A Human Detector Based on Flexible Pattern Matching of Silhouette Projection

Hajime Ohata*, Nobuyoshi Enomoto**, Akio Okazaki**
*Yanagicho Works, **Multimedia Engineering Lab.
Toshiba Corporation
70 Yanagi-cho Saiwai-ku Kawasaki, 210, Japan

Hiroaki Kawasumi, Shigeo Sudo, Yoshiaki Yamada
Computer and Communications R & D Center
Tokyo Electric Power Company
4-1 Egasaki-cho Tsurumi-ku Yokohama, 230, Japan

ABSTRACT

In this paper, we propose a method for detecting a human being using silhouette projection pattern matching, and present experimental results of its application in video surveillance systems.

There are some studies on detecting intruders in video images by extracting changing regions which follow moving targets. However, detection errors caused by small animals, changes in brightness due to sunshine, or the swaying of trees in the wind cannot be avoided. Several methods based on additional constraints such as the area or the rectangle surrounding the changing region are proposed, but they are not sophisticated enough to overcome the above problems.

Our method is based on silhouette projection pattern matching of the changing regions to reduce erroneous detection. Although the shape of human silhouettes change as the person moves, the widths of certain parts of the body, which are elements of the projection, are stable. Therefore, we propose a flexible pattern matching algorithm where different allowance levels are set to different parts of the body. This method is simple and takes short calculating time. The algorithm was implemented in specially designed hardware which was able to extract and evaluate the changing regions in real-time. The experimental results prove the effectiveness of this method.

1. INTRODUCTION

Recently, many remote video surveillance systems using TV cameras have been installed in important areas such as public facilities. In these conventional systems, security guards had to continuously watch CRT monitors in order to not miss any security breaches. To reduce security guards' labor, automated video surveillance systems have come into demand.

Some studies have been made on automatic detection of changing regions (by changing regions, we mean large regions in a video frame where the pixel values differ from those of the previous frame) in a scene using a moving image analysis technique. However, automatic systems based on the methods in these studies are not always acute. They detect small animals such as dogs and cats, changes in brightness due to sunshine, or the swaying of trees in the wind, in addition to intruding human beings. To reduce erroneous detection, a human detection method using information of the size of the minimal rectangle surrounding the block of changing regions is often adopted for processing simplicity.[1] However, it is not a very satisfactory technique because there could be many spurious events in the real world, especially in outdoor scenes.

One approach to improve human detection accuracy is to extract more shape information from each changing region during the recognition stage. Silhouettes of changing regions hold a large quantity of shape information that can be recognized as human. However, silhouette pattern matching is not only time consuming, but is also difficult to implement due to the problems in constructing flexible template patterns.

2. FLEXIBLE PATTERN MATCHING FOR HUMAN DETECTION

Fig. 1 shows a variety of silhouettes and their projection patterns of people walking. Though the silhouette projections have different patterns due to changes in motion and clothing, sub-patterns corresponding to several parts such as the head and the neck are relatively stable. Our basic idea is to use the projection patterns of silhouettes for flexible pattern matching and obtain a score (= similarity) to detect human objects.

Scores of the observed projection patterns are as follows: high scores should be given to the portions where the projection value is stable, such as the head or neck, if the observed value matches the reference value. On the contrary, low scores should be given to

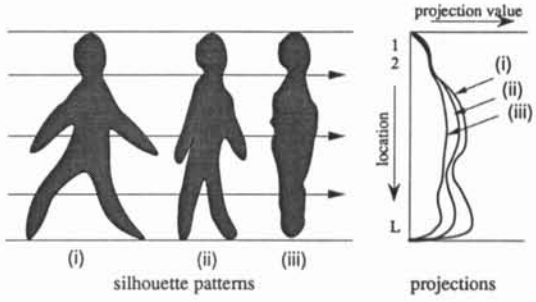


Fig. 1 Examples of Silhouette Patterns and Projections

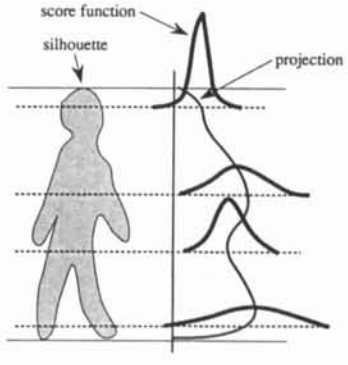


Fig. 2 Score Functions

unstable portions such as the torso or legs because the matching range is wide. We can obtain the above scores by adopting different similarity evaluation functions for different parts of the body. In Fig. 2, similarity is evaluated using score function g_i s, where i refers to the different part of the body. The score functions are high and variance scores are small for the head and neck parts, and score functions are low and variance scores wide for the torso and legs. Using g_i , the total similarity score S' is given by Eq.1.

$$S' = \sum_{i=1}^{L} g_i (p_i) \qquad \text{(Eq. 1)}$$

where $P = [p_1, p_2, ..., p_L]$: projection pattern

g_i : score function at position i

The final determination of whether a human has been detected is carried out by thresholding S'. That is, if $S' > T$ (T: threshold value), a human object has been detected. In the digital image, each element p_i of the projection pattern P is an integer, and Eq.1 is converted into Eq. 2 using L x M (M is an upper limit of p_i) matrix G (i, j) whose (i, j) element is g_i (j). Here we refer to G as a score matrix.

$$S' = \sum_{i=1}^{L} G (i , p_i) \qquad \text{(Eq. 2)}$$

In our method, similarity evaluation is accomplished by obtaining the projection, normalization, and reference of score matrix G. Therefore, this method gives us both high processing speed and detection accuracy.

3. CONSTRUCTION OF SCORE MATRIX

A method to construct the score matrix G is shown in Fig. 3. First, a variety of human motion data was collected and silhouettes extracted. Since the sizes in the sample data vary in height from each other, normalization is carried out on the basis of this height. Next, projection patterns are derived and converted into L x M matrices G'_k (k=1,2,..,N; N is the number of human samples). Finally, the score matrix G is calculated by the iterative operation in Eq. 3.

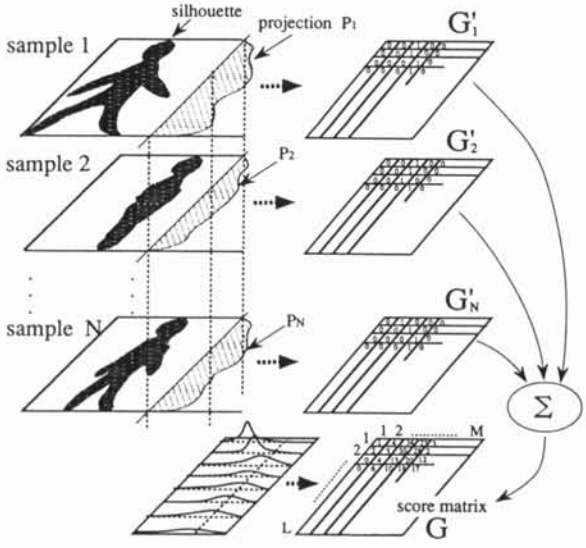


Fig. 3 An Example of Score Matrix G

$$G = \sum_{k=1}^{N} G'_k \qquad \text{(Eq. 3)}$$

Each row of matrix G is the score function for each part of the body. It is easily observed that the functions are not similar, indicating that the proposed pattern matching method uses a flexible technique in dealing with these various types of sub-patterns. In Eq. 3, the score depends on L and N, so that the final total score S of the specified changing region is given by Eq. 4.

$$S = \frac{S'}{L \times N} \qquad \text{(Eq. 4)}$$

4. EXPERIMENT - I

To evaluate our method's effectiveness, we made a simulation on a work station using human and non-human silhouettes. Sequential differential extraction using three successive images as shown in Fig. 4 was used to extract changing regions. Making vertical and horizontal projections and thresholding the projection value was done twice to get a minimal rectangle surrounding the changing region. Projection patterns derived in the second horizontal direction were the projection patterns of the changing region. Normalization as explained in section 3 was accomplished in the projection pattern.

Since silhouette patterns depend on the direction the person is walking, two score matrices were constructed that correspond to the two directions; one for left/right movement and the other for near/far movement. About 450 projection patterns from 15 people were used to construct the score matrices. Environments where score matrices were constructed, and the score matrices themselves, are shown in Figs. 5 and .6. In the environments, we evaluated actual human and non-human objects which had almost the same size for the surrounding rectangle; about 350 patterns from 12 people for left/right movement and about 250 patterns from 9 people for near/far movement, which had not been used to construct cost matrices. We also evaluated 27 non-human patterns such as parts of vehicles or folklifts. Fig. 7 shows the distribution of scores for human and non-human objects in left/right movement. The number of intersections of human and non-human objects was few, and 70% of the non-human objects could be eliminated (false-detection ratio was 30%) if the score threshold was set at the minimum score for human objects. Since the results were almost the same for the near/far movement, sufficient flexibility of this pattern matching method was proved.

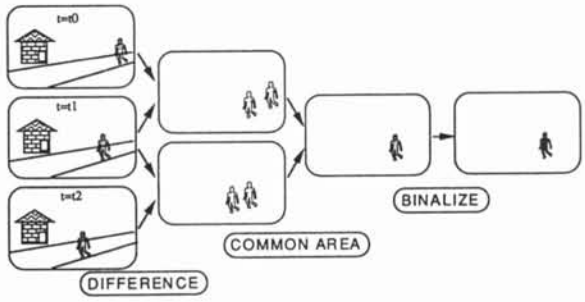


Fig. 4 A Method to Extract Moving Objects

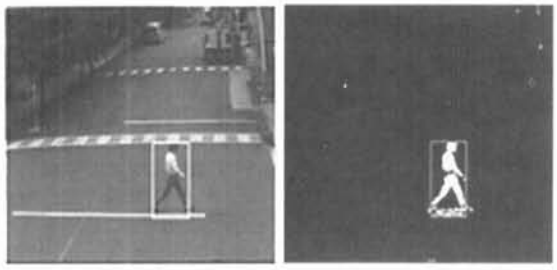


Fig. 5 Environments where Score Matrices were Constructed

(a) for left/right direction (b) for far/near direction

Fig. 6 Score matrices (High contrast areas represent High scores)

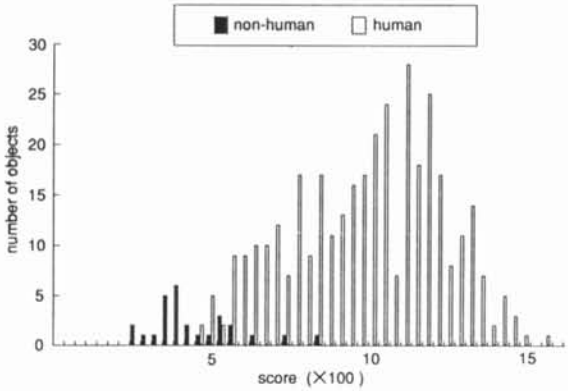


Fig. 7 Distribution of Scores

5. SPECIALLY DESIGNED HARDWARE

To apply the proposed matching method to actual surveillance sites, a normal work station is not appropriate for implementation. A prototype system was constructed instead by installing the above algorithm on specially designed hardware.[2] This system consists of a video I/F, an image memory unit, a main processor, an image processor, and image processing hardware, as shown in Fig. 8. For high processing speed and algorithm flexibility, difference-binary-images were processed on the processing hardware, and the changing regions extraction and projection pattern evaluation was executed by the main processor and the image processor. The difference-binary-image was made on the processing hardware within 1 v-sync period using plural image memories, and then transferred to the image processor. At the image processor, minimal rectangles surrounding the changing regions were derived and classified by size. All changing regions that have rectangle sizes within a pre-determined range were classified as suspicious objects (=human candidates). This information was sent to the main processor for pattern matching evaluation. The main processor normalized the projection pattern to obtain the score S from a table selection and determined whether the object was human by thresholding S by T. Finally, information of the area detected, such as the location and the size of the object, was indicated on the monitor through the video I/F.

Here, the direction the object was moving was not known, and the projection patterns were evaluated by both of the score matrices, with the higher score being chosen. To shorten calculation time, normalization was also done by table sampling. Processor computing time depended on the number of changing regions; it was almost 40-90msec for the image processor, and 40-60msec for the main processor. Spare time remains for the main processor can be used for other jobs such as camera control, image recording, and for the image processor to do additional analysis.

At the field testing, computed results and corresponding images should be recorded to analyze performance, therefore log recording onto FD and image recording onto VCR through image delay equipment were attached during the experiment.

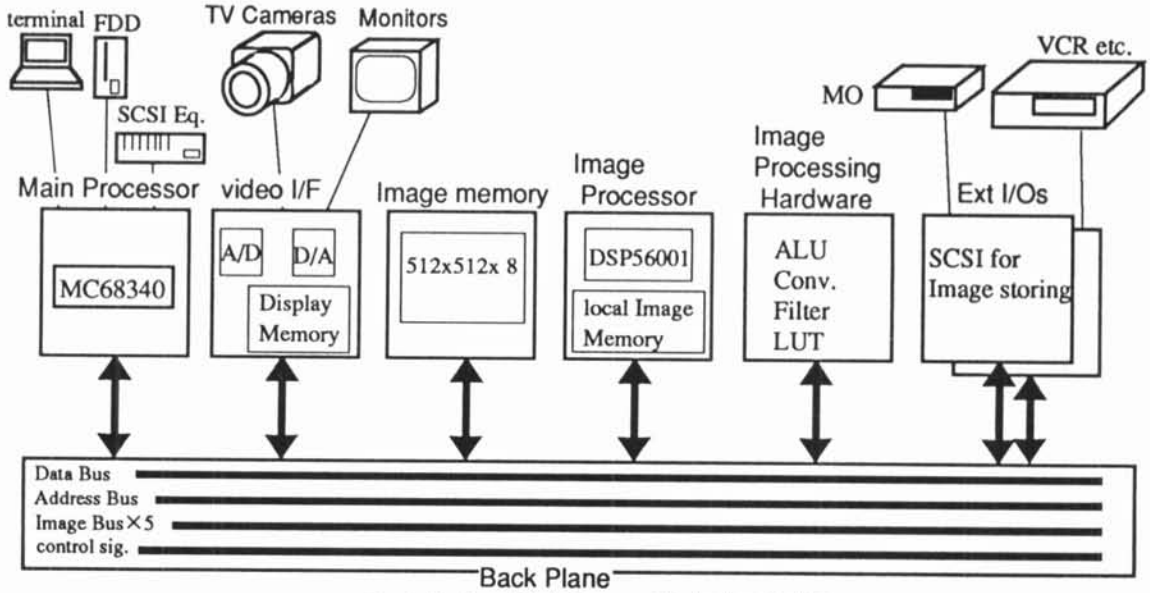


Fig. 8 Hardware Block Diagram of Proto-type System

6. EXPERIMENT - II

The second experiment was carried out at the entrance of a building using the proto-type surveillance system at night time in August. Table 1 shows the experimental conditions and the results. Fig. 9 shows example images of human detection and non-human detection (the reflection of lamps in puddles).

There are two types of detection errors; non-detection and false-detection. Our method is based on evaluating changing regions, these ratios were determined by the number of suspicious areas which were actually human or non-human objects. We calculated the error ratio as shown in Eq. 5.

$$\text{non-detection error ratio} = \frac{Hs}{Ht}$$

$$\text{false detection error ratio} = \frac{NHs}{NHt} \qquad \text{(Eq.5)}$$

where
Ht: total number of human objects
NHt: total number of non-human objects
Hs: number of human objects whose score < T
NHs: number of non-human objects whose score > T

The environmental conditions were too severe for the extraction stage to work well because of darkness, vehicle headlights, and autoiris lens. Especially, the normalization process would not work correctly if the heights of human objects were wrong. Therefore, we also evaluated the non-detection error ratio for humans whose head and legs were successfully extracted.

Fig. 10 shows the above error ratios against the threshold level T of the scores. If we choose T as 0.04, the non-detection error ratio based on the total number of human objects was 20%, the non-detection error ratio for restricting human objects whose rectangle had covered the head and legs was 2%, and the false-detection error ratio was 50%.

7. CONCLUSION

A flexible pattern matching method for detecting human objects and reducing erroneous detection was proposed and evaluated by experiments.

The experimental results for normal environmental conditions were quite satisfactory : the false-detection error was 30% for cases in which the human detection accuracy was 100%. However, the results in poor environmental conditions were not as good : the false-detection error was 50% for cases in which the non-detection error was 20%. The major cause of the non-detection error was failure to extract silhouettes. Our method does not use specific pattern information of projection changes in successive video frames during motion. Therefore, this system is not acute enough to extract imperfect silhouettes and to process spurious images such as the reflection of lamps in puddles. Improvement in accuracy by the introduction of time-dependent techniques is the next step in making a truly robust human detection system.

References

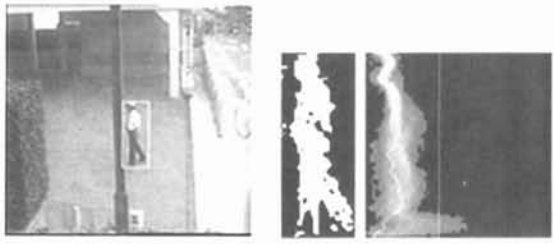
[1] M.Kaneta et al.,"Image Processing Method for Intruder Detection around Power Line Towers" Proceedings of MVA92 IAPR Workshop on Machine Vision Applications pp.353-356

[2] Y.Togashi et al.,"An Image Processing Platform with Adaptable Functions" Proceedings of the 24th Joint Conference on Imaging Technology, 1993, pp.289-292 (In Japanese)

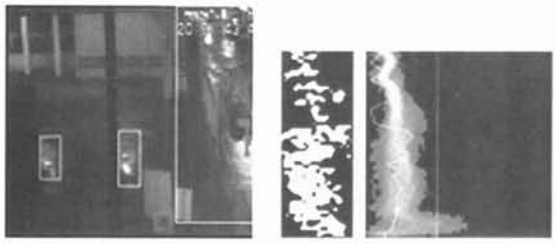
Table1 Conditions & Results of Experiment - II

CONDITIONS	
-DATE:	8 days in August
-TIME:	8:00PM - 7:00AM / 8:00AM
-LENS:	Autoiris
-Size of Image:	360 x 240
-Threshold level for binalization of the image:	10 fixed (density level 256)
-Size of rectangle surrounding for suspicious object:	20 - 120 (H), 20 - 80 (W)

RESULTS	
EXTRACTION Results	
-Total number of suspicious areas:	417
due to Humans:	110 ..(A)
(the head and legs successfully extracted:	53) ..(B)
due to non-Humans:	307 ..(C)
HUMAN DETECTION Results (T= 0.04)	
-Non-detection error ratio for (A):	20%
-Non-detection error ratio for (B):	2%
-False-detection error ratio for (C):	50%



(a) an Example of Human Object (score 0.126)



(b) an Example of Non-Human object (score 0.074)

Fig. 9 Examples of Detected Objects

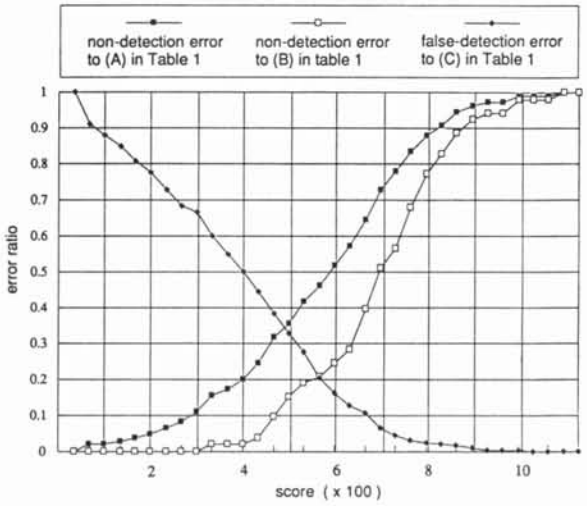


Fig. 10 Error ratio at Experiment - II