IJICT International Journal of Information & Communication Technology

ITRC

Volume 2- Number 2- July 2010 (pp.1- 8)

A Novel Descriptor for Pedestrian Detection in Video Sequences

Narges Armanfard
Dept. of Electrical and Computer Eng.
Tarbiat Modarres University
Tehran, Iran
armanfard@modares.ac.ir

Majid Komeili
Dept. of Electrical and Computer Eng.
Tarbiat Modarres University
Tehran, Iran
komeili@modares.ac.ir

Ehsanollah Kabir
Dept. of Electrical and Computer Eng.
Tarbiat Modarres University
Tehran, Iran
kabir@modares.ac.ir

Received: February 28, 2009- Accepted: July 26, 2010

Abstract—This paper presents a novel Texture-Edge Descriptor, TED, for background modeling and pedestrian detection in video sequences which models texture and edge information of each image block simultaneously. Each block is modeled as a group of adaptive TED histograms that are calculated for pixels of the block over a rectangular neighborhood. TED is an 8-bit binary code which is independent of the neighborhood size. Experimental results over real-world sequences from PETS database clearly show that TED outperforms LBP.

Keywords- pedestrian detection, texture, edge, block-based approach, background subtraction, surveillance systems.

I. INTRODUCTION

Pedestrian detection is one of the most challenging tasks in surveillance systems. The output of pedestrian detection can be the input of higher level processes such as pedestrian tracking [1, 2]. Pedestrian detection with a single fixed camera usually involves foreground detection and moving region classification. In [3], at first, moving regions are extracted, then they are classified into single person, people in a group and other objects. In [4], a background model is computed to detect foreground regions in each frame. Each region is classified as either human or other object based on its shape and appearance. The overall performance of pedestrian detection depends on how accurately the foreground regions are detected.

Background subtraction [5, 6], optical flow [7, 8] and temporal differencing [9] are conventional approaches for moving region extraction. Background subtraction is a popular method in surveillance systems with single fixed camera. In order to detect moving objects, each incoming frame is compared with a background model learned from previous frames. The performance of background subtraction depends mainly on the background modeling.

One of the difficulties of background modeling is that the backgrounds are usually non-stationary. When a background subtraction technique is applied for a surveillance system which captures outdoor scenes, it detects not only the moving objects but also a lot of noise since it shows great sensitivity to small changes. These changes are caused by, for example, waving leaves, fluttering flags and ripple water. Furthermore, shadows [10] and sudden lighting changes [11, 12] could cause some limitation to background modeling.

Dealing with these problems, a number of methods have been proposed for background modeling which utilize different features and descriptors. Most background modeling methods are pixel-based and have the advantage of extracting detailed shapes of moving objects. However their drawback is that their segmentation results are sensitive to non-stationary scenes. In [3] each pixel is modeled during training phase by three values; its minimum and maximum intensity and the maximum intensity difference between consecutive frames. In order to obtain clean background images, a pixel-wise median filter over time is applied. In [13], Mixture of Gaussian, MoG, is proposed in color space which uses K Gaussian distributions with different means and standard deviations. In [14], intensity values are modeled by support vector regression. The algorithm presented in [15], segments foreground objects from dynamic textured background by using Kalman filter. In [16], a framework for Hidden Markov Model topology and parameter estimation was proposed. In [17] color and edge information are fused to detect foreground regions. In [18], each pixel is modeled as a group of adaptive Local Binary Pattern, LBP, histograms that are calculated over a circular region around the pixel. In [19], normalized coefficients of five orthogonal transforms (DCT, DFT, Haar, SVD and Hadamard) are utilized to detect moving regions.

In many applications such as surveillance systems, there is no need to detect the detailed shapes of moving objects. Recently some researchers used block-based methods instead of pixel-based. In these methods, each image is divided into either overlapping or non-overlapping blocks. Since each block monitors more global changes in the scene, these methods are more suitable for non-stationary scenes. Besides, these methods decrease the complexity. For example, for a 320×240 image, pixel-based methods make 76800 models for each image, whereas by dividing it into 8×8 blocks, only 1200 models are needed. In [20], a normalized vector distance is used as a measure for blocks correlation. In [21], edge and color histograms are used for modeling each block. In [22], texture of each block is modeled by LBP. In [23], contrast histogram for each block is used for background modeling. In [24], each N×N block is represented by an N^2-dimensional vector which its elements are intensity values of the pixels in that block.

Our interest is to estimate the moving regions occupied by a pedestrian in video sequences. In this paper a new descriptor, called TED, is proposed which models texture and edge information, simultaneously. TED is an 8-bit binary code. This small size is an important property from the implementation point of view. Similar to LBP, TED is based on intensity difference, hence robust against illumination changes. We use TED in a block-based approach; so it is more capable of dealing with non-stationary backgrounds. A preliminary version of this paper has appeared in [25].

The paper is organized as follows: Section II discusses the texture description with Local Binary Pattern. Section III introduces our proposed Texture-Edge Descriptor. Section IV presents the experimental results and the performance evaluation. Conclusion and future work are given in Section V.

II. TEXTURE DESCRIPTION WITH LOCAL BINARY PATTERN

Local Binary Pattern, LBP, is vastly used for texture description and has good performance in texture classification [26], fabric defect detection [27] and moving region detection [18]. In this approach, texture feature assigned to a pixel is local feature which consider its neighboring pixels. The common version of LBP operator is defined as follows [26]:

$$LBP(P_c) = \sum_{n=0}^{P-1} s(g_n - g_c)2^n$$
$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (1)$$

Where g_c is the intensity value of center pixel, P_c and g_ns are the intensities of neighboring pixels. Basic LBP considers a 3×3 neighborhood as shown in Fig. 1 [28].

III. PROPOSED TEXTURE-EDGE DESCRIPTOR

LBP is a general texture descriptor. Having the advantages of LBP, we proposed a new descriptor customized for pedestrian detection. Our descriptor is developed using the fact that vertical edge is a significant feature of pedestrian image [29].

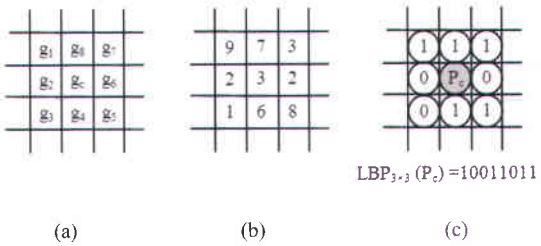

Fig. 1. (a) a 3×3 neighborhood of g_c. (b) Their intensity values. (c) Binary number assigned to P_c; start from top left pixel anticlockwise.

A. Proper neighborhood for pedestrian detection

The neighborhood of P_c for computing TED is considered in such a way that it includes not only texture but also vertical edge information. Human shape features more vertical edges than the background [29, 30]. Furthermore, vertical edges of pedestrian are short and fragmentary. The intensity profile over a typical row of pedestrian vertical edge shows some small-scale intensity variations in addition to the large-scale variation. To model large-scale intensity variation and disregard small-scale variations, we extend the neighborhood region horizontally as shown in Fig. 2. Where g_c is the intensity value of central pixel P_c and $g_n\, n=1,...,16$ is the intensity value of neighboring pixels. Extension along horizontal direction lets us model vertical edges corresponding to the large-scale intensity variations. Small vertical neighborhood is because of the fact that vertical edges of pedestrian image are short and fragmentary. To make addressing neighboring pixels easier, the neighborhood is divided into four regions:

Upper Row, UR, Lower Row, LR, Right Column, RC, and Left Column, LC (see Fig. 2).

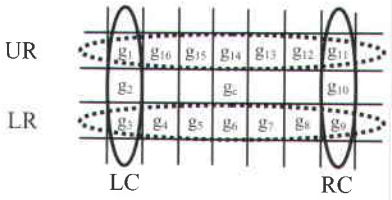

Fig. 2. UR, LR, RC and LC regions in a 3×7 rectangular neighborhood

These four regions represent texture of central pixel. Furthermore, the intensity values on LC and RC can represent vertical edges. If P_c is an ideal vertical edge, the gray values of pixels on the RC (LC) together are bigger or smaller than g_c .With respect to section II, the straightforward way of texture description over the rectangular neighborhood is LBP operator. But the downside is that the complexity increases exponentially with the number of neighboring pixels. Hence other solution has to be found.

B. TED code

The main disadvantage of LBP over a rectangular neighborhood is that the complexity increases exponentially with the number of bits, P. For more illustration, considering a neighborhood of size $w \times l$, the number of neighboring pixels is $P = 2w + 2l - 4$. For example in a 3×7 neighborhood, LBP returns a 16-bit binary number as shown in Fig. 3. So texture of each pixel can have 2^{16} different values which is not applicable in real time applications.

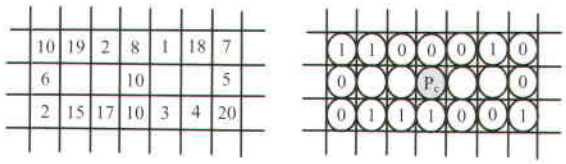

Fig. 3. (a) Intensity values of a sample 3×7 neighborhood. (b) binary number assigned to P_c; starting from top left pixel anticlockwise

To cope with this complexity while modeling texture and edge simultaneously, we propose an efficient descriptor called TED. As shown in Fig. 4, TED is an 8-bit binary code defined over binary representation of LBP. For simplicity, the rectangular neighborhood where TED is computed over it is named TED window.

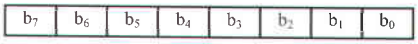

Fig. 4. 8-bit Texture-Edge binary code (TED code)

b_0 and b_1 stand for the texture of upper neighborhood of P_c and are defined as:

$$b_0 = \begin{cases} 1 & UR1s > (l-1)/2 \\ 0 & UR1s \le (l-1)/2 \end{cases}$$
$$b_1 = \begin{cases} 1 & URT \ge (l-1)/2 \\ 0 & URT < (l-1)/2 \end{cases} \quad (2)$$

Where UR1s indicates the number of ones in the upper row, UR; and URT denotes the number of transitions between 0 and 1 in the upper row. b_3 and b_4 describe texture of lower row neighborhood and are defined as follows:

$$b_3 = \begin{cases} 1 & LR1s > (l-1)/2 \\ 0 & LR1s \le (l-1)/2 \end{cases}$$
$$b_4 = \begin{cases} 1 & LRT \ge (l-1)/2 \\ 0 & LRT < (l-1)/2 \end{cases} \quad (3)$$

Where LR1s and LRT are defined in a similar way as UR1s and URT, respectively. The remaining bits of TED code represent edge pixel, single point and flat area in addition to describing texture of P_c.

$$b_2 = \begin{cases} 1 & RCT = 0 \\ 0 & RCT \ne 0 \end{cases}$$
$$b_5 = \begin{cases} 1 & LCT = 0 \\ 0 & LCT \ne 0 \end{cases} \quad (4)$$

$$b_6 = \begin{cases} 1 & UR1s + LR1s + RC1s + LC1s = 0 \\ 0 & UR1s + LR1s + RC1s + LC1s \ne 0 \end{cases}$$
$$b_7 = \begin{cases} 1 & UR1s + LR1s + RC1s + LC1s = P \\ 0 & UR1s + LR1s + RC1s + LC1s \ne P \end{cases} \quad (5)$$

Where RCT and LCT are the number of transitions between 0 and 1 in the right column, RC, and left column, LC, respectively. RC1s and LC1s are the number of ones in the right and left column, respectively. P is the number of neighborhood pixels. For more illustration see Table 1.

Table 1. binary values of b_2, b_5, b_6 and b_7 correspond to edge pixel, single point and flat area

b_2	b_5	b_6	b_7	
1	X	0	0	**Vertical edge pixel**
X	1	0	0	**Vertical edge pixel**
1	1	1	0	**Single point**
1	1	0	1	**Single point or Flat area**

The first two rows of Table 1 depict the case that pixels intensity on RC (LC) together are bigger or smaller than the intensity of P_c; as a result, all binary values on RC (LC) return 1 or 0, respectively. In other words, this case leads to RCT=0 (LCT=0) or equally $b_2 = 1$ $(b_5 = 1)$. The third and fourth rows show the case that P_c is a single point or belongs to a flat area as shown in Fig. 5.

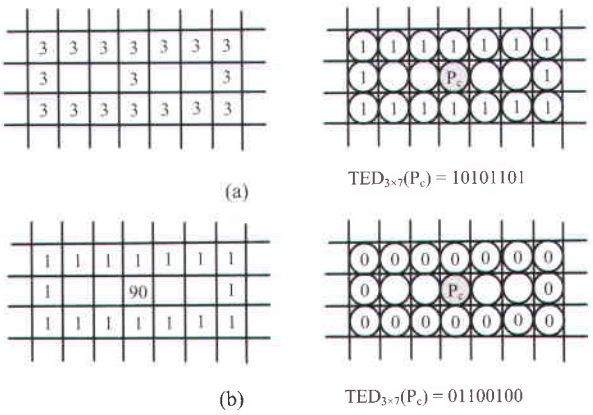

Fig. 5. a) a dark flat region b) a bright single point

For example in Fig. 5(a) UR1s=7, URT=0, LR1s=7, LRT=0, RCT=0, LCT=0, RC1s=3 and LC1s=3. Considering equations (2) to (5), it results in $TED_{3\times7}(P_c)$ =10101101. In a similar way, it is obvious that in Fig. 5(b) $TED_{3\times7}(P_c)$ =01100100. Note that, in these two cases, although there is not any transition between 0 and 1 in the RC and LC, with respect to Table 1, P_c is not considered as edge pixel.

The main advantage of TED is that it always has eight bits independent of the TED window size. These bits include implicitly texture and edge information. We demonstrate the effectiveness of TED for pedestrian detection in a block-based approach introduced in the next section.

IV. Experimental Results

We demonstrate the effectiveness of our descriptor on three indoor and outdoor real-world sequences from PETS data base [31]. As shown in the first column of Fig. 6. The first sequence, *Seq. 1*, is from *PETS2001* which is an outdoor scene containing small moving objects and its resolution is 768×576. The second one, *Seq. 2*, is selected from *PETS2002* which is an indoor sequence. This sequence is dynamic because of the reflection of light from the shop window and its resolution is 640×240. The other sequence, *Seq. 3*, is from *PETS2006* which is an indoor sequence and its resolution is 768×576. Light reflection is the main problem with this sequence. Each frame of these sequences is divided into non-overlapped blocks. The Size of blocks depends on the size of moving object. In PETS2001 and PETS2006, the block size is 8×16 and in PETS2002 is 16×24.

In this section, our proposed method is compared with the basic LBP described in Sec. II. We use the block-based foreground detection framework such as what is presented in [22] which is the earlier version of [18].

1) *Block based foreground detection*

Suppose that gray level images are divided into non-overlapping blocks. The history of each block is modeled by K adaptive weighted TED (LBP) histograms, $\{\vec{q}_1, \vec{q}_2, ..., \vec{q}_K\}$. The weight of the kth histogram is denoted as w_k. Let $\vec{h}$ be the TED (LBP) histogram of block B in the current frame.

Step 1) Distance between $\vec{h}$ and the kth histogram is measured using L1-distance as follows:

$$D_k = \sum_{n=0}^{N-1} \left| \overrightarrow{h_n} - \overrightarrow{q_{k,n}} \right| \qquad k = 1,2,...,K \tag{6}$$

Where h_n is the nth bin of $\vec{h}$, $\overrightarrow{q_{k,n}}$ the is nth bin of kth histogram and N is the number of histogram bins.

Step 2) If the distance between $\vec{h}$ and at least one of the existing histograms is lower than threshold T_l, block B is defined as background, otherwise it is considered as foreground.

Step 3) If block B is defined as background, the best matching histogram is adapted with $\vec{h}$ as follows:

$$\overrightarrow{q_k} = \alpha\vec{h} + (1-\alpha)\overrightarrow{q_k} \qquad \alpha \in [0,1] \tag{7}$$

Where α is a user-settable learning rate. The weights of all model histograms are updated according to (8) and are normalized to ensure $\sum_{k=1}^{K} w_k = 1$.

$$w_k = (1-\beta)w_k + \beta M \qquad \beta \in [0,1] \tag{8}$$

Where β is another user-settable learning rate. M=1 is set for the best matching histogram and M=0 is set for the remaining histograms.

Step 4) If block B is defined as foreground, the histogram with the lowest weight is replaced with $\vec{h}$ and a low initial weight is given to the new histogram. In our experiments, a value of 0.01 was used. No further processing is required in this case.

Step 5) The model histograms are sorted in decreasing order according to their weights and the first B histograms are selected as background histograms.

$$B = \arg\min_{b}\left(\sum_{k=1}^{b} w_k > T_B\right) \tag{9}$$

Where T_B is a user predefined threshold.

The number of model histograms, K, is in proportion to scene complexity. In non-stationary scenes a higher value of K is needed which is computationally expensive and uses more memory. In our experiments, *Seq. 1* and *Seq. 2*, K parameter is 4 and T_B is 0.9 and in *Seq. 3*, K and T_B are 3 and 0.85, respectively. The learning rates α and β are set as 0.01.

The bigger the learning rate, the faster the adaptation. The neighborhood size for computing TED is set experimentally. In our experiments, the proper neighborhood size is 3×7 for all three sequences.

As mentioned earlier, TED is an 8-bit binary code independent of the neighborhood size; while the number of LBP code bits is equal to P. So in order to have a fair comparison between these two descriptors, in our experiments, we compute LBP codes over a 3×3 neighborhood size so that P=8 (i. e $LBP_{3\times3}$).

2) Evaluation

Original images, ground truth images and detected regions for two different methods are shown in Fig. 6. It can be seen that the results by using TED have higher true detected blocks and lower false detected blocks. We used numerical evaluations in addition to visual interpretations. Let A_d be a detected region and A_g be the corresponding ground truth. The similarity between A_d and A_g is defined as

$$S(A_d, A_g) = \frac{A_d \cap A_g}{A_d \cup A_g} \tag{10}$$

Where $S(A_d, A_g)$ lies between 0 and 1. If A_d and A_g are the same, $S(A_d, A_g)$ is 1, otherwise 0 if A_d and A_g have the least similarity.

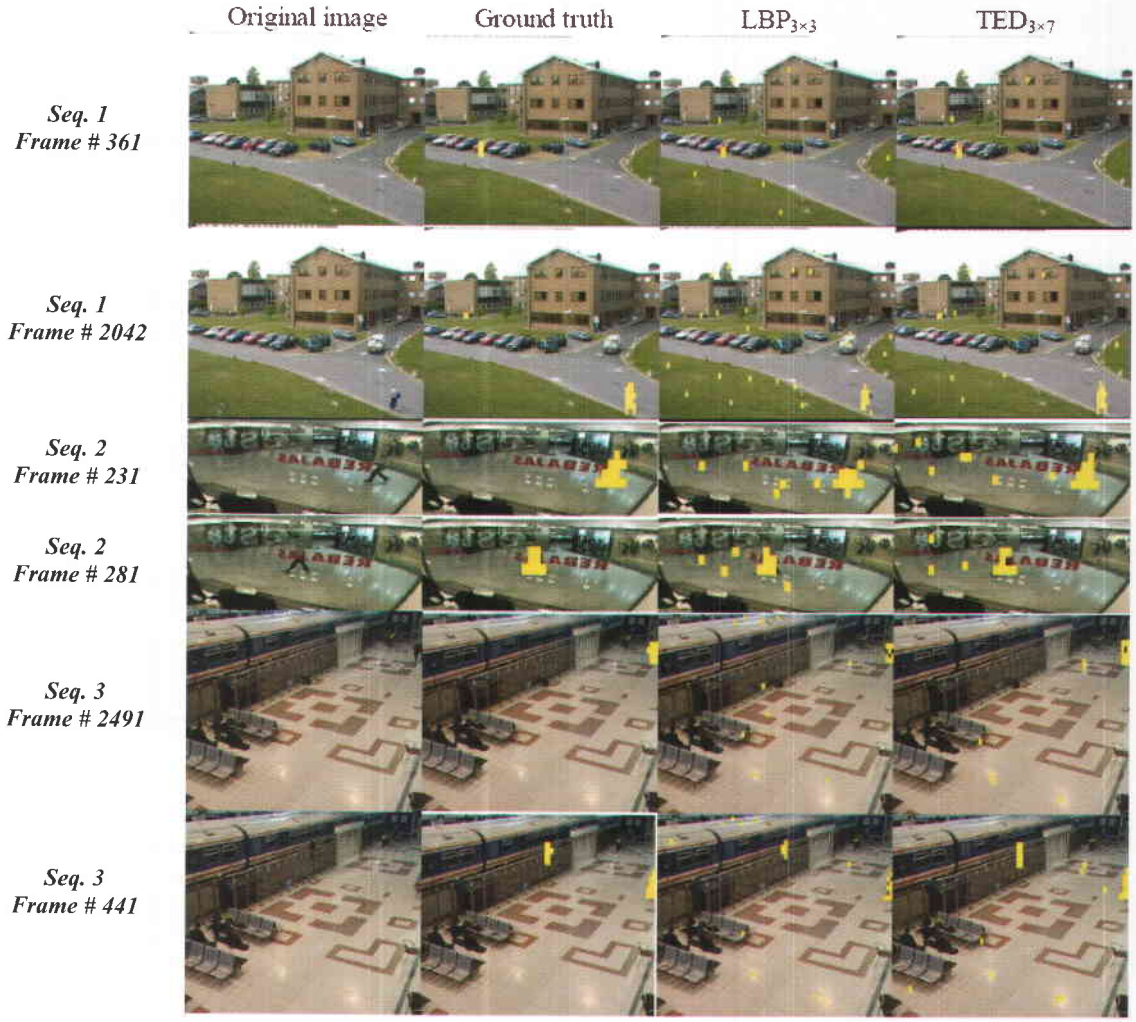

Fig. 6 Detection results using $LBP_{3\times3}$ and $TED_{3\times7}$ for six frames from *PETS* database. From left to right, the first column is original image, the second column is ground truth, the third and fourth columns are detection results by $LBP_{3\times3}$ and $TED_{3\times7}$, respectively

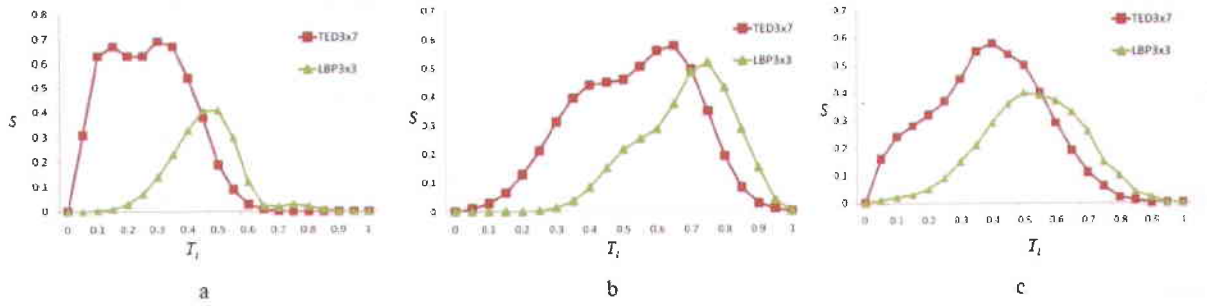

Fig. 7 S versus T_l for $LBP_{3\times3}$ and $TED_{3\times7}$ a) Frame # 361 of *Seq. 1*, b) Frame #231 of *Seq. 2* c) Frame # 2491 of *Seq. 3*

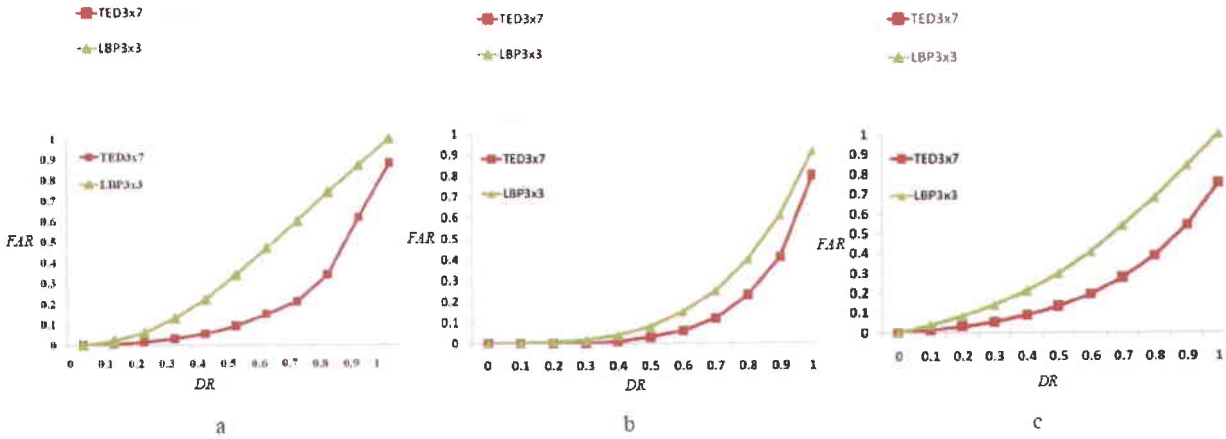

Fig. 8 FAR versus DR for $LBP_{3\times3}$ and $TED_{3\times7}$ a) Frame # 361 of *Seq. 1*, b) Frame # 231 of *Seq. 2* c) Frame # 2491 of *Seq. 3*

Table 2 Detection results for LBP3×3 and our proposed method, $TED_{3\times7}$.

	DR		FAR		S		T_l		Block size
	$LBP_{3\times3}$	$TED_{3\times7}$	$LBP_{3\times3}$	$TED_{3\times7}$	$LBP_{3\times3}$	$TED_{3\times7}$	$LBP_{3\times3}$	$TED_{3\times7}$	
Seq. 1 #361	0.66	0.83	0.61	0.32	0.32	0.6	0.5	0.3	8×16
Seq.1 #2042	0.7	0.9	0.58	0.3	0.35	0.65			
Seq. 2 #231	0.75	0.9	0.37	0.38	0.52	0.58	0.75	0.65	16×24
Seq. 2 #281	0.75	0.88	0.39	0.33	0.51	0.61			
Seq. 3 #2491	0.7	0.84	0.51	0.37	0.4	0.57	0.55	0.4	8×16
Seq. 3 #441	0.75	0.9	0.4	0.31	0.5	0.64			

Detection and false alarm rates are also used for evaluation. Suppose, TP is the number of detected blocks that correspond to moving objects, FP is the number of detected blocks that do not correspond to a moving object and FN is the number of moving blocks not detected. These parameters are combined to define Detection Rate, DR, and False Alarm Rate, FAR. The ground truths are marked manually.

$$DR = \frac{TP}{FN + TP}$$
$$FAR = \frac{FP}{FP + TP} \quad (11)$$

Fig. 7 shows the S parameter versus T_l for $LBP_{3\times3}$ and $TED_{3\times7}$. It illustrates that the value of S for $TED_{3\times7}$ is higher than LBP3×3 for a wide range of T_l. In addition, the maximum S for $TED_{3\times7}$ is higher than the $LBP_{3\times3}$. It shows better performance of TED compare to LBP especially in *PETS2001* which is an outdoor sequence with dynamic scene. It confirms that our proposed method is more robust in dynamic scenes. Detection results given in Table 2 are provided by the value of T_l which has maximum S obtained from Fig. 7. As is shown in Table 2, both DR and S for our proposed method, $TED_{3\times7}$, are higher than $LBP_{3\times3}$ and the FAR is lower. Fig. 8 shows FAR versus DR. It can be understood that for a constant DR, $TED_{3\times7}$ has lower FAR than $LBP_{3\times3}$.

V. Conclusion and Future Work:

In this paper we proposed a novel Texture-Edge Descriptor for pedestrian detection in video sequences. The strengths of TED are as follows: 1) TED encodes texture and edge information simultaneously over a neighborhood. 2) The size of TED, in contrast to LBP,

is always 8 bits independent of the neighborhood size which is an important property from the implementation point of view. Therefore TED can model a larger neighborhood than LBP and gives better detection results. 3) Like LBP, TED is based on the intensity difference so it is robust against illumination changes.

Currently, each frame is divided into non-overlapping blocks. We plan to extend our work in order to extract features from overlapping blocks to increase the detection rate. In this paper we used gray images for pedestrian detection, while one can use color information to improve the detection rate.

ACKNOWLEDGEMENTS

This work was supported in part by Iran Telecommunication Research Center, ITRC, under contact T500-7213, TMU-78-06-32.

REFERENCES

[1] M. Komeili, N. Armanfard, and E. Kabir, "Adaptive visual tracking by decision level fusion of features in a particle filtering frame work," in *The Fifth Iranian Conference on Machine Vision and Image Processing*, 2008.

[2] M. Komeili, N. Armanfard, and E. Kabir, "A fuzzy approach for multi-feature pedestrian tracking with particle filter," in *IEEE International Symposium on Telecommunications*, 2008.

[3] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: Real-time surveillance of people and their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, pp. 809–831, 2000.

[4] C. Dai, Y. Zheng, and X. Li, "Pedestrian detection and tracking in infrared imagery using shape and appearance," *Computer Vision and Image Understanding,* vol. 106, pp. 288–299, 2007.

[5] D. S. Lee, "Effective gaussian mixture learning for video background subtraction," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, pp. 827-835, No. 5, 2005.

[6] P. Spagnolo, T. D. Orazio, M. Leo, and A. Distante, "Moving object segmentation by background subtraction and temporal analysis," *Image and Vision Computing,* vol. 24, pp. 411-423, 2006.

[7] J. Barron, D. Fleet, and S. Beauchemin, "Performance of optical flow techniques," *Computer Vision,* vol. 12, pp. 42-77, No. 1, 1994.

[8] S. Fejes and L. S. Davis, "What can projections of flow fields tell us about the visual motion," Bombay, India, 1998.

[9] N. Paragios and R. Deriche, "Geodesic active contours and level sets for the detection and tracking of moving objects," *IEEE Transactions on Pattern Analysis and Machine Interface,* vol. 22, pp. 266-280, 2000.

[10] N. Martel-Brisson and A. Zaccarim, "Moving cast shadow detection from a gaussian mixture shadow model," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[11] M. Komeili and E. Kabir, "Robust color-based pedestrian tracking in varying illumination environment," in *Iranian Conferance on Electrical Engineering*, 2008.

[12] L. Li, W. Huang, I. Y.-H. Gu, and Q. Tian, "Statistical modeling of complex backgrounds for foreground object detection," *IEEE Transactions on Image Processing,* vol. 13, pp. 1459-1472, No. 11, 2004.

[13] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1999.

[14] J. Wang, G. Bebis, and R. Miller, "Robust Video-Based Surveillance by Integrating Target Detection with Tracking," in *Proceedings of the IEEE Computer Vision and Pattern Recognition Workshop*, 2006.

[15] J. Zhong and S. Sclaroff, "Segmenting Foreground Objects from a Dynamic Textured Background via a Robust Kalman Filter," in *Ninth IEEE International Conference on Computer Vision*, 2003.

[16] C. Ridder, O. Munkelt, and H. Kirchner, "Adaptive background estimation and foreground detection using kalman-filtering," in *Proceedings of International Conference on Recent Advances in Mechatronics*, 1995.

[17] S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld, "Detection and location of people in video images using adaptive fusion of color and edge information," in *15th International Conference on Pattern Recognition*, 2000.

[18] M. Heikkilä and M. Pietäikinen, "A texture-based method for modeling the background and detecting moving objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28, pp. 657-662, No. 4, 2006.

[19] W. Zhang, X. Z. Fang, and Y. Xu, "Detection of moving cast shadows using image orthogonal transform," in *18th International Conference on Pattern Recognition*, 2006.

[20] T. Matsuyama, T. Ohya, and H. Habe, "Background subtraction for nonstationary scenes," in *Proceedings of Asian Conference on Computer Vision*, 2000.

[21] M. Mason and Z. Duric, "Using histograms to detect and track objects in color video," in *Proceedings of 30th Applied Imagery Pattern Recognition Workshop*, 2001.

[22] M. Heikkilä, M. Pietäikinen, and J. Heikkilä, "A texture-based method for detecting moving objects," in *Proceedings of British Machine Vision Conference*, 2004.

[23] Y.-T. Chen, C.-S. Chen, C.-R. Huang, and Y.-P. Hung, "Efficient hierarchical method for background subtraction," *Pattern Recognition,* vol. 40, pp. 2706-2715, No. 10, 2007.

[24] M. Seki, T. Wada, H. Fujiwara, and K. Sumi, "Background Subtraction based on Cooccurrence of Image Variations," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2003.

[25] N. Armanfard, M. Komeili, and E. Kabir, "TED: A Texture-Edge Descriptor based on LBP for pedestrian detection," *IEEE International Symposium on Telecommunications,* 2008.

[26] T. Ojala, M. Pietikainen, and T. Mäenpää, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, pp. 971-987, No. 7, 2002.

[27] F. Tajeripour and E. Kabir, "Defect Detection in Patterned Fabrics Using Modified Local Binary Patterns," in *IEEE International Conference on Computational Intelligence and Multimedia Applications*, 2007.

[28] T. Ojala, M. Pietäikinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition,* vol. 29, pp. 51-59, No. 1, 1996.

[29] M. Bertozzi, A. Broggi, C. Caraffi, M. D. Rose, M. Felisa, and G. Vezzoni, "Pedestrian detection by means of far-infrared stereo vision," *Computer Vision and Image Understanding,* vol. 106, pp. 194-204, No. 2-3, 2007.

[30] L. Malagón-Borja and O. Fuentes, "Object detection using image reconstruction with PCA," *Image and Vision Computing,* vol. 27, No. 1-2, 2009.

[31] http://www.cvg.rdg.ac.uk/slides/pets.html.

Narges Armanfard received her B.Sc. degree in Electronics Engineering from Shahid Chamran University, Ahvaz, Iran in 2006. She received her M.Sc degree in Electronics Engineering from Tarbiat Modarres University, Tehran, Iran, in 2008. Currently she is a Ph.D. student at Tarbiat Modarres University. Her research interests include Machine Vision, Pattern Recognition and Blind Source Separation.

Majid Komeili received his B.Sc. degree in Electronics Engineering from Iran University of Science and Technology, Tehran, in 2006. He received his M.Sc degree in Electronics Engineering from Tarbiat Modarres University, Tehran, Iran, in 2008. Currently he is a Ph.D. student at Tarbiat Modarres University. His research interests include Machine Vision, Pattern Recognition and video surveillance.

Ehsanollah Kabir is a professor of electrical engineering at Tarbiat Modarres University. He obtained his B.Sc. and M.Sc. degrees in electrical and electronics engineering from the University of Tehran. He received his Ph.D. degree in 1990 from the University of Essex, where he worked on the recognition of handwritten postal addresses. His main areas of research are document image analysis and handwriting recognition.