Identification of Shape-Signature of Moving Objects in Video Frames: A review

Kalpana Martha, Chandra Sekhar Panda

Department of Computer Science and Application, Sambalpur University,

Jyoti Vihar, Odisha 768019, India

Abstract

Detecting and tracking object is applied in diverse areas of computer vision,including video surveillance, person tracking, vehicle navigation, and robotics. Object detection and classification is essential prior to trackan object from a video sequence.The detectionprocess focuses on position, shape, and size of an object and can be achieved by using various approaches such asbackground subtraction, optical flow and spatio-temporal filtering. Once detected, the object can be categorizedusing shape-based, motion-based, and texture-based classification techniques. Over the past decades, there has been rapid development in the evaluation and analysis of object detection and tracking. This review highlights a comprehensive and up to date comparisons on available techniques, the underlying ideas, and their limitations.

Key words:***Object detection; Segmentation; Background subtraction; Object classification***

1. Introduction

Videos are frames of images displayed sequentially in fast frequency so thatwe can sense continuity of its content. Detecting and tracking object is an important and fast-growing area of computer vision applications, including video surveillance, person tracking, vehicle navigation, and robotics. However, identification of a moving object is a challenging task because of variations in appearance and poses of the object, imaging quality, background and illumination conditions.Manycomputer vision systems take a static camera environment in which image taken from a video sequence is divided into foreground and background objects. Specifying standards to identify and mark foreground and background objects are difficult and problematic.But generally, foreground objects are moving objects and everything else is background [1].Starting witha low resolution image from video to high-level analysis, computer vision process proceeds through three steps: detection, tracking from frame to frame, and evaluating tracking objects[2]. Objects can be anything of interest for analysis and depicted by their shapes and appearances.Object shapes can be represented by object region, elliptical region, silhouette (contour), part-based, skeletal models[3]. The representation of a shape derived from shape boundary is called shape-signature.

2. Techniques

In general, there is a correlation between the object representation, detection, and tracking.The successful analysis of a moving object of interest from video footage depends on precise detection. The detection process

generally occurs in two steps: object detection and object classification. The schematic representation of object detection and classification is given in Figure 1.

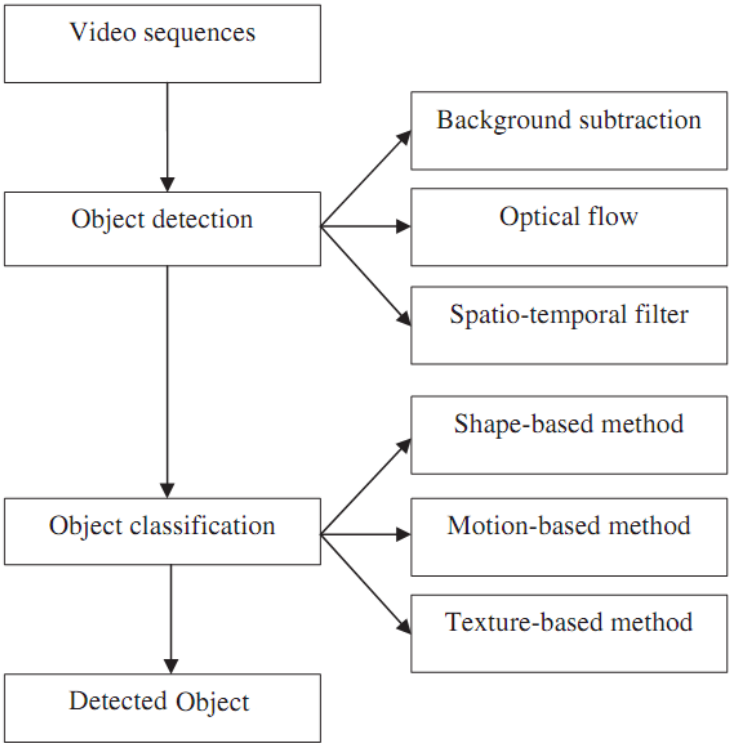

Figure 1. Flow chart of object detection and classification [4]

2.1. Object detection

Object detection comprises detection of moving objects, extraction or separation from immobilized background and recognizing the patterns. Segmentation, process of partitioning the image into some non-intersecting regions, is the essential step in detecting objects from in a video sequence[5]. Presently, most of the segmentation methods use eithertemporal or spatial information in the image sequence. Several conventional approaches arebackground subtraction, optical flow and spatio-temporal filtering method [4, 6].

2.1.1. Background subtraction

Background substation is widelyused method to detect moving object as foreground object from a static background. The rationale of this approach is todetectmoving objects from the difference between the currentframe and the reference frame (known as background image, background model or environment model) in a pixel-by-pixel orblock-by-block fashion. This is a three stage process; background initialization, foreground extraction, and background maintenance[7]. The methods of background subtraction to detect moving objects are well studied and reported in the literatures [8-10]. Extracting the foreground by subtracting a background image can be done by different methods, such as the running Gaussian average [11], temporal median filter[12], mixture of Gaussians (MoGs)[13, 14], kernel density estimation (KDE) [15, 16].Few available methods are discussed in the following section:

- **Mixture of Gaussian (MoGs) model**:Since single Gaussian model fails to accommodate the oscillating background, Stauffer and Grimson introduced a multi-label background using a mixture of Gaussians

(MoG)[13], which describes the probability of observing a certain pixel value, *x*, at time *f* by means of a mixture of Gaussians:

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} \times \eta(X_t, \mu_{i,t}, \sum_{i,t})$$

Where *K* is the number of distributions, $\omega_{i,t}$ is an estimate of the weight of the i^{th} Gaussian in the mixture at time *t*, $\mu_{i,t}$ is the mean value of the i^{th} Gaussian in the mixture at time *t*,$\Sigma_{i,t}$ is the covariance matrix of the i^{th} Gaussian in the mixture at time *t*, and η is a Gaussian probability density function.

Zivkovic and Heijden providedan algorithm to compute the correct number of Gaussian distributions at each pixel, based on their sample variation over time[17, 18].Shimada et al. presented analgorithm to control the number of Gaussiansadaptively in order to improve the computationaltime without losingquality of the background modelling[19].Wang et al. modelled each pixel by supportvector regression[20]. In another work, Ridder et al. used Kalman filter for adaptive background estimation and foreground detection[21].

- **Non-parametric model**:Sometimes, temporal pixel sequence at any location might not follow the default Gaussian distribution [22]. Unlike Gaussian model, the non-parametric background models do not assume prior shape distribution. Thus, many researchers introduced non-parametric model. These models consider the statistical behaviour of image features to segment the foreground from the background[4]. Elgammal et al. proposed model based on Kernel Density Estimation (KDE) to represent the colour distribution of each background pixel[15].Parag et al. proposed a boosting algorithm, namely RealBoost, to select appropriate features for the KDE methods [23]. Zhang and Xu suggested an approach that uses fuzzy integral to fuse thetexture and color features for background subtraction[24]. In another work, a fuzzy approach using the Choquet integral is used to get better results [25]. Kim and Kim proposed a clustering-based feature, called Fuzzy Colour Histogram (FCH) to construct the background model[26].

- **Frame differencing**:Frame diff erencing, popularly known as temporal diff erence, uses the difference between the video frames at time *t*-1 (as the background or reference image) and frame at time *t*[27]. It computes the pixel wise diff erence between the current frameand the previous frame to extract desired moving objects. If the resulting difference value is greater than a threshold value, then it can be referred as foreground object otherwise background.

$$Foreground\ pixel = (F_t - F_{t-1}) > T$$

Where, F_t is the current frame at time *t*, F_{t-1} is the previous (background) frame at time *t-1*, and *T* is the threshold value.

- **Warping background subtraction**:Ko et al. modeled the background as a set of canonical images to capture the different layers of background that appear or become occluded as background object moves. Foreground regions are thus defined as those that cannot be modelled by some composition of some warping of these background layers [28].

- **Hierarchical background model**: Chen et al. proposed a hierarchical background model (HBM) based on segmented background images and pixel descriptors to detect and track foreground. The HBM method first segments the background into different regions by the mean-shift algorithm and then builds region models and pixel models. Benefiting from the background segmentation, the region models and pixel models corresponding

to different regions can be set to different parameters. The pixel descriptors are calculated only from neighboring pixels belonging to the same object [29].

2.1.2. Optical flow

Theoptical flow technique is a vector based approachthat can be used to identify independently moving objects in a video such as in crowd analysis and conditions that contain dense motion [30]. The method uses characteristics of flow vectors of moving objects over time to detect moving regions in an image sequence [6].The key advantage is that it shows effective results to multiple objects as well as cameras when both are in motion[31]. This technique is very sensitive towards image noise, color non-uniform lighting and motion discontinuities. Most of the optical flow computation methods have large computational requirements[32].

2.1.3. Spatio-temporal filtering

Spatio-temporal filtering analysis is an unsupervised technique developed for detecting unusual activities in a large video set. This method generally considers motion as a whole to characterize its spatio-temporal distributions and utilize extremely simple features [33]. Zhong et al. proposed an algorithm which divides the video into equal length segments. Then detects moving objects and extract motion and colour/texture histogram for each frame, quantize every histogram into prototypes, and from which a prototype-segment co-occurrence matrix is calculated. The approach is efficient and stable, but are susceptible to noise and variations of the timings of the movements.

Aperformance analysis of different object detection methods based on the work of Paul et al. [4] is summarized in Table 1.

Table 1 Comparison of object detection methods

Methods	Accuracy	Computational time	Comments	Reference
Background subtraction				
Adaptive Gaussian Mixture model	Moderate	Moderate	- Easy execution and good performance - Can capture multi-modal states - Not well supportive for dynamic background	[13]
Non-parametric background model	Moderate to high	Low to moderate	- NP Performs well in dynamic background over MoGs - Does not perform well in occlusion situation - Requires post-processing	[15]
Frame differencing (Temporal differencing)	High	Low to moderate	- Good with sudden illumination changes in indoor environment	[34]
Warping background	High	Moderate to high	- Good in outdoor environment with high background motion - Computationally intensive in some variation	[28]
Hierarchical background model (HBM)	High	Low to moderate	- Uses block-based and pixel-based approaches - Pixel-based approach is quicker but quality is less	[29]

Optical flow	Moderate	High	- Performs well with cameramotion - Useful in crowd analysis - Computational time is very high	[32]
Spatio-Temporal filter	Moderate to high	Low to moderate	- Performs well with low-resolution states - Affected by noise	[33]

2.2. Object classification

A moving object needs to be classified accurately for its recognition and further analysis. The object classification methods can be divided into three categories: shape-based, motion-based and texture-based methods[4].

2.2.1. Shape-based method

For classifying moving objects, different descriptions of shape such as points,boxes, silhouettes and blobs are available.Accuracy of the classification depends on the type of classifier and the extracted object features. In general, shape-based recognition methods find the best match between these features with a priori statistics about the objects of interest. Kuno et al. used the extraction function based on the brightness transformation to extract the shapes of the human silhouette patterns[35].Lipton et al. used the dispersedness and area of image blobs asclassification metrics to classify all moving objects into three categories: humans, vehicles and clutter [36].Wang et al. investigated human silhouettes deformations during the articulated motionas discriminating features to implicitly capture motion dynamics. They used two signal transform methods i.e. Discrete Fourier Transform (DFT) andDiscrete Wavelet Transform (DWT) to characterize and recognize human motion sequences[37]. Generally motions are of two types; composite motions that can be divided into different temporal segments and then primitive motions that cannot be further decomposed. They focus on primitive motion recognition from short videos. Lin and Davis proposed a shape-based, hierarchical part template matching approach to match human shapes with images to detect and segment human simultaneously.In this approach, local part-based and global shape-template-based methods are combined to detect and segment humans from images[38].

2.2.2. Motion-based method

This classification method uses the periodic property of the images to recognize moving objects. The object motion characteristics and patterns are unique between objects provide a basic idea of classification. Bobick and Davis developed a view-based approach for the recognition of human movements. The basis of the representation is a temporal template, which is a static vector image where the vector value at each point is a function of motion properties at the corresponding spatial location in an image sequence[39]. Efros et al. introduced a new motion descriptor based on optical flow measurements over a spatio-temporal volume centered on a moving figure. Any residual motion within the spatio-temporal volume is due to the relative motions of different body parts such as limbs, head, torso etc.[31].

2.2.3. Texture-based method

In image processing, texture analysis plays an important role for classification or segmentation of images. A successful classification or segmentation requires an efficient description of image texture. Ojala et al. derived a generalized gray scale and rotation invariant operator. This is known as Local Binary Pattern (LBP). It is used in detecting the 'uniform' patterns for any quantization of the angular space. This operator is very efficient and used in multi resolution analysis[40].Dalal and Triggs introduced another texture-based method i.e. Histogram of Oriented Gradients (HOG) descriptor. This method provides excellent performance and counts the occurrences of gradient orientation in localized portions of an image. They adopt linear SVM based human detection as a test case for robust object recognition. First, gathered a data set which provide the human body deformations clearly, even in difficult illumination and then applied the HOG descriptor[41].By combining both the methods Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP) as the feature set, Wang et al. proposed an efficient human detection approach which is also capable of handling partial occlusion. The occlusion likelihood map is then segmented by 'Mean-Shift' approach. The main demerit of this method is the detector cannot handle the articulated deformation of people[42].

A comparison among object classification methods in terms of accuracy and computational time is presented in Table 2.

Table 2 Comparison of object classification methods

Methods	Accuracy	Computational time	Comments	Reference
Shape-based method	Moderate	Low	- It is a simple pattern-matching approach. - It does not work well in dynamic situations. - It is unable to determine internal movements well	[35-38]
Motion-based method	Moderate	High	- It does not require predefined pattern templates. - It fails to identify a non-moving human	[31, 39]
Texture-based method	High	High	- It provides improved quality with the expense of additional computation time	[40-42]

3. Conclusion

In this paper various detection phases for detecting an object viz. object detection and object classification has been studied. Available methods for these phases have been reviewed and a number of limitations and shortcomings were highlighted. The various method for object detection are background subtraction, optical flow and spatio-temporal filtering. Again back-ground subtraction method is performed in three different methods such as: adaptive Gaussian mixture model, non-parametric background model and frame differencing method. Object classification can be performed using shape-based, motion-based and texture-based method. It can be concluded that advanced study may be carried out to find more efficient algorithm to reduce computational cost and time.

References

1. Kamath, C. and S. Cheung, *Robust techniques for background subtraction in urban traffic video*, 2003, Lawrence Livermore National Laboratory (LLNL), Livermore, CA.

2. Porikli, F. and A. Yilmaz, *Object detection and tracking.* Video Analytics for Business Intelligence, 2012: p. 3-41.

3. Yilmaz, A., O. Javed, and M. Shah, *Object tracking: A survey.* Acm computing surveys (CSUR), 2006. **38**(4): p. 13.

4. Paul, M., S.M. Haque, and S. Chakraborty, *Human detection in surveillance videos and its applications-a review.* EURASIP Journal on Advances in Signal Processing, 2013. **2013**(1): p. 176.

5. Pal, N.R. and S.K. Pal, *A review on image segmentation techniques.* Pattern recognition, 1993. **26**(9): p. 1277-1294.

6. Hu, W., et al., *A survey on visual surveillance of object motion and behaviors.* IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2004. **34**(3): p. 334-352.

7. Kumar, S. and J.S. Yadav. *Background Subtraction Method for Object Detection and Tracking*. in *Proceeding of international conference on intelligent communication, control and devices*. 2017. Springer.

8. Piccardi, M. *Background subtraction techniques: a review*. in *Systems, man and cybernetics, 2004 IEEE international conference on*. 2004. IEEE.

9. Brutzer, S., B. Höferlin, and G. Heidemann. *Evaluation of background subtraction techniques for video surveillance*. in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. 2011. IEEE.

10. Bouwmans, T., *Recent advanced statistical background modeling for foreground detection-a systematic survey.* Recent Patents on Computer Science, 2011. **4**(3): p. 147-176.

11. Wren, C.R., et al., *Pfinder: Real-time tracking of the human body*. IEEE Transactions on pattern analysis and machine intelligence, 1997. **19**(7): p. 780-785.

12. Cucchiara, R., et al., *Detecting moving objects, ghosts, and shadows in video streams.* IEEE Transactions on pattern analysis and machine intelligence, 2003. **25**(10): p. 1337-1342.

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

14. Lee, D.-S., *Effective Gaussian mixture learning for video background subtraction.* IEEE Transactions on pattern analysis and machine intelligence, 2005. **27**(5): p. 827-832.

15. Elgammal, A., D. Harwood, and L. Davis, *Non-parametric model for background subtraction.* Computer Vision—ECCV 2000, 2000: p. 751-767.

16. Mittal, A. and N. Paragios. *Motion-based background subtraction using adaptive kernel density estimation*. in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. 2004. Ieee.

17. Zivkovic, Z. *Improved adaptive Gaussian mixture model for background subtraction*. in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. 2004. IEEE.

18. Zivkovic, Z. and F. Van Der Heijden, *Efficient adaptive density estimation per image pixel for the task of background subtraction.* Pattern recognition letters, 2006. **27**(7): p. 773-780.

19. Shimada, A., D. Arita, and R.-i. Taniguchi. *Dynamic control of adaptive mixture-of-Gaussians background model.* in *Video and Signal Based Surveillance, 2006. AVSS'06. IEEE International Conference on*. 2006. IEEE.

20. Wang, J., G. Bebis, and R. Miller. *Robust video-based surveillance by integrating target detection with tracking.* in *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*. 2006. IEEE.

21. Ridder, C., 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.

22. Choudhury, S.K., et al., *An evaluation of background subtraction for object detection vis-a-vis mitigating challenging scenarios.* IEEE Access, 2016. **4**: p. 6133-6150.

23. Parag, T., A. Elgammal, and A. Mittal. *A framework for feature selection for background subtraction*. in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. 2006. IEEE.

24. Zhang, H. and D. Xu, *Fusing color and texture features for background model.* Fuzzy Systems and Knowledge Discovery, 2006: p. 887-893.

25. El Baf, F., T. Bouwmans, and B. Vachon. *Foreground detection using the Choquet integral*. in *Image Analysis for Multimedia Interactive Services, 2008. WIAMIS'08. Ninth International Workshop on*. 2008. IEEE.

26. Kim, W. and C. Kim, *Background subtraction for dynamic texture scenes using fuzzy color histograms.* IEEE Signal processing letters, 2012. **19**(3): p. 127-130.

27. Cheung, S.-C.S. and C. Kamath. *Robust techniques for background subtraction in urban traffic video.* in *Proceedings of SPIE*. 2004.

28. Ko, T., S. Soatto, and D. Estrin. *Warping background subtraction.* in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. 2010. IEEE.

29. Chen, S., et al., *A hierarchical model incorporating segmented regions and pixel descriptors for video background subtraction.* IEEE Transactions on Industrial Informatics, 2012. **8**(1): p. 118-127.

30. Candamo, J., et al., *Understanding transit scenes: A survey on human behavior-recognition algorithms.* IEEE Transactions on Intelligent Transportation Systems, 2010. **11**(1): p. 206-224.

31. Efros, A.A., et al. *Recognizing action at a distance*. in *null*. 2003. IEEE.

32. Barron, J.L., D.J. Fleet, and S.S. Beauchemin, *Performance of optical flow techniques.* International journal of computer vision, 1994. **12**(1): p. 43-77.

33. Zhong, H., J. Shi, and M. Visontai. *Detecting unusual activity in video*. in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. 2004. IEEE.

34. Cheng, F.-C., S.-C. Huang, and S.-J. Ruan, *Scene analysis for object detection in advanced surveillance systems using Laplacian distribution model.* IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2011. **41**(5): p. 589-598.

35. Kuno, Y., et al. *Automated detection of human for visual surveillance system*. in *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*. 1996. IEEE.

36. Lipton, A.J., H. Fujiyoshi, and R.S. Patil. *Moving target classification and tracking from real-time video*. in *Applications of Computer Vision, 1998. WACV'98. Proceedings., Fourth IEEE Workshop on*. 1998. IEEE.

37. Wang, L., et al. *Moving shape dynamics: A signal processing perspective*. in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. 2008. IEEE.

38. Lin, Z. and L.S. Davis, *Shape-based human detection and segmentation via hierarchical part-template matching.* IEEE Transactions on pattern analysis and machine intelligence, 2010. **32**(4): p. 604-618.

39. Bobick, A.F. and J.W. Davis, *The recognition of human movement using temporal templates.* IEEE Transactions on pattern analysis and machine intelligence, 2001. **23**(3): p. 257-267.

40. Ojala, T., M. Pietikainen, and T. Maenpaa, *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns.* IEEE Transactions on pattern analysis and machine intelligence, 2002. **24**(7): p. 971-987.

41. Dalal, N. and B. Triggs. *Histograms of oriented gradients for human detection*. in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. 2005. IEEE.

42. Wang, X., T.X. Han, and S. Yan. *An HOG-LBP human detector with partial occlusion handling*. in *Computer Vision, 2009 IEEE 12th International Conference on*. 2009. IEEE.