Detection of Small-Waving Hand by Distributed Camera System

Kenji Terabayashi*, Hidetsugu Asano†, Takeshi Nagayasu‡, Tatsuya Orimo†, Mutsumi Ohta†, Takaaki Oiwa*, and Kazunori Umeda‡

* Department of Mechanical Engineering, Shizuoka University
3-5-1 Johoku, Naka-ku, Hamamatsu-shi, Shizuoka 432-8561, Japan
Email: tera@eng.shizuoka.ac.jp
† PIONEER CORPORATION
1-1 Shinkokura, Saiwai-ku, Kawasaki-shi, Kanagawa 212-0031, Japan
Email: hidetsugu_asano@post.pioneer.co.jp
‡ Department of Precision Mechanics, Chuo University
1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan
Email: umeda@mech.chuo-u.ac.jp

***Abstract*—This paper proposes a method to detect small-waving hand from image sequences. The method is based on frequency analysis of intensity value change for each pixel of low-resolution images converted from input images. Handwaving makes periodical change of intensity value of image pixels located in the moving region. The periodical change is detected by thresholding of a value accumulating frequency features obtained through FFT calculation. The proposed method can stably detect small-waving hand with 3-5 Hz in a range of 2-6 m. We demonstrate an application of spatial memory through measuring 3D position of small-waving hand detected by a distributed camera system.**

I. Introduction

Gesture recognition from image sequences is one of the most promising technologies in human computer/robot interaction. The first step of gesture recognition is to extract or detect motion of body part such as a hand from input image sequences. Then gestures are recognized based on shape and motion trajectory of the extracted body part [1].

There are two main approaches to extract body part for gesture recognition; (i) color-based extraction, (ii) motion-based extraction. In color-based extraction, color information about skin [2][3] or image background [4][5] is used for distinguishing between body part and background. This approach is not robust to change in illumination because the color information is usually preliminarily given. In motion-based extraction, motion part in image sequences is detected by optical flow [6] and periodicity detection [7][8][9]. This approach is more robust to illumination change than color-based extraction while restricting gesture types.

In a distributed camera system, cameras are different in position, orientation, and illumination condition. This difference makes an image object different in color, appearance, and size, between images captured from the cameras. Motion-based extraction is more appropriate to this case than color-based extraction. Detection of waving hands from images [9] is applied to Intelligent Room that consists of distributed cameras and provides gesture recognition to users for operating home appliances by hand gestures [10][11].

The Intelligent Room can detect waving hands robustly to illumination and appearance changes in input images captured by distributed cameras. However, it requires that width of waving hand is more than 30 cm for the robust detection. This is not easy for the elderly to use because of large movement. Spatial memory applications [12] also need to minimize the width of waving hand for improving resolution performance.

This paper proposes a method to detect small-waving hand from image sequences. In this method, degree of periodic motion is defined based on case-based knowledge. The integral of the degree of periodic motion over time is used for detection of small-waving hand.

This paper is organized as follows: Section II defines the degree of periodic motion based on case-based knowledge and proposes algorithm to detect small-waving hands from image sequences. Section III details experimental results for evaluating detection of waving hand with various conditions of motion speed, distance from a camera and illumination. Section IV shows a demonstration of spatial memory applications by using a distributed camera system with detection of small-waving hand. Section V presents conclusions.

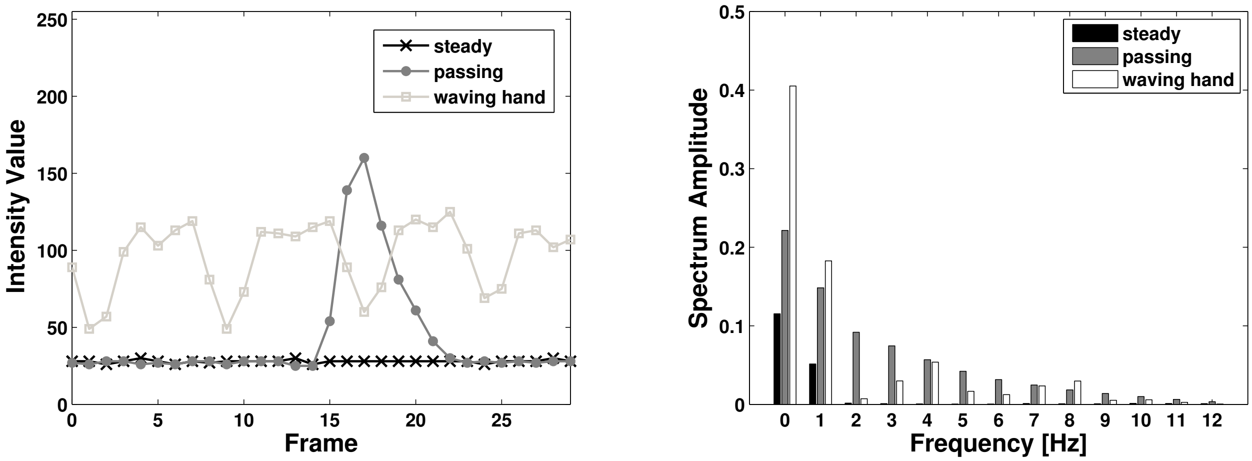
II. Detection of Small-Waving Hand by Degree of Periodic Motion

In this section, we proposes a method to detect small-waving hands from image sequences. This method is based on pixel-wise frequency analysis (II-A). Through case-based knowledge of periodic motion (II-B), degree of periodic motion is defined (II-C). Small-waving hands are detected by thresholding of the degree of periodic motion (II-D).

A. Frequency Analysis of Image Pixels

Amplitude spectrum of frequency analysis is calculated for intensity changes of input images as frequency features.

978-1-4673-1786-3/12/$31.00 ©2012 IEEE



(a) Intensity change $I_L(i, j, t)$.

(b) Spectrum amplitude $A(i, j, t, f)$.

Fig. 1. Comparisons of intensity change and spectrum amplitude among 'steady', 'passing', and 'waving hand' conditions. (a) Intensity change of a pixel on a low-resolution image converted from an input image (30 fps). (b) Spectrum amplitude corresponding to the intensity change shown in Fig.1(a).

Input image $I(t)$ captured from a distributed camera at current time t is converted to low-resolution image $I_L(t)$. This conversion is for smoothing change of intensity value $I_L(i, j, t)$ at a point with image coordinates (i, j) including gesture motion and is for reducing computational cost of frequency analysis and noise included in an input image $I(t)$.

At each pixel on low-resolution image $I_L(i, j, t)$ ($i = 1, 2, ..., W_L, j = 1, 2, ..., H_L$), amplitude spectrum is analyzed by FFT (Fast Fourier Transform) with recent N intensity values $I_L(i, j, t - N + 1)$ ... $I_L(i, j, t)$. W_L and H_L are the pixel numbers of width and height of the low-resolution image $I_L(t)$, N is the number of intensity samples for frequency analysis. Amplitude spectrum $A(i, j, t, f)$ is calculated by

$$A(i, j, t, f) = \left| \sum_{k=0}^{N-1} I_L(i, j, t + k) \mathbf{W}_N^{fk} \right| \quad (1)$$

where f is frequency, $\mathbf{W}_N$ is the twiddle factor of DFT (Discrete Fourier Transform).

B. Frequency Characteristics of Periodic Motion

Small-waving hand of periodic motion (labeled as 'waving hand') is compared with other two typical conditions (labeled as 'steady' and 'passing'). The details of these conditions are as follows: 'waving hand' is that waving hand periodically passes through a low-resolution image pixel (i, j). 'steady' is that variation of $I_L(i, j, t)$ is quite small such as background. 'passing' is that a moving object once passes through a low-resolution image pixel (i, j) only once.

Examples of time series variation of intensity values are shown in Fig.1(a), comparing among 'steady', 'passing', and 'waving hand' conditions. Handwaving makes periodical change of intensity value of image pixels located in the moving region while keeping intensity values almost flat in 'steady' and 'passing' conditions.

Fig.1(b) shows spectrum amplitude of frequency analysis corresponding to the intensity change shown in Fig.1(a). In 'steady' condition, low-frequency spectrums below 1 Hz are mainly observed. In 'passing' condition, with increasing frequency, spectrum amplitude decreases monotonically according to the hamming window for frequency analysis of finite samples. In 'waving hand' condition, in contrast to 'passing' condition, the spectrum amplitude at 2 Hz is quite small, and there is a peak spectrum amplitude at 4Hz in the frequency range more than 1 Hz. The frequency of this peak corresponds to actual frequency of waving hand.

C. Degree of Periodic Motion

In this paper, we define the degree of periodic motion $P(i, j, t)$ at a point with image coordinate (i, j) at current time t as

$$P(i, j, t) = \int_0^t \left[\max_{3 \leq f \leq 6} A(i, j, s, f) - A(i, j, s, 2) \right] ds. \quad (2)$$

The degree of periodic motion $P(i, j, t)$ integrates the frequency characteristics of periodic motion described in II-B over time and is used for detection of small-waving hands from image sequences.

D. Detection of Small-Waving Hand

Small-waving hands are detected in low-resolution image sequences $I_L(i, j, t)$ by thresholding of the degree of periodic motion $P(i, j, t)$ at each image pixel (i, j). For each image pixel (i, j), the detection algorithm is as follows:

1) Initialization of $P(i, j, t = 0)$ to be 0.
2) Update of $P(i, j, t)$ with the current $A(i, j, t, f)$ as

$$P(i, j, t) = P(i, j, t - 1) + \max_{3 \leq f \leq 6} A(i, j, t, f) - A(i, j, t, 2). \quad (3)$$

3) Judgment of small-waving hand by thresholding of $P(i, j, t)$.

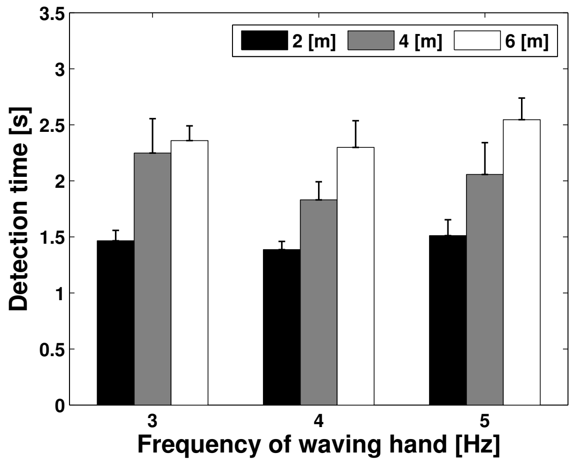


Fig. 2. Average and standard deviation of detection time when changing conditions of waving hand and distance from a camera.

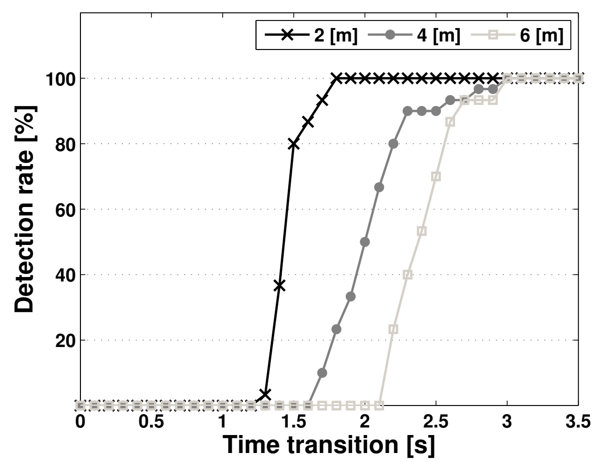


Fig. 3. Relationship between detection rate and detection time.

III. Evaluation of Detecting Small-Waving Hand

The proposed method of detecting small-waving hand is evaluated on various conditions of waving speed, illumination, and distance from a camera. The relationship between detection rate and detection time is investigated in the section III-A. The robustness to illumination change is examined in the section III-B.

A camera of Axis 233D is used in these experiments. The horizontal angle of view is 55.8 deg, and the vertical angle of view is 43.3 deg. The size of input image is 640×480. The frame rate is 30 fps.

A computer with Intel Core i7 975 processer, 6 GB memory, and GeForce GTX 285 of GPU (Graphics Processing Unit). The input image is converted to low-resolution image of 80×60. The detection of small-waving hand is processed at each pixel on low-resolution image, which is implemented in parallel using the GPU. The calculation time to process one image is about 5 ms.

A. Detection Rate and Detection Time

Detecting small-waving hand is evaluated on the relationship between detection rate and detection time. A subject makes small-waving hand which width is about 5 cm. The speed of waving hand is controlled with a metronome to be 3, 4, and 5 Hz. The distance of subject's hand from a camera is changed as 2, 4, and 6 m. For each condition, the number of trials is 10.

The average and standard deviation of detection time is shown in Fig.2. In this figure, the vertical axis represents detection time for the average and standard deviation of error bar, the horizontal axis is frequency of waving hand, and the color of each bar means different condition of distance from a camera to the hand. This figure shows that the detection time increases with increasing the distance while it is independent of the speed of waving hand.

The transition of detection rate of small-waving hand is shown in Fig.3. In this figure, each line means different distance from a camera to a waving hand. The detection rates of all distance conditions rises to 100 % over time, which is from the degree of periodic motion $P(i, j, t)$ accumulating frequency characteristics of periodic motion as defined in (2). Even a small movement of 5 cm width can be detected at 6 m far from a camera.

False detection occurred when a person in striped or checked clothes was walking in a captured scene. Movement of structured color pattern makes periodical change in intensity value of a image pixel of input image.

B. Robustness to Illumination Change

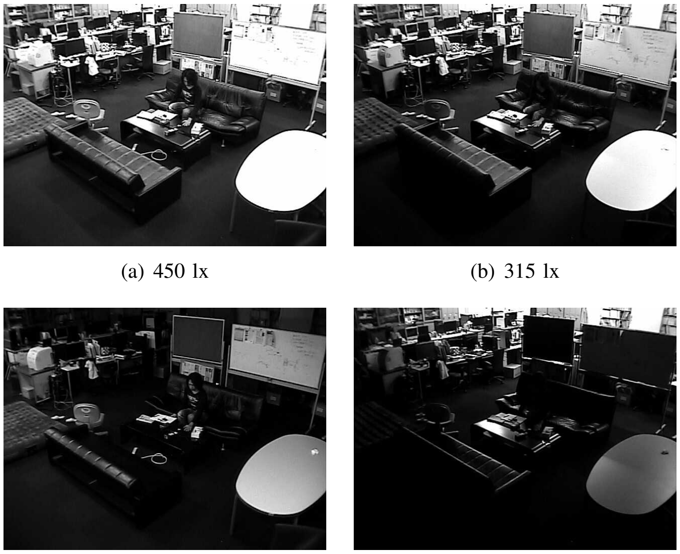
Detecting small-waving hand is evaluated on robustness to illumination change. Luminance at waving hand is changed as 450, 315, 160, 70, and 43 lx. A subject makes a small movement of a hand at 5 m far from a camera. Examples of input images are shown in Fig.4. At each illumination condition, the number of trials is 20.

All the waving hands is detected correctly, and the average of detection time with the standard deviation is shown in Fig.5. In environments with more than 160 lx, the average detection time keeps about 1.2 s. Even in an environment of 43 lx, small-waving hand can be detected in about 3 s.

IV. Spatial Memory Application of Detecting Small-Waving Hand

In this section, detecting small-waving hand is applied to a spatial memory system. This system has several commands to operate a TV such as switch-on, switch-off, and changing channels. These commands are assigned to 3D spatial regions and are activated by giving a trigger of waving a hand to an assigned region.

The system consists of four networked cameras of Axis 233D and a computer same as used in the previous section. These cameras are installed at the corners of a square with 6 m width on the ceiling of a room and are calibrated about the intrinsic and extrinsic parameters to measure 3D positions of detected waving hand.



(a) 450 lx (b) 315 lx

(c) 160 lx (d) 43 lx

Fig. 4. Input images with illumination change: 450, 315, 160, and 43 lx at the point of a waving hand.

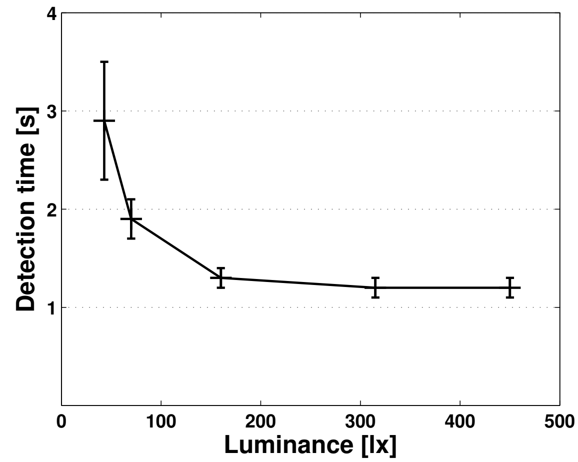


Fig. 5. Average and standard deviation of detection time when changing illumination conditions as shown in Fig.4.

There are 6 commands in this system: (1) Ch. 1, (2) Ch. 2, (3) Ch. 3, (4) Ch. 4, (5) Ch. 5, (6) TV on/off. Each command is assigned to a cube with 30 cm width, and 6 cubes are arranged in 2 rows and 3 columns on a table. Examples of input images from distributed cameras are shown in Fig.6. In the figure, 6 white papers on a table are areas of the assigned commands. The surrounding people makes small-waving hand, including finger tapping shown in Fig.7, in the command-assigned regions to operate a TV.

The system detects small-waving hand in image sequences from distributed cameras and measures the 3D position by stereo vision to select a command assigned to the handwaving location. If the measured 3D position is in a cube defined before, the system outputs the assigned command.

The spatial memory system is evaluated on relationship between input commands by handwaving and detected commands by this system. Each command was tested 4 times by a subject. The number of subjects is 5. Each command assigned to a certain 3D region has 20 trials in total.

The confusion matrix of actual and detected locations is visualized in Fig.8. In this figure, black cells mean high co-occurrence frequency of actual location in the column and detected location in the row. The rate of correct detections was 94.2 %. The reason of false detection is the shift of measured 3D position by detection of waving arms involved with handwaving.

V. Conclusion

In this paper, we proposes a method to detect small-waving hand in image sequences captured from a distributed camera system. The degree of periodic motion is defined as accumulation of frequency features over time and is effective to detect small-waving hand even in far and dark environments. The proposed method can also detect a finger tapping with 5 cm gap, which is 6 times smaller than conventional way. Detection of small-waving hand is applied to a spatial memory system through measuring the 3D position by distributed cameras.

References

[1] V. I. Pavlovic, R. Sharma, and T. S. Huang, “Visual interpretation of hand gestures for human-computer interaction: a review,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, pp. 677–695, 1997.

[2] S. M. Dominguez, T. Keaton, and A. H. Sayed, “A Robust Finger Tracking Method for Multimodal Wearable Computer Interfacing,” IEEE Transactions on Multimedia, vol. 8, pp. 956–972, 2006.

[3] Y. Sato, M. Saito, and H. Koike, “Real-Time Input of 3D Pose and Gestures of a Userfs Hand and Its Applications for HCI,” Proc. of IEEE Virtual Reality Conference, pp. 79–86, 2001.

[4] L. Hong and M. Greenspan, “Multi-scale gesture recognition from time-varying contours,” Proc. of Tenth IEEE International Conference on Computer Vision (ICCV2005), vol. 1, pp. 236–243, 2005.

[5] J. Davis and A. Bobick, “The representation and recognition of human movement using temporal templates,” Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 928, 1997.

[6] M. Bayazit, A. Couture-Beil, and G. Mori, “Real-time Motion-based Gesture Recognition using the GPU,” IAPR Conference on Machine Vision Applications (MVA2009), pp. 9–12, 2009.

[7] M. Takahashi, K. Irie, K. Terabayashi, and K. Umeda, “Gesture recognition based on the detection of periodic motion,” International Symposium on Optomechatronic Technologies (ISOT2010), pp. 1–6, 2010.

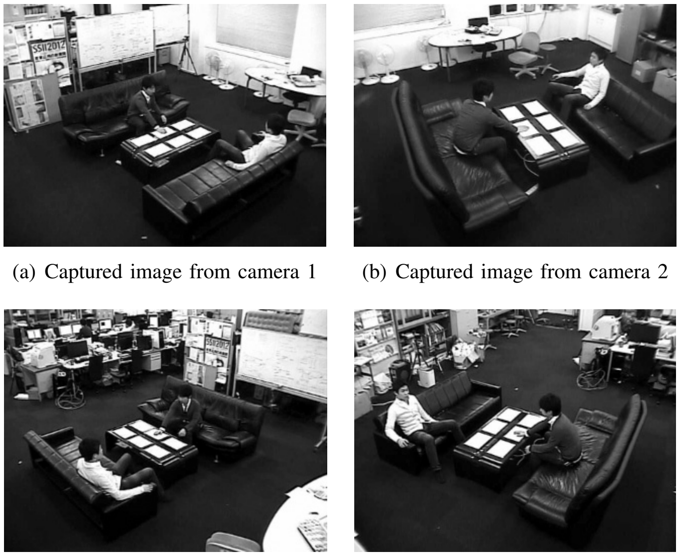
[8] A. Briassouli, and N. Ahuja, “Estimation of Multiple Periodic Motions from Video,” Proc. of Seventh European Conference on Computer Vision (ECCV2006), pp. 147–159, 2006.

[9] K. Irie and K. Umeda, “Detection of Waving Hands from Images - Application of FFT to time series of intensity values,” Proc. of Third China- Japan Symposium on Mechatronics, pp. 79–83, 2002.

[10] K. Irie, N. Wakamura, and K. Umeda, “Construction of an intelligent room based on gesture recognition: operation of electric appliances with hand gestures,” Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004), vol. 1, pp. 193–198, 2004.

[11] K. Irie, M. Wada, and K. Umeda, “3D Measurement by Distributed Camera System for Constructing an Intelligent Room,” Proc. of Fourth International Conference on Networked Sensing Systems (INSS2007), pp. 118–121, 2007.

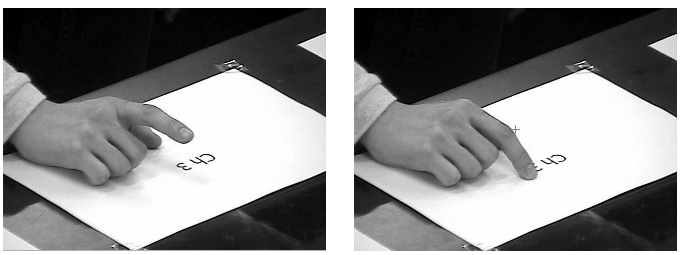
[12] M. Niitsuma and H. Hashimoto, “Spatial Memory as an Aid System for Human Activity in Intelligent Space,” IEEE Transactions on Industrial Electronics, vol. 54, pp. 1122–1131, 2007.



(a) Captured image from camera 1 (b) Captured image from camera 2

(c) Captured image from camera 3 (d) Captured image from camera 4

Fig. 6. Input images captured from distributed cameras.



(a) Initial phase (b) Opposite phase

Fig. 7. Zoomed images of a finger tapping as small-waving hand.

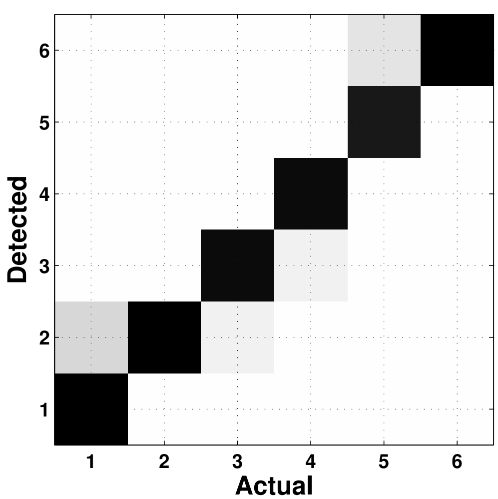


Fig. 8. Visualized confusion matrix of actual and detected locations of waving hand.