View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Universidad de Castilla-La Mancha: Repositorio Universitario Institucional de Recursos...

Available online at www.sciencedirect.com

SCIENCE DIRECT

Pattern Recognition 39 (2006) 2194–2211

PATTERN RECOGNITION

THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Visual surveillance by dynamic visual attention method

María T. López[a], Antonio Fernández-Caballero[a,*], Miguel A. Fernández[a], José Mira[b], Ana E. Delgado[b]

[a] *Departamento de Sistemas Informáticos, Escuela Politécnica Superior de Albacete, and Instituto de Investigación en Informática de Albacete (I3A), Universidad de Castilla-La Mancha, 02071 Albacete, Spain*

[b] *Departamento de Inteligencia Artificial, E.T.S. Ingeniería Informática, Universidad Nacional de Educación a Distancia, 28040 Madrid, Spain*

Received 6 September 2005; received in revised form 22 February 2006; accepted 11 April 2006

Abstract

This paper describes a method for visual surveillance based on biologically motivated dynamic visual attention in video image sequences. Our system is based on the extraction and integration of local (pixels and spots) as well as global (objects) features. Our approach defines a method for the generation of an *active attention focus* on a dynamic scene for surveillance purposes. The system segments in accordance with a set of predefined features, including gray level, motion and shape features, giving raise to two classes of objects: vehicle and pedestrian. The solution proposed to the selective visual attention problem consists of decomposing the input images of an indefinite sequence of images into its moving objects, defining which of these elements are of the user's interest at a given moment, and keeping attention on those elements through time. Features extraction and integration are solved by incorporating mechanisms of charge and discharge—based on the permanency effect—, as well as mechanisms of lateral interaction. All these mechanisms have proved to be good enough to segment the scene into moving objects and background.

© 2006 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.

Keywords: Dynamic visual attention; Visual surveillance; Segmentation from motion; Feature extraction; Feature integration

1. Introduction

Visual input is probable the most powerful source of information used by man to represent a monitored scene. Visual information is composed of a great deal of redundant sets of spatial and temporal data robustly and quickly processed by the brain. Visual information was entirely processed by human operators in first generation video-based surveillance systems. But when a human observes a set of monitors connected to a set of cameras his performance quickly decays in terms of a correct alarm detection ratio. Modern digital computation and communication technologies have enabled a complete change in the perspective of the design of surveillance systems architectures. Surveillance is a multidisciplinary field, which affects a great number of services and users. Some examples where surveillance is necessary by computing media are: intelligent traffic management [1–3], prevention of non-desired situations by means of closed systems of surveillance, such as vandalism in metro stations [4], management of traffic lights in pedestrian crossings [5], automatic and simultaneous visual surveillance of vehicles and persons [6]. This short review does not deal with all areas where camera surveillance is used, but rather centers in vehicles and persons surveillance.

Let us start with some previous vehicle surveillance systems. These systems may be categorized by their capabilities of counting vehicles, measuring speeds and monitoring traffic queues. The traffic research using image processing (TRIP) [7] system is a system designed to count vehicles running on a bi-directional freeway. Another system [8] uses sampling points able to detect the presence of a vehicle with the purpose of counting the number of vehicles. The wide area detection system

* Corresponding author. Tel.: +34 967 599200; fax: +34 967 599224.
E-mail address: caballer@info-ab.uclm.es (A. Fernández-Caballero).

0031-3203/$30.00 © 2006 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2006.04.018

(WADS) system is able to detect, to count and to measure the speed of the vehicles in movement [9]. Motion detection for frame differentiation is also is the nucleus of a system able to count vehicles, to measure their speeds and to track them in complex highway crossings [10]. IMPACTS [11] system operates at a macroscopic level, offering a qualitative description of the space distribution of moving and stationary traffic in the scene. Another system, able to measure parameters of traffic queues [12], operates in small regions of the image. Now, surveillance exclusively dedicated to persons is also a growing field of interest. Broadly speaking, there are different approaches ranging from active vision algorithms [13] to model-based tracking methods [14], from active contour processes[15] to different features integration (numeric or semantic) [16]. Lastly, let us highlight the more recent works in vehicle and person surveillance integration. In this case motion segmentation is also used in many cases to exploit image difference techniques (generally, using a reference image) [17–19]. And, to clearly differentiate among vehicles and pedestrians a great number of methods are based in models [20–24].

In this paper, we introduce a method for visual surveillance based on dynamic visual attention in video image sequences. Our system, inspired in human vision, is based on the extraction and integration of local (pixels and spots) as well as global (objects) features. Our approach defines a method for the generation of an *active attention focus* on a dynamic scene for surveillance purposes. The system segments in accordance with a set of predefined features, including gray level, motion and shape features, giving raise to classes of objects: vehicles and pedestrians. In Section 2 a solution to the dynamic visual attention method in visual surveillance is described. Section 3 offers the results of segmenting a pedestrian and a car, depending on the input parameters to the attention system. Lastly, Section 4 discusses on the performance of the method proposed, and Section 5 offers the more prominent conclusions.

2. Dynamic visual attention method in visual surveillance

2.1. Visual attention

The human attentional system is a complex matter. Findings in psychology and brain imaging have increasingly suggested that it is better to view attention not as a unitary faculty of the mind but as a complex organ system sub-served by multiple interacting neuronal networks in the brain [25]. The images are built habitually as from the entries of parallel ways that process distinct features: motion, solidity, shape, color. Desimone and Ungerleider indicate in Ref. [26] that the representations of the different properties from an object are distributed through multiple regions partially specialized of cortex (shape, color, motion, location). A mechanism must intervene in such a way that the brain associates momentarily the information that is being processed independently at distinct cortical regions. This mechanism is denominated as integration mechanism.

The architecture models for selective attention may be divided into two broad groups: (a) models based exclusively on the scene (bottom–up), and, (b) models based on the scene (bottom–up) and on the task (control top–down).

The first bottom–up neurally plausible architecture of selective visual attention was proposed by Koch and Ullman [27], and it is related to the feature integration theory [28]. The MORSEL model [29] links visual attention to object recognition, to provide an explicit account of the interrelations between these two processes. MORSEL essentially contains two modules, one for object recognition and one for visual attention. In Ref. [30] a neural network (connectionist) model called the selective attention for identification model (SAIM) is introduced. The function of the suggested attention mechanism is to allow translation-invariant shape-based object recognition. The model of guided-search (GS) by Wolfe [31] uses the idea of "saliency map" to realize the search in scenes. GS assumes a two-stage model of visual selection. The first, pre-attentive stage of processing has great spatial parallelism and realizes the computation of the visual simple features. The second stage is spatially serial and it enables more complex visual representations to be computed, involving combinations of features. In Ref. [32] a model is presented that is able to obtain objects separated of the background in static images, in which bottom–up and top–down processes are combined. The bottom–up processes mainly obtain the edges to be able to form the objects. The top–down processes compare the shapes obtained in the bottom–up processes with known forms stored in a database. A very recent model of attention for dynamic vision has been introduced by Backer and Mertsching [33]. In this model there are two selection phases. Previous to the first selection a saliency map is obtained as the result of integrating the different features extracted. Concretely the features extracted are symmetry, eccentricity, color contrast, and depth. The first selection stage selects a small number of items according to their saliency integrated over space and time. These items correspond to areas of maximum saliency and are obtained by means of dynamic neural fields. The second selection phase has top–down influences and depends on the system's aim.

2.2. Our method

Our approach defines a method for the generation of an *active attention focus* on a dynamic scene to monitor a scene for surveillance purposes. The aim is to obtain the objects that keep the user's attention in accordance with a set of predefined features, including gray level, motion and shape features. On the opposite to computational models based on the space (spotlight, zoom),

where attention is paid on one zone of the image, this paper proposes an object-based computational model, which enables to capture attention on one or various objects of the image. One of the mostly referenced selective attention models based on the spotlight metaphor is the Koch and Ullman model [27]. Its disadvantage is that it is restricted to static images. In dynamic environments the model of Backer and Mertsching [33] is of a great interest to us. Indeed, the approach we propose for feature extraction and integration is similar to that of Backer and Mertsching.

In previous papers of our research team some algorithms for the segmentation of the image in different objects have been proposed based on the detection of motion, the permanency effect and lateral interaction [34,35]. Thus, based on the satisfactory results of the algorithms commented, to solve the current problem we propose to incorporate mechanisms of charge and discharge (based on the permanency effect), as well as mechanisms of lateral interaction. These mechanisms are good enough to segment the scene into moving objects and background.

The solution proposed to the selective visual attention problem consists of decomposing the input images of an indefinite sequence of images into its moving objects, defining which of these elements are of the observer's—or user's—interest, and keeping attention on those elements through time. In the system proposed it is mandatory that the observer may define the features of the objects on which attention is focused. The commands (or indications) that the observer introduces into the system in order to adjust parameters which define the attention focus are of a top–down modulation. This modulation is included in a static way during the process of feature selection, as well as in a dynamic form established as a feedback from the attention focus where parameters which define the interest may be modified to center the focus on objects that are of a real interest.

Our solution consists of two processes: bottom–up processes, where pixel and object features are extracted, and, top–down processes, where the observer organizes mechanisms and search parameters to satisfy his expectations respect to the attention focus. The selection of the interest elements of the scene necessarily starts with setting some criteria based on features extracted from the elements (*feature extraction and integration*). Firstly, all necessary mechanisms to provide sensitivity to the system are included in order to succeed in centering the attention. Frame to frame it will be possible to capture attention (*attention capture*) on elements constructed from image pixels that fulfill the requirements established by the user. On the other hand, stability has been provided to the system. This has been gotten by including mechanisms to reinforce attention (*attention reinforcement*), in such a way that elements that assemble the user's predefined requirements are strengthened up to be configured as the system attention center. Therefore, three subtasks are needed: (a) *Feature extraction*: obtains the features (gray level, motion and shape features) of the image necessary to capture attention. (b) *Attention capture*: applies the user-defined criteria (values provided to parameters) to the features extracted and obtains the different parts of the objects of potential interest. (c) *Attention reinforcement*: maintains attention on certain elements (or objects) of the image sequence that are of real interest.

2.3. Feature extraction

Subtask *feature extraction and integration* is composed of two parts. The first one is related to feature extraction (*gray level, motion and shape features extraction*), whilst the second one is feature integration (*feature integration*).

2.3.1. Gray level feature extraction

The aim of subtask *gray level feature extraction* is to get the chromatic features associated to the image pixels. We work with 256 gray level input images and transform them to a lower number of levels $n < 256$. Generally, good results are usually obtained with $n = 8$ levels in normal illumination indoor and outdoor scenes. A higher value rarely gives better results, whilst lower values (say, 2 or 4) may be used for night vision.

Let $GL[x, y, t]$ be the gray level of a pixel (x, y) of the input image at time instant t, GL_{max} the maximum gray level value (generally, 255), GL_{min} the minimum gray level value (generally, 0), n the number of gray level bands, and, $GLB[x, y, t]$ the gray level band of pixel (x, y) at t. Let also ovl be the overlap (or minimum value of the difference in the gray levels between two consecutive time instants required to produce a change in the gray level band of a pixel). Then:

$$GL_{diff} = GL_{max} - GL_{min} + 1, \tag{1}$$

$$GLB[x, y, t] = \begin{cases} GLB[x, y, t-1] & \text{if } max\left(\dfrac{(GLB[x, y, t-1] - 1) * GL_{diff}}{n} - ovl,\ GL_{min}\right) \\ & \leqslant GL[x, y, t] < min\left(\dfrac{GL[x, y, t-1] * GL_{diff}}{n} + ovl,\ GL_{max}\right), \\ \left\lfloor \dfrac{GL[x, y, t] * n}{GL_{diff}} \right\rfloor + 1 & \text{otherwise.} \end{cases} \tag{2}$$

Fig. 1. (a) Input image; (b) image segmented into eight gray level bands; (c) image segmented into four gray level bands.

Fig. 2. (a) Image segmented into eight gray level bands at $t-1$; (b) image segmented into eight gray level bands at t; (c) motion presence at t.

Eq. (2) checks if gray level value $GLB[x, y, t]$ produces a variation of band in relation to the gray level band value obtained at $t-1$, that is to say, $GLB[x, y, t-1]$. For this aim, the criterion used is to check if $GLB[x, y, t]$ has sufficiently changed in its gray level between time instants t and $t-1$ (use of overlap *ovl*).

Fig. 1 shows an example of segmenting the input image (a) into eight gray level bands (b) and four gray level bands (c). To conclude, *gray level feature extraction* transforms a 256 gray level input image into an image composed of patches belonging to one of n possible gray level bands. This will be very valuable information in *motion feature extraction* as well as in *shape feature extraction*.

2.3.2. Motion feature extraction

The aim of subtask *Motion feature extraction* is to calculate the dynamic (motion) features of the image pixels, that is to say, in our case, the presence of motion. Due to our experience [36–41] we know some methods to get that information. Indeed, to diminish the effects of noise due to the changes in illumination in motion detection, variation in gray level bands at each image pixel is performed. Motion presence $motion[x, y, t]$ is easily obtained as a variation in gray level band between two consecutive time instants t and $t-1$:

$$motion[x, y, t] = \begin{cases} 0 & \text{if } GLB[x, y, t] = GLB[x, y, t-1], \\ 1 & \text{if } GLB[x, y, t] \neq GLB[x, y, t-1]. \end{cases} \tag{3}$$

Fig. 2 shows the result of obtaining the motion presence (c) from image differencing between the image segmented into eight gray level bands at time t (b) and the image segmented into eight gray level bands at time $t-1$ (a).

2.3.3. Shape feature extraction

Firstly, subtask *shape feature extraction* extracts different shape features of the patches obtained by segmenting the image into gray level bands, and labeled as interesting during the *attention capture* subtask (see Fig. 3, where the interest pixels are drawn in white color on black background for each one of eight gray level bands). We may call this the *spot shape feature extraction*. These features are the size $s\_spot[v_{label}]$, the width $w\_spot[v_{label}]$ and the height $h\_spot[v_{label}]$. For the running

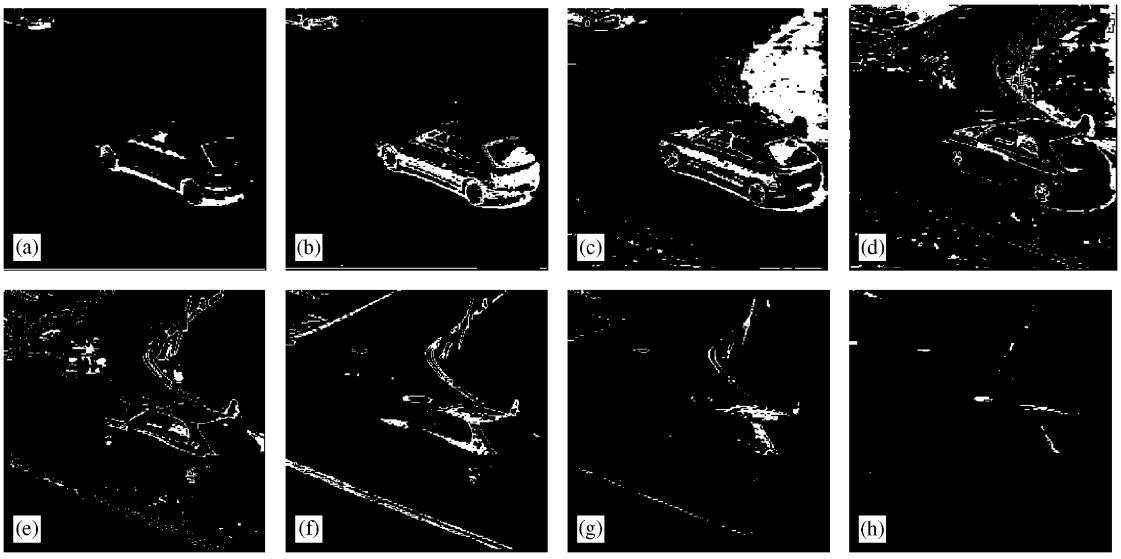

Fig. 3. Pixels of interest at each gray level band: (a) band 1 to (h) band 8.

Table 1
Values of the spot shape features for the running example

Band #	Number of spots	Maximum $s\_spot[v_{label}]$	Maximum $w\_spot[v_{label}]$	Maximum $h\_spot[v_{label}]$
1	48	271	66	30
2	68	2131	151	63
3	215	6269*	87	132
4	424	738	62	95
5	384	139	41	22
6	162	539	59	47
7	101	333	70	32
8	28	62	19	18

example the maximum values for these features extracted are shown in Table 1. Obviously, the asterisk (*) in Table 1 shows a non-valid value for a spot corresponding to a car.

In a similar way the features of the objects stored during the *attention reinforcement* are obtained (the size $s\_object[v_{label}]$, the width $w\_object[v_{label}]$, the height $h\_object[v_{label}]$, the width–height ratio $hw\_object[v_{label}]$ and the compactness $c\_object[v_{label}]$). These are now complete objects united by a common identifying label. So, let us talk about an *object shape feature extraction*. The compactness is obtained as

$$c\_object[v_{label}] = \frac{s\_object[v_{label}]}{h\_object[v_{label}] * w\_object[v_{label}]}. \quad (4)$$

For the case of the running example, consider the output shown in Fig. 4, where the car has been segmented. The object shape features extracted have the following values: $s\_object[v_{label}] = 11,654$, $w\_object[v_{label}] = 162$, $h\_object[v_{label}] = 114$, $hw\_object[v_{label}] = 0.7$ and $c\_object[v_{label}] = 0.6$.

Although the shape features are obviously very simple, the results obtained are good enough, as you may appreciate in the results section.

2.3.4. Feature integration

The output of subtask *Feature integration* is an interest map (in the sense used in visual attention methods) obtained by integrating motion and shape features. In the interest map each image pixel is classified depending on the system's parameters. The states of "active" or "inhibited" are reserved for those pixels where motion presence has been detected at current time t, or for pixels belonging to an object—or object spot—of interest at time instant $t - 1$. Now, "neutral" pixels are the rest of the

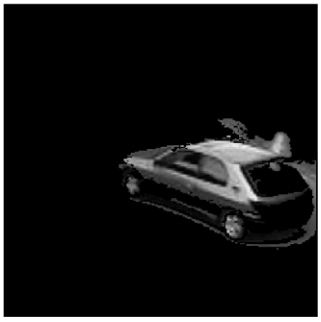

Fig. 4. Result of *attention reinforcement.*

image pixels. "Active" pixels are those that fulfill the requirements imposed by the user, whilst "inhibited" pixels do not fulfill the requirements.

$$interest[x, y, t] \in \{v_{active}, v_{inactive}, v_{neutral}\}. \tag{5}$$

2.4. Attention capture

The objective of subtask *attention capture* is to select image zones (or patches) included in objects of interest. It has been decided to construct these patches from image pixels that fulfill the systems' requirements.

Some research groups solve the problem of defining the elements that decompose the scene by initial border extraction and further obtaining complex objects from more simple ones by looking for families of shapes. Our approach starts obtaining the object's parts from their gray level bands. Later on, these objects parts (also called zones, patches or spots) will be treated as whole objects.

To conclude, the aim of this subtask is to construct object spots from image pixels that possess the requirements established by the observer. Remember that, firstly, the image has been segmented into gray level bands in regions composed of connected pixels whose illumination level belongs to a common interval (gray level band). Secondly, only those connected regions that include an "active" pixel in the interest map have been selected. Each one of these regions (or silhouettes) of a uniform gray level band is defined as a scene spot belonging to a potentially interesting object.

In order to obtain the patches that contain "active" pixels in the interest map, the process consists of overlapping the image segmented in gray level bands of the current frame (at t) with the image of the interest map constructed in the previous frame (at $t-1$), just as done with superimposed transparencies. The output of this subtask is a number $attention_i[x, y, t]$ assigned to each pixel belonging to a selected spot (the label of the spot). Value 0 is for pixels that do not belong to a patch of interest.

As the model works with n gray level bands, the value at each pixel will be the maximum value of all the values calculated at each gray level band:

$$attention[x, y, t] = \arg \max_i \ attention_i[x, y, t]$$
$$\forall i \in [1 \ldots n]. \tag{6}$$

The initial value (patch label) for each pixel (x, y) at gray level band i is the pixel's position within the image (coordinate x multiplied by the number of image columns (NC) + coordinate y) whenever the pixel is in state "active". A maximum value (v_{max} = number of columns $*$ number of rows $+1$) is assigned if the pixel is labeled as "neutral" and a minimum value ($v_{min} = 0$) if the pixel is "inhibited".

$$v_i[x, y] = \begin{cases} (x * NC + y) + 1 & \text{if } GL[x, y, t] = i \ \wedge \ interest[x, y, t] = v_{active} \\ v_{max} & \text{if } GL[x, y, t] = i \ \wedge \ interest[x, y, t] = v_{neutral} \\ v_{min} & \text{otherwise} \end{cases} \quad \forall i \in [0 \ldots n]. \tag{7}$$

This initial value is compared to the neighbors' values that are at the same gray level band i in an iterative way up to reaching a common value for all the pixels of a same element:

$$v_i[x,y] = \begin{cases} v_{min} & \text{if } v_i[x,y] = v_{min}, \\ min(v_i[\alpha,\beta]) & \text{if } v_{min} < min(v_i[\alpha,\beta]) < v_i[x,y] \leqslant v_{max}, \\ v_i[x,y] & \text{if } v_{min} < v_i[x,y] < min(v_i[\alpha,\beta]) \leqslant v_{max}, \\ v_{max} & \text{if } (v_i[x,y] = v_{max}) \wedge \neg(v_{min} < v_i[\alpha,\beta] < v_{max}) \end{cases} \quad \forall[\alpha,\beta] \in [x \pm 1,\ y \pm 1]. \tag{8}$$

The first row of Eq. (8) means that pixel $[x, y]$ remains inhibited, independently of the values of its neighbors. The second row of the equation shows how pixel $[x, y]$ changes to the lowest neighbor value different from inhibited. The third case says that the value of the pixel remains the same (and active) if its value is lower than the values of all its (non-inhibited) neighbors. The fourth row tells that pixel $[x, y]$ remains neutral in the solely case where it was previously neutral and all its neighbors are either neutral or inhibited.

Finally the value established by consensus is assigned as output at each gray level band:

$$attention_i[x,y,t] = \begin{cases} 0 & \text{if } (v_i[x,y] = v_{min}) \vee (v_i[x,y] = v_{max}), \\ v_i[x,y] & \text{otherwise.} \end{cases} \tag{9}$$

2.5. Attention reinforcement

The mechanisms used to generate the output of *attention capture* endow the system of sensitivity, as it enables resources to include elements associated to interest pixels in the memory. Unfortunately, scene object patches whose shape features do not correspond to those defined by the observer may appear at time instant t. This is precisely because their shape characteristics have not yet been obtained. But, if these spots shape features really do not seem to be interesting for the observer, they will appear as "inhibited" in $t + 1$ in the interest map (now, in $t + 1$ their shape features will have been obtained). And, this means that in $t + 1$ they will disappear from *attention capture.* Scene object spots appear and disappear at each input image frame, as they fulfill or do not fulfill the desired spot shape features. In the same way we have gotten sensitivity, we need some mechanism to endow stability to the system.

In order to provide stability to the system, that is to say, in order to obtain at each frame only objects with the desired features, we have to provide *attention reinforcement* by means of accumulative mechanism followed by a threshold. Accumulation is performed on pixels that have a value different from 0 (pixels that do not belong to labeled zones) as input to *attention capture*. Concretely, pixels that appear with a value different from 0 reinforce attention, whilst those that appear as a 0 decrement the attention value. This accumulative effect followed by a threshold (θ) maintains "active" a set of pixels that belong to a group of scene object of interest to the system.

$$focus[x,y,t] = \begin{cases} max(Ch_{AM}[x,y,t-1] - D_{AM}, Ch_{min}) & \text{if } attention[x,y,t] = 0, \\ min(Ch_{AM}[x,y,t-1] + C_{AM}, Ch_{max}) & \text{if } attention[x,y,t] > 0. \end{cases} \tag{10}$$

Now, objects need to be labeled in *attention reinforcement*. This is performed using an initial value at each pixel as shown in Eq. (11):

$$v[x,y] = \begin{cases} (x * NC + y) + 1 & \text{if } focus[x,y,t] \geqslant \theta \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

This initial value is contrasted with the values of the neighbors until a common value for all pixels of a same moving object is reached:

$$v[x,y] = \begin{cases} 0 & \text{if } v[x,y] = v[\alpha,\beta] = 0, \\ min(v[\alpha,\beta]) & \text{if } 0 < min(v[\alpha,\beta]) < v[x,y], \\ v[x,y] & \text{if } 0 < v[x,y] < min(v[\alpha,\beta]). \end{cases} \quad \forall[\alpha,\beta] \in [x \pm 1, y \pm 1], \tag{12}$$

Finally, the value agreed is assigned as the image output (output of *attention reinforcement*):

$$focus[x,y,t] = v[x,y]. \tag{13}$$

3. Data and results

To test the performance of the model proposed, 60, 256 × 512 pixel images have been used from the PETS 2001 dataset. The images contain a scene where a vehicle (a car) and a pedestrian are moving. The sequence is a subset of a test database of the *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance* called "*PETS2001 Datasets*, *Dataset* 1: *Moving people and vehicles*", and has been downloaded via ftp://pets.rdg.ac.uk/PETS2001/. This series offers some moving objects; specifically, there is a car and a pedestrian in motion. First the car and next the pedestrian are selected through their major shape features, showing this way the versatility of our *dynamic visual attention* method in visual surveillance of vehicles and pedestrians in indefinite sequences of video images. A vehicle, as well as a pedestrian, is modeled as a rectangular image element with known dimensions.

Notice that the method is dependent on the specific scenario chosen, in the sense that the parameters have to be tuned for the scenario and for each class of object to pay attention on. Fortunately, this parameter tuning does not depend on each different situation stored in a video sequence taken from the camera, but only on the predefined attention focuses. And this operation has only to be performed once, when installing the surveillance camera.

3.1. Segmentation of the car

Firstly, parameters to detect and results of selecting car as the attention focus are shown. This is precisely the running example as shown in Section 2. In this case, an overlap $ovl = 8$ has been used. Table 2 shows the parameters used (as well as their values) to get the patches' shapes. Similarly, in Table 3 we show the parameters and values for the object's shapes.

As it may be appreciated in Fig. 5, in row (b) "Active" pixels in the interest, pixels where motion has been detected between two consecutive time instants are shown. As the overlap $ovl = 8$ only pixels belonging to the pedestrian and the car are present. Finally, let us highlight the excellent result of segmenting the car produced as output in the *attention focus* (Table 4).

3.2. Segmentation of the pedestrian

The attention focus may be changed, simply by varying the values of *spot shape features* and *object shape features*. In this case, the overlap has been $ovl = 0$. Table 5 shows the parameters used (as well as their values) to get the patches' shapes. Similarly, in Table 6 we show the parameters and values for the object's shapes. Lastly, the parameters used to calculate the *Attention Focus* are offered in Table 7. Results are shown in Fig. 6.

As it may be appreciated in Fig. 6, in row (b) "Active" pixels in the interest, pixels where motion has been detected between two consecutive time instants are shown. The major difference is probably in the "active" pixels obtained in the interest map. As $ovl = 0$, there are pixels belonging to both moving objects, pedestrian and car, as well as to other parts of the image due to

Table 2
Spot shape features parameters

Parameter	Value (in number of pixels)
Spot maximum size: s_spot_{max}	2600
Spot maximum width: w_spot_{max}	180
Spot maximum height: h_spot_{max}	150

Table 3
Object shape features parameters

Parameter	Value (in pixels)	Value (ratios)
Object minimum size: s_object_{min}	5000	
Object maximum size: s_object_{max}	12,000	
Object minimum width: w_object_{min}	50	
Object maximum width: w_object_{max}	180	
Object minimum height: h_object_{min}	50	
Object maximum height: h_object_{max}	150	
Object minimum width–height ratio:hw_object_{min}		0.4
Object maximum width–height ratio: hw_object_{max}		1.2
Object minimum compactness: c_object_{min}		0.3
Object maximum compactness: c_object_{max}		1.0

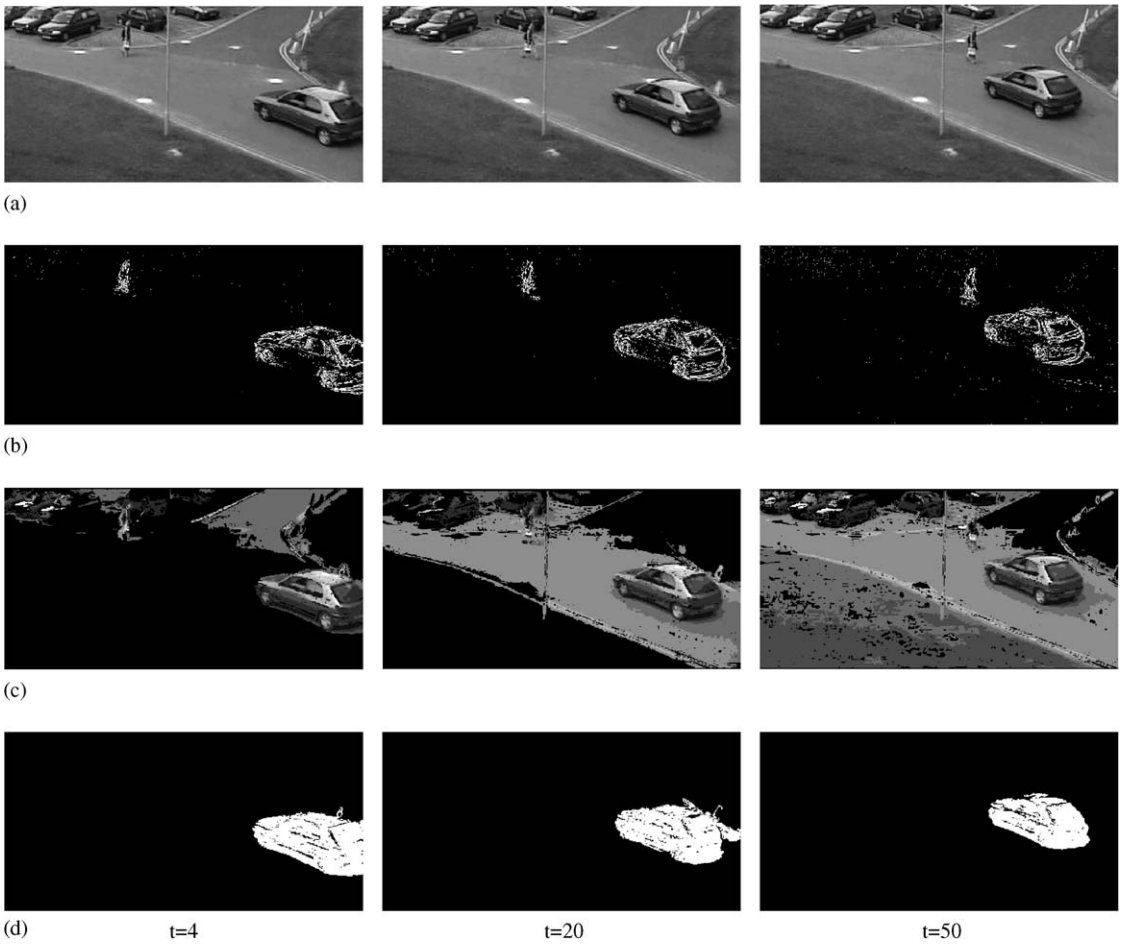

Fig. 5. Sequence of selective attention on the car in different time instants: (a) input image; (b) "Active" pixels of the interest map; (c) output image of *attention capture*; (d) output image of *attention reinforcement.*

Table 4
Attention reinforcement parameters

Parameter	Values
Charge constant: C_{AM}	50
Discharge constant: D_{AM}	250
Threshold: θ	200

Table 5
Spot shape features parameters

Parameter	Value (number of pixels)
Spot maximum size: s_spot_{max}	300
Spot maximum width: w_spot_{max}	20
Spot maximum height: h_spot_{max}	60

the great variations in illumination in the scene. In row (c) due to the high number of false interesting pixels obtained due to noise, some elements that do not belong to the desired attention focus appear. The restrictions imposed as *spot shape features* and *object shape features* enable that in row (d) *attention focus* only the pedestrian appears.

Table 6
Object shape features parameters

Parameter	Value (in pixels)	Value (ratios)
Object minimum size: s_object_{min}	100	
Object maximum size: s_object_{max}	1800	
Object minimum width: w_object_{min}	10	
Object maximum width: w_object_{min}	80	
Object minimum height: h_object_{min}	80	
Object maximum height: h_object_{max}	280	
Object minimum width–height ratio:hw_object_{min}		0.5
Object maximum width–height ratio: hw_object_{max}		3.5
Object minimum compactness: c_object_{min}		0.3
Object maximum compactness: c_object_{max}		1.0

Table 7
Attention reinforcement parameters

Parameter	Values
Charge constant: C_{AM}	100
Discharge constant: D_{AM}	250
Threshold: θ	200

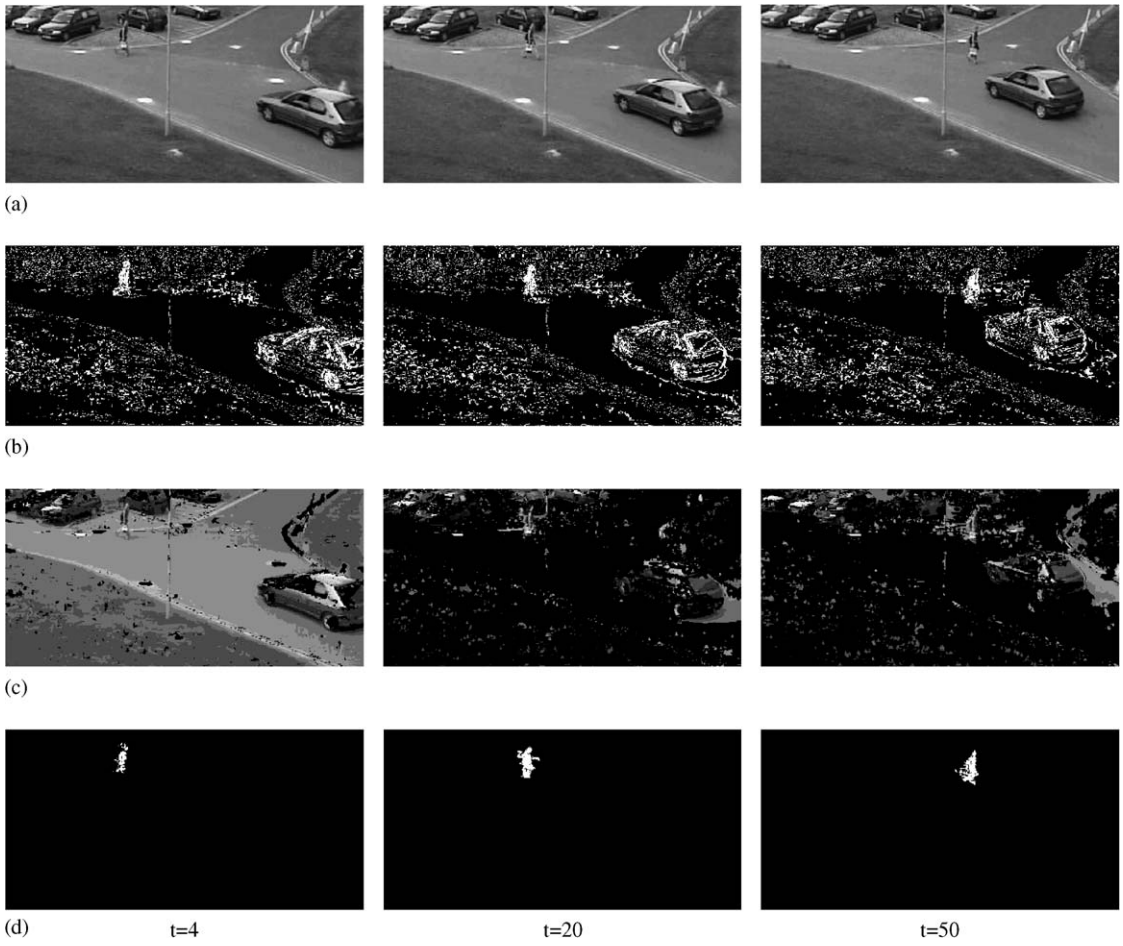

Fig. 6. Sequence of selective attention on the pedestrian in different time instants: (a) input image; (b) "Active" pixels of the interest map in white color; (c) output image of *attention capture*; (d) output image of *attention reinforcement*.

4. Discussion

A well-known image sequence to demonstrate the performance of the technique proposed is the *Hamburg Taxi* sequence (ftp://csd.uwo.ca). In this sequence there are four moving objects: a taxi turning around the corner, a car moving to the right, a van moving to the left and a pedestrian in the upper left. In this discussion section we will show the behavior of our proposal when changing the main parameters of the algorithms described before. In order to provide quantitative performance evaluation values we have defined hits (h), false positives (fp), and false negatives (fn) as:

$$h = \frac{\text{pixels that belong to the attention focus and to the ground truth}}{\text{number of pixels of the ground truth}} 100,$$

$$fp = \frac{\text{pixels that belong to the attention focus, but not to the ground truth}}{\text{number of pixels of the attention focus}} 100,$$

$$fn = \frac{\text{pixels that belong to the ground truth, but not to the attention focus}}{\text{number of pixels of the ground truth}} 100.$$

Firstly, we provide in Tables 8–10, the approximate values for segmenting the "car" and the "taxi" from the *Hamburg Taxi* sequence, calculated by taking direct measures in the input image sequence, and running the algorithms with different values of charge, discharge and threshold.

4.1. Influence of the spot maximum size parameter

In this section we show the influence of changing the value of parameter spot maximum size s_spot_{max}. The rest of the values are number of gray level bands $n = 8$, overlap $ovl = 16$ and those provided in Tables 8–10.

Table 11 shows the results for several spot maximum sizes. Study case 1 corresponds to the case where the spot maximum size is equal to the maximum size of the objects (greater values are non-sense). You may observe in the table that it is possible to diminish the spot maximum size down to 2400 without loss of results. Notice that the percentage of hits (in most cases over an 80%), even without having looked for the optimal values, is really good, and perfectly comparable to other segmentation

Table 8
Spot shape features parameters for segmenting the "car" and the "taxi"

Parameter	Value (in number of pixels)
Spot maximum width: w_spot_{max}	85
Spot maximum height: h_spot_{max}	70

Table 9
Object shape features parameters for segmenting the "car" and the "taxi"

Parameter	Value (in pixels)	Value (ratios)
Object minimum size: s_object_{min}	400	
Object maximum size: s_object_{max}	3000	
Object minimum width: w_object_{min}	40	
Object maximum width: w_object_{min}	85	
Object minimum height: h_object_{min}	35	
Object maximum height: h_object_{max}	70	
Object minimumwidth–height ratio: hw_object_{min}		0.5
Object maximum width–height ratio: hw_object_{max}		1.5
Object minimum compactness: c_object_{min}		0.4
Object maximum compactness: c_object_{max}		1.0

Table 10
Attention reinforcement parameters for segmenting the "car" and the "taxi"

Parameter	Values
Charge constant: C_{AM}	50
Discharge constant: D_{AM}	250
Threshold: θ	201

Table 11
Influence of maximum spot size feature

Frame	Study case 1 ($s\_spot_{max} = 3000$)			Study case 2 ($s\_spot_{max} = 2400$)			Study case 3 ($s\_spot_{max} = 2000$)		
	h	*fp*	*fn*	*h*	*fp*	*fn*	*h*	*fp*	*fn*
5	67.02	10.40	32.98	67.02	10.40	32.98	34.46	7.15	65.54
6	73.77	11.46	26.23	73.77	11.46	26.23	36.07	7.61	63.93
7	84.98	15.07	15.02	84.98	15.07	15.02	36.83	8.90	63.17
8	84.32	12.15	15.68	84.32	12.15	15.68	36.08	5.41	63.92
9	83.20	16.17	16.80	83.20	16.17	16.80	37.54	5.99	62.46
10	88.90	19.92	11.10	88.90	19.92	11.10	40.14	8.83	59.86
11	84.67	18.89	15.33	84.67	18.89	15.33	37.96	9.68	62.04
12	85.94	16.62	14.06	85.94	16.62	14.06	37.28	11.12	62.72
13	83.82	13.78	16.18	83.82	13.78	16.18	36.33	7.85	63.67
14	84.50	14.01	15.50	84.50	14.01	15.50	35.82	7.06	64.18
15	83.31	16.05	16.69	83.31	16.05	16.69	34.77	10.72	65.23
16	86.50	18.85	13.50	86.50	18.85	13.50	34.65	14.59	65.35
17	84.50	15.77	15.50	84.50	15.77	15.50	32.82	11.10	67.18
18	89.24	18.04	10.76	89.24	18.04	10.76	33.62	13.53	66.38

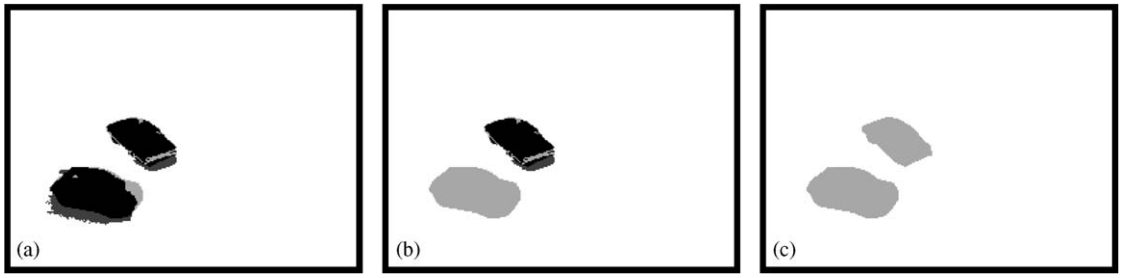

Fig. 7. Graphical results when varying spot maximum size parameter for frame 18: (a) $s\_spot_{max} = 3000$ and/or $s\_spot_{max} = 2400$; (b) $s\_spot_{max} = 2400$, (c) $s\_spot_{max} = 550$.

techniques, after having tuned the values. But if the value of spot maximum size continues diminishing the error percentage augments as it may be seen in study case 3.

Fig. 7 represents graphically the hits, the false positives and the false negatives. In black color the hits (*h*) are represented; the false negatives (*fn*) are drawn in light gray color, whereas dark gray color is deserved for false positives (*fp*). Fig. 7b shows an increment of false negatives with $s\_spot_{max} = 2000$, due to the fact that the "car" does not appear in the attention focus. Fig. 7c indicates that for $s\_spot_{max} = 550$, that is to say, when the spot maximum size continues diminishing, the percentage of hits is equal to 0%—no any object appears in the attention focus.

4.2. Influence of the overlap feature

In this section we show the influence of the overlap parameter *ovl*. Again, the values are number of gray level bands $n = 8$ and those provided in Tables 8–10. The value for spot maximum size is fixed to $s\_spot_{max} = 2400$, as this has been proved to be the optimal value in the previous subsection.

As shown in Table 12 and in Fig. 8, the percentages of *h* and *fn* are similar for the different overlaps, but the percentage of *fp* rises significantly when $ovl = 8$ compared to overlaps 16 and 32. The increment when $ovl = 8$ is due to the appearing of pixels that offer motion information, where there has been no real motion in them. We are rather in front of noise in form of a change in luminosity eliminated by greater overlap values (16 or 32). Remember that the overlap allows filtering noise by making it more difficult to jump from one gray level band to another one (see Eq. (2)).

4.3. Influence of the height–width ratio and compactness features

This section demonstrates how it is easy to change the searching objective of the attention focus by only varying the *height–width ratio* and the *compactness* features and maintaining constant the rest of the parameters. In this case we try to

Table 12
Influence of overlap parameter

Frame	Study case 1 ($ovl = 32$)			Study case 2 ($ovl = 16$)			Study case 3 ($ovl = 8$)		
	h	*fp*	*fn*	*h*	*fp*	*fn*	*h*	*fp*	*fn*
5	65.78	10.57	34.22	67.02	10.40	32.98	70.10	31.19	29.90
6	73.51	11.50	26.49	73.77	11.46	26.23	74.03	32.64	25.97
7	84.92	15.08	15.08	84.98	15.07	15.02	85.11	33.95	14.89
8	84.32	12.15	15.68	84.32	12.15	15.68	84.57	30.89	15.43
9	83.20	16.17	16.80	83.20	16.17	16.80	83.44	34.77	16.56
10	88.90	19.92	11.10	88.90	19.92	11.10	88.96	37.16	11.04
11	84.67	18.89	15.33	84.67	18.89	15.33	84.76	37.07	15.24
12	85.94	16.62	14.06	85.94	16.62	14.06	85.94	35.55	14.06
13	83.82	13.78	16.18	83.82	13.78	16.18	83.82	34.17	16.18
14	84.50	14.01	15.50	84.50	14.01	15.50	84.50	33.58	15.50
15	83.24	16.06	16.76	83.31	16.05	16.69	83.31	35.70	16.69
16	86.46	18.85	13.54	86.50	18.85	13.50	86.50	37.48	13.50
17	84.50	15.77	15.50	84.50	15.77	15.50	84.57	35.49	15.43
18	89.24	18.04	10.76	89.24	18.04	10.76	89.31	36.74	10.69

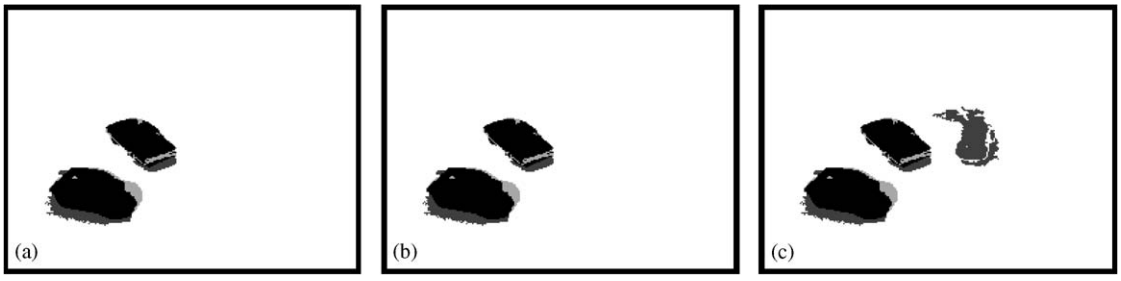

Fig. 8. Graphical results when overlap parameter for frame 18: (a) $ovl = 32$; (b) $ovl = 16$; (c) $ovl = 8$.

clearly discriminate among the "car" and the "taxi". The values of these fixed parameters are those provided in the previous section.

Table 13 shows the results where for study case 1 only the ground truth of the "car" has been used and for study case 2 only the ground truth of the "taxi" has been taken into account. The optimal parameters for case studies 1 and 2 are also offered in the same table. Fig. 9 shows graphically the change in the attention focus.

4.4. Influence of charge and discharge constants, and threshold parameter

For this study, values taken are number of gray level bands $n = 8$, overlap $ovl = 16$, $s\_spot_{max} = 2400$ and the values of Tables 8 and 9.

It may be concluded, as you may observe in Table 14 that a greater relation between threshold and charge (by taking the minimum value of k such that $C_{AM} * k > \theta$, we have for study case 1, $k = 3$, for study case 2, $k = 4$, and for study case 3, $k = 5$), the percentage of *fn* augments, whereas the percentage of *fp* diminishes. The reason for the diminishment of the percentage for *fp* is that when the relation between threshold and charge augments, it is necessary that a pixel appears more times in *attention capture* task to be included in the attention focus. Therefore, the probability of this pixel to belong to the desired objective diminishes, and the percentage of *fp* augments. The values shown for frame 5 perfectly show thus issue.

Notice that the parameters charge and discharge constants, as well as the threshold parameter permit to get better results if well tuned (Fig. 10). When comparing the results of Table 11, where only gross shape parameter values were provided, and Table 14, where we are looking for the best permanency effect parameters, a great enhancement in performance may be observed. Now the results offer all hits (*h*) around a 90%.

4.5. Influence of number of gray level bands parameter

In the previous sections the number of gray level bands has always been set to $n = 8$. In this section different number of gray level bands will be used and the results are shown.

Table 13
Influence of height–width ratio and compactness features

Frame	Study case 1 ($hw\_object_{min} = 0.5$; $hw\_object_{max} = 1.0$; $c\_object_{min} = 0.5$ $c\_object_{max} = 1.0$)			Study case 2 ($hw\_object_{min} = 0.65$; $hw\_object_{max} = 1.0$; $c\_object_{min} = 0.4$ $c\_object_{max} = 0.65$		
	h	*fp*	*fn*	*h*	*fp*	*fn*
5	71.94	12.44	28.06	61.83	59.78	38.17
6	71.55	14.86	28.45	76.25	56.71	23.75
7	88.22	19.25	11.78	80.93	63.19	19.07
8	85.51	16.59	14.49	82.62	62.41	17.38
9	81.12	23.02	18.88	85.87	5.99	14.13
10	88.61	27.21	11.39	89.25	8.83	10.75
11	81.34	25.10	18.66	89.16	9.68	10.84
12	83.04	20.39	16.96	90.05	11.12	9.95
13	81.29	17.82	18.71	87.25	7.85	12.75
14	83.72	18.49	16.28	85.59	55.38	14.41
15	79.83	19.49	20.17	88.55	10.72	11.45
16	84.91	21.46	15.09	88.98	14.59	11.02
17	84.06	18.48	15.94	85.20	11.10	14.80
18	88.67	20.55	11.33	88.92	13.53	11.08

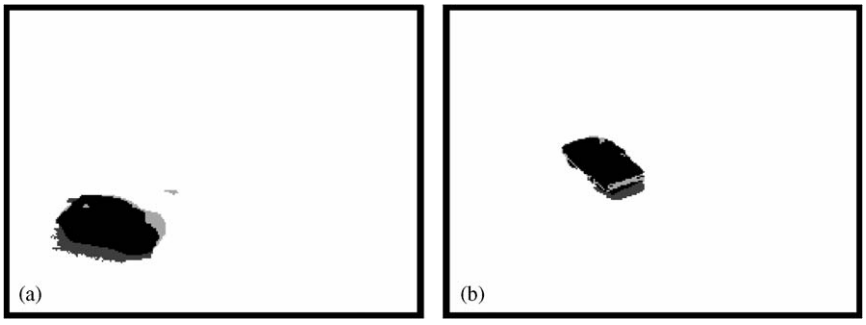

Fig. 9. Graphical results when varying the attention focus for frame 18: (a) study case 1 obtains the "car" and (b) study case 2 obtains the "taxi".

Table 14
Influence of charge and discharge constants, and threshold parameter

Frame	Study case 1 ($C_{AM} = 100$, $D_{AM} = 200$, $\theta = 201$)			Study case 2 ($C_{AM} = 50$, $D_{AM} = 200$, $\theta = 151$)			Study case 3 ($C_{AM} = 50$, $D_{AM} = 250$, $\theta = 201$)		
	h	*fp*	*fn*	*h*	*fp*	*fn*	*h*	*fp*	*fn*
5	88.77	16.95	11.23	78.74	11.33	21.26	67.02	10.40	32.98
6	88.09	16.34	11.91	85.59	15.27	14.41	73.77	11.46	26.23
7	90.89	40.62	9.11	88.83	41.65	11.17	84.98	15.07	15.02
8	89.01	12.87	10.99	88.20	12.36	11.80	84.32	12.15	15.68
9	86.01	18.24	13.99	85.26	16.86	14.74	83.20	16.17	16.80
10	90.43	22.39	9.57	90.14	21.28	9.86	88.90	19.92	11.10
11	87.02	21.24	12.98	85.83	19.64	14.17	84.67	18.89	15.33
12	89.74	17.36	10.26	88.50	16.97	11.50	85.94	16.62	14.06
13	85.94	17.20	14.06	84.77	14.66	15.23	83.82	13.78	16.18
14	87.57	16.71	12.43	86.67	15.37	13.33	84.50	14.01	15.50
15	86.55	18.48	13.45	84.94	16.95	15.06	83.31	16.05	16.69
16	89.06	21.14	10.94	88.23	20.26	11.77	86.50	18.85	13.50
17	87.54	18.29	12.46	86.01	16.40	13.99	84.50	15.77	15.50
18	92.42	20.39	7.58	91.52	19.37	8.48	89.24	18.04	10.76

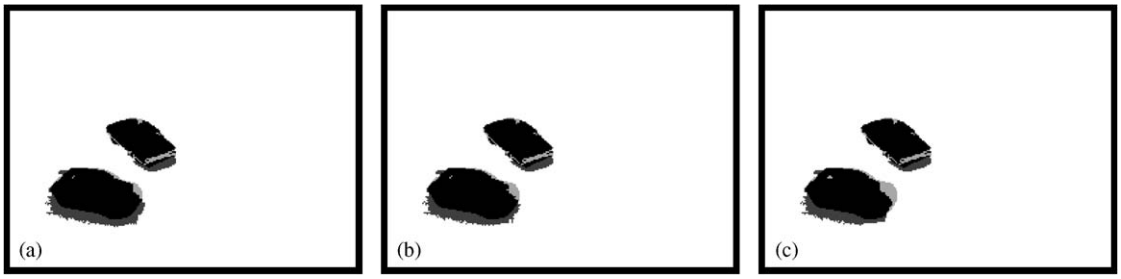

Fig. 10. Graphical results when varying charge and discharge constants and threshold parameter for frame 18: (a) $C_{AM} = 100$, $D_{AM} = 200$, $\theta = 201$; (b) $C_{AM} = 50$, $D_{AM} = 200$, $\theta = 151$; (c) $C_{AM} = 50$, $D_{AM} = 250$, $\theta = 201$.

Table 15
Influence of number of gray levels parameter

Frame	Study case 1 ($n = 4$)			Study case 2 ($n = 32$)		
	h	*fp*	*fn*	*h*	*fp*	*fn*
5	0.00	100.00	100.00	35.00	75.19	65.00
6	29.75	46.99	70.25	61.48	69.14	38.52
7	46.34	42.87	53.66	71.39	66.41	28.61
8	47.49	39.28	52.51	76.48	15.41	23.52
9	58.89	38.47	41.11	84.52	10.80	15.48
10	74.46	36.81	25.54	86.97	11.77	13.03
11	77.47	35.73	22.53	90.44	11.20	9.56
12	81.53	16.74	18.47	86.29	11.08	13.71
13	76.65	14.76	23.35	86.37	11.35	13.63
14	78.50	14.89	21.50	88.65	6.94	11.35
15	78.49	16.87	21.51	67.43	4.63	32.57
16	83.01	19.28	16.99	71.41	8.23	28.59
17	79.15	16.49	20.85	68.25	11.18	31.75
18	83.67	18.96	16.33	71.73	16.50	28.27

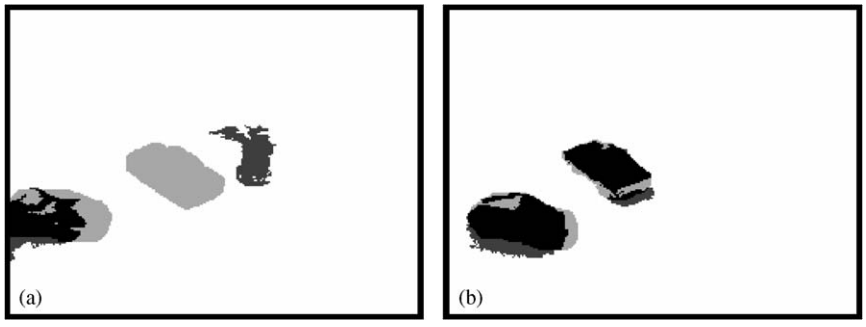

Fig. 11. Graphical results with number of gray level bands $n = 4$: (a) frame number 6, (b) frame number 12.

As it may be observed in Table 15, study case 1, with $n = 4$, in comparison to the results of Table 11, the results are visibly worse. With four gray level bands it is not possible to segment the cars from the road in all frames, and, therefore, they cannot be configured as the attention focus. The results of Table 15 for the study case 1 have been calculated using the ground truth of the "car" and the "taxi". The results may graphically be studied in Fig. 11. There have been performed some more tests for number of gray level bands 16 and 32, but no better results have been gotten for the aim of obtaining the "car" or the "taxi" as the attention focus.

However, when using the case $n = 32$, and by varying the values as described in Table 16, it has been possible to obtain as the attention focus the "van" present in the scene in the lower right corner. This was not possible with a different number of gray level bans (e.g. $n = 4$ and 8). Notice that the "van" can only be obtained when varying at the same time the values for the fundamental shape features. This is the cause for the loss of the "car" and the "taxi" in this new example. The results of

Table 16
Spot shape features parameters for segmenting the "van" (32 gray level bands)

Parameter	Value (in number of pixels)
Spot maximum size: s_spot_{max}	500
Spot maximum width: w_spot_{max}	60
Spot maximum height: h_spot_{max}	50

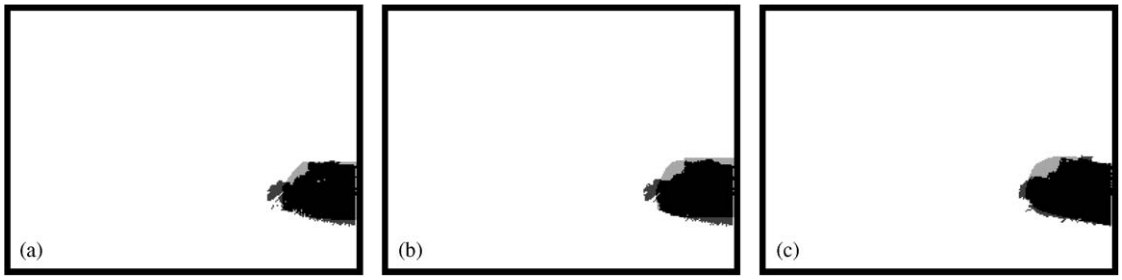

Fig. 12. Graphical results with number of gray level bands $n = 32$: (a) frame number 11; (b) frame number 12; (c) frame number 14.

Table 15 for the study case 2 have been calculated using the ground truth of the "van". Graphically the results are shown in Fig. 12.

5. Conclusions

A model of dynamic visual attention capable of segmenting vehicles and pedestrians in a real visual surveillance scene has been introduced in this paper. The model enables focusing the attention at each moment at shapes that possess certain characteristics and eliminating shapes that are of no interest. The features used are related to motion and shape of the elements present in the dynamic scene. Features extraction and integration are solved by incorporating mechanisms of charge and discharge—based on the permanency effect—, as well as mechanisms of lateral interaction. All these mechanisms prove to be good enough to segment the scene into moving objects—vehicles and pedestrians—and background, without using a reference image or modeling the background. The dynamic visual attention method proposed, due to its image differencing between two consecutive frames step, may be used under changing illumination conditions. Moreover, the model may be used to monitor real environments indefinitely in time.

An example has been offered where, by using a same selection criterion, namely, motion detection between two consecutive image frames, and by slightly changing values corresponding to the desired shape features, the attention focus has been directed towards a pedestrian or a car. The system is stable enough to detect and classify broad classes of elements—such as vehicles and pedestrians—while requiring minimal scene-specific knowledge. Thus, our method applied to visual surveillance may be classified into the more recent works in integration of vehicle and person surveillance. The discussion section has also offered a general view of the capabilities of the method described to tune the parameters in order to change the attention focus to different objects present in a video sequence in accordance to their gray level, motion and shape features.

Acknowledgements

This work is supported in part by the Spanish CICYT TIN2004-07661-C02-01 and TIN2004-07661-C02-02 grants, as well as the Junta de Comunidades de Castilla-La Mancha PBI06-0099 grant.

References

[1] J.-M. Blosseville, Image processing for traffic management, in: Advanced Video-Based Surveillance Systems, Kluwer Academic Publishers, Dordrecht, 1999, pp. 67–75.
[2] M. Pellegrini, P. Tonani, Highway traffic monitoring: main problems and current solutions, in: Advanced Video-Based Surveillance Systems, Kluwer Academic Publishers, Dordrecht, 1999, pp. 27–33.

[3] A. Pozzobon, G. Sciutto, Security in ports: the user requirements for surveillance systems, in: Advanced Video-Based Surveillance Systems, Kluwer Academic Publishers, Dordrecht, 1999, pp. 18–26.
[4] D. Huts, J.-P. Mazy, K. Graf, The prevention of vandalism in metro stations, in: Advanced Video-Based Surveillance Systems, Kluwer Academic Publishers, Dordrecht, 1999, pp. 34–43.
[5] A. Anzalone, A. Machi, Video-based management of traffic light at pedestrian road crossing, in: Advanced Video-Based Surveillance Systems, Kluwer Academic Publishers, Dordrecht, 1999, pp. 49–57.
[6] P. Remagnini, S. Maybank, R. Fraile, K. Baker, R. Morris, Automatic visual surveillance of vehicles and people, in: Advanced Video-Based Surveillance Systems, Kluwer Academic Publishers, Dordrecht, 1999, pp. 96–105.
[7] K.W. Dickinson, C.L. Wan, C.L., Road traffic monitoring using the TRIP II system, Second International Conference on Road Traffic Monitoring, 1989, pp. 56–60.
[8] S. Takaba, M. Sakauchi, T. Kaneko, B. Won-Hwang, T. Sekine, Measurement of traffic flow using real-time processing of moving pictures, 32nd Conference on Vehicular Technology, 1982, pp. 488–494.
[9] R.M. Inigo, Application of machine vision to traffic monitoring and control, IEEE Trans. Veh. Technol. 38 (3) (1989) 112–122.
[10] A.T. Ali, E.L. Dagless, A parallel processing model for real-time computer vision-aided road traffic monitoring, Parallel Process. Lett. 2 (2) (1992) 257–264.
[11] P. Briquet, Video processing applied to road and urban traffic monitoring, Sixth International Conference on Road Traffic Monitoring, 1992, pp. 153–157.
[12] M. Fathy, M.Y. Siyal, A real-time image processing approach to measure traffic queue parameters, IEE Proc. Vision Image Signal Process. 142 (5) (1995) 297–303.
[13] J. Badenas, J.M. Sanchiz, F. Pla, Motion-based segmentation and region tracking in image sequences, Pattern Recognition 34 (3) (1999) 661–670.
[14] J. Badenas, M. Bober, F. Pla, Segmenting traffic scenes from gray level and motion information, Pattern Anal. Appl. 4 (2001) 28–38.
[15] S. Gupte, O. Masoud, R.F.K. Martin, N.P. Papanikiolopoulos, Detection and classification of vehicles, IEEE Trans. Intell. Transport. Syst. 3 (1) (2002) 37–47.
[16] J.R. Bergen, P.J. Burt, R. Hingorani, S. Peleg, A three-frame algorithm for estimating two component image motion, IEEE Trans. Pattern Anal. Mach. Intell. 14 (9) (1992) 886–896.
[17] J.M. Ferryman, S.J. Maybank, A.D. Worrall, Visual surveillance for moving vehicles, Proceedings of the IEEE Workshop on Visual Surveillance, 1998, pp. 73–80.
[18] F. Leymarie, M.D. Levine, Tracking deformable objects in the plane using an active contour model, IEEE Trans. Pattern Anal. Mach. Intell. 15 (6) (1992) 617–634.
[19] T. Darrell, G. Gordon, J. Woodfill, H. Baker, M. Harville, Robust, real-time people tracking in open environments using integrated stereo, color and face detection, Proceedings of the IEEE Workshop on Visual Surveillance, 1998, pp. 26–33.
[20] A. Baumberg, D. Hogg, An efficient method for contour tracking using active shape models, IEEE Workshop on Motion of Non-rigid and Articulated Objects, 1994, pp. 194–199.
[21] M. Loria, A. Machi, Automatic visual control of a pedestrian traffic light, IAPR Workshop on Machine Vision and Application, 1996, pp. 183–186.
[22] R. Howarth, H. Buxton, An analogical representation of spatial events for understanding traffic behaviour, Tenth European Conference on Artificial Intelligence, 1992, pp. 785–789.
[23] T. Huang, S.J. Russell, Object identification: a Bayesian analysis with application to traffic surveillance, Artificial Intelligence 103 (1–2) (1998) 77–93.
[24] M. Haag, H.-H. Nagel, Incremental recognition of traffic situations from video image sequences, ICCV-98 Workshop on Conceptual Description of Images, 1998.
[25] M.I. Posner, M.E. Raichle, Images of Mind, Scientific American Library, New York, 1994.
[26] R. Desimone, L.G. Ungerleider, Neural mechanisms of visual perception in monkeys, in: Handbook of Neuropsychology, Elsevier, Amsterdam, 1989, pp. 267–299.
[27] C. Koch, S. Ullman, Shifts in selective visual attention: towards the underlying neural circuitry, Hum. Neurobiol. 4 (1985) 219–227.
[28] A.M. Treisman, G. Gelade, A feature-integration theory of attention, Cognitive Psychol. 12 (1980) 97–136.
[29] M. Mozer, The Perception of Multiple Objects: A Connectionist Approach, MIT Press, Cambridge, MA, 1991.
[30] D. Heinke, G.W. Humphreys, G. diVirgilo, Modeling visual search experiments: selective attention for identification model (SAIM), Neurocomputing 44 (2002) 817–822.
[31] J.M. Wolfe, Guided Search 2.0. A revised model of visual search, Psychonom. Bull. Rev. 1 (1994) 202–238.
[32] S.P. Vecera, Toward a biased competition account of object-based segregation and attention, in: Brain and Mind, Kluwer Academic Publishers, The Netherlands, 2000, pp. 353–384.
[33] G. Backer, B. Mertsching, Two selection stages provide efficient object-based attentional control for dynamic vision, Proceedings of the International Workshop on Attention and Performance in Computer Vision, 2003, pp. 9–16.
[34] M.A. Fernández, A. Fernández Caballero, M.T. López, J. Mira, Length speed ratio (LSR) as a characteristic for moving elements real-time classification, Real-Time Imag. 9 (2003) 49–59.
[35] M.A. Fernández, J. Mira, Permanence memory: a system for real time motion analysis in image sequences, IAPR Workshop on Machine Vision Applications, 1992, pp. 249–252.
[36] M.A. Fernández, J. Mira, M.T. López, J.R. Álvarez, A. Manjarrés, S. Barro, Local accumulation of persistent activity at synaptic level: application to motion analysis, in: From Natural to Artificial Neural Computation, Springer, Germany, 1995, pp. 137–143.
[37] M.T. López, M.A. Fernández, A. Fernández-Caballero, A.E. Delgado, Neurally inspired mechanisms for the dynamic visual attention map generation task, in: Computational Methods in Modeling Computation, Springer, Germany, 2003, pp. 694–701.
[38] A. Fernández-Caballero, J. Mira, M.A. Fernández, M.T. López, Segmentation from motion of non-rigid objects by neuronal lateral interaction, Pattern Recognition Lett. 22 (14) (2001) 1517–1524.
[39] A. Fernández-Caballero, J. Mira, A.E. Delgado, M.A. Fernández, Lateral interaction in accumulative computation: a model for motion detection, Neurocomputing 50 (2003) 341–364.
[40] A. Fernández-Caballero, M.A. Fernández, J. Mira, A.E. Delgado, Spatio-temporal shape building from image sequences using lateral interaction in accumulative computation, Pattern Recognition 36 (5) (2003) 1131–1142.
[41] A. Fernández-Caballero, J. Mira, M.A. Férnandez, A.E. Delgado, On motion detection through a multi-layer neural network architecture, Neural Networks 16 (2) (2003) 205–222.

About the Author—MARIA T. LOPEZ received her degree in Physics from the University of Valencia (Spain) in 1991 and her Ph.D. from the Department of Artificial Intelligence of the National University for Distance Education (Spain) in 2004. Since 1991, she is an Assistant Professor with the Department of Computer Science at the University of Castilla-La Mancha, Spain. Her research interests are in image processing, computer vision, and neural networks.

About the Author—ANTONIO FERNANDEZ-CABALLERO received his M.Sc. in Computer Science from the Polytechnic University of Madrid (Spain) in 1993 and his Ph.D. from the Department of Artificial Intelligence of the National University for Distance Education (Spain) in 2001. His research interests are mainly in Image Processing, Neural Networks and Agents Technology. He is currently an Associate Professor in the Department of Computer Science of the University of Castilla-La Mancha (Spain).

About the Author—MIGUEL A. FERNANDEZ received his M.Sc. in Physics from the University of Granada (Spain) in 1987 and his Ph.D. from the Department of Artificial Intelligence of the National University for Distance Education (Spain) in 1996. His research interests are mainly in Image Processing and Neural Networks. He is currently an Assistant Professor in the Department of Computer Science of the University of Castilla-La Mancha (Spain).

About the Author—JOSE MIRA is a Professor of Computer Science and Artificial Intelligence and Head of the Department of Artificial Intelligence, National University for Distance Education (Spain). His current research interests include AI fundamentals from the perspective of a knowledge-modeling discipline similar to electronics engineering, neural modeling of biological structures (and application of these models to the design of more realistic artificial neural nets), and integration of symbolic and connectionist problem solving methods in the design of hybrid knowledge based systems. He is the General Chairman of the International Work-conferences on the Interplay between Natural and Artificial Computation.

About the Author—ANA E. DELGADO is a Professor of Computer Science and Artificial Intelligence of the Department of Artificial Intelligence, National University for Distance Education (Spain). Her current research interests include neural modeling of biological structures, algorithmic lateral inhibition, cooperative processes of biological inspiration and fault-tolerant computation.