

Detección de eventos en secuencias con multitudes

Fernando de la Calle Silos

24 de septiembre de 2012

Outline

- 1 Introducción
- 2 Detección
- 3 Estimación del movimiento
- 4 Reconocimiento de eventos
- 5 Resultados
- 6 Conclusiones

Objetivos

Sistema de reconocimiento automático de eventos anómalos en secuencias de vídeo con multitudes. Utilizando la base de datos PETS 2010 [5].

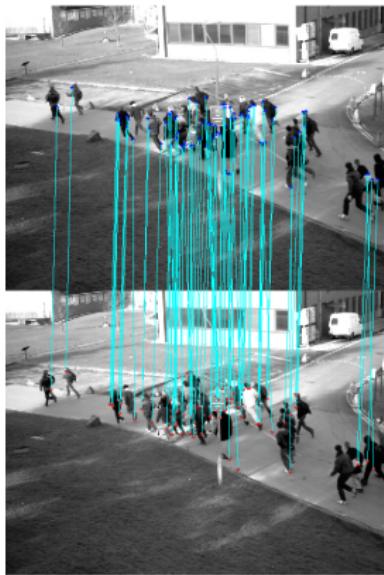
Eventos:

- *Walking*
- *Running*
- *Evacuation*
- *Crowd Formation*
- *Crowd Splitting*
- *Local Dispersion*

División del sistema

Tres etapas:

- ① Detección de objetos de interés (Foreground).
- ② Estimación del movimiento (Tracking).
- ③ Reconocimiento de eventos.



Outline

1 Introducción

2 Detección

- Resta del plano anterior
- Modelo de mezcla de Gaussianas
- Combinación de los métodos

3 Estimación del movimiento

4 Reconocimiento de eventos

5 Resultados

6 Conclusiones

Sistema de extracción de fondo

Dos fases:

- ① Modelado del fondo.
- ② Detección de objetos de interés.

Dos tipos de modelado del fondo:

- ① Modelos de actualización temporal. El modelo del fondo se calcula a partir de planos anteriores. Necesaria la existencia de movimiento.
- ② Modelos entrenados. El modelo del fondo se calcula a partir de datos de entrenamiento donde no estén presentes los objetos de interés.

El sistema implementado combina estos dos tipos de modelos.

Resta del plano anterior

- 1 Como fondo estimado se toma el plano anterior:

$$B_t(x, y) = I_{t-1}(x, y) \quad (1)$$

- 2 Se calcula la imagen diferencia:

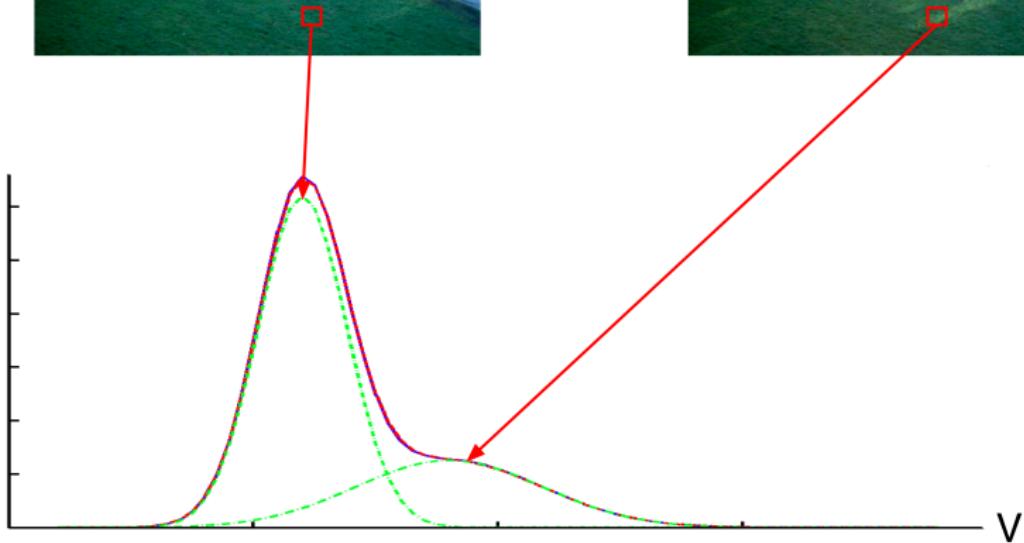
$$D(x, y) = |I_t(x, y) - I_{t-1}(x, y)| \quad (2)$$

- 3 La máscara binaria $BW_t(x, y)$ se obtiene umbralizando:

$$BW_t(x, y) = \begin{cases} 0 & \text{si } D(x, y) < T \\ 1 & \text{si } D(x, y) \geq T \end{cases} \quad (3)$$

Donde T es un umbral elegido de forma empírica.

Modelo mezcla de Gaussianas



Modelo GMM

Una mezcla de Gaussianas es una suma ponderada de k Gaussianas, se define como:

$$p(I(x, y) | \theta) = \sum_{j=1}^k \omega_j(x, y) \phi(I(x, y) | \mu_j(x, y), \Sigma_j(x, y)) \quad (4)$$

El reto es encontrar:

$$\theta(x, y) = \{(\omega_j(x, y), \mu_j(x, y), \Sigma_j(x, y))\}_{j=1}^k \quad (5)$$

Utilizando el algoritmo EM (*Expectation Maximization*) se obtiene el θ que maximiza la función de verosimilitud:

$$\hat{\theta}_{MLE} = \arg \max_{\theta \in \Theta} \log P(I_t(x, y) | \theta(x, y)) \quad (6)$$

Modelo GMM

Para estimar si un píxel $I_t(x, y)$ pertenece al fondo, se umbraliza la verosimilitud de un dato $p(I_t(x, y)|\theta(x, y))$ en el GMM:

$$BW_n(x, y) = \begin{cases} 0 & \text{si } p(I_t(x, y)|\theta(x, y)) > Th \\ 1 & \text{si } p(I_t(x, y)|\theta(x, y)) \leq Th \end{cases} \quad (7)$$

donde Th se ajustó empíricamente para obtener los mejores resultados.

Combinación: Resta plano anterior



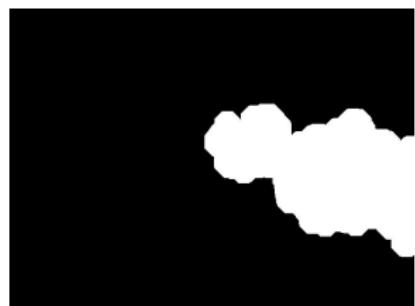
Imagen original



Resta plano anterior



Dilatación resta plano anterior



Postprocesado

Combinación: Modelo mezcla de Gaussianas



Imagen original



Mezcla de Gaussianas



Apertura



Dilatación GMM

Máscara final



Resta plano anterior

&



GMM



Máscara Final

Outline

1 Introducción

2 Detección

3 Estimación del movimiento

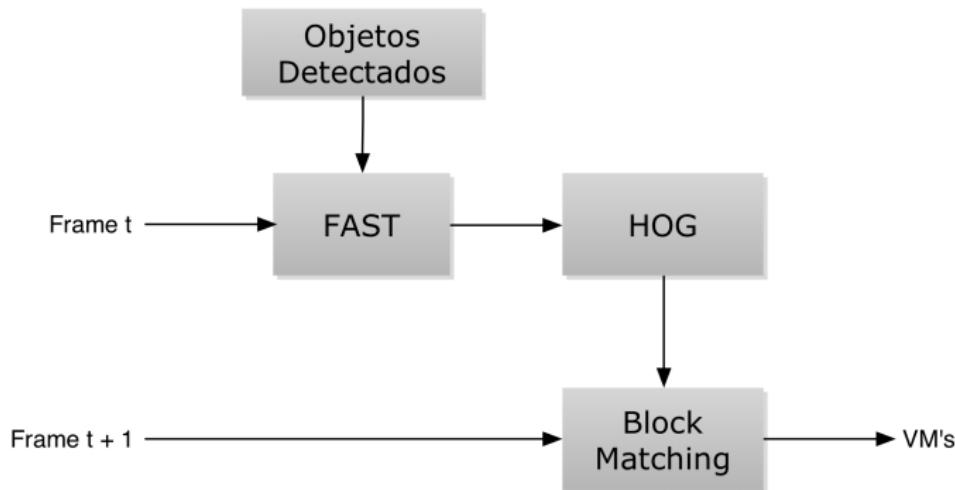
- Extracción de puntos característicos
- Descriptor de puntos
- Block Matching

4 Reconocimiento de eventos

5 Resultados

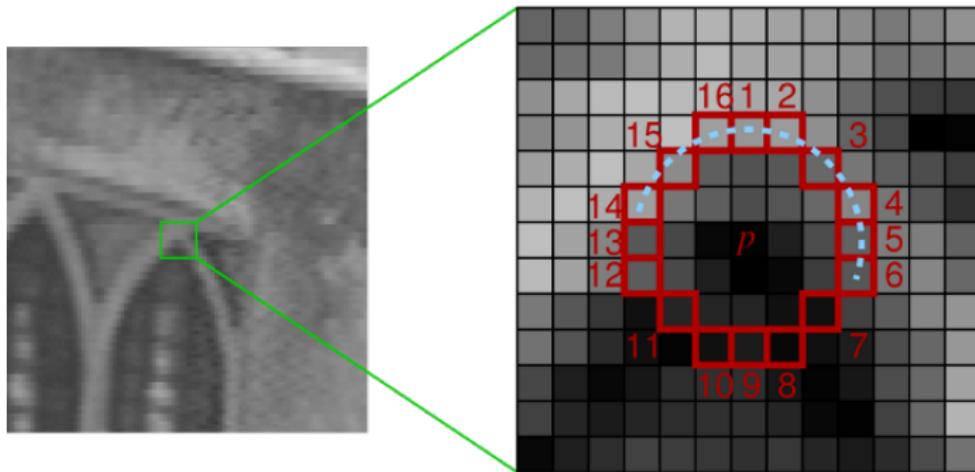
6 Conclusiones

Esquema del sistema de estimación de movimiento



Extracción de puntos característicos: FAST

Se buscan variaciones grandes de la señal en dos dimensiones. FAST consiste en buscar los puntos donde convergen los bordes de la imagen (esquinas).



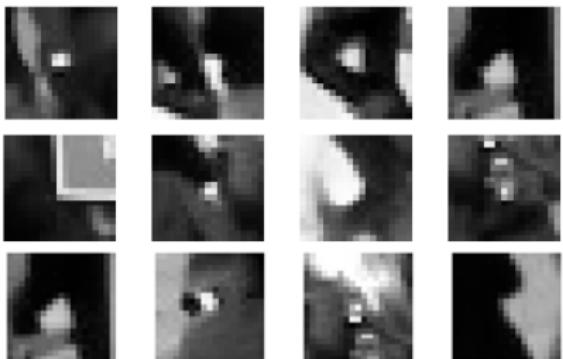
Condición de detección:

La intensidad de al menos n píxeles contiguos deberá ser mayor que $I_p + t$ o menor que $I_p - t$.

Ejemplo de puntos

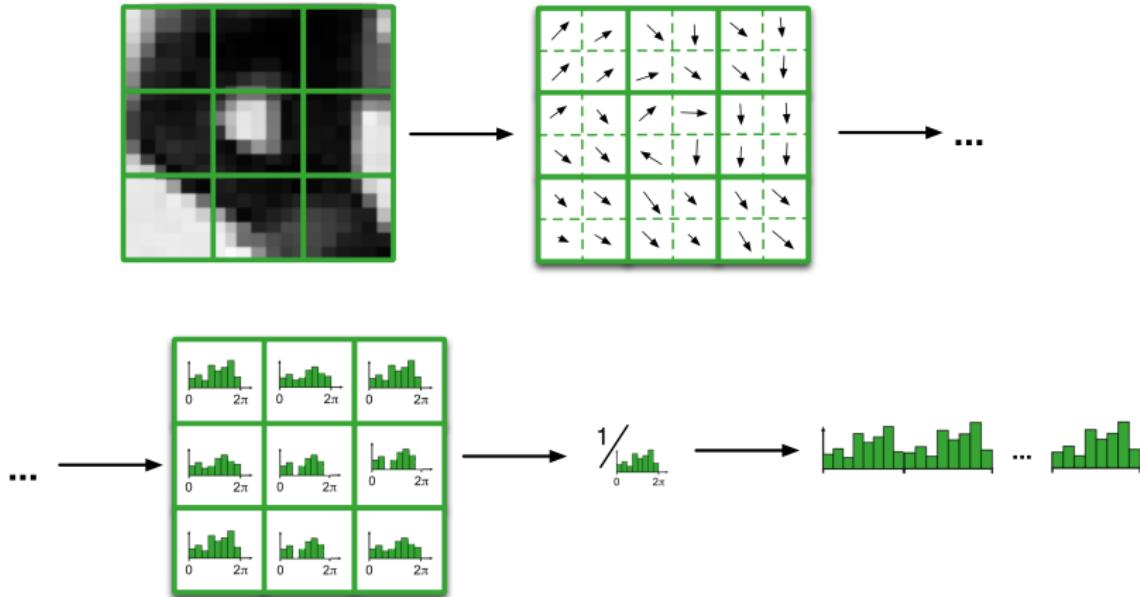


Puntos extraídos por FAST

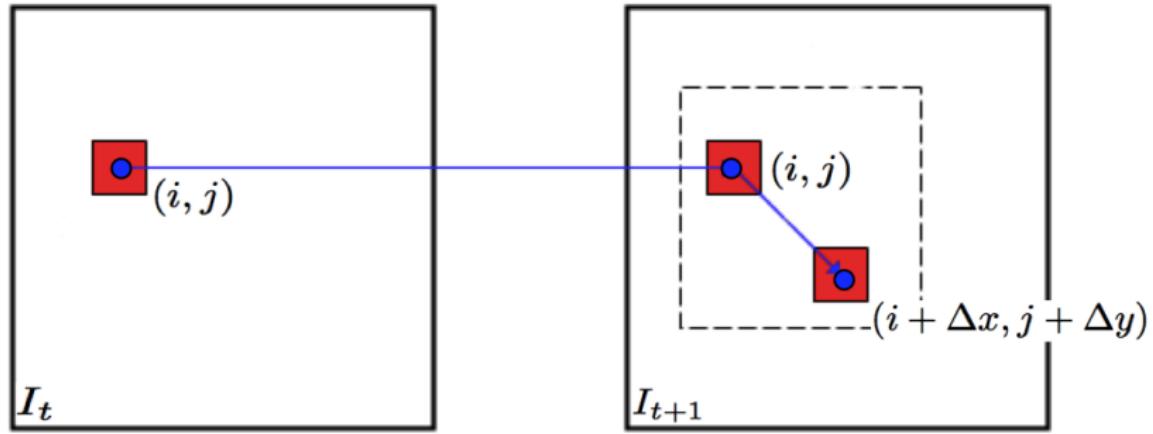


Algunos parches

Descriptor de puntos: HOG



Block Matching



$$d(h_1, h_2) = 1 - \sum_{i=1}^N \min\{h_1(i), h_2(i)\} \quad (8)$$

Outline

1 Introducción

2 Detección

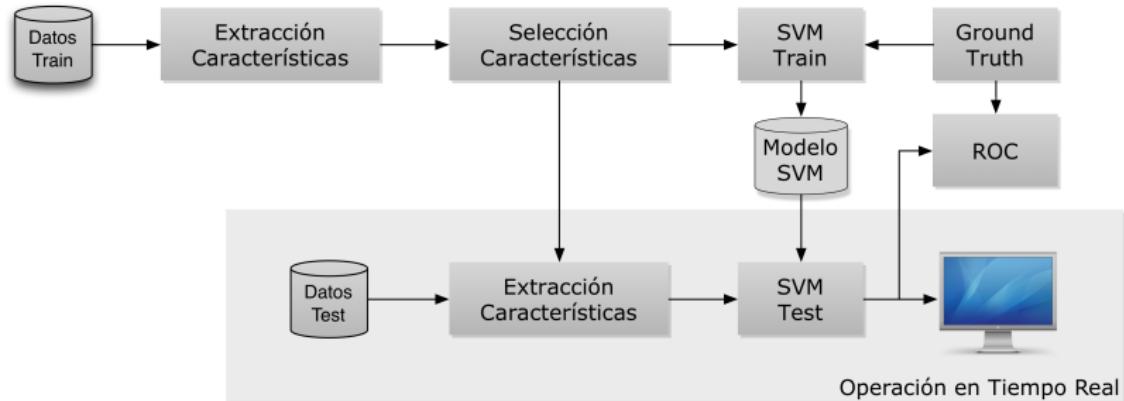
3 Estimación del movimiento

4 Reconocimiento de eventos

- Características
- Selección de las características mediante el estudio de información mutua
- Clasificador

5 Resultados

Diagrama del sistema de reconocimiento



Características

Características implementadas:

- Velocidad Media.
- Histograma de orientaciones del movimiento.
- Parámetro de divergencia.
- Ratio de distancias
- Densidad de puntos

Velocidad media

Velocidad media de N puntos:

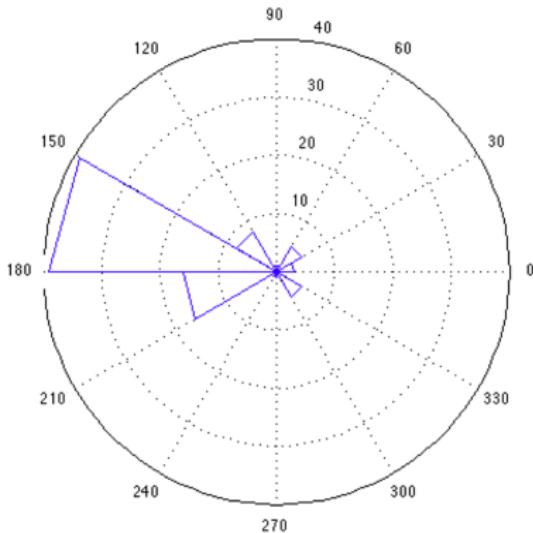
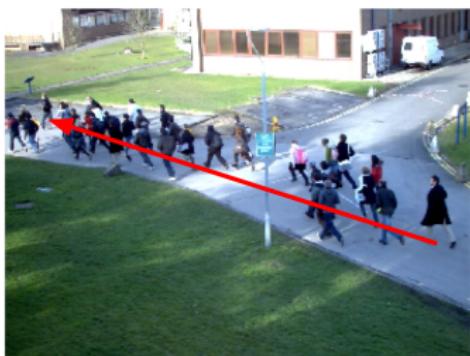
$$\bar{v}_n = \frac{1}{N} \sum_{i=1}^N v_i \quad (9)$$

En una ventana temporal de k planos:

$$\bar{v} = \begin{bmatrix} \bar{v}_n \\ \bar{v}_{n-1} \\ \bar{v}_{n-2} \\ \vdots \\ \bar{v}_{n-k} \end{bmatrix} \quad (10)$$

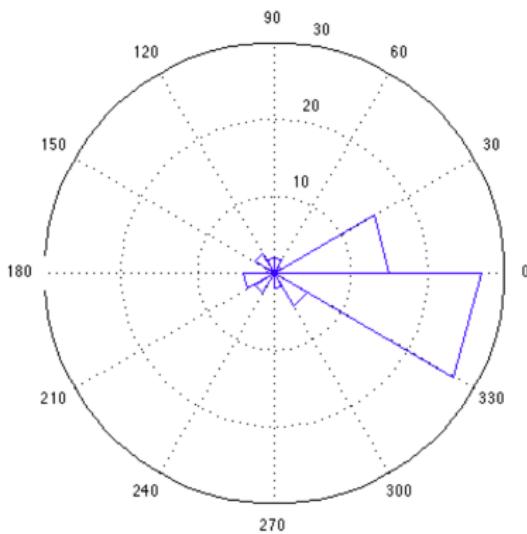
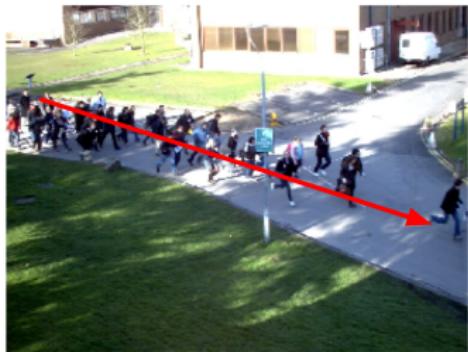
Se tomó un tamaño de ventana $k = 5$.

Histograma de orientaciones del movimiento

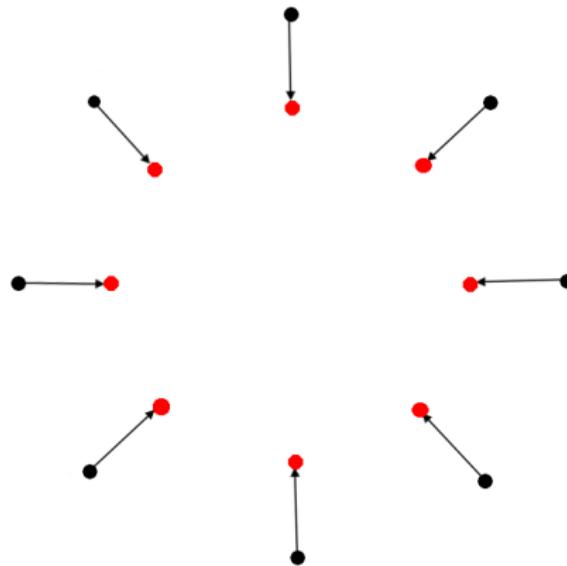


$$\theta = \arctg \left(\frac{v_y}{v_x} \right) \quad (11)$$

Histograma de orientaciones del movimiento



Parámetro de divergencia



$$p' = p + v \quad (12)$$

Parámetro de divergencia

Estimar la matriz de transformación \mathbf{H} .

$$P' = \mathbf{H}P \quad (13)$$

$$\begin{bmatrix} p'_{x1} & p'_{x2} & \dots & p'_{xn} \\ p'_{y1} & p'_{y2} & \dots & p'_{yn} \\ 1 & 1 & \dots & 1 \end{bmatrix} = \begin{bmatrix} \epsilon \cos \theta & -\epsilon \sin \theta & t_x \\ \epsilon \sin \theta & \epsilon \cos \theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_{x1} & p_{x2} & \dots & p_{xn} \\ p_{y1} & p_{y2} & \dots & p_{yn} \\ 1 & 1 & \dots & 1 \end{bmatrix}$$

Dos formas:

- ① Utilizando la matriz pseudoinversa.
- ② Utilizando el algoritmo **RANSAC** [3].

El parámetro se define como:

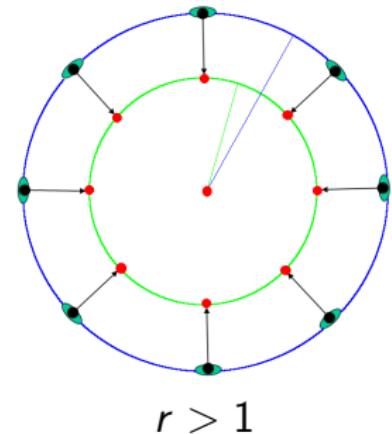
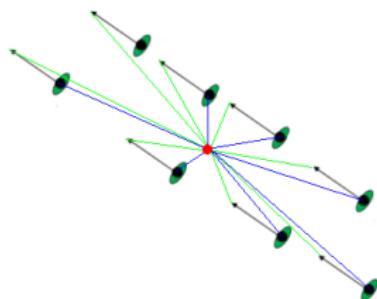
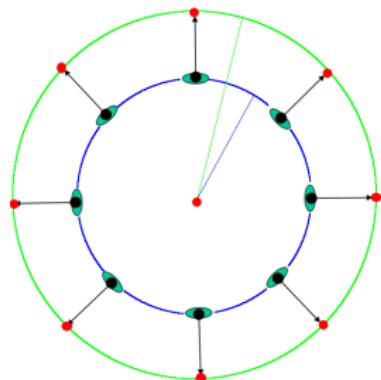
$$div = 2\epsilon \cos \theta \quad (14)$$

Ratio de distancias

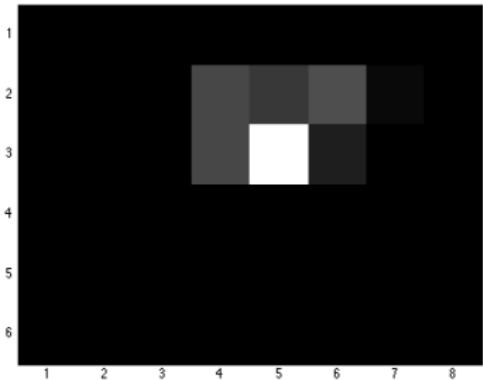
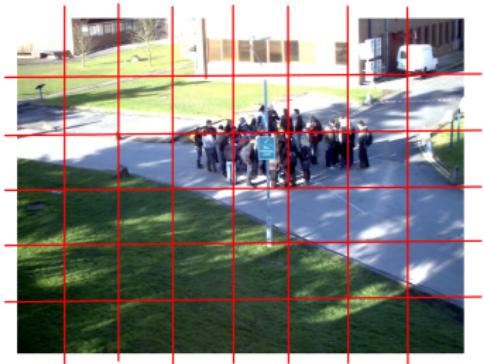
Ratio entre la distancia media de los puntos al punto medio en el plano t y la distancia media de los puntos al punto medio en el plano $t + 1$

$$r = \frac{\frac{1}{n} \sum_{i=1}^n d(p_i, \bar{p})}{\frac{1}{n} \sum_{i=1}^n d(p'_i, \bar{p})}$$

Ratio de distancias



Densidad de puntos



Selección de las características mediante información mutua

La información mutua $I(X; Y)$ se define como:

$$MI(X; Y) = H(Y) - H(Y|X) \quad (15)$$

Selección de las características para el evento *Evacuation*:

Velocidad media	0.0925	—	—	—	—	—
Densidad de puntos	0.0525	0.1338	—	—	—	—
Ratio de distancias	0.0525	0.0813	0.1398	—	—	—
Parámetro de divergencia	0.0456	0.0821	0.1322	0.1329	—	—
Histograma orientaciones	0.0867	0.1006	0.1051	0.0980	0.1013	—

Selección de las características mediante información mutua

La información mutua $I(X; Y)$ se define como:

$$MI(X; Y) = H(Y) - H(Y|X) \quad (15)$$

Selección de las características para el evento *Evacuation*:

Velocidad media	0.0925	—	—	—	—	—
Densidad de puntos	0.0525	0.1338	—	—	—	—
Ratio de distancias	0.0525	0.0813	0.1398	—	—	—
Parámetro de divergencia	0.0456	0.0821	0.1322	0.1329	—	—
Histograma orientaciones	0.0867	0.1006	0.1051	0.0980	0.1013	—

Selección de las características mediante información mutua

La información mutua $I(X; Y)$ se define como:

$$MI(X; Y) = H(Y) - H(Y|X) \quad (15)$$

Selección de las características para el evento *Evacuation*:

Velocidad media	0.0925	—	—	—	—	—
Densidad de puntos	0.0525	0.1338	—	—	—	—
Ratio de distancias	0.0525	0.0813	0.1398	—	—	—
Parámetro de divergencia	0.0456	0.0821	0.1322	0.1329	—	—
Histograma orientaciones	0.0867	0.1006	0.1051	0.0980	0.1013	—

Selección de las características mediante información mutua

La información mutua $I(X; Y)$ se define como:

$$MI(X; Y) = H(Y) - H(Y|X) \quad (15)$$

Selección de las características para el evento *Evacuation*:

Velocidad media	0.0925	—	—	—	—	—
Densidad de puntos	0.0525	0.1338	—	—	—	—
Ratio de distancias	0.0525	0.0813	0.1398	—	—	—
Parámetro de divergencia	0.0456	0.0821	0.1322	0.1329	—	—
Histograma orientaciones	0.0867	0.1006	0.1051	0.0980	0.1013	—

Selección de las características mediante información mutua

La información mutua $I(X; Y)$ se define como:

$$MI(X; Y) = H(Y) - H(Y|X) \quad (15)$$

Selección de las características para el evento *Evacuation*:

Velocidad media	0.0925	—	—	—	—
Densidad de puntos	0.0525	0.1338	—	—	—
Ratio de distancias	0.0525	0.0813	0.1398	—	—
Parámetro de divergencia	0.0456	0.0821	0.1322	0.1329	—
Histograma orientaciones	0.0867	0.1006	0.1051	0.0980	0.1013

Selección de las características mediante información mutua

La información mutua $I(X; Y)$ se define como:

$$MI(X; Y) = H(Y) - H(Y|X) \quad (15)$$

Selección de las características para el evento *Evacuation*:

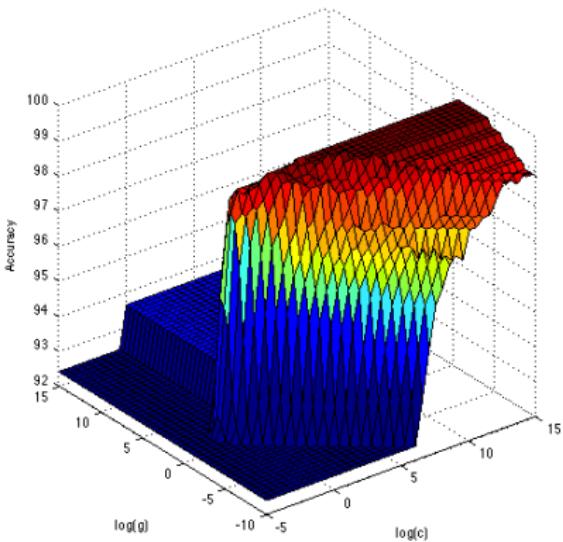
Velocidad media	0.0925	—	—	—	—
Densidad de puntos	0.0525	0.1338	—	—	—
Ratio de distancias	0.0525	0.0813	0.1398	—	—
Parámetro de divergencia	0.0456	0.0821	0.1322	0.1329	—
Histograma orientaciones	0.0867	0.1006	0.1051	0.0980	0.1013

Características seleccionadas

	VM	Densidad	Ratio	Divergencia	Histograma
<i>Walking</i>	✓	✓			
<i>Running</i>	✓	✓	✓		
<i>Evacuation</i>	✓	✓	✓	✓	
<i>Crowd Formation</i>	✓	✓		✓	
<i>Crowd Splitting</i>	✓		✓		✓
<i>Local Dispersion</i>	✓	✓			✓

Clasificador

- ① Dividir los planos en entrenamiento y test.
- ② Normalizar los datos linealmente.
- ③ Considerar la función de kernel RBF.
- ④ Validación cruzada para encontrar los parámetros C y γ .
- ⑤ Utilizar los mejores C y γ para entrenar.
- ⑥ Test.



Outline

- 1 Introducción
- 2 Detección
- 3 Estimación del movimiento
- 4 Reconocimiento de eventos
- 5 Resultados
 - Resultados detección explícita
 - Vídeo resultados
 - Errores
- 6 Conclusiones

Resultados detección explícita

Evento	TPR		
	[4]	[2]	MP
Walking	0.92	—	0.94
Running	0.92	0.80	0.98
Evacuation	0.97	0.89	0.98
Crowd Formation	0.60	1.00	0.94
Crowd Splitting	0.79	0.75	1
Local Dispersion	0.77	0.88	0.99

[4] → Sistema parecido pero sin usar técnicas de aprendizaje máquina.

[2] → Histogramas de orientaciones de movimiento, para la detección de los distintos eventos.

Resultados detección explícita

Evento	ACC	
	[1]	MP
Walking	0.87	0.96
Running	0.88	0.97
Evacuation	0.94	0.99
Crowd Formation	0.68	0.94
Crowd Splitting	0.77	0.99
Local Dispersion	0.80	0.97

[1] → Se detectan grupos de personas presentes en la escena, se divide la imagen en regiones y se cuentan las personas en cada una de éstas.

Resultados

(Resultados.mp4)

Los errores existentes en el sistema se producen principalmente por dos motivos:

- ① No detección de los planos en donde está empezando a ocurrir el evento.
- ② El sistema no logra modelar correctamente algunas situaciones al no disponer de suficientes datos de entrenamiento.

Outline

1 Introducción

2 Detección

3 Estimación del movimiento

4 Reconocimiento de eventos

5 Resultados

6 Conclusiones

- Aportaciones
- Líneas futuras

Aportaciones

- Empleo del modelo de mezcla de Gaussianas para la sustracción del fondo.
- Diseño de un mecanismo de combinación de modelos de actualización temporal y modelos entrenados.
- Inclusión de características no presentes en la literatura.
- Selección de las distintas características mediante información mutua.
- Empleo de SVM para las tareas de clasificación.

- Ampliación base de datos.
- Detección de más eventos.
- Implementación del sistema en una situación real.
- Implementar el sistema con tecnologías que permitan su uso en tiempo real.
- Aumentar el número de características.

Preguntas



A. B. Chan, M. Morrow, and N. Vasconcelos.

Analysis of crowded scenes using holistic properties.

In *In IEEE Intl. Workshop on Performance Evaluation of Tracking and Surveillance (PETS 2009)*, june 2009.



H.M. Dee and A. Caplier.

Crowd behaviour analysis using histograms of motion direction.

In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 1545 –1548, sept. 2010.



Martin A. Fischler and Robert C. Bolles.

Readings in computer vision: issues, problems, principles, and paradigms.

chapter Random sample consensus: a paradigm for model fitting with applications to image analysis and automated

cartography, pages 726–740. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1987.

 C. Garate, P. Bilinsky, and F. Bremond.

Crowd event recognition using hog tracker.

In *Performance Evaluation of Tracking and Surveillance (PETS-Winter), 2009 Twelfth IEEE International Workshop on*, pages 1 –6, dec. 2009.

 IEEE Computer Society, IEEE Signal Processing Society, and Boston University.

Performance evaluation of tracking and surveillance 2010 database.

<http://www.cvg.rdg.ac.uk/PETS2010/>, 2010.