چهارمین کنفرانس ماشین بینایی و پردازش تصویر ایران

۲۵-۲۶ بهمن ماه ۱۳۸۵

A Background Model Initialization Algorithm Based on QR-Decomposition

M. Fathy
Iran University of Science and Technology, Iran
mahFathy@iust.ac.ir

M. Analoui
Iran University of Science and Technology, Iran

N. Mozayani
Iran University of Science and Technology, Iran

M. Amintoosi
Iran University of Science and Technology, Iran
mAminToosi@iust.ac.ir

F. Farbiz
Iran University of Science and Technology, Iran
farbiz@iust.ac.ir

Abstract: *Background subtraction is a major part of many motion detection, tracking and surveillance systems. In this paper a new algorithm for the purpose of the background model initialization has been presented. The key idea of the proposed method lies in the identification of the background based on QR-Decomposition method in linear algebra. R-values produced with QR-Decomposition can be applied to decompose a given system to indicate the degree of the significance of the decomposed parts. We split the image into small blocks and select the background blocks with the weakest contribution, according to the assigned R-values. The main advantage of the proposed method is that in contrast to many other methods, here, there is no need for an empty scene with no foreground object. Simulation results showed that the proposed method produced better background model with respect to some others.*

Keywords*:* *Background Subtraction, QR-Decomposition, Gaussain Mixture Model, Singular Value Decomposition.*

1 Introduction

Segmentation of the moving regions, so called as foreground, from the static part of a scene, commonly named as background, is one of the most fundamental tasks in computer vision with a wide spectrum of applications from compression to scene understanding. Background subtraction is a widely used approach for detecting moving objects in videos from static cameras. The rationale in the approach is that of detecting the moving objects from the difference between the current frame and a reference frame, often called the "background image" or "background model". As a basic, the background image must be a representation of the scene with no moving objects. Several methods for performing background subtraction have been proposed in the recent literatures [1, 2, 3, 4, 5]. All of these methods try to effectively estimate the background model from the temporal sequence of the frames.

Background model initialization is a problem which has received little attention in literatures [6]. Often the assumption is made that an initial model can be obtained by using a short training sequence in which no foreground objects are present. However, in some situations, e.g., public areas, it is difficult or impossible to control the area being monitored. In such cases it may be necessary to train the model using a sequence which contains foreground objects.

This paper presents a new algorithm which is able to learn a model of the background when moving (foreground) objects are present in the scene. Moreover the proposed method can handle a problem in methods based on Gaussian Mixture Model: distinguishing which kernel is belongs to background and which is that of the foreground.

Even, in a sequence of frames in which the scene is not empty at any moment, each small block of background can be seen in some frames. The key idea of the proposed method lies in the identification of the background based on these blocks with QR-Decomposition, a known method in Linear Algebra. R-values produced with QR-Decomposition can be applied to decompose a given system and indicate the degree of the significance of the decomposed parts. Selection of background blocks is conceptually obtained by choosing those parts which have weak contribution, according to the assigned R-values. The background model is constructed with these selected blocks.

The algorithm is applied to a traffic and some synthesized movies, and its performance is compared with some other background modeling techniques.

The reminder of this paper organized as follows: Section 2 provides a summary of previous work relevant to background modeling. Section 3 describes some definitions such as background initialization problem and QR decomposition. In sections 4 and 5, the proposed method and the simulation results are explained in more details. Finally section 6 describes the conclusion and future works.

2 Previous Work

In the surveillance system of Stauffer and Grimson [2], which has become the standard formulation for the mixture approach in the field, an online EM approximation based on recursive filter was used to train the mixture background model. The rate of adaptation is controlled by a global parameter α that ranges between 0 and 1. In order to preserve a reasonably long learning history and maintain system stability, a very small constant is typically used for video applications. In [1], a single Gaussian model is used per pixel and the parameters are updated by alpha blending. Unfortunately, these approaches fail in case the distribution of the background color values do not fit into a single model. Mixture models were proposed to handle the backgrounds that exhibit multimodal characteristics. Porikli and Tuzel [5] modeled each pixel as a set of layered normal distributions. Lee [3] proposed a Gaussian mixture learning method, which produced good estimate of the background; even the room was never empty at any moment.

Gaussian Mixture Model (GMM) has been used for modeling the background in many literatures [1, 2, 3, 4, 5]. If we monitor the intensity value of a pixel over time in a completely static scene (i.e., with no background motion), then the pixel intensity can be reasonably modeled with a Normal distribution $N(\mu,\sigma^2)$. This Normal distribution model for the intensity value of a pixel is the underlying model for many background subtraction techniques.

At the pixel level, background segmentation involves a binary classification problem based on $P(B|x)$,where x is the pixel value at time t, and B represents the background class. With an explicit representation of the temporal distribution $P(x)$ as a mixture [3],

$$P(x) = \sum_{k=1}^{K} P(G_k)P(x \mid G_k) = \sum_{k=1}^{K} \omega_k . g(x;\mu_k,\sigma_k) \qquad (1)$$

the posterior probability can be expressed in terms of the mixture components $P(G_k)$ and $P(x|G_k)$ and a density estimate $P(B|G_k)$ as follows:

$$P(B \mid x) = \sum_{k=1}^{K} P(B \mid G_k)P(G_k \mid x) = \frac{\sum_{k=1}^{K} P(x \mid G_k)P(G_k)P(B \mid G_k)}{\sum_{k=1}^{K} P(x \mid G_k)P(G_k)} \qquad (2)$$

One of the concentrating aspects of background subtraction in the literatures is the estimation of $P(B|G_k)$, or in other words, distinguishing which kernel belongs to background and which is that of the foreground. Some of them such as [1] suppose that we have a training phase, which the scene is empty. Others try to make decision heuristically. In [2], $P(B|G_k)$ equals to 1 for Gaussians with the highest ω/σ covering a certain percentage of observations, and 0 for all others. Lee [3] trained a sigmoid function on ω/σ to approximate $P(B|G_k)$. In [7] the Gaussians are manually labeled and remain fixed; the darkest component is labeled as shadow, of the remaining of two components, the one with the largest variance labeled as vehicle and the other as road.

In the proposed method dealing with the aforementioned problem is omitted.

3 Definitions

This section introduces the elementary definitions and concepts utilized in later sections.

3.1 The Background Initialization Problem

The background initialization problem is defined as follows. The input is a short video sequence in which any number of moving objects may be present. The goal is to output a single background model describing the scene. Several assumptions are necessary to make the task feasible [6]:

1. Each pixel in the image will reveal the background for at least a short interval of the sequence.
2. The background is approximately stationary; only small background motion may occur.
3. A short processing delay is allowed subsequent to acquiring the training sequence.

The first assumption is necessary to avoid randomly choosing background appearance. If an object occludes a certain area for the entire training period, it is impossible to accurately estimate the background pixel values in that area. Allowing parts of the background to be moving would require discrimination between different types of motion. Motion of the foreground objects would have to be distinguishable from that of the background, and task-specific assumptions would be necessary to distinguish them. The recognition of background motion could be added to this algorithm, which is a topic of future research. The final assumption allows batch processing of the sequence, which gives the algorithm certain advantages over sequential processing. Batch algorithms make use of information from the entire sequence and output a single decision, while background updating schemes must perform sequential decision-making using information from past frames only. For most applications, the penalty of a small processing delay at system startup caused by an initialization procedure is outweighed by the resulting performance improvement [6].

Moreover, we have a stronger restrict on assumption 1, "Each small block in the image will reveal the background for at least a short interval of the sequence, and this short interval must be longer than any foreground interval".

3.2 Review of QR-Decoposition Technique

The singular value decomposition (SVD) of a matrix is a factorization of the matrix into a product of three matrices. For an NxM matrix A, the decomposition can be written as $P=UDV^T$ where $U\in R^{NxN}$ and $V\in R^{MxM}$ are orthogonal matrices, $D=diag(\sigma_1, \sigma_2, \ldots, \sigma_M) \in R^{NxM}$ is a diagonal matrix with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_M \geq 0$. The diagonal elements of D are called the **singular values** of A [8].

The key idea of using SVD in complexity reduction is that the singular values can be applied to decompose a given system and indicate the degree of the significance of the decomposed parts. Reduction is conceptually obtained by the truncation of those parts which have weak or no contribution at all to the output according to the assigned singular values [9]. This advantageous feature of SVD has been used by many people for rule base reduction.[9, 8, 10]. The problem of picking the most influential columns of a given matrix is known as **subset selection** [10].

The column vector X is said to be a **combination** of n other column vectors V_j, if there exist numbers W_j (j=1,2,…,n) not all of them zero, for which: $X = \sum_j W_j V_j$ holds. A set of columns, of which none can be expressed as a combination of the others, is said to be **independent** from each other. The **rank** of a matrix can be revealed by the number of independent columns which can not be expressed as a combination of its columns.

Some of the more obvious cases of singularity are: proportionality of two rows, proportionality of two columns, a zero row, and a zero column.[11]

The **QR decomposition** of P is given by $P\Pi=QR$, where $\Pi\in R^{MxM}$ is a permutation matrix, $Q\in R^{NxM}$ has orthonormal columns, and $R\in R^{MxM}$ is upper triangular. The QR decomposition is uniquely determined by the permutation matrix Π, and many techniques have been proposed to compute it. The values $|R(kk)|$ on the diagonal of R, called the R values, are decreasing and they tend to track the singular values of P [10].

If there is a well defined gap $\sigma_{r+1}(P) << \sigma_r(P)$, then the subset selection will tend to produce a subset containing the most important columns (rules) of P. However, often, the singular values tend to decrease smoothly without any clear gap. In such cases, r is determined by counting the number of (close to) zero singular values in the SVD of P, resulting in a conservative rule reduction method that degenerates to a means for detecting equal rules and rules that do not fire. To help in such situations, it has been claimed, that "the smaller are the singular values, the less important are the associated rules" [10]. As for the singular values, the values of the R-values can help to determine the number of rules to pick [10].

For an example suppose that matrix P has the values shown in Table 1.

Table 1. Matrix P that used in the example in the text; which is constructed such that its four first columns are independent. Columns 5,6,7 are the same as the first column, but with a random small noises, and the last column is the same as the second with small noises

1	2	3	4	5	6	7	8
17.00	24.00	1.00	8.00	17.02	17.04	17.03	24.05
23.00	5.00	7.00	14.00	22.97	23.01	23.01	4.96
4.00	6.00	13.00	20.00	3.98	3.98	4.02	6.04
10.00	12.00	19.00	21.00	10.01	10.00	9.96	12.03
11.00	18.00	25.00	2.00	11.00	11.01	10.99	17.99
35.00	1.00	6.00	26.00	34.96	34.97	34.99	1.01
3.00	32.00	7.00	21.00	2.96	3.03	2.97	32.05
31.00	9.00	2.00	22.00	30.98	31.05	31.03	9.01

8.00	28.00	33.00	17.00	7.99	8.01	8.03	27.97
30.00	5.00	34.00	12.00	30.00	29.95	30.00	4.99

Matrix P is constructed such that its four first columns are independent. Columns 5,6,7 are the same as the first column, but with a random small noises, and the last column is the same as the second with small noises. Although its rank is equal to 8 but as Table 2 shows, four of its singular and R-values are close to zero.

Table 2. Singular values and R-Values of Matrix P shown in Table 1.

Singular Values	Absolute of R-Values
150.67	64.93
69.75	48.77
37.23	40.49
24.95	26.94
0.07	0.09
0.05	0.06
0.03	0.04
0.02	0.04

The logarithm of both the singular values and absolute R-values of P are plotted in Figure 1, and it is seen that the R-values track the singular values well. Further, we notice that P has four small (close to zero) singular values, but also a distinct gap after four values (as could be expected with our knowledge about the structure of P).

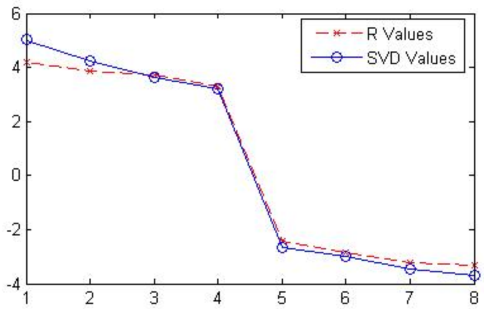

Figure 1. The singular values (o) and R values (x) of matrix P shown in Table 1.

In the proposed method we have used this advantageous features of QR-decomposition for distinguishing background pixels from foreground ones.

4 The Proposed Background Modeling Initializing Method

In our proposed method we split the input image frame into M small blocks and apply QR-decomposition method on each block to identify the background part.

Even in a sequence of frames, which the scene is not empty at any moment such as Figure 2, each small block of background can be seen in some frames, depending on the traffic (assumption no. 1 in previous section).

The key of the proposed method lies in the identification of the background based on these blocks.

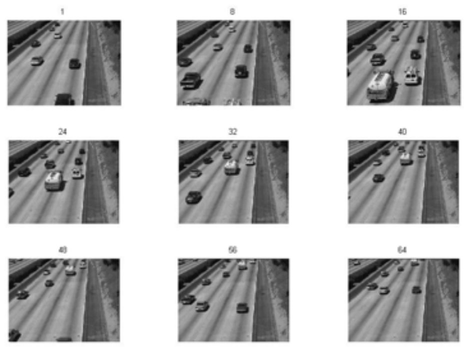

Figure 2. Sample frames of a traffic movie[1].

We consider the values of a particular pixel over time as a "pixel process", a time series of pixel values as follows:

$$\{X_{i,1}^b, X_{i,2}^b, \cdots, X_{i,t}^b\}$$

This is a series of intensity values X of a particular pixel $\{x_i, y_i\}$ at b^{th} block (b=1,…,M), at time t. We construct matrix A^b for b^{th} block as:

$$A^b = \begin{bmatrix} X_{1,1}^b & X_{1,2}^b & \cdots & X_{1,t}^b \\ X_{2,1}^b & X_{2,2}^b & \cdots & X_{2,t}^b \\ \vdots & \vdots & \vdots & \vdots \\ X_{N,1}^b & X_{N,2}^b & \cdots & X_{N,t}^b \end{bmatrix} \qquad (3)$$

Because the proposed method is the same for all blocks, in the following, for simplicity we name A^b as A. Each R-value of A is related to one of the columns of A. Since the columns containing only background data are almost similar, the R-values corresponding to these columns will be smaller than those containing moving objects.

If {1,2,…,t} is the frame numbers in the original order, we define $\{f_1, f_2, \ldots, f_t\}$ as the frame number indices in the ordered list. Suppose that $\{X_{i,f_1}^b, X_{i,f_2}^b, \cdots, X_{i,f_t}^b\}$ is the sorted list of i^{th} pixel's intensity values at b^{th} block, according to the mentioned R-values, we consider:

---

[1] *This movie can be downloaded from:* http://cvrr.ucsd.edu/aton/press/movies/voigtclip_short.avi

$$P(B \mid X^b_{i,f_j}) = \begin{cases} 1, & \text{if } j>\beta * t \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where β shows the percentage of the blocks containing foreground objects in the image sequence and/or (1-β) shows the percentage of the blocks representing background in the image sequence. Since the block sizes are small, each block shows nothing but background data in many image frames while there are some moving objects at each moment in the image sequence. Based on our experiments, 2/3 is a proper value for β.

For modeling the background we used Gaussian Mixture Model. As mentioned earlier, one of the concentrating aspects of background subtraction with GMM is distinguishing which kernel of GMM belongs to background and which is that of the foreground. In the proposed method, since the background pixels are recognized before using GMM, we do not deal with the mentioned problem.

If we use K=1 in GMM, we can estimate the average and standard deviation of the background model at ith pixel of bth block by $Mean(\{X^b_{i,f_{s+1}}, X^b_{i,f_{s+2}}, \cdots, X^b_{i,f_t}\})$ and $STD(\{X^b_{f_{i,s+1}}, X^b_{f_{i,s+2}}, \cdots, X^b_{i,f_t}\})$ of blocks belonging to background based on equation (4). For K>1, one can use EM algorithm to estimate the kernels' parameters.

5 Experimental Results

We applied our method on a video data from a traffic scene (see Figure 2). Figure 3 shows 64 consequence frames of an instance block in original time order of the video. The blocks rearrangement result based on their R-values' order is illustrated in Figure 4. As we expected blocks that contain only background data, have been shifted to the end of the list.

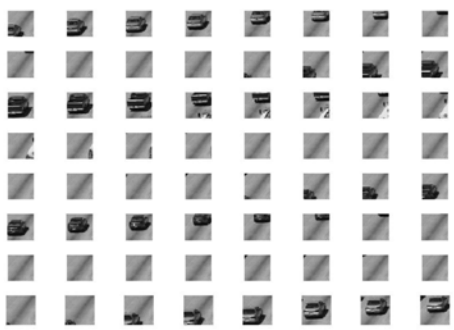

Figure 3. 64 Consequence frames of an instance block in its original order.

As mentioned earlier, the R-values track the singular values well. Figure 5 shows both of them for the illustrated blocks in Figure 3.

For speedup the implementation we did not use all of the pixels in each block; Rader than we used only the major and minor diagonal elements and the middle row and column of each block.

Experimental results on sample movies showed that at least one third (β=2/3) of the end of rearranged blocks can be considered as background (in movies such that illustrated in Figure 2). We build the background model from these blocks. For example in a whole of 121 frames of a sample movie (Figure 2) we consider the last 40 blocks based on QR's order, as background.

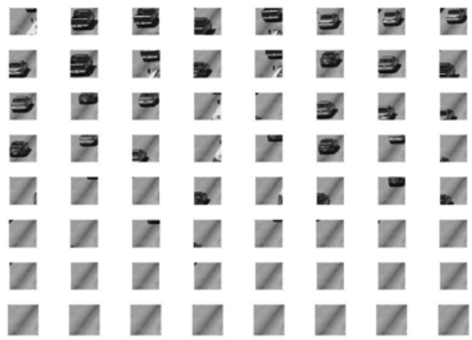

Figure 4. Rearrangement of blocks in Figure 3 based on their QR's order; as can be seen background frames has been shifted to the end.

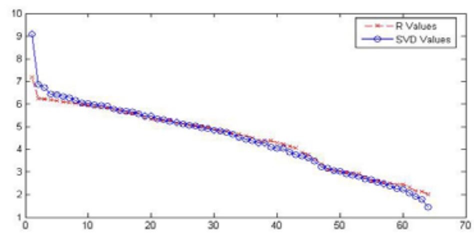

Figure 5. R-values and singular values for the illustrated blocks in Figure 3. as can be seen R-values track the singular values well.

Figure 6 shows the result of the proposed method executed on total 121 frames of movie Figure 2, when we used a single gaussian kernel for modeling the background.

Figure 6: (Detected background) Result of the proposed method for the shown movie in Figure 2.

Every background model should be examined with a foreground detection problem. Figure 7 shows the detected foreground corresponding for the frames illustrated in Figure 2. Pixel values that don't fit the background distribution are considered as foreground. In the mentioned figure every pixel that its intensity value is more than 2.5*σ far from its mean in background model, is considered as foreground. The shown results are raw and can be improved by image processing techniques such as morphological operations.

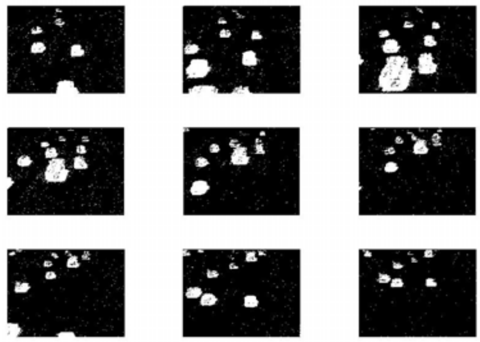

Figure 7. Detected foreground (moving objects) with the proposed modeled background.

Figure 8 shows the comparison results between the GMM (Stauffer) method, temporal averaging and the proposed method, without any post processing. This figure confirms the high performance of the proposed method on background subtraction. For an image size of 320 x 240 we obtained an average processing speed of 16 fps. We used the first 100 frames of the video for the batch processing. The batch processing took 6 seconds on our 1.83 GHz Intel Centrino Duo processor based PC, where the block size was 40x40.

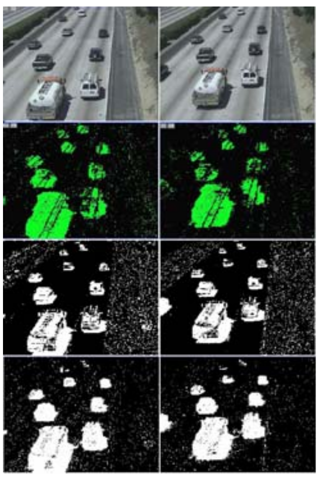

Figure 8. From top to bottom: original frames, foreground object detection using GMM (taken from [12]), temporal averaging and the proposed approach.

As can be seen on Figure 8, the small cars on the top of the scene did not detected properly with the proposed method. We can deal with such situations by changing the block size to a smaller one but it will increase the process time slightly. Simulation results on the same video data with block size equal to 10, showed the better detection with average speed of 9 fps. Figure 9 shows the experimental results with different block sizes. In this figure dilation and erosion has been used as post processing enhancements.

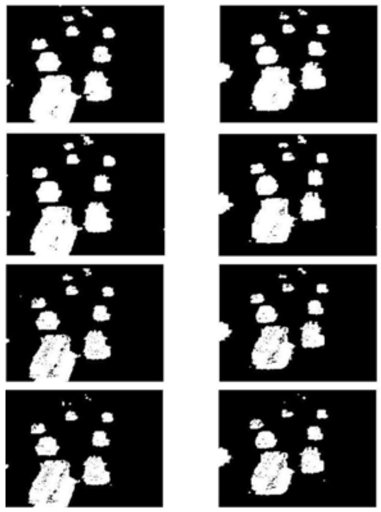

Figure 9. The result of the proposed method with different block sizes: 10x10, 16x16, 20x20, and 40x40. The smaller block size gives finer result but taking more processing time.

We also compared the proposed method and temporal averaging with another comparison criterion. Our criterion was Mean Square Error, MSE, between background model and the ground truth background. Because we need the ground-truth background, 30 movies with known background were synthesized. Figure 10 shows the comparison results based on the logarithm of MSE. This figure also confirms the higher accuracy of the proposed method in compare with temporal averaging algorithm.

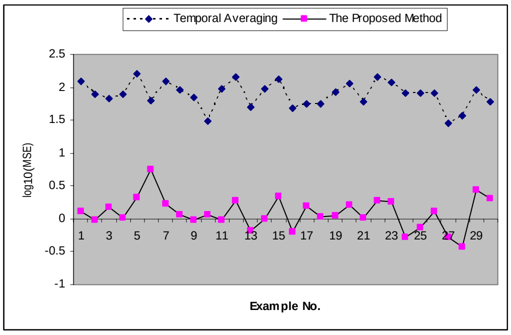

Figure 10. Comparison between temporal averaging method and the proposed method.

Moreover, the simulation results showed that the QR-decomposition can be computed faster than the singular values, and hence the QR-Decomposition is superior to the SVD from an implementation point of view.

5 Concluding Remarks and Future Works

In this paper, a new method for background subtraction based on QR-decomposition technique has been proposed. We can summarize the advantages of the proposed method in compare with other background subtraction techniques as follows:

- Initialization with moving objects: there is no need for an empty scene with no foreground object for the algorithm initialization. The system can model the background by analyzing few video frames even if there are foreground objects in every frame.
- One of the drawbacks of background subtraction techniques based on Gaussian Mixture Model, GMM, is that they assume Gaussian models for both background and foreground objects while it is not logical to consider a Gaussian model for moving objects in the scene. For instance, in a video showing a traffic scene, different cars pass the road and we cannot assume a Gaussian model that includes all cars. In the proposed method, this problem is solved by assuming Gaussian model only for background.
- Since the background pixels are recognized before using GMM, the problem of distinguishing which kernel of GMM belongs to background and which is that of the foreground is eliminated in the proposed method.
- Advantageous features of QR-decomposition susceptible our method to act on light changing environments.

The experimental results on foreground detection showed better performance of the proposed method with respect to temporal averaging and GMM.

As future works, we plan to apply the proposed method on RGB data, exploring possible modifications to the algorithm for handling background adaptation, while the current approach uses batch processing and to find a proper threshold for cutting the R-values.

References

[1] Wren, Christopher R., Ali Azarbayejani, Trevor Darrell, and Alex Pentland. "Pfinder: Real-Time Tracking of the Human Body", In *IEEE Trasactions on Pattern Analysis and Machine Intelligence*, vol 19, no 7, pp.780-785, 1997.

[2] C. Stauffer and E. Grimson. Adaptive background mixture models for real-time tracking. *In CVPR*, pages 246–252, 1999.

[3] D.S.Lee, "Effective Gaussian Mixture Learning for Video Background Subtraction", *In IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol 27, no 5, pp. 827-832, 2005.

[4] A. Elgammal, D. Harwood, and L. Davis. "Non parametric model for background subtraction", In *Computer Vision ECCV 2000*, pp. 751–767, 2000.

[5] Fatih Porikli, Oncel Tuzel, "Bayesian Background Modeling for Foreground Detection", in Proceedings of *the third ACM international workshop on Video surveillance* & sensor networks, pp. 55-58, 2005.

[6] D. Gutchess, M. Trajkovic, E. Cohen-Solal, D. Lyons, A. K. Jain, "A Background Model Initialization Algorithm for Video Surveillance", *IEEE Int. Conf. on Computer Vision*, 2001.

[7] N. Friedman and S. Russell, "Image Segmentation in Video Sequences: A Probabilistic Approach," *Proc. 13th Conf. Uncertainty in Artificial Intelligence*, 1997.

[8] J. Yen and L.Wang, "Simplifying fuzzy rule-based models using orthogonal transformation methods," *IEEE Trans. Syst., Man, Cybern*. B, vol. 29, pp. 13–24, Feb. 1999.

[9] Kaynak, O., et al., "Complexity reduction of rule based models: a survey", in *Proceedings of the 2002 IEEE International Conference on Fuzzy Systems*, pp. 1216-1221, 2002.

[10] Magne Setnes, Robert Babuska, "Rule Base Reduction: Some Comments on the Use of Orthogonal Transforms", *IEEE Trans. Syst., Man, Cybern,* Part C, vol. 31, no. 2, pp. 199-206, May 2001.

[11] A R G Heesterman , *MATRICES and their ROOTS - A Textbook of Matrix Algebra*, World Scientific Publishing, 1990.

[12] Q. Zang, R. Klette, "Robust Background Subtraction and Maintenance", *in Proceedings of the 17th International Conference on*

Pattern Recognition (ICPR'04), Vol. 2, pp.90-93, Aug 2004.