Carried Object Detection Using Ratio Histogram and its Application to Suspicious Event Analysis

Chi-Hung Chuang, Jun-Wei Hsieh, Luo-Wei Tsai, Sin-Yu Chen, and Kuo-Chin Fan

***Abstract*— This letter proposes a novel method to detect carried objects from videos and applies it for analysis of suspicious events. First of all, we propose a novel kernel-based tracking method for tracking each foreground object and further obtaining its trajectory. With the trajectory, a novel ratio histogram is then proposed for analyzing the interactions between the carried object and its owner. After color re-projection, different carried objects can be then accurately segmented from the background by taking advantages of Gaussian mixture models. After bag detection, an event analyzer is then designed to analyze various suspicious events from the videos. Even though there is no prior knowledge about the bag (such as shape or color), our proposed method still performs well to detect these suspicious events. As we know, due to the uncertainties of the shape and color of the bag, there is no automatic system that can analyze various suspicious events involving bags (such as robbery) without using any manual effort. However, by taking advantages of our proposed ratio histogram, different carried bags can be well segmented from videos and applied for event analysis. Experimental results have proved that the proposed method is robust, accurate, and powerful in carried object detection and suspicious event analysis.**

***Index Terms*— Carried bag detection, finite state machines, gaussian mixture models, ratio histogram, suspicious event detection.**

I. Introduction

DETECTION OF CARRIED objects [1], [2] in video sequences is an important task in video surveillance. It can be used for security monitoring, crime detection, and anti-terrorist surveillance. Its challenge comes from the gap between the high-level semantic concepts of events and machine-understandable low-level features extracted from videos. Since humans are the most concerning target over other objects, most approaches that have been proposed for bridging the gap almost focus on human behavior (or action) analysis. For example, Oliver *et al.* [3] developed a visual surveillance system that models and recognizes human behavior using hidden Markov models (HMMs). Cucchiara *et al.* [5] proposed a probabilistic posture classification scheme for classifying human behaviors to different types such as walking, running, squatting, or sitting. Park and Aggarwal [6] proposed a dynamic Bayesian network to segment a body into different parts and then used HMMs to analyze various interactions among persons. In addition to individual human behaviors, dangerous accidents are sometimes caused by abandoned objects such as a packet bomb. In public places (such as airports, train stations, or bus terminals), people will usually carry different types of suitcases, luggage, or bags. The owner of the object would possibly leave his/her package in a public place. These abandoned objects are usually considered potential security breaches. Thus, in the literature, there have been several approaches [10], [11] proposed for the detection of abandoned objects. For example, Stringa [10] located moving objects and then recognized suspicious objects from their background according to their positions, moments, and shape features. In [11], Foresti *et al.* used a long-term change detection algorithm to detect abandoned objects and then classified video sequences to four dangerous events. In addition to detection of abandoned objects, another challenging work is to analyze the interactions between objects and humans. Due to the difficulty of object segmentation, there are very few works addressing this topic. In [2], Haritaoglu *et al.* presented a silhouette-based approach to analyze the symmetry of body shape for determining whether a person carried an object, e.g., backpack. However, the silhouette feature is not good for detecting other objects (such as a handbag) and is easily vulnerable to noise.

This letter proposes a novel approach to detect carried objects and analyze their transferring conditions for suspicious event analysis. In public places, people will carry different bags or packages for traveling, shopping, or other private purposes. When a suspicious event happens, the carried bags or packages will undergo various changes in their condition (like being transferred or left). For example, in a robbery event, the carried object will be transferred from the victim to the robber. For a terrorism event, the carried package will be on purpose left at an airport. To detect these events, the object's transference conditions should be first detected. However, this task is difficult since there is no prior knowledge about the object's shape and color. The major contribution of this letter is to present a novel histogram-based framework for detecting the conditions and then segmenting each carried object from videos irrespective of its shape and color. Fig. 1 shows the flowchart of the whole system. Assume that the video is captured by a still camera. The system first uses a subtraction technique to detect the foreground objects from

Manuscript received September 2, 2007; revised May 9, 2008. First version published March 16, 2009; current version published June 19, 2009. This work was supported in part by the National Science Council of Taiwan under Grant NSC96-2221-E-155-021, Taiwan, and the Ministry of Economic Affairs under contract no. 97-EC-17-A-02-S1-032. This paper was recommended by Associate Editor D. Schonfeld.

Chi-Hung Chuang, Luo-Wei Tsai, and Kuo-Chin Fan are with the Department of Computer Engineering, National Central University, Chung-Li 320, Taiwan (e-mail: chi_hung_chuang@hotmail.com; echoo@fox1.csie.ncu.edu.tw; kcfan@csie.ncu.edu.tw).

Jun-Wei Hsieh and Sin-Yu Chen are with the Department of Electrical Engineering, Yuan Ze University, Chung-Li 320, Taiwan (e-mail: shieh@saturn.yzu.edu.tw; s958507@mail.yzu.edu.tw).

Digital Object Identifier 10.1109/TCSVT.2009.2017415

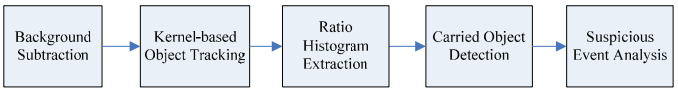

Fig. 1. Flowchart of the proposed system to detect and analyze suspicious objects and events.

videos. Then, a novel kernel-based tracking technique is proposed for tracking each foreground object. Then, a novel ratio histogram is proposed for finding all missing colors between two pedestrians if they have interactions. After detecting the missing colors, the location of each carried object can be easily found using a color re-projection technique. Furthermore, we take advantage of a set of Gaussian mixture models (GMMs) to segment the object more accurately from its background. By checking its subsequent speed and direction, an event analyzer can be then designed for the suspicious event analysis. The experimental results demonstrate the feasibility and superiority of our method in suspicious event detection.

The remainder of the letter is organized as follows. In Section II, the kernel-based tracking technique is described. In Section III, the techniques of ratio histogram extraction and missed color detection are proposed. Then, details of object modeling are discussed in Section IV. In Section V discusses our suspicious event analyzer. The experiment results are given in Section VI. We then present our conclusions in Section VII.

II. Kernel-Based Object Tracking from Multiple Frames

This letter assumes that a still camera is used for capturing all the analyzed videos. Then, different moving objects can be detected using a GMMs-based background subtraction technique [6]. After detecting the foreground objects, we should track them and obtain their trajectories for analyzing properly their transferring conditions.

Let $O_{k,t}$ denote the kth object O_k at the tth frame I_t. Let $\{p_i\}_{i=1,\ldots,n}$ be the normalized pixel locations of $O_{k,t-1}$. We can use an m-bin color density function f to represent $O_{k,t-1}$ [7]

$$f_u^{t-1} = \frac{1}{C}\sum_{p_i \in O_{k,t-1}} K(\|p_i\|^2)\delta[b(p_i)-u], \quad \text{for } u = 1,\ldots,m \tag{1}$$

where $C = \sum_{p_i \in O_{k,t-1}} K(\|p_i\|^2)$, is an isotropic kernel function, δ the Kronecker delta function, and $b(p_i)$ the color index of p_i. Let $\theta_{k,t}$ denote the position of $O_{k,t}$ in I_t, and $\{q_i\}_{i=1,\ldots,n_s}$ the set of normalized pixels of $O_{k,t}$ centered at the position $\theta_{k,t}$ in I_t. Then, we also can use another density function g to represent $O_{k,t}$ in I_t

$$g_u(\theta_{k,t}) = \frac{1}{C_s}\sum_{q_i \in O_{k,t}} K\left(\left\|\frac{q_i-\theta_{k,t}}{s}\right\|^2\right)\delta[b(q_i)-u], \quad \text{for } u = 1,\ldots,m$$

where $C_s = \sum_{q_i \in O_{k,t}} K(\|\frac{q_i}{s}\|^2)$ and s is a scaling change between $O_{k,t-1}$ and $O_{k,t}$. Then, from [7], the optimal position $\theta_{k,t}$ of $O_{k,t}$ is predicted as

$$\theta_{k,t} = \frac{\sum_{p_i \in O_{k,t-1}} \beta_i p_i K\left(\|\frac{p_i}{s}\|^2\right)}{\sum_{p_i \in O_{k,t-1}} \beta_i K\left(\|\frac{p_i}{s}\|^2\right)}$$
$$\text{with } \beta_i = \sum_{u=1}^{m}\sqrt{\frac{f_u}{g_u(0)}}\delta[b(p_i)-u]. \tag{2}$$

When the object has occlusions or large size changes, only the frame I_{t-1} cannot provide enough information for predicting $O_{k,t}$. Different from [7], this letter uses multiple frames to estimate the optimal position $\theta_{k,t}$ of $O_{k,t}$. Assume that there are F previous frames used for tracking $O_{k,t}$. Then, we want to find $O_{k,t}$ from I_t by solving the equation

$$O_{k,t} = \arg\max_{R \in I_t} P(R|I_t, O_{k,t-1}, O_{k,t-2}, \ldots, O_{k,t-F}). \tag{3}$$

Using the independence of $O_{k,t-i}$, we can rewrite (3) as the maximum likelihood problem

$$O_{k,t} = \arg\max_{R} \prod_{i=1}^{F} P(R|I_t, O_{k,t-i}) \tag{4}$$

where $P(R|I_t, O_{k,t-i})$ is defined as $\exp(-d^2(R|I_t, O_{k,t-i})/(\rho_{k,i}^2))$, $\rho_{k,i}$ is a weight to measure the contribution of $d(R|I_t, O_{k,t-i})$, and $d(R|I_t, O_{k,t-i}) = 1 - \sum_{u=1}^{m}\sqrt{f_u^{t-i} g_u(\theta_{k,t})}$. Through calculations, the optimal position of $O_{k,t}$ can be found as follows:

$$\theta_{k,t}^{optimal} = \sum_{i=0}^{F-1} w_{k,i}\theta_{k,t-i}$$
$$\text{with } w_{k,i} = \frac{\{d^2(R|I_t, O_{k,t-i})\}^{-1}}{\sum_{i=1}^{F}\{d^2(R|I_t, O_{k,t-i})\}^{-1}}. \tag{5}$$

III. Missed Color Detection Using Ratio Histogram

Persons appearing in video frames will have different interactions such as approaching, contacting, and leaving across different frames. As in Fig. 2, the two persons A and B have an "approaching" condition which then becomes a "leaving" condition after contacting. If a suspicious event (caused by carried objects) happens, there should be some object transfer happening between A and B. Let $A_{t=a}$ and $A_{t=l}$ denote the cases in which A appears at the "approaching" state and the "leaving" state, respectively. The symbols $B_{t=a}$ and $B_{t=l}$ are defined similarly for describing B. Then, this letter proposes a novel ratio histogram for analyzing their transfer conditions. Assume that $H_{A_{t=a}}(i)$ and $H_{A_{t=l}}(i)$ are the color histograms of $A_{t=a}$ and $A_{t=l}$, respectively. Two kinds of ratio histogram of A are defined, respectively, as

$$RatioH_{A_{a,l}}(i) = \frac{H_{A_a}(i)}{H_{A_l}(i)+1}$$
$$\text{and } RatioH_{A_{l,a}}(i) = \frac{H_{A_l}(i)}{H_{A_a}(i)+1}. \tag{6}$$

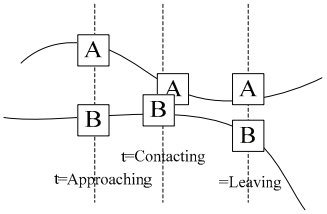

Fig. 2. Interactions between two persons A and B.

Fig. 3. Bag identification using color re-projection. (a) Person carrying a handbag. (b) Person without any bag. (c) Result of color re-projection.

This letter calls $RatioH_{A_{a,l}}(i)$ and $RatioH_{A_{l,a}}(i)$ as the *forward* and *backward* ratio histograms, respectively. Furthermore, the power of a ratio histogram is defined as follows:

$$|RatioH| = \sum_i RatioH(i).$$

With the ratio histograms, the missed colors between $A_{t=a}$ and $A_{t=l}$ can be easily found. For example, Fig. 3(a)–(b) show two cases of a person A with or without a handbag, respectively. Their corresponding color histograms are shown in Fig. 4(a)–(b), respectively. The ratio histograms $RatioH_{A_{a,l}}$ and $RatioH_{A_{l,a}}$ between (a) and (b) are shown in Fig. 4(c)–(d), respectively. Then, the handbag area can be identified by finding the missing colors between (a) and (b) which have large responses in $RatioH_{A_{a,l}}$. It is noticed that if the handbag is transferred from $A_{t=a}$ to $A_{t=l}$, the power of the forward ratio histogram $|RatioH_{A_{a,l}}|$ will be larger than the power of $|RatioH_{A_{l,a}}|$, and vice versa. Thus, whether the bag is transferred from A to B or from B to A can be easily determined by analyzing the powers of $|RatioH_{A_{a,l}}|$ and $|RatioH_{A_{l,a}}|$. In what follows, for convenience, the carried object is defined as a bag.

In order to analyze various suspicious events caused by bags more accurately, the bag position should be further located. Assume that the bag is transferred from A to B. We can find all the higher responses in $RatioH_{A_{a,l}}$ and re-project their corresponding colors on $A_{t=a}$ to find the bag's location. Let $T^H_{A_{a,l}}$ be the average value of all bins in$RatioH_{A_{a,l}}$. For a color bin k, if $RatioH_{A_{a,l}}(k) > 1.5T^H_{A_{a,l}}$, the kth color will be an important missing color for highlighting the bag. Fig. 3(c) shows the re-projection result of the missing colors found from Fig. 3(b) and re-projected on Fig. 3(a). Then, the desired bag location can be easily found through a connected component analysis. Only the region with the maximum area is the bag candidate. To filter out noisy regions, this letter requires this candidate to be larger than 50 pixels. In addition, a "position" constraint is used to filter out unlikely bag candidates. It means that a bag will not appear in the head of the observed pedestrian. Thus, we cannot find a bag from the extreme positions of its owner.

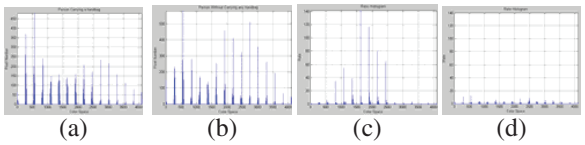

Fig. 4. Ratio Histograms. (a) and (b): Color histograms of Fig. 3(a)–(b). (c) Forward ratio histogram between (a) and (b). (d) Backward ratio histogram between (a) and (b).

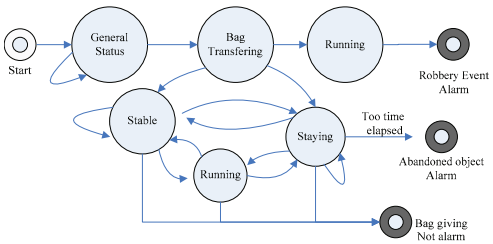

Fig. 5. Finite state machine for suspicious event detection through a bag-transferring trigger.

IV. Carried Object Detection Using GMMs

This section will use the GMMs-based approach to segment the desired bags from videos more accurately. Assume R is the bag region extracted using the above re-projection technique. Since it has different orientations, we first map its original coordinates (x, y) into an elliptic coordinate (s, t) as follows:

$$\begin{pmatrix} s \\ t \end{pmatrix} = \begin{pmatrix} cos\varphi_R & -sin\varphi_R \\ sin\varphi_R & cos\varphi_R \end{pmatrix} \begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix} \tag{7}$$

where (μ_x, μ_y) is its center of R, and φ_R is its major orientation obtained through a moment-based approach [12]. Let σ_s and σ_t denote its position variances in the s and t coordinates, respectively. We first model it using the model

$$G(s,t) = \exp\left(-\varepsilon_C^2(s,t) - \frac{s^2}{\sigma_s^2} - \frac{t^2}{\sigma_t^2}\right) \tag{8}$$

where ε_C denotes the color distance between (s, t) and the color mean μ_C of R, i.e.

$$\varepsilon_C(s,t) = (I_C(s,t) - \mu_C)^T \Sigma_R^{-1} (I_C(s,t) - \mu_C). \tag{9}$$

Here, $I_C(s, t)$ is the color of (s, t) and Σ_R the color covariance matrix of R in the (R, G, B) channels. Assume that $p_{bag}(z)$ denotes the probability of a pixel z belonging to a bag. The GMMs-based method using K Gaussian distributions to model $p_{bag}(z)$ has the form

$$p_{bag}(z) = \sum_{j=1}^{K} p(z|j)P(j) \tag{10}$$

TABLE I
STATE TRANSITION PROBABILITIES FOR SUSPICIOUS EVENT ANALYSIS

	Current State C_j		
	Stable	Running	Staying
$p(Stab\|C_j)$	0.55	0.25	0.3
$p(Run\|C_j)$	0.15	0.5	0.25
$p(Stay\|C_j)$	0.3	0.25	0.45

(a) (b) (c) (d)

Fig. 6. Bag segmentation when the bag has multiple colors. (a) Original image. (b) Result of color re-projection. (c) Result of bag location using GMM. (d) Result of bag segmentation.

(a) (b) (c) (d)

Fig. 7. Bag segmentations when the bag has a slanting orientation or a handbag has a handle. (a) and (c): Original images. (b) and (d): Results of bag segmentation using GMM.

where $P(j)$ the weight of the jth model and $p(z|j)$ is a single Gaussian parameterized with the mean μ_j and variance σ_j. K is set to three in this letter. Let $\alpha_j = P(j)$. Then, we can form a vector θ containing $3K$ elements: $\theta = (\alpha_1, \mu_1, \sigma_1, \ldots, \alpha_K, \mu_K, \sigma_K)$, where μ_j is a six-dimensional feature vector including three for (R, G, B) channels, two for (x, y) coordinates, and one for the major axis orientation. Then, the likelihood of a data z_n estimated from θ can be expressed by

$$P_{bag}(z_n|\theta) = \sum_{j=1}^{K} \alpha_j p(z_n|\mu_j, \sigma_j) \tag{11}$$

where $p(z_n|\mu_j, \sigma_j) = \exp(-(1/2)(z_n - \mu_j)^T \sigma_j^{-1}(z_n - \mu_j))$. Given the training data $\{z_1, \ldots, z_n, \ldots, z_N\}$, θ can be obtained by the maximum likelihood estimation as follows:

$$\theta^* = \arg\max_{\theta} \sum_{n=1}^{N} \log P_{bag}(z_n|\theta). \tag{12}$$

Let θ^i, μ_j^i, and σ_j^i denote the parameters θ, μ_j, and σ_j estimated at the ith iteration, respectively. In addition, let $q(j, z_n, \theta^i) = (\alpha_j p_j(z_n|\mu_j^i, \sigma_j^i))/(p(z_n|\theta^i))$, $\alpha_j^{i+1} = (1/N)\sum_{n=1}^{N}(\alpha_j^i p_j(z_n|\mu_j^i, \sigma_j^i))/(p(z_n|\theta^i))$, and $c_{n,j}^i = (z_n - \mu_j^i)(z_n - \mu_j^i)^T$. According to the expectation maximization algorithm [8], the $(i+1)$th estimates of μ_j and σ_j are obtained as follows:

$$\mu_j^{i+1} = \frac{\sum_{n=1}^{N} q(j, z_n, \theta^i) z_n}{\sum_{n=1}^{N} q(j, z_n, \theta^i)}$$

$$\text{and } \sigma_j^{i+1} = \frac{\sum_{n=1}^{N} c_{n,j}^i q(j, z_n, \theta^i)}{\sum_{n=1}^{N} q(j, z_n, \theta^i)}. \tag{13}$$

The initial estimate of θ can be obtained by clustering the carried bag region into different areas using the ISO-data algorithm [8].

V. SUSPICIOUS EVENT ANALYSIS USING FINITE STATE MACHINES

This section will use the finite state machines for suspicious event analysis if the event is caused by a carried bag R. Fig. 5 shows the diagram of our finite state machine. Usually, if a robbery event happens, the bag's speed will have obvious changes and a bag transferring condition will happen between the victim and the robber. Thus, in Fig. 5, the finite state machine always stays in the general state until a bag transferring (BT) state is enabled. In the BT state, if a running state is further enabled, a robbery event is then detected and an alarm sent. If the BT state enters into a "staying" state for a long time, the bag will be considered as an abandoned object. Other than the above two states, the BT state will enter into a stable state. A "stable" state means that R_t belongs to neither the running nor the staying states like jumping or walking. This condition makes a "bag giving" event to be enabled. Of course, there are some frequent state transitions between the "staying" and "stable" states. The state transitions are guided by two inputs: the bag's motion energy and its moving direction. Let $(v_x(R_t), v_y(R_t))$ be the motion vector of R at the tth frame. Then, its average motion direction is obtained from J consecutive frames as follows:

$$\bar{v}(R_t) = (\bar{v}_x(R_t), \bar{v}_y(R_t)) = \frac{1}{J}\sum_{i=0}^{J-1}(v_x(R_{t-i}), v_y(R_{t-i})) \tag{14}$$

where J is set to 8. The motion energy of R is defined by

$$ME(R_t) = \sqrt{(\bar{v}_x(R_t))^2 + (\bar{v}_x(R_t))^2}. \tag{15}$$

Since $\bar{v}(R_t)$ and $ME(R_t)$ are estimated from J frames, single-frame errors then can be filtered out for preventing possible transition errors between states. To categorize the states of R_t, hundreds of samples are first collected. Then, the K-mean algorithm [8] is used for obtaining different parameters of each bag state. Three categories are used to cluster the states of R_t, i.e., running, staying, and stable. In the recognition phase, the best state of R_t can be determined

(a) (b) (c)

Fig. 8. Detection results when the bag's colors are the same or similar to the person. (a) Original image. (b) Result using 2^{12} color bins (c) Result using 2^{15} color bins.

TABLE II
ACCURACY ANALYSIS OF THE RATIO HISTOGRAM TO ANALYZE OBJECT TRANSFERRING CONDITIONS

Conditions	Transferring	Interaction
Transferring	382	9
Interaction	0	75

TABLE III
ACCURACY ANALYSIS OF BAG DETECTION WHEN THE MAXIMUM REGION IS SELECTED AS A BAG

Total no. of bags	No. of detected bags	No. of missed bags
391	381	10

(a) (b) (c) (d)

Fig. 9. Results of carried bag segmentation when the person ran very quickly. (a) and (c): Contacting. (b) and (d): Results of bag detection.

(a) (b) (c)

Fig. 10. Result of abandoned object detection. (a) Leaving the bag on the ground. (b) Leaving. (c) Detection result of an abandoned object.

by searching the minimum distance between R_t and each bag state.

The general model shown in Fig. 5 sets all state transitions with the same probability. Nevertheless, some transitions are quite unequal in real cases. To fix this problem, the Baum–Welch algorithm [11] can be used to estimate the conditional probabilities of a state going into another state. The estimated state transition probabilities are listed in Table I.

VI. EXPERIMENTAL RESULTS

For all of our experiments, the video dimension used is 320 × 240, and each color channel (R, G, and B) is quantized into 16 levels such that 4096 color bins were used for the ratio histogram calculation. Fig. 6 shows the result when a bag with multiple colors was handled. Fig. 6(a) is the original image, (b) is the result of color re-projection, and (c) shows the result of the GMM-based method. The bag region is completely detected and shown in (d). Fig. 7 shows two other cases when the bag had a slanting orientation or a handbag had a handle. Clearly, even if the bags have various orientations or types, our method still works well to detect them.

Fig. 8 shows the case when the bag's colors are almost similar to those of the clothes of the carrier. In Fig. 8(b), when 4096 color bins were used, only a few bag pixels were detected. The detection precision of bag pixels was 51%. Fig. 8(c) is the result of bag pixel detection when 2^{15} color bins were used. Its detection precision is 83%. Obviously, we can increase the number of color bins for detecting a bag when its colors are almost similar to those of the clothes. However, more bins means more inefficiency. After trading off between the efficiency and accuracy, the number of used color bins is 2^{12}.

Table II lists the accuracy of our proposed ratio histograms to analyze object transferring conditions. A total of 466 video sequences were collected for this experiment. In this table, "interaction" means that only person interaction happens without any object transferring. After that, we seek the maximum missed region as possible bag candidates. Table III shows the accuracy of this criterion to detect bag candidates. Clearly, this criterion works very well.

Fig. 9 shows two cases of carried bag detection when the observed persons ran fast. The quick movements will blur the video quality. However, each bag still was successfully detected [see Fig. 9(b)–(d)]. Sometimes, the carried object will become an "abandoned" object. Fig. 10 shows the result of an abandoned object detection. Table IV summaries the accuracy analysis of carried object detection among different event types. After a carried object detection, an event classifier then can be designed for suspicious event analysis. Three event types were analyzed in this letter, i.e., "robbery," "abandoned," and "interchange." Table V lists the accuracy analysis among the three event types. All the above results have proved that our method is a robust, accurate, and powerful tool for carried object detection and suspicious event analysis.

TABLE IV
ACCURACY ANALYSES OF CARRIED OBJECT DETECTION AMONG DIFFERENT EVENTS

Conditions	Robbery	Abandoned	Interchange	Average
Number of carried objects	165	74	152	391
Number of correctly detected objects	159	74	148	381
Accuracy (%)	96	100	97	97.4

TABLE V
ACCURACY ANALYSIS OF SUSPICIOUS EVENT DETECTION

Event types	Abandoned	Interchange	robbery
Abandoned	74	0	0
Interchange	2	148	2
Robbery	4	8	153

VII. Conclusion

Detailed contributions of this letter can be summarized as follows: 1) A novel tracking method was proposed for tracking objects not only from a single frame but also multiple frames; 2) A novel ratio histogram was proposed for analyzing suspicious object transferring conditions. 3) A novel bag segmentation algorithm was proposed for extracting various bags from videos; 4) An event analyzer was designed for suspicious event detection using finite state machines.

References

[1] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Comput. Vision Image Understand.*, vol. 81, no. 3, pp. 231–268, Mar. 2001.

[2] I. Haritaoglu, R. Cutler, D. Harwood, and L. Davis, "Backpack: Detection of people carrying objects using silhouettes," *Comput. Vision Image Understand.*, vol. 81, no. 3, pp. 385–397, Mar. 2001.

[3] N. Oliver, B. Rosario, and A. Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE Trans. Pattern Recognition Mach. Intell.*, vol. 22, no. 8, pp. 831–843, Aug. 2000.

[4] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani, "Probabilities posture classification for human-behavior analysis," *IEEE Trans. Syst., Man Cybernetics-Part A: Syst. and Humans*, vol. 35, no. 1, pp. 42–54, Jan. 2005.

[5] J. Park, S. Park, and J. K. Aggarwal, "Model-based human motion tracking and behavior recognition using hierarchical finite state automata," in *Proc. Int. Conf. Computational Sci. Applicat.*, Assisi, Italy, May 2004, pp. 311–320.

[6] C. Stauffer and E. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Recognition Mach. Intell.*, vol. 22, no. 8, pp. 747–757, Aug. 2000.

[7] D. Comaniciu, V. V. Ramesh, and P. Meer, "Kernel based object tracking," *IEEE Trans. Pattern Recognition Mach. Intell.*, vol. 25, no. 5, pp. 564–577, May 2003.

[8] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. New York, NY: Wiley, 2001.

[9] E. Stringa and C. S. Regazzoni, "Real-time video-shot detection for scene surveillance applications," *IEEE Trans. Image Process.*, vol. 9, no. 1, pp. 69–79, Jan. 2000.

[10] G. L. Foresti, "Automatic detection and indexing of video-event shots for surveillance applications," *IEEE Trans. Multimedia*, vol. 4, no. 4, pp. 459–471, Dec. 2002.

[11] L. R. Rabiner and B. H. Juang, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[12] M. Sonka, V. Hlavac, and R. Boyle, *Image Process., Anal. and Mach. Vision*, London, U.K. Chapman & Hall, 1993.