

QR DECOMPOSITION-BASED ALGORITHM FOR BACKGROUND SUBTRACTION

Mahmood Amintoosi^{1,2}, Farzam Farbiz^{1,3}, Mahmood Fathy¹, Morteza Analoui¹, Naser Mozayani¹

¹ Computer Engineering Department, Iran University of Science and Technology, Tehran, Iran

² Mathematics Department, Sabzevar Teacher Training University, Iran, 397

³ A*STAR Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613

ABSTRACT

This paper presents a new algorithm for background subtraction that can model the background image from a sequence of images, even if there are foreground objects in each image frame. In contrast with Gaussian Mixture Model algorithm, in our proposed method the problem of distinguishing between background and foreground kernels becomes trivial. The key idea of our method lies in the identification of the background based on QR-Decomposition method in linear algebra. R-values taken from 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. Simulation results show the better background detection performance with respect to some others.

Index Terms—Image processing, Matrix decomposition, Linear algebra, Image segmentation, Object detection

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 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, called the “background image” or “background model”. The background image is a representation of the scene with no moving objects.

Several methods for performing background subtraction have been proposed in literatures. All of these methods try to effectively estimate the background model from the temporal sequence of the frames. If we monitor the intensity value of a pixel over time in a completely static scene, then the pixel intensity can be reasonably modeled by 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.

To model the background where the scene is not completely static (such as moving trees branches and leaves), a mixture model techniques are used [1, 2, 3, 4, 5]. For instance, 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. However, when the distribution of background color values does not fit into a single model, these approaches fail. 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.

In the Gaussian Mixture Model, GMM, method for background subtraction, 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,

$$P(x) = \sum_{k=1}^K P(G_k)P(x|G_k) = \sum_{k=1}^K \omega_k \cdot g(x; \mu_k, \sigma_k) \quad (1)$$

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 [3]:

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

The estimation of $P(B|G_k)$, (or in other words distinguishing which kernel of GMM belongs to the background and which is that of the foreground) is one of the concentrating aspects of background subtraction in the literatures [2, 3, 6]. 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 [6] the Gaussians are manually labeled and remain fixed: the darkest component is labeled as shadow, the one with the largest variance is labeled as vehicle and the remaining one is labeled as road.

In this paper, we present a novel algorithm for background subtraction based on QR-decomposition, a known method in linear algebra. Using this technique, the background model can be easily distinguished from foreground model. Moreover, the proposed method can handle “Initialization with moving objects”. Many algorithms require a scene with no moving objects during initialization while the proposed method can model the background image even all image sequences include some foreground objects. Simulation results showed that the proposed

method produced better estimation of background with respect to some others.

2. THE PROPOSED METHOD

In this section, we describe our proposed background subtraction algorithm. We start by reviewing QR-decomposition technique and then we explain how this technique can be applied for background modeling.

2.1. Review of QR-decomposition technique

The singular value decomposition (SVD) of a matrix is a factorization of the matrix into a product of three matrices. For an $N \times M$ matrix A , the decomposition can be written as $P = UDV^T$ where $U \in \mathbb{R}^{N \times N}$ and $V \in \mathbb{R}^{M \times M}$ are orthogonal matrices, and $D = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_M) \in \mathbb{R}^{N \times M}$ is a diagonal matrix with $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_M \geq 0$. The diagonal elements of D are called the singular values of A [7].

The singular value decomposition has been used by many researchers for rule base reduction [8, 7, 9]. The key idea of using SVD in complexity reduction is that SVD decomposes a given system into separate parts and indicates the degree of the significance of each decomposed parts. We can reduce the size of the matrix by picking the most influential columns of the matrix, which is known as subset selection problem [9]. To do so, we can easily truncate the vectors with least degree of significance according to SVD [8].

The QR decomposition of matrix P is given by $P = QP\Pi = QR$, where $\Pi \in \mathbb{R}^{M \times M}$ is a permutation matrix, $Q \in \mathbb{R}^{N \times M}$ has orthonormal columns, and $R \in \mathbb{R}^{M \times M}$ is upper triangular. The QR decomposition is uniquely determined by the permutation matrix Π . The values $|R(kk)|$ on the diagonal of R , called the *R-values*, are in decreasing order and they tend to track the singular values of P [9].

If there is a well defined gap in singular values of P , $\sigma_{r+1}(P) \ll \sigma_r(P)$, then the subset selection will tend to produce a subset containing the most important columns (rules) of P . However, the singular values often decrease smoothly without any clear gap. In such cases, the truncation index r is determined by counting the number of (close to) zero singular values in the SVD of P since it has been claimed that ‘‘the smaller are the singular values, the less important the associated rules will be’’. As for the singular values, the R-values also help to determine the number of rules to pick [9].

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

2.2. The proposed background subtraction 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. We consider the values of a particular pixel over time as a ‘‘pixel process’’, a time series of pixel values as follows:

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

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

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

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

$$P(b | X_{i,f_j}^b) = \begin{cases} 1, & \text{if } j > (1-\beta) * t \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where $1-\beta$ shows the percentage of the blocks containing foreground objects in the image sequence and/or β shows the percentage of the blocks representing background in the image sequence. Since the block sizes are small, even when there are moving object in each frames, some blocks will contain only background data. As a result, in an image sequence, each block shows nothing but background data in many image frames. Hence, we found that $1/3$ would be a proper value for β .

If we use $K=1$ in GMM, we can estimate the average and standard deviation of the background model at i^{th} pixel of b^{th} block by $\text{Mean}\{X_{i,f_{t+1}}^b, X_{i,f_{t+2}}^b, \dots, X_{i,f_t}^b\}$ and $\text{STD}\{X_{i,f_{t+1}}^b, X_{i,f_{t+2}}^b, \dots, X_{i,f_t}^b\}$ of blocks belonging to background based on equation (4). For $K>1$, one can use a version of EM algorithm to estimate the kernels' parameters.

3. EXPERIMENTAL RESULTS

We applied our method on a video data from a traffic scene (see Figure 1). Figure 2 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 3. As we expected blocks that contain only background data, have been shifted to the end of the list.

Figure 4 shows both R-values and the singular values for the illustrated blocks in Figure 2. As can be seen, the R-values track the singular values well.



Figure 1. Sample frames of a traffic movie¹.

¹ This movie can be downloaded from:

http://cvrr.ucsd.edu/aton/press/movies/voigtclip_short.avi

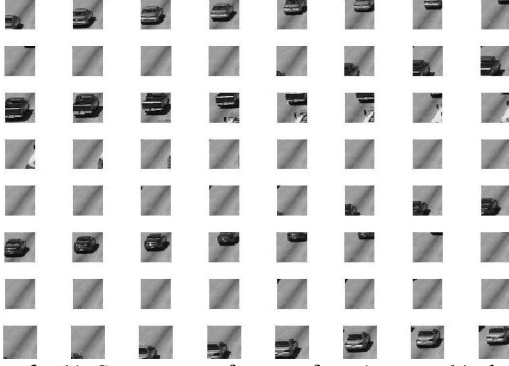


Figure 2. 64 Consequence frames of an instance block in its original order.

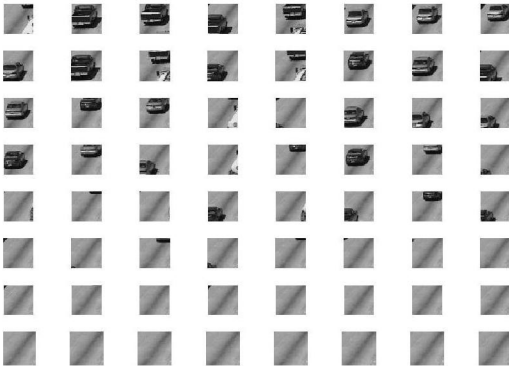


Figure 3. Rearrangement of blocks in Figure 2 based on their R-values' order. A can be seen, background frames has been shifted to the end.

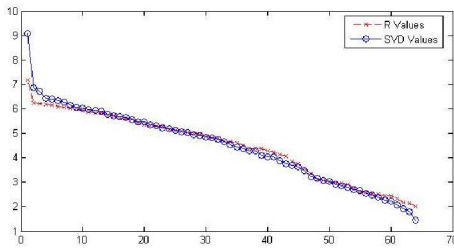


Figure 4. R-values and singular values for the illustrated blocks in Figure 2.

Every background model should be examined with a foreground detection problem. We consider each pixel with intensity value more than 2.5σ far from its mean in background model, as a foreground pixel. Figure 5 shows the result of foreground detection for the test video data. The middle row shows results of GMM (taken from Figure 2 of [11]), and the bottom row illustrates the result of the proposed method. This figure confirms the high performance of the proposed on background subtraction. For an image size of 320×240 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. The block size for this experiment was chosen as 40×40 . As can be seen on Figure 5,

the small cars on the top of the scene did not detected properly. We can deal with such situations by changing the block size to a smaller one. However, it will slightly increase the process time. Simulation results on the same video data with block size equal to 10, showed the better detection with average speed of 9 fps (see Figure 6).

We made 30 artificial movies to have ground-truth data in order to evaluate the result of the proposed method in compare with the temporal averaging algorithm. Our criterion was Mean Square Error, MSE, between background model resulted by the algorithm and the ground truth. Figure 7 shows the comparison results based on the logarithm of MSE. As can be seen in this figure, the proposed method has higher accuracy in background modeling in compare with temporal averaging algorithm.

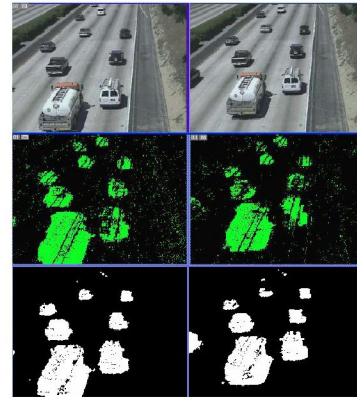


Figure 5. From top to bottom: original frames, foreground object detection using GMM (taken from [11]) and the proposed approach.

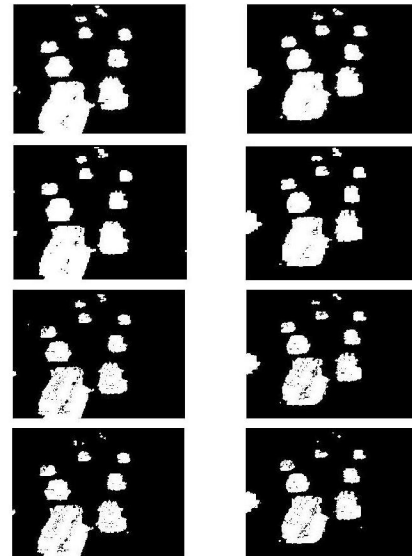


Figure 6. The result of the proposed method with different block sizes: 10×10 , 16×16 , 20×20 , and 40×40 . The smaller block size gives finer result but taking more processing time.

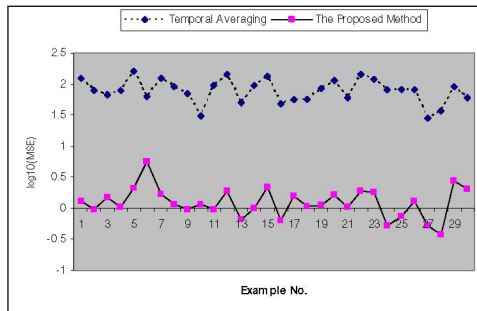


Figure 7. Comparison between Temporal Averaging method and the proposed method.

4. 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 frames.
- Distinguishing background data from foreground ones by using R-values obtained from QR-decomposition.
- 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 a moving object 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.

The experimental results on foreground detection showed better performance of the proposed method with respect to temporal averaging and GMM. The proposed method can be also used as an initialization step in a hybrid method with other algorithms.

As future works, we plan to apply the proposed method to test on different lighting conditions and shadows and to find an adaptive threshold for the R-values.

5. REFERENCES

- [1] C. R. Wren, A. Azarbayejani, T. Darrell and A.P. Pentland. "Pfindex: Real-Time Tracking of the Human Body", In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol 19, no 7, pp.780-785, July 1997.
- [2] C. Stauffer and E. Grimson. "Adaptive background mixture models for real-time tracking". In *CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, p. 2246, 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, May 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] F. Porikli, O. 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] N. Friedman and S. Russell, "Image Segmentation in Video Sequences: A Probabilistic Approach," in *Proc. 13th Conf. Uncertainty in Artificial Intelligence*, Aug 1997.

- [7] 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.

- [8] O. Kaynak, 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.

- [9] M. Setnes, R. 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.

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

- [11] 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.