Optimized motion detection method based on modified three-frames and stationary wavelets

Oussama Boufares [1*], Noureddine Aloui [1], Adnen Cherif [1]
[1] Laboratory of Analysis and Processing of Electrical and Energy Systems (ATSEE), Department of physics, Faculty of Sciences Tunis, University of Tunis El Manar Tunis, Tunisia

Abstract:
In visual surveillance, the object detection is a fundamental step for further processing such as segmentation, tracking, and extraction of a scene's contextual information. A lot of approaches of motion detection have been proposed in the literature. In this paper, we have proposed a new method which is more efficient and computationally inexpensive compared to the others ones to detect moving objects in video sequence. The proposed algorithm begins with computation of difference temporal using difference between three frames successive. And then, we propose to apply to fixed number of alternate frames centralized around the actual frames instead of making difference images using the traditional approach. The new approach helps to reduce the computational complexity without reducing quality of the obtained images. After calculating the modified three-frame difference, the obtained results are decomposed using discrete stationary wavelet transform 2D and the coefficients are thresholded using Birge-Massart strategy in order to extract the foreground. The evaluation tests show that the proposed approach reaches the better performance of detection than the other approaches.
Keywords
moving objects detection, modified temporal differencing, SWT, three-frame difference.

1. Introduction

The process of change detection is one of the most essential steps in computer vision and in video surveillance, so it attracts a huge attention from researchers in the video processing's domain. Detection step can be as well applied to image compression [1], environment monitoring [2], exception alert[3] virtual presence[4] etc. It gives the ability to extract the target from a background of sequences in order to proceed with follow-up procedures for surveillance. Generally there are three fundamental factors that make it more difficult to detect moving objects in scene. The presence of complex background is the first factor; the second is the camera motion and thirdly the requirement of prior knowledgement. Furthermore, most existing moving object detection algorithms are not intelligent or not robust enough for what they are applied. Combining the modified temporal differencing and the stationary wavelet transform is the solution proposed to deal with these challenges.

Generally, motion detection methods utilizing temporal information can be classified into two major approaches, temporal frame differencing methods [5, 6] and background subtraction methods [7]. The background subtraction methods find out the moving objects by computing the absolute difference between current frame and the background model. This absolute difference is used to calculate the binary moving objects with the object threshold function. Additionally, the generated background model might not be applied in some scenes with some specified challenges, but are not limited to the conditions: 1) Adaptability to illumination change, 2) dynamic textures adaptation, 3) Noise tolerance 4) Sensitivity to clutter motion, 5) Bootstrapping and 6) Convenient implementation.

Many algorithms in this area have been developed for background subtraction —background modeling-based method [8, 9, 10] and filter background estimation-based method [11, 12]. As mentioned in [11, 12], the algorithms are not robust to dynamic backgrounds. The statistical background modeling is generally computationally expensive and cannot manage fast dynamic backgrounds as presented above [13, 14]. Jodoin et al. [14] utilizes global mixture of Gaussians for background modeling which results in a reduction in detection precision for an augmentation in the number of moving objects in foreground. Baf et al. [15] proposed a background modeling based a Fuzzy which is more robust and adapts perfectly to changes in the scene. Baocai et al. [16] proposed a non-parametric background of dynamic scenes based on dual model. One of them is called the self-model and the other one is called the neighbourhood-model.

Generally, temporal difference approaches select two adjacent frames I_t and I_{t-1}, or select two frames separated by an interval, and then difference these two selected frames to detect if the pixel of result image has been modified. I_t and I_{t-1} respectively show the current and the previous frame at time t and t-1. The output frame is presented by:

Manuscript received December 5, 2017
Manuscript revised December 20, 2017

$$ob_t = \begin{cases} 1, |I_t - I_{t-1}| > \tau & \text{change} \\ 0, |I_t - I_{t-1}| < \tau & \text{non change.} \end{cases} \quad (1)$$

The advantage of this method is rapid in adapting illumination changes or moving camera. Furthermore, the moving targets that appear suddenly do not leave behind ghosts. However, simple frame differencing can only detect the trailing and leading edge of a uniform colored object. Additionally, detecting an object moving towards or away from the camera becomes difficult. According to algorithms of motion detection using frame difference method, this algorithm calculates difference between previous frame and current frame at time t-1 and t and current frame and next frame at time t and t+1, after that are combined by a logical AND operator. On the other hand, this approach can only be utilized to precise motion detection if enough separation in location of the foreground objects exists between the consecutive frames. Collins et al. [17] integrate a novel motion detection approach based on three-frame differencing methods. This algorithm solves the ghosting problem, which is presented as the incorrect detection of the foreground objects when none exists. . Lei and Gong [18] proposed a robust and efficient method of detection based on three-frame difference, Gaussian mixture model in order to solve the problem that can't detect the entirety of the moving objects in three frame difference.

In this paper, we propose an algorithm based an improved three frame difference approach with the help of non-overlapping blocks. In addition, it is combined with local adaptive thresholding using Stationary Wavelet Transform (SWT) so that an incorporated MDI-SWT algorithm is realized to perfectly detect moving objects. The suggested algorithm gives better calculation efficiency and speed as compared to traditional temporal difference approaches. In addition, the proposed algorithm has been applied to higher accuracy in detecting moving objects in sudden illumination changes and bad weather conditions.

The proposed contributions are organized as follows: Section. II presents the strategy of the proposed motion detection. In this step, the first difference image generated by calculating the result of the difference between modified three-frame, and then decomposed using SWT and the coefficients are thresholded using Birge-Massart strategy. Section III presents a detailed experimental test of proposed approach qualitatively as well as quantitatively. Finally, Section VI concludes this paper with a discussion and some directions of our future work.

2. Proposed algorithm using modified temporal difference image for moving object detection

2.1 Modified temporal difference image

2.1.1 Three-frame differencing generation:

Very recently, Ding and Gong [18] have proposed a three-frame difference method, to determine the foreground challenges in inter-frame difference. Inspired by this approach, the difference image is provided using the frame difference method.

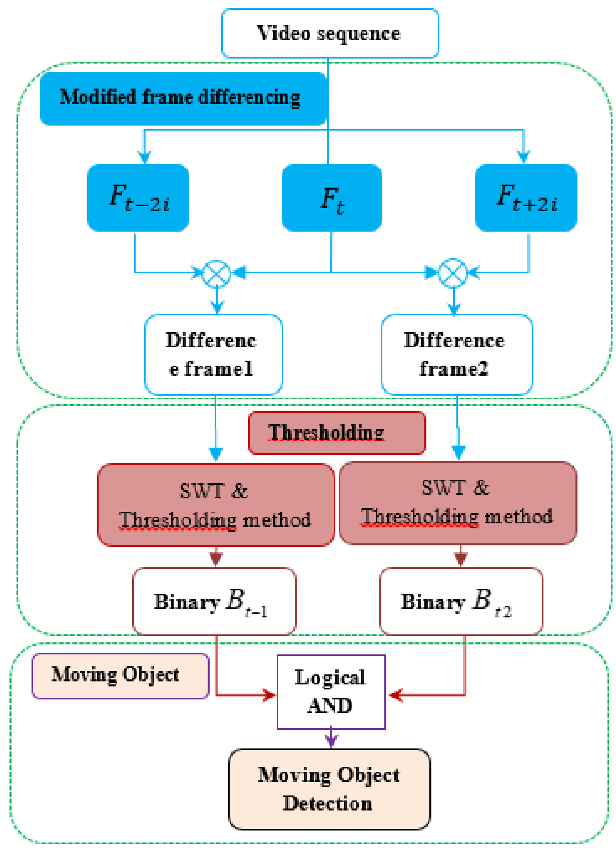

Fig.1 Block diagram of the proposed algorithm

For an image sequence, we can construct a frame set $[..., I_{t-n}(x,y), ..., I_{t-1}(x,y), I_t(x,y), I_{t+1}(x,y), ..., I_{t+n}(x,y)...]$

Where I_t presents a frame in a video scene at time t, and both adjacent frames have the congruent scope. Therefore, I_{t-1} and I_{t+1} constitute the first adjacency pair, I_{t-2} and I_{t+2} constitute the second adjacency pair and so on.

A target appearing in image I_t can be defined as the intersection of the change areas that occur between two consecutive frames I_t and I_{t-1} and between frames I_t and I_{t+1}. The result frame is obtained by computing the Euclidean distance between frames I_t and $I_{t\pm}$ of the corresponding pixels in RGB color space, illustrated as follows:

$$\mathrm{ob} = \sqrt{\sum_{d=R,G,B}\left(I_{t,d}(x) - I_{t+i,d}(x)\right)^2} \cap \sqrt{\sum_{d=R,G,B}\left(I_{t,d}(x) - I_{t-i,d}(x)\right)^2}, \quad (2)$$

where I_t (x) is the intensity value of pixel x of frame at time t, ob is the moving targets at time t, $I_{t,cl}$ is the cl-color channel of frame at time t. The scheme of three-frame differencing is given in fig.1.

The intersection operation between two consecutive frames over a period of time is defined as the corresponding point to point multiplication of the pixel intensity values across the frames. The summation process of pixel values in the image I_t gives the foreground objects. This thereby, raise up the intensity values of the moving objects that have changed position as opposed to the relatively static background, which still with low intensity. A unique frame difference does not provide a well representation of the moving object as long as pixel intensity values are very low. By repeating this difference operation inter-frames over an ensemble of 'N' frames, the intensity of the foreground object is rised, moreover the low intensity of the static background is conserved in the difference frame. The result of each difference operation is resumed to obtain the difference frame as given by:

$$\mathrm{ob} = \sum_{i=1}^{N}\left(\sqrt{\sum_{d=R,G,B}\left(I_{t,d}(x) - I_{t+i,d}(x)\right)^2} \cap \sqrt{\sum_{d=R,G,B}\left(I_{t,d}(x) - I_{t-i,d}(x)\right)^2}\right). \quad (3)$$

2.1.2 Modified difference three-frame:

While this approach of three-frame difference does provide good results, there is redundancy observed in the selection of adjacency pairs. To enhance the detection results, we suggest using N/2 pairs instead of selecting 'N' adjacency pairs that are selected alternately. This is shown in the following equation:

$$\mathrm{ob} = \sum_{i=1}^{N/2}\left(\sqrt{\sum_{d=R,G,B}\left(I_{t,d}(x) - I_{t+2i,d}(x)\right)^2} \cap \sqrt{\sum_{d=R,G,B}\left(I_{t,d}(x) - I_{t-2i,d}(x)\right)^2}\right) \quad (4)$$

For modification of proposed algorithm, instead of selecting eight images around the It image up to images I_{t+8} and I_{t-8}, we select four images, I_{t+8} and I_{t-8} ,, with a one-frame gap inter two successive images. The motivation behind this approach is the reduction of computational overhead, which in this case, is reduced by 50 %. The modification of proposed diagram has been shown in Fig. 2.

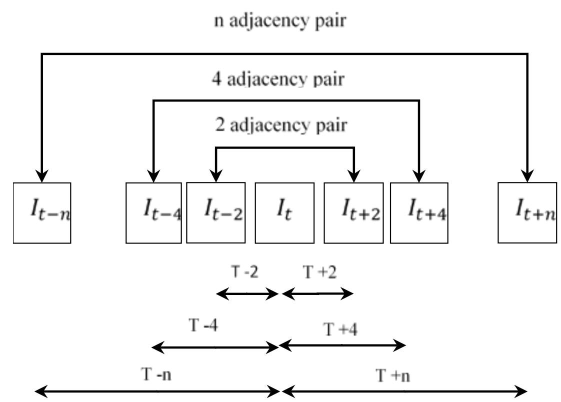

Fig.2 Proposed diagram for adjacency pairs

2.2 Adaptive thresholding for binarization of final difference frame

With previous study [19], the MDI-SWT approach starts with modified difference temporal method. This latter is based on a difference between three frames successive instead of background subtraction. In this section, the thresholding step is based on stationary wavelet transform and Birge-Massart strategy. The divers steps of frame analysis are explained in the following:

Step1: Input final difference.
Step2: Decompose the frame exploiting SWT.
Step3: Thresholding using universal threshold.
Step4: Reconstruction using ISWT.

2.2.1 Decomposition of three difference frame:

The Stationary wavelet transform (SWT) is similar to the DWT except the signal is never subsampled in SWT but the filters are upsampled at each level of decomposition, instead. The SWT generates four subbands at each level of decomposition, called HH, LL, LH and HL. This SWT is the simplest way to decompose a frame. It necessitates decimation by a factor 2N, where N stands for the level, of the transformed signal at each stage of the decomposition. .Our difference continuous frames are decomposed into LL2-band frames by using SWT. Figure 4 shows the decomposition of SWT.

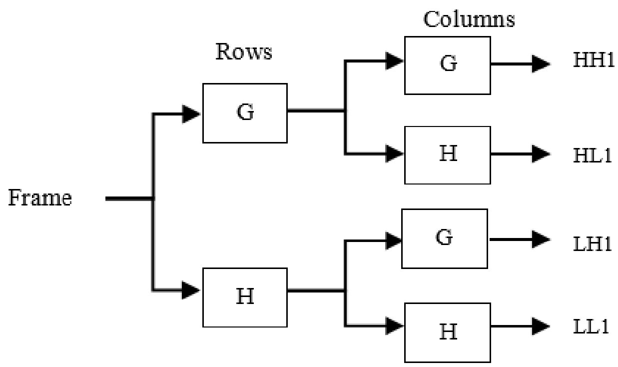

Fig.3 Diagrams of SWT frame decomposition: G: high-pass filter, H: low-pass filter

2.2.2 Hard thresholding technique:

After decomposition of final difference frame, the obtained coefficients subbands are adaptively thresholded to reduce the preserved noise in high frequency. It consists to reject the subbands coefficients inferior to a given threshold. In this paper we have used the hard threshold given in this equation:

$$diff_{(x,y)} = \begin{cases} 1, & for \quad diff_{(x,y)} > T \\ 0, & for \quad others. \end{cases} \quad (5)$$

The universal threshold (Visu Shrink) $\boldsymbol{T = \sigma\sqrt{2ln(N)}}$ has been proposed by Donoho[20]. $\sigma = median(\mathrm{LL})$ is the noise variance and N is the number of the wavelet coefficients in that frame, with an absolute median deviation (MAD) converged to 0.6745 times σ as the sample size.

2.2.3 Reconstruction difference frame (Binarization):

In this part, the result image is the difference between (Ft-2i, Ft, and Ft+2i, Ft) after the steps of decomposition and thresholding. This later is converted to binary image using formula (6). This conversion is done by taking threshold value. If pixel intensity is above the threshold value then the pixel value becomes 1 or white and if the pixel value is less than the threshold value then the pixel is converted into black or 0. Formulas (6) and (7) represent how to obtain a binary image by subtraction between the three images. The reconstruction frame using ISWT have been shown in Fig. 4.

$$\mathrm{B}_{t-1} = \begin{cases} 1, \text{ if } \sum_{i=1}^{N/2}\left(\sqrt{\sum_{d=R,G,B}\left(\mathrm{I}_{t,d}(x) - \mathrm{I}_{t-2i,d}(x)\right)^2}\right) > \mathrm{T} \\ 0, \text{ if otherwise.} \end{cases} \quad (6)$$

$$B_t = \begin{cases} 1, \text{ if } \sum_{i=1}^{N/2}\left(\sqrt{\sum_{d=R,G,B}\left(\mathrm{I}_{t,d}(x) - \mathrm{I}_{t+2i,d}(x)\right)^2}\right) > \mathrm{T} \\ 0, \text{ if otherwise.} \end{cases} \quad (7)$$

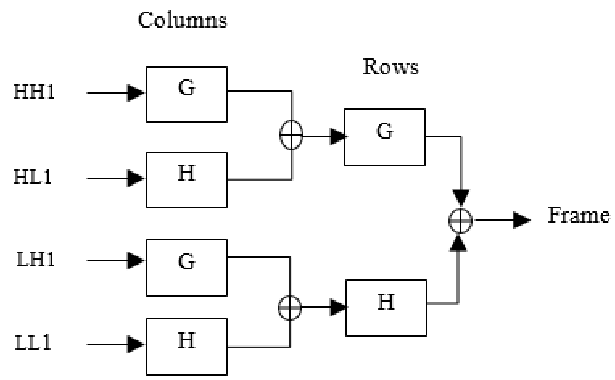

Fig.4 Diagrams of ISWT frame Reconstruction

2.3 Object detection

In this stage, after obtaining the binary mask Bt-1 and Bt from these modified three successive frames and a threshold value T, the moving object detection (MMt) can be generated using the intersection operation (logical AND) between Bt-1 and Bt. The detection result presents the foreground in white with a black background. The function is represented as follows:

$$MMt = B_{t-1} \cap B_t \quad (8)$$

The resulted frames of AND operation are further processed by utilizing morphological filter in order to remove the noise of the thresholded image to detect the moving objects correctly and quickly of a video sequences.

3. Experimental result and analysis

We have implemented this proposed method in on a computer running 4GB RAM, Intel core i3 processor of 2.4GHz frequency and Windows 10 operating system, utilize MATLAB software (version R2013a). Some of the common challenges often faced during object detection are: contains of repeated motions in the background, a scenes with sudden and varying illumination and indoor environments with shadows. The tested different video sequences are namely 'intelligent room', 'traffic', 'hall monitor', 'laboratory', 'pts2006' , 'Bungalow ', 'Pedestrians', ' Highway', ' Trees' , ' Skating ' and ' Snow'. To test the performance of the proposed approach, the results obtained by it are evaluated by comparing some images with the ground truth frame. All video sequences utilized for the test study have been taken by a public database. Table 1 presents auxiliary information of each video utilized in our experimental analysis.

The rest of this part is organized as follow: Section III.1 provides a qualitative study of proposed algorithm, showing the obtained results using our method under various test conditions. Section III.2 introduces a comprehensive comparison with some existing schemes both qualitatively as well as quantitatively between our algorithm and other motion detection methods, including by Mixture of Gaussians (MOG) [14], Block-Based Subtraction (BBS) [21] and Multiple Difference Images (MTI) [22] methods. In section III.3, quantitative analysis is applied to measure the accuracy using various detection parameters including Recall, Precision, F1 and Similarity. Finally, the time consumption of each method is reported in terms of performance evaluation.

Table 1: INFORMATION ABOUT VIDEOS USED FOR EXPERIMENTAL

Video	Category	Video source	Frames rate (fps)	Number of frames used
Trees (TR)	Dynamic background	perception.i2r	24	1400
PETS2006 (PT)	baseline	changedetection.net	24	310
Bungalow (BN)	shadows	changedetection.net	30	350
Pedestrians (PD)	baseline	changedetection.net	30	310
Highway (HG)	baseline	changedetection.net	24	350
Snow (SN)	Bad weather	changedetection.net	30	100
Wetsnow (WS)	Bad weather	changedetection.net	30	100
Skating (SK)	Bad weather	changedetection.net	30	250
Laboratory (LB)	shadows	visor	10	887
Intelligent room (IR)	shadows	visor	10	300
Traffic (TF)	Illumination change	MATLAB	25	300

3.1 Qualitative study

The proposed detection algorithm was tested on different scenes. Some results which are depicted in figure 5 and figure 6 have been compared with ground truths. The moving object is identified a yellow rectangle. Fig.5 PD, Fig.5 HW and Fig.6 HW present baseline sequences, (a) shows the original framers, (b) presents the detection results of foreground object and (c) is the ground truths. From the results, proposed method is very close to their corresponding ground truth. Fig. 6 BN represents video with shadow area while Fig. 6 SW presents challenging weather. The detection results of the foreground object shown in row (c) and ground truths shown in row (b). In these two sequences, we can see that our approach can detect the foreground with high accuracy and robustness.

3.2 Comparative study

Fig. 7 presents the results of the foreground object under some typical challenging surroundings: TR: Dynamic background, PT: baseline, IR: intelligent room and SN: snow. The detection results of the proposed method are shown in column E. We also compared our approach with binary mask of moving objects obtained by the (MOG), (BBS) and (MTI) methods. The first three rows show the results of two test sequences including ‘Intelligent room and Pets2006’. It may be seen that the proposed algorithm can detect the moving objects and eliminate shadow and noise in successive frames almost perfectly. Trees video is a scene with a dynamic background in the form of flowing branches and leaves. Results show that the proposed algorithm can easily eliminate the dynamic motions in background. The snow sequence contains small outliers; the foreground object has been detected almost perfectly.

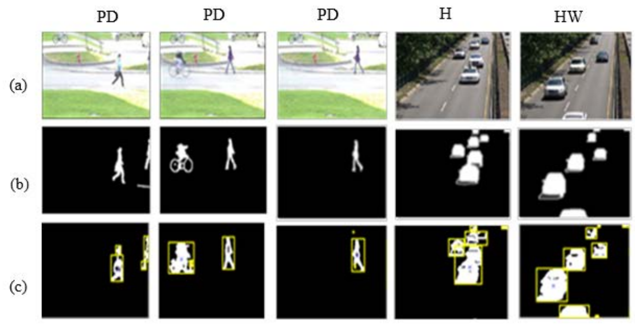

Fig.5 Algorithm detection results of sequence with pedestrians (PD) and for sequence with highway (HW)

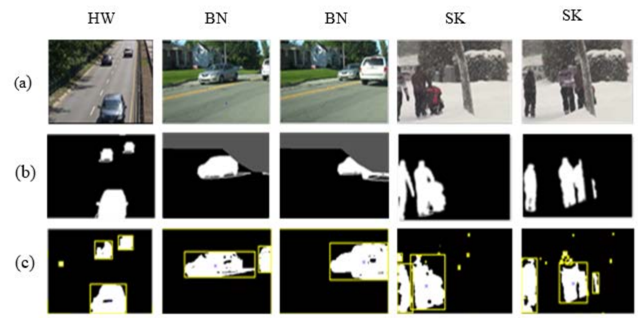

Fig.6 Algorithm detection results of sequence with highway (HW), bungalows (BN) and skating (SK)

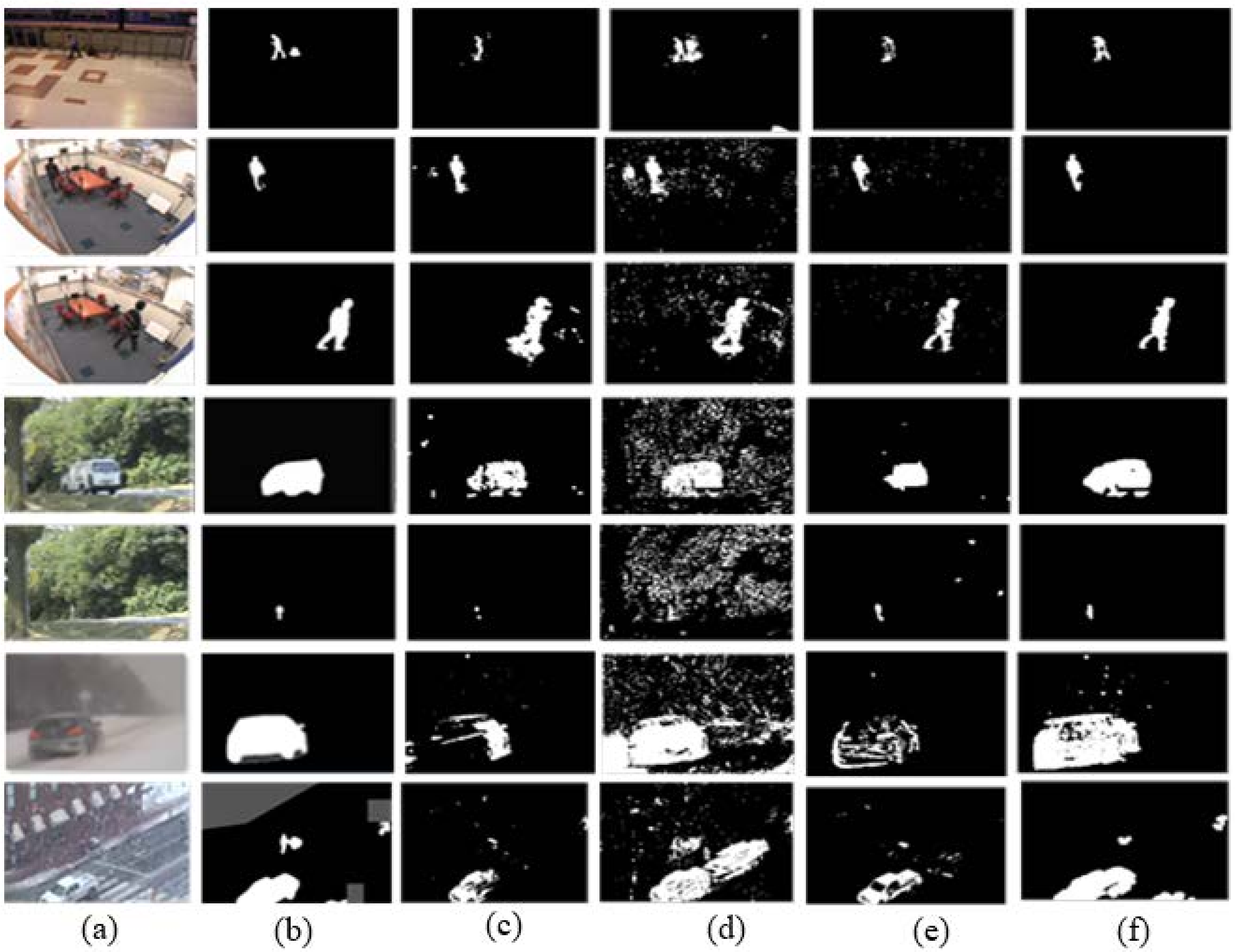

Fig.7 Extracted foreground from qualitative comparison using proposed approach and all approaches for different videos: a. Original frames, b. Ground truths, c. MOG, d. BBS, e. MTI and f. proposed approach

Fig.8 Algorithm moving object detection results of (1) scene with illumination change, (2), (3) and (4) three indoor scenes with shadows

In figure 8 shows the performance and efficiency of detecting moving objects of the proposed method on four other sequences with the various challenges discussed above.

3.3 Quantitative study

To assess the detection results quantitatively of the proposed algorithm, we used out the experiments on four sequence videos called pedestrians, trees. The performance of proposed method is evaluated in terms of recall, precision, similarity, and f-measure. Among these four, Tp represents the total number of the true positive pixels, Fp depicts the number of the false positives and Fn indicates the number of the false negatives. The four indexes are defined as:

$$Recall = \frac{Tp}{Tp+Fn} \quad (9)\ ;$$

$$Precision = \frac{Tp}{Tp+Fp} \quad (10)\ ;$$

$$\text{F-measure} = 2*\frac{\text{Recall*Precision}}{\text{Recall + Precision}} \quad (11)\ ;$$

$$\text{Similarity} = \frac{\text{Tp}}{\text{Tp + Fp + Fn}} \quad (12)$$

Table 2 highlights the robustness and efficacy of the average detection results on the four datasets, which illustrates numerically that the proposed algorithm outperforms the other three methods. Table 4 shows the comparison of the computation self-time. The computation time per frame is obtained by averaging the processing time of different videos. Our method takes less time than other BSS and MTI methods, and as near as time in MOG method.

Table 2: AVERAGE PRECISION, RECALL AND SIMILARITY FOR DATA SETS ON PTS2006, INTELLIGENT ROOM, TREES AND HIGHWAY

Methods	**Recall**	**Precision**	**F-measure**	**Similarity**
MOG	0,7174	0,7519	0,7275	0,5476
BBS	0,5627	0,4201	0,3953	0,2343
MDI	0,7014	0,6915	0,6826	0,5127
MDI-SWT	0,7425	0,8331	0,7798	0,6464

Table 3: DETECTION RESULTS ON PTS2006, INTELLIGENT ROOM, TREES AND HIGHWAY

Video	Recall	Precision	F-measure	Similarity
Trees (TR)	0.6531	0.8342	0.7319	0.5862
PETS2006 (PT)	0.7363	0.9216	0.8115	0.6835
Pedestrians (PD)	0.8285	0.7603	0.7929	0.6480
Intelligent room (IR)	0.7522	0.8164	0.7829	0.6679
Average	0.7425	0.8331	0.7798	0.6464

Table 4: COMPUTATIONAL COMPARISON RESULTS IN TERMS OF EXECUTION TIME

Methods	Intelligent room (IR)	Trees	Highway	Average
MOG	1.937	3.369	2.656	2.654
BSS	3.337	4.647	4.518	4.167
MTD	1.593	3.988	3.462	3.014
MDI-SWT	2.267	3.594	2.811	2.890

4. conclusion

We propose a novel robust algorithm for motion detection using modified temporal differencing. We solely utilize the three frame difference to generate frame difference along with adaptive thresholding based on SWT and Birge-Massart methodology to efficiently extract the detect moving object. Our approach has been validated on different public datasets. The analysis results demonstrate that the proposed approach can detect moving objects with high precision and robustness. Compared with other methods, we have verified that the proposed method has reached the most satisfactory results based on qualitative evaluation and quantitative measurement. Further research focuses on how to deal with the occlusion problems between the consecutive moving objects.

References

[1] M. C. Liu, I. Tsukagoshi, and M. A. Kutner: 'Real-time MPEG video encoder with embedded scene change

detection and telecine inverse' Proc Intl. Conf. Consumer Electron. 2002 ICCE, pp. 246–247 ~2002!.
[2] M. Brand and V. Kettnaker: 'Discovery and segmentation of activities in video' IEEE Trans. Pattern Anal. Mach. Intell. 22~8!, 844–851 ~2000!.
[3] I. Pavlidis, V. Morellas, P. Tsiamyrtzis, and S. Harp: 'Urban surveillance systems: from the laboratory to the commercial world' Proc. IEEE 89~10!, 1478–1497 ~2001!.
[4] Y. Raja, S. T. McKenna, and S. Gong: 'Tracking and segmenting people in varying lighting conditions using colour' Proc. Third IEEE Intl. Conf. Auto. Face Gesture Recog., pp. 228–233 ~1998!.
[5]
[6] D. J. Dailey and L. Li: 'An algorithm to estimate vehicle speed using un-calibrated cameras, intelligent transportation systems' Proc. 1999 IEEE/IEEJ/JSAI Intl. Conf., pp. 441–446 ~1999!.
[7] M. Tekalp: 'Digital Video Processing' Prentice-Hall, Englewood Cliffs, NJ ~1995!.
[8] D.-M. Tsai and S.-C. Lai: 'Independent component analysis-based background subtraction for indoor surveillance' IEEE Trans. Image Process., Jan. 2009, vol. 18, no. 1, pp. 158–167.
[9] N.J.B. McFarlane and C.P. Schofield S: 'Segmentation and tracking of piglets in images', Machine Vision and Applications 1995, vol. 8, pp.187–193.
[10] Zheng. J and Wang. Y: 'Extracting roadway background image', a mode based approach. J Transp Res Board.,2006, 1944:82–88.
[11] Wei, L., Hongfei, Y., Huai, Y., Hong, Z., Xiaowei, X. :' Effective background modelling and subtraction approach for moving object detection'. IET Computer Vision, 2015, Vol. 9, Issue: 1, 2 ,pp 13-24.
[12] Toyama. K, Krumm. J: 'Wallflower: principles and practice of background maintenance'. In: Proc. 7th. IEEE Int. Conf. on Computer Vision, 1999, pp 255–261
[13] Messelodi. S and Modena. C: ' A Kalman filter based background updating algorithm robust to sharp illumination changes'. Lect Notes Comput Sci, 2005, 3617:163–170
[14] Wren. C and Azarbayejani. A: ' Pfinder: Real-time tracking of the human body'. IEEE Trans Pattern Anal Mach Intell, 1997, 19(7):780–785
[15] Jodoin. PM et al: 'Statistical background subtraction using spatial cues'. IEEE Trans Circuits Syst Video Technol, 2007, 17(12) pp 1758–1763.
[16] Baf. FE, Bouwmans. T and Vachon. B: ' Fuzzy integral for moving object detection'. In: IEEE Int. Conf. on Fuzzy Systems, 2008, pp 1729–1736.
[17] Baocai. Y, Jing. Z and Zengfu. W: 'Background segmentation of dynamic scenes based on dual model'. IET Computer Vision, 2014 ,Vol. 8, Issue: 6, pp 545 - 555.
[18] Collins. R, Lipton. A et al:' A system for video surveillance and monitoring'. Techrep. Carnegie Mellon University, Pittsburg, pp1-15
[19] D. Lei, and N. Gong :'Moving objects detection based on improved three frame difference'. Video Engineering, 2013, vol. 37, issue I, pp. 151- 153.
[20] Oussama. B, Noureddine. and adnene. C :' Adaptive Threshold for Background Subtraction in Moving Object Detection using Stationary Wavelet Transforms 2D'. International Journal of Advanced Computer Science and Applications, 2016, Vol. 7, Issue 8, pp 29-36.
[21] 17. D. L. Donoho and I. M. Johnstone, :' Denoising by soft thresholding'. IEEE Trans. on Inform. Theory, 1995, Vol. 41, pp. 613-627.
[22] Matsuyame T, Ohya T, Habe H (2000) :' Background subtraction for non-stationary scene'. In: Proc. Asian Conf. on Computer Vision, pp 622–667.
[23] J.-E. Ha and W.-H. Lee:'Foreground objects detection using multiple difference images'. Opt. Eng., Apr.2010,vol. 49, no. 4, p. 047 201.