A New Illumination-Invariant Method of Moving Object Detection for Video Surveillance Systems

Elham Kermani[1], Davud Asemani[2]

[1] Electrical Eng. Faculty, K.N. Toosi Univ. of Technology, Tehran, Iran, Kermani@ee.kntu.ac.ir
[2] Electrical Eng. Faculty, K.N. Toosi Univ. of Technology, Tehran, Iran, Asemani@eetd.kntu.ac.ir

Abstract**—**Visual surveillance especially for humans and vehicles is currently one of the most active research topics in computer vision. In this paper, a new method is introduced for the detection of moving objects in surveillance applications. The proposed method relies on a model assigning a vector of gray levels to every pixel location of the current image. The vector represents information on the neighborhood region of that pixel. Using norm of the vectors in two consecutive frames and the Bayesian change detection algorithm, we introduce a novel method for moving object detection which is robust to noise and illumination changes. Also, it is insensitive to repeated motions in the background.

Keywords**—**Moving Object Detection, Bayesian, Vector Model, Surveillance.

I. Introduction

Stationary cameras are extensively used today for perimeter surveillance, motor traffic analysis and environmental monitoring. One of the most important tasks in such applications is change detection, i.e., automatic segmentation of a video sequence into static and changed (e.g., moving) areas [1]. The capability of extracting moving objects from a video sequence is a fundamental and crucial problem of vision systems [2]. Detection of the regions that correspond to moving objects such as people and vehicles in video is the first processing step of almost every vision system, because the other processing tasks such as tracking and activity analysis are implemented in the regions of moving objects [3].

Among the numerous algorithms developed up to now, the simplest ones use mostly a thresholding operation on the intensity differences (e.g., between consecutive video frames or between the current frame and a background frame). These basic methods result often in a poor perform [1]. To achieve a better performance, other methods have been proposed employing probabilistic models [4-7] and statistical tests [8-10]. In these methods, probabilistic models and statistical tests are used to model the background. The performance of these detection algorithms is highly influenced by the method of threshold selection. To improve the performance, it is necessary to adaptively modify threshold value. Up to now, several threshold adaptation methods have been proposed [1]. The most successful algorithms of detection are those exploit frame differencing and modeling of change labels using Markov random field (MRF) in a Bayesian framework [10].

In parallel, change detection methods were developed based on the maximum a posteriori probability (MAP) criterion which use MRFs as prior models [1]. Although MAP-inspired change detection performs well, it is computationally complex, because MAP estimation is an optimization problem which requires special algorithms such as simulated annealing and graph-cuts [11]. We introduce an algorithm which is based on local MAP estimation. Though our approach cannot find the global optimum of the posterior probability but is very accurate and robust. Moreover, it doesn't suffer from heavy computational complexity required in global MAP estimation.

A vector model of images and the related new idea is introduced in Section II-A. Section II-B provides a review on the basic Bayesian change detection algorithm and section II-C describes the proposed algorithm. The illumination variation is handled in Section III. Simulation results are presented in section IV. Finally, section V includes the conclusion.

II. Methodology

A. Vector Representation

To decide whether or not a change did occur at a certain pixel (*m*), we use the gray levels of successive frames which lie within a small sliding window, which is centered around that pixel. These gray levels are ordered into column vectors $\vec{x}$ and $\vec{y}$, respectively. If the window contains *N* pixels (25 in this work), each of these vectors contains *N* components. The vector model for *N*=9 is illustrated in Fig. 1.

Neglecting noise effects, if no scene change occurs within the window and no illumination variation exists between two desired frames, both vectors would be identical. In other words, both vectors will have the same norms in this case. A measure of deviation from this identity case may be defined as ratio of two vectors' norms, namely $k(m)$, as following:

$$k(m) = \frac{|\vec{y}|}{|\vec{x}|} \tag{1}$$

It can be easily deduced that $k(m)$ equals unity if and only if $\vec{x}$ and $\vec{y}$ have the same magnitudes. Otherwise, $k(m)$ is larger or less than one. So, we can say that the probabilities of $k(m)$ being less than or greater than one are equal. It is here supposed that $\ln(k)$ represents a random Gaussian variable having a mean equivalent to zero.

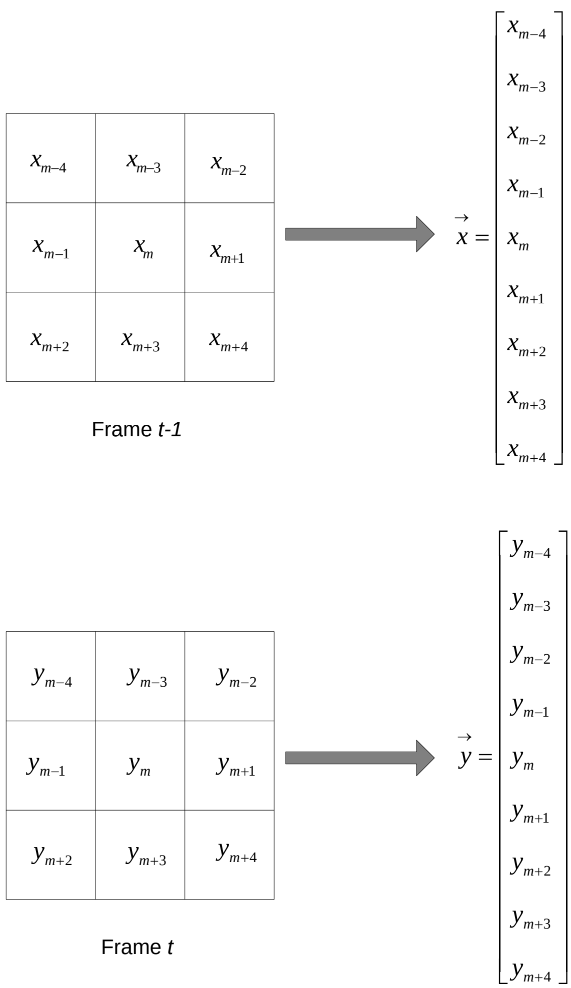

Fig. 1. Illustration of how the vectors are constructed. The region of support for the sliding window is 3×3.

B. The Bayesian Approach

The goal of a motion detection system is to divide each image frame into moving and still segments. It is realized through generating a mask Q consisting of binary labels $q(m)$ for each pixel m on the image grid. The labels take either the label "u" ('unchanged') or "c" ('changed'). In order to determine the label $q(m=i)$ of pixel i, it may be started with the gray-level difference image $D=\{d(m)\}$ between two successive frames, and then comparing the sum of squared differences $\overline{\Delta_i^2}$ through a sliding window w_i having N pixels and center i with a threshold T [10]:

$$\overline{\Delta_i^2} = \frac{1}{\sigma_0^2}\sum_{m\in w_i} d^2(m) \underset{u}{\overset{c}{\gtrless}} T \tag{2}$$

σ_0 is the noise standard deviation of the gray-level differences in stationary areas, which is assumed to be constant over space.

Though, this method performs well but the interior parts of the foregrounds are not detected in the case of big, uniform or slow objects. This is because of differencing two successive frame values. Moreover, it has significant difficulties with changing illumination conditions. Some methods have been proposed to handle the illumination problem of this algorithm [12-15]. The latest technique presented by Liu et al. [15], uses the illumination model for adjusting the image difference. Applying this model, an image difference is obtained being independent of illumination variations. Then, employing this adjusted difference image in the Bayesian framework, a new modified value is presented instead of conventional $\overline{\Delta_i^2}$.

We propose an illumination invariant method based on MAP estimation, which provides an accurate and robust detection of moving objects and then compare the detection results with the presented algorithm in [15].

C. The Proposed Method

Equation 2 is the result of estimating the change mask Q such that its a posteriori probability $P(Q|D)$ given the difference image D is maximized (MAP estimate). Here, we try to maximize $P(Q|K)$, where, K stands for the ratio image. K is a matrix with $k(m)$ values as its components. Now, we repeat the stages applied in [10] to solve the problem:

- First, we assume that the values of the labels $q(m)$ are known for all picture elements except for one element i. Estimating Q then reduces to deciding between $q(i)=u$ and $q(i)=c$ with the resulting change masks Q_u^i, Q_c^i respectively. The decision rule can thus be written as:

$$\frac{P(Q_u^i | K)}{P(Q_c^i | K)} \underset{c}{\overset{u}{\gtrless}} t \tag{3}$$

With t being a decision threshold.

- Using Bayes' theorem and assuming that $k(m)$ values are conditionally independent (i.e. $P(K \mid Q) = \prod_m P(k(m) \mid q(m))$, the decision rule can be rewritten as:

$$\frac{P(k(i) \mid q(i) = u)}{P(k(i) \mid q(i) = c)} \underset{c}{\overset{u}{\gtrless}} t \frac{P(Q_c^i)}{P(Q_u^i)} \tag{4}$$

$P(k|q)$ is the likelihood function and $P(Q_c^i)$, $P(Q_u^i)$ are the prior distributions for Q_c^i, Q_u^i.

- Considering the zero-mean Gaussian distributions with variances σ_c^2, σ_u^2 for the conditional probability distributions results in the following equation:

$$\ln^2(k(i)) \underset{u}{\overset{c}{\gtrless}} 2\frac{\sigma_c^2\sigma_u^2}{\sigma_c^2 - \sigma_u^2}(\ln t\frac{\sigma_c}{\sigma_u} + \ln\frac{P(Q_c^i)}{P(Q_u^i)}) \tag{5}$$

σ_u^2 is the variance of $\ln(k)$ in the pixels which their variations are only due to noise and σ_c^2 is related to the pixels in which the variations are caused by anything except noise, such as illumination, motion, etc. It is clear that σ_c^2 is much larger than σ_u^2 because of the large magnitude of differences in changed areas. In this paper, we have considered a very small value for σ_u^2 and a hundred to a thousand times as big for σ_c^2.

Equation 5 has two considerable properties:

1- Low sensitivity to the selection of σ_c, σ_u: The constant term in threshold value is proportional to $\ln t(\sigma_c/\sigma_u)$. Since σ_c is always much larger than σ_u, the term $\ln t(\sigma_c/\sigma_u)$, almost always places at saturation region of ln(.) function (region 1 in Fig. 2). So, you can consider a fixed value for the constant term in different sequences.

2- High detection sensitivity: k varies around 1 with small deviations. So, $\ln(k)$ often appears at rising region of ln(.) function (region 2 in Fig. 2). Large slope of this curve at region 2 leads to a high sensitivity in change detection even in the case of small k values.

To reduce the errors produced by global thresholding, we locally adapt the decision threshold using the prior knowledge which is expressed by Markov Random Field models [10]. Thus, we obtain a threshold T that adapts to the label constellation within a pixel's neighbourhood:

$$T = T_0 + (4 - n_i).B \tag{6}$$

Where T_0 is equal to the constant term in right side of equation 5. The parameter B is a positive-valued potential and n_i is the number of changed pixels in 3×3 neighbourhood of each pixel. Finally, the decision rule is expressed as:

$$\ln^2(k(i)) \underset{u}{\overset{c}{\gtrless}} T_0 + (4 - n_i)B \tag{7}$$

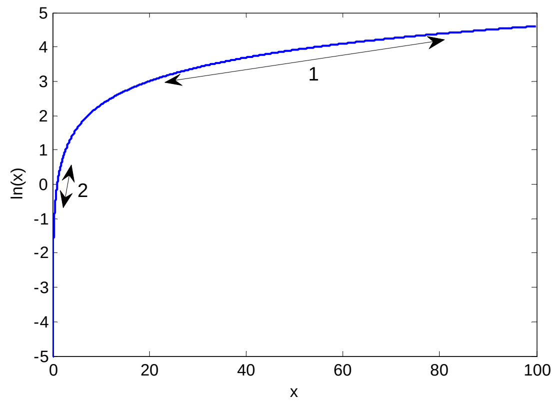

Fig. 2. plot of function ln(x) versus x. Region 1 and 2 are corresponding to the variation range of the terms $\ln t(\sigma_c/\sigma_u)$ and $\ln(k)$, which are present in the proposed algorithm.

III. Illumination Variations

One of the main problems of this algorithm is its sensitivity to illumination variations. Because every change that causes $\ln^2(k)$ to become larger than T, is considered as a motion change. Thus, since the difference caused by illumination is much larger than changes due to noise, $\ln^2(k)$ exceeds T and consequently the algorithm will not be able to differentiate illumination variations from motion changes.

To overcome this problem, we have modeled the illumination as an offset in each window. A structure has been proposed which approximates the offset value using the difference vector of $\vec{x}$ and $\vec{y}$ (i.e. $\vec{d} = \vec{x} - \vec{y}$) and then removes it. The structure is shown in Fig. 3. u_d is the mean of difference vector $\vec{d}$. When illumination variations occur, the mean of the difference vector approaches zero in the area of a moving object, else it has an absolute value greater than zero.

If we remove the offset before computing the k value, no false detection will be produced by illumination changes. The experiments show that even one operation ensures an acceptable performance and there is no need to feedback u_d and repeat the process.

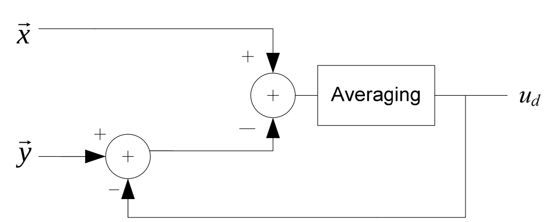

Fig. 3. The proposed structure to model the illumination effect as an offset.

IV. RESULTS

The performance of the proposed method was evaluated using several indoor and outdoor sequences and then compared to the proposed algorithm in [15].

Fig. 4 shows the results for two of the indoor test sequences. As can be seen, the proposed algorithm can separate the moving objects almost perfectly. The overall performance of our approach seems to be better than the performance of the other method.

Fig. 5(a,b) shows two successive frames with varying illumination in a video sequence. Fig. 5(c) represents the result after applying the modified basic Bayesian algorithm of [15]. Fig. 5(d) is the detection result for the proposed algorithm.

A considerable property of the proposed algorithm is that it can handle repeated motions in background such as waving trees (Fig. 6). Moreover it is not very sensitive to shadows as can be seen in Fig. 4 and Fig. 5. These properties derive from using two consecutive frames for the detection algorithm.

The variances σ_u^2, σ_c^2 were set to 0.001 and 0.2 experimentally and the parameter B to B=0.003. The simulation results are not very sensitive to the variance values because the logarithm reduces range of the changes significantly. This property allows using a fixed value for T_0 in equation 7.

To quantitatively evaluate and better compare the algorithms, we used the similarity measure introduced in [6]. The similarity measure is defined as:

$$S(A,B) = \frac{A \cap B}{A \cup B} \tag{8}$$

where A is a detected region and B is the corresponding ground truth. $S(A,B)$ approaches to a maximum value of 1 if A and B are the same.

Some frames from the sequences used (intelligent room, walk, traffic and trees) were randomly selected to evaluate algorithms based on manually-produced ground truth. The averaging values of similarity measures for mentioned video sequences are shown in Table 1. Last row shows the average of results for four sequences. Comparison illustrates that the proposed method leads to highly-improved detection results rather than the other method.

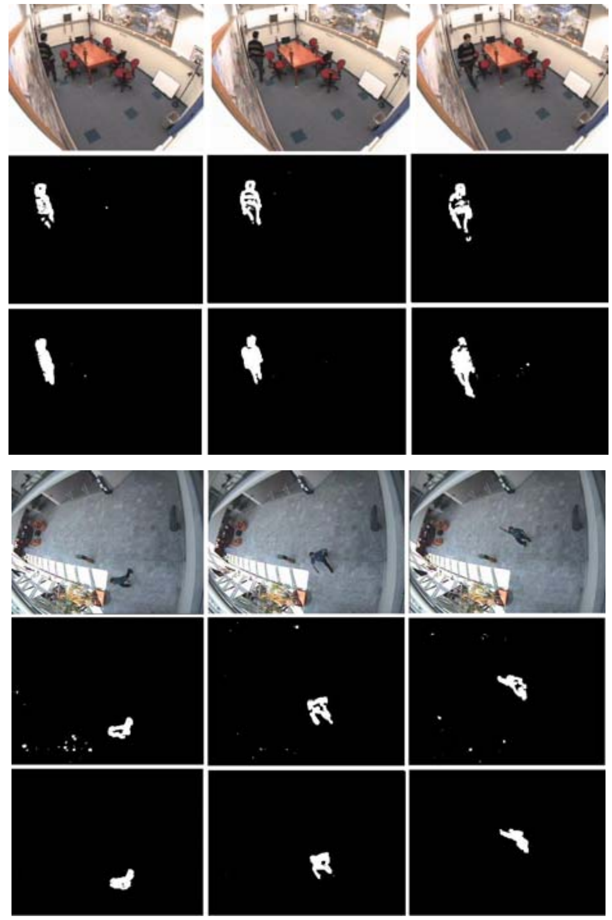

Fig. 4. Detection results for two indoor sequences: The top row of each image shows three frames from an indoor test sequence where a person is walking in a room**.** The two other rows show the results of the proposed algorithm and change detection algorithm proposed by [15], respectively.

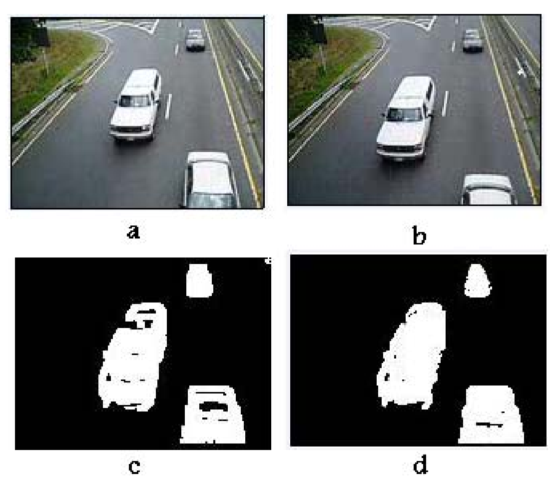

Fig. 5. Detection results: (a),(b) two successive frames of traffic sequence with varying illumination (c) modified Bayesian change detection algorithm [15] (d) proposed method

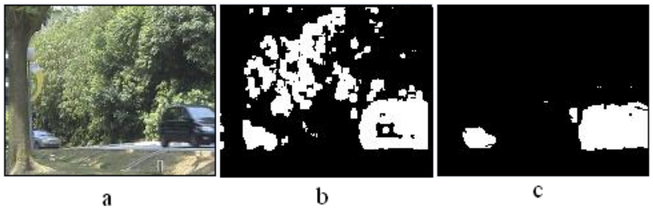

Fig. 6. Experimental results on a campus environment containing wavering tree branches in strong winds: (a) a frame from the sequence (b) modified Bayesian change detection algorithm of [15] (c) proposed method

TABLE 1. QUANTITATIVE EVALUATION: S(A,B) VALUES FROM THE TEST SEQUENCES

	Proposed Method	[15]
Intelligent Room	0.73	0.49
Walk	0.61	0.41
Traffic	0.71	0.64
Trees	0.69	0.23
Average	0.68	0.44

V. CONCLUSION

A new algorithm for motion detection was proposed in this paper which is almost robust against noise, illumination variations and repeated motions in the background. The algorithm provides good detection results, i.e., detects objects and suppresses local and global illumination changes successfully. Quantitative evaluation and comparison with the existing methods have shown the effectiveness of the proposed method. The proposed method results in a similarity measure of 68% compared to 44% associated with previously presented method.

REFERENCES

[1] J.M. McHugh, J. Konrad, V. Saligrama, P. Jodoin, "Foreground-Adaptive Background Subtraction," IEEE Signal Processing Letters, vol. 16, issue 5, pp. 390-393, 2009.

[2] L. Ovsenik, A. Kolesarova, J. Turan, "Video Surveillance Systems," Acta Electrotechnica et Informatica, Vol. 10, No. 4, 2010.

[3] H.R. Shayegh, N. Moghadam, " A New Background Subtraction Method in Video Sequences Based on Temporal Motion Windows," International Conference on IT, 2009.

[4] C.R. Wren, A. Azarbayejani, T. Darrell., A. Pentland, "Pfinder: Realtime tracking of the human body," IEEE Trans Pat Anal Mach Intel, pp. 780-785, 1997.

[5] C. Stauffer, W. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking," CVPR, IEEE Comput Society Conf, 1999.

[6] A. Elgammal, R. Duraiswami, D. Harwood, L. Davis, "Background and foreground modeling using nonparametric kernel density for visual surveillance," Proc. IEEE, vol. 90, pp. 1151-1163, 2002.

[7] 7. L. Li, W. Huang, I.Y. Gu, Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," IEEE Trans. Image Processing, vol. 13, no. 11, pp. 1459-1472, 2004.

[8] Y.Z. Hsu, H.H. Nagel, G. Refers, "New likelihood test method for change detection in image sequences," Computer Vision, Graphics and image Processing, vol. 26, pp. 73–106, 1984.

[9] T. Aach, A. Kaup, R. Mester, "Statistical model-based change detection in moving video. Signal Processing," vol. 31, pp. 165–180, 1993.

[10] T. Aach, A. Kaup, "Bayesian algorithms for adaptive change detection in image sequences using Markov random fields," Signal Process: Image Comm, 7(2): 147-160, 1995.

[11] Z. Kato, "Multiresolution Markovian models in computer vision. Application on segmentation of SPOT images," PhD thesis, INRIA, Sophia Antipolis, France, 1994.

[12] D. Toth, T. Aach, V. Metzler, "Bayesian spatio-temporal motion detection under varying illumination," European Signal Processing Conference, Tampere, Finland, 2000.

[13] R. Mester , T. Aach , L. Dümbgen, "Illumination-Invariant Change Detection Using a Statistical Colinearity Criterion, Proceedings of the 23rd DAGM-Symposium on Pattern Recognition," pp. 170-177, September, 2001.

[14] T. Aach, L. D¨umbgen, R. Mester, D. Toth, "Bayesian illumination-invariant motion detection," IEEE Intl.Conf. Im. Proc. (ICIP):in press, Thessaloniki, Oct., 2001.

[15] Q. Liu, M. Sun, R.J. Sclabassi, "Illumination-invariant change detection model for patient monitoring video," Conf. Eng. Med. Biol. Soc., 1:1782-1785, 2004.