ICIC Express Letters
Part B: Applications
Volume **7**, Number **6**, June **2016**

ICIC International ©2016 ISSN 2185-2766
pp. 1359–1364

MULTI-TARGET SEGMENTATION IN A MULTI-CAMERA SURVEILLANCE ENVIRONMENT

Cheng-Hung Chuang[1], Chao-Ching Lee[2], Ye-Chi Wu[2], Kuan-Kai Huang[4,5]
Ching-Kan Lo[1,2,3] and Yung-Chen Chou[1,*]

[1]Department of Computer Science and Information Engineering
Asia University
No. 500, Liufeng Rd., Wufeng, Taichung 41354, Taiwan
chchuang@asia.edu.tw; drsimonlo@gmail.com; *Corresponding author: yungchen@gmail.com

[2]Department of IT
[3]Orthopedic Department
[4]Department of Administration, Show Chwan Memorial Hospital
Show Chwan Health Care System
No. 542, Sec. 1, Chung-shan Rd., Changhua 50008, Taiwan
{ johnson10723; yechiwu; hydokykk }@gmail.com

[5]Department of Information Management
Chien-Kuo Technology University
No. 1, Chiehshou North Rd., Changhua 50094, Taiwan

Received November 2015; accepted February 2016

Abstract. *This paper presents a multi-target segmentation system in a multi-camera surveillance environment, which can be used in smart video surveillance applications. The system is composed of three subsystems, i.e., the foreground and background segmentation subsystem (FBSS), the human body detection subsystem (HBDS), and the human feature matching subsystem (HFMS). By means of the integration of these subsystems, moving body detection in the multi-sensor images of the same scene using video cameras is performed. Experimental results show that the proposed system provides acceptable moving body detection rate.*

Keywords: Video segmentation, Foreground-background segmentation, Moving object segmentation, Object tracking, Video surveillance

1. **Introduction.** Currently, video cameras are widely used for detecting and monitoring the behavior, activities, or other status information of people or objects in an indoor or outdoor environment to ensure human safety. Various smart video surveillance technologies have been developed for proactive security [1], e.g., real-time threat alerts [2], rapid video search and retrieval [3], object detection and tracking [4], and object classification [5]. Moreover, many smart home technologies have also been developed to detect or monitor activities of people at home and help their independent living [6]. The motion detection in the extraction of moving objects from a video sequence captured by a video camera is a crucial step in smart video surveillance [7]. A large variety of image and video processing techniques exist for the extraction of moving objects in video sequences [6-20].

Block matching and optical flow algorithms [8] that are widely used to find changes in object intensity by successive frames are the two major motion estimation techniques. Because the computational cost of these methods is high, it is seldom used in real-time application. In [9], a camera hand-off filter in a real-time surveillance tracking system was proposed for tracking individual across cameras via a surveillance network. In [10], a skin detection method to extract facial features for human tracking in a video surveillance system was presented. In [11], three probabilistic models, i.e., Naive Bayes classifier, forward procedure of a Hidden Markov Model and Viterbi algorithm based on a Hidden

Markov Model, were proposed to represent and recognize human activities from observed sensor sequences. In [12], the off-time swimming pool surveillance using thermal imaging system to enhance the security of the surrounding area was proposed. In [13], a set of moving points is used and clustered according to their positions and color and motion around them. Finally, a segmentation of moving objects is obtained from each of these clusters of moving points in complex scenes. Active contour model [14,15] that is used to accurately detect the irregular object boundaries is also employed to segment moving objects. However, this method requires an initial contour for the initialization settings and needs a long computational time to achieve the energy convergence.

The temporal or frame differencing [6,16] and background subtraction [17,18] are two common methods for motion or moving object detection. The temporal differencing method computes the difference between adjacent frames in a video sequence to identify the moving or stationary regions for motion detection. This method is suitable for application in the real-time moving object detection due to its fast computation. However, this method is very sensitive to noise, highly affected by the light illumination, and not suitable for the video sequences with a moving background. Background subtraction, which can construct a reference background in advance and subtract it in the video frame to extract accurate foreground objects, is easily implemented and widely used in practical application. However, it is very difficult to produce the exact background for subtraction and also not suitable for the video sequences with a moving background.

In our previous study [6], the single moving human body detection was achieved using frame differencing and background subtraction methods in which the background images had to be prior captured manually. However, it is essential to perform automatic background subtraction and detect multiple targets for action recognition and event detection. Therefore, we developed a multi-target segmentation system based on the codebook model to detect moving bodies for video surveillance. The codebook model [19,20] for automatic and real-time foreground and background segmentation is efficient in memory and speed compared with other background modeling techniques. It can handle scenes containing moving backgrounds or illumination variations, and achieve robust detection for different types of videos. Three subsystems are integrated in our multi-target segmentation system, i.e., the foreground and background segmentation subsystem (FBSS), the human body detection subsystem (HBDS), and the human feature matching subsystem (HFMS).

The content of this paper is presented as follows. Section 2 describes the proposed multi-target segmentation system. Section 3 covers some experimental results to demonstrate the performance of the proposed system. Finally, Section 4 concludes this study.

2. The Proposed System.

In a large space, the monitoring is mostly performed with a set of multiple cameras which can cover the environment due to the limitation of the camera view range. Another benefit of the use of multiple cameras is to handle occlusions of persons by other persons or objects. Figure 1 shows a scenario of video surveillance for monitoring multi-target in a multi-camera environment. Therefore, we developed a multi-target segmentation system in a multi-camera surveillance environment. The distinguishing contribution of our study is listed as follows.

1. The foreground and background segmentation is performed by means of the codebook background model to achieve automatic background subtraction. This segmentation model can handle scenes containing illumination variations and extract the regions of moving bodies.
2. The video frame is transformed from RGB to Lab color space. Thus the illumination and color features can be separated using the Lab color model. The color feature of moving body regions can be calculated by the ab components and used to identify the different targets.

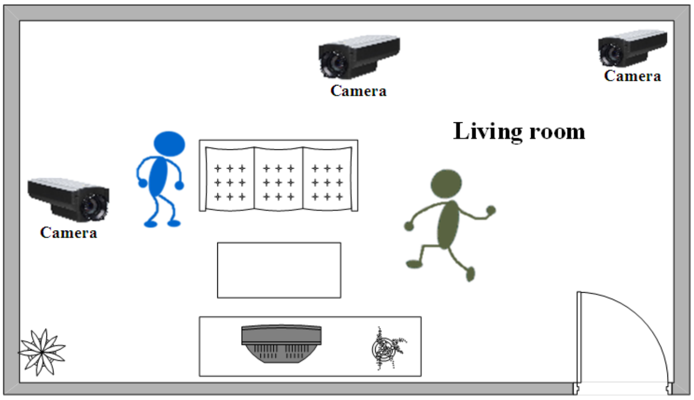

Figure 1. A scenario of multi-target in a multi-camera surveillance environment

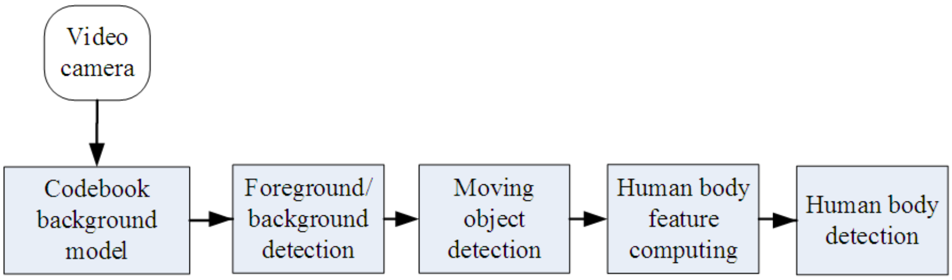

Figure 2. The flowchart of human body detection procedure

Figure 2 shows the flowchart of human body detection procedure. The aim of this procedure is to perform a strategy for dealing with the detection of human bodies. Therefore, the proposed system can be separated into three subsystems, i.e., the foreground and background segmentation subsystem (FBSS), the human body detection subsystem (HBDS), and the human feature matching subsystem (HFMS).

2.1. **Foreground and background segmentation subsystem (FBSS).** The FBSS is composed of an image smoothing processing, a codebook training procedure, a foreground extraction module, and a connected component analysis. The image smoothing processing is applied to reduce noise effect. The codebook training procedure is used to model a background from a period of video sequence. The pixel is detected as foreground by using the background subtraction. Then the connected component analysis is used to make the foreground regions more complete.

In the codebook training procedure, it builds a codebook consisting of one or more codewords for each pixel. Let X be a training sequence for a single pixel consisting of n RGB-vectors, i.e., $X = \{x_1, x_2, \ldots, x_n\}$. Let $C = \{c_1, c_2, \ldots, c_L\}$ represent the codebook for the pixel consisting of L codewords. Each pixel has a different codebook size L based on its sample variation. Each codeword c_i, $i = 1, 2, \ldots, L$ consists of an RGB vector $v_i = \left(\overline{R}_i, \overline{G}_i, \overline{B}_i\right)$ and a 6-tuple $aux_i = \left\{\breve{I}_i, \widehat{I}_i, f_i, \lambda_i, p_i, q_i\right\}$, where $\breve{I}_i$ and $\widehat{I}_i$ are the minimum and maximum brightness of all pixels assigned to this codeword, respectively, f_i is the frequency with which the codeword has occurred, λ_i is the maximum negative run-length defined as the longest interval during the training period that the codeword has not recurred, and p_i and q_i are the first and last access times that the codeword has occurred, respectively.

In the initial setting, $L = 0$ and $C = \{\}$. For the first sampled time $t = 1$, $t \in \{1, 2, \ldots, n\}$, the codeword c_1 is computed. For the each other sampled time $t = 2, 3, \ldots, n$, the RGB vector and the brightness are computed, i.e., $x_t = (R, G, B)$ and $I_t = \sqrt{R^2 + G^2 + B^2}$.

They are compared to the current codebook to determine which codeword it matches by the following rules.

$$dist(x_t, v_m) = \sqrt{\left(R - \overline{R}_m\right)^2 + \left(G - \overline{G}_m\right)^2 + \left(B - \overline{B}_m\right)^2} \leq \varepsilon \tag{1}$$

$$brightness\left(I_t, \left[\breve{I}_m, \widehat{I}_m\right]\right) = \begin{cases} \text{true}, & \text{if } \breve{I}_m < I_t < \widehat{I}_m \\ \text{false}, & \text{otherwise} \end{cases} \tag{2}$$

If the matched codeword exists, the matched codeword c_m is updated by the following functions. If there is no matched codeword, then create a new codeword.

$$v_m = \left(\frac{f_m\bar{R}_m + R}{f_m + 1}, \frac{f_m\bar{G}_m + G}{f_m + 1}, \frac{f_m\bar{B}_m + B}{f_m + 1}\right) \tag{3}$$

$$aux_m = \left\{\min\left(I_t, \breve{I}_i\right), \max\left(I_t, \widehat{I}_i\right), f_m + 1, \max(\lambda_m, t - q_m), p_m, t\right\} \tag{4}$$

After the training procedure, the codebook for all pixels is completed. Then a detection threshold ε_{TH} can be determined to perform the foreground and background segmentation. A pixel is detected as foreground if there is no match by the following rules; otherwise it is detected as background.

$$dist(x, c_m) \leq \varepsilon_{TH} \tag{5}$$

$$brightness\left(I, \left[\breve{I}_m, \widehat{I}_m\right]\right) = \begin{cases} \text{true}, & \text{if } \breve{I}_m < I < \widehat{I}_m \\ \text{false}, & \text{otherwise} \end{cases} \tag{6}$$

2.2. **Human body detection subsystem (HBDS).** After the foreground and background segmentation, the foreground regions are extracted and their illumination and color features are separated using the Lab color model. First, the current video frame is transformed from RGB to Lab color space. The color feature of moving body regions is calculated by the ab components and used to identify the different human bodies. Figure 2 shows the block diagram of the HBDS subsystem. The human feature is computed by the following equation.

$$F_{i,t} = \left(\frac{1}{N}\sum_{(x,y)\in\Omega_i} a_{i,t}(x, y), \frac{1}{N}\sum_{(x,y)\in\Omega_i} b_{i,t}(x, y)\right) \tag{7}$$

where $F_{i,t}$ denotes the ith human feature in the current video frame t, $a_{i,t}$ and $b_{i,t}$ represent the ab components of ith human feature in the current video frame t, Ω_i represents the pixel set of ith human region, and N is the total number of pixels in the ith human region.

2.3. **Human feature matching subsystem (HFMS).** After the human body detection and human feature computation, the human features are matched by the HFMS. The goal of HFMS is to develop a simple strategy for finding human body in the video sequence. In this subsystem, the difference of different human features is computed. Then the mask of corresponding human features of the current video frame t and the previous video frame $t-1$, is determined as follows.

$$D_{i,j} = |F_{i,t} - F_{j,t-1}| \tag{8}$$

$$M_{i,j} = \begin{cases} 1, & \text{if } D_{i,j} < TH_D \\ 0, & \text{if } D_{i,j} \geq TH_D \end{cases} \tag{9}$$

where $D_{i,j}$ denotes the difference of corresponding ith and jth human features of the current video frame t and the previous video frame $t-1$, respectively, TH_D is a threshold value, $M_{i,j}$ represents the mask of corresponding ith and jth human features of the current video frame t and the previous video frame $t-1$, respectively.

Table 1. Accuracy of human body detection by the proposed and active contours methods

Video sequence	Proposed method			Active contours method [15]		
	Maximal accuracy (%)	Minimal accuracy (%)	Average accuracy (%)	Maximal accuracy (%)	Minimal accuracy (%)	Average accuracy (%)
No. 1	99.64	97.82	99.07	96.38	91.67	93.18
No. 2	99.50	95.54	98.08	94.62	89.23	91.44
No. 3	99.15	97.57	98.37	95.93	90.56	92.73

Figure 3. Sampled video frames of human body detection (the top row shows original video frames, the second row shows standard segmented video frames, the third row shows the results of the proposed method, and the bottom row shows the results of the active contours method [15])

3. **Experimental Results.** In this section, we present the experimental results of the proposed multi-target segmentation system. We set up three cameras in our laboratory and two persons were walking around the laboratory. There are three video sequences captured by the three cameras. To verify the performance of the proposed method, these video sequences are manually segmented to obtain the human body regions for the standard evaluation. The active contour model [15] is also performed for comparison. The accuracy which is the proportion of both true positives and true negatives is computed for evaluation. Table 1 shows the accuracy of human body detection (foreground and background segmentation) by the proposed and active contours methods. The maximal, minimal and average of accuracy are evaluated for comparison. It is difficult to completely segment human arms or legs. However, we can obtain an average accuracy of 98.51% in the test video sequences. The average accuracy is only 92.45% using the active contours method. Figure 3 shows the sampled video frames and detection results. The first row shows the sampled original video frames. The second row shows the sampled manually segmented video frames. The third and last rows show the sampled resultant segmented video frames using the proposed method and the active contours method, respectively. It is obvious that the human arms or legs are failed to detect in the results of active contours method.

4. **Conclusions.** In this paper, a multi-target segmentation system, which is now designed for monitoring indoor environments, is proposed for human body detection. The proposed system is composed of three subsystems, i.e., the foreground and background

segmentation subsystem (FBSS), the human body detection subsystem (HBDS), and the human feature matching subsystem (HFMS). Experimental results show that the proposed system performs well for human body detection. Future work will mainly focus on the detection and tracking of more targets.

Acknowledgment. This work was supported by the Ministry of Science and Technology of Taiwan under Grant No. MOST 103-2221-E-468-008-MY2.

REFERENCES

[1] A. Hampapur, Smart video surveillance for proactive security, *IEEE Signal Processing Magazine*, vol.25, no.4, pp.136-134, 2008.

[2] H. Gupta, L. Yu, A. Hakeem, T. E. Choe, N. Haering and M. Locasto, Multimodal complex event detection framework for wide area surveillance, *Proc. of 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp.47-54. 2011.

[3] Z. Xiong, R. Radhakrishnan, A. Divakaran, Y. Rui and T. S. Huang, *A Unified Framework for Video Summarization, Browsing & Retrieval: With Applications to Consumer and Surveillance Video*, Academic Press, 2006.

[4] Z. Musa and J. Watada, Video tracking system: A survey, *ICIC Express Letters*, vol.2, no.1, pp.65-72, 2008.

[5] Y. Gurwicz, R. Yehezkel and B. Lachover, Multiclass object classification for real-time video surveillance systems, *Pattern Recognition Letters*, vol.32, no.6, pp.805-815, 2011.

[6] C.-H. Chuang and Z.-Y. Lian, A video surveillance system for home care applications, *ICIC Express Letters*, vol.8, no.4, pp.1111-1118, 2014.

[7] S.-C. Huang, An advanced motion detection algorithm with video quality analysis for video surveillance systems, *IEEE Trans. Circuits and Systems for Video Technology*, vol.21, no.1, pp.1-14, 2011.

[8] J. T. Philip, B. Samuvel, K. Pradeesh and N. K. Nimmi, A comparative study of block matching and optical flow motion estimation algorithms, *Proc. of 2014 Annual International Conference on Emerging Research Areas: Magnetics, Machines and Drives*, Kottayam, India, pp.1-6, 2014.

[9] C.-Y. Lee, S.-J. Lin, C.-W. Lee and C.-S. Yang, An efficient camera hand-off filter in real-time surveillance tracking system, *International Journal of Innovative Computing, Information and Control*, vol.8, no.2, pp.1397-1417, 2012.

[10] I. Hemdan, S. Karungaru and K. Terada, Video surveillance using facial features-based tracking, *International Journal of Innovative Computing, Information and Control*, vol.8, no.4, pp.2761-2776, 2012.

[11] H. Fang, R. Srinivasan and D. J. Cook, Feature selections for human activity recognition in smart home environments, *International Journal of Innovative Computing, Information and Control*, vol.8, no.5(B), pp.3525-3535, 2012.

[12] W. K. Wong, J. H. Hui, C. K. Loo and W. S. Lim, Thermal imaging based off-time swimming pool surveillance system, *International Journal of Innovative Computing, Information and Control*, vol.9, no.3, pp.1293-1320, 2013.

[13] A. Bugeau and P. Pérez, Detection and segmentation of moving objects in complex scenes, *Computer Vision and Image Understanding*, vol.113, no.4, pp.459-476, 2009.

[14] C. Ma, H. Yang, X. Li, Y. Ling, D. Wu and J. Wang, Multi-target segmentation in multi-sensor images in complex scenes, *Opto-Electronic Engineering*, vol.36, no.1, 2009.

[15] C.-H. Chuang, Y.-L. Chao and Z.-P. Li, Moving object segmentation and tracking using active contour and color classification models, *Proc. of the IEEE International Symposium on Multimedia*, Taichung, Taiwan, pp.73-80, 2010.

[16] T. Barbu, Pedestrian detection and tracking using temporal differencing and HOG features, *Computers & Electrical Engineering*, vol.40, no.4, pp.1072-1079, 2014.

[17] R. Cucchiara, C. Grana, M. Piccardi and A. Prati, Detecting moving objects, ghosts, and shadows in video streams, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.25, no.10, pp.1337-1342, 2003.

[18] O. Barnich and M. V. Droogenbroeck, ViBe: A universal background subtraction algorithm for video sequences, *IEEE Trans. Image Processing*, vol.20, no.6, pp.1709-1724, 2011.

[19] K. Kim, T. H. Chalidabhongse, D. Harwood and L. S. Davis, Real-time foreground-background segmentation using codebook model, *Real-Time Imaging*, vol.11, no.3, pp.172-185, 2005.

[20] J.-M. Guo, Y.-F. Liu, C.-H. Hsia, M.-H. Shih and C.-S. Hsu, Hierarchical method for foreground detection using codebook model, *IEEE Trans. Circuits and Systems for Video Technology*, vol.21, no.6, pp.804-815, 2011.