Motion Segmentation Based on Structure-Texture Decomposition and Improved Three Frame Differencing

Sandeep Singh Sengar(✉)

Department of Computer Science and Engineering,
SRM University-AP, Amaravati 522 502, India
sandeepsingh.s@srmap.edu.in

Abstract. Motion segmentation from the video datasets has several important applications like traffic monitoring, action recognition, visual object tracking, and video surveillance. The proposed technique combines the structure-texture decomposition and the improved three frames differencing for motion segmentation. First, the Osher and Vese approach is employed to decompose the video frame into two components, viz., structure and texture/noise. Now, to eliminate the noise, only the structure components are employed for further steps. Subsequently, the difference between (i) the current frame and the previous frame as well as (ii) the current frame and the next frame are estimated. Next, both the difference frames are combined using pixel-wise maximum operation. Each combined difference frame is then partitioned into non-overlapping blocks, and the intensity sum as well as median of each block is computed. Successively, target objects are detected with the help of threshold and intensity median. Finally, post-processing in the form of morphology operation and connected component analysis is carried out to accurately find the foreground. Our technique has been formulated, implemented and tested on publicly available standard benchmark datasets and it is proved from performance analysis that our technique exhibit efficient outcomes than existing approaches.

Keywords: Structure-texture decomposition · Motion segmentation · Frame difference · Morphology · Block

1 Introduction

Motion segmentation is one of the basic and critical approaches of computer vision system. Motion segmentation technique is motivated by the assumption that foreground objects commonly involve changes in the intensity between successive frames. This approach is mandatory in a real environment and should be

SRM University-AP, Amaravati.

Published by Springer Nature Switzerland AG 2019
J. MacIntyre et al. (Eds.): AIAI 2019, IFIP AICT 559, pp. 609–622, 2019.
https://doi.org/10.1007/978-3-030-19823-7_51

applicable to difficult circumstances, like background clutter, fake motion, and changes due to noise and illumination deviation. All of these situations lead to the false region detection for the moving targets.

Currently, background subtraction [1], optical flow [2] and frame differencing [3] are commonly used techniques for motion segmentation. The background subtraction technique depends on the accuracy of generated background; if the background is not accurate then there will be large segmentation error. The optical flow technique is complex and there are high amount of noise in detected result due to involvement of differential operations in it. On the other side, frame differencing is a simple and effective technique to segmenting the moving objects by calculating the pixel-wise difference between successive frames. However, there are two most common problems in two frames differencing viz., (i) background clutter (ii) foreground aperture. To reduce these problems, Sengar et al. [4] proposed an enhancement of frame differencing technique. This is a straightforward method, which detects the moving objects by employing block based three frame differencing. However, detection accuracy of this method is low for some frames of video data due to the randomly generated noise. To improve the detection accuracy, Sengar et al. [5] proposed another approach for motion segmentation using block based optical flow. However, processing time of this work is high due to the local processing. In order to improve the detection accuracy, reduce the false positives and processing time, the structure-texture decomposition and three frame differencing based motion detection approach is proposed. The proposed technique first extracts the frames from the video datasets, followed by color to gray-scale processing. Next, the structure component of the gray-scale video frame is extracted using structure-texture decomposition technique so as to remove the texture/noise. Subsequently, block based improved three frame differencing algorithm is employed on extracted structure component for motion segmentation. Finally, post-processing steps using mathematical morphology operation and connected component analysis are performed to accurately detect the foreground. Our method is simple and has considerably low computational cost. We have tested the proposed method on different publically available benchmark video datasets and also compared with existing techniques. Our approach has higher detection accuracy than widely used recent techniques.

As discussed earlier, there are certain problems available in existing motion segmentation techniques. Here, in this approach we have reduced those. It is an important technique for visual surveillance, traffic monitoring and other several related applications. This scheme can be the initial step for visual object tracking, where future position of target can be predicted after detecting it in previous video frames.

The organization of remaining part of this paper is as follows. The survey on related methodologies are presented in Sect. 2. The proposed motion segmentation method and it's implementation details are provided in Sects. 3 and 4 respectively. Experimental results with the qualitative as well as quantitative evaluations are shown in Sect. 5. Finally, conclusion of our technique is provided in Sect. 6.

2 Related Work

Several schemes for motion segmentation have been presented by image processing and computer vision groups to deal with challenges like ghosting, illumination variations, object pose variation, occlusion and out-of-plane rotation [6–8]. One of the methods for textured image segmentation has been presented by Sandberg et al. [9]. This scheme has employed active contour model within Gabor filter environment. Elharrouss et al. [1] proposed an approach for motion detection using the structure-texture decomposition and the background subtraction. Block-wise low rank texture characterization is employed for cartoon-texture image decomposition model [10]. This model works like a convex optimization problem. Surface normals and structure-texture decomposition are used for intrinsic image decomposition [11]. High quality outcomes are produced by this approach with less number of constraints. Wells et al. [12] presented a work for image decomposition in the context of classification and adaptive segmentation with the help of different statistical techniques. Malgouyres [13,14], Candès et al. [15] have employed total variation minimization framework in a wavelet domain for restoration of textured images. Casadei et al. [16] and Zhu et al. [17] have presented a texture modeling with the help of statistical techniques.

Halidou et al. [18] proposed an approach for pedestrian detection by employing multi-block local binary pattern descriptors and region of interest (ROI). Here, three frame differencing and optical flow techniques are used to compute ROI. Neural network based classification approach and frame differencing technique are employed to detect moving objects in the indoor environments [19]. In this technique, robust classification is achieved using finite state automation. Caballero et al. [20] presented a technique for human detection using frame differencing and optical flow. This technique is proposed for an infrared camera mounted on a mobile robot. The non-pyramidal based Lucas-Kanade and the frame differencing techniques are employed to detect target objects by using thermal signatures [21]. A weighted mixture of Gaussians based adaptive background method is proposed for real time tracking [22]. However, this technique is not good for non-static backgrounds and sudden illumination changes. The W4 and histogram based frame differencing approaches are used to detect the moving objects [23]. Maddalena et al. [24] presented a self-organizing approach for background subtraction. In this approach, the background of the frame is learned with the help of labeling decision and neural network at neighboring pixels. Principal component analysis is used to projecting the high-dimensional data into a lower dimensional subspace for the background modeling [25]. This technique is suitable for global illumination variations, but slow moving objects in local illumination variation environment cannot be detected accurately. Sengar et al. [26] proposed two methods, viz., (i) improved three frame differencing and (ii) combined improved three frame differencing and background subtraction to accurately detect the slow/fast moving objects of varying number and size in the illumination variations and background clutter environments. Quaternion is used with modified optical flow technique to increase the accuracy of information of unclear pixels [27]. Schwarz et al. [28] employed the depth information

obtained by a time of flight camera to compute the location of the foreground object. Reduced processing time and accurate results are obtained from the above techniques [27,28], but these are not suitable for real time environments. Sengar et al. [29] used the normalized self adaptive optical flow technique for moving object detection. Here, self adaptive window approach and Otsu's [30,31] method are used to select the target object area and to adapt the threshold value respectively. This method accurately detects the foreground, but not suitable for moving objects of small size.

Algorithm 1. Algorithm of Osher-Vese for structure-texture decomposition [32]

Input : Video frame F, No_of_iteration, Tuning parameter $\lambda > 0$, Step space $h > 0$

1. $U \leftarrow F$, function $f_1 = -\frac{F_x}{2\lambda|\nabla V_f|}$, function $f_2 = -\frac{F_y}{2\lambda|\nabla V_f|}$
2. **for** *t=1 to No_of_iteration* **do**
   for *i=2 to w-1* **do**
   for *j=2 to h-1* **do**

$$K_1 \leftarrow \frac{1}{\sqrt{\left(\frac{U(i+1,j)-U(i,j)}{h}\right)^2+\left(\frac{U(i,j+1)-U(i,j-1)}{2h}\right)^2}}$$

$$K_2 \leftarrow \frac{1}{\sqrt{\left(\frac{U(i,j)-U(i-1,j)}{h}\right)^2+\left(\frac{U(i-1,j+1)-U(i-1,j-1)}{2h}\right)^2}}$$

$$K_3 \leftarrow \frac{1}{\sqrt{\left(\frac{U(i+1,j)-U(i-1,j)}{2h}\right)^2+\left(\frac{U(i,j+1)-U(i,j)}{h}\right)^2}}$$

$$K_4 \leftarrow \frac{1}{\sqrt{\left(\frac{U(i+1,j-1)-U(i-1,j-1)}{2h}\right)^2+\left(\frac{U(i,j)-U(i,j-1)}{h}\right)^2}}$$

$$U(i,j) \leftarrow \left(\frac{1}{1+\frac{1}{2\lambda h^2}(K_1+K_2+K_3+K_4)}\right) \times \left[F(i,j) - \frac{f_1(i+1,j)-f_1(i-1,j)}{2h} - \frac{f_2(i,j+1)-f_2(i,j-1)}{2h} + \frac{1}{2\lambda h^2}\big(K_1U(i+1,j) + K_2U(i-1,j) + K_3U(i,j+1) + K_4U(i,j-1)\big)\right]$$

   end
   end
3. $V = F - U$
   end

Output: Structure component (U) and texture/noise component (V)

3 Proposed Method

The steps for the proposed motion segmentation technique are as follows:

- Extraction of frames from the video data.
- Conversion of video frame to gray-scale, if it is required.
- Separation of structure and texture/noise components.
- Application of three-frame differencing techniques on the structure components. Two difference frames are obtained.

- Computation of pixel-wise maximum value between estimated difference frames.
- Division of computed result of previous step in non-overlapping blocks.
- Calculation of intensity sum and intensity median of each block.
- Separation of background and foreground using threshold and block's median.
- Application of post processing steps to accurately detect the foreground objects.

The steps for the proposed technique is explained below:

3.1 Structure-Texture Image Decomposition

A set of information in the form of structure, texture, and noise are contained by an image and it can be extracted using structure-texture image decomposition method. The inverse estimation problem i.e. structure-texture decomposition approach is mandatory for analyzing and understanding images. This method split a given video frame or image (F) into a structure (bounded variation) component U and a texture and/or noise (oscillating) components V i.e. $F = U + V$. Many models have been presented for image decomposition, in which one method is proposed by Osher and Vese [32]. This method has shown its efficiency by preserving discontinuities and removing oscillations in the image. The algorithmic steps for Osher-Vese [32] method are shown in Algorithm 1. Here in the algorithm, w and h denote the width and height of video frame.

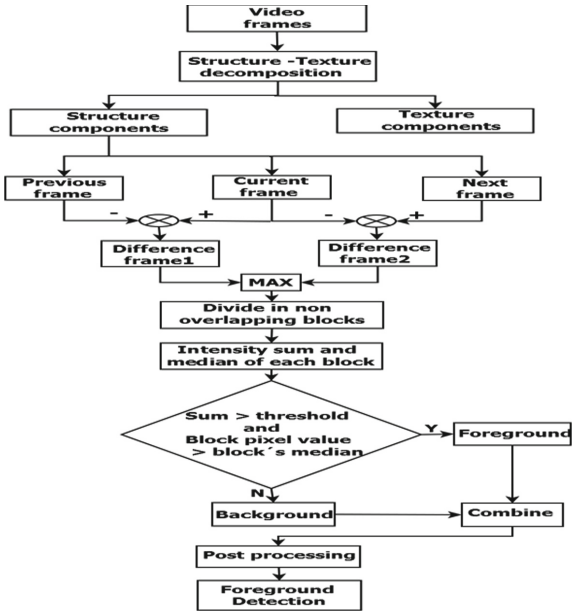

Fig. 1. Schematic diagram of the proposed technique

Table 1. Description of test video dataset

Description of video data	Frame size in pixels	Number of frames	Frame rate	Bit rate
Video 1: In this dataset two men are moving in hall with briefcase in their hands. The color of the foreground and background are not much varied, but there is variation in illumination. (**Color**)	240×352	299	30	24
Video 2: In this dataset a man is walking slowly from one side of the frame to other side. The color of foreground is lighter than the background. (**Color**)	240×368	132	25	24
Video 3: In this dataset a lady is running. There are high amount of noise and illumination variation. (**Color**)	92×144	42	35	24

4 Implementation

The sequence of steps is presented below for the implementation of our technique:

4.1 Extraction of Frames from Video Data and Conversion to Gray-Scale

Frames from the input video data $D : Z \times Z \times T \rightarrow Z$ are extracted. In case, $D(x, y, t|x, y \in Z, t \in T) = N \in Z$. The j^{th} frame can be represented as $F_j(x, y) = D(x, y, t = tj)$. Here in the next step, we focus on gray-scale frames in place of colored ones, due to the high complexity of three dimensional structures of color frames. Therefore, gray-scale frames can be obtained from colored one by employing [5]:

$$F(x, y) = 0.2126R(x, y) + 0.7152G(x, y) + 0.0722B(x, y) \tag{1}$$

Here R, G and B represent the red, green and blue channel components of each colored pixel, and scalar multiples represent the weights of these components.

4.2 Structure-Texture Decomposition

The individual video frame extracted in the preceding step can have noise, and our target is to decrease the effect of noise for detection of accurate moving objects in the subsequent step. For that, structure and texture/noise decomposition technique is employed on individual frame with the help of Algorithm 1 and for the further processing, only the structure component (U) are being utilized.

4.3 Improved Three-Frame Differencing

We employ the improved three frame differencing technique to find the pixel-wise difference of consecutive frames. It can be summarized in the following steps:

- Difference frame $Df_{t,t-1}(x,y)$ between frames $U_{t-1}(x,y)$ and $U_t(x,y)$ as well as second difference frame $Df_{t+1,t}(x,y)$ between frames $U_{t+1}(x,y)$ and $U_t(x,y)$ are estimated:

$$Df_{t,t-1}(x,y) = |U_{t-1}(x,y) - U_t(x,y)| \tag{2}$$
$$Df_{t+1,t}(x,y) = |U_{t+1}(x,y) - U_t(x,y)| \tag{3}$$

  Here $U_{t-1}(x,y)$, $U_t(x,y)$, and $U_{t+1}(x,y)$ are the structure components of three successive video frames.

Fig. 2. Results with Video 1 for the frames number (a) 72^{th} (b) 191^{th} (c) 222^{th} (d) 265^{th}; row wise, up to down: original frame, ground truth, Sukumaran [33], Fei [3], Yin [34], Sengar [5], and the proposed method.

- Pixel-wise maximum intensity value between two difference frames is computed:

$$Df_t(x, y) = max[Df_{t,t-1}(x, y), Df_{t+1,t}(x, y)] \quad (4)$$

4.4 Divide in Non-overlapping Blocks

Computed difference frame in previous step is partitioned into 8×8 size non-overlapping blocks.

4.5 Detection of Moving Objects

This process can be summarized in the following steps:

- Intensity sum and median of each non-overlapping block are computed.

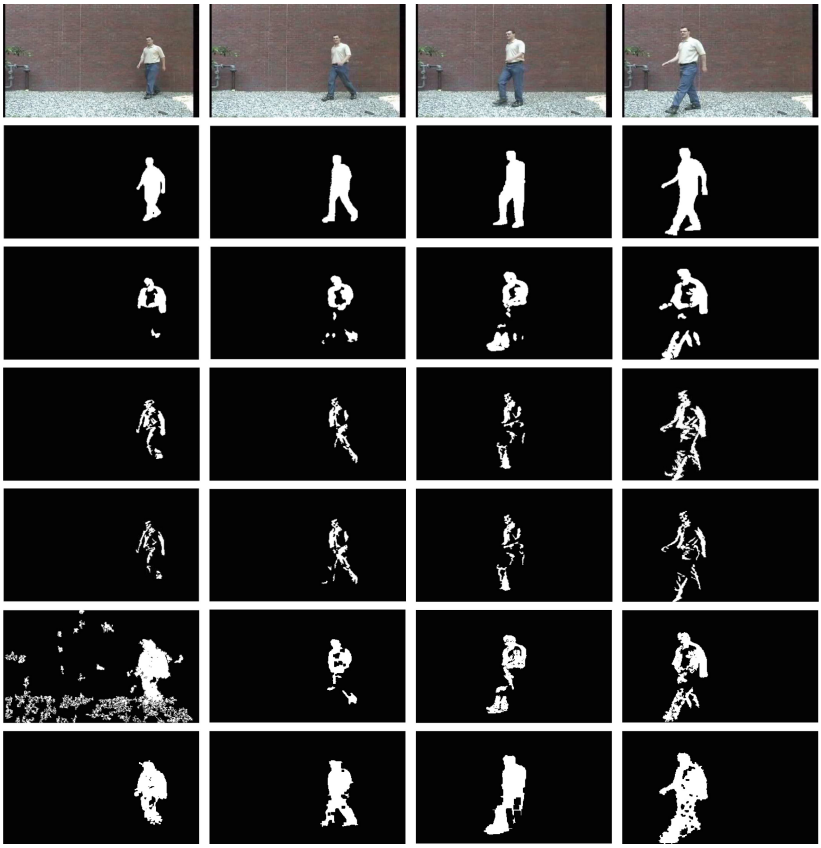

Fig. 3. Results with Video 2 for the frames number (a) 4^{th} (b) 34^{th} (c) 81^{th} (d) 119^{th}; row wise, up to down: original frame, ground truth, Sukumaran [33], Fei [3], Yin [34], Sengar [5], and the proposed method.

- Classification of background and foreground on the basis of block's median (Bm) and threshold (Th) i.e. if the value of block's intensity and block's sum are less than block's median and threshold respectively then it will be background (0), otherwise foreground (1).

$$FG_t(x,y) = \begin{cases} 0, & Df_t(x,y) < Bm \text{ and } sum(B_{tj}) < Th \\ 1, & \text{otherwise} \end{cases} \tag{5}$$

  Here FG_t, $sum(B_{tj})$ denote the foreground objects for the t^{th} frame, sum of intensities of the j^{th} block of the t^{th} frame respectively.
- Motion is detected by combining both the foreground and background pixels.

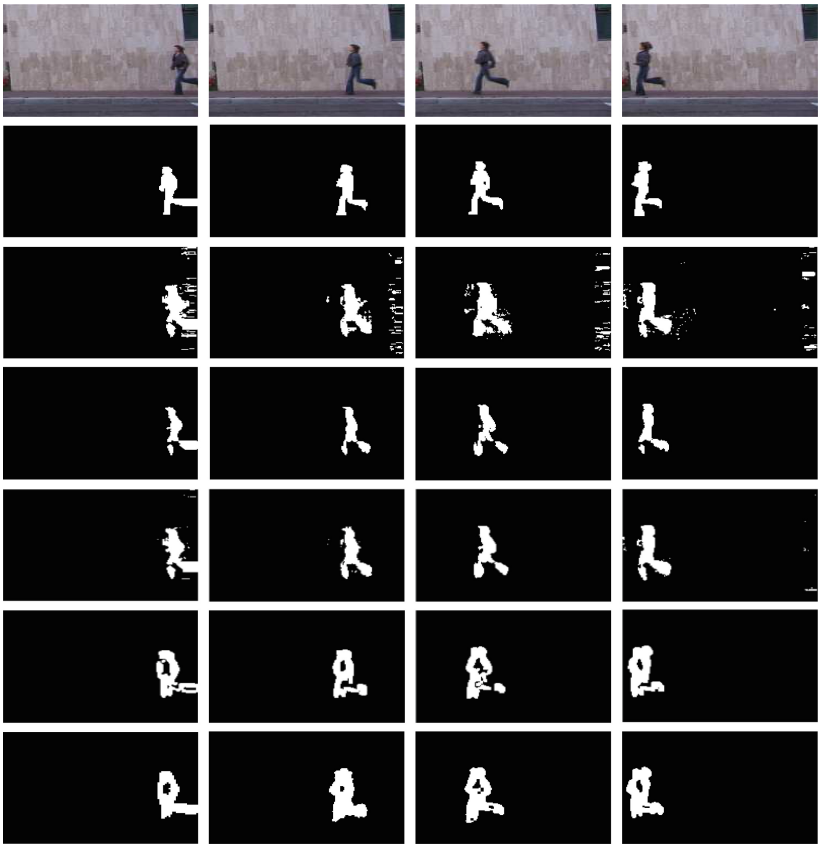

Fig. 4. Results with Video 3 for the frames number (a) 3^{th} (b) 11^{th} (c) 28^{th} (d) 38^{th}; row wise, up to down: original frame, ground truth, Sukumaran [33], Fei [3], Yin [34], Sengar [5], and the proposed method.

4.6 Post-processing

In morphology, closing is a cascaded combination of first dilation followed by erosion [35] and it is employed to fill the gap within the foreground object. Connected component labeling is employed to remove the isolated noisy blobs. Finally, morphological closing operations of square structuring element (SE) of size 3×3 (as shown in Eq. 6) followed by connected components labeling are employed to accurately segment the motion.

$$MV_t = (FG_t \oplus SE) \ominus SE \tag{6}$$

The schematic diagram of our technique is shown by Fig. 1.

5 Experimental Results and Analysis

We have done qualitative and quantitative evaluation to prove the efficiency of our technique.

5.1 Qualitative Evaluation

To prove the effectiveness of our method, the proposed and existing techniques have been implemented and tested on standard benchmark video datasets namely: Video 1 (Hall monitoring) [36], Video 2 (Walk) [37], and Video 3 (Daria) [38]. The details of used video sequences are given in Table 1. We have iterated the structure-texture decomposition algorithm 100 times. The qualitative results of our algorithm as well as other tested methods on different video sequences are shown in Figs. 2, 3 and 4. These above mentioned figures display the original video frames, ground truth, and the motion segmentation results in each video frame by the five different techniques independently. For Video 1 we have displayed the experimental results on frame number 72, 191, 222, and 265. Similarly the related frame numbers are (4, 34, 81, 119), and (3, 11, 28, 38) for Video 2 and Video 3 respectively. The original frames and their ground truths are displayed by the first two rows of Figs. 2, 3 and 4 respectively. Next five rows (from up to down) show the outcomes of Sukumaran [33], Fei [3], Yin [34], Sengar [5] and the Proposed technique respectively. It is proved from the aforementioned results that our approach in every case has considerably higher likeness with the ground truth as compared to existing approaches. All the existing approaches mis-classify most of the foreground pixels as background and vice-versa for all the tested video datasets. Hence, target objects are detected with holes and foreground pixels cannot be separated from the background. Furthermore, we have also measured and compared the performance of our technique based on quantitative evaluation in next section.

5.2 Quantitative Evaluation

Quantitative evaluation and comparison with the help of accuracy and processing time are also done in our work. The experimental results of the proposed and existing techniques are compared with the ground truth (Figs. 2, 3 and 4). Accuracy of the results can be computed using following equation:

$$count = \sum_{x=1}^{h} \sum_{y=1}^{w} G_t(x, y) \ (XOR) \ MV_t(x, y) \tag{7}$$

$$accuracy = 100 - \frac{count}{framesize} \times 100 \tag{8}$$

Where, the ground truth and obtained result of t^{th} frame are represented by G_t and MV_t respectively. The number of mismatched pixels between ground truth and result are denoted by variable count. The results of quantitative evaluation are given in Tables 2 and 3, these results are computed using different approaches over different datasets. The comparable results are also shown with the help of bar chart in Figs. 5 and 6 for accuracy and processing time respectively. It has been agreed from accuracy and processing time based quantitative evaluation that the proposed technique gives better results in comparison with other techniques.

Table 2. Average accuracy of existing and proposed schemes.

Input video	Scheme				
	Sukumaran [33]	Fei [3]	Yin [34]	Sengar [5]	Proposed
Video 1	88.9	89.4	86.7	92.4	96.8
Video 2	93.2	92.9	92.5	89.9	96.6
Video 3	90.5	93.3	92.5	93.8	94.6

Table 3. Processing time of existing and proposed schemes.

Input video	Scheme				
	Sukumaran [33]	Fei [3]	Yin [34]	Sengar [5]	Proposed
Video 1	1.01	0.90	0.87	0.95	0.83
Video 2	1.05	1.04	0.98	0.99	0.91
Video 3	0.79	0.75	0.73	0.80	0.65

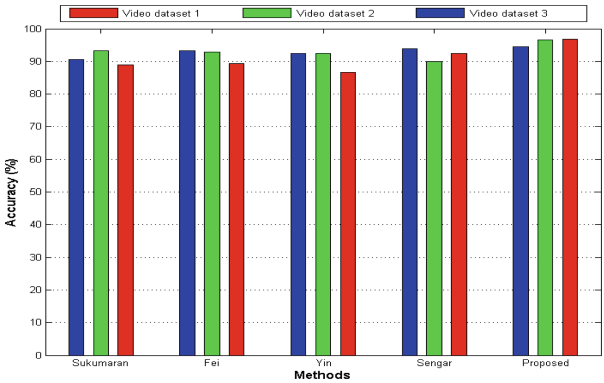

Fig. 5. Comparison in the form of detection accuracy

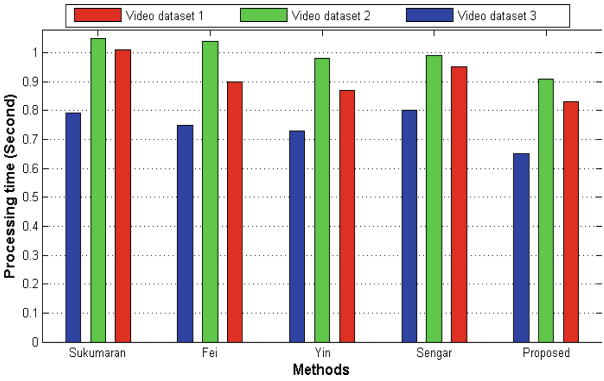

Fig. 6. Comparison in the form of processing time

6 Conclusion

Different techniques for motion segmentation have been analyzed and we came at conclusion that detection accuracy of these techniques is not suitable because of large amount of noise in the input video dataset. For that, we have proposed a simple and efficient approach for motion segmentation from complex noise scenes. With the help of proposed structure-texture decomposition and improved three frames differencing based scheme, we eliminate almost all the background noises without loss of foreground. Therefore, optimal results are obtained finally. This approach has improvements in comparison with other recent methods in the form of enhanced accuracy and decreased processing time. Experimental results and analyses prove that the proposed algorithm can effectively handle a slow/fast moving objects, camera shake problem and variation of object size with noisy datasets. In future attempt, our technique will be improved for visual object tracking in moving camera environment.

References

1. Elharrouss, O., Moujahid, D., Tairi, H.: Motion detection based on the combining of the background subtraction and the structure-texture decomposition. Optik-Int. J. Light Electron Opt. **126**(24), 5992–5997 (2015)
2. Sengar, S.S., Mukhopadhyay, S.: Detection of moving objects based on enhancement of optical flow. Optik-Int. J. Light Electron Opt. **145**, 130–141 (2017)
3. Fei, M., Li, J., Liu, H.: Visual tracking based on improved foreground detection and perceptual hashing. Neurocomputing **152**, 413–428 (2015)
4. Sengar, S.S., Mukhopadhyay, S.: A novel method for moving object detection based on block based frame differencing. In: 3rd International Conference on Recent Advances in Information Technology, pp. 462–472. IEEE (2016)
5. Sengar, S.S., Mukhopadhyay, S.: Motion detection using block based bi-directional optical flow method. J. Vis. Commun. Image Represent. **49**, 89–103 (2017)
6. Sahoo, P.K., Kanungo, P., Mishra, S.: A fast valley-based segmentation for detection of slowly moving objects. Signal Image Video Process. **12**, 1–8 (2018)
7. Bouwmans, T., Sobral, A., Javed, S., Jung, S.K., Zahzah, E.H.: Decomposition into low-rank plus additive matrices for background/foreground separation: a review for a comparative evaluation with a large-scale dataset. Comput. Sci. Rev. **23**, 1–71 (2016)
8. Sengar, S.S., Mukhopadhyay, S.: Moving object tracking using Laplacian-DCT based perceptual hash. In: International Conference on Wireless Communications, Signal Processing and Networking, pp. 2345–2349. IEEE (2016)
9. Sandberg, B., Chan, T., Vese, L.: A level-set and Gabor-based active contour algorithm for segmenting textured images. In: UCLA Department of Mathematics CAM Report. Citeseer (2002)
10. Ono, S., Miyata, T., Yamada, I.: Cartoon-texture image decomposition using blockwise low-rank texture characterization. IEEE Trans. Image Process. **23**(3), 1128–1142 (2014)
11. Jeon, J., Cho, S., Tong, X., Lee, S.: Intrinsic image decomposition using structure-texture separation and surface normals. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8695, pp. 218–233. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10584-0_15
12. Wells, W.M., Grimson, W.E.L., Kikinis, R., Jolesz, F.A.: Adaptive segmentation of MRI data. IEEE Trans. Med. Imag. **15**(4), 429–442 (1996)
13. Malgouyres, F.: Mathematical analysis of a model which combines total variation and wavelet for image restoration. J. Inf. Process. **2**(1), 1–10 (2002)
14. Malgouyres, F.: Combining total variation and wavelet packet approaches for image deblurring. In: IEEE Workshop on Variational and Level Set Methods in Computer Vision, Proceedings, pp. 57–64. IEEE (2001)
15. Candès, E.J., Guo, F.: New multiscale transforms, minimum total variation synthesis: applications to edge-preserving image reconstruction. Signal Process. **82**(11), 1519–1543 (2002)
16. Casadei, S., Mitter, S., Perona, P.: Boundary detection in piecewise homogeneous textured images. In: Sandini, G. (ed.) ECCV 1992. LNCS, vol. 588, pp. 174–183. Springer, Heidelberg (1992). https://doi.org/10.1007/3-540-55426-2_20
17. Zhu, S.C., Wu, Y., Mumford, D.: Filters, random fields and maximum entropy (frame): towards a unified theory for texture modeling. Int. J. Comput. Vis. **27**(2), 107–126 (1998)

18. Halidou, A., You, X., Hamidine, M., Etoundi, R.A., Diakite, L.H.: Fast pedestrian detection based on region of interest and multi-block local binary pattern descriptors. Comput. Electr. Eng. **40**(8), 375–389 (2014)
19. Foresti, G.L., Micheloni, C., Piciarelli, C.: Detecting moving people in video streams. Pattern Recogn. Lett. **26**(14), 2232–2243 (2005)
20. Caballero, A.F., Castillo, J.C., Cantos, J.M., Tomas, R.M.: Optical flow or image subtraction in human detection from infrared camera on mobile robot. J. Rob. Auton. Syst. **58**, 1273–1281 (2010)
21. Bouguet, J.Y.: Pyramidal implementation of the affine Lucas kanade feature tracker description of the algorithm. Intel Corporation **5**, 1–10 (2001)
22. Stauffer, C., Grimson, W.: Adaptive background mixture models for real-time tracking. In: International Conference On Computer Vision and Pattern Recognition. IEEE (1999)
23. Sengar, S.S., Mukhopadhyay, S.: Moving object detection based on frame difference and W4. Signal Image Video Process. **11**(7), 1357–1364 (2017)
24. Maddalena, L., Petrosino, A.: The SOBS algorithm: what are the limits? In: Workshop on Computer Vision and Pattern Recognition, pp. 21–26. IEEE (2012)
25. Oliver, N.M., Rosario, B., Pentland, A.P.: Bayesian computer vision system for modeling human interactions. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 831–843 (2000)
26. Sengar, S.S., Mukhopadhyay, S.: Foreground detection via background subtraction and improved three-frame differencing. Arab. J. Sci. Eng. **42**(8), 3621–3633 (2017)
27. Chen, E., Xu, X., Yang, X., Zhang, W.: Quaternion based optical flow estimation for robust object tracking. J. Digit. Signal Proc. **23**, 118–125 (2013)
28. Schwarz, L.A., Mkhitaryan, A., Mateus, D., Navab, N.: Human skeleton tracking from depth data using geodesic distances and optical flow. J. Image Vis. Comput. **30**, 217–226 (2012)
29. Sengar, S.S., Mukhopadhyay, S.: Moving object area detection using normalized self adaptive optical flow. Optik-Int. J. Light Electron Opt. **127**(16), 6258–6267 (2016)
30. Liu, D., Yu, J.: Otsu method and k-means. In: 9th International Conference on Hybrid Intelligent Systems, pp. 344–349. IEEE (2009)
31. Liao, P., Chen, T., Chung, P.: A fast algorithm for level thresholding. J. Inf. Sci. Eng. **17**, 713–727 (2001)
32. Luminita, A.V., Stanley, J.O.: Modeling textures with total variation minimization and oscillating patterns in image processing. J. Sci. Comput. **19**(1–3), 553–572 (2003)
33. Sukumaran, A.N., Sankararajan, R., Swaminathan, M.: Compressed sensing based foreground detection vector for object detection in wireless visual sensor networks. AEU-Int. J. Electron. Commun. **72**, 216–224 (2017)
34. Yin, J., Liu, L., Li, H., Liu, Q.: The infrared moving object detection and security detection related algorithms based on W4 and frame difference. Infrared Phys. Technol. **77**, 302–315 (2016)
35. Dougherty, E.R., Lotufo, R.A.: Hands-on Morphological Image Processing, vol. 71. SPIE Optical Engineering Press, Washington (2003)
36. Database: Images & video clips (2), Collected by the HDTV group, July 2006. http://see.xidian.edu.cn/vipsl/database_Video.html
37. vidme, videodata, July 2015. https://vid.me/videodata
38. Action Recognition. https://github.com/hueihan/Action_Recognition/tree/master/data/WIS/video/run