PAPER

High-Speed Human Motion Recognition Based on a Motion History Image and an Eigenspace

Takehito OGATA†, Joo Kooi TAN†, *Nonmembers*, *and* Seiji ISHIKAWA†, *Member*

SUMMARY This paper proposes an efficient technique for human motion recognition based on motion history images and an eigenspace technique. In recent years, human motion recognition has become one of the most popular research fields. It is expected to be applied in a security system, man-machine communication, and so on. In the proposed technique, we use two feature images and the eigenspace technique to realize high-speed recognition. An experiment was performed on recognizing six human motions and the results showed satisfactory performance of the technique.
key words: *motion recognition, high-speed recognition, motion history images, eigenspace*

1. Introduction

When we communicate with a person, visual information such as hand or body gestures often helps understanding his/her intension. In the same way, human motion recognition by a computer vision system will give us a new type of man-machine communication. A human motion recognition technique has been attracting much attention in the computer vision field, and it is expected in surveillance, man-machine communication, and other various fields. However, what is important in developing a real-time recognition system is that the system should recognize successive motions. Although a large number of studies have been performed on vision-based human motion recognition, only a few attempts have focused on recognition of successive motions.

There are some ways of classifying prior works related to human motion recognition [1], [2]. For instance, they can be classified into a model-based method (which includes both 2-D and 3-D methods), and an appearance-based method. The fundamental strategy of model-based method achieves motion recognition by using estimated or recovered human posture. Hogg [3] recovered pedestrian's posture from a monocular camera, by using a cylinder model. Wren *et al.* [4] estimated human pose by using a color based body parts tracking technique. Recovering human pose is an effective approach for motion recognition, since the human motion is related to its posture. However, model-based method generally needs a large amount of computation cost for pose estimation, or it needs some assumptions such as human skin color is already known in order to reduce computation cost.

Manuscript received January 11, 2005.
Manuscript revised July 20, 2005.
†The authors are with the Department of Mechanical and Control Engineering, Kyushu Institute of Technology, Kitakyushu-shi, 804–8550 Japan.
DOI: 10.1093/ietisy/e89–d.1.281

On the other hand, an appearance based method attempts to recognize human motion without pose estimation. In this approach, motion recognition belongs to a kind of pattern recognition problem. A *Hidden Markov Model* (HMM) is one of the learning based recognition techniques, and Yamato *et al.* [5] are the first researchers who applied it for motion recognition. However, one of the difficulties in the use of the HMM is that it needs to construct the HMM by manual work, and performance of the motion recognition largely depends on the constructed HMM.

An eigenspace technique [6] is also one of the learning based recognition techniques and it is also used in the motion recognition field [7]. A motion given by successive video image frames is expressed as a curve (called a motion curve) in an eigenspace, and, by adopting a similarity measure, it can be used in judging if an unknown motion is similar to any of the memorized motion curves. However, there remains an unsolved issue: how to realize an efficient judgment of the unknown motion employing motion curves.

Another well-known 2-D based method is the method using integration of past images, or integrated images, such as *a motion history image* [8] and *a recursive filtering* [9]. These approaches focus on the fact that a motion is expressed by some image sequences. However, these images include not only the movement of a body itself but also movement of the position of the person in the image. These movements should be separated because the former information is more important than the latter information in many situations.

In this paper, we propose an appearance based high-speed motion recognition technique based on superposed images and an eigenspace technique. We create superposed images by using only movement of a body. In order to exclude movement of position in the image, we generate superposed images from extracted human posture images. We employ two kinds of superposed images to represent a human motion; a motion history image (MHI) and a superposed motion image (SMI). Employing these images, a human motion is described in an eigenspace as a set of points, and each SMI plays a role of reference point. An unknown motion image is transformed into the MHI and then a match is found with images described in the eigenspace to realize high-speed motion recognition. Experimental results using six motions are shown and discussion is given on the performance of the proposed technique.

2. Overview of the Technique

2.1 The Developed System

The developed system has following distinctive features compared to existent motion recognition systems.

- **High-speed recognition:** The developed system is an online system, and it works every 50 ms.
- **Use of differential images for creating integrated images:** Unlike the existent techniques [9] that employ gray value images for creating integrated images, the proposed technique employs differential images of original gray value images. This is because differential images receive less effect on a dress problem of a target person than using gray value images in the sense of color and weak texture elimination. (Here a dress problem is the difficulties in recognizing human postures or motions, caused by texture, shape and color of the dress a person wears.) Moreover, differential images can keep past motions efficiently when integrating them into a motion history image, for example, since these images are composed of only lines and exclude unnecessary part of the images. On the other hand, silhouette images may also be useful in escaping from the dress problem. There are, however, motions that cannot easily be recognized only from the contour of a silhouette image. Therefore, differential images are more suitable for creating an integrated image than human silhouette images.
- **Employment of human posture images:** The technique employs a human posture image, i.e., human shape itself, so as to realize precise motion description and recognition. The employment of human shape may be more promising than using simpler features such as velocity information on human motions [10], when we think of performing the motion recognition not only by a static camera but also by a mobile camera.
- **Recognition of successive motions:** Unlike existent systems that recognize separate motions, the proposed system can perform high-speed recognition of successive motion changes.

Figure 1 shows configuration of the entire system.

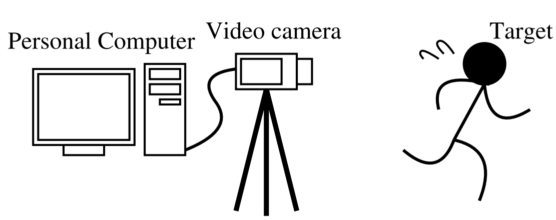


Fig. 1 System configuration.

2.2 A Motion History Image and a Superposed Motion Image

A motion history image (MHI) is a kind of temporal template. It is the weighted sum of past successive images and the weights decay as time lapses. Therefore, a MHI contains past images in itself, and the latest image is brighter than past ones. A MHI includes the direction of image motion because past images vanish little by little. Normally, a MHI is defined by the following equation [8].

$$H_\tau(x, y, k) = \begin{cases} \tau & \text{if } D(x, y, k) = 1 \\ \max(0, H_\tau(x, y, k-1) - 1) & \text{otherwise.} \end{cases} \quad (1)$$

where $D(x, y, k)$ is a binary image obtained from subtraction frames, and τ is a duration. However, we use in this paper a multi-valued differential image to extract information about a human posture because differential images are more suitable for creating an integrated image than binary image as was explained in 2.1. Differential image includes human posture information more than a binary image such as a silhouette image. Thus, we define modified MHIs corresponding to a multi-valued image as follows:

$$H_\alpha(x, y, k) = \max(f_i(x, y, k), \alpha H_\alpha(x, y, k-1)) \quad (2)$$

where $f_i(x, y, k)$ is an input image (a multi-valued differential image), H_α is a modified MHI, and parameter α is a vanishing rate which is set at $0 < \alpha < 1$.

A superposed motion image (SMI), on the other hand, is the maximum value image generated from summing past successive images with an equal weight. A SMI $S(x, y, k)$ is generated as follows:

$$S(x, y, k) = \max(f_i(x, y, k), S(x, y, k-1)) \quad (3)$$

Figure 2 shows examples of MHIs and a SMI.

MHIs contain not only a current image but also some past images. These are more suitable to use in motion recognition than original differential images. One of the reasons is that human motion consists of several human postures. Original images have information of only one posture, however MHIs have some postures in itself. The other reason is that MHIs are less sensitive to some noises than original images. Figure 2 (a) and (b) show examples of original differential images and MHIs, respectively. As one can see, back parts of the person's contour are missing for 4 successive frames from the 10th frame in Fig. 2 (a). Actually the values of these pixels are zero at the back parts. However same pixels of the same frames in MHIs are not zero as shown in (b), because past information is included in them. In the same way, MHIs are less sensitive to the small position deviation than original images. Figure 2 (c) is an example of a SMI.

Motion recognition by using MHIs is not suitable for high-speed recognition, however, because several MHIs are needed to express one motion. Therefore, in order to realize high-speed recognition, we use SMIs as reference points.

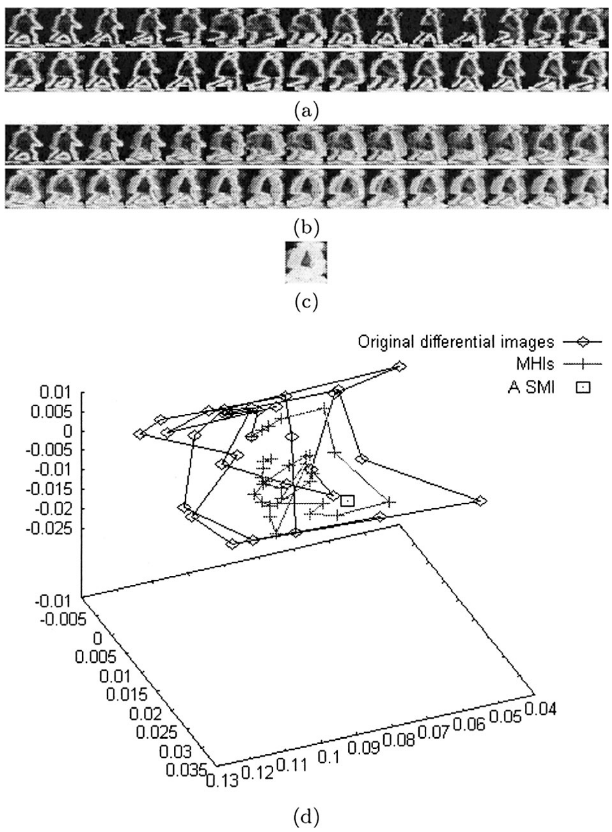


Fig. 2 An example of an eigenspace that is created from normalized differential images, and MHIs and a SMI: (a) Normalized differential images; (b) MHIs; (c) the SMI, and (d) the obtained eigenspace drawn with the first three eigenvectors. In (a) and (b), the time flows from the upper left frame to the lower right frame.

A SMI created by a single motion sequence shows the features of the motion well by only a single image. Moreover, if MHIs and a SMI are generated by a single motion sequence, these images have strong correlation between them, because Eqs. (2) and (3) show that the MHIs are included in the SMI. SMIs are not suitable to generate from unknown motion sequences that contain more than 2 motions, since SMIs keep past images forever. In contrast, MHIs are suitable to generate from unknown sequences, because these images discard past information little by little.

Hence, we use SMIs for reference images of respective motions and we generate MHIs for recognizing from unknown motion sequence. The recognition is performed by calculating the correlation between reference SMIs and the MHI generated from an unknown motion sequence.

2.3 The Eigenspace Technique

The eigenspace technique is one of linear dimensionality reduction techniques, and it has already been employed in object recognition [6], human face recognition [11], and so on. We use this technique for recognizing human motions.

Karhunen-Loeve transformation is employed to create an eigenspace. First, we create a data matrix X from a set of normalized learning image data $\boldsymbol{x}_n$ $(n = 1, \cdots, N)$.

$$X = \left[\ \boldsymbol{x}_1 - \boldsymbol{c} \quad \boldsymbol{x}_2 - \boldsymbol{c} \quad \cdots \quad \boldsymbol{x}_N - \boldsymbol{c}\ \right] \tag{4}$$

Here $\boldsymbol{x}_n$ satisfies $|\boldsymbol{x}_n| = 1$ $(n = 1, \cdots, N)$, and its dimension is P. The constant $\boldsymbol{c}$ is a mean image vector of $\boldsymbol{x}_n$ $(n = 1, \cdots, N)$. Then, $P \times P$ covariance matrix Q is defined as follows:

$$Q = XX^T \tag{5}$$

The eigenvalues and the corresponding eigenvectors of covariance matrix Q are obtained by solving the following eigen equation:

$$Q\boldsymbol{x} = \lambda \boldsymbol{x} \tag{6}$$

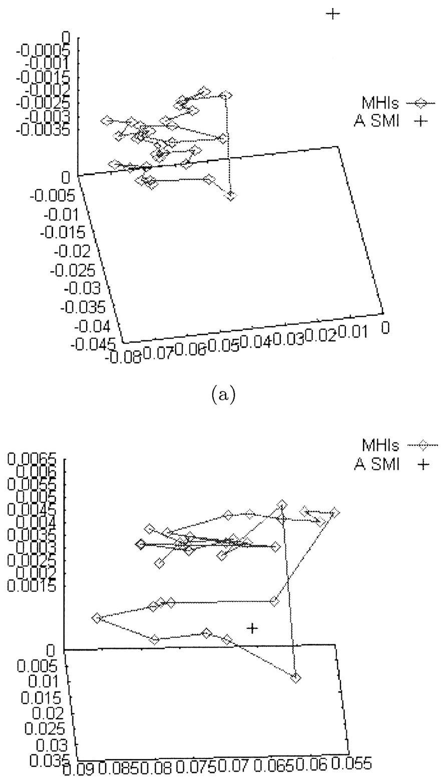
An eigenspace is created from k eigenvectors chosen corresponding to the largest k eigenvalues out of P. In the k-dimensional eigenspace, a learning image vector $\boldsymbol{x}_n$ is expressed as a point $\boldsymbol{g}_n$. If an unknown image vector $\boldsymbol{x}$ is projected near the point $\boldsymbol{g}_n$, we recognize that the unknown image $\boldsymbol{x}$ and a learning image $\boldsymbol{x}_n$ have strong correlation between them.

3. Strategy of Recognition

The entire procedure of the technique consists of two parts. One is a learning part, and the other is a recognizing part. The learning part performs generating MHIs and SMIs from learning image data, creating an eigenspace, and calculating reference points. The recognition part generates MHIs from unknown motion images in a periodic way, projects them into the eigenspace, and recognizes the motion.

The basic idea of this approach is to calculate the distances between reference SMIs and a MHI generated from an unknown motion image, in the eigenspace. Reference SMIs are generated from the learning data. We use only SMIs for reference points. On the other hand, we use MHIs in the recognition part. In the eigenspace, images that have strong correlation with each other are projected on mutually close points. MHIs contain not only one image but also some past images. Thus, MHIs expressing a single motion have mutually stronger correlation than the correlation among original differential images. Therefore, in an eigenspace, the region of MHIs becomes smaller than that of original differential images. For example, Fig. 2 (d) shows the first three components of the eigenspace that is created from a human walking motion employing original normalized differential images, MHIs, and a SMI shown in Fig. 2 (a), (b), (c), respectively. In Fig. 2 (d), a red line shows the sequence of original normalized differential images, a green line shows the sequence of MHIs, and a blue point is the SMI. As is seen in Fig. 2 (d), the green line is projected into narrower region than the red line, and the green line is closer to the point of SMI than the red line.

In the learning part, we use both MHIs and SMIs for creating an eigenspace, although we generate only MHIs in the recognition part. This is necessary to make the distances



(a)

(b)

Fig. 3 Points of the MHIs and the SMI projected into the first three components of different eigenspaces: eigenspace (a) is created from the MHIs only (Fig. 2 (b)); and eigenspace (b) is created from the MHIs and the SMI (Fig. 2 (b) and (c)).

between MHIs and a SMI of the same motion small in the eigenspace. Examples are shown in Fig. 3 (a) and (b). These two figures show the points of MHIs and a SMI in the first three components of eigenspaces. The eigenspace (a) is created from the MHIs only, and (b) is created from both the MHIs and the SMI. Here, the MHIs and the SMI in Fig. 2 (b) and (c) are used. A red line shows the sequence of the MHIs, and a blue point shows the SMI. Although the shapes of the red lines are similar to each other, the point of the SMI in (a) is projected into farther point from the red line. This is because the eigenspace technique is one of linear dimensionality reduction techniques. In the eigenspace (a), all MHIs can be presented by eigenvectors with a combination of linear functions. The MHIs and the SMI are generated from original differential images by using the maximum function, therefore the SMI can be generated from the MHIs by using the maximum function. However, since the maximum function is one of non-linear functions, it is impossible to present the SMI from the MHIs by using linear functions, and it is also impossible to present the SMI from the eigenvectors of the eigenspace (a) by using linear functions. Therefore, if we want to calculate the distances among MHIs and SMIs in the eigenspace, we need to create the eigenspace by employing both of them.

Thus, in the eigenspace, MHIs and a SMI, expressing a same motion, are projected in a narrow region around the SMI. Therefore, if a MHI obtained by an unknown motion is projected near the point of SMI of motion m, we can judge the unknown motion as motion m.

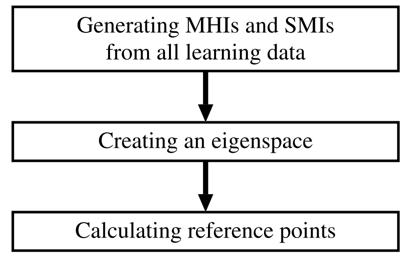


Fig. 4 Flow of the learning part.

3.1 Flow of the Learning Part

Figure 4 shows the flow of the learning part. To create the eigenspace, the MHIs and SMIs are generated by all learning motion data. For example, if an image sequence expressed by G image frames shows motion m, G MHIs and a single SMI are generated. The eigenspace is created by Eqs. (4)–(6) employing all MHIs and SMIs corresponding to the interested motions.

An average image generated by averaging all MHIs is similar to the SMI of the motion. It can therefore be used as a reference point in the eigenspace instead of the SMI. However, we don't use it because it has a problem in generating MHIs from a cyclic motion such as walking. The problem is that the position of the generated average image of MHIs in the eigenspace depends on the cycle and the frequency of original motion images. On the other hand, a SMI doesn't depend on the cycle and the frequency, because it is generated by the max-value computation as shown in Eq. (3). Therefore, if we obtain a motion image sequence longer than one cycle of the motion, we can always generate a unique SMI. It means that a SMI is stably created from learning data, and it is therefore advantageous for automatic generation of learning database in future.

Actually reference points in the eigenspace are given by the average point of SMIs that are generated by an identical motion of various persons.

3.2 Flow of the Recognition Part

Flow of the recognition part is shown in Fig. 5. Background subtraction is used for extracting a mask image containing a human region. In the proposed system, we perform background subtraction by using a differential image, because the MHI is created by differential images. In the differential image, the border between the human region and the background region always has high values: Actually it is an edge. Thus, the human region extraction using differential images performs better than using non-processed images such as color images or gray images. Let us denote an original differential image by $g(x, y)$, a background differential

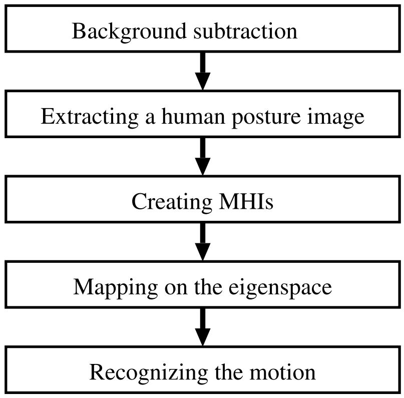


Fig. 5 Flow of the recognition part.

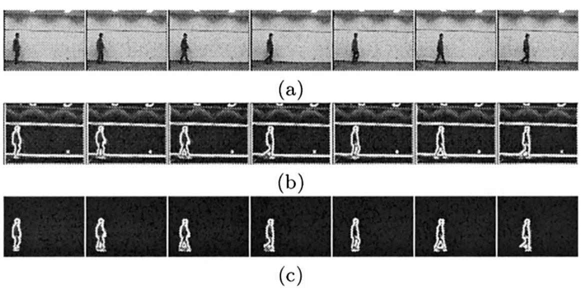


Fig. 6 Example of background subtraction by using differential images: (a) Original images; (b) differential images; and (c) background subtracted images.

image by $g_b(x, y)$, and a mask image by $g_m(x, y)$. The equation is then as follows:

$$g_m(x, y) = \begin{cases} 0 & \text{if} \quad g(x, y) - g_b(x, y) \leq T_b \\ 1 & \text{otherwise.} \end{cases} \tag{7}$$

Here T_b is a threshold value. The point of this method is to use equation $g(x, y) - g_b(x, y)$ for subtraction, unlike the absolute value $|g(x, y) - g_b(x, y)|$. If we use the absolute value, it will extract not only those edges between a human region and the background but also the background edges hidden by the human region. Figure 6 shows the examples of background subtraction by using differential images.

Binary image processing, such as expanding, contracting and labeling are performed to the mask image for eliminating noises and the largest blob is selected as a person's mask image. Then only a human differential image is extracted from the logical AND operation between the differential image and the mask image.

Size normalization is performed on the extracted human posture image for reducing influence on figure size of the target person. The latest MHIs are created by the normalized image and past MHIs, and magnitude normalization is performed on it. Generated images are projected on the eigenspace that is prepared in advance by learning data. Finally, motion recognition is done employing the distance with the reference points.

In the developed system, V MHIs are created at each sample time in order to adapt to the small change of a motion speed and to realize robust recognition. Each MHI has its own vanishing rate $\alpha_1, \alpha_2, \cdots, \alpha_V$. All MHIs are projected into the eigenspace, and the distance between the projected points and the reference points are calculated. Let us denote the reference point of motion m by $\boldsymbol{p}_m$ $(m = 1, \cdots, M)$, and the MHI with vanishing rate α_v by $\boldsymbol{g}_v$ $(v = 1, \cdots, V)$. The distance of the MHI with vanishing rate α_v is then calculated by

$$d_{vm} = \sqrt{(\boldsymbol{p}_m - \boldsymbol{g}_v)^T (\boldsymbol{p}_m - \boldsymbol{g}_v)} \tag{8}$$

and the result of recognition r_v is obtained by the following equation;

$$r_v = \begin{cases} m & \text{if} \quad \min_{1 \leq m \leq M} d_{vm} < D_m \\ \text{unknown} & \text{otherwise.} \end{cases} \tag{9}$$

where D_m is the distance threshold of motion m. Finally, the result of motion recognition $r(k)$ at sample time k is determined by the majority decision of the result r_v $(v = 1, \cdots, V)$. If the number of r_v which outputs motion m is defined as $\phi(m)$, it is expressed by

$$r(k) = m_{\max} \quad \text{where} \quad \phi(m_{\max}) = \max_{1 \leq m \leq M} \phi(m) \tag{10}$$

Above equation means that the most frequent r_v is reported as the final decision at sample time k.

4. Experiment

4.1 Experimental Environment

The RT-Linux is employed as an operating system of the used computer to achieve stable real-time performance of the overall system. In the developed system, the sampling rate is set at 50 milliseconds by the used 2.4 GHz computer. Usually, RT-Linux cannot capture images using Video for Linux by a real-time program directly. Therefore, we perform only image capture by a non real-time periodic program that sends an image to a shared memory space.

An experiment on motion recognition is done to investigate the effectiveness of the proposed system. The photograph of the experimental environment is shown in Fig. 7. For simplicity, we assume a uniform background in order to extract a human region with less difficulty. We use the recorded video sequences of 9 persons' 6 motions, each motion being repeated 3 times. The employed motions are walking, running, crouching, picking up, careful walking and side walking. Figure 8 shows examples of each motion. In order to calculate the recognition rate, we use the leave out method. We separate all data into 3 classes, each class containing 3 persons' data. One class is used as learning data for creating an eigenspace, determining a distance threshold D_m, and the others are used as test data.

In the experiment, we have to define how to judge the recognition result, since our system outputs the recognition result with each frame. Thus, we use the duration of a motion cycle to decide the recognition result. If the same result

Fig. 7 The experimental environment.

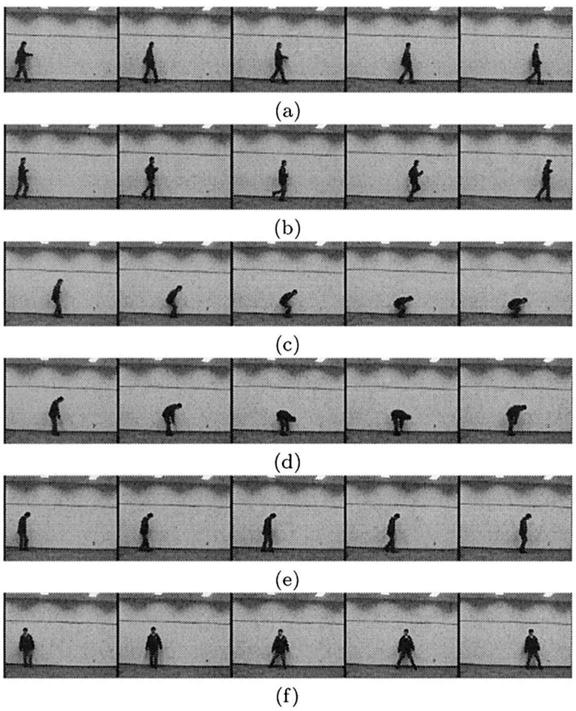


Fig. 8 Examples of each motion; (a) walking; (b) running; (c) crouching; (d) picking up; (e) careful-walking; (f) side-walking.

'motion m' continues during a time threshold T_m, the system judges that the result is 'motion m'. Here, the time threshold T_m of motion m is obtained by the duration of the motion cycle in the learning data.

4.2 Experimental Results

We performed the experiment twice by changing the number of eigenspace dimension. The numbers of dimension we used were 20 and 40. We used Sobel operator for generating a multi-valued differential image, and the resolution of the image employed for an eigenspace creation was set to 25 horizontal pixels by 25 vertical pixels. Note that the employment of this image with reduced resolution contributes to escaping from dress and body shape change to a large extent. The number of MHIs created at every sample time is 4, and their vanishing rates are set at 0.98, 0.95, 0.9, and 0.85, respectively. These rates were decided experimentally.

The average recognition rate was about 76% by the 20-dimensional eigenspace, whereas it was 80% by the 40-dimensional eigenspace. The details of the recognition result are shown in Table 1. In the experiment, not only single motions but also two or three successive motions were recognized successfully. Examples of the recognition results are shown in Fig. 9. In Fig. 9, (a) shows the developed application window including the original image, the processed image, motion history images, and the recognition result by a color panel: (b) and (c) report the recognition results. In (b), the unknown motion is reported as 'walking' by the indication of a 'red' panel. On the other hand, in (c), the motion is reported as 'walking before crouching' by the 'red and blue' panel. Computation time from feeding in a motion frame to reporting the recognition result was 50 milliseconds by the used 2.4 GHz computer. Thus, the system achieved high-speed motion recognition.

Table 1 Recognition results.

No. of dimensions	20	40
Walking (%)	88.9	92.6
Running (%)	64.8	68.5
Crouching (%)	74.1	77.7
Picking up (%)	77.8	83.3
Careful-walking (%)	79.6	83.3
Side-walking (%)	72.2	74.1
Recognition rate (%)	76.2	79.9

Figure 10 shows more complex result. We attempted to recognize successive motion of 'walking, picking up, and then crouching.' In the figure, (a) shows the original image sequence (it shows some selected frames), and (b) shows the result of recognition by the graph in which the x-coordinate is the sampling time, and the y-coordinate indicates the recognized motion. In (b), we can see that the system recognized the three motions correctly. In this way, the developed system showed satisfactory results.

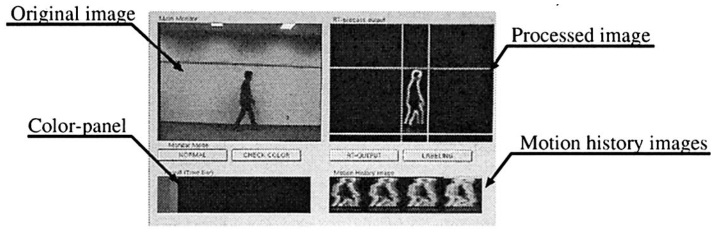
4.3 Experimental Results in a Mobile Environment

In the second experiment, we attempted this recognition technique in mobile camera situation. We have applied this technique to an aerial robot system. The aerial robot has 4 rotors to move in the air, and a single CCD camera to observe the ground. Figure 11 shows the configuration of the system.

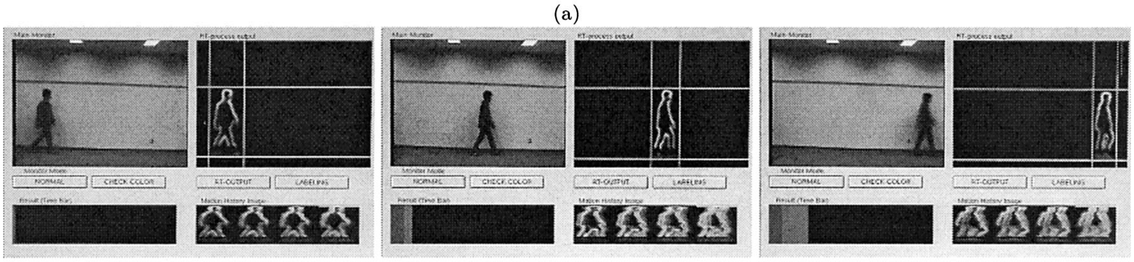
The aerial robot controlls itself to find a person on the ground and to recognize his/her motion. In order to extract the person on the ground from the air, we use color of the clothes the person wears. The system binarizes an original image employing the color of the clothes, and the binarized images are used as the mask images described in 3.2.

In the learning stage, we use sequences of a calling motion and a standing motion. These sequences are acquired in advance. The calling motion is the motion swinging both arms, and the standing motion is just standing on the ground. Therefore the robot distinguishes calling from standing of the person.

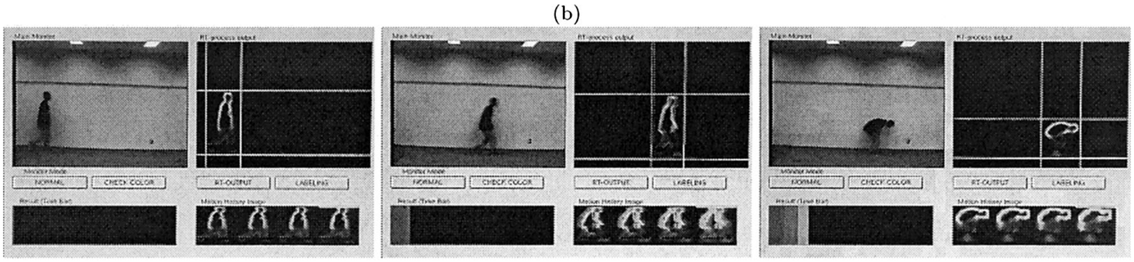
The experiment was performed in a high-ceiled room. Some photos of the experiment are shown in Fig. 12.



(a)

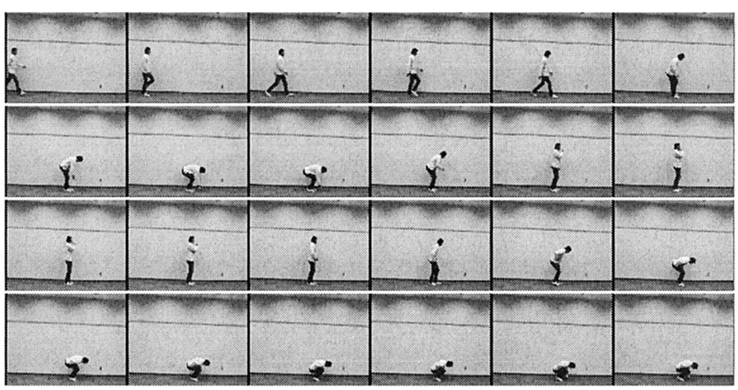


(b)

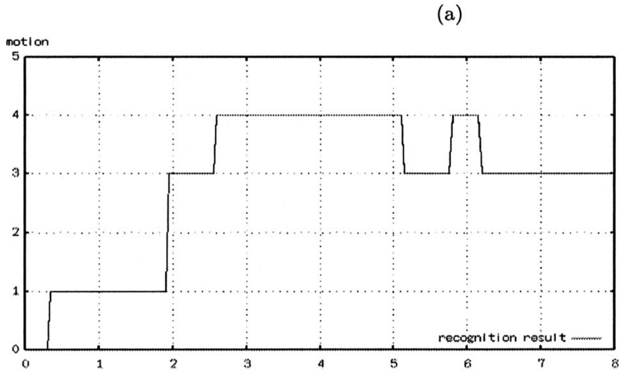


(c)

Fig. 9 Experimental results: (a) The developed application window; (b) walking; and (c) walking followed by crouching.



(a)



0	Unknown
1	Walking
2	Running
3	Crouching
4	Picking up
5	Careful-walking

(b)

Fig. 10 Experiment on sequential motions: (a) Some original image sequences, and (b) recognition result.

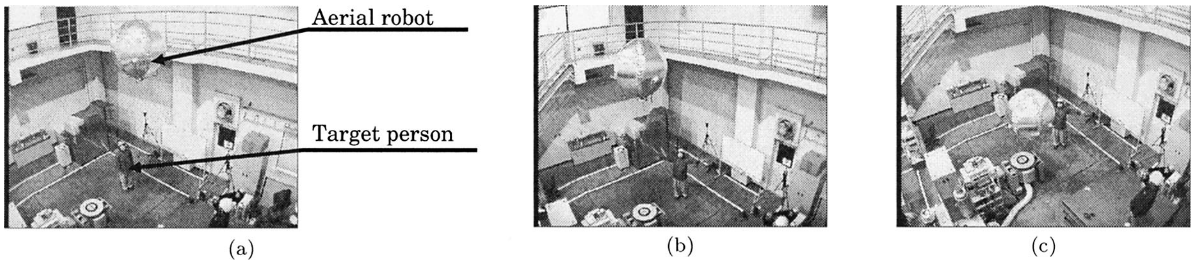


Fig. 12 Experimental results on the developed aerial robot system: Time flows from (a) to (c).

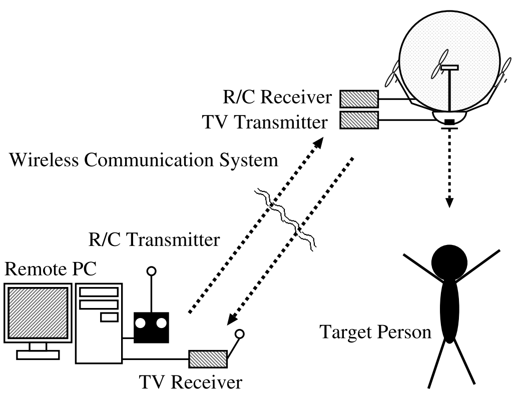


Fig. 11 Configuration of an aerial robot system.

The developed aerial robot started from its initial position to try to find a person wearing red-colored clothes (See Fig. 12 (a)). Once it has found such a person, it approached the person and hovered above him (Fig. 12 (b)). Then the aerial robot recognized the person's calling motion and came down near him (Fig. 12 (c)).

Although this experiment is still on a simpler stage, the experimental result shows that the proposed technique can be used in mobile camera situation. Details of the results are presented in [12].

5. Discussion and Conclusions

A technique was proposed for recognizing human motions by a computer vision system employing superposed images and an eigenspace. A motion was represented by motion history images and a superposed motion image, and they were used for creating an eigenspace. Six motions were represented by the reference points in the eigenspace and recognition of unknown motions was performed using them. The average recognition rate was about 80%. The recognition was done in 50 [ms] (20 [fps]) at each sample time. The proposed system therefore achieved high-speed human motion recognition. It should be noted that the system can recognize not only a single motion but also successive motions.

Although we have achieved high-speed human motion recognition, we also confirm some limitations of the proposed technique from the experimental results. One of these is that the proposed technique is difficult to recognize the motion whose speed changes frequently or suddenly. Although the proposed technique uses several MHIs to adapt to the motion speed varying, it may not adapt to big speed change. This is because our proposed technique performs simple recognition method, calculating correlations between known SMIs and the MHI, in order to reduce computation costs.

The angle that the optical axis of the camera and the direction of the person make is also one of the limitations of the proposed system. The angle was 90° in the performed experiment. If we attempt similar experiments with different angles, the recognition rate will change because our method is appearance-based method. However, since our system is learning-based method, the system can adapt to other angles if the system has learning data. In fact, the system worked well in aerial robot situation, in which the angle was 0°.

One of the main advantages of the proposed system is that we perform the recognition by using extracted and normalized human posture images. In other words, we only use relative posture change to recognize a human motion. We may add that the change of human postures appearance can be described in the eigenspace as well [13]. All of these mean that if we extract a human region stably, the system can recognize the motion even using mobile cameras.

One of possible applications of the presented system is that a corridor surveillance system in elementary schools or hospitals. For instance, since our system can recognize 'running' and 'crouching', our system is able to detect running people for giving them warning or is able to inform a medical treatment room for giving crouching people help. If we use a multi-camera system and connect them with networks, the system can cover the whole building. The proposed technique will be applied to a surveillance system tracking a person who behaves in a suspicious manner, or to an intelligent security system finding an elderly person who is, for example, carrying a heavy bag or sitting down on the road as he/she feels bad.

On the other hand, the current system still has some difficulties. One of them is that the system contains many parameters that cannot be tuned automatically. In the proposed system, most of the parameters were decided experimentally. We have to reduce these experimentally tuned parameters to realize a fully automatic human motion recognition system.

Another difficulty is how to judge the recognition result. In the experiment, we define that, if the result 'motion *m*' continues during a certain time interval, then the system judges the overall action as 'motion *m*'. However this may sometimes go wrong if the system makes misjudgment even once during the local recognition. Therefore, further consideration needs to be given to this issue.

If the above difficulties are improved, the system will show better performance.

Acknowledgements

This study was partly supported by JSPS KAKENHI (17700201), which is greatly acknowledged. The authors wish to thank the students who helped them in acting the motions used in the experiment.

References

[1] J.K. Aggarwal and Q. Cai, "Human motion analysis: A review," Comput. Visi. Image Underst., vol.73, no.3, pp.428–440, March 1999.
[2] D.M. Gavrila, "The visual analysis of human movement: A survey," Comput. Vis. Image Underst., vol.73, no.1, pp.82–98, Jan. 1999.
[3] D. Hogg, "Model-based vision: A program to see a walking person," Image Vis. Comput., vol.1, no.1, pp.5–20, Feb. 1983.
[4] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-time tracking of the human body," IEEE Trans. Pattern Anal. Mach. Intell., vol.19, no.7, pp.780–785, July 1997.
[5] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using Hidden Markov Model," Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp.379–385, June 1992.
[6] H. Murase and S.K. Nayar, "Visual learning and recognition of 3-D objects from appearance," Int. J. Comput. Vis., vol.14, no.1, pp.5–24, Jan. 1995.
[7] H. Murase and R. Sakai, "Moving object recognition in eigenspace representation: Gait analysis and lip reading," Pattern Recognit. Lett., no.17, pp.155–162, Feb. 1996.
[8] A.F. Bobick and J.W. Davis, "The recognition of human movement using temporal templates," IEEE Trans. Pattern Anal. Mach. Intell., vol.23, no.3, pp.257–266, March 2001.
[9] O. Masoud and N.P. Papanikolopoulos, "A method for human action recognition," Image Vis. Comput., vol.21, no.8, pp.729–743, Aug. 2003.
[10] M. Thonnat and N. Rota, "Image understanding for visual surveillance applications," Proc. 3rd Int. Workshop on Cooperative Distributed Vision, pp.51–82, Nov. 1999.
[11] M. Turk and A. Pentland, "Eigenfaces for recognition," J. Cognitive Neurosci., vol.3, no.1, pp.71–86, March 1991.
[12] T. Ogata, S. Matsuda, J.K. Tan, and S. Ishikawa, "Real-time human motion recognition by an aerial robot," Proc. IEEE Int. Conf. on Systems, Man and Cybernetics, pp.5290–5295, Oct. 2004.
[13] S. Houman, M.M. Rahman, J.K. Tan, and S. Ishikawa, "Representing human motions in an eigenspace based on surrounding cameras," Proc. ICCAS, pp.1808–1813, Aug. 2004.

Takehito Ogata received B.E. and M.E. degrees in Mechanical and Control Engineering from Kyushu Institute of Technology, Kitakyushu, Japan, in 2003 and 2004, respectively. From April 2004, he is a Ph.D student at Kyushu Institute of Technology. His research interests include computer-human interaction, pattern recognition and computer vision. He received the IEEJ Kyushu Branch Excellent Student Award in 2000, the SICE Kyushu Branch Young Author's Award in 2004, and the Excellent Student Award of The IEEE Fukuoka Section for 2004. He is a student member of IEEE, and The Society of Instrument and Control Engineers.

Joo Kooi Tan received B.E. and M.E. degrees in Computer Science and Ph.D. in Control Engineering from Kyushu Institute of Technology, in 1995, 1997, and 2000, respectively. From 2000 to 2003, she was a postdoctoral researcher at Satellite Venture Business Laboratory of Kyushu Institute of Technology. She is presently Research Assistant with Department of Mechanical and Control Engineering in the same university. Her current research interests are 3D reconstruction, image processing, pattern recognition, and applications of computer vision. She received the SICE Kyushu Branch Young Author's Award in 1999, the AROB 10th Young Author's Award in 2004, and Young Author's Award from IPSJ of Kyushu Branch for 2004. Dr. Tan is a member of IEEE, The Society of Instrument and Control Engineers, and IPS of Japan.

Seiji Ishikawa obtained B.E., M.E., and Ph.D. from The University of Tokyo in 1974, 1976, and 1980, respectively, where he majored in Mathematical Engineering and Instrumentation Physics. In 1979, he joined Department of Computer Science in Kyushu Institute of Technology and he is currently Professor of Department of Control and Mechanical Engineering. Professor Ishikawa was a visiting research fellow at Sheffield University, U.K., from 1983 to 1984, and a visiting professor at Utrecht University, The Netherlands, in 1996. His research interests include three-dimensional shape/motion recovery, human motion analysis, medical image analysis, and a multiple robots system. He is a member of IEEE, The Society of Instrument and Control Engineers, and The Institute of Image Electronics Engineers of Japan.