Robust Segmentation and Classification of Moving Objects from Surveillance Video

Chandrajit M¹*, Shobha Rani N² and Manohar N²

¹ Maharaja Research Foundation, Maharaja Institute of Technology Mysore, India; ² Department of Computer Science, Amrita School of Arts & Science, Mysuru Campus, Amrita, Vishwa Vidyapeetham, India.

*Email: chandrajithmmca@mitmysore.in

Abstract. Object segmentation and classification in the video sequence is a classical and critical problem that is constantly addressed due to vast real-time applications such as autonomous vehicles, smart surveillance, etc. Segmentation and classification of moving objects in video sequences captured from real-time non-constraint sequences is a challenging task. This paper presents a moving object segmentation and classification method for video sequences captured in a real-time environment. The key contributions of the paper include a method to detect and segment motion regions by applying the non-parametric Kolmogorov–Smirnov statistical test in the Spatio-temporal domain and a probabilistic neural network-based classification method to classify the moving objects into various classes. Promising results are obtained by experimentation using challenging PETS and Change Detection datasets. To corroborate the efficacy, a comparative analysis with contemporary method is also performed.

1. Introduction

Enormous amount of work in the domain of visual analysis has been reported in the last two decades due to the huge demand in applications such as robotics, video surveillance, driver assist system, video indexing, traffic video monitoring etc. Nevertheless, success of any visual analysis application directly depends on accurate segmentation and classification of moving objects [1, 2]. Surveillance video feed generally suffers from illumination variations, object occlusion, noise, cluttered environment etc. These problems in turn increase the complexity for devising an accurate solution to the problem. Therefore, research in this direction focus to contribute substantial evidence based solutions.

In this paper emphasis is given to accurately extract moving objects from surveillance video sequence and further classify the objects into predefined classes.

The article is organized as follows. Section 2 reports the related work. Section 3 and 4 elaborates the proposed motion segmentation and object classification method respectively, Section 5 presents the experimental and comparative results and Section 6 presents the conclusion.

2. Related Work

Background subtraction, optical flow and temporal differencing are the major category of techniques for motion segmentation. Background subtraction technique constructs the
background model for motion segmentation. Optical flow technique computes the direction vectors for every pixel using information from successive frames to achieve motion segmentation. Temporal differencing based technique achieves motion segmentation on the basis of pixel difference between frames using a threshold. Statistical tools are also used in these techniques for correlating the pixel values from consecutive frames.

Stuaffer and Grimson [3] proposed GMM based background subtraction method for motion segmentation. Chen et al. [4] combined pixel-level and block-level features to construct background model for motion segmentation. Armanfard et al. [5] proposed texture and edge feature based background subtraction for pedestrian segmentation. Elgammal et al. [6] proposed kernel density estimation based background modeling for moving object segmentation. Sadaf et al. [7] proposed a temporal domain based background subtraction method. Subudhi et al. [8] proposed Wronskian framework based background construction technique. Sergi et. al. [9] proposed convolution network based semi-supervised method for motion segmentation. Variation of convolution network based motion segmentation work is reported in [10-12].

Denman et al. [13] proposed a optical flow based method for motion segmentation. Variation of optical flow based technique for motion segmentation can be found in the work reported by [14, 15].

Murali and Girisha [16] proposed temporal differencing based method for motion segmentation. Variation of temporal differencing based techniques can be found in the work reported by [17-19]. Cheng and Chen [20] proposed achieved temporal based motion segmentation in frequency transformed domain. Liu et al. [21] proposed information theory based temporal differencing motion segmentation method.

The methods reported can be classified as online or learning based methods. Online based methods use geometric properties of objects for classification. Whereas the learning based method uses a trained model for classification. Learning based method is more suitable due to increased accuracy. Javed and Shah [22] proposed a method to classify objects into three classes by using temporal motion features. Liang and Juang [23] proposed a method to classify the objects into three classes using histogram feature with SVM classifier. Zang and Klette [24] proposed a method for two class classification using geometric features. Asaidi et al. [25] proposed a method for vehicle classification using shape features. Zhang et al. [26] proposed a three class classification using shape and motion features. Zhang et al. [27] proposed adaptive boosting based method for four class classification using block based local binary pattern. Vijaya and Venkata [28] proposed war filed object classification using SIFT features. Mahalingam and Subramoniam [29] proposed texture based decision tree classification method for object classification.

In summary, a great amount of work has been dedicated for motion segmentation using background subtraction. Generally, the distribution of data is assumed to be normal distribution in background subtraction based methods. Nevertheless, in real-time environment, this assumption need not be valid [6]. Further, the background subtraction based methods cannot handle dynamic and cluttered environment. Moreover, the background subtraction technique cannot be applied for moving and Pan Tilt and Zoom (PTZ) cameras. Optical flow technique based methods accurately segments moving objects. However, these methods are not suitable for illumination variation and cluttered environment sequences. Temporal differencing based methods can extract motion objects satisfactorily in the case of dynamic and cluttered environment. However, these methods fail to accurately extract the objects.

Further, the methods reported on moving objects classification are limited in numbers. Most of them classify objects into few categories. However, in real-time scenario the scene
will be composed of many kinds of objects. Therefore, effort is made in this work for multi-class classification.

3. The Proposed Motion Segmentation Method

Figure 1 presents the schematic representation of proposed method devised for moving object segmentation from video sequence. The input frames $I_t$ of video are smoothened by 2D $3 \times 3$ Gaussian filter [30] to reduce the noise caused by sensors and illumination variations. Consequently, the smoothened frame $F_t$ is obtained, where, $t$ is the temporal frame. Subsequently, RGB mean value for each pixel $P(w,h)$ is considered as pixel value.

![Figure 1](image-url)

Figure 1. Schematic representation of proposed method for segmentation

The detection of motion pixels can be determined by capturing the variations among pixel values from successive frames. However, capturing the variations using single pixel value from successive frame will not lead to accurate motion segmentation. Therefore, for each pixel its $3 \times 3$ neighbors are considered in this work.

The displacement of object in frames of the video sequence in the case of fast moving object will be very high. Therefore, successive temporal frames are used for motion segmentation.

Kolmogorov–Smirnov (KS) test [31] is a popular nonparametric test in statistics used to compare a sample with a reference probability distribution, or it can also be used to compare two samples. The variation in pixel values from successive temporal frames for moving objects segmentation is determined by employing the KS test. In the case of dynamic environment, pixel data obtained from video frame cannot be assumed to follow any specific distribution. Therefore, this test has been chosen for moving object segmentation in dynamic environment.

The KS test requires sample values for establishing the variation. The neighborhood pixel values in smoothened frames $F_t$ and $F_{t+1}$ which is denoted as $F_t^{P N_9 (RGB)}$ and $F_{t+1}^{P N_9 (RGB)}$ respectively are used as sample values.

The KS test is applied on the pixel values in smoothened frames $F_t$ and $F_{t+1}$ using Eq.(1) for the segmentation of moving objects.

$$D_{F_t, F_{t+1}} = \sup_x \left| F_t^{P N_9 (RGB)} - F_{t+1}^{P N_9 (RGB)} \right|$$

where, $D$ is KS test statistic for each pixel and $\sup$ is supremum function.
The variation among pixel values is captured by comparing the test statistic $D_{F_t,F_{t+1}}$ with threshold value $\tau$ chosen from table for $\alpha = 0.5$. The null hypothesis is rejected at level $\alpha$ based on Eq.(2).

$$D_{F_t,F_{t+1}} > c(\alpha) \sqrt{\frac{n+m}{nm}}$$

(2)

where, $(n, m)$ are size of two samples and

$$c(\alpha) = \sqrt{-\frac{1}{2} \ln \frac{\alpha}{2}}$$

(3)

Consequently, the motion pixels are detected using KS test. The output frames $O_t$ and $O_{t+1}$ that contain extracted foreground objects from input frames $I_t$ and $I_{t+1}$ respectively are generated using Eq.(4).

$$O_{mP(w,h)} = \begin{cases} \text{RGB of } I_{mP(w,h)} & \text{if } (D_{F_t,F_{t+1}} \geq \tau) \\ 0 & \text{otherwise} \end{cases}$$

(4)

where, $m$ is the temporal frame and $I_{mP(w,h)}$ is input frame pixel.

4. The Proposed Classification Method

Schematic representation of proposed method to classify objects is shown in figure 2. Objects are classified into predefined classes viz., single person, group of people, two wheeler, small vehicle or big vehicle using PNN classifier [32].

![Schematic representation of proposed method for classification](image)

Figure 2. Schematic representation of proposed method for classification

PNN is feed forward neural network based on Bayesian framework with Kernel Fisher discriminant analysis. In this work, PNN consists of four layered network which estimate PDF for each of the classes. The first layer is the input feature values. The second layer consists of the Gaussian functions estimated by given set of feature values. Subsequently, averaging of values is performed in third layer for each class. Finally, class label is determined for maximum value among the distribution value output in the fourth layer.

In general, a PNN for M classes is defined as:

$$y_j(X) = \frac{1}{n_j} \sum_{i=1}^{n_j} \exp\left(-\frac{\|x_i - \mu_j\|^2}{2\sigma^2}\right)$$

(5)

$$j = 1, 2, ..., M$$

$n_j$ is the number of features.
The PNN classifier assigns $X$ to class $j$ having maximum $y_j(X)$.

Object representation using features plays a pivotal role in classification. In this work, object is represented by the following features: Seven invariant Hu moments features [33], Gabor Features, geometric features such as, centroid, major axis, minor axis, orientation.

5. Experiments

Experiments on the challenging benchmark IEEE PETS and IEEE CD [34] and DARPA Neovision datasets have been conducted to evaluate the proposed method. The dataset contains outdoor and indoor sequences captured from normal, thermal, night vision and PTZ sensors. The result of motion segmentation is shown in figure 3 and figure 4. Similarly, the result of classification is shown in figure 5 and figure 6.

Quantitative evaluation has been performed using metrics viz. precision, recall and F-measure [34]. Further, sample groundtruth frames from the same dataset have been used for evaluation. Table 1 presents the obtained results.

![Figure 3. Motion Segmentation result for (a) PETS 2013 People Tracking (b) Thermal](image)

![Figure 4. Motion Segmentation result for (a) Night vision Fluid Highway (b) PTZ frames](image)

![Figure 5. Results of Classification for (a) PETS 2001 (b) PETS 2013](image)

![Figure 6. Results of classification for DARPA Neovision2](image)
Evaluation results in table 1 indicate the accurate extraction of foreground object blobs from sequences captured from variety of sensors. Further, the results presented are raw and no post processing has been incorporated for enhancement.

Comparative results from the SOBS [35], Chi-Square test based [36] and proposed methods are presented. Table 2 presents the comparative performance metrics. It indicates the significant performance of proposed KS test based segmentation. This corroborates the suitability of proposed method for video sequences captured using various kinds of sensors.

A quantitative result of classification is presented in table 3 which clearly demonstrates the promising results obtained by the proposed method for real-time challenging surveillance video feed. The combination of geometric, shape and texture features has proven in retaining intra and inter class variations. An F-measure of 96% and 71% is the maximum and minimum score obtained.

**Table 1.** Performance metrics values obtained by the proposed KS test based motion segmentation

| Dataset     | Sequence     | Precision | Recall | F-measure |
|-------------|--------------|-----------|--------|-----------|
| PETS 2013   |              | 0.79      | 0.75   | 0.77      |
| PETS 2006   |              | 0.86      | 0.72   | 0.78      |
| CD night vision |          | 0.54      | 0.63   | 0.58      |
| CD PTZ      |              | 0.63      | 0.6    | 0.61      |
| CD thermal sensor | | 0.81      | 0.69   | 0.75      |
| Average     |              | 0.73      | 0.68   | 0.70      |

**Table 2.** Comparative performance metrics values

| Dataset     | Sequence     | Precision | Recall | F-measure |
|-------------|--------------|-----------|--------|-----------|
| Proposed    | SOBS Chi-   | Proposed  | SOBS   | Proposed  | SOBS   | Chi-   |
|             | Square Test  | Square Test| Square Test| Square Test| Square Test|         |
| PETS 2013   |              | 0.79      | NA     | 0.76      | 0.75    | NA     | 0.77    | NA     | 0.73    |
| PETS 2006   |              | 0.86      | 0.71   | 0.83      | 0.72    | 0.74   | 0.68    | 0.78    | 0.72    | 0.75    |
| CD night vision |          | 0.54      | 0.48   | 0.51      | 0.63    | 0.75   | 0.60    | 0.58    | 0.58    | 0.55    |
| CD PTZ      |              | 0.63      | 0.02   | 0.59      | 0.60    | 0.70   | 0.57    | 0.61    | 0.05    | 0.58    |
| CD thermal sensor | | 0.81      | 0.96   | 0.78      | 0.69    | 0.43   | 0.65    | 0.75    | 0.60    | 0.71    |

6. Conclusion

In this paper, a method is proposed for moving object segmentation and classification from surveillance videos. Moving object segmentation is based on statistical modeling combined with temporal domain. Suitability of the non-parametric statistical hypothesis testing tool for moving object segmentation has been explored in this work. Further, object classification is achieved by ensemble of geometric, shape and texture features with PNN classifier. A five class classification has been attempted in this work. Experimentation has been carried out by using the challenging benchmark IEEE PETS, IEEE CD and DARPA Neovision2 datasets.
which contain sequences captured from thermal, PTZ, aerial and night vision sensors. Qualitative results indicate the robust nature of proposed method in the case of cluttered and illumination variations environment. Further, comparative analysis with related works has also been presented for showcasing the stability of proposed method. In future, application of deep multi-layered networks shall be investigated for segmentation and classification.

Table 3. Quantitative results of proposed classification method

| Dataset Sequence          | Training Samples | Precision | Recall | F-measure |
|---------------------------|-------------------|-----------|--------|-----------|
| PETS 2013 S2.L1 View 1    |                    |           |        |           |
| 70%                       | 0.91              | 0.9       | 0.90   |
| 50%                       | 0.87              | 0.9       | 0.88   |
| 30%                       | 0.9               | 0.89      | 0.89   |
| PETS 2013 S2.L1 View 7    |                    |           |        |           |
| 70%                       | 0.9               | 0.73      | 0.81   |
| 50%                       | 0.91              | 0.86      | 0.88   |
| 30%                       | 0.85              | 0.85      | 0.85   |
| PETS 2001 Camera 1        |                    |           |        |           |
| 70%                       | 0.94              | 0.88      | 0.91   |
| 50%                       | 0.95              | 0.87      | 0.91   |
| 30%                       | 0.91              | 0.84      | 0.87   |
| PETS 2001 Camera 2        |                    |           |        |           |
| 70%                       | 0.76              | 0.75      | 0.75   |
| 50%                       | 0.75              | 0.74      | 0.74   |
| 30%                       | 0.72              | 0.71      | 0.71   |
| Neovision2 002            |                    |           |        |           |
| 70%                       | 0.76              | 0.89      | 0.82   |
| 50%                       | 0.75              | 0.87      | 0.81   |
| 30%                       | 0.7               | 0.94      | 0.8    |
| Neovision2 012            |                    |           |        |           |
| 70%                       | 0.99              | 0.94      | 0.96   |
| 50%                       | 0.98              | 0.9       | 0.94   |
| 30%                       | 0.95              | 0.9       | 0.92   |

References
[1] Zappella, Luca, Xavier Lladó, and Joaquim Salvi, Motion segmentation: a review, inProc. Int. conference on Artificial Intelligence Research and Development: Proceedings of the 11th International Conference of the Catalan Association for Artificial Intelligence, 2008, IOS Press.
[2] Olatunji I.E., Cheng CH, Video Analytics for Visual Surveillance and Applications: An Overview and Survey. In: Tsihrintzis G., Virvou M., Sakkopoulos E., Jain L. (eds) Machine Learning Paradigms. Learning and Analytics in Intelligent Systems, (2019), vol 1. Springer.
[3] Stauffer C and Grimson, Adaptive background mixture models for real-time tracking,
in Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 1999, vol. 2.

[4] Chen Y., Chen C., Huang C and Hung Y, Efficient hierarchical method for background subtraction. Pattern Recognition, 2007, vol. 40, no. 10, pp. 2706-2715.

[5] Armanfard N., Komeili M and Kabir E, Ted: A texture-edge descriptor for pedestrian detection in video sequences. Pattern Recognition, 2012, vol. 45, no. 3, pp. 983-992.

[6] Elgammal A., Duraiswami R., Harwood D and Davis L, Background and foreground modeling using nonparametric kernel density estimation for visual surveillance, in Proc. of the IEEE, 2002, vol. 90, no. 7, pp. 1151-1163.

[7] Sadaf A., Ali J and Irfan, Human identification on the basis of gaits using time efficient feature extraction and temporal median background subtraction. International Journal of Image, Graphics and Signal Processing, 2014, vol. 6, no. 3, pp. 35-42.

[8] Subudhi B., Ghosh S and Ghosh, Change detection for moving object segmentation with robust background construction under Wronskian framework. Machine Vision and Applications, 2013, vol. 24, no. 4, pp. 795-809.

[9] Sergi Caelles, Kevis-Kokitsi Maninis, Jordi Pont-Tuset, Laura Leal-Taixe, Daniel Cremers, Luc Van Gool; One-Shot Video Object Segmentation, in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 221-230.

[10] Pavel Tokmakov, Kartek Alahari, Cordelia Schmid; Learning Video Object Segmentation With Visual Memory, in Proc. IEEE International Conference on Computer Vision (ICCV), 2017, pp. 4481-4490.

[11] Jeong, J.; Yoon, T.S.; Park, J.B.: 'MOSnet: moving object segmentation with convolutional networks', Electronics Letters, IET Digital Library, 2018, 54, (3), p. 136-138.

[12] Vecchio, S. Palazzo, D. Giordano, F. Rundo and C. Spampinato, MASK-RL: Multiagent Video Object Segmentation Framework through Reinforcement Learning, in Proc. IEEE Transactions on Neural Networks and Learning Systems. 2020.

[13] Denman S., Fookes C and Sridharan S, Improved simultaneous computation of motion detection and optical flow for object tracking. In Proceedings of the International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2009, pp. 175-182.

[14] Verri A., Uras S and De Micheli E, Motion segmentation from optical flow, in Proc. of Alvey Vision Conference, 1989, pp. 1-6.

[15] Klappstein J., Vaudrey T., Rabe C., Wedel A and Klette R, Moving object segmentation using optical flow and depth information, in Proc. Pacific-Rim Symposium on Image and Video Technology, 2009, pp. 611-623.

[16] Murali S and Girisha R, Segmentation of motion objects from surveillance video sequences using temporal differencing combined with multiple correlation, in Proc. Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2009, pp. 472-477.

[17] Lim T., Han B and Han J. H, Modeling and segmentation of floating foreground and background in videos. Pattern Recognition, 2012, vol. 45, no. 4, pp. 1696-1706.

[18] Lee ChaoYang C. L., ShouJen Lin and Yang C, An efficient continuous tracking system in real-time surveillance application. Journal of Network and Computer Applications, Special Issue on Trusted Computing and Communications, 2012, vol. 35, no. 3, pp. 1067-1073.

[19] Ahn J., Choi C., Kwak S., Kim K and Byun H, Human tracking and silhouette extraction for human robot interaction systems. Pattern Analysis and Applications, 2009, vol. 12, no. 2, pp.167-177.

[20] Cheng F and Chen Y, Real time multiple objects tracking and identification based on
discrete wavelet transform. Pattern Recognition, 2006, vol. 39, no. 6, pp. 1126-1139.

[21] Liu C., Yuen P. C and Qiu G, Object motion detection using information theoretic spatio temporal saliency. Pattern Recognition, 2009, vol. 42, no. 11, pp. 2897-2906.

[22] Javed O and Shah M, Tracking and object classification for automated surveillance, inProc. European Conference on Computer Vision, 2002, pp. 343-357.

[23] Liang Chung-Wei and Juan Chia-Feng, Moving object classification using local shape and HOG features in wavelet-transformed space with hierarchical SVM classifiers. Applied Soft Computing, 2015, vol. 28, pp. 483-497.

[24] Zang Q and Klette R, Object classification and tracking in video surveillance, inProc. Tenth International Conference on Computer Analysis of Image and Patterns (CAIP), 2003, vol. 128, pp. 198-205.

[25] Asaidi H., Aarab A and Bellouki M, Shadow elimination and vehicles classification approaches in traffic video surveillance context. Journal of Visual Languages & Computing, 2014, vol. 25, Issue 4, pp. 333-345.

[26] Zhang Z., Huang K., Wang Y and Li M, View independent object classification by exploring scene consistency information for traffic scene surveillance. Neurocomputing, 2013, vol. 99, pp. 250-260.

[27] Zhang L., Li S. Z., Yuan X and Xiang S, Real-time object classification in video surveillance based on appearance learning, inProc.the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007, pp. 1-8.

[28] Vijaya Lakshmi T.R., Venkata Krishna Reddy C. Object Classification Using SIFT Algorithm and Transformation Techniques. In: Mallick P., Balas V., Bhoi A., Zobaa A. (eds) Cognitive Informatics and Soft Computing. Advances in Intelligent Systems and Computing, Springer, 2019, vol 768.

[29] T. Mahalingam, M. Subramoniam, A robust single and multiple moving object detection, tracking and classification, Applied Computing and Informatics, 2018.

[30] Shapiro L. G. and Stockman G. C, Computer vision, (Prentice Hall, 2001).

[31] Dodge and Yadolah, The concise encyclopedia of statistics, (Springer Science & Business Media, 2008).

[32] Donald F. Specht, Probabilistic neural networks, Neural Networks, Volume 3, Issue 1, Science Direct, 1990, Pages 109-118.

[33] Hu M. K, Visual pattern recognition by moment invariants. IRE Transactions on Information Theory, 1962, vol. IT-8, pp.179-187.

[34] Goyette N., Jodoin P., Porikli F., Konrad J and Ishwar P, Changedetection.net: A new change detection benchmark dataset, inProc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2012, pp. 1-8.

[35] Maddalena L and Petrosino A, The SOBS algorithm: What are the limits?, inPorc. IEEE Computer Society Conference Computer Vision and Pattern Recognition Workshops (CVPRW), 2012, pp. 21-26.

[36] Chandrajit M., Girisha R. and Vasudev T, Motion Segmentation using Chi-Square Test Statistics. In Proceedings of the 2nd International Conference on Emerging Research in Computing, Information, Communication and Applications (ERCICA 2014), Elsevier, 2014, vol. 2, pp. 365-372.