Clustering Based Non-parametric Model for Shadow Detection in Video Sequences

Ehsan Adeli Mosabbeb[1], Houman Abbasian[2], Mahmood Fathy[1]
[1]Iran University of Science and Technology, Tehran, Iran
[2]University of Ottawa, Ottawa, Ontario, Canada
eadeli@iust.ac.ir, habba057@uotawa.ca, mahfathy@iust.ac.ir

Abstract

In the area of computer vision, tracking moving objects in video sequences is a challenge. Most techniques in the literature use background subtraction to extract moving objects and their shadows, as the first step, after which they use shadow detection algorithms to remove shadows form foreground objects. In this paper, a novel method for shadow detection in traffic video images is presented, a simple background modeling technique and Deterministic Non-Model based shadow detection algorithm (DNM1) are used together with the spatial information along with color information. This algorithm uses K-means clustering to improve DNM1 classification and to separate shadow region form object region. Initial cluster centers are determined using color information from DNM1. The proposed algorithm is tested on highway video sequences and showed considerable improvement in comparisons with DNM1.

Keywords: Clustering, Computer vision, Detection, Shadow detection.

1. Introduction

Designing a vision based system for traffic analysis is an essential and challenging problem in computer vision systems regarding Intelligent Transportation Systems. An important component of such a system is the extraction process of foreground objects in video sequences. Background subtraction is common approach to deal with this problem. But foreground contains both moving objects and their shadows. Therefore, a robust segmentation algorithm has become a challenging problem, where moving objects can appropriately be separated form their shadows. Many works reported in the literature, working on different areas of applications, have taken this into account and done lots of researches.

In order to extract foreground objects, background model needs to be extracted form the video sequence, and then background subtraction is used to determine the differences between background and any frame in the sequence. Although, there are many techniques for background modeling and subtraction, common method to background subtraction is to compare each new frame with the background. Significant differences between a frame and background scene are usually considered as foreground objects. The aim of background subtraction is to extract the moving objects only, but in practice shadows are also detected as moving objects. Thus, it is necessary to remove shadow from foreground objects.

Detection of shadow pixels, as the foreground object, can cause undesirable problems in tracking. For instance due to the locating of cast shadow of one object over another object, shadows can cause objects to be merged in one group. They also may cause object shape distortion or loss. There are several techniques for shadow detection in literatures. Although most of these techniques are based on color video sequences, there are several approaches which can be applied to grayscale video sequences. Some of the approaches are based on RGB color space. Some are based on HSV color space and some others are based on the location of light source. In fact, most of the shadow detection methods are based on color information. In this work we modify an existing background modeling technique described in [1], to use in color video sequence. Also we propose an improvement in deterministic non-model based shadow detection algorithm, DNM1, described in [2], using spatial information along with the color information.

The remainder of this paper is organized as follows. Section 2 presents related works regarding different background modeling and shadow detection algorithms. Our proposed method is described in section 3, some experimental results on a traffic video sequence, and some direction for future work are provided on section 4 and finally conclusion is given in section 5.

2. Related Works

Several methods for background modeling and shadow detection are proposed. Background modeling techniques can be used on both grayscale and color video sequences but shadow detection techniques take into use the color information to detect shadow pixels and therefore they are not applicable to grayscale images. Here, some of these techniques are described.

C. Stauffer and W. Grimson used an adaptive background modeling for real time tracking [3]. In their work, they represented each pixel with a mixture of Gaussians. They also used motion information along with the color information to model dynamics of background. Recently automatic background estimation has been proposed [3, 4, 5]. A. Elgammal, et al. in [4] used non-parametric prediction algorithm instead of Gaussian mixture to estimate probability density function of each pixel. This technique succeeded to better model the behavior of each pixel, while needed several thresholds. In [1], Haritaoglu et al. proposed an approach to model background in grayscale video sequences. They used three values for each pixel: Minimum intensity, Maximum intensity and maximum intensity difference between consecutive frames observed during training period. Then pixel wise median filter over time was applied to each pixel to distinguish between stationary and moving pixels. Only stationary pixels are considered as initial background model. In this work, we modify this work for color video sequences.

Shadow detection algorithms are also widely explored in many researches. In general, shadow detection algorithms are divided into two major groups. *Deterministic* and *Statistical* approaches. *Deterministic approaches* use an on/off decision process [6, 7, 8, 9], whereas *Statistical approaches* use probability functions to describe the class membership [5, 10, 11, 12, 13]. Each of these classes of methods solely has two subclasses. For the case of deterministic approaches, we can categorize them to model based and non-model based subclasses, concerning whether the decision comes from model based knowledge or not. Statistical approaches, based on selection of parameters, are divided into parametric and non-parametric subclasses. Examples of deterministic model-based approaches are reported in [9, 14, 15], while methods presented in [6, 7] are examples of deterministic non-model-based ones. [16] presents a statistical parametric model, whereas [12, 13] are examples of statistical non-parametric methods. A few authors have studied shadow detection in monochromic video sequences [7, 17, 18, 19].

A revision of works done indicates that there are many background modeling techniques applicable on both gray scale and color video sequences. Here we choose background modeling technique described in [1] because it is effective and requires less computational cost than Gaussians and other complex methods and modify it in order to use in color video sequences. In addition, a deterministic non-model based shadow detection algorithm (DNM1) [4], and an improvement on it, based on spatial information using K-means algorithm, are implemented for classification of shadow pixels and moving object pixels.

3. The proposed algorithm

In this section, first we describe the background model W^4 [1], modifying it for colored video sequences. After that, we propose a novel shadow detection algorithm based on DNM1 and spatial information in each video frame. This is done using K-means algorithm to classify shadow and moving object pixels. A flowchart of our algorithm is illustrated in Figure 1.

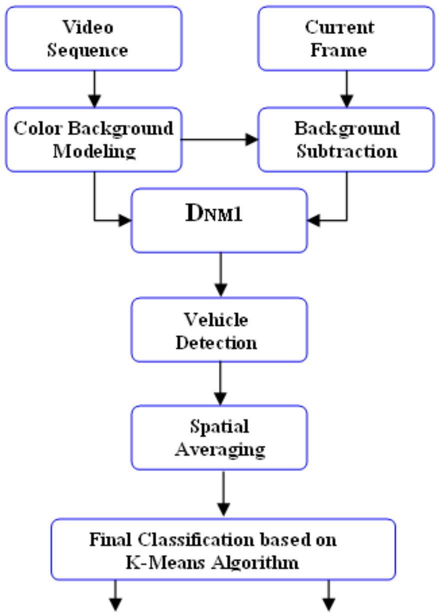

Figure 1: Our proposed shadow detection algorithm.

As it can be seen, first background modeling is applied on the video sequence, extracting colored background. In the next step, a background subtraction for taking out moving objects along with their shadows is used. Then in order to distinguish between shadow pixels and moving objects, DNM1 is taken in use. Two final stages are to improve DNM1. Here, we use spatial averaging around shadow pixels and object pixels to find two initial centers for the process of K-Means clustering. Then K-Means clustering is applied using these two initial centers. Thus, the modification to DNM1 is done through using spatial information along with color information which leads to an improvement to this aforementioned shadow detection algorithm.

3.1. Background modeling

W^4 is a real-time visual surveillance system for detecting and tracking of moving objects and monitoring their activities in an outdoor environment. It operates on monocular grayscale video images. In this work, it is modified for colored video sequences. We apply W^4, on each channel of RGB color space of the video sequence. Having three backgrounds for each channel, the results can be integrated forming a single RGB background done by concatenating the three matrixes acquired.

W4, to model the background, uses a two-stage method. The first stage consists of applying a pixel wise median filter over time to several portions of video (typically 20-30 seconds) to distinguish moving pixels from stationary pixels. In this experiment we use 100 frames, which is about 3.3 seconds. And then in the second stage only those stationary pixels are chosen as background pixels. Let *A* be a colored video sequence containing *N* consecutive frames, $A^k(i, j)$ be the intensity of pixel (i, j) in k-th frames of video sequence *A*, $\sigma(i, j)$ *and* $\mu(i, j)$ be the standard deviation and median values at pixel (i, j) in all frames in video sequence *A* respectively.

According to W^4 [1], initial background for pixel (i, j) is formed by three dimensional vector *B* including $m(i, j)$ minimum, $n(i, j)$ maximum intensity values and $d(i, j)$ maximum intensity difference between frames observed during the training period. Background model $B(i, j) = [m(i, j), n(i, j), d(i, j)]$ can be calculated as follows [1, 20]:

$$B(i,j)=\begin{bmatrix} m(i,j) \\ n(i,j) \\ d(i,j) \end{bmatrix} = \begin{bmatrix} \min_z A^z(i,j) \\ \max_z A^z(i,j) \\ \max_z \left| A^z(i,j) - A^{z-1}(i,j) \right| \end{bmatrix} \qquad (1)$$

Where *z*, are frames satisfying:

$$\left| A^z(i,j) - \mu(i,j) \right| \le 2\sigma(i,j) \qquad (2)$$

According to [1, 20], condition (2) assures that only stationary pixels are selected as background pixels. This is because when a moving object moves across a pixel, the intensity of that pixel decreases or increases sharply. Then if we choose the median value of that pixel over time, we can model image without any moving object, this is the background image we are seeking for. After training phase initial background model for each pixel is obtained (i.e. $B(i, j)$). Then each pixel of the input frame, $I_t(i, j)$, is compared with $B(i, j)$, and a pixel is classified as background pixel if:

$$I_t(i,j) - m(i,j) \le \mu.k \quad or \quad I_t(i,j) - n(i,j) \le \mu.k \qquad (3)$$

where μ is the median value of the largest inter-frame absolute difference image $d(i, j)$, and *k* is a constant which chosen empirically (authors suggest the value k=2 [20]). Here we apply this procedure on each channel of RGB color space to generate background in each channel and then integrate the results to generate a single color background image. To determine foreground pixels, we will only need to detect foreground pixels for one channel, as far as foreground pixels are the same in all channels. Here we use channel B.

Figure 2 illustrates an example of this background modeling. Background image of each RGB channel, are shown in figures 2(a), 2(b) and 2(c). Final background image, which are the result of integrating of these three channels, is shown in figure 2(d). In addition, a frame of video sequence and corresponding foreground pixels are shown in figure 2(e) and 2(f) respectively. As it can be seen, final background image is a colored background image. After background subtraction, background image and a frame of video sequence act as inputs to the deterministic non-model based shadow detection algorithm DNM1, discussed in the following section.

3.2. Deterministic non-model based shadow detection algorithm DNM1

System described in [21] is an example of deterministic non-model based shadow detection algorithm (we called it DNM1). Because HSV color space is very close to human perception of color, it can distinguish between object and shadow pixels more accurately [22]. Therefore this algorithm is designed based on HSV color space. According to experiment results in [2], saturation of shadow points is lower than background point. In addition, shadow cast on background doesn't change significantly in hue. The resulting decision process is reported in the following equation:

$$SP_K(x,y) = \begin{cases} 1 & if \quad \alpha \le \dfrac{I_K^V(x,y)}{B_K^V(x,y)} \le \beta \\ & \wedge \quad (I_K^S(x,y) - B_K^S(x,y)) \le \tau_S \\ & \wedge \quad \left| I_K^H(x,y) - B_K^H(x,y) \right| \le \tau_H \\ 0 & Otherwise \end{cases} \qquad (4)$$

Where $I_K(x, y)$ and $B_K(x, y)$ are the pixel values in coordinates (x, y) in the input image (frame K) and in the background model (computed at frame K), respectively. The use of β prevents the identification as shadows of those points where the background was slightly changed by noise,

Whereas, α takes into account the strength of the shadow, i.e. how strong the light source is. Thus the stronger and the higher the sun, the lower α should be chosen. The choice of parameter of τ_S and τ_H is less straightforward and, for now should be chosen empirically. Here in our work the values for $\alpha, \beta, \tau_S, \tau_H$ are chosen 0.1, 0.58, 0.01, 0.07 respectively.

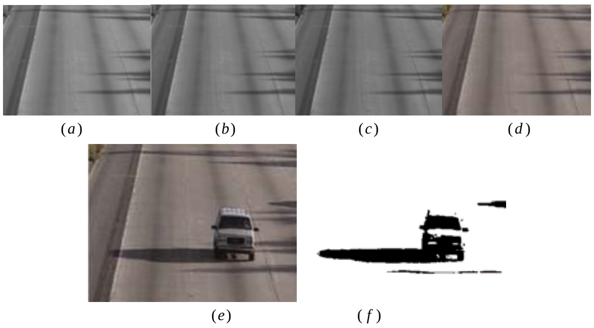

Figure 2: (a) Background image of channel R. (b) Background image of channel G. (c) Background image of channel B. (d) Final background image which is the result of integrating backgrounds of channels R,G and B. (e) Frame 101 of Highway video sequence. (f) Result of background subtraction on channel B.

3.3. Vehicle detection

To use K-means clustering algorithm, we need to find the place of the cars and shadows. After using DNM1 algorithm, shadow and foreground candidate pixels are determined. Here we first use background subtraction technique to produce binary image which includes foreground and shadow pixels. Then we remove shadow pixels, which was detected by DNM1 algorithm, form binary image. So now we have a binary image which includes only foreground pixels. After this step, by applying morphological operators, we fill existing probable gaps in the binary image. Then we find the place of each connected component in the binary image. As a result, the places of vehicles are determined. To find the location of shadows, we repeat this procedure for shadow pixels determined by DNM1. Overall procedure results in finding the place of shadow and foreground connected components. Finally by spatial averaging of each of the connected components, we can find cluster centroids which are the inputs fed into the K-means clustering algorithm.

3.4. Spatial averaging

Here we use spatial information to improve shadow detection algorithm DNM1. Using DNM1 approach, we have two groups of errors in detecting shadow and foreground pixels. First group of errors are false negatives (FN), the shadow pixels classified as foreground pixels, and the second group of errors are false positives (FP), the foreground pixels detected as shadow points. But most pixels are detected truly. So the geometrical centroids of shadow pixels and moving object pixels are separated from each other. In the other words, shadow pixels are closer to the centroid of shadow points than centroid of foreground pixel.

Therefore after finding each connected component, we use spatial averaging to determine the geometrical centroid of shadow and foreground points. Figure 3 illustrates the spatial averaging. A frame of a video sequence before and after applying DNM1 approach are shown in figure 3(a) and 3(b) respectively and the result of spatial averaging, after finding shadow and foreground connected components, is represented in figure 3(c). As it can be seen from figure 3(c) result of spatial averaging is two centroids for shadow and foreground pixels. In the following section K-Means Clustering algorithm will be explained.

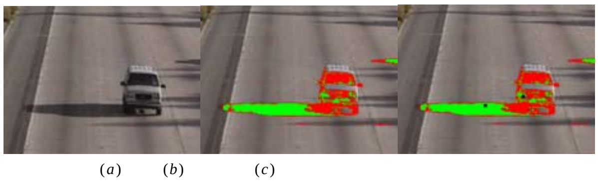

Figure 3: (a) Frame 101 of video sequence (b) Result of DNM1 approach(c) The result of spatial averaging. Two black points are the centers of shadow and foreground pixels.

3.5. K-Means clustering

Clustering algorithms, partition data into certain number of groups based on some similarity/dissimilarity measures. K-Means algorithm is one of the most popular clustering algorithms. One of the problems of K-Means algorithm is that the number of clusters should be determined by user. Here for our purpose, only we have two clusters. Shadow cluster and foreground cluster. In fact the number of clusters is obvious. In what follows, we describe the steps of K-Means algorithm in brief:

Step 1: Choose K initial cluster centers $z_1, z_2, \ldots, z_K$ randomly from the n points $\{x_1, x_2, \ldots, x_n\}$.

Step 2: Assign point x_i, $i = 1, 2, \ldots, n$ to cluster C_j, $j \in \{1, 2, \ldots, K\}$ iff

$$\|x_i - z_j\| < \|x_i - z_p\|, p = 1, 2, \ldots, K, \text{and } j \neq p.$$

Step 3: Compute new cluster centers z_i^* as follows:

$$z_i^* = \frac{1}{n} \sum_{x_j \in C_i} x_j, \; i = 1, 2, \ldots, K$$

where n_i is the number of elements belonging to cluster C_i.

Step 4: If there is no more changes in cluster centers, terminate otherwise continue from step 2.

As it can be seen from the first step of this algorithm, we need to randomly choose initial cluster centers. Sometimes, K-Means clustering algorithm converges to suboptimal values depending on the choice of initial cluster centers [23]. On the other hand, K-Means algorithm returns different answers for different initial cluster centers. Here in this work first DNM1 is applied to a frame of video sequence and initial shadow and foreground pixels are determined. Then by spatial averaging of these points, two initial centers for shadow and foreground pixels are determined. Therefore initial cluster centers for each frame depends on DNM1 approach and because this approach always returns one answer for each frame, initial cluster centers are always the same. Therefore, the result of the K-Means algorithm will be always the same for each frame.

In the process of K-Means algorithm, the cluster centers will move, they keep moving till there is no need to move any more. Here shadow and foreground centers are separated from each other as much as possible. So shadow and foreground pixels are easily classified based on the results drawn from the clustering process. This process is illustrated in figure 4. As it can be seen from this figure, first, centers of shadow pixels and foreground pixels are computed and then by using final K-Means clustering algorithm, these centers move and therefore shadow and foreground pixels are classified.

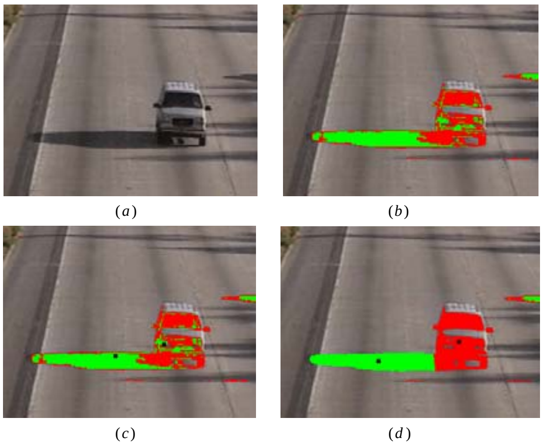

Figure 4: (a) Frame 101 of video sequence (b) Result of DNM1 approach(c) The result of spatial averaging. Two black points are the centers of shadow and foreground pixels (d) Result of shadow detection and the location of shadow and foreground centers after using K-Means clustering algorithm.

4. Experimental results and future works

In this section we analyze the performance of our shadow detection algorithm in a video sequence. In this work, we have used a highway video sequence available for download at *http://cvrr.ucsd.edu/aton/shadow*. Figure 5 illustrates four different frames of highway video sequence. In the first column original frames of video sequence are displayed. The results of DNM1 approach are shown in the second column and the results of K-Means algorithm after DNM1 method are illustrated in the third column. It can be noticed that our approach detects shadows more effectively and there are less misclassification of shadow pixels.

One drawback of our approach is the misclassification of foreground pixels as shadow pixels if an object is located in the shadow of the other object. The reason is that K-Means clustering algorithm returns valid result on a convex dataset. If an object is located in the shadow of another object, the data is said not to be convex. An example of this situation is illustrated in figure 6. As it can be seen, the smaller vehicle is located in the shadow of the bigger vehicle; our data set is not convex. It can be noted form this figure, that the bigger vehicle is classified as foreground while the smaller car is classified as part of its shadow. So as a future work this situation can be improved, using a clustering algorithm returning valid results for not convex datasets.

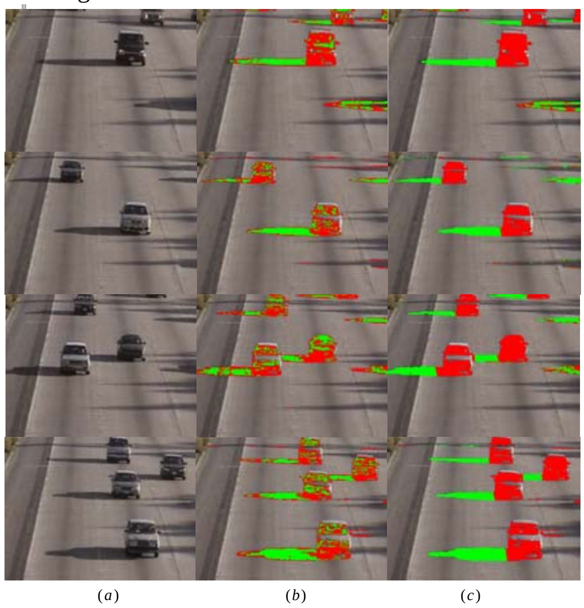

Figure 5: (a) Original frames of video sequence (b) Result of DNM1 approach (c) Result of shadow detection after using K-Means clustering algorithm.

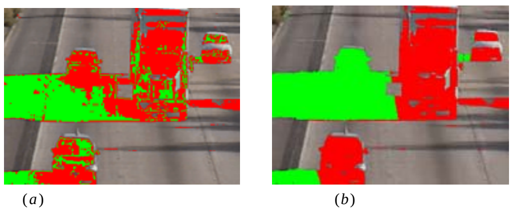

Figure 6: Example of unsuccessful shadow detection. (a) Frame 360 of video sequence after using DNM1. (b) Result of shadow detection after using K-Means clustering algorithm.

5. Conclusions

In this work, we utilized a background modeling, the deterministic non-model based shadow detection algorithm (DNM1) and spatial information along with the color information. To do this, we used K-Means clustering algorithm to separate shadow pixels from foreground pixels. Initial cluster centers of K-Means algorithm was determined by averaging shadow and foreground pixels determined by DNM1.

Our algorithm was tested on an outdoor highway color video sequence and showed considerable improvement in comparison with DNM1 method. Also the drawback of our method was the misclassification of foreground pixels as shadow pixels due to the disability of K-Means algorithm in classifying of non convex datasets.

Further work will concentrate on using a new algorithm which is able to return valid results on non convex datasets.

References

[1] I. Haritaoglu, D. Harwood, and L. S. David. W^4: Real-time Surveillance of People and Their Activities. IEEE Transaction on Pattern Recognition and Machine Intelligence, 22(8):809-830, 2000.

[2] A. Prati, R. Cucchiara, I. MiKic, and M. Trivedi. Analysis and Detection of Shadows in Video Streams: A Comparative Evaluation. CVPR, (2) 571-576, 2001.

[3] C. Stauffer and W. Grimson. Adaptive Background Mixture Models for Real-Time Tracking. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, (2) 246-252, 1999.

[4] A. Elgmmal, D. Harwood, and L. Davis, Non-parametric Model for background subtraction. European Conference on Computer Vision (ECCV), v2 pp. 751-767, 2000.

[5] N. Friedman and S. Russell. Image Segmentation in Video Sequences: A Probabilistic approach. Proceeding of 13th IEEE Conference on Uncertainty in Artificial Intelligence. Aug 1997.

[6] C. Jiang and M.O. Ward. Shadow identification. Proceedings of IEEE Int’l Conference on Computer Vision and Pattern Recognition, pp. 606–612, 1992.

[7] J. Stauder, R. Mech, and J. Ostermann. Detection of moving cast shadows for object segmentation. IEEE Transactions on Multimedia, 1(1):65–76, 1999.

[8] R. Cucchiara, C. Grana, G. Neri, M. Piccardi, and A. Prati. The Sakbot system for moving object detection and tracking. In Video-based Surveillance Systems - Computer Vision and Distributed Processing. pp. 145–157, Kluwer Academic, 2001.

[9] D. Koller, K. Daniilidis, and H. H. Nagel. Model-based object tracking in monocular image sequences of road traffic scenes. International Journal of Computer Vision, vol. 10, pp. 257–281, 1993.

[10] I. Mikic, P. Cosman, G. Kogut, and M.M. Trivedi. Moving shadow and object detection in traffic scenes. In Proceedings of Int’l Conference on Pattern Recognition, vol. 1, pp. 321–324, Sept 2000.

[11] M. M. Trivedi, I. Mikic, and G. Kogut. Distributed video networks for incident detection and management. In Proceedings of IEEE Int’l Conference on Intelligent Transportation Systems, pp. 155–160, Oct. 2000.

[12] A. Elgammal, D. Harwood, and L.S. Davis. Non-parametric model for background subtraction, In Proceedings of IEEE ICCV’99 FRAME-RATE Workshop, 1999.

[13] T. Horprasert, D. Harwood, and L.S. Davis. A statistical approach for real-time robust background subtraction and shadow detection, In Proceedings of IEEE ICCV’99 FRAME-RATE Workshop, 1999.

[14] R. B. Irvin and D.M. McKeown. Methods for exploiting the relationship between buildings and their shadows in aerial imagery. IEEE Transactions on Systems, Man, and Cybernetics, vol. 19, pp. 1564-1575, 1989.

[15] M. Kilger. A shadow handler in a video-based real-time trace monitoring system. Proceedings of IEEE Workshop on Applications of Computer Vision, pp. 11-18, 1992.

[16] D. S. Lee, J. J. Hull and B. Erol. A Bayesian Framework for Gaussian Mixture Background Modeling. IEEE International Conference on Image Processing, vol. 3, pp. 973-976, 2003.

[17] S.-Y. Chien, S.-Y.Ma, and L.-G. Chen. Efficient moving object segmentation algorithm using background registration technique. IEEE Transactions on Circuits and Systems for Video Technology, 12(7):577–586, 2002.

[18] P. L. Rosin and T. Ellis. Image difference threshold strategies and shadow detection. In 6th British Machine Vision Conf., Birmingham, pages 347–356, 1995.

[19] D. Xu, X. Li, Z. Liu, and Y. Yuan. Cast shadow detection in video segmentation. Pattern Recognition Letters, 26(1):5–26, 2005.

[20] J. C. S. Jacques Jr. and C. R. Jung. Background Subtraction and Shadow Detection Video Sequences. In Proceeding of SIBGRAPI, 189-196. 2005.

[21] R. Cucchiara, C. Grana, A. Prati and M. Piccardi. Detecting objects, sahdows and ghosts in video streams by exploiting color and motion information. In Proceedings of the IEEE Int’l Conference on Image Analysis and Processing (ICIAP’01). 360-365, 2001.

[22] N. Herodotou, K. Plataniotis, and A. Venetsanopoulos. A color segmentation scheme for object-based video coding. In Proceedings of the IEEE Symposium on Advances in Digital Filtering and Signal Processing, pages 25–29, 1998.

[23] E. Gose, R. Johnsonbrugh and S. Jost. Pattern Recognition and Image Analysis. Prentice Hall, 1996.