Research Article

A Conditional Entropy-Based Independent Component Analysis for Applications in Human Detection and Tracking

Chin-Teng Lin,[1] Linda Siana,[1] Yu-Wen Shou,[2] and Tzu-Kuei Shen[1]

[1] *Department of Electrical and Control Engineering, National Chiao Tung University, Hsinchu 300, Taiwan*
[2] *Department of Computer and Communication Engineering, China University of Technology, Hsinchu 303, Taiwan*

Correspondence should be addressed to Yu-Wen Shou, owen@cute.edu.tw

We present in this paper a modified independent component analysis (mICA) based on the conditional entropy to discriminate unsorted independent components. We make use of the conditional entropy to select an appropriate subset of the ICA features with superior capability in classification and apply support vector machine (SVM) to recognizing patterns of human and nonhuman. Moreover, we use the models of background images based on Gaussian mixture model (GMM) to handle images with complicated backgrounds. Also, the color-based shadow elimination and head models in ellipse shapes are combined to improve the performance of moving objects extraction and recognition in our system. Our proposed tracking mechanism monitors the movement of humans, animals, or vehicles within a surveillance area and keeps tracking the moving pedestrians by using the color information in HSV domain. Our tracking mechanism uses the Kalman filter to predict locations of moving objects for the conditions in lack of color information of detected objects. Finally, our experimental results show that our proposed approach can perform well for real-time applications in both indoor and outdoor environments.

1. Introduction

Video-based human detection and tracking has been a popular research area and widely applied in various applications such as homecare, security, and patient monitoring. With the increasing criminal rate, the development of automatic visual surveillance with computer visions has attracted more and more researchers' attentions. Therefore, the ability to distinguish people from other moving objects such as animals or vehicles has become an important issue for tracking targets and analyzing their behaviors.

Human detection system could be divided into two parts, segmentation of the moving objects from backgrounds and discrimination of humans from nonhuman objects. There have been several methods for segmenting moving objects from backgrounds such as the optical flow, stereo-based vision, and temporal difference method. The optical flow method could succeed in detecting independent moving objects, but would be more computational and sensitive to the change of intensity. Zhao and Thorpe [1] exploited the stereo based segmentation algorithm to extract objects

from backgrounds and identified the extracted objects by neural networks. Although the stereo-based vision technique has been proved to be more robust, it required at least two cameras and could be used only for the short distance detection. Orrite Uruñuela et al. [2] used multiple cameras to analyze 3D skeletal structure in gait sequences and 3D skeletons to extract human body shapes completely and constructed the point distribution model (PDM) by using Principal Component Analysis (PCA). Jiang et al. [3] used the background subtraction method to segment an isolate human and took advantage of the homogeneous properties of shadows and background objects to reduce the shadowing effects. An area threshold was also used to avoid a sudden change of light and interfering the results of moving object extraction by illumines. Tian and Hampapur [4] combined the background subtraction and optical flow methods to locate the motion area and to remove the false foreground pixels. They modeled the background image as Gaussian distributions to adapt to the gradual change of light by recursively updating the arguments of models with an adaptive filter. However, this basic model would sometimes
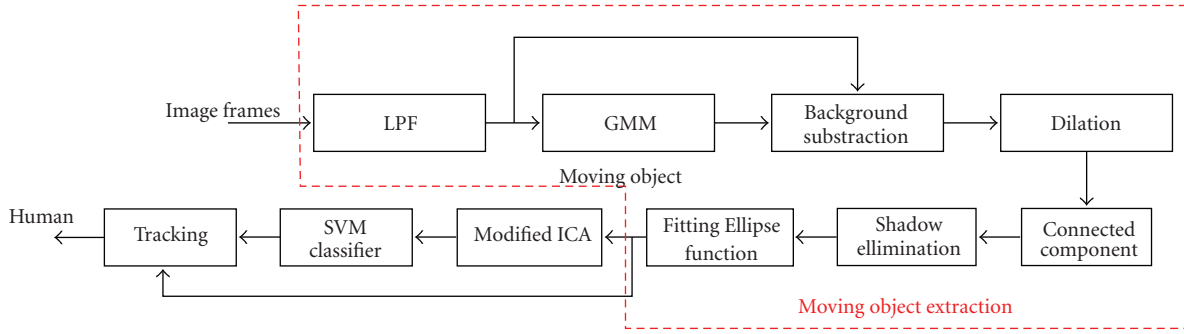
Figure 1: System architecture.

fail to handle complicated backgrounds such as water wave and tree shaking. Stauffer and Grimson [5] constructed a mixture of Gaussian model by modeling each pixel as a mixture of Gaussians and using an online approximation to update the extracted backgrounds. Our proposed real-time system firstly used a simpler way to segment moving objects to reduce the time complexity, and applied Gaussian mixture model (GMM) to constructing a dynamic background model as to handle dynamic backgrounds or unstable illumination in images.

After moving objects have been segmented, the next process would be human recognition. There have been several kinds of methods for human recognition like shape-based, motion-based, and multicue-based ones. Zhou and Hoang [6] used the shape information of human bodies to construct a codebook and to tell human beings from other objects. This method obviously would work well if the extracted human shape was obvious. However, this shape-based would usually fail for the cases of partially occluded humans or the detected humans carrying something. Histograms of Oriented Gradients (HOG) [7, 8], the algorithms based on Fast Fourier Transform, extracted features from the shape information. Curio et al. [9] carried out the detection process based on the geometrical features of human at the first step, and then used motion patterns of limb movements to determine the initial hypotheses of objects. Yoon and Kim [10] made use of the robust skin color, background subtraction, and human upper body appearance information to classify human or other objects with similar skin color regions. For the approaches based on neural networks for human identification [11], used the back-propagation model to recognize the pedestrians, to analyze the shape of object, and to classify human beings from other objects. Mostly, researchers have focused on the issue of feature extraction but paid much less attention to the field of feature selection. In this paper, we presented a modified ICA approach based on conditional entropy. In the recent years, ICA has been applied to human feature extraction for constructing a sufficient set of features describing human beings. ICA is a high-order statistical analysis method, and can be usually regarded as an extension of PCA, which addresses only the second-order statistical arguments. Unlike PCA features, the ICA features are not sorted, thus the conditional entropy is applied to feature selection, the sorting process, and choosing

an appropriate subset of ICA features. Sorting variables may be an important step to enhance the high-dimensional dataset, which gave us the idea to place correlated or similar dimensions close to each other in high-dimensional visual space to help human users perceive relationships among those variables easier [11]. The remainder of this paper could be organized as follows. Section 2 described the moving object extraction, including shadow elimination and occlusion handling. Section 3 introduced the modified ICA. Section 4 described the color-based tracking method. Section 5 showed the experimental results. We finally summarized discussions and conclusions in Section 6.

2. Moving Object Extraction

The architecture of our moving object extraction was indicated in the dotted-line block of Figure 1 and the remained blocks represented our processes in human feature extraction and classification. For the moving object extraction, we used the background subtraction method in order to meet the real-time acquirements. Besides, we built up a dynamic background model based on GMM algorithm to deal with more complicated backgrounds. Our background model was constructed by using three different Gaussian distributions. We in this paper took the difference of luminance in images since human eyes would be more sensitive to luminance than chrominance. The difference DI for each pixel (x, y) could be calculated by

$$\mathrm{DI}(x, y) = \left| I_c(x, y) - I_b(x, y) \right|, \tag{1}$$

where I_c and I_b denote the luminance of the current and background image, respectively. Practically, the moving objects would have larger variances than the background, so the determined threshold was set by the variance of each Gaussian background model and the possible foreground image PFI could be described in the following equation

$$\mathrm{PFI}(x, y) = \begin{cases} 1 & \text{if } \mathrm{DI}(x, y) \geq 3\sigma(x, y), \\ 0 & \text{if } \mathrm{DI}(x, y) < 3\sigma(x, y). \end{cases} \tag{2}$$

Each Gaussian distribution $N \sim (\mu, \sigma)$ could adapt to the gradual change of light by recursively updating each pixel over time. In the practical conditions, the captured

background might be in gray scale or in an edge map. Both of the background types had their individual advantages and disadvantages. The background image in gray scale might take longer in the updating process than that in the edge map, but could model the background in more details. Relatively, the edge-type background was less sensitive than the gray-type one, and would be more suitable for noisy images or environments with unstable intensities. For the strategic design in modeling a background in this paper, the Gaussian low-pass filter would be carried out in the consecutive input frames before processed at the GMM stage so as to reduce the influences of noises and disturbances.

2.1. Color-Based Shadow Elimination. Our color based shadow elimination are based on RGB-color channels. It can be easily observed that the luminance of shadow pixels is lower than that of the corresponding pixels in the background image. Thus, if we denote I_{CF} and I_B the intensity of current frame and the background image, respectively, the pixel (x, y) satisfying (3) may be in the shadowed region

$$I_{CF}(x, y) < I_B(x, y). \tag{3}$$

Some other observed characteristics of shadows can be arranged as follows. First, the texture of shadows like edge would have a smaller fluctuation than that of the corresponding pixels in the background image. Similarly, the chromaticity value of shadows would have a slighter change than that of the corresponding pixels in the background image. These observations are described in

$$\text{Between-pixel invariant} \longrightarrow \frac{I_{CF}(x, y)}{I_{CF}(x + 1, y)} = \frac{I_B(x, y)}{I_B(x + 1, y)},$$

$$d_h(x, y) = \ln \frac{I(x, y)}{I(x + 1, y)}, \qquad d_v(x, y) = \ln \frac{I(x, y)}{I(x, y + 1)}, \tag{4}$$

where $I(x, y)/I(x + 1, y)$ is the ratio between pixel (x, y) and its neighboring pixel $(x + 1, y)$, $d_h(x, y)$, and $d_v(x, y)$ denote the ratio maps which can keep the texture- and edge-information without the interferences of shadows. We will consider the pixel (x, y) in the shadow region if its ratio map is similar to that of the background pixel. The error in discriminating the pixel (x, y) from shadows can be calculated by

$$\Psi(x, y) = \sum_{(i,j) \in W} |d_{CF,h}(i, j) - d_{B,h}(i, j)|$$
$$+ |d_{CF,v}(i, j) - d_{B,v}(i, j)|, \tag{5}$$

where $\Psi(x, y)$ denotes the sum of difference of the ratio map in a small neighborhood window W with the center at (x, y)

$$\text{Within-pixel invariant} \longrightarrow r_{CF}(x, y) = r_B(x, y) = \ln \frac{R(x, y)}{B(x, y)},$$

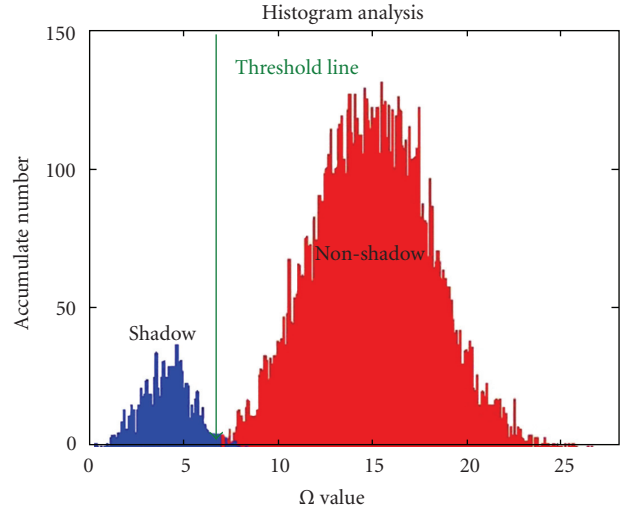$$g_{CF}(x, y) = g_B(x, y) = \ln \frac{G(x, y)}{B(x, y)}, \tag{6}$$

FIGURE 2: The distribution of POI pixels.

r and g denote the spectral ratios of *R-B* and *G-B*, respectively. As what we have observed, the shadow on the background pixel may result in a bigger change of brightness than color. Assume that the color of illumination may not change with the effect of shadows, thus the spectral ratio $r(x, y)$ is invariant to the magnitude of illumination. Similarly, the spectral ratio $g(x, y)$ is invariant under shadows or different conditions of illumination. Thus, pixel (x, y) is in the shadow region if both the current and background spectral ratios are the same. The error of spectral ratios can be computed by $\Theta(x, y)$ defined in

$$\Theta(x, y) = |r_{CF}(x, y) - r_B(x, y)| + |g_{CF}(x, y) - g_B(x, y)|. \tag{7}$$

The total error in discriminating (x, y) from shadows is described in

$$\Omega(x, y) = \alpha \cdot \Psi(x, y) + (1 - \alpha) \cdot \Theta(x, y), \tag{8}$$

where α denotes the weighting parameter. Finally, a thresholding operation will be applied on $\Omega(x, y)$ to determine whether the pixel (x, y) belongs to the shadow or foreground object.

The distribution of pixels of the possible object image (POI) which contains both of the moving object and shadowed pixels is shown in Figure 2. Figure 2 illustrates a smaller distribution for the shadowed region than the extracted region of moving objects. We hence take advantage of this $\Omega(x, y)$ observation to determine a threshold for discriminating the shadowed regions. The threshold value decides if a pixel (x, y) is in a shadowed region, and can be denoted in

$$\text{Th}_s = \mu_{PO} - \beta \cdot \sigma_{PO}, \tag{9}$$

where β is a weighting value, μ_{PO} and σ_{PO} are the mean and standard deviation of POI, respectively. And the region of shadow image SI would be described in (10). To enhance
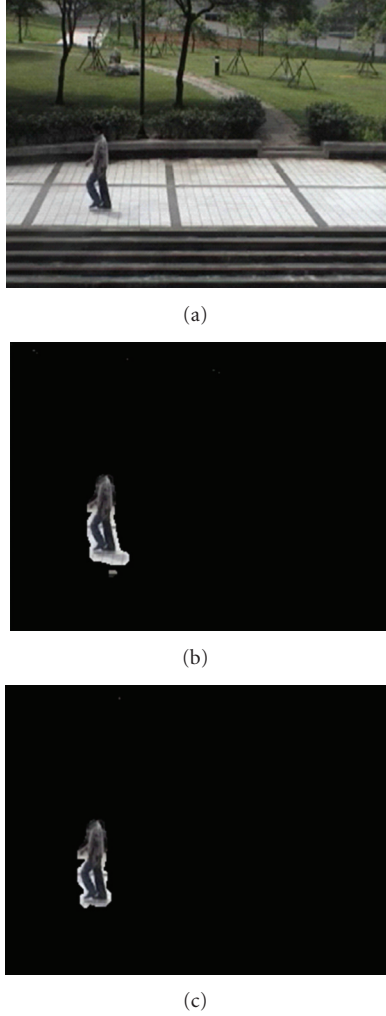
(a)

(b)

(c)

FIGURE 3: The results in shadow elimination, (a) the original image, (b) the extracted object before shadow elimination, (c) the extracted object after shadow elimination.

the results by shadow removal, we have the results during the process of shadow elimination in Figure 3

$$
\text{SI}(x, y) = \begin{cases} 1 & \text{if } I_{\text{PO}}(x, y) > I_B(x, y), \\ & \Omega(x, y) < \mu_{\text{PO}} - \beta \cdot \sigma_{\text{PO}}, \\ 0, & \text{otherwise.} \end{cases} \tag{10}
$$

2.2. Occlusion Handling. The moving objects could be detected as a group of people who may move together or may be partially occluded by each other. In this case, the moving object extraction system will label the group of people as one object by connected components. Without separating the group of people into each individual, the classification process may usually fail to identify the human beings. In most conditions, however, the heads are usually separate when the human bodies have been occluded. Besides, the shape of human heads is almost invariant even though a person rotates his head in different phases. Therefore, as long
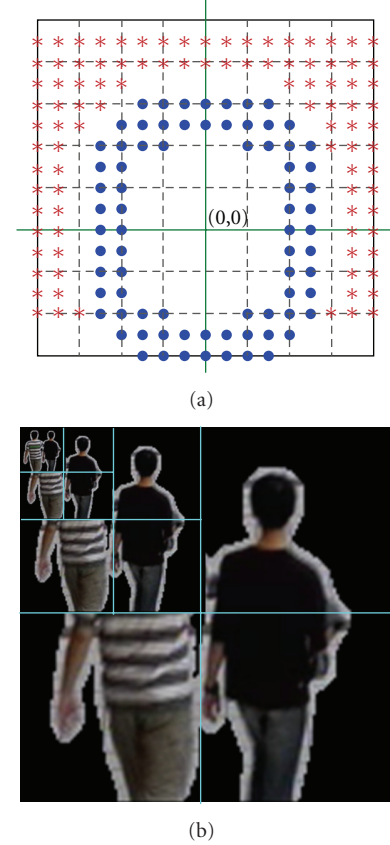

(a)

(b)

FIGURE 4: (a) The ellipse head model. (b) The pyramid down sampling process.

as only the human bodies are occluded we can use the head information to overcome the occluded problem. If heads are partially or fully occluded with each other, then our ellipse head model will find the head with the best match.

The proposed head model is shown in Figure 4(a) where the dot "•" represents the pixels of a head, the star "∗" represents the pixels of background, and the point $(0, 0)$ is the center of the ellipse. The process in down sampling is applied to fit the ellipse model in different sizes of a moving object. By setting a threshold of the similarity value, we can decide which point is a possible center of the head. Consequently, there may be more than one center detected in the real head region, which would be illustrated in the group of green points in Figure 5. Thus, we have to project the original head region into x-axis and y-axis, and to group these points to determine the final representative center as shown in blue points of Figure 5. We also show some results in our human detection mechanism for individual humans in Figure 6.

3. Modified ICA Based on Conditional Entropy

The independent component analysis (ICA) is a statistical method for transforming an observed multidimensional random vector into components that are statistically independent. ICA can be considered as a generalization of
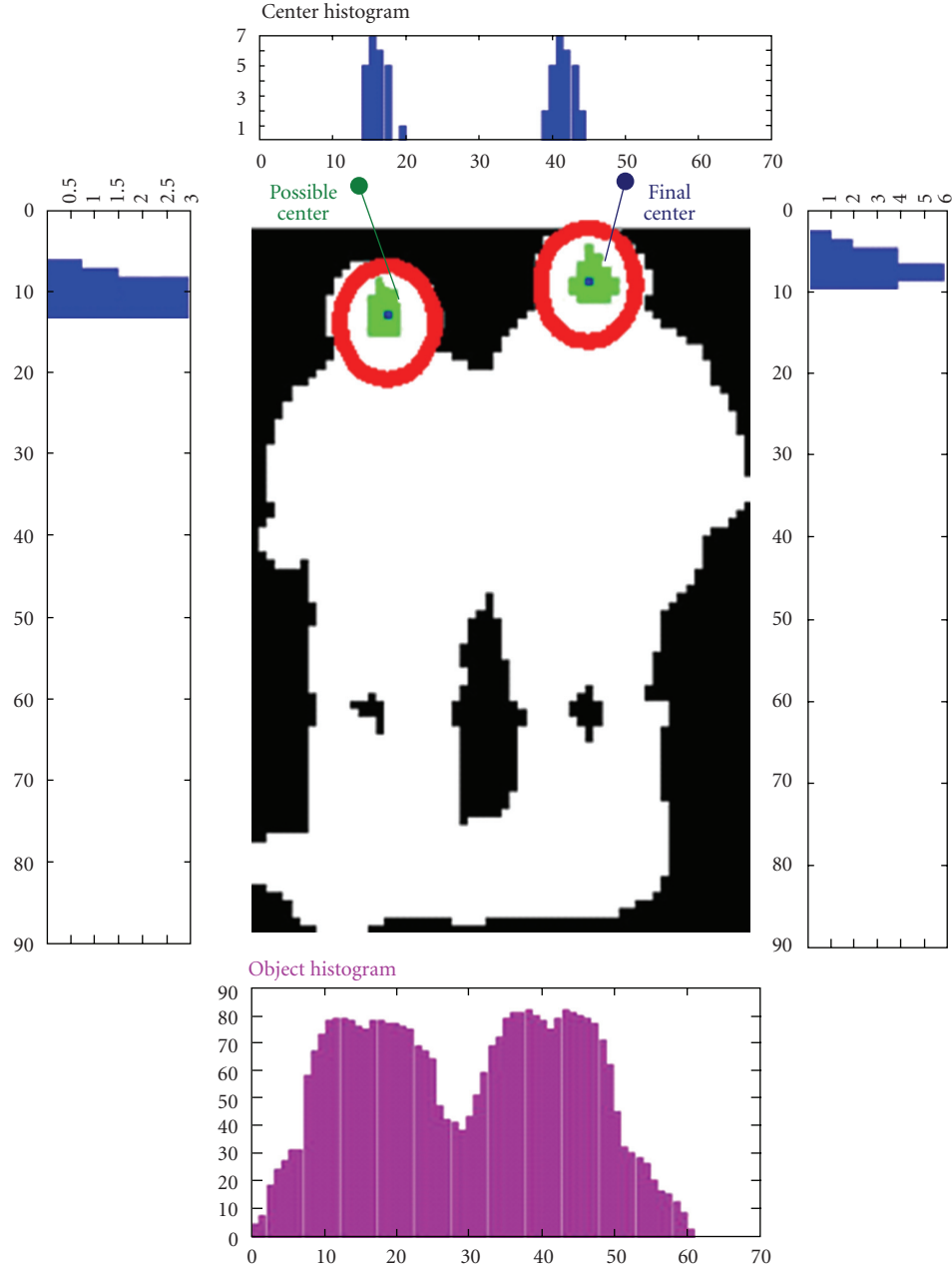
FIGURE 5: Projection of head region.

principal component analysis (PCA) with appended independent properties in the second order equations. In the field of signal processing, ICA can separate the waveform of the original source from the sensor array without uses of the characteristics of the source signal. The main purpose in this work is to separate the patterns of humans and nonhumans. As Figure 7 shows, ICA is a statistic approach in the higher order and can transform each input image to the combination of bases. For the two major problems we are confronting, one is how to choose the bases with the higher capability in classification. The other is how to enhance the discriminability of independent components between classes

which might not be sorted by the creating sort and might not depend on the binary classification.

Let us have m-training images including both humans and nonhumans with the size $(n_r \times n_c)$. Figure 8 displays the bases of our image set. Reshape all the training data into an N-length vector, and the mixture data $\mathbf{X}$ is an $m \times N$ matrix. Also, the mixture data $x_1, x_2, \ldots, x_m$ are the linear combination of n independent and zero-mean of the source signal $s_1, s_2, \ldots, s_n$ (typically $m \geq n$) as described in

$$x_j = h_{j1}s_1 + h_{j2}s_2 + \cdots + h_{jn}s_n. \tag{11}$$
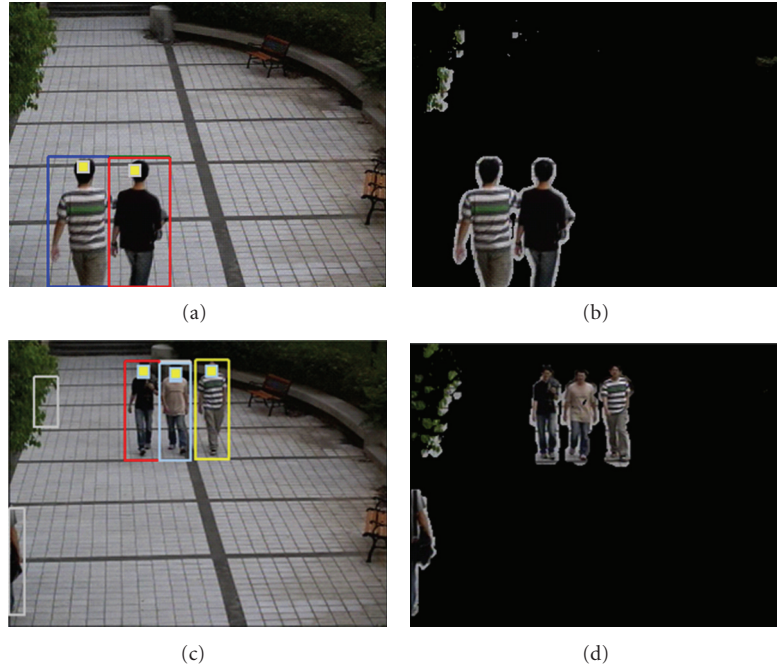
(a)

(b)

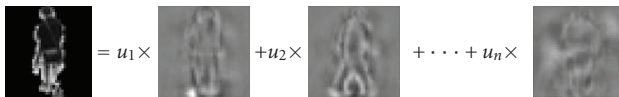(c)

(d)

FIGURE 6: The results of separate humans.

FIGURE 7: ICA-image decomposition.

The matrix $\mathbf{H}$ is expressed in terms of the elements h_{ij}, and it is an unknown full rank $(m \times N)$ mixture matrix. Since all vectors are column vectors and the transpose of $\mathbf{X}$ is a row vector, we can rewrite (11) to (12) by using vector-matrix notations

$$\mathbf{X} = \mathbf{HS}. \tag{12}$$

Without loss of generality, we assume that both the mixture variables and independent components have zero mean and non-Gaussian distributions. For the nonzero mean distributions, the observable variables x_j can always be centered by subtracting the sample mean to become the zero mean distributions. If $\mathbf{W}$ denotes the inverse of the basis matrix $\mathbf{S}$, the coefficients matrix $\mathbf{U}$ for training matrix $\mathbf{X}^{\mathrm{T}}$ will be expressed in

$$\mathbf{U} = \mathbf{WX}^{\mathrm{T}}. \tag{13}$$

The n-component base vectors which have the best distinguishability for detecting humans and nonhumans should be chosen from many candidate components. It can be achieved by calculating the ratios of between-class and within-class variability r for each coefficient, and the largest ratio r implies the best distinguishability. Or the base vectors can be selected by using perceptions in neural network. These two methods

depend on the capability of binary classifiers. In Figure 9, the solid line and dashed line indicate the positive and negative values of ICA coefficients, respectively. If the distribution is like Figure 9(a), we can separate humans from nonhumans by the dotted threshold line in an easier way. Unfortunately, the information provided by the binary classifier is too insufficient to select ICA features. Like what is shown in Figure 9(b), we cannot easily separate the distributions into two classes by using a threshold line. That is also the major reason for us to modify the original ICA by using the information of conditional entropies for selecting the optimal ICA bases in this paper.

If the entropy is the amount of information provided by a random variable, then our conditional entropy can be defined as the amount of information about one random variable provided by another random variable. The entropy of a random variable reflects the more truthful information of the observed variable. If the variable is more random, it means unpredictable and unstructured, which may result in the large entropy value. Figure 10 illustrates how the entropy values are relevant to the distributions of variables. In such a case, the higher entropy value in Figure 10(a) reveals that the variable Z_1 is more random than Z_2.

The 2-D data space obtained from ICA feature extraction needs to be discretized into a matrix of grid cells by separating each dimension into a set of intervals or bins. The discretization process begins with calculating the mean value of data in one dimension and dividing the data into two halves with that mean value. Recursively, each half is divided into halves with its own mean value. The recursion will stop when we obtain the required number of intervals or meet the constraint of total bins. Let a discrete random variable Z be
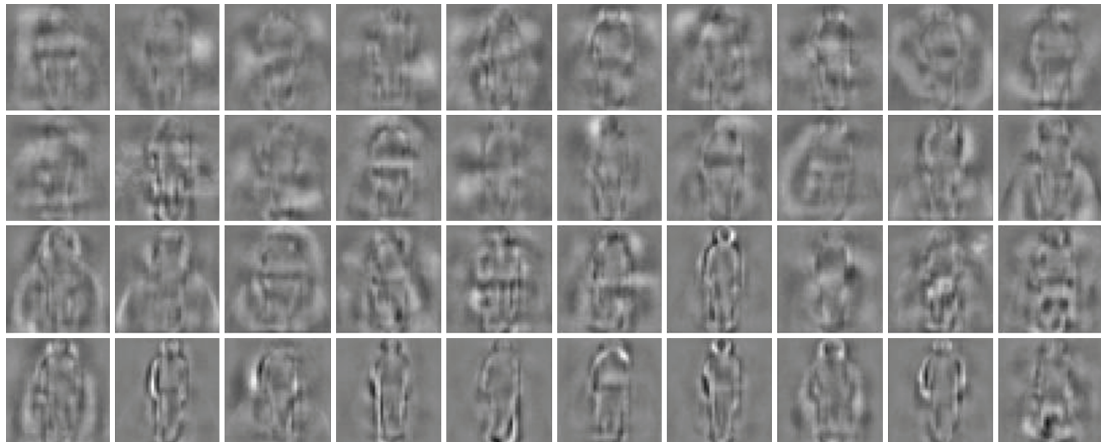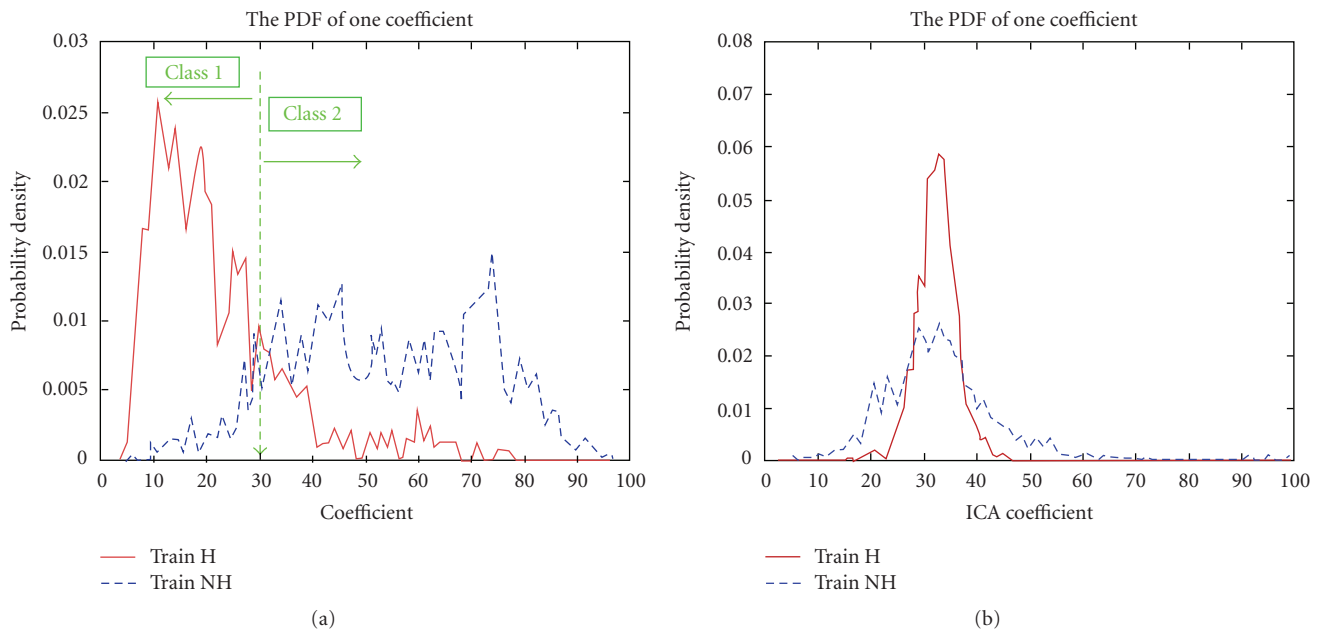
FIGURE 8: The bases of image set.

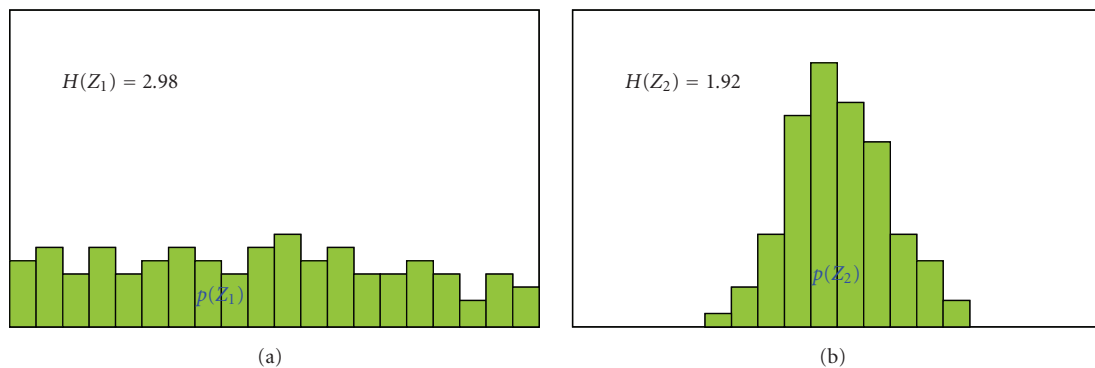FIGURE 9: Different distributions of ICA coefficient in (a) the ideal case and (b) the real case.

FIGURE 10: The information entropy. (a) Z_1 and (b) Z_2.

TABLE 1: Results for feature selection and classification.

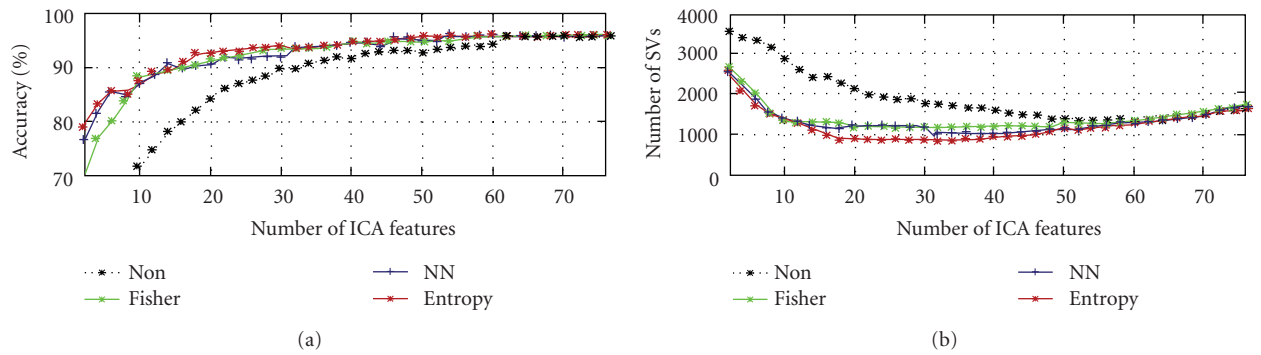| Feature-classifier	Method	Number of SV	Accuracy (%)
20 IC-SVM	Entropy	895	92.58
	Fisher's criterion	1197	91.24
	Neural Network	1198	90.57
	Non	2166	84.07
30 IC-SVM	Entropy	825	93.88
	Fisher's criterion	1154	93.21
	Neural Network	1137	92.20
	Non	1800	89.58

FIGURE 11: Analysis of feature selection. (a) Accuracy rate. (b) Number of SV.

with possible values $\{z_1, z_2, \ldots, z_n\}$. The information entropy of Z with the probability density $p(z)$ is defined in

$$H(Z) = -\sum_{i=1}^{n} p(z_i) \log p(z_i). \tag{14}$$

The conditional entropy quantifies the uncertainty of a random variable Y if given that the value of a second random variable Z is known. Each coefficient has to be normalized to $[-1, 1]$ and quantized to n bins. Let $Y = \{-1, 1\}$ be the desired class, then the conditional entropy can be described in

$$H(Y \mid Z) = -\sum_z \sum_y p(y, z) \log p(y \mid z) = H(Y, Z) - H(Z). \tag{15}$$

The conditional entropy $(Y|Z)$ is a weighted sum of the entropy values in all columns, where the joint entropy is defined by

$$H(Z, Y) = -\sum_z \sum_y p(z, y) \log p(z, y). \tag{16}$$

We sort the conditional entropy $(Y \mid Z)$ and use the sorted results to select corresponding independent components. The coefficients or independent components with the better classification ability are associated with the small conditional entropy. The selected ICA features will be used in the

SVM classifier to indentify humans or nonhumans. Table 1 and Figure 11 showed the comparisons of results by our conditional entropy based feature selection approach with those by others for feature selection and classification. All the comparisons in this paper used the same training and testing database. The training database consisted of 1843 human and 840 nonhuman images. Meanwhile, 3178 human and 2847 nonhuman images were used in the testing database. The same ICA algorithm was used for feature extraction and SVM in classification, and the only difference for obtaining reasonable compared results lied in the feature selection method. Our feature selection approach was based on the conditional entropy, and was compared with Fisher's criterion, neural networks, and without feature selection. Our used parameters in the comparison process were the number of support vectors (SV) and the accuracy rate which was obtained from each method with respect to the number of independent components. Figure 11 showed the accuracy rate and the number of SV for all number of ICA features and indicated the maximum number of ICA to be 76. We chose two subsets of independent components as 20 and 30 and displayed the accuracy rate and number of SV in more details in Table 1. Table 1 exhibited that the conditional entropy based approach had the accuracy rate in more than 90% but needed the smallest number of support vectors (SV). For all approaches, the accuracy rate will increase and the corresponding number of SVs will decrease when the number of ICA features increases from zero to the specific
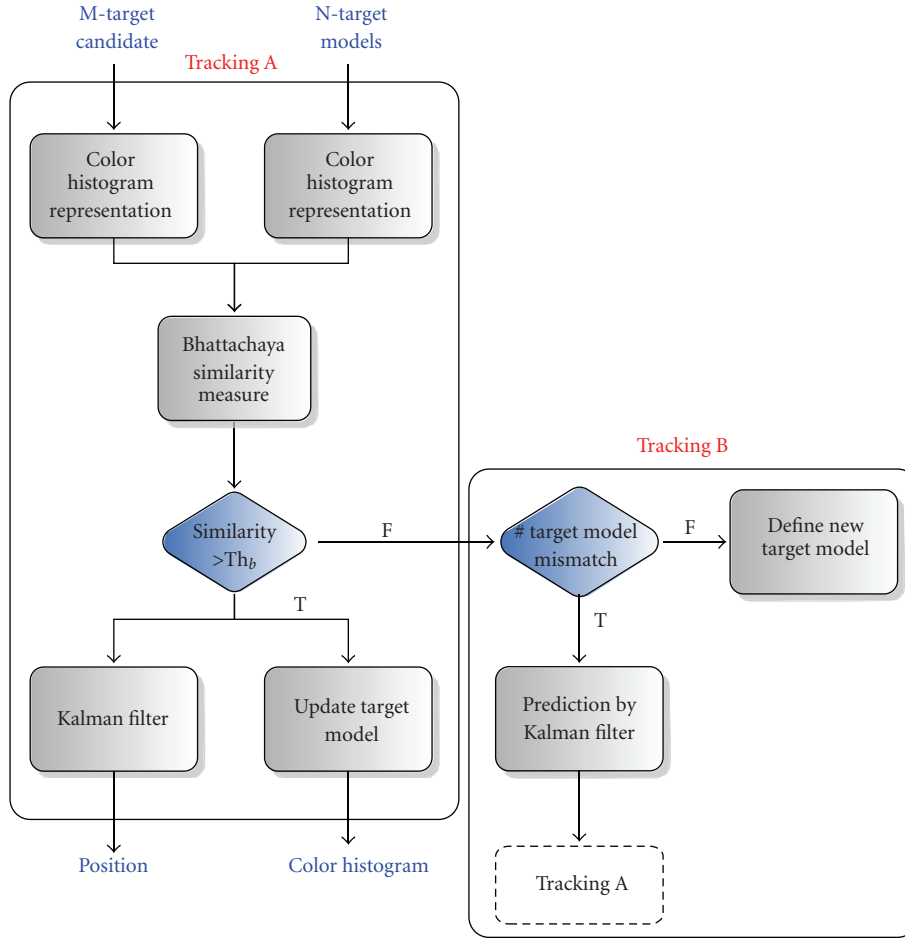
FIGURE 12: The tracking module.

value. When the number of ICA features increases from the specific value to the maximum number, we can use more SVs as to maintain the accuracy rate.

4. Tracking

Our tracking module is depicted in Figure 12. The proposed tracking system is based on the color appearance model because the color distribution will be typically stable under rotations, scaling, or partial occluded conditions. At the same time, Kalman filter is applied to calculate and predict new locations of each moving object, and to solve the occlusion problems which the color models may be invalid with. Let $\text{hist}(i)$ represent the ith bin of total N bins of the color histogram, and the PDF of target models can be computed by

$$p_i = \frac{\text{hist}(i)}{\sum_{i=1}^{N} \text{hist}(i)}. \tag{17}$$

Most of the color features is unstable under the change of lightness. The HSV color channel extracts the lightness information from the RGB color channel, therefore the sensitivity to illumination can be reduced. But the problem
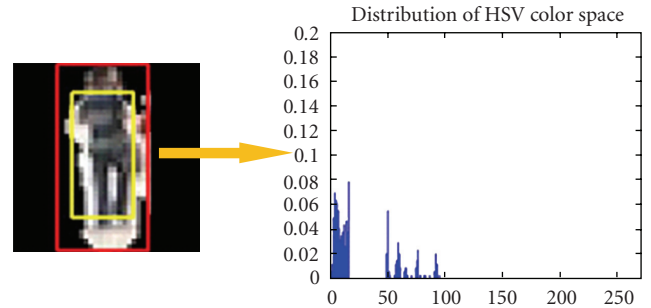

FIGURE 13: The PDF of color histograms.

of HSV color channel occurs when the saturation value S is close to 0. In this condition, the hue H will become quite noisy. Therefore, in practical applications, the HS-histogram will be used only when S is larger than a threshold value 0.1. Otherwise, only intensity V-histogram is used, and the total number of histogram's bin becomes $N_H N_S + N_V$. In order to reduce the computational time and increasing the accuracy of object tracking, we use three of fourth of the original moving object region with the same centers as shown in Figure 13. Moreover, Bhattacharya similarity measure is

FIGURE 14: The positive database.

FIGURE 15: The negative database.

applied to compute the similarity value between two PDF, the target model p_i and target candidate q_i as shown in

$$\mathrm{BC}(p, q) = \sum_{i=1}^{N} \sqrt{p_i \cdot q_i}. \tag{18}$$

When the target candidate q_i and target model p_i are similar, the PDF of target models of moving objects can be updated by the weighting factor γ in the tracking process, which is expressed in

$$p_i = (1 - \gamma) \cdot p_i + \gamma \cdot q_i, \tag{19}$$

where $\gamma = \mathrm{BC}/4$.

5. Experimental Results

Our training database captured from 16 different videos included 1843 positive and 2066 negative data, and the database in the testing phase were captured from 18 videos in 3178 positive and 2847 negative data. The images used in our

database were acquired by considering various conditions and activities such as the detected images contained part of lateral or frontal human shapes, the detected humans were walking or running, the detected moving object did not have a complete human shape, and so forth. We also took the cases under both indoor and outdoor environments into account and meanwhile some nonhuman targets in complicated conditions such like trees, animals, and vehicles were used in the testing database in this paper. All the image data were normalized to the 40×40 block size. The normalization algorithm used in our work was carried out by comparing the width and height of moving-object regions. If the width of moving-object region was larger than the height, the moving object would be centralized by shifting horizontally, otherwise by shifting vertically. We showed several positive and negative images in our database after normalization in Figures 14 and 15.

We also listed the compared results in the number of required features, the accuracy rate, and the detection time by our proposed conditional entropy-based feature selection approach with those by others in Table 2 and
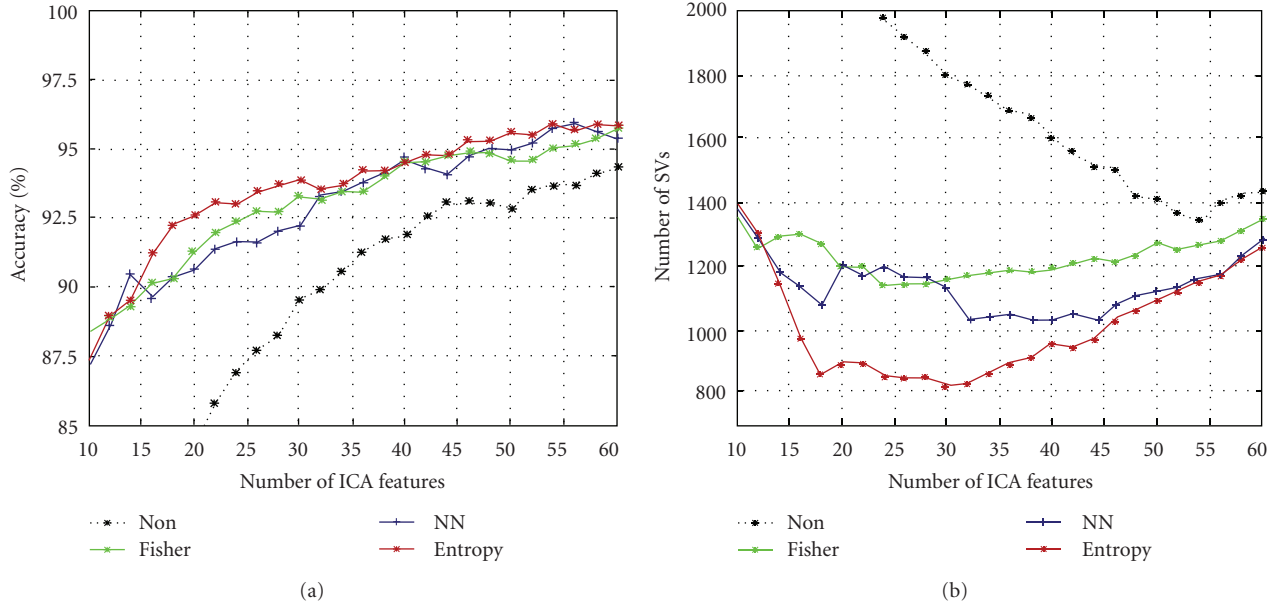
FIGURE 16: Analysis of different methods of feature selection in (a) Accuracy rate, and (b) Number of SVs.

TABLE 2: Comparisons in the computational time.

IC selection method	Number of IC	Number of SV	Accuracy (%)	Detection (ms/object)
Entropy	30	825	93.88	1.13
Fisher	30	1157	93.21	1.33
Entropy	40	958	94.51	1.41
Fisher	40	1194	94.40	1.65
NN	40	1028	94.58	1.51

TABLE 3: Accuracy of human detection system (%).

Method	Training Data		Testing Data	
	Human	Nonhuman	Human	Nonhuman
mICA+SVM	97.72	95.84	94.15	93.57
ICA+cosine	90.87	85.73	90.34	85.49
ICA+SVM	97.55	93.90	93.17	91.13
Codebook	87.95	92.83	90.88	93.68
PCA+BP	99.18	99.46	89.65	94.09

Figure 16. We had 5 videos with a total number of frames, 14056, and the computational time indicated in our entropy based method would be 1.13–1.41 miniseconds depending on the number of independent components (IC). With the increasing number of IC in human feature extraction, the number of support vectors (SV) would also increase, which made the system take longer to detect a human. Moreover, in Table 3, we compared the accuracy of our mICA+SVM approach with that of some others both in the training and testing data. The codebook matching approach in Table 3 used the human shape as the features, and matched the moving object by the code vectors in the codebook. The PCA+BP method used PCA for feature extraction and the back-propagation model in neural networks for classification. In the other two approaches, ICA + Cosine and ICA + SVM, the IC-features were determined by calculating the ratios of between-class and within-class variables r for each coefficient and choosing a larger r as the features with the better distinguishability. After the features have been determined, they used the cosine similarity measurement and SVM for classification, respectively. Table 3 showed the higher accuracy of our mICA+SVM approach in the training part than all the others except PCA+BP. However, in the testing part, our mICA+SVM approach demonstrated the highest accuracy to identify humans among all the compared methods.
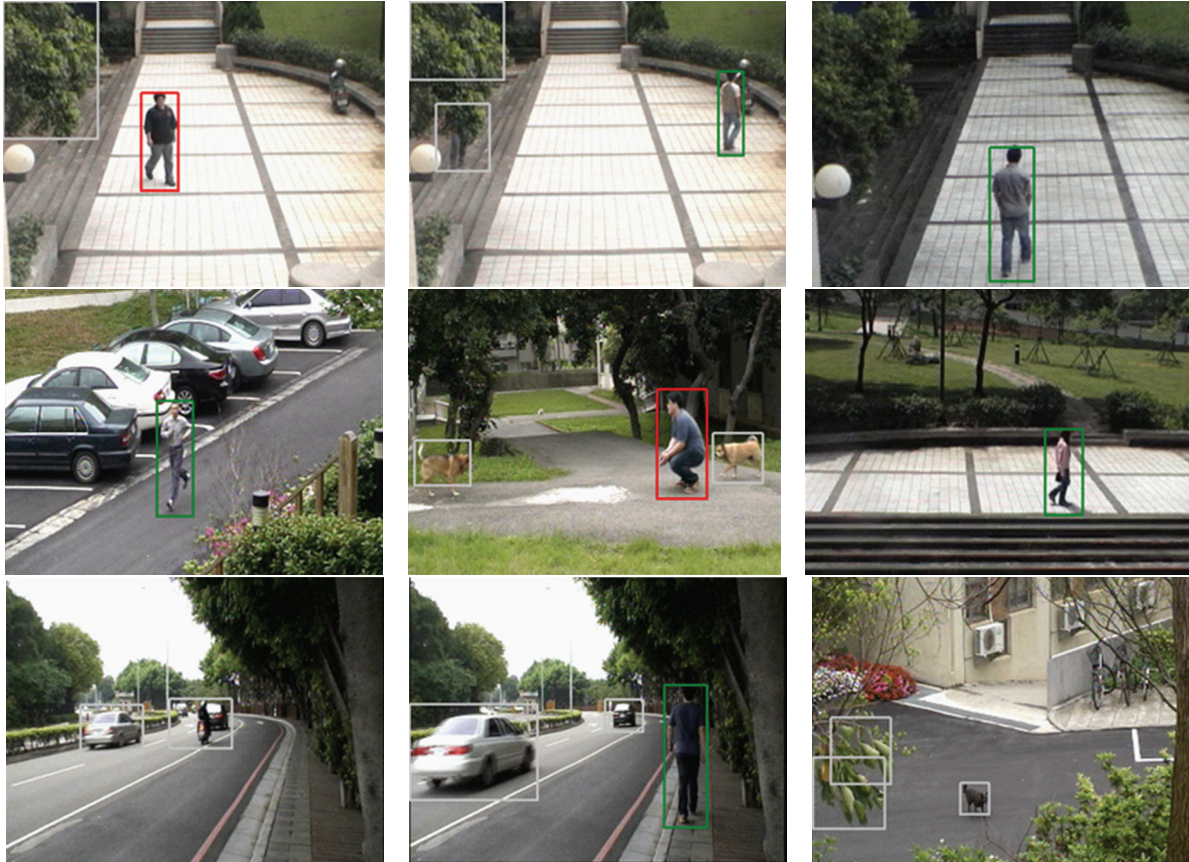
FIGURE 17: The detection results for humans and nonhumans.

FIGURE 18: The processed results for some occlusion cases.

Figures 17–20 showed the human detection results in different conditions, where the white color blocks described the nonhuman moving objects and the blocks in other colors indicated the moving humans. Figure 17 revealed that the proposed human detection system could work well in many kinds of conditions, and our approach would accurately detect humans for cases that the humans were running, walking in different positions and directions, and could correctly recognize the vehicles, moving tree leaves, or animal as nonhuman objects. Figure 18 showed our experimental
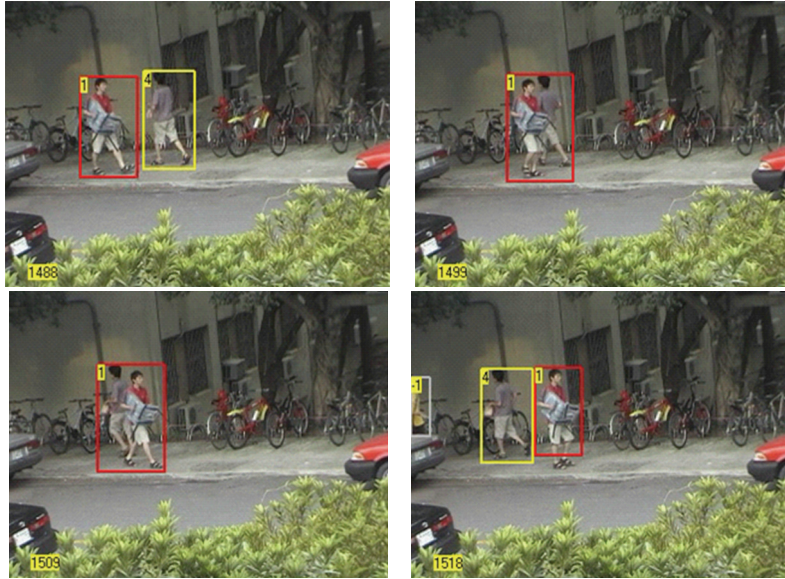
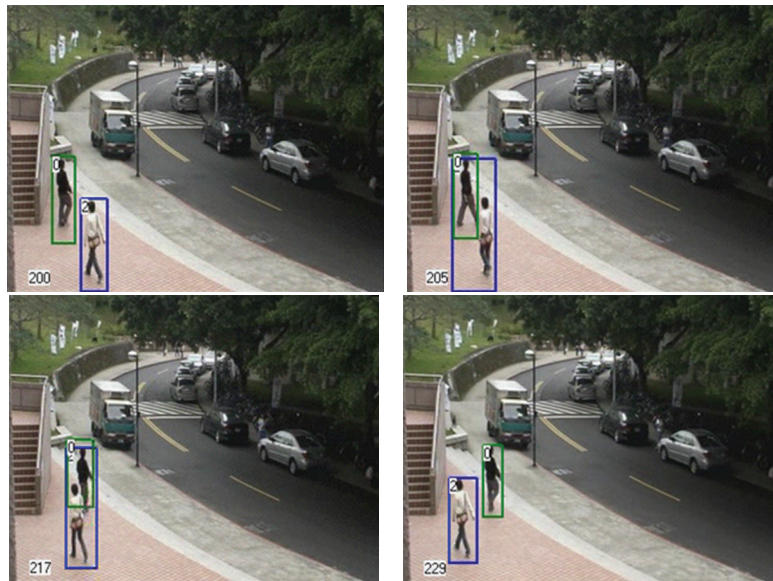FIGURE 19: The results in human tracking—Environment 1.

FIGURE 20: The results in human tracking—Environment 2.

results in the occluded cases where people were occluded by each other or by other objects. Figures 19 and 20 displayed the results of human tracking in consecutive frames where we indicated the number of frames and the label of identified humans in the lower left and the upper left in each image, respectively.

6. Conclusions and Discussions

The modified ICA approach using conditional entropy has been proposed for human detection in this paper. The experimental results have proved the conditional entropy to be effective in sorting features with the better classification ability. The SVM classifier is applied to classify the features into two classes, humans and nonhumans. The Kalman filter and Bhattacharya color similarity measurement are both used to predict and track the humans in the consecutive frames. Our experiments also indicate the higher performance in human detection and tracking. Besides, we use the GMM method which is used to model and update a background image for moving object segmentation to handle the dynamic backgrounds. The color-based shadow elimination algorithm is also implemented in our work to
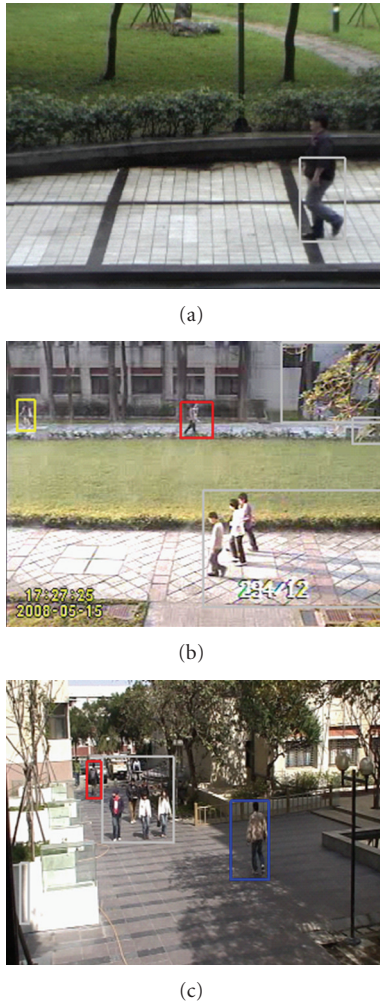
(a)

(b)

(c)

Figure 21: The negative examples in human detection.

reduce the influences of grouping shadows by connected components effectively. In order to make our approach much more practical and perfect, in the near future we may consider more conditions such as the clothing colors of detected humans are close to those of the backgrounds (Figure 21(a)), the shadowed regions of detected humans are much larger than the truthful moving objects (Figure 21(b)), and the heads of detected humans in the sampled images are too small to be detected more accurately (Figure 21(c)). To sum up, the conditional entropy-based mICA approach has solved most problems in human detection and provides the better discriminability in classes for ICA which may not depend on the binary classification in an efficient computational time, 1.13–1.41 ms/object, and in the accuracy of more than 93% for real-time applications.

Acknowledgments

References

[1] L. Zhao and C. E. Thorpe, "Stereo and neural network-based pedestrian detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, no. 3, pp. 148–154, 2000.

[2] C. Orrite-Uruñueta, J. M. del Rincón, J. E. Herrero-Jaraba, and G. Rogez, "2D silhouette and 3D skeletal models for human detection and tracking," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, pp. 244–247, Cambridge, UK, August 2004.

[3] Z.-L. Jiang, S.-F. Li, and D.-F. Gao, "A time saving method for human detection in wide angle camera images," in *Proceedings of the International Conference on Machine Learning and Cybernetics*, pp. 4029–4034, Dalian, China, August 2006.

[4] Y.-L. Tian and A. Hampapur, "Robust salient motion detection with complex background for real-time video surveillance," in *Proceedings of the IEEE Workshop on Motion and Video Computing (MOTION '05)*, pp. 30–35, Breckenridge, Colo, USA, August 2005.

[5] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, pp. 246–252, Fort Collins, Colo, USA, June 1999.

[6] J. Zhou and J. Hoang, "Real time robust human detection and tracking system," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pp. 246–252, San Francisco, Calif, USA, 2005.

[7] R. Polana and R. Nelson, "Detecting activities," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2–7, New York, NY, USA, June 1993.

[8] M. Bertozzi, A. Broggi, M. D. Rose, M. Felisa, A. Rakotomamonjy, and F. Suard, "A pedestrian detector using histograms of oriented gradients and a support vector machine classifier," in *Proceedings of the 10th International IEEE Conference on Intelligent Transportation Systems (ITSC '07)*, pp. 143–148, Seattle, Wash, USA, October 2007.

[9] C. Curio, J. Edelbrunner, T. Kalinke, C. Tzomakas, and W. von Seelen, "Walking pedestrian recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 1, no. 3, pp. 155–163, 2000.

[10] S. M. Yoon and H. Kim, "Real-time multiple people detection using skin color, motion and appearance information," in *Proceedings of the IEEE International Workshop on Robot and Human Interactive Communication*, pp. 331–334, Tokyo, Japan, September 2004.

[11] D. Guo, "Coordinating computational and visual approaches for interactive feature selection and multivariate clustering," *Information Visualization*, vol. 2, no. 4, pp. 232–246, 2003.