** que indica la prominencia visual de cada pixel de una imagen de entrada.



Imagen de entrada

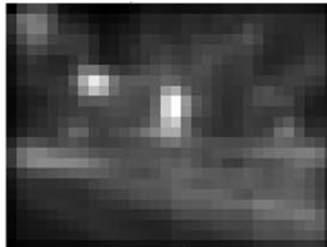


Imagen de prominencia

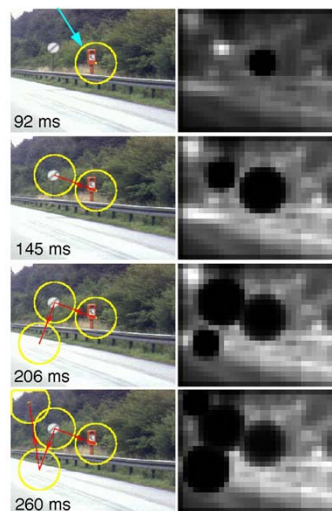
Imágenes extraídas de "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", Itti, Koch, Niebur, PAMI 98

■ Atención visual basada en prominencia

- Secuencia de Puntos de Atención de Foco:
 - (1) PAF → Pixel de mayor prominencia.
 - (2) Inhibición local de prominencia alrededor de PAF actual.
 - (3) Ir a (1) hasta número máximo de PAFs o prominencia por debajo de umbral.



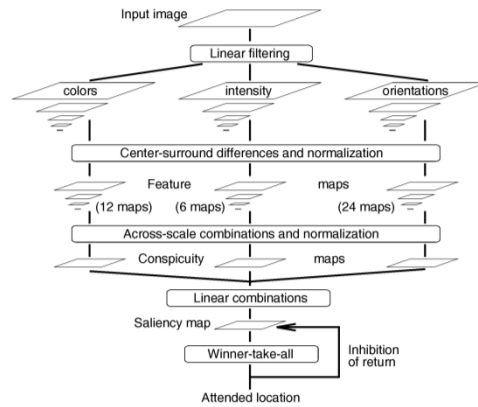
Imagen de entrada



Imágenes extraídas de "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", Itti, Koch, Niebur, PAMI 98

■ Atención visual basada en prominencia

- "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", Itti, Koch, Niebur, PAMI 98.

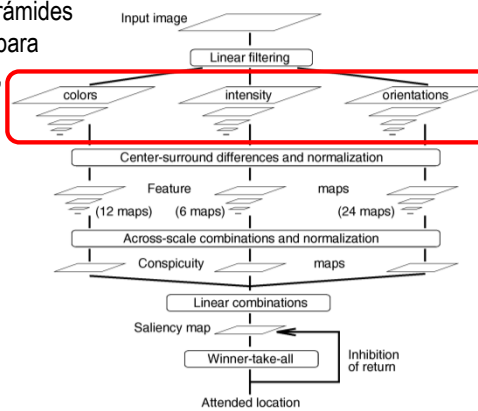


Imágenes extraídas de "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", Itti, Koch, Niebur, PAMI 98

■ Atención visual basada en prominencia

- "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", Itti, Koch, Niebur, PAMI 98.

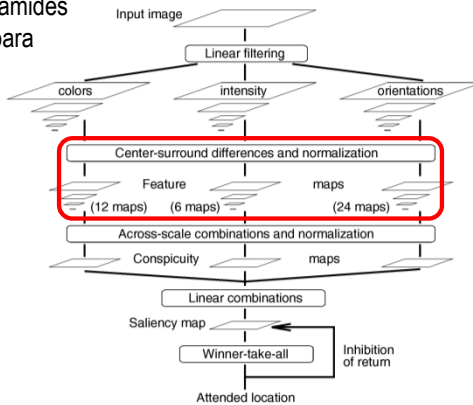
- Análisis multiescala ➡ Pirámides Gaussianas independientes para cada modalidad visual (color, intensidad, orientaciones)



Imágenes extraídas de "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", Itti, Koch, Niebur, PAMI 98

Atención visual basada en prominencia

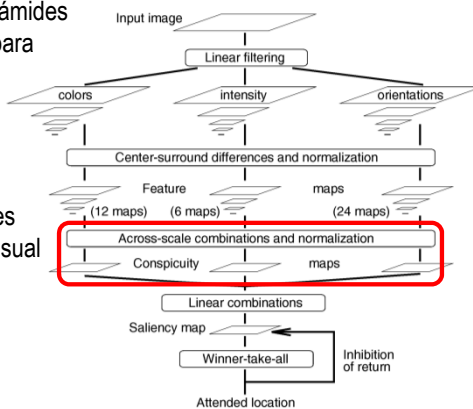
- "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", Itti, Koch, Niebur, PAMI 98.
- Análisis multiescala → Pirámides Gaussianas independientes para cada modalidad visual (color, intensidad, orientaciones)
- Detección de bordes → Diferencias centro-periferia.



Imágenes extraídas de "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", Itti, Koch, Niebur, PAMI 98

Atención visual basada en prominencia

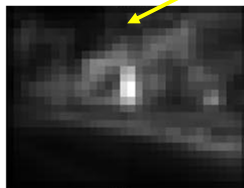
- "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", Itti, Koch, Niebur, PAMI 98.
- Análisis multiescala → Pirámides Gaussianas independientes para cada modalidad visual (color, intensidad, orientaciones)
- Detección de bordes → Diferencias centro-periferia.
- Fusión de bordes de diferentes escalas en cada modalidad visual → mapas de visibilidad (conspicuity maps).



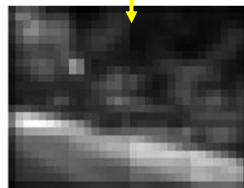
Imágenes extraídas de "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", Itti, Koch, Niebur, PAMI 98

■ Atención visual basada en prominencia

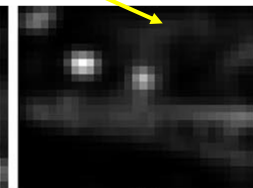
- Mapas de visibilidad \equiv mapas de prominencia para cada modalidad visual.



Color



Intensidad

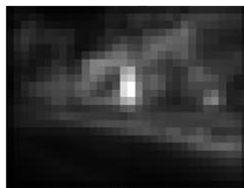


Orientaciones

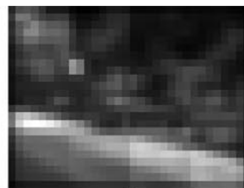
Imágenes extraídas de "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", Itti, Koch, Niebur, PAMI 98

■ Atención visual basada en prominencia

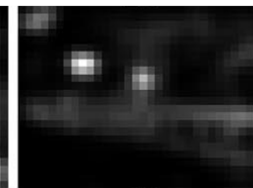
- Imagen de prominencia generada promediando mapas de visibilidad.



Color



Intensidad



Orientaciones

Imágenes extraídas de "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", Itti, Koch, Niebur, PAMI 98

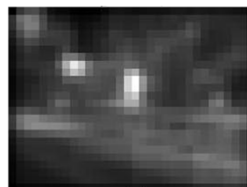
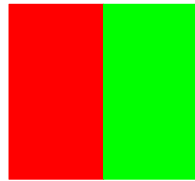


Imagen de prominencia

■ Atención visual basada en prominencia

- Canales de color e intensidad son procesados independientemente.
- Problema ➡ Percepción de intensidad y color no está desacoplada.
- Ejemplo: igual color (Matiz, Saturación), diferente intensidad (Luminosidad)

Alta diferencia perceptual



(0, 255, 128) (85, 255, 128)

Baja diferencia perceptual

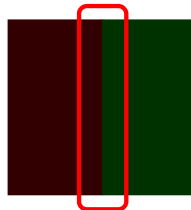


(0, 25, 25) (85, 255, 25)

- Misma prominencia de color elevada en la frontera entre ambas regiones.

■ Atención visual basada en prominencia

- Canales de color e intensidad son procesados independientemente.
- Problema ➡ Percepción de intensidad y color no está desacoplada.
- Sistema de seguimiento focaliza erróneamente en características visuales no distintivas visualmente, aunque tengan elevada prominencia de color.



- Negativo para video-vigilancia ➡ objetos oscuros, escenarios pobremente iluminados...
- Solución ➡ Procesar color e intensidad conjuntamente.

■ Espacio perceptual conjunto de color e intensidad

- Modelo **teórico** bio-inspirado propuesto por Izmailov y Sokolov (1991) en el ámbito de la Psicofísica.
- Definido experimentalmente a partir de diferencias de color observadas por sujetos humanos y análisis de escalado multidimensional.
- Espacio métrico de color:
 - Espacio cromático 3D (X_1, X_2, X_3)
 - Espacio acromático 2D (Y_1, Y_2)
 - Diferencia cromática entre dos puntos 3D $\rightarrow \Delta C_{ij}$
 - Diferencia acromática entre dos puntos 2D $\rightarrow \Delta W_{ij}$
- Diferencia perceptual conjunta entre dos colores $\rightarrow \Delta S_{ij}$

$$(\Delta S_{ij})^2 = (\Delta C_{ij})^2 + X_3^i X_3^j (\Delta W_{ij})^2$$

■ Espacio perceptual conjunto de color e intensidad

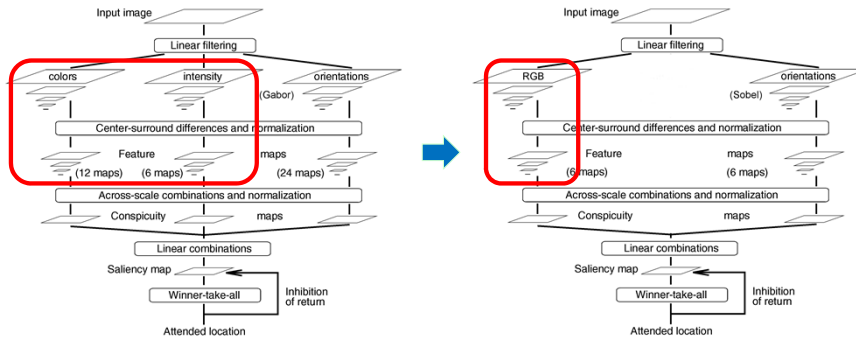
- Modelo **teórico** bio-inspirado propuesto por Izmailov y Sokolov (1991) en el ámbito de la Psicofísica.
- Definido experimentalmente a partir de diferencias de color observadas por sujetos humanos y análisis de escalado multidimensional.
- Espacio métrico de color:
 - Espacio cromático 3D (X_1, X_2, X_3)
 - Espacio acromático 2D (Y_1, Y_2)
 - Diferencia cromática entre dos puntos 3D $\rightarrow \Delta C_{ij}$
 - Diferencia acromática entre dos puntos 2D $\rightarrow \Delta W_{ij}$
- Diferencia perceptual conjunta entre dos colores $\rightarrow \Delta S_{ij}$

■ Modelo computacional para modelo teórico de Izmailov y Sokolov

- Mapeo analítico desde el espacio RGB al espacio $(X_1, X_2, X_3, Y_1, Y_2)$.
- Modelo perceptual de Izmailov y Sokolov aplicable a imágenes de color.

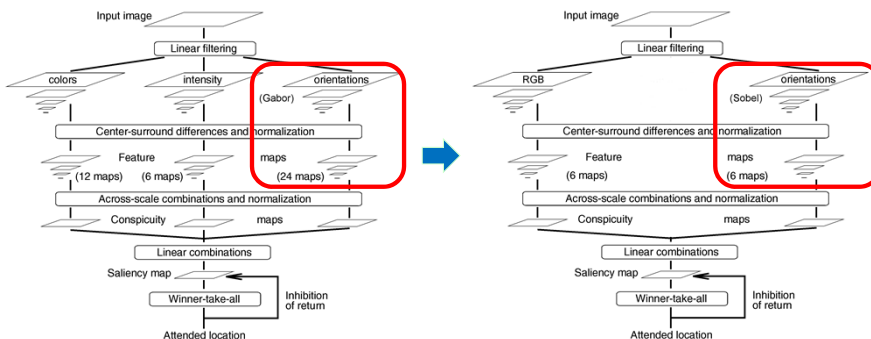
Adaptación del modelo de atención visual basado en prominencia

- Procesamiento conjunto de color e intensidad.
- Aplicación de modelo computacional propuesto a imágenes RGB.
- Diferencias perceptuales con modelo de Izmailov y Sokolov.



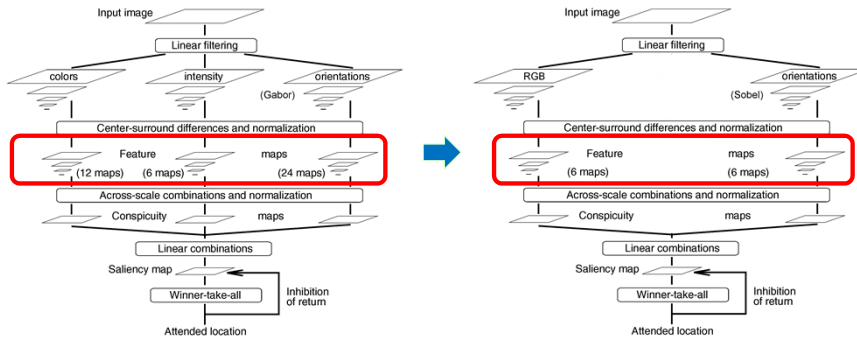
Adaptación del modelo de atención visual basado en prominencia

- Orientaciones con detector de Sobel en lugar de banco de filtros de Gabor (4 filtros).



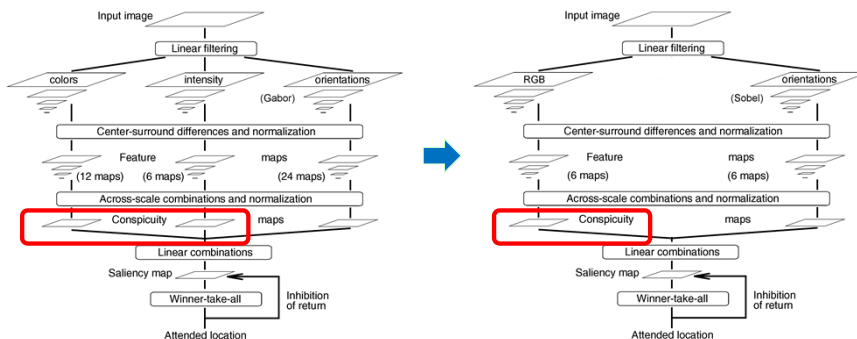
Adaptación del modelo de atención visual basado en prominencia

- Menos mapas de características: 42 ➡ 12.



Adaptación del modelo de atención visual basado en prominencia

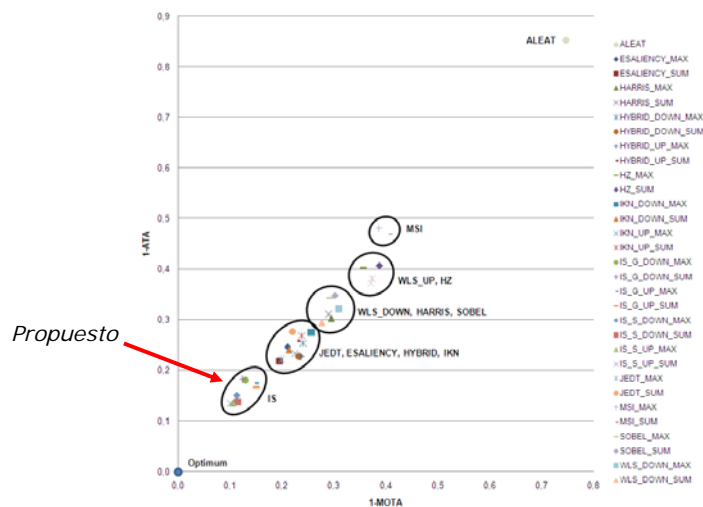
- Mapa de visibilidad (prominencia) conjunto color / intensidad.



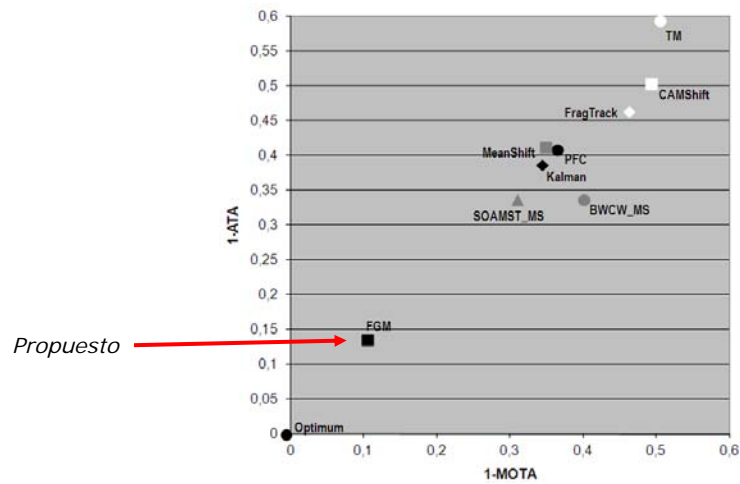
■ Esquema simple de seguimiento de blobs

- Lista de blobs previos seguidos durante las ultimas N imágenes (ej.: N=25).
- Para cada nuevo blob:
 - Centrar ventanas en los principales PAFs dentro del blob (excluir borde del blob).
 - Reconocimiento de bloques (Suma de Diferencias Absolutas) entre ventanas de PAFs y cada blob previo.
 - Promediar SDAs para cada blob previo.
- Ordenar emparejamientos (nuevo blob, blob previo) en orden ascendente de SDA promedio.
- Establecer asociaciones (nuevo blob, blob previo) recorriendo la lista ordenada desde su cabeza.
- Nuevos blobs con SDA promedio mínimo por encima de umbral etiquetados como nuevos objetos.

■ Rendimiento para diferentes modelos de atención visual



■ Rendimiento para diferentes algoritmos de vídeo-seguimiento



■ Experimentos con el conjunto de datos PETS 2006



Strategies for Object Segmentation, Detection and Tracking in Complex Environments for Event Detection in Video Surveillance and Monitoring
TEC2011-25995 EventVideo (2012-2014)

Sesión 5: DETECCIÓN DE EVENTOS

Actividades e interacciones de personas

Juan Carlos San Miguel
juancarlos.sanmiguel@uam.es

José M. Martínez
josem.martinez@uam.es



Introducción



■ Interacciones y actividades de personas

- Interacción: acción realizada por dos personas o entre persona y objeto
- Actividad: acción realizada por una persona

■ Entorno controlado

- Conocimiento a priori de objetos, eventos, contexto espacial, ...

KTH dataset



HERMES dataset



■ Limitaciones actuales de la detección de actividades e interacciones

- Se suelen usar secuencias de prueba con una única persona
 - Evitando los efectos causados por la presencia de otras personas u objetos
- Se asume que los datos de análisis son completos y fiables
 - Evitando los efectos de oclusiones, fallos (intermitentes) de los algoritmos, ...
- Se asume aprendizaje de modelos sobre grandes bases de datos
 - Exigiendo grandes bases de datos de entrenamiento, capacidad de proceso, ...
 - Creando modelos sobrentrenados y no pudiendo detectar acciones no presentes en los datos de entrenamiento
- No existe un marco de referencia global para modelar el contexto
 - Impidiendo la aplicación en entornos distintos al de entrenamiento
- No permiten detección en tiempo real
 - Aproximaciones de alta complejidad

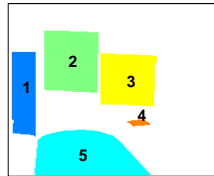
■ Utilización de un modelo para la representación formal del contexto

- Incorporando los diferentes tipos de contexto
- Mejora la tasa de detección
- Permite que los algoritmos se adapten al contexto
 - Frente a soluciones ad-hoc para cada contexto

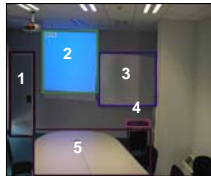
■ Utilización de un modelo de eventos híbrido (determinístico y probabilístico)

- D: modelado de eventos semántico
 - Evita los datos de entrenamiento
- P: modelado de eventos Bayesiano
 - Elimina la incertidumbre tras el análisis basado en el modelo determinístico

Contexto espacial



Objetos (Blobs)



Objetos (Bounding Boxes)

Categorías de objetos

- 1 Obj\ContextObject\FixedObject\Door
- 2 Obj\ContextObject\FixedObject\ProjectionA
- 3 Obj\ContextObject\FixedObject\Blackboard
- 4 Obj\ContextObject\MovableObject\Table
- 5 Obj\ContextObject\FixedObject\Table

Contexto de eventos

- Eventos mutuamente exclusivos
- Orden en la ocurrencia de eventos

Contexto de objetos



Persona Mesa auxiliar



Relación de tamaño

Umbrales para detección de agrupamiento

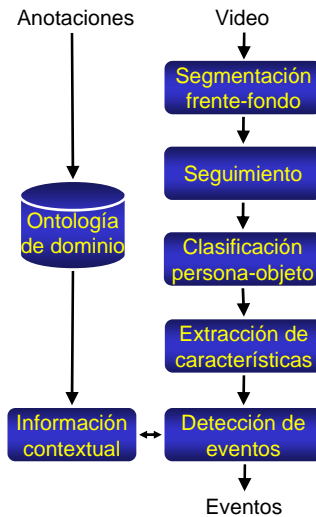
Sistema de análisis 2D en tiempo real

- Segmentación frente-fondo
- Seguimiento de blobs
 - Regiones conexas de frente
- Clasificación persona-objeto
- Detección de eventos
 - Basada en características extraídas de objetos de frente y contextuales
 - Modelos de eventos

utilizando información contextual

- Ontología de dominio
 - A partir de modelos de contexto
 - Anotación para cada entorno
 - Escenario, objetos, eventos, ...

para adaptarse a diversos entornos

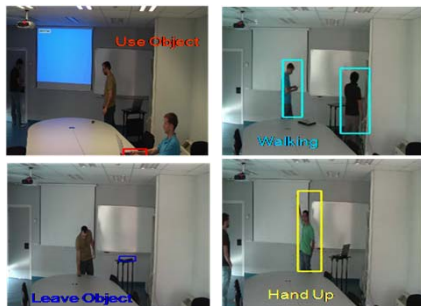


■ Detección de eventos

- Datos de entrada: Blobs y sus características y objetos contextuales
- Tipos de eventos modelados y modelos:

<p>Interacciones persona-objeto/persona</p> <ul style="list-style-type: none"> - Condiciones (C) entre características de blobs y objetos contextuales - Combinación Bayesiana <p>GetObject</p> <ul style="list-style-type: none"> •C1: Blob appears now •C2: Blob belongs to background •C3: Blob classified as object •C4: There is an associated cont. object •C5: A person is doing the action •C6: Distance person-object less than th 	<p>Actividades de personas (andar, levantar la mano, ...)</p> <ul style="list-style-type: none"> - Evolución temporal y relaciones de características (espaciales) de blobs tipo persona  <p>Skin Areas</p> <p>Legs Mass center</p>	<p>Estado (presencia, conteo, ...)</p> <ul style="list-style-type: none"> - Maquina de Estados Finitos (<i>Finite State Machine</i>) - Media temporal para filtrar espúreos  <p>$F > \beta$ $F < \alpha$ $F < \alpha$</p> <p>Presence No Presence</p> <p>$F > \beta$</p> <p>$F \rightarrow$ person exists in the last N frames</p>
---	--	---

- Interacciones persona-objeto (dejar, coger, abrir puerta, usar,...)
- Interacciones persona-persona (darse la mano, conversación,...)
- Actividades (andar, levantar la mano,...)



Examples of the events proposed in the LIRIS dataset

- ICPR-HARL 2012 challenge
- 828 acciones in total
 - D1-depth+color (461)
 - D2-color (367)
- 70 participants
(only 4 submitted results)

Nr.	Team	Dataset	Recall	Precision	F-Score
49	ADSC-NUS-UIUC	D1	0.74	0.41	0.53
59	TATA-ISI	D1	0.08	0.17	0.11
13	VPULABUAM	D2	0.36	0.66	0.46
51	IACAS	D2	0.30	0.46	0.36

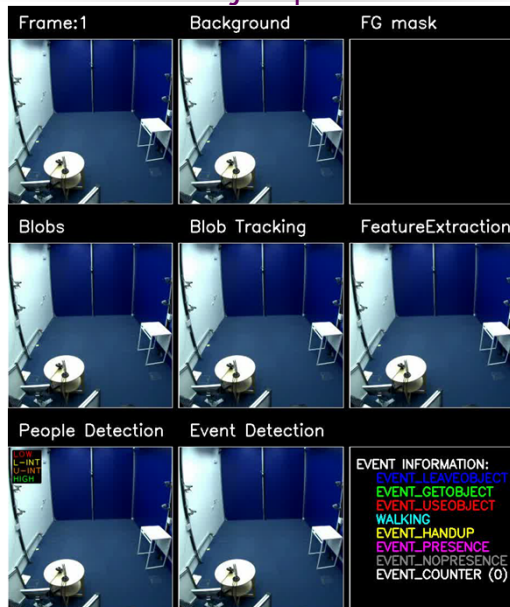
- VPULab único en tiempo real



Ejemplos



Ejemplos



- SanMiguel, J., Martínez J.M. and García. "An Ontology for Event Detection and its Application in Surveillance Video", in AVSS2009.
- SanMiguel, J.C. , Martínez J.M, "Dynamic video surveillance systems guided by domain ontologies", in ICDP 2009
- SanMiguel, J.C. Escudero, M., Martínez J.M. and J.C. Bescós, ., "Real-time single-view video event recognition in controlled environments", in CBMI 2011
- SanMiguel, J.C. and Martínez J.M, "A semantic-based probabilistic approach for real-time video event recognition", Computer Vision and Image Understanding, Sept. 2012,
- SanMiguel, J.C. and Suja S., "Skin detection by dual maximization of detectors agreement for video monitoring", Pattern Recognition Letters, December 2013

Strategies for Object Segmentation, Detection and Tracking in Complex Environments for Event Detection in Video Surveillance and Monitoring
TEC2011-25995 EventVideo (2012-2014)

Sesión 5: Detección de eventos Detección de regiones estáticas: robo y abandono

Diego Ortego Hernández, Carlos Sánchez Bueno, Juan Carlos San Miguel Avedillo

{diego.ortego, carlos.sanchezb, juancarlos.sanmiguel}@uam.es



Introducción



■ Motivación y objetivo

- La **detección de objetos estáticos** (SFO) en tareas de vídeo-vigilancia es una etapa importante en diversas aplicaciones...
- SFO: Todo objeto o persona que habiendo estado en movimiento, permanece inmóvil durante un cierto tiempo.
- Motivación → **Bajo rendimiento en entornos concurridos**
- Objetivo → Mejora estado del arte

Abandono



Robo



Estacionamiento ilegal



■ Idea base

- Etapa sustracción de fondo (BGS):
 - Imágenes → Frente (FG) y fondo (BG)
- Detección estática (SD):
 - FG → Máscara estática (aplicación de reglas)

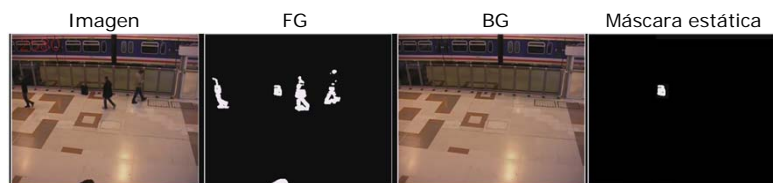
Imágenes



FG

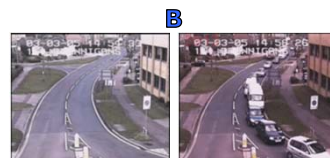


Máscara estática



■ Limitaciones

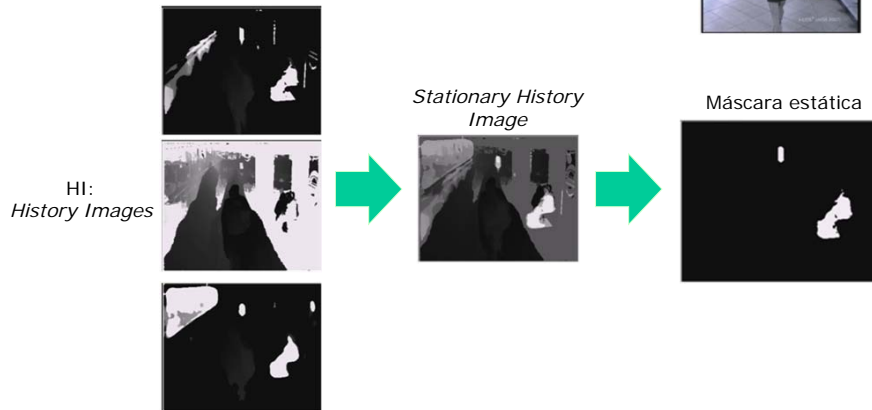
- Zonas elevado movimiento **A**
- Sombras y cambios de iluminación **B**
- Inicialización de BG **C**



■ Esquema multi-característica

- Robustez a problemática → Distinta información

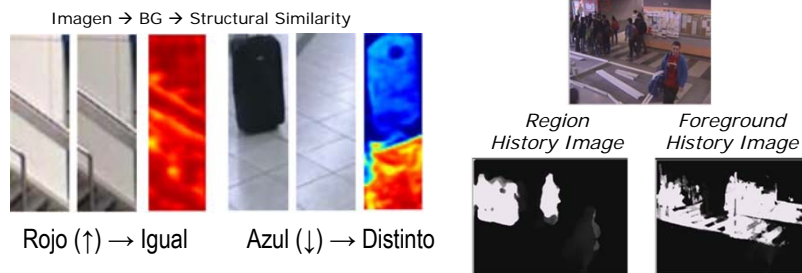
- Frente + Movimiento + Estructura
- Combinación y umbralizado



■ Detección de NO MOVIMIENTO en zonas en oclusión



■ Robustez frente a SOMBRAS y cambios de ILUMINACIÓN





- Parámetros
 - Configuración óptima
- Tiempo de alarma
 - Tiempo para considerar un objeto como estático

- D. Ortego and J.C. SanMiguel, "Stationary foreground detection for video-surveillance based on foreground and motion history images," in Proc. of IEEE Conf. Adv. Video Signal Based Surveill. (AVSS), Aug. 2013, pp. 75–80.
- D. Ortego and J.C SanMiguel, "Multi-Feature Stationary Foreground Detection for Crowded Video-Surveillance" in Proc. of 21st IEEE International Conference on Image Processing (ICIP 2014), Paris (France), Sep. 2014. (accepted May 2014).
- W. Hassan, P. Birch, B. Mitra, N. Bangalore, R. Young, and C. Chatwin, "Illumination invariant stationary object detection", IET Computer Vision, vol. 7, no. 1, pp. 1–8, Feb. 2013.
- Q. Fan and S. Pankanti, "Modeling of temporarily static objects for robust abandoned object detection in urban surveillance," in Proc. of IEEE Conf. Adv. Video Signal Based Surveill. (AVSS), Sep. 2011, pp. 36–41.
- A. Cavallaro, Steiger O., and Ebrahimi T., "Semantic video analysis for adaptive content delivery and automatic description," IEEE Trans. Circuits Syst. Video Technol., vol. 15, no. 10, pp. 1200–1209, Oct. 2005.
- H. R. Sheikh Z. Wang, A. C. Bovik and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Process., vol. 13, no. 4, pp. 600–612, Apr. 2004.

Strategies for Object Segmentation, Detection and Tracking in Complex Environments for Event Detection in Video Surveillance and Monitoring
TEC2011-25995 EventVideo (2012-2014)

Sesión 5: DETECCIÓN DE EVENTOS Detección de Anomalías

Ana Huélamo Vallejo
Luis Caro Campos

ana.huelamo@uam.es
luis.caro@uam.es



Introducción



■ Detección de movimientos anómalos



■ Detección de Anomalías

- Localizar movimiento que resulte anómalo respecto al movimiento normal
- Aplicaciones:
 - Seguridad y vigilancia
 - Control de tráfico (anomalías = *infracciones*)
- Problemas:
 - La anomalía puede no ser de interés
 - El algoritmo sólo es capaz de alertar que ha ocurrido movimiento anómalo
 - Sólo aplicable a escenarios donde el movimiento normal sigue unos patrones
 - La detección depende de las características/atributos extraídos

■ Anomalías en función de tamaño y velocidad de objetos en movimiento

- Es posible detectar las siguientes anomalías:
 - Movimiento de objetos de tamaño inusual (ej: muy grandes)
 - Movimiento de objetos normales a velocidades inusuales (muy lentos)
 - Movimiento de cualquier objeto en zonas inusuales o prohibidas
- El objetivo es descartar todo el movimiento normal

Técnica utilizada:

- Se crea una "imagen de comportamiento" para cada escena que representa el movimiento normal a partir del tamaño (explícito) y la velocidad (implícita) de los objetos en movimiento
- El movimiento que no encaje en este modelo, se considera anomalía

[1] P. M. Jodoin, V. Saligrama, and J. Konrad, "Behavior Subtraction," *Image Processing, IEEE Transactions on*, vol. 21, no. 9, pp. 4244–4255, 2012.

Video frame I_t



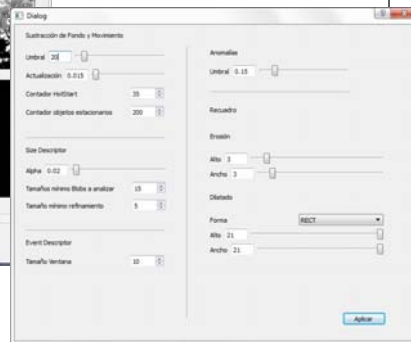
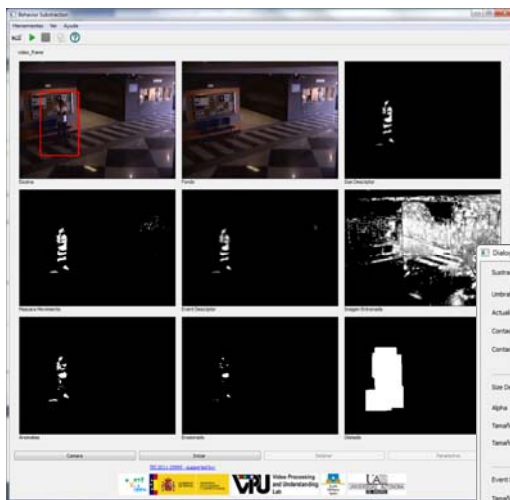
Motion label field I_t



B image



Anomaly map



■ Referencias

Background subtraction

- [1] M. Piccardi, "Background subtraction techniques: a review," in *Systems, Man and Cybernetics*, 2004 IEEE International Conference on, pp. 3099–3104, 2004.

Anomaly detection / behavior subtraction

- [1] P. M. Jodoin, V. Saligrama, and J. Konrad, "Behavior Subtraction," *Image Processing, IEEE Transactions on*, vol. 21, no. 9, pp. 4244–4255, 2012.