Human Verification in a Video Surveillance System Using Statistical Features

Sanpachai Huvanandana

Abstract—A human verification system is presented in this paper. The system consists of several steps: background subtraction, thresholding, line connection, region growing, morphlogy, star skelatonization, feature extraction, feature matching, and decision making. The proposed system combines an advantage of star skeletonization and simple statistic features. A correlation matching and probability voting have been used for verification, followed by a logical operation in a decision making stage. The proposed system uses small number of features and the system reliability is convincing.

Keywords—Human verification, object recognition, video understanding, segmentation.

I. INTRODUCTION

VIDEO understanding is the process of understanding a video sequence. The procedure can be included any of the following process: detecting, segmenting, representing, and understanding an object. A human verification is a challenge research in machine vision, especially in an opened environment, which requires a process of image and video understanding. In general, a human verification system consists of several steps i.e. object segmentation, feature extraction, and human verification. Several researches topics have been explored in human verification [1-15]. A human body is segmented from a background and a human shape has been used for verification. Object segmentation and human verification can be done in several ways. Rosin et. al. and Borghys et. al. proposed thresholding after the frame difference to reduce noise in object segmentation. Haritaoglu et. al. categorized the background by representing each pixel into maximum intensity value, minimum intensity value and intensity difference values between consecutive pixels. This method is sensitive to illumination changes. Oliver et. al. proposed an eigenspace model for moving object segmentation. A drawback of this method is incomplete dynamic scenes modeling. A statistical method was proposed in [14]. Each point in a scene is modeled by using a Gaussian distribution with an estimated mean intensity value. The proposed method can only handle unimodal distributions. A mixture of Gaussians is also proposed by [7].
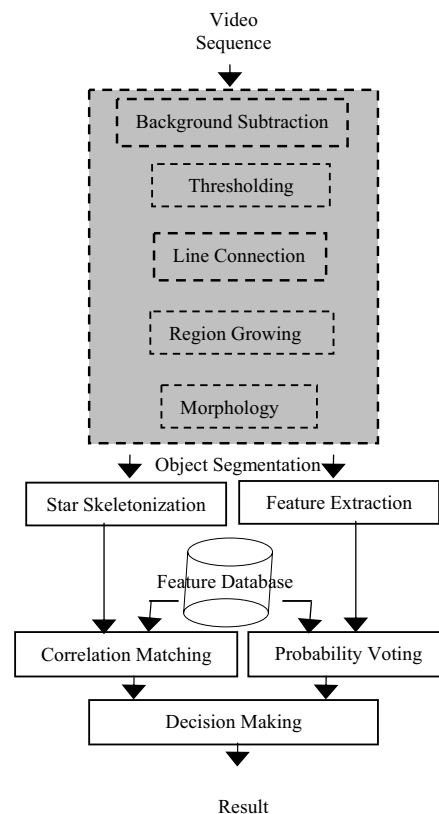

Fig. 1 The proposed human verification system

Elgammal et. al. used sample background images to estimate a probability of observing pixel intensity values in a nonparametric manner. A geometrical structure has been used for human detection in [1] and Mahalanobis distance is calculated as a human score. Constantine et. al. proposed a wavelet-based human detection method. Chen et. al. introduced a human detection based on fuzzy rules therefore a system is suitable for one moving object. Fengliang et. al. extracted depth and gray information followed by split and merge strategy for human detection. A combination of optical flow and background subtraction for human extraction has been used in [10] followed by neural network in verification stage. Our proposed human verification system is shown in Figure 1.

S. H. Author is with the Department of Electrical Engineering and Information Technology, Chulachomklao Royal Military Academy, Nakon Nayok, 26001 Thailand (corresponding author to provide phone: 6637-393-024; e-mail: shuvanan@crma.ac.th).

This paper has been categorized into 6 sections. In section 2, an object segmentation algorithm is explained. Section 3 discusses a proposed feature extraction framework. A human verification system is introduced in section 4. The experimental results are shown in section 5, followed by conclusions in section 6.

II. OBJECT SEGMENTATION

A. Background Subtraction

Background subtraction can be done by a Principle Component Analysis (PCA) which dimension can be reduced. After applying PCA, non-moving pixels have been distinguished. The resulted image has been subtracted from each frame to create a moving object. Let $I_{t+N\Delta}(x,y)$ be a Nth frame, a background can be found as following steps:

- Calculate an average (μ_b) of N sampling images
- Normalize an average image and transform into column matrix (A)
- Calculate covariance matrix (C)
- Use Covariance matrix to calculate eigenvalues and eigenvector
- Collect M highest Eigen-vectors and use them as background information

Let $I_t(x,y)$ be a new image entered into a system, a reconstructed image $I_t'(x,y)$ is obtained using M eigenvectors information. A moving object, $D_i(x,y)$, is a difference between $I_t(x,y)$ and $I_t'(x,y)$, as the following equation:

$$D_i(x,y) = \left| I_{t_i}(x,y) - I_t'(x,y) \right| \tag{2.1}$$

The details can be found in [11].

B. Thresholding

The difference image, $D_i(x,y)$, is divided into a block of m by n. An ith threshold image, $f_i(x,y)$, is obtained from

$$f_i(x,y) = \begin{cases} 0 \ (background) \ if \ D_i(x,y) \le th \\ 1 \ (foreground) \ if \ D_i(x,y) > th \end{cases} \tag{2.2}$$

$$th = \frac{(1+\rho)}{mn} \sum_{i=1}^{m}\sum_{j=1}^{n} B_k(x,y) \tag{2.3}$$

where th is a threshold and $B_k(x,y)$ is a kth block of difference image.

C. Line Connection and Region Growing

Line connection is the process of enhancing a binary image. Let G_x and G_y be the gradient of x and y-axis, a magnitude gradient (∇f) can obtained from

$$\nabla f = \left(G_x^2 + G_y^2 \right)^{1/2} \tag{2.4}$$

and a line connection can be calculated from

$$\left| \nabla f(x,y) - \nabla f(x_0,y_0) \right| \le E \tag{2.5}$$

$$\left| \alpha(x,y) - \alpha(x_0,y_0) \right| \le A \tag{2.6}$$

where $\nabla f(x_0,y_0)$ be the magnitude gradient at x_0^{th} and y_0^{th} position, α is the angle of the gradients, E and A is a constant. A pixel value is updated as a part of object if (2.4) and (2.5) is true. After line connection, a border of the object has been connected and region growing [12] has been used to filled object hole(s).

D. Morphology

A dilation $(\oplus)$ and erosion $(\ominus)$ technique in image morphology has been used to enhance threshold image respectively. The 4-connected neighboring of one has been used as a structure element. Figure 4 shows an example of object segmentation.

III. FEATURE EXTRACTION

After segmentation, the following features have been extracted from each image.

A. Star Skeletonization

A centroid (c_x, c_y) is obtained by averaging object positions. A distance at i^{th} frame (d_i) between a centroid and a border has been sampled to create signature [12]. The resolution of signature depends on the number of angle sampling (θ_i). In our experiment, we start from 0 degree and keep a distance between centroid and 5 highest concaves. The details can be found in [9].

B. Movement rate

A movement rate (v_i) can be found by calculating a major and minor axis. A block of object at i^{th} frame (b_i), a human body movement, usually has stable rate of velocity, i.e. left to right, right to left, back to front, and front to back. Figure 3(a) shows an example of object block.

C. Major and Minor Axis Slope

Let (x_1, y_1) and (x_2, y_2) be the highest and lowest coordinate of the object block (b_i), a slope at i^{th} frame (φ_i) can be obtained from

$$\varphi_i = \tan^{-1}\frac{y_2 - y_1}{x_2 - x_1} \tag{3.1}$$

A major and minor axis is shown in figure 2.

D. Major and Minor Axis Ratio

The ratio between a major and minor axis can described as

$$\gamma_i = \frac{l_{major}}{l_{minor}} \tag{3.2}$$

where l_{major} and l_{minor} be a length of major and minor axis, respectively.

E. Body and Head Ratio

The object block (b_i) is divided into 8 vertical sub-blocks, an approximated head size can be found by calculating the length of object in a first line from the top, see Figure 2. The ratio between body length and head length, (κ_i), is obtained.

F. The Feature Sets

The features are divided into 2 separate groups. Let $f_1(x)$ be a set of star skeletonization features which are sampled into 5 features that have highest concaves [9], and $f_2(x)$ be a set of features in section 3.B thru section 3.E which has 4 features. Both set of ith human features can be written as

$$f_{1i}(x) = \left\{ d_{i\theta_1} \; d_{i\theta_2} \; ..... \; d_{i\theta_5} \right\} \tag{3.3}$$

$$f_{2i}(x) = \left\{ v_i \; \varphi_i \; \gamma_i \; \kappa_i \right\} \tag{3.4}$$

The second set of feature $f_2(x)$ has been recorded in term of averaging value of each video frame.

G. A Verification System

A verification system has been separated into 2 sub-systems, a correlation matching and a probability voting. A correlation between two set of feature is written as:

$$f_{1i} \circ f_{1j} = \sum_{m=0}^{4} f_{1i}^*(m) \, f_{1j}(x+m) \tag{3.5}$$

A Probability P_1 is defined as:

$$P_1 = \begin{cases} = f_{1i} \circ f_{1j} & if \; f_{1i} \circ f_{1j} \geq 0 \\ = 0 & otherwise \end{cases} \tag{3.6}$$

Let P_{f2} be a probability of accept criterion, i.e. the feature that an object is verified as a human, P_2 be an average probability voting of all statistic features, the probabilities P_{f2} can be written as:

$$P_{f2} = \frac{\text{number of accept criteria}}{\text{total sampling}} \tag{3.7}$$

$$P_2 = \frac{1}{N} \sum_{i=0}^{N-1} P_{f2} \tag{3.8}$$
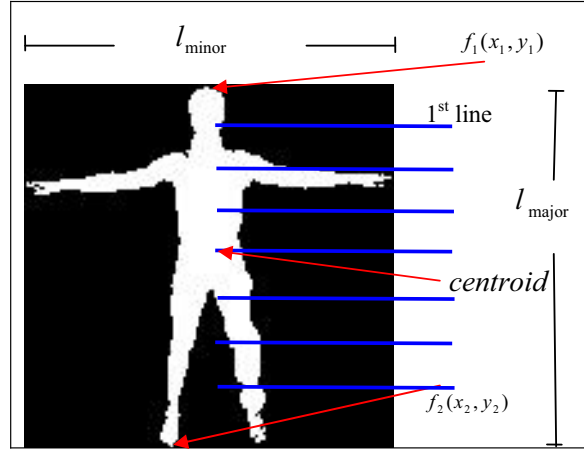
where N is the number of features used.

Fig. 2 Statistic features extraction details.

H. A Decision Model

A decision model uses 3 different logical operation criteria for decision making, AND, OR, and probability weighting. A probability weighting (P_w) can be written as:

$$P_w = aP_{f1} + bP_{f2} \tag{3.9}$$
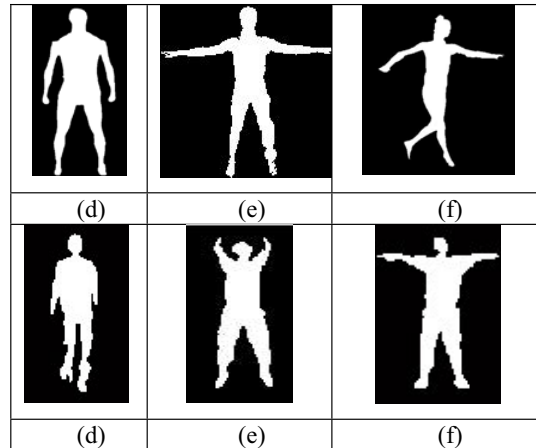

Fig. 3: (a) - (c) example of tested images.

Fig. 4: (a)-(f) example of segmented images.

IV. EXPERIMENTAL AND EVALUATION RESULTS

For our experiments, 300 video sequences have been collected, 270 video sequences contain human body and 30 video sequences contain non-human. Figure 3 show an example of video input. One hundred and fifty video sequences have been used as a training set. We have calculated an average of each feature in the training set and used them as referenced features for human and non-human. For the testing set, all features are obtained as described in section 3. We calculate a suitable threshold for optimum correlation efficiency in training database and use it as a threshold for correlation matching as in equation (3.6). For statistic features, as described in section 3.B thru section 3.E, we use probability voting in equation (3.7) and (3.8). The constraint we have used for acceptance criteria come from the averaging feature values found in training database with $\pm 20\%$ variation. A logical operation is used in a decision model. For AND operation, the verification criteria need to be accepted in both decision models, while either P_1 or P_2 criteria is accepted for OR operation. For probability weighting, P_w is calculated, if a decision value is greater than a certain threshold, the system is accepted as a human. The simulation results with separate system has been shown in Table 1 and the simulation results of different decision making criteria are shown in Table 2. As the results, the proposed system achieves a good verification rate.

TABLE 1 SIMULATION RESULTS USING SEPARATE VERIFICATION SYSTEM

Verification System	False Acceptance (%)	False Reject (%)
Correlation Matching	0.6	2.3
Probability Voting	0.4	3.4

TABLE II SIMULATION RESULTS OF A DIFFERENT DECISION MAKING CRITERIA.

Decision Making Criteria	False Acceptance (%)	False Reject (%)
OR	0.8	3.5
AND	0.3	3.7
Probability Weighing	0.3	2.6

V. CONCLUSIONS

We have developed a human verification system. The system consists of several sequential steps: background subtraction, thresholding, line connection, region growing, morphlogy, star skelatonization, feature extraction, correlation matching, probability voting, and decision making. The proposed system combines an advantage of star skeletonization and simple statistic features. Our system distinguishes itself from the type of features used. The experimental results show that our proposed verification system needs less computation to verify moving object movement as a human than traditional methods. It is also practical for a real-time verification system. The drawback of our system is a sensitivity of an opened environment which causes some noises in background subtraction. An additional process could be added for noise reduction.

REFERENCES

[1] Akira U. and Nobuji T., "Human Detection using Geometrical Pixel Value Structure", IEEE Internation Conference on Automatic Face and Gesture Recognition, FGR'02, 2002.
[2] Borghys, D. , Verlinde, P., Perneel, C. , Acheroy, M. , "Multi-level Data Fusion for the Detection of Targets using Multi-Spectral Image Sequences", Opt. Eng., Vol. 37, No. 2, pp. 477-484, February 1998.
[3] [3a] Chen X, He Z, Anderson D, Keller J & Skubic M, "Adaptive Silhouette Extraction and Human Tracking in Complex and Dynamic Environments", International Conference on Image Processing, Atlanta, Georgia, October 8-13, 2006.
[4] Constantine P., Theodoros E., and Tomaso P., "A trainable pedestrian detection system", In *Proc. of Intelliggent Vehicles*, pages 241–246, 1998.
[5] Elgammal A., Harwood D., and Davis L.S., "Non-Parametric Model for Background Subtraction.", International Proc. IEEE ICCV'99 Frame-Rate Workshop, 1999.
[6] Fengliang Xu, Kikuo Fujimura," Human Detection Using Depth and Gray Images.",IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS' 03), 2003.
[7] Grimson W. E. L. and Stauffer C., "Adaptive Background Mixture Models for Real-time Tracking.", Proc. IEEE Conference CVPR, Vol. 1 ,pp.22-29, 1999.
[8] Haritaoglu I., Harwood D., and Davis L.S., "W4: Real-Time System for Detection and Tracking People in 2 ½ D.", 5th European Conference on Computer Vision, 1998.
[9] Hironobu Fujiyoshi and Alan J. Lipton, "Real-time Human Motion Ananlysis by Image Skeletonization", IEICE Transanction on Information & System, vol.E87-D, No.1, 2004.
[10] Jianpeng Z. and Jack H., "Real Time Robust Human Detection and Tracking System", Proceeding of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition workshop (CVPR'05), Vol.3.
[11] Oliver N., B. Rosario, and A. Pentland, "A Baysian Computer Vision System for Modeling Human Interactions.", International Conference on Vision System, 1999.
[12] Rafael C. Gonzalez and Richard E. Woods, "Digital Image Processing", Prentice Hall, 2002.
[13] Rosin, P.L. , Ellis, T. , "Image Difference Threshold Strategies and Shadow Detection", British Machine Vision Conf., pp. 347-356, 1995.
[14] Wren C.R., Azarbayejani A., Darrell T., and Pentland A., "Pfinder: Real-time Tracking of Human Body.", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 19, No.7, pp.780-785, July 1997.
[15] Wu B. and Nevatia R, "Detection of Multiple, Partially Occluded Humans in a Single Image by Basian Combination of Edgelet Part Detectors" ICCV'05, 2005.

Sanpachai Huvanandana received a scholarship from the Royal Thai Army to pursue a MS.EE. and Ph.D. in the United States of America. He received a Ph.D. in electrical engineering at university of Washington in 2002. His research interests are biometrics technology, machine vision, pattern recognition, and data mining. He is currently an instructor in electrical engineering and information technology at Chulachomklao Royal Military Academy.