

A new strategy based on adaptive mixture of gaussians for real-time moving objects segmentation

Carlos Cuevas, Narciso García, and Luis Salgado
Grupo de Tratamiento de Imágenes – E.T.S. Ing. Telecomunicación
Universidad Politécnica de Madrid – Madrid – Spain
{ccr, narciso, lsa}@gti.ssr.upm.es

ABSTRACT

Here, a new and efficient strategy is introduced which allows moving objects detection and segmentation in video sequences. Other strategies use the mixture of gaussians to detect static areas and dynamic areas within the images so that moving objects are segmented [1], [2], [3], [4]. For this purpose, all these strategies use a fixed number of gaussians per pixel. Typically, more than two or three gaussians are used to obtain good results when images contain noise and movement not related to objects of interest. Nevertheless, the use of more than one gaussian per pixel involves a high computational cost and, in many cases, it adds no advantages to single gaussian segmentation. This paper proposes a novel automatic moving object segmentation which uses an adaptive variable number of gaussians to reduce the overall computational cost. So, an automatic strategy is applied to each pixel to determine the minimum number of gaussians required for its classification. Taking into account the temporal context that identifies the reference image pixels as background (static) or moving (dynamic), either the full set of gaussians or just one gaussian are used. Pixels classified with the full set are called MGP (Multiple Gaussian Pixel), while those classified with just one gaussian are called SGP (Single Gaussian Pixel). So, a computation reduction is achieved that depends on the size of this last set. Pixels with a dynamic reference are always MGP. They can be Dynamic-MGP (DMGP) when they belong to the dynamic areas of the image. However, if the classification result shows that the pixel matches one of the gaussian set, then the pixel is labeled static and therefore it is called Static-MGP (SMGP). Usually, these last ones are noise pixels, although they could belong to areas with movement not related to objects of interest. Finally, pixels with a static reference that still match the same gaussian are SGP and they belong to the static background of the image. However, if they do not match the associated gaussian, they are changed either to SMGP or DMGP. In addition, any pixel can maintain its status and SMGP can be changed to DMGP and SGP. A state diagram shows the transition schemes and its characterizations, allowing the forecasting of the reduction of the computational cost of the segmentation process. Tests have shown that the use of the proposed strategy implies a limited loss of accuracy in the segmentations obtained, when comparing with other strategies that use a fixed number of gaussians per pixel, while achieving very high reductions of the overall computational cost of the process.

Keywords: segmentation, background subtraction, mixture of gaussians, adaptive dynamic number of gaussians, real-time.

1. INTRODUCTION

Nowadays, the segmentation analysis in video sequences is essential in a big number of applications. Some of these applications are, for example: video-surveillance, monitoring, people's movement analysis, human-machine interaction, and video coding based on objects (MPEG-4). In all these applications, the segmentation process is a key step that avoids foreground objects identification and makes possible object features extraction, and the measuring of some of their main characteristics. Also, segmentation process makes feasible to proceed with image analysis and to obtain a high-level image comprehension too [4].

An important number of problems appears in these segmentation processes, and many methods that try to avoid them can be found. One possible list of the most relevant problems is the following [1]: all kind of changes in the illumination (abrupt and slow changes), shadows, slow variations at the background, repetitive movements (tree leaves, flicking flags, snow...), objects that at the beginning of the sequence are part of the background and later disappear, new objects that first were dynamic objects and later are part of the background (static objects), and others.

To avoid all these problems, a large number of investigations have been developed. These investigations can be classified in two main groups [5]: techniques based on optical flow analysis, and techniques based on background subtractions. The first one makes use of the analysis of the local velocity of the images (optical flow), and the work proposed by Horn and Schunck [6] has been the reference work for the most posterior investigations based on optical flow. In the other hand, background subtraction techniques make use of the movement information along the sequence and an analysis of the differences in each pixel along the time [7].

In the literature, many background subtraction methods can be found. The main differences between them depend on the obtained results. These results are different attending to the characteristics of the sequences and to the environment conditions. Usually, the quality of these methods is evaluated measuring their speed (computational cost) and their memory requirements. In the review proposed in [5], a comparison of the most important background subtraction techniques is presented. All of these works make possible generating a binary mask that classifies the pixels of the images as static or dynamic. Some of these methods are based on the average of images over time (these are useful only for short sequences where there are not important changes), other ones, like shown in [8] and [9], make use of adaptive gaussians to represent color values of each pixel of the background (the main problems of these techniques are that they are not efficient in the presence of illumination changes and that they are unimodal). In [10], a non-parametric Kernel density strategy that is used instead of a gaussian model is proposed which offers better results in presence of illumination changes, but whose main limitation is that it is not a multimodal method. The most popular method was proposed by Stauffer and Grimson [1]. It makes use of a mixture of gaussians to obtain an adaptive model of each one of the image pixels. The use of more than one gaussian makes that the proposed background subtraction technique will be a multimodal one. The popularity of this method results in a big number of works that use the mixture of gaussians to model backgrounds ([2], [3], [11], [12], [13]).

In many cases, the background areas of the sequences are not totally static. This is the case, for example, of sequences where appear trees, flags, sea waves, rain, snow, etc. These kinds of elements, commonly found in outdoor video sequences, present cyclic (non-static) movements, which need to be classified as part of the background. Making use of techniques with a unique mode (unimodal techniques), the classification of these areas as background is not possible: due to with one mode, a unique state can be modeled for each pixel, and pixels with cyclic movements can't be modeled taking into account all their possible states. On the other hand, using a mixture of gaussians (multimodal technique), it is possible to model more than one state for each pixel, and then it is possible to classify the elements with cycling movements as background objects [1].

Nevertheless, the main drawback of the mixture of gaussians method is its associated high computational cost. This cost increases when the number of used gaussians grows. Typically, to obtain good results, between 3 and 5 gaussians per pixel are required, a number which can seriously limit the use of this method in some applications where accurate segmentations are required to operate in real time (video-surveillance, monitoring...).

In this paper, a new background subtraction technique is proposed. This technique uses a mixture of gaussians to model each one of the pixels of the images, but instead of any fixed number of gaussians per pixel, an adaptive variable number of gaussians is computed for each pixel. The chosen number of gaussians depends on the characteristics of the sequence areas. Pixels on areas without movement will be modeled with a lower number of gaussians than those on areas with cyclical movements or areas with movement. This adaptive selection of the number of gaussians makes possible a correct classification as background, of all the static areas and areas with cyclical movement, while an overall reduction of the computational cost. The final number of gaussians used in each image will be smaller in comparison with the traditional case, where a fixed number of gaussians per pixel is assigned. Therefore, will be possible the reduction of the overall computational cost.

In this way, and depending on the characteristics of each area of the images, pixels could have assigned up to 5 gaussians. Usually, the areas where more than one gaussian is needed (small areas where objects with cyclical movements appear and areas exposed to light changes) are smaller than areas that need a simple gaussian. Results obtained show very good segmentations, closer to the original segmentation using 5 gaussians for each pixel, while significantly reducing the computational cost. The proposed technique has been evaluated over different sequences and, compared with other strategies that use a fixed number of gaussians. The computational cost has been reduced while

obtained results are similar to background subtraction using 5 gaussians per pixel in all the images. The used sequences may be found in PETS data base (<ftp://ftp.pets.rdg.ac.uk>).

This paper is organized as follows. Section 2 introduces de Mixture of Gaussians (MoGs) segmentation algorithm. Section 3 describes in detail, the proposed method and its implementation, while section 4 is dedicated to the analysis of results. Finally, section 5 shows the conclusions of the paper.

2. SEGMENTATION BASED ON MIXTURE OF GAUSSIANS

The work presented in this paper is based on the adaptive multimodal background subtraction method, proposed by Stauffer and Grimson [1]. In this method, a mixture of K gaussian distributions is used to model luminance changes, at each pixel location, in the imaging scene over the time. If the current image is $I_n(x,y)$, where 'n' is the image position into the sequence and (x,y) are the coordinates of each one of the pixels, the correspondent probability density function for this image is defined as the weighted sum of K gaussian distributions:

$$f(I_n(x,y)) = \sum_{i=1}^K \omega_{n,i}(x,y) \cdot \eta_i(I_n(x,y); \mu_{n,i}(x,y); \sigma_{n,i}(x,y)) \quad (1)$$

where η_i is the i -th gaussian probability density function, $\omega_{n,i}$ is the weight of the i -th gaussian distribution at current image, $\mu_{n,i}$ is its mean and $\sigma_{n,i}^2$ is its variance. The gaussian probability density function is formulated as follows:

$$\eta_i(I_n(x,y); \mu_{n,i}(x,y); \sigma_{n,i}(x,y)) = \frac{1}{\sigma_{n,i}(x,y)\sqrt{2\pi}} \cdot \exp\left[-\frac{(I_n(x,y) - \mu_{n,i}(x,y))^2}{2\sigma_{n,i}^2(x,y)}\right] \quad (2)$$

With each new incoming frame the gaussian distributions are updated, and then used to determine which pixels are most likely to result from a background process. The update equations are:

$$\rho = \alpha \cdot \eta_i(I_n(x,y); \mu_{n,i}(x,y); \sigma_{n,i}(x,y)) \quad (3)$$

$$\mu_{n,i}(x,y) = (1 - \rho) \cdot \mu_{n-1,i}(x,y) + \rho \cdot I_n(x,y) \quad (4)$$

$$\sigma_{n,i}^2(x,y) = (1 - \rho) \cdot \sigma_{n-1,i}^2(x,y) + \rho \cdot (I_n(x,y) - \mu_{n-1,i}(x,y))^2 \quad (5)$$

where $\mu_{n-1,i}$ and $\sigma_{n-1,i}^2$ are the mean and the variance corresponding to the image $n-1$, α is the learning rate (reflects the speed of the current pixel fused into background), and ρ defines the importance of the new image values compared with the current values of the stored gaussians. Finally, the weighting factors of the K gaussian distributions are updated by:

$$\omega_{n,i}(x,y) = (1 - \alpha) \cdot \omega_{n-1,i}(x,y) + \alpha \quad (6)$$

This model allows a proper representation of the background scene undergoing slow lighting, and scene changes as well as momentary variations. Also, this method can eliminate misclassifications due to cyclical motion in the background (such as moving foliage, wavering tree branches, sea waves, snow, rain, flags movements...). According to different authors, it is required to use between 3 and 5 gaussians per pixel to obtain good results. However, the main problem of this model is the high computational cost that results of its application with a large number of gaussians as, for each pixel

of the image, it is required to store three values per gaussian (mean, variance and weight) and these values need to be updated at each new image.

The proposed method, which is detailed in the next section, is a variation of the original Stauffer and Grimson method. While the original strategy uses a fixed number of gaussians per pixel, the presented approach uses a variable number of gaussians depending on the characteristics of the pixels, and therefore reducing the computational cost of the segmentation process. Also, further investigations show that, for some specific applications, MoGs method with more than 2 gaussians can degrade the performance for foreground detection ([14], [15]). Therefore, adapting the number of gaussians might result in a segmentation improvement.

3. VARIABLE NUMBER OF GAUSSIANS STRATEGY

3.1 State diagram

This paper proposes a novel automatic moving object segmentation which uses an adaptive variable number of gaussians to reduce the overall computational cost. So, an automatic strategy is applied to each pixel to determine the minimum number of gaussians required for its classification. Taking into account the temporal context that identifies the reference image pixels as background (static) or moving (dynamic), either the full set of gaussians or just one gaussian are used. Pixels classified with more than one gaussian are called MGP (Multiple Gaussian Pixel), while those classified with just one gaussian are called SGP (Single Gaussian Pixel). The main computation reduction depends on the size of this last set (SGP). Moreover, this reduction depends of the number of gaussians used in the MGP.

Pixels with a dynamic reference are always MGP. They can be Dynamic-MGP (DMGP) when they belong to the dynamic areas of the image. However, if the classification result shows that the pixel matches one of the gaussian set, then the pixel is labeled static and therefore it is called Static-MGP (SMGP), and belongs to the background. Finally, pixels with a static reference that still match a unique gaussian are SGP and they belong to the static background of the image too. However, if they do not match the associated gaussian, they are changed either to SMGP or DMGP. In addition, any pixel can maintain its status and SMGP can be changed to DMGP and SGP.

The state diagram showing the transition schemes and their characterizations is shown in Fig. 1. So, for each pixel, the sub-states (named as $\{S_i\}_{i=1..6}$) of the three main states (SGP, SMGP and DMGP) are identified.

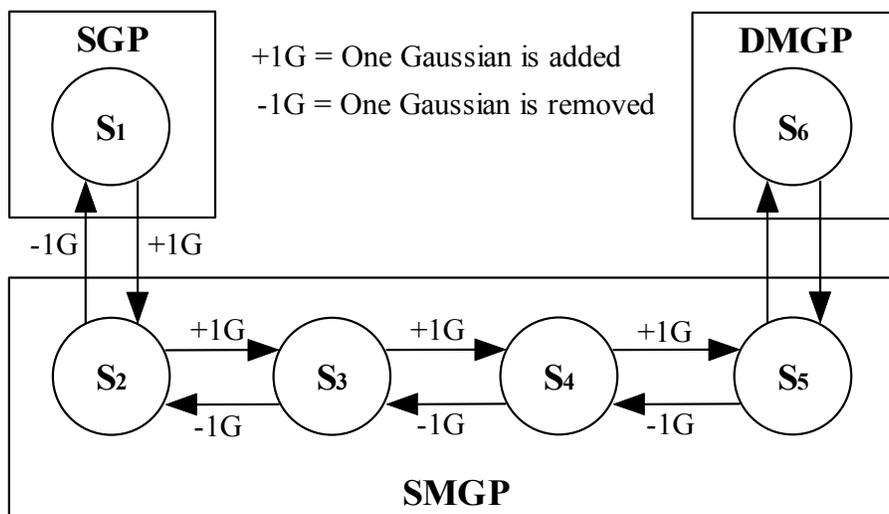


Fig. 1. State diagram of the pixels: three main states (SGP, SMGP and DMGP) and six sub-states (S₁, S₂, S₃, S₄, S₅ and S₆).

The procedure runs as follows. All pixels are initialized in the first image of the sequence as S_1 . In this sub-state only one gaussian is assigned to the pixel, so it is classified as a SGP. Typically, pixels belonging to areas without movement will show this sub-state for most part of the sequence, so that this sub-state will be, usually, the most frequent appeared sub-state.

When the intensity value of a pixel changes (e.g. due to moving objects) a new gaussian is added and the pixel change to the state SMGP (sub-state S_2). If two gaussians are not enough, another one is added (sub-state S_3). In the same way the pixels could have four gaussians (sub-state S_4) and five gaussians (sub-state S_5), if necessary. When one of these changes occurs, pixels are classified as dynamic, but if in the following image they continue in the same sub-state, they are reclassified as static (background pixels). The maximum number of gaussians is five, due to the really high computational cost implied by the use of more gaussians.

However, if five gaussians are not enough to represent a pixel variation, it will pass to the sub-state S_6 . Although in this sub-state pixels have five gaussians assigned, they are classified as dynamic (foreground pixels). A pixel continues in this state until its value can be fully modeled with its five associated gaussians. In this case, the pixel returns to S_5 and it is classified as SMGP.

An SMGP pixel can reduce its number of gaussians from five to two. This occurs when it is detected that one gaussian is not necessary to model the pixel values. If this happens, the total number of gaussians used for the pixel is reduced, and the pixel changes its associated sub-state from S_i to S_{i-1} . If $i=2$, and the same conditions occur (one of the two gaussians is considered not to be necessary), the pixel changes to SGP with only one gaussian assigned.

3.2 Implementation

In this section, the criteria and operations used to apply the variable number of gaussians strategy are detailed. In the first frame, according to the actual intensity value of the pixels ($I_1(x,y)$), one gaussian is assigned to each pixel ($\mu_{1,1}(x,y)$, $\sigma_{1,1}(x,y)$, $\omega_{1,1}(x,y)$). For the next images, Eq. (7) is evaluated for each one of them pixels. If some of the pixels satisfy this equation, the gaussian is updated and the pixel stands in S_1 . In the other hand, according to the new pixel value, a new gaussian is added. First, the pixel passes to S_2 with two associated gaussians. If, for the following images, new pixel values can't be represented by the existing gaussians (the gaussians associated to the pixel do not satisfy Eq. (7)), a new gaussian is added and the pixel passes from S_2 to S_3 (three gaussians). In the same way, if new gaussians are needed, the pixel changes from S_3 to S_4 (four gaussians), and from S_4 to S_5 (five gaussians).

$$\left| I_n(x,y) - \mu_{n-1,i}(x,y) \right| < 2.5 \cdot \sigma_{n-1,i}(x,y) \quad (7)$$

Nevertheless, if the pixel is in one of the sub-states and for its new values new gaussians are not needed (one of the existent gaussians satisfy the previous expression), the pixel will have two options: it can stand in its actual sub-state (S_i), or it can moves to the previous sub-state (S_{i-1}). This last option implies the suppression of one of the existing gaussians that are assigned to the pixel. One gaussian is removed when two conditions are satisfied, which reflects that the gaussian has low relevance to model the pixel values.

- In the first condition (Eq. (8)), for each new pixel at each new image, the weight of each one of the existing gaussians is compared with a weighted sum of all the gaussian weights. If one of the existing gaussians has a smaller weight that the weighted sum, this gaussian will be a candidate to be removed. To do this, a new parameter is used (β), which determine the minimum weight of one gaussian to be removed, taking values between '0' and '1'. If β is close to '0' the weight of one gaussian should be very small to satisfy Eq. (8) and, therefore, the reduction of the number of gaussians will be more difficult. On the other hand, if β value is increased, it is easier to satisfy this equation, and gaussians will be deleted more frequently.

$$\omega_{n,i}(x, y) < \beta \cdot \sum_{i=0}^K \omega_{n,i}(x, y) \quad (8)$$

- In the second condition, to be sure that one candidate gaussian should be removed, Eq. (8) should be satisfied for the next N values of the pixel. Depending on the value of α , N should be larger or smaller. Obviously, the appropriate N value to be used is relative to the selected α value. The lower values of α , the longer number of images is required to update the initial weighting factor of a new gaussian to stop satisfy Eq. (8).

Additionally, a small number of gaussians will be removed along the sequence if these conditions are not satisfied (for low values of β and high values of N), and the computational cost will be closer to that using a fixed number of gaussians. On the other size, if these values are too high for β , and too low for N, the obtained segmentations will have more misclassifications.

4. RESULTS

The proposed method has been evaluated against the original technique based on mixture of gaussians. Extracted from the PETS data base (<ftp://ftp.pets.rdg.ac.uk>), two sequences with different characteristics have been used.

The first sequence (henceforth Sequence 1) is composed of 2823 grey level images of 576x768 pixels. It is an outdoor recording, where several moving elements appear and some background elements make repetitive movements (waving trees due to the wind and illumination changes in some points of the existing buildings). The second sequence (henceforth Sequence 2) is composed of 5336 grey level images with 576x768 pixels. The recorded scenario is the same than in Sequence 1 but now there are very significant illumination changes along it besides many moving objects appear.

First, a comparison between three background subtraction methods is done: the traditional MoGs using 5 gaussians per pixel, segmentation using only one gaussian per pixel, and the proposed method (variable number of gaussians). Due to the nature of the sequences (outdoor scenes), a small value of α is used ($\alpha = 0.001$), while $N = 80$ and $\beta = 0.07$. To remove the resulting noise after the segmentation process to get compact object masks, two morphological operations have been applied after pixels classification: first an “opening” operation with a square mask of 3x3 pixels, and then a “closing” operation using a square mask of 11x11 pixels.

Segmentation results from the three aforementioned methods, for one image of each sequence, are shown in Fig. 2 (image 1411 of Sequence 1 and image 725 of Sequence 2). It presents two original images and their binary masks obtained after the segmentation process and after the morphological operations. In the masks white pixels correspond to those that have been classified as dynamic and black pixels classified as background.

The second row of Fig. 2 shows the segmentation obtained using only one gaussian per pixel (unimodal segmentation). In the result obtained for the Sequence 1 (left column), many objects appear which are part of the background appear. Some of these objects have been segmented correctly, but others are the result of a misclassification of some pixels: objects which show up on the outline of the tree situated in the centre of the image. Using only one gaussian per pixel (unimodal segmentation), it is not possible to model the variations is there areas. For Sequence 2 (right column), there are many pixels whose result from a wrong segmentation: large part of the resultant binary mask displays incorrectly segmented areas, which correspond to illumination changes, which affects to a very high number of pixels. Using only one gaussian per pixel, these changes cannot be modeled.

The third row of Fig. 2 shows results from the application of MoGs method with 5 gaussians per pixel. For Sequence 1 (left column), the obtained results show less segmented objects compared to the previous method (with one gaussian per pixel). This is due to 5 gaussians per pixel can model the pixels variations in background areas with small movements (like edges of the tree), while one gaussian per pixel can't do it. In the case of Sequence 2, occurs the same: there are less foreground pixels in compare to the previous case, due to illumination changes can be modeled better with more than one gaussian.

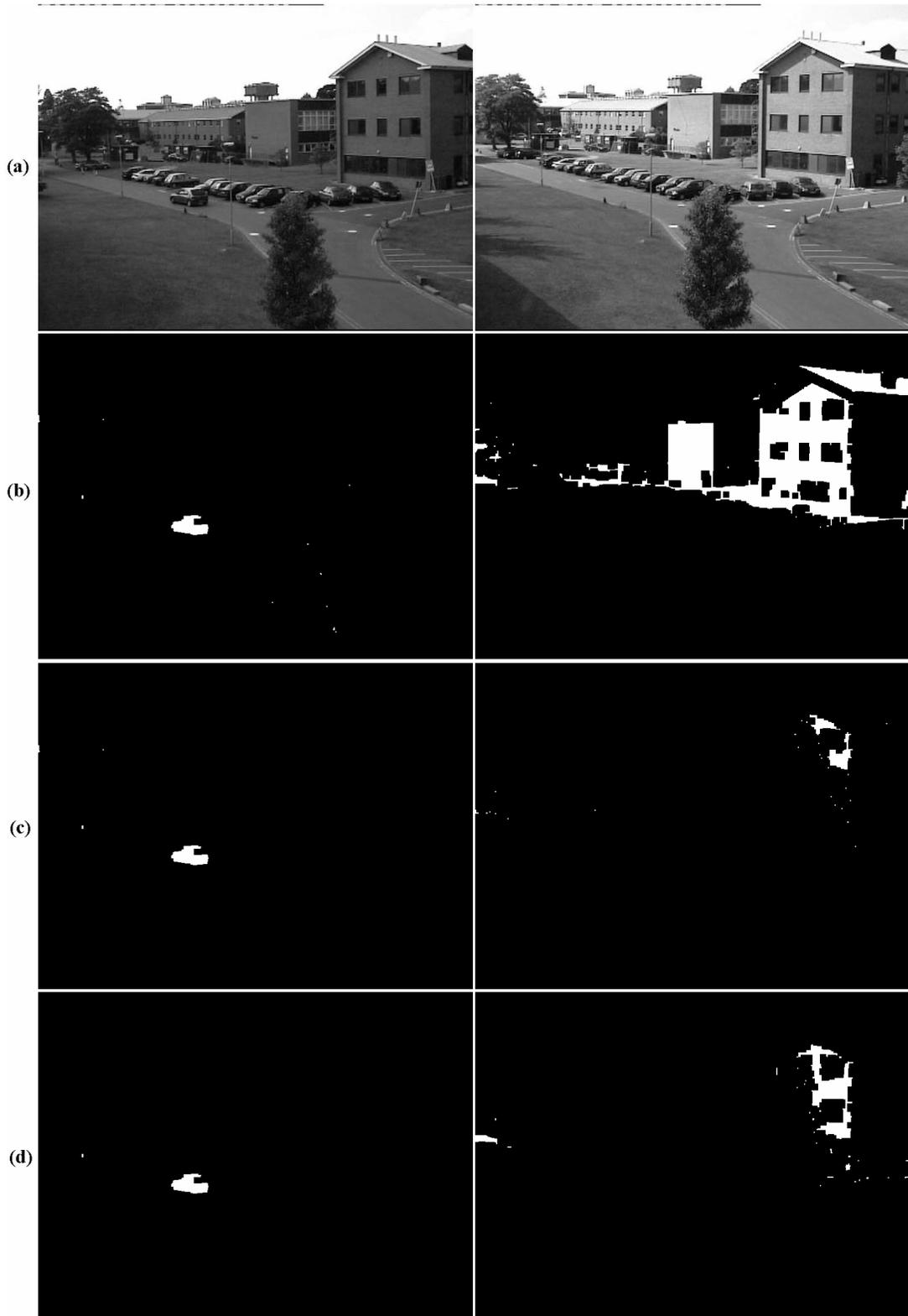


Fig. 2. Two original images (a) and the resulting binary masks of three different segmentation methods: (b) one gaussian per pixel segmentation, (c) five gaussians per pixel segmentation, (d) variable number of gaussians segmentation.

The last row of Fig. 2 shows the results obtained with the proposed method which consists, like it is known, in a variable number of gaussians assigned to each pixel. Comparing these results with the previously exposed: for Sequence 1, a correct classification of the non-static areas has been done (the areas around the tree are not in the segmented image and therefore this result is very close to the which one obtained with the MoGs); for Sequence 2, the obtained segmentation is better than the segmentation using only a gaussian per pixel and it is very similar to the segmentations with 5 gaussians. This is due to with a variable number of gaussians, has been possible to assign more gaussians to the pixels with more variations along the time and this make possible the correct classification of most of pixels.

Therefore, results over Sequence 1 show that, using only one gaussian per pixel, some background objects can be classified as foreground, but using more gaussians only in the areas where these objects appear, a correct segmentation can be done. In the same way, results over Sequence 2 show that, using a variable number of gaussians per pixel, illumination changes are modeled better than using one gaussian at each pixel, and the results obtained with this method are similar to the which ones obtained with 5 gaussians, with a very significant reduction in the computational cost.

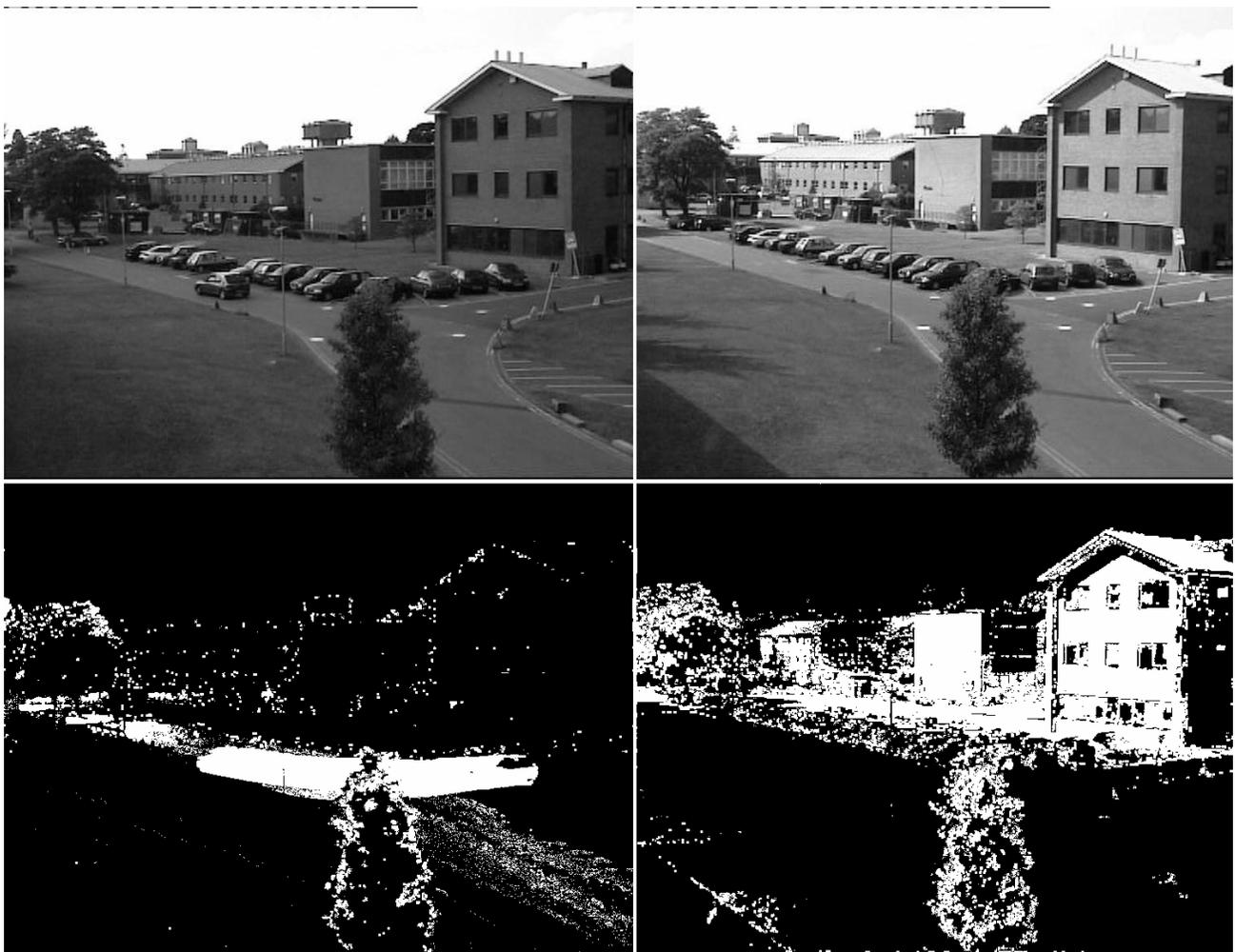


Fig. 3. Original images and their associated binary masks of assigned gaussians.

To achieve a better understanding on the image areas that have more than one gaussian assigned and those which have a single one, Fig. 3 displays (same images as before) the pixels in which more than one gaussian have been assigned in the segmentation process (second row). These images represent as white color the image areas where more than one gaussian per pixel are being assigned, while the black areas show the pixels modeled with just one gaussian. At the image of the Sequence 1 (left column), the 92.21 % of the pixels have been modeled using only one gaussian and the 7.79 % are pixels modeled with more than one gaussian (these pixels appear in the trees, in the trail of the car, and in the edges of the buildings where illumination changes appear). On the other hand, for the image of the Sequence 2 (right column), the 73.37 % of the pixels have been modeled by solely one gaussian and the 26.63 % with more gaussians (areas with important lighting changes).

These data states that, in both of cases a big number of pixels need only one gaussian to be modeling appropriately (these pixels are those in which important changes do not occur). The remaining pixels have more than one gaussian allotted for treating to model their variations.

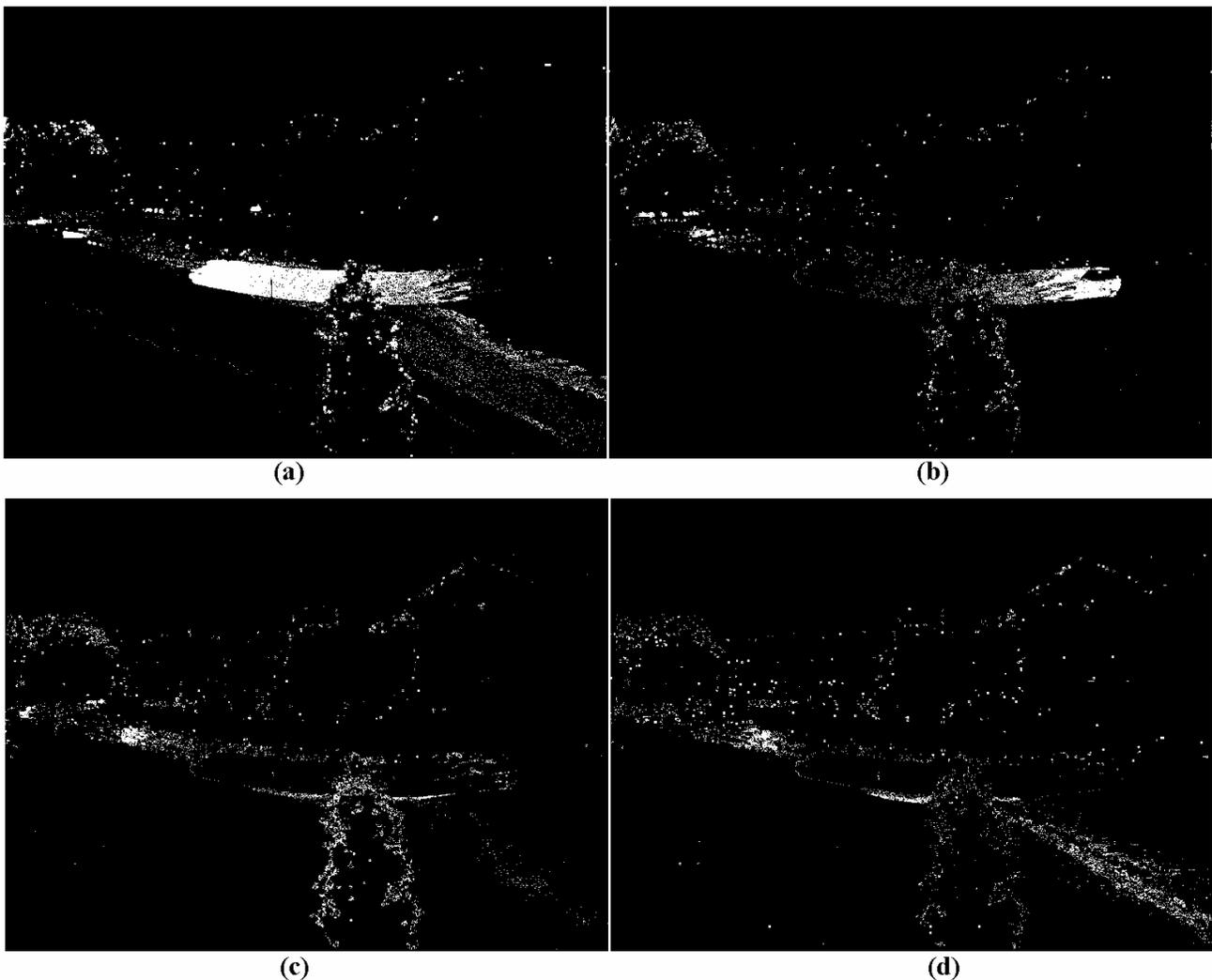


Fig. 4. For Sequence 1: (a) pixels with five associated gaussians, (b) pixels with four associated gaussians, (c) pixels with three associated gaussians, (d) pixels with two associated gaussians.

Fig. 4 shows four different binary masks where pixels with 5, 4, 3 and 2 gaussians are showed. The first mask (a) shows the pixels that have been modeled by 5 gaussians. So, this image allows realizing that pixels with 5 associated gaussians are those ones where mobile objects appear (besides, the areas where those objects appeared in previous images) and some of the background pixels: pixels pertaining to not totally static objects (like the trees) and areas with illumination changes (for example, some edges in the buildings).

The second mask (b) of Fig. 4 shows the pixels with 4 assigned gaussians. These pixels appear in the same background areas than pixels with 5 gaussians (trees and buildings). In this image, in possible to see that part of the trail of the car is marked. This is due to when one pixel has 5 associated areas (dynamic objects like the car), and in the following images the pixel is static, its number of associated gaussians is reduced one by one (pass from 5 to minimum necessary visiting the intermediate sub-states with 4, 3 and 2 gaussians).

The third (c) and the fourth (d) masks represent the pixels with 3 and 2 associated gaussians respectively. These images show that there are pixels in areas with little movements and illumination changes where, only 2 or 3 gaussians have been needed to model the pixels variations.

Table 1. Distribution of number of used gaussians for an example image.

	Sequence 1 (frame 1411)	Sequence 2 (frame 725)
5 gaussians	3.48 %	10.37 %
4 gaussians	1.19 %	4 %
3 gaussians	0.96 %	4.16 %
2 gaussians	2.16 %	4.10 %
1 gaussian	92.21 %	77.37 %

Table 2. Distribution of number of used gaussians for complete sequences.

	Sequence 1	Sequence 2
5 gaussians	2.85 %	12.87 %
4 gaussians	1.94 %	6.91 %
3 gaussians	2.42 %	9.47 %
2 gaussians	3.43 %	6.87 %
1 gaussian	89.36 %	63.88 %

Table 1 shows the amount of pixels with 1, 2, 3, 4 and 5 assigned gaussians with respect to the total image pixels. The first column of the table contains information about the previous used image for the Sequence 1, and the second about the previous used image for the Sequence 2. Therefore, it is possible to see that, in both images, only one gaussian per pixel is assigned usually (92.21 % in the image 1411 of the first sequence, and 77.37 % in the image 725 of the second sequence).

Table 2 shows similar results that Table 1, but for the complete sequences. Now, the pixels with a unique associated gaussian continue being the most frequents. Sequence 1 has a total of 2823 images where only appear some real dynamic objects. In some areas of the images, this sequence suffer changes in the trees (because of the wind) and in some areas of the buildings (because of the illumination changes). To model these areas has been necessary more than one gaussian per pixel, but the rest of the images areas have been needed one gaussian per pixel. Therefore, in all the sequence an 89.36 % of the pixels have been modeled by one gaussian. In Sequence 2, pixel changes occur more frequently (there are important variations in the illumination and dynamic objects appear too). For this reason it is necessary the use of more than one gaussian in a bigger number of pixels in compare to Sequence 1. In this case the 77.37 % of the pixels have been modeled by one gaussian and the rest of the pixels by more than one gaussian attending to them necessities.

The obtained results show that, in most of outdoor sequences, a very large amount of pixels is correctly modeled using only one gaussian per pixel. Therefore, only small areas of the images need more gaussians to avoid misclassifications

after the segmentation process. This reduction of the number of used gaussians implies an important reduction in the computational cost. Experiments have been performed to measure the cost of segmentation. Results show that, compared to MoGs with 5 gaussians per pixel, the proposed segmentation obtains significant reductions in the computational cost: 37.5 % in Sequence 1 and 18.65 % in Sequence 2, computed as:

$$\Delta T(\%) = \frac{T_M - T_5}{T_5} \cdot 100 \quad (9)$$

where T_M is the measured time in milliseconds spent by the proposed method; and T_5 is the measured time for the MoGs method.

As a consequence of all the tests conducted and therefore the results presented in this section, it is possible to obtain an important reduction of the required computational cost using the Variable Number of Gaussians Technique. In addition, it is quite important to realize that the achieved segmentations offer similar results compared to those obtained with the original Mixture of Gaussians Method using a large number of gaussian per pixel.

5. CONCLUSIONS

A novel automatic moving object segmentation technique based on background subtraction has been exposed. Although it is based on the work proposed by Stauffer and Grimson in [1], it contains significant differences. Nevertheless, it is based on the utilization of a variable assignment of the number of gaussians used in each one of the pixels.

The proposed automatic technique uses an adaptive variable number of gaussians to reduce the overall computational cost. So, an automatic strategy is applied to each pixel to determine the minimum number of gaussians required for its classification. Taking into account the temporal context that identifies the reference image pixels as background (static) or moving (dynamic), either the full set of gaussians or just one gaussian are used.

Tests have shown that the use of the proposed strategy implies a limited loss of accuracy in the segmentations obtained, when comparing with other strategies that use a fixed number of gaussians per pixel, while achieving very high reductions of the overall computational cost of the process.

6. ACKNOWLEDGMENTS

This work has been partially supported by the Ministerio de Ciencia y Tecnología of the Spanish Government and FEDER/ERDF funds through project TIN2004-07860 (Medusa) and by the Comunidad de Madrid under project P-TIC-0223-0505 (Pro-Multidis).

REFERENCES

- ¹ W. E. L. Grimson, and C. Stauffer. "Adaptive background mixture models for real-time tracking", CVPR, Vol. 2, pp. 246-252, 1999.
- ² S. Atev, O. Masoud, and N.Papanikopoulos. "Practical Mixture of Gaussians with Brightness Monitoring", IEEE Intelligent Transportation Systems Conference, pp. 423-428, October 2004.
- ³ P. Dickinson, and A. Hunter. "Scene Modelling Using An Adaptive Mixture of Gaussians in Colour and Space", Advanced Video and Signal Based Surveillance, pp. 64-69, September 2005.
- ⁴ W. Chuan-xu, Z. Xiang-guang, Y. Chun-feng, and L. Yun. "Video Segmentation of Illuminance Abrupt Variation Based on MOGs and Interframe Gradient Cross-correlation", ICSP2006 Proceedings, DOI 10.1109/ICOSP.2006.344456, November 2006.

- ⁵ M. Piccardi, "Background subtraction techniques: a review". International Conference on Systems, Man and Cybernetics, pp. 3099-3104, 2004.
- ⁶ B. Horn, y B. Schunck, "Determining optical flow". Artificial Intelligence, No. 17, pp. 185-203, 1981.
- ⁷ A. K. Jain, "Image Data Compression: A Review". Proceedings of the IEEE, Vol. 69, No. 3, March 1981.
- ⁸ C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfindex: Real-Time Tracking of the Human Body". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 780785, July 1997.
- ⁹ T. Zhao, and R. Nevatia, "Tracking Multiple Humans in Complex Situations". IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26, No. 9, pp. 1208-1221, September 2004.
- ¹⁰ A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and Foreground Modeling Using Non-Parametric Kernel Density Estimation for Visual Surveillance". Proceedings of the IEEE, Vol. 90, No. 7, pp. 1151-1163, July 2002.
- ¹¹ H. Wang, and D. Suter, "A re-evaluation of mixture of Gaussian background modeling [video signal processing applications]". IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, pp. 18-23, March 2005.
- ¹² A. A. Farag, A. S. El-Baz, and G. Gimel'farb, "Precise segmentation of multimodal images". IEEE Transactions on Image Processing, Vol. 15, No. 4, pp. 952-968, April 2006.
- ¹³ J. L. Landabaso, and M. Pardàs, "Foreground Regions Extraction and Characterization Towards Real-Time Object Tracking". Proceedings of Multimodal Interaction and Related Machine Learning Algorithms, Lecture Notes in Computer Science, Springer, 2005.
- ¹⁴ T. E. Boult, R. Micheals, X. Gao, P. Lewis, C. Power, W. Yin, and A. Erkan, "Frame-rate omnidirectional surveillance & tracking of camouflaged and occluded targets". Proceedings IEEE Workshop on Visual Surveillance, pp. 48-55, 1999.
- ¹⁵ X. Gao, T. Boult, F. Coetzee, and V. Ramesh, "Error analysis of background adaptation". Proceedings of IEEE Computer Vision and Pattern Recognition, pp. 503-510, 2000.