A survey on human detection for video surveillance system

Dharmendrakumar Viradiya

Post Graduate Student
Computer Science & Engineering Department
B.H.Gardi College of Engineering & Technology Rajkot, Gujarat, India

***Abstract*- Detecting human beings perfectly in a visual surveillance system is crucial for diverse application areas including abnormal event detection, human gait characterization, person identification, gender classification and fall detection for elderly people. The first step of the detection process is to detect an object which is in motion. Object detection is used three methods: background subtraction, optical flow and spatio-temporal filtering techniques. Once detected, a moving object could be classified as a human being using shape-based, texture-based and motion-based features.**

***Index Terms*- Background subtraction, Motion based, Optical flow, Shape based, Spatio-temporal filtering, Texture based**

1. INTRODUCTION

Over the recent years, detecting human beings in a video scene of a surveillance system is attracting more attention due to its wide range of applications in abnormal event detection, person counting in a dense crowd, gender classification, person identification fall detection for old people, etc.

The scenes obtained from a surveillance video are usually with low resolution. Most of the scenes captured by a static camera are with minimal change of background. Objects in the outdoor surveillance are often detected in far field. Most existing digital video surveillance systems rely on human observers for detecting specific activities in a real-time video scene. However, there are limitations in the human capability to monitor simultaneous events in surveillance displays [1]. Hence, human motion analysis in automated video surveillance has become one of the most active and attractive research topics in the area of computer vision and pattern recognition.

The detection process generally occurs in two steps: object detection and object classification. Object detection could be performed by background subtraction, optical flow and spatio-temporal filtering. The object classification methods could be divided into three categories: shape-based, motion-based and texture-based.

1.1 Techniques

Human detection in a smart surveillance system aims at making distinctions among moving objects in a video sequence. The successful interpretations of higher level human motions greatly rely on the precision of human detection [2]. The detection process occurs in two steps: object detection and object classification.

1.1.1 Object detection

An object is generally detected by segmenting motion in a video image. Most conventional approaches for object detection are background subtraction, optical flow and spatio-temporal filtering method. They are outlined in the following subsections.

1.1.1.1 Background subtraction Background subtraction is a popular method to detect an object as a foreground by segmenting it from a scene of a surveillance camera. The camera could be fixed, pure translational or mobile in nature [3]. Background subtraction attempts to detect moving objects from the difference between the current frame and the reference frame in a pixel-by-pixel or block-by-block fashion. The reference frame is commonly known as ‘background image’, ‘background model’ or ‘environment model’. A good background model needs to be adaptive to the changes in dynamic scenes. Updating the background information in regular intervals could do this [4], but this could also be done without updating background information [5]. Few available

approaches have been discussed in this section:

- **Mixture of Gaussian model:** Stauffer and Grimson [6] introduced an adaptive Gaussian mixture model, which is sensitive to the changes in dynamic scenes derived from illumination changes, extraneous events, etc. Rather than modeling the values of all the pixels of an image as one particular type of distribution, they modeled the values of each pixel as a mixture of Gaussians. Over time, new pixel values update the mixture of Gaussian (MoG) using an online K-means approximation. In the literature, many approaches are proposed to improve the MoG [7]. In [8], authors presented an algorithm to control the number of Gaussians adaptively in order to improve the computational time without sacrificing the background modeling quality.
- **Temporal differencing:** The temporal differencing approach [9] involves three important modules: block alarm module, background modeling module and object extraction module. The block alarm module efficiently checked each block for the presence of either a moving object or background information. This was accomplished using temporal differencing pixels of the Laplacian distribution model and allowed the subsequent background modeling module to process only those blocks that were found to contain background pixels. Next, the background modeling module is employed in order to generate a high-quality adaptive background model using a unique two-stage training procedure and a mechanism for recognizing changes in illumination. As the final step of their process, the proposed object extraction module computes the binary object detection mask by applying suitable threshold values. This is accomplished using their proposed threshold training procedure.
- **Warping background:** Ko et al. [10] presented a background model that differentiates between background motion and foreground objects. Unlike most models that represent the variability of pixel intensity at a particular location in the image, they modeled the underlying warping of pixel locations arising from background motion. The background is modeled as a set of warping layers where at any given time, different layers may be visible due to the motion of an occluding layer. Foreground regions are thus defined as those that cannot be modeled by some composition of some warping of these background layers.

1.1.1.2 Optical flow Optical flow is a vector-based approach [11] that estimates motion in video by matching points on objects over image frame(s). Under the assumption of brightness constancy and spatial smoothness, optical flow is used to describe coherent motion of points or features between image frames. Optical flow-based motion segmentation uses characteristics of flow vectors of moving objects over time to detect moving regions in an image sequence. One key benefit of using optical flow is that it is robust to multiple and simultaneous cameras and object motions, making it ideal for crowd analysis and conditions that contain dense motion. Optical flow-based methods can be used to detect independently moving objects even in the presence of camera motion. Apart from their vulnerability to image noise, color and non-uniform lighting, most of flow computation methods have large computational requirements and are sensitive to motion discontinuities. A real-time implementation of optical flow will often require a specialized hardware due to the complexity of the algorithm and moderately high frame rate for accurate measurements [11].

1.1.1.3 Spatio-temporal filter For motion recognition based on spatio-temporal analysis, the action or motion is characterized via the entire 3D spatio-temporal data volume spanned by the moving person in the image sequence. These methods generally consider motion as a whole to characterize its spatio-temporal distributions [12]. Zhong et al. [12] processed a video sequence using a spatial Gaussian and a derivative of Gaussian on the temporal axis. Due to the derivative operation on the temporal axis, the filter shows high responses at regions of motion. These responses were then used to generate thresholds to yield a binary motion mask, followed by aggregation into spatial histogram bins. Such a feature encodes motion and its corresponding spatial information compactly and is useful for far-field and medium-field surveillance videos. As these approaches are based on simple convolution operations, they are fast and easy to implement. They are quite useful in scenarios with low-resolution or poor-quality video where it is difficult to extract other features such as optical flow or silhouettes. Spatio-temporal motion-based methods are able to better capture both spatial and temporal information of gait motion. Their advantage is low computational complexity and a simple implementation. However, they are susceptible to noise and to variations of the timings of movements.

1.1.2 Object classification

An object in motion needs to be classified accurately for its recognition as a human being. The available classification methods could be divided into three main categories: shape-based method, motion-based method and texture-based method.

1.1.2.1 Shape-based method Shape-based approaches first describe the shape information of moving regions such as points, boxes

and blobs. Then, it is commonly considered as a standard pattern recognition issue [11]. However, the articulation of the human body and differences in observed viewpoints lead to a large number of possible appearances of the body, making it difficult to accurately distinguish a moving human from other moving objects using the shape-based approach. Eishita et al. [13] proposed a simple but effective method for object tracking after full or partial occlusion using shape, color and texture information even if the color and textures are the same for the objects. Wang et al. [14] investigated how the deformations of human silhouettes (or shapes) during articulated motion could be used as discriminating features to implicitly capture motion dynamics and exploited the applicability of discrete wavelet transform and DFT for the purpose of human motion characterization and recognition.

1.1.2.2 Motion-based method This classification method is based on the idea that object motion characteristics and patterns are unique enough to distinguish between objects. Motion-based approaches directly make use of the periodic property of the captured images to recognize human beings from other moving objects. Bobick and Davis [15] developed a view-based approach for the recognition of human movements by constructing a vector image template comprising two temporal projection operators: binary motion-energy image and motion-history image. Cutler et al. [16] presented a self-similarity-based time-frequency technology to detect and analyze periodic motion for human classification. Unfortunately, methods based on periodicity are restricted to periodic motion. Efros et al. [17] characterized the human motion within a spatio-temporal volume by a descriptor, which was based on computing the optical flow, projecting the motion onto a number of motion channels and blurring with a Gaussian. Recognition was performed in a nearest-neighbor framework. By computing a spatio-temporal cross correlation with a stored database of previously labeled action fragments, the most similar to the motion descriptor of the query action fragment could be found.

1.1.2.3Texture-based method Local binary pattern (LBP) is a texture-based method that quantifies intensity patterns in the neighborhood of the pixel [18]. Zhang et al. [19] proposed the multi-block local binary pattern (MB-LBP) to encode intensities of the rectangular regions by LBP. HOG [20] introduced another texture-based method which uses high-dimensional features based on edges and then applies SVM to detect human body regions. This technique counts the occurrences of gradient orientation in localized portions of an image, is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy. Zhu et al. [21] applied the HOG descriptors in combination with the cascade of rejecters algorithm and introduced blocks that vary in size, location and aspect ratio. In order to isolate the blocks best suited for human detection, they applied the AdaBoost algorithm to select those blocks to be included in the rejecter cascade. Moctezuma et al. [22] proposed HOG with Gabor filter and showed improved performances in both person counting and identification.

2. CONCLUSIONS

Detecting human beings accurately in a surveillance video is one of the major topics of vision research due to its wide range of applications. It is challenging to process the image obtained from a surveillance video as it has low resolution. A review of the available detection techniques is presented. The detection process occurs in two steps: object detection and object classification. In this paper, all available object detection techniques are categorized into background subtraction, optical flow and spatio-temporal filter methods. The object classification techniques are categorized into shape-based, motion-based and texture-based methods. The characteristics of the benchmark datasets are presented, and major applications of human detection in surveillance video are reviewed.

3.

A discussion is made to point the future work needed to improve the human detection process in surveillance videos. These include exploiting a multi-view approach and adopting an improved model based on localized parts of the image.

3. ACKNOWLEDGEMENT

I am deeply indebted & would like to express gratitude to my thesis guide Prof. Amit Maru , B. H. Gardi College of Engineering & Technology for his great efforts and instructive comments in the dissertation work.

I would also like to extend my gratitude to Prof.Hemal Rajyaguru, Head of the Computer Science & Engineering Department, B. H. Gardi College of Engineering & Technology for his continuous encouragement and motivation.

I would also like to extend my gratitude to Prof. Vaseem Ghada, PG Coordinator, B. H. Gardi College of Engineering & Technology for his continuous support and cooperation.

I should express my thanks to my dear friends & my classmates for their help in this research; for their company during the research, for their help in developing the simulation environment.

I would like to express my special thanks to my family for their endless love and support throughout my life. Without them, life would not be that easy and beautiful.

REFERENCES

[1] N Sulman, T Sanocki, D Goldgof, R Kasturi," How effective is human video surveillance performance?" in 19th International Conference on Pattern Recognition, (ICPR 2008) (IEEE, Piscataway, 2008), pp. 1–3

[2] G Lavee, E Rivlin, M Rudzsky, "Understanding video events: a survey of methods for automatic interpretation of semantic occurrences in video". IEEE Trans. Syst., Man, Cybern. C 39(5), 489–504 (2009)

[3] W Hu, T Tan, L Wang, S Maybank, A survey on visual surveillance of object motion and behaviors. IEEE Trans. Syst., Man, Cybern. Part C, Appl. Rev. 34(3), 334–352 (2004)

[4] H-H Lin, T-L Liu, J-H Chuang, "Learning a scene background model via classification". IEEE Trans. Signal Process. 57(5), 1641–1654 (2009)

[5] T Du-Ming, L Shia-Chih, "Independent component analysis-based back-ground subtraction for indoor surveillance". IEEE Trans. Image Process. 18(1), 158–167 (2009)

[6] C Stauffer, W Grimson, "Adaptive background mixture models for real-time tracking", in IEEE Conference on Computer Vision and Pattern Recognition (CVPR 1999) (IEEE, Piscataway, 1999), pp. 246–252

[7] DS Lee, "Effective Gaussian mixture learning for video background subtraction". IEEE Trans. Pattern Anal. Mach. Intell. 27(5), 827–835 (2005)

[8] A Shimada, D Arita, "Dynamic control of adaptive mixture-of-Gaussians background model", in IEEE International Conference on Video and Signal Based Surveillance (AVSS'06) (IEEE, Piscataway, 2006), p. 5

[9] F-C Cheng, S-C Huang, S-J Ruan, "Scene analysis for object detection in advanced surveillance systems using Laplacian distribution model". Syst. Man Cybern. Part C Appl. Rev. IEEE Trans. 41(5), 589–598 (2011)

[10] T Ko, S Soatto, D Estrin, "Warping background subtraction", in 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010) (IEEE, Piscataway, 2010), pp. 1331–1338

[11] J Candamo, M Shreve, DB Goldgof, DB Sapper, R Kasturi, "Understanding transit scenes: A survey on human behavior-recognition algorithms". IEEE Trans. Intell. Transp. Syst. 11(1), 206–224 (2010)

[12] H Zhong, J Shi, M Visontai, "Detecting unusual activity in video", in 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004) (IEEE, Piscataway, 2004), pp. 819–826

[13] FZ Eishita, A Rahman, SA Azad, A Rahman, "Occlusion handling in object detection. Multidisciplinary computational intelligence techniques: applications in business, engineering, and medicine". IGI Global. (2013). doi: 10.4018/978-1-4666-1830-5.ch005

[14] L Wang, X Geng, C Leckie, R Kotagiri, "Moving shape dynamics: a signal processing perspective", in IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008) (IEEE, Piscataway, 2008), pp. 1–8

[15] AF Bobick, JW Davis, "The recognition of human movement using temporal templates". IEEE Trans. Pattern Anal. Mach. Intell. 23(3), 257–267 (2001)

[16] R Cutler, LS Davis, "Robust real-time periodic motion detection, analysis, and applications". IEEE Trans. Pattern Anal. Mach. Intell. 22(8), 781–796 (2000)

[17] A Efros, A Berg, G Mori, J Malik, "Recognizing action at a distance", in Ninth IEEE International Conference on Computer Vision (ICCV 2003) (IEEE, Piscataway, 2003), pp. 726–733

[18] T Ojala, M Pietikinen, T Maenpaa, "Multi-resolution grayscale and rotation invariant texture classification with local binary patterns". PAMI 24(7), 971–987 (2002)

[19] L Zhang, SZ Li, X Yuan, S Xiang, "Real-time object classification in video surveillance based on appearance learning", in IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2007 (CVPR 2007) (IEEE, Piscataway, 2007), pp. 1–8

[20] N Dalal, B Triggs, "Histograms of oriented gradients for human detection", in IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005) (IEEE, Piscataway, 2005), pp. 886–893

[21] Q Zhu, S Avidan, M-C Yeh, K-T Cheng, "Fast human detection using a cascade of histograms of oriented gradients", in IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2006 (CVPR '06) (IEEE, Piscataway, 2006), pp. 1491–1498

[22] D Moctezuma, C Conde, IM Diego, E Cabello, "Person detection in surveillance environment with HoGG: Gabor filters and histogram of oriented gradient", in ICCV Workshops (IEEE, Piscataway, 2011), pp. 1793–1800