JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 25, 1575-1592 (2009)

Illegal Entrant Detection at a Restricted Area in Open Spaces Using Color Features

JAU-LING SHIH, YING-NONG CHEN[1], KAI-CHIUN YAN[1] AND CHIN-CHUAN HAN[2]
Institute of Computer Science and Information Engineering
Chung Hua University
Hsinchu, 300 Taiwan
[1]*Institute of Computer Science and Information Engineering*
National Central University
Chungli, 320 Taiwan
[2]*Department of Computer Science and Information Engineering*
National United University
Miaoli, 360 Taiwan
E-mail: cchan@nuu.edu.tw

Digital video recording (DVR) systems are widely used in our daily life because of cost-down of capturing devices. Developing an automatic and intelligent system to detect, track, recognize, and analyze moving objects could save human power in monitoring centers. In this study, the color features of an employee's uniform were extracted to identify the entrance legality in a restricted area of an open space. First of all, a background subtraction technique was used to detect moving objects in image sequences. Three key object features, the position, the size and the color, were extracted to track the detected entrants. After that, the body of an entrant was segmented into three parts for locating the region of interest (ROI) using a watershed transform. Dominant color features extracted from the ROI were classified for preventing the illegal entrance. Some experiments were conducted to show the feasibility and validity of the proposed system. In the final part of the paper, conclusions are drawn and future work is suggested.

Keywords: video surveillance, legality detection, color structure descriptor, color feature, watershed transform

1. INTRODUCTION

Due to cost-down of capturing devices, monitoring cameras are widely used in public areas. Currently, many realizable systems using computer vision and image processing techniques intelligently detect, track, recognize, and analyze objects, especially human beings [1-3]. In an open space, people can move around freely without any limitation. However, there is no entrance control in restricted areas like at information desks in airports, cash counters in shops, pharmacy/nursing stations in hospitals, …, *etc*. Important data or money must be protected in these areas. Surveillance systems observe the entrants and try to grab their face images for ID identification or legality verification. Grabbing clear frontal face images is a common and effective approach for computer-based recognition systems. Different from the entrance control applications, face images in low resolution, in side views, or in a back view, are frequently grabbed from still cameras in an

Received October 25, 2007; revised March 3, 2008; accepted April 3, 2008.
Communicated by Tong-Yee Lee.
* This work was supported by Technology Development Program for Academia of DOIT, MOEA, Taiwan under grant No. 97-EC-17-A-02-S1-032.

open space. Successfully grabbing recognizable face images for legality verification has a low probability. Poor image quality decreases the monitoring performance. Fortunately, there is a common point in restricted areas: all legal entrants dress in uniforms. In this study, the color features of a uniform were extracted for verifying the legality of entrants.

Illegal entrance could be considered as an abnormal event. Due to the poor quality of face-based verification, other object features such as trajectory, shape, color, and motion data, are adopted to detect the abnormal events. These features have been widely used in content-based image retrieval (CBIR) fields [4]. Trajectory features are the most used for rare event detection at intersections, parking lots, or freeways for intelligent transportation systems (ITS) [5, 6], at airports [7], or on campuses [8]. Motion features are also popular for human behavior analysis. They are extracted from video streams in both un-compressed and compressed domains. Human motion extracted from motion-history images [9], motion-energy images [9], moment features [10], and polar-based histograms [11] are used to analyze human actions. It is known that human behavior is represented as a sequence of actions in the spatio-temporal domain. All approaches use the hidden Markov model (HMM) to solve this spatio-temporal problem. However, humans seldom behave as robbers with fierce motion or an abnormal trajectory in restricted areas. Their actions frequently appear to be similar to those of legal users. That means motion and trajectory features are not suitable in restricted areas. Uniform color is another cue for the verification of entrance legality, especially at the cases of lacked frontal face images.

In many pattern recognition applications, three common steps, *preprocessing*, *feature extraction*, and *classification*, are sequentially performed. In this study, moving object detection, tracking, segmentation, and region of interest (ROI) location were classified as the preprocessing step. A background subtraction based object detection was first performed and then targets were tracked based on the color structure descriptors. Since the object images are varied and poor in an open space, it is insufficient to determine the legality of an entrant using a single image. Using the tracking module, a sequential object images were grabbed for increasing the verification performance of legality. Next, the detected region image was segmented using watershed transform in four steps: *image simplification*, *gradient computation*, *rain-falling processing*, and *region merging*. In identifying the human body parts, circularity and human body model-based schemes were designed for identifying the head region, the upper body, and the lower body. The ROI was thus determined for extracting the colors of the uniform. A neural network-based verification was conducted to determine the legality of one object image. The legality of an entrant is determined by voting those sequential outputs.

The rest of this paper is organized as follows: Moving object detection and tracking are presented in section 2. Next, human body segmentation to find the ROI was designed using watershed transform as described in section 3. The legality of an entrant was determined in section 4 using color features of the uniform. In section 5, some experimental results were conducted to show the validity of the proposed approach. Finally, some conclusions are given in section 6.

2. MOVING OBJECT DETECTION AND TRACKING

Moving object detection and tracking are essential processes in video surveillance.

They determine the system performance. Traditionally, three methodologies, *the background subtraction-based*, *the temporal difference-based*, and *the optical flow-based approaches*, are used to detect moving objects in cluttered backgrounds. In this study, the background subtraction-based technique was adopted to identify the changing pixels. However, many detection algorithms have a shadow problem. Prati *et al.* [12] surveyed the shadow detection algorithms and made a comparative evaluation of four algorithms. The approach proposed by Davis *et al.* [1] was utilized to solve the shadow problem in this study. Image pixels were classified as the foreground pixels, the background pixels and the shadow pixels. Next, a labeling process was performed to cluster the components of objects. Two morphological operations, erosion and dilation, were performed to connect the components and to eliminate the noise. Each object was bounded in a bounding box from the connected components.

Subsequently, the detected objects were tracked and re-labeled. Generally, the position of objects and the relations among objects can be identified by using the motion data, template matching, and overlapping information. Many complex algorithms have been proposed for prediction [1, 13, 14]. In this study, the previous motion vectors [2] were used for the position prediction. The objects were re-labeled around the predicted position to get the new position. In addition to the labeling process, it was necessary to identify the interaction between objects in the video streams. Region-based template matching algorithm is an effective method to obtain the relation between objects in the consecutive image frames. However, it is a time consuming procedure to use the region-based template matching algorithm. This can be improved by using the overlapping information, table *OR*, between the current image frame and those in the histogram [3]. Using this table, several relations, including the 'merging', 'separating', 'entering', and 'leaving' between the moving objects can be easily obtained. The information of merging objects was kept, and a new object of type 'merged' was created. Next, the template matching procedure was performed to assign the exact labels at the 'separating' case. Histogram-based matching is an effective approach for identification. Collins *et al.* [13] kept the data of moving objects, such as the color histogram, the shape, the size, the trajectory, and the grey-image histogram.

MPEG-7 based descriptors represent the audio-visual meta-data for many surveillance applications [4, 15, 16]. Texture, color, and shape descriptors were frequently represented and retrieved subjects from a database. In this study, the color structure descriptor (CSD) of MPEG-7 standard was adopted for object representation to identify the objects during the separation. In addition to the statistical properties of the histogram-based approach, color structure descriptor possesses the regional structure of color distribution. The color data of an object in RBG color space were first converted to HMMD color space data (*Hue*, *Max*, *Min*, *Different*). The *hue* value ranged from 0° to 360°. The *min* and *max* features represent the minimum and the maximum values of the R, G, and B values. The *diff* and *sum* features denote the different feature ($diff = max - min$) and the summation feature (*e.g*. $sum = (max + min)/2$), respectively. According to the MPEG-7 standard, the CSD expresses the local color structures of a region. In this study, the color data within a window were quantized into 64 bins to obtain the CSD of an object image. This window of size 8 by 8 was called the *structuring element* (SE) and slid through the object image. If one color appeared within the SE, the number in the corresponding bin was increased by one. The SE slid from top to bottom and from left to right. The CSD of

size 64 was generated. The distance metric of CSD between two objects was defined as the summation of the absolute difference values, *e.g.*, the L_1-norm based distance. More details can be found in [4].

3. HUMAN HEAD/BODY SEGMENTATION AND IDENTIFICATION

In this section, the human body segmentation algorithm proposed by Park and Aggarwal [17] was modified to find the head, the upper body, and the lower body for identifying the ROI.

3.1 Segmentation of Human Body Using Watershed Transform

Park and Aggarwal [17] proposed an expectation maximization (EM) algorithm to estimate the Gaussian components of color distribution for image segmentation. Each region with a similar color belongs to the same cluster. It takes a lot of computational time to obtain the segmentation results. The watershed transform-based approach was applied for efficiently segmenting the human body regions. Four steps, *image simplification, gradient computation*, *rain-falling processing*, and *region merging*, were performed in watershed transform for body segmentation. The metric in CIELab color space is similar to that of human vision perception. Object images in RGB color space are thus converted to CIELab color space images before the process.

The main function of image simplification was to filter out noise and to blur the images. For the sake of simplification, morphology-based closing operations by partial reconstruction $\varphi^{(\text{res})}(\delta_n(f), \varphi_k(f))$ were performed as follows [18]:

$$\varphi^{(\text{res})}(f, r) = \varepsilon^{(\infty)}(f, r) = \ldots\varepsilon^{(1)}(\ldots\varepsilon^{(1)}(f, r))\ldots, r). \tag{1}$$

Here, two operations, dilation $\delta_n(f)$ and closing $\varphi_k(f)$ operations, were operated by the window structuring elements of size n and k, respectively. In addition, the geodesic erosion of size one was defined as $\varepsilon^{(1)}(f, r) = \max(\varepsilon_1(f), r)$. That means: image f was first dilated by a structuring element of size n. The dilated image $\delta_n(f)$ was repeatedly eroded by the geodesic erosion operation. The 'max' operation of $\varphi_k(f)$ and the eroded image was done in each time. This process was repeated until convergence.

In the second step, the morphological gradient [19] was generated for obtaining the gradient magnitude defined as $G(f) = \delta(f) - \varepsilon(f)$. Three gradient images G_L, G_a, and G_b on channels L, a, and b were generated and integrated to obtain an effective gradient image $g = \max(w_L G_L, w_a G_a, w_b G_b)$, where weights were set as $(w_L, w_a, w_b) = (1, 2, 2)$, respectively.

Third, a rain-falling process was executed for region growing using a pre-labeling strategy. Two advantages were presented in this strategy: the over-segmentation problem was solved, and the rain-falling process for the pre-labeled pixels became unnecessary. Therefore, the watershed transform process was performed. Initially, if the gradient magnitude of a pixel was lower than a specified drowning level t, this pixel was marked as the candidate of an initial region. Here, t was set as 0.6 of the average value of the gradient image. After all pixels were examined in this step, a labeling process was performed to obtain the initial regions and the new labels were assigned under the drowning

level t. The direction of unmarked pixels was defined to point to the neighbor with the lowest gradient magnitude. All unmarked pixels were assigned a region label by following their gradient downhill to the initial region.

The region merging rules were based on the region features such as size, color, and adjacency in the fourth step. These features for a specified region A_i were represented by the region size $|A_i|$, the mean vector of region color $\mu_i = [\mu_L, \mu_a, \mu_b]$, the boundary set $\Psi(A_i)$, and the adjacency rate $\beta_i(A_j) = \frac{|\Psi(A_i \cap A_j)|}{|\Psi(A_j)|}$ of two regions A_i and A_j. The merging rules were designed as follows:

Rule Adjacency: If the adjacency rate was larger than a given threshold T_1, *e.g.*, $\beta_i(A_i) > T_1$ or $\beta_j(A_i) > T_1$, the two regions were merged.

Rule Color Consistency: If the color distance between two adjacent regions was smaller than a pre-defined value, *e.g.*, $|\mu_i - \mu_j| < T_2$, the two regions were merged. The L_2-norm based distance was adopted in this step.

Rule Size: If the size of region A_i was smaller than a threshold T_3, *e.g.*, $|A_i| < T_3$, region A_i was merged to the adjacent region with the highest color similarity.

Three merging parameters T_1, T_2, and T_3 were set as 0.7, 15, and 50, respectively. For instance, Figs. 1 (a) and (b) show an original image and its simplified result. Fig. 1 (c) shows the results of watershed transform, and Fig. 1 (d) shows the merged result. There are seven regions in this figure.

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

Fig. 1. The segmentation results using watershed transform; (a) a detected object, (b) the simplified result, (c) the segmentation result using watershed transform, (d) the region merging results, and (e) the region map.

3.2 Identification of Human Body Parts

In order to find the ROI, a circularity-based measure and a merging scheme were designed for identifying the head, the upper body, and the lower body. Compared with the other body parts, the human head is a region with stable and obvious features [20]. In this study, the circularity measurement was adopted to evaluate the merged regions for finding the head region. The lower and the upper body parts were determined based on the ratio of the human body.

Fig. 2. The identification of a head region using a circularity-based measure.

Consider the pixels (x, y) of a region A (*e.g.* regions A_0, …, A_6 in Fig. 2 (b) or regions A_0, …, A_4 in Fig. 2 (e)). The centroid $O(\bar{x}, \bar{y})$ of region A was first calculated as follows: $\bar{x} = \frac{1}{|A|}\sum_{(x,y)\in A} x$, and $\bar{y} = \frac{1}{|A|}\sum_{(x,y)\in A} y$. Two circularity measurements, C_1 and C_2, were introduced in a textbook [21]. The first one represents the compactness of a region: $C_1 = \frac{|\Psi(A)|^2}{|A|}$. Here symbols $|\Psi(A)|$ and $|A|$ denote the perimeter and the size of region A, respectively. The second one is formulated as $C_2 = \frac{\mu_A}{\sigma_A}$, μ_A and σ_A represent the average distance and the standard derivation of distances from all boundary points to the centroid of region A as follows:

$$\mu_A = \frac{1}{|\Psi(A)|}\sum_{(x,y)\in\Psi(A)} \left|(x,y)-(\bar{x},\bar{y})\right|, \tag{2}$$

$$\sigma_A^1 = \frac{1}{|\Psi(A)|}\sum_{(x,y)\in\Psi(A)} \left(\left|(x,y)-(\bar{x},\bar{y})\right|-\mu_A\right)^2. \tag{3}$$

Since metric C_1 was much affected by the un-smoothing boundary, metric C_2 was adopted in this study. In addition, two criteria of the human model must be satisfied. First, the size of the head region must be smaller than $N/3$. Value N is the size of a moving object detected and tracked in the previous section. Second, the region with skin color possesses the higher priority for merging into the head region. The merging sequence of regions was based on the y coordinates of the regions' centroids. The finding steps are described below:

Step 1: Sort the y coordinates of the regions' centroids from top to bottom, initialize a list $\mathcal{R}_0$ as an empty set, and set an iteration index $k = 1$.

Step 2: Add the unselected and the top region A_i to list $\mathcal{R}_{k-1}$, *i.e.* $\mathcal{R}_k = \mathcal{R}_{k-1} \cup A_i$, to obtain a new region.

Step 3: If the size of region $\mathcal{R}_k$ is larger than $N/3$, terminate this procedure, and set the head region by choosing the region with the skin color and the largest circularity value. If not, choose the region with the largest circularity value.

Step 4: Compute the circularity value $C(\mathcal{R}_k)$ for region $\mathcal{R}_k$. Increase the iteration index k by one and repeat the above steps 2 to 4.

Table 1. The circularity values of the merged regions in Fig. 2.

Figures	Regions	Circularity
(b)	A_0	2.83
	$A_0 \cup A_1$	4.81
(e)	A_0	11.54
	$A_0 \cup A_1$	8.01
	$A_0 \cup A_1 \cup A_2$	4.42
	$A_0 \cup A_1 \cup A_2 \cup A_3$	3.20
	$A_0 \cup A_1 \cup A_2 \cup A_3 \cup A_4$	2.46

Two examples, a frontal image and a back one, are given to show the merged results as shown Fig. 2. The regions were numbered and the circularity values were calculated as tabulated in Table 1. In the first illustration, two regions A_0 (hair) and A_1 (face) with circularity value C_2 = 4.81 were assigned as the head region. Since the skin region possessed the higher priority, the A_0 (hair) and A_1 (face) regions whose circularity value C_2 = 8.01 were assigned as the head region even though region A_0 possessed the highest circularity value in the second example.

The information of the head region was thereafter obtained from the following equations [21]: The (p, q) order of central moments were computed as

$$\mu_{p,q} = \sum_{(x,y)\in A} (x-\overline{x})^p (y-\overline{y})^q. \tag{4}$$

Three second order moments $\mu_{1,1}$, $\mu_{2,0}$, and $\mu_{0,2}$ were computed for determining the lengths and orientations of the major and minor axes. The orientation of the major axis and two lengths of the major and minor axes were thus generated from the following equations:

$$\theta = \frac{1}{2}\tan^{-1}\left[\frac{2\mu_{1,1}}{\mu_{2,0}-\mu_{0,2}}\right], \tag{5}$$

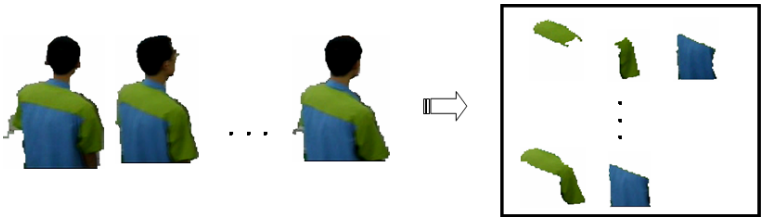
$$l = \sum_{(x,y)\in A} [(y-\overline{y})\cos\theta - (x-\overline{x})\sin\theta]^2, \tag{6}$$

$$L = \sum_{(x,y)\in A} [(y-\overline{y})\sin\theta - (x-\overline{x})\cos\theta]^2. \tag{7}$$

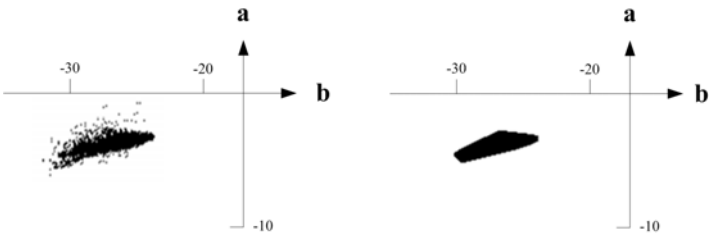
After finding the head region, the region of interest (ROI) was easily found. Approximately, the upper ROI was set as the region of length 2L, *e.g.* twice that of the head region. The rest of the regions of a moving object were classified as the lower ROI. When the color of the upper and the lower parts was similar, two parts were merged in the segmentation process. This scheme can also separate it into two parts, *e.g.* an upper ROI and a lower ROI. The upper ROI was considered in a 7-11 shop, and the entire part was treated as the ROI in a laboratory space. Only the pixels in the ROI were countered for legality verification. Three regions for the illustrated images in Figs. 2 (a) and (d) were identified as shown in Figs. 2 (c) and (f), respectively.

4. CONSTRUCTION OF UNIFORM COLOR MODEL

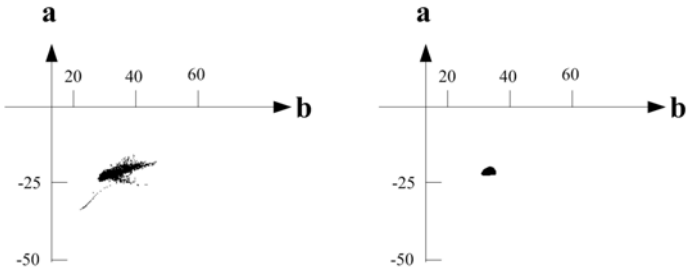
In this section, the color features of a uniform are extracted and classified for determining legality. An example for dominant color extraction is given. The image data of legal entrants were collected in the training phase. Using the segmentation algorithm for the human body in the previous section, the ROIs of human bodies were successfully identified as shown in Fig. 3 (a). The colors of the uniform were automatically modeled in CIELab color space. Since the number of dominant colors within the ROI was unknown, it had to be determined first. The number can be automatically determined using an un-supervised clustering algorithm. The images as shown in Fig. 3 (a) were segmented



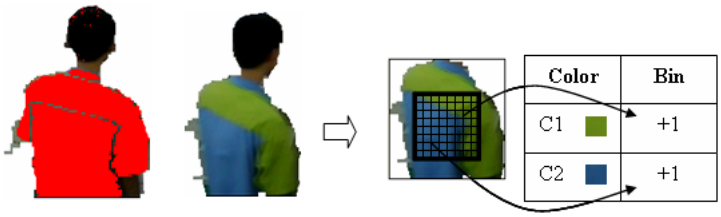
(a) The ROI images of training samples.



(b) The prototype of a blue color.



(c) The prototype of a green color.



(d) The pixels belonging to the dominant colors. (e) The CSDs of dominant colors within the ROI.

Fig. 3. The generation of dominant colors within the ROI.

into several regions with similar color pixels. Each region could be represented by its mean vectors and classified using the components of channels a and *b*. Moreover, regions smaller than 5% of the ROI were ignored. The *maxmin-distance-based* clustering algorithm was adopted to cluster regions as shown in Fig. 3 (b). Each small circle represents a segmented region. Two dominant color prototypes within the ROI were obtained in this illustration.

The distribution of each color prototype was computed from the pixels belonging to the same cluster. It could be formulated in the following steps.

1. Compute the mean μ and the standard derivation σ of each color prototype.
2. Remove those pixels whose distances to the cluster center are larger than a threshold value($\alpha\sigma$), $\alpha = 3$.
3. Remove the isolated pixels using the morphology-based clustering technique.
4. Bound the region using a convex hull.

The above steps identified the range of each dominant color. The range was bounded by a convex hull represented in several polynomial functions. In the given illustration, the distributions of two dominant color prototypes, blue and green, are displayed in Figs. 3 (b) and (c), respectively. Since the range of dominant colors has been identified, the pixels belonging to the dominant colors were counted as shown in Fig. 3 (d).

Since the entrants freely moved in a monitored space, the poses were varied with the time. Besides, the appearance of every entrant was also different. Because of the above causes, the ratios of dominant colors within the ROI were different at different times. Using the ratios of dominant colors within the uniform regions was not enough to determine legality. The spatio relations between color pixels also helped discrimination. Using the CSD-based representation, a structure element of size 8 by 8 slid the ROI, and the numbers of dominant colors were automatically counted to generate the feature vectors of the uniform. The color data within a window were quantized into 64 bins to obtain the features of an entrant. In the given example, the numbers of two dominant colors within the structuring element were counted as shown in Fig. 3 (e).

Backpropagation neural network (BPNN) classifier has been widely utilized in the filed of pattern recognition. In this study, this well-known classifier was applied to perform the verification task. Consider the feature vectors of dimensionality *n* to be inputted into a BPNN classifier. Value *n* was set as 64, *e.g.* the dimensionality of CSD of an ROI in the experiments. The BPNN architecture was designed as a three layer-based network including an input, a hidden, and an output layer. There were n, $\frac{n+1}{2}$, and 1 neurons in each layer, respectively. Fully connected links were established and trained for finding the better weights. The training process of BP neural network includes *sample collection*, *weight initialization*, *forward learning*, and *backward learning* steps. The scaled conjugate gradient algorithm [22] was utilized to adapt the weights with a specified error function in the last three steps. Given several video sequences, moving entrants were tracked, and the ROI of an entrant in every frame was located and identified. The features in a CSD-based vector were extracted from an ROI for representing the entrant's uniform. This vector was considered as a training sample for a BPNN classifier. Positive and negative training samples were collected from video streams. In our experiments, both

positive and negative samples were equally created for the simplification of training process. In addition, negative samples play a crucial role in determining the threshold value. When the video sequences of each entrant's uniform features were extracted and sequentially inputted the trained BPNN, the sequential outputs were obtained. The legality of an entrant was determined by voting the outputs.

5. EXPERIMENTAL RESULTS

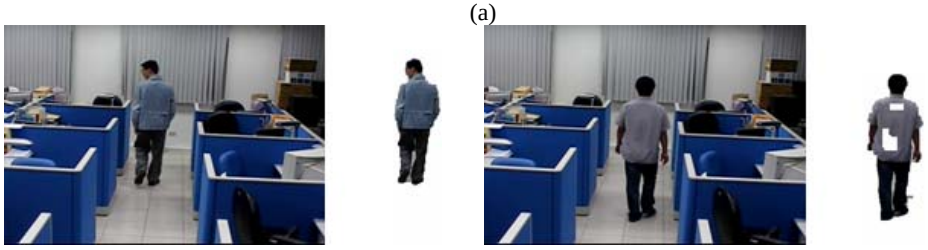
The proposed method was implemented in a 7-11 shop and in a laboratory as shown in Fig. 4. In the 7-11 shop, the restricted area was set at the cash counter, and in the laboratory, it was set at several personal desks. When the centroid of an entrant was located at an area, the detection procedure was triggered. Only the upper body was set as the ROI from the monitoring data in a 7-11 shop; the entire body was set as the ROI in the laboratory. Shown in Fig. 5 is the moving object detection and tracking results. There were several fragments in these figures. The results of body segmentation are shown in Fig. 6. The detection and segmentation results are shown in Figs. 6 (a) and (b), respectively. Eight illustrations are given from Figs. 7 and 8. In each figure, two image sequences were identified for the legality. Illegal entrants were detected and bounded in the red rectangles. In order to show the detection rate, 65 video sequences, 40 legal and 25 illegal, were collected in a 7-11 shop. 15 legal and 15 illegal entrants were randomly selected in the training phase. The others were used for testing. The false acceptance rate (FAR) and false rejection rate (FRR) were 9.8% (689/7000 frames) and 0.4% (30/7000 frames), respectively. Similarly, 30 video sequences were collected in a laboratory; 16 were used for training. The FAR and FRR for the other video data were both 6.7% (200/3000 frames). Moreover, the false detected frames were manually checked for comparing the errors of segmentation and classification. The segmentation errors are the main culprits of the FAR and FRR because more than 93% errors were generated from them.

(a) (b)

Fig. 4. The experimental spaces (a) in a 7-11 shop and (b) a laboratory.

In order to show the capability of the proposed method, five conditions of entering were simulated. The testing space was chosen at a stairway with two desks and several chairs. In the first case, a T-shirt in black color was considered as the legal uniform as shown in Fig. 9 (a). The first test is to verify an entrant wearing a gray T-shirt, *e.g.*, in a

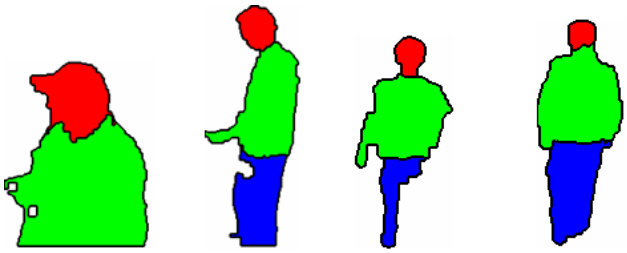
(a)



(b)

Fig. 5. The detected moving objects (a) in a 7-11 shop, and (b) a laboratory.

(a) The original images.



(b) The segmented results.

Fig. 6. The body segmentation results in a 7-11 shop and a laboratory.

(a) Image sequence one.

Fig. 7. Four legal entrants in (a) a 7-11 shop and (b) a laboratory.

(b) Image sequence two.

(c) Image sequence three.

(d) Image sequence four.

Fig. 7. (Cont'd) Four legal entrants in (a) a 7-11 shop and (b) a laboratory.

(a) Image sequence one.

(b) Image sequence two.

(c) Image sequence three.

(d) Image sequence four.

Fig. 8. Four illegal entrants in (a) a 7-11 shop and (b) a laboratory.

similar color. In this case, he was considered as a legal one as shown in Fig. 9 (b). From cases two to five, the video sequences of entrants wearing the seven-horizontal-stripe clothes in black-and-white colors were collected for training the detectors. Two samples are shown in Figs. 9 (c) and (d). The upper part of an entrant was set as the ROI. The dominant colors were determined from the training video sequences. The second case is to show the detection capability of multiple persons. Two persons simultaneously appeared in the image frame as shown in Fig. 9 (e). Since each entrant was correctly segmented, multiple persons can be verified by the trained classifier respectively. If the entrants were legal, they were drawn in white rectangles. Otherwise, they were drawn in red rectangles. The occlusion problem as shown in Fig. 9 (f) can't be solved because the illegal entrants occluded the legal one. The third and fourth cases show the robustness of a detector verifying the clothes in different patterns. Fig. 9 (g) shows an entrant wore the similar clothes but in different patterns. The CSDs of two ROIs, *e.g.* Figs. 9 (g) and (c), were quite different. The person in Fig. 9 (g) wore a black-white T-shirt was identified as an illegal entrant. Similarly, two persons wore the clothes both in similar colors and patterns. One person wore a T-shirt in ten black-and-white horizontal stripes, and the other one wore the clothes in vertical stripes as shown in Figs. 9 (h) and (i), respectively. In these conditions, the rates of dominant colors and the CSDs between two ROIs were almost equal and similar with those of legal entrants. They were both identified as the legal entrants. The last one is to show the case of wearing a back-bag. If an entrant wore a back-bag as shown in Figs. 9 (j) and (k), *e.g.* another occlusion problem, the BPNN outputted an illegal value because of the low rate of dominant colors. Since the tracking module was used in the proposed approach, they could be solved using a voting strategy. When the frontal view of an entrant was grabbed, he was considered as a legal one as shown in Fig. 9 (l).

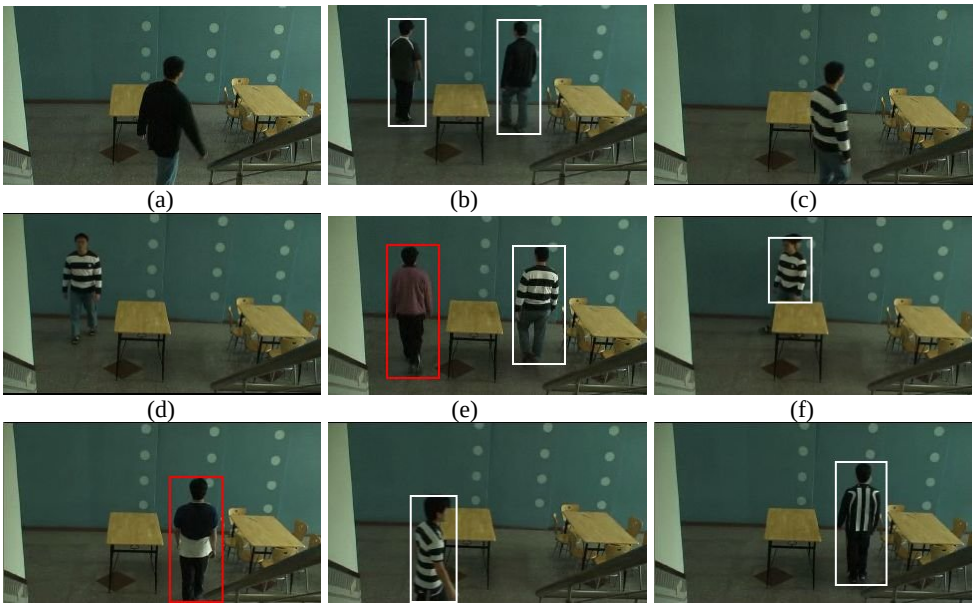
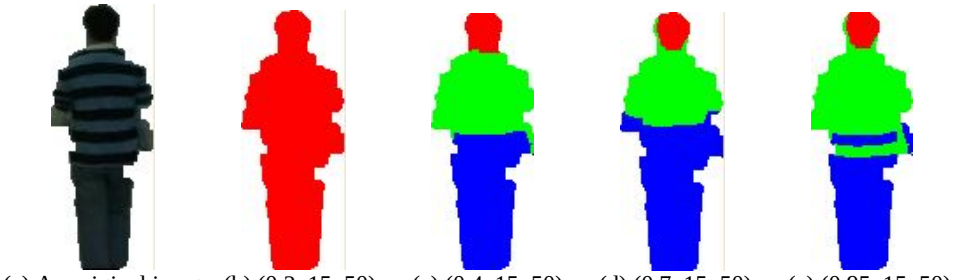


Fig. 9. The verification results of uniforms in various colors and patterns.

(j) (k) (l)

Fig. 9. (Cont'd) The verification results of uniforms in various colors and patterns.



(a) An original image. (b) (0.2, 15, 50) (c) (0.4, 15, 50) (d) (0.7, 15, 50) (e) (0.95, 15, 50)

(f) (0.7, 5, 50) (g) (0.7, 100, 50) (h) (0.7, 15, 10) (i) (0.7, 15, 75) (j) (0.7, 15, 200)

Fig. 10. The segmentation results using various parameters $T = (T_1, T_2, T_3)$.

The segmentation results depend on the parameters T_1, T_2, and T_3. An original object image is shown in Fig. 10 (a). During the merging process, value T_1 located in a range (0, 1) results in the segmentation results as shown in Figs. 10 (b) to (e). If value T_1 was too small, all regions are merged into an entire object region. On the contrary, small regions are isolated as shown in Fig. 10 (e). Next, parameter T_2 is a merging threshold using the color consistency criterion. If two regions belong to the same region of a uniform, the color difference should be small. The results shown in Figs. 10 (f) and (g) are similar by setting parameter T_2 as 5 and 100, respectively. The upper part and lower part are not merged because of the large color difference between them. Parameter T_3 is a threshold to merge fragmented regions. It can't be a large value. The upper and lower parts are merged by setting parameter T_3 as a large value, *e.g.* 200, as shown in Fig. 10 (j). The un-merged regions will be ignored in identifying the ROI. In summary, parameter T_1 is more sensitive than parameters T_2 and T_3. The over-segmentation of watershed transformation generates lots of fragmented regions. The first merging rule, rule adjacency, merges most of the small regions with high adjacency. The second rule determines the color difference between the upper and lower parts. Rule three merges the fragmented regions to obtain the complete region of a uniform.

Next, the false detections generated by the segmentation errors are presented in the following cases. In the first case, the detection errors occurred due to segmentation errors. They resulted in incorrect features of the uniform. Another factor causing the segmentation error was the basket used by an entrant, as shown in Fig. 11 (a). It changed the ROI's size and resulted in the incorrect rate of dominant uniform colors. In the second case, human posture resulted in segmentation errors, as shown in Fig. 11 (b). One hand of an entrant, for example, was put on his head. It generated an incorrect head region and an incorrect uniform feature.

(a) (b)

Fig. 11. The detection errors generated by segmentation errors.

Now, let us discuss the execution time of the proposed approach. According to the performance requirement, the event decision should be made within a few seconds. In the proposed approach, detecting and tracking moving objects needed 0.11 seconds per frame in both experiments. The legality identification included two sub-processes: the body segmentation and the uniform color extraction. The needed time depended on the object sizes and the region number within the ROI. The object sizes and the region number in a 7-11 shop were larger than those in a laboratory. Therefore, the legal identification in a 7-11 shop needed 0.37 seconds per frame, and it needed 0.14 seconds per frame in the laboratory. In summary, two image frames in a 7-11 shop and four image frames in a laboratory were processed in one second for illegality detection.

6. CONCLUSIONS

In this study, the color features of the uniform were extracted to determine the legality of entrants. This approach is suitable for use in a large space where the image resolution is low, the image quality is bad, or where there are back face images. In addition to the detection and tracking of moving objects, ROI identification and dominant color extraction were the main tasks used for representing the uniform. The domain colors were automatically determined using an un-supervised clustering algorithm. The CSDs of the uniform colors were extracted and inputted into the NN-based classifier for the legality decision. Since the BPNN classifier is trained in few minutes, this detector of any legal uniform can be easily constructed even though the illumination is varied. One important thing should be emphasized: This simple and useful approach can be imple-

mented on the existing systems without any extra equipment. This early warning system warns the employees to keep an eye on the cash counter when they leave out. In future, the video clips of illegal entrances will be collected and indexed for further retrieval.

REFERENCES

1. I. Haritaoglu, D. Harwood, and L. Davis, "W^4: Real-time surveillance of people and their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligent*, Vol. 22, 2000, pp. 809-830.
2. F. Brémond and M. Thonnat, "Tracking multiple nonrigid objects in video sequences," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 5, 1998, pp. 575-584.
3. L. M. Fuentes and S. A. Velastin, "People tracking in surveillance application," in *Proceedings of the 2nd IEEE International Workshop on PETS*, 2001, pp. 141-149.
4. B. S. Manjunath, P. Salembier, and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Interface*, John Wiley and Sons, New York, 2002.
5. G. Medioni, I. Cohen, F. Brémond, S. Hongeng, and R. Nevatia, "Event detection and analysis from video streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, 2001, pp. 873-889.
6. W. M. Hu, D. X. Tieniu, and S. Maybank, "Learning activity patterns using fuzzy self-organizing neural network," *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*, Vol. 34, 2004, pp. 1618-1626.
7. M. T. Chan, A. Hoogs, J. Schmiederer, and M. Petersen, "Detecting rare events in video using semantic primitives with HMM," in *Proceedings of the 17th International Conference on Pattern Recognition*, 2004, pp. 150-154.
8. D. Makris and T. Ellis, "Learning semantic scene models from observing activity in visual surveillance," *IEEE Transactions on System, Man, and Cybernetic – Part B: Cybernetics*, Vol. 35, 2005, pp. 397-408.
9. A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, 2001, pp. 257-267.
10. R. Rosales, "Recognition of human action using moment-based features," Technical Report BU 98-020, Department of Computer Science, Boston University, 1998.
11. R. V. Babu, B. Anantharaman, K. R. Ramakrishnan, and S. H. Srinivasan, "Compressed domain action classification using HMM," *Pattern Recognition Letters*, Vol. 23, 2002, pp. 1203-1213.
12. A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, "Detecting moving shadows: Algorithms and evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligent*, Vol. 25, 2003, pp. 918-923.
13. R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, and Y. Tsin, "A system for video surveillance and monitoring," Technical Report CMU-RI-TR-00-12, The Robotics Institute, Carnegie Mellon University, 2000.
14. P. Perez, J. Vermaak, and A. Blake, "Data fusion for visual tracking with particles," *Proceedings of the IEEE*, Vol. 92, 2004, pp. 495-513.
15. J. Annesley and J. Orwell, "On the use of MPEG-7 for visual surveillance," in *Pro-*

ceedings of the 6th IEEE International Workshop on Visual Surveillance, 2006.
16. J. Annesley, V. Leung, S. Velastin, A. Colombo, and J. Orwell, "Fusion of multiple features for identity estimation," in *Proceedings of International Conference on Imaging for Crime Detection and Prevention, Visual Information Engineering IET*, 2006, pp. 534-539.
17. S. Park and J. K. Aggarwal, "Segmentation and tracking of interacting human body parts under occlusion and shadowing," in *Proceedings of IEEE Workshop on Motion and Video Computing*, 2002, pp. 105-111.
18. P. Salembier and M. Pardas, "Hierarchical morphological segmentation for image sequence coding," *IEEE Transactions on Image Processing*, Vol. 3, 1994, pp. 639-651.
19. M. Kim, J. Choi, D. Kim, H. Lee, M. H. Lee, C. Ahn, and Y. Ho, "A VOP generation tool: Automatic segmentation of moving objects in image sequences based on spatio-temporal information," *IEEE Transactions on Circuits and System for Video Technology*, Vol. 9, 1999, pp. 1216-1226.
20. S. Birchfield, "An elliptical head tracker," in *Proceedings of IEEE Conference on Signal, System and Computers*, 1997, pp. 1710-1714.
21. L. G. Shapiro and G. C. Stockman, *Computer Vision*, Prentice Hall, New Jersey, 2001.
22. M. F. Moller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Network*, Vol. 6, 1993, pp. 525-533.

Jau-Ling Shih (石昭玲) was born on December 13, 1969 in Tainan, Taiwan. She received the B.S. degree in Electrical Engineering from National Sun Yat-Sen University, Kaohsiung, Taiwan in 1992, the M.S. degree in Electrical Engineering from National Cheng Kung University, Tainan, Taiwan in 1994, and the Ph.D. degree in Computer and Information Science from National Chiao Tung University, Hsinchu, Taiwan in 2002. She is currently an Associate Professor in the Department of Computer Science and Information Engineering, Chung Hua University, Hsinchu, Taiwan. Her main research interests include image processing and image retrieval.

Ying-Nong Chen (陳映濃) was born in Taipei, Taiwan, in 1976. He received the B.S. and M.S. degrees in information management and informatics from the Nan Hua University, Taiwan and the Fo Guang University, Taiwan, in 2000 and 2003, respectively. He is currently pursuing the Ph.D. degree in Computer Science and Information Engineering at the National Central University, Taiwan. His research interests include pattern recognition, computer vision and machine learning.

Kai-Chiun Yan (袁凱群) was born in Taoyuan, Taiwan, in 1980. He received the B.S. degree in Computer Science from Chung Hua University, Taiwan, in 2003. and the M.S. degree in Computer Science from National Central University, Taiwan, in 2005. He is currently a software engineer in Coretronic Corporation.

Chin-Chuan Han (韓欽銓) received the B.S. degree in Computer Engineering from National Chiao Tung University in 1989, and an M.S. and a Ph.D. degree in Computer Science and Electronic Engineering from National Central University in 1991 and 1994, respectively. From 1995 to 1998, he was a postdoctoral fellow in the Institute of Information Science, Academia Sinica, Taipei, Taiwan. He was an assistant research fellow in the Telecommunication Laboratories, Chunghwa Telecom Co. in 1999. From 2000 to 2004, he worked with the department of Computer Science and Information Engineering, Chunghua University, Taiwan. In 2004, he joined the department of Computer Science and Information Engineering, National United University, Taiwan, where he became a professor in 2007. Prof. Han is a member of IEEE, SPIE, and IPPR in Taiwan. His research interests are in the areas of face recognition, biometrics authentication, video surveillance, and pattern recognition.