IJBSCHS
[Original article]

Biomedical Soft Computing and Human Sciences, Vol.24, No.1, pp.1-7
Copyright©1995 Biomedical Fuzzy Systems Association
(Accepted in April 2019)

A Method of Describing a Self-Occlusive Motion – A Reverse Motion History Image

Joo Kooi TAN[1], Sayaka OKAE[2], Youtaro YAMASHITA[2] and Yuta ONO[2]

[1]*Faculty of Engineering, Kyushu Institute of Technology*
[2]*Graduate School of Engineering, Kyushu Institute of Technology*

Abstract: *This paper proposes a new method of describing a self-occlusive human motion, particularly in the depth direction, which has been considered little in the motion/action recognition studies to date in spite of its importance in our daily life. A Motion History Image (MHI) is a well-known method of describing a motion by a single gray value image, but it suffers from a self-occlusion problem in which present motion overwrites past motion. To solve this difficulty, a Reverse description MHI (RMHI) is proposed in the paper. RMHI and the original MHI are both employed for motion representation in the proposed method; the former for approach motion, whereas the latter for leave motion. In the experiment on motion recognition, motions are described by RMHI or MHI according to motion direction, transformed then to Hu moment vectors, and finally recognized employing the k-nearest neighbor. Experimental results show effectiveness of the RMHI description.*
Keywords: *MHI, RMHI, Motion recognition, Action recognition, Hu moments, k-nearest neighbor, Heuristic approach*

1. Introduction

Automatic human motion (or action) recognition by a camera and computer system has become an increasingly important technology in our society, since working population has been decreasing. In near future, various human jobs will be replaced by or conducted jointly with such a system. Application fields may include robot teaching in a factory, surveillance of abnormal behavior for preventing crimes, finding a person with sudden sickness for prompt help, watching elderly people at home or in a nursing home to give hands to them when necessary, and others.

In order to realize automatic human motion recognition, motion should be described properly in the first place by a single image, for example. Among various motion description techniques, Motion History Image (MHI) [1] is a well-known method. It layers foreground images in an image sequence of a motion so that past images become darker, yielding a single gray value image representing the motion. One of its drawbacks is, however, that it cannot describe a motion with self-occlusion which occurs when a recent motion overlaps a former motion such as walking first to the right and then suddenly to the left or crouching followed by standing. There are some variations of the MHI such as motion history histograms [2], hierarchical MHIs [3], a kernel-based method [4], multiple key MHIs [5], pseudo-color MHI [6], Histogram of Oriented Gradients (HOG) of MHI [7]. However, they do not discuss the self-occlusion problem. The idea of directional MHIs [8,9] and the variations of the DMHI, such as DMHI with an MEI (Motion Energy Image) histogram [10] and histogram of DMHI and LBP (Local Binary Pattern) [11], solves this difficulty by separating a motion into four directions (right, left, up, down) using optical flow. However, it does not deal with the motion toward depth.

The motion toward depth is another self-occlusion problem, which occurs in a single direction. Leaving from or approaching to an observer is an important motion. In particular, an approaching person might have some interest in the observer. However, these motions are still excluded in most of motion description and recognition studies. All the related studies take account of the motions to the right/left direction such as walking,

Sensuicho 1-1, Tobata, Kitakyushu 804-8550, Japan
e-mail: etheltan@cntl.kyutech.ac.jp

running, or the motions facing an observer or a camera like radio exercises. It is because those figures have the largest information with the motion interested. This fact can be found even in a standard human action dataset such as KTH dataset [12]. The motion toward depth direction has less information on motion on account of its self-occlusive nature.

To solve the difficulty of recognizing a motion toward depth, 3-D MHI description was proposed [13,14]. However, the computation time of those methods is large for practical use. It may be more advantageous to describe a motion in a 2-D way than in a 3-D way.

This paper proposes a method of describing a self-occlusive human motion particularly in a depth direction by a modified MHI. The method solves the self-occlusion problem by use of a Reverse description MHI (abbr., RMHI). It makes an MHI in a reverse way according to the motion direction. Proposal of the RMHI discriminates the present paper from others. By use of the RMHI, one can expect that a broader class of human motions is within the scope of automatic recognition. This may contribute to practical use of the human motion recognition technology.

Definition of the RMHI is given in Section 2. The strategy for motion recognition is explained in Section 3. Experimental results are shown in Section 4. The results and some issues on the RMHI are discussed in Section 5. Finally the paper is concluded in Section 6.

2. Reverse Description MHI

The definition of an MHI [1] is given in Equation (1).

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

Here $H_\tau(x,y,t)$ is the MHI at frame t (t =1,2,…,T), and $D(x,y,t)$ is the binarized image frame at t, D = 1,0 meaning the foreground and the background region on the image, respectively. Parameter τ is a positive integer defining the brightness of the most recent foreground region, whereas δ is the amount by which the former MHI reduces the brightness.

Overwrite can arise in making an MHI by the first condition of Equation (1), since it gives priority to the foreground region of the present image frame.

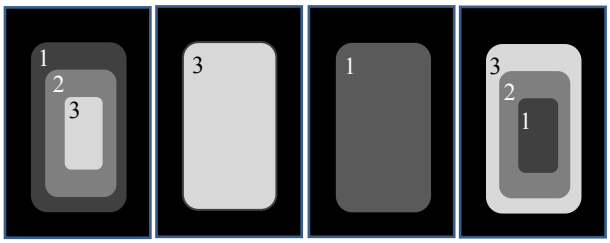
The proposed description of a motion, RMHI, leaves past foreground regions preferentially, although the highest brightness is given to the present foreground. It is defined by

$$H_\tau(x,y,t)=\begin{cases}\tau & \text{if } D(x,y,t)=1\cap H_\tau(x,y,t-1)=0\\ \max\{0,\ H_\tau(x,y,t-1)-\delta\} & \text{otherwise}\end{cases} \quad (2)$$

The first condition of Equation (2) means that the area satisfying $D(x,y,t)=1$ and $H_\tau(x,y,t-1)>0$ is not overwritten by the latest foreground.

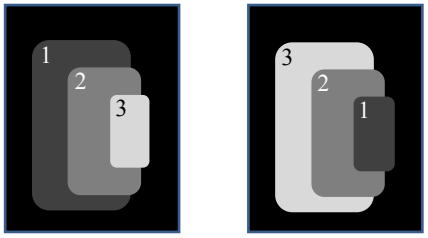
Let us take two motions of a rectangle in the 3-D space, leaving from and approaching to an observer. The difference between an MHI and an RMHI given by Equation (1) and Equation (2), respectively, can be seen in Figure 1. In Figure 1, (a) shows MHIs and (b) are RMHIs, whereas (1) is 'leaving' and (2) is 'approaching'. Three foregrounds (rectangles at 3 sample times) are layered: 1 is the oldest and 3 is the latest. As is seen in (a2), the information on past motion disappears on approach w.r.t. the MHI, whereas it is kept on leave in the form of a gray value image as shown in (a1). On the contrary, the proposed RMHI shows the past images on approach as in (b2), but it shows only the oldest image on leave as in (b1), containing no motion information.

Thus the recognition strategy needs to employ MHI for leaving and RMHI for approaching motion. Motion vectors can be employed for the judgment on leave or approach.



(a1) (a2) (b1) (b2)

Figure 1. Examples of MHIs and RMHIs: (a) MHIs, (b) RMHIs; (1) leaving, (2) approaching



(a) (b)

Figure 2. MHI and RMHI w.r.t. a motion in a diagonal direction: (a) MHI, leaving, (b) RMHI, approaching

It is noted that motion description by MHI or RMHI is indifferent to visual impression of those images. In Figure 2, two motions in diagonal directions are shown respectively by (a) an MHI (leave) and (b) an RMHI (approach). Brighter regions may be seen as closer to an observer in (a), but they are actually farther. On the other hand, darker regions seem closer, but they are more distant.

3. Motion Recognition

The procedure for human motion recognition consists of three main stages, (1) acquisition of a data set, (2) training of a classifier and (3) test of the classifier. In order to train a chosen classifier, a data set S (the set of pairs of a feature vector and its label) is prepared in the first place. It is then divided into a training set S^{tr} and a test set S^{te}. A chosen classifier is trained employing the set S^{tr} and then tested its performance by the set S^{te} to report the recognition rate. Flow of the procedure is illustrated in Figure 3, in which the stage of data set acquisition is shown in detail as the main interest of this paper.

Stage (1) contains 4 main steps, i.e., (1_1) human region extraction from an image frame, (1_2) motion vectors computation for finding an FoE (Focus of Extension) and judgment of motion direction, (1_3) motion description by MHI or RMHI, and (1_4) transform of the description into a feature vector and store the vector and the label into a set S.

Each step is explained in the following.

(1_1) A human region is extracted from each image frame in a given video employing Gaussian mixture model.

(1_2) Points are arranged with equidistance on the contour of the extracted human region in the initial frame. Motion vectors are calculated by finding the points' correspondence on the next frame using LK tracker [16]. Outliers in the vectors are removed by RANSAC [13] under the assumption of projective transform between frames. An FoE is computed by extending the motion vectors. If the extension is beyond the head of the motion vector, the human motion is 'leave', whereas, if it is beyond the tail, the motion is 'approach'. This can be judged by comparing the directions of the motion vector and the direction of FoE. Weighted vote [14] is employed for the computation of an FoE to raise its precision.

(1_3) MHIs or RMHIs are made employing the sequence of extracted human regions by changing the first frame in the video, as the initial frame of a motion is somewhat ambiguous.

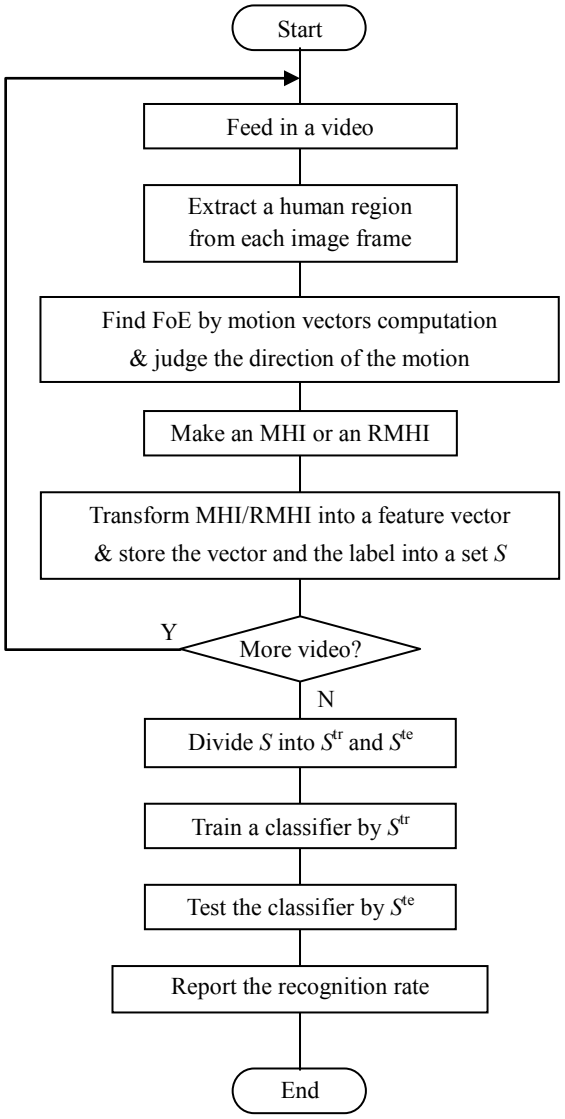


Figure 3. Flow of the procedure

(1_4) The obtained MHIs or RMHIs are then described by 7 Hu moments [17] yielding 7-dimensional feature vectors and stored into a set S along with their labels. The reason why Hu moment expression is employed is that it is invariant to translation, rotation and scale w.r.t. the shape concerned.

In stage (2), the set S is divided into two subsets, S^{tr} for training and S^{te} for test. A classifier is chosen and it is trained employing S^{tr}. The k-nearest neighbor method, which is effective to ill-separated multiclass classification, is chosen as a classifier in the performed experiment. All the vectors in S^{tr} are plotted in a feature space defined by the vector dimension.

Finally, in stage (3), all the feature vectors in S^{te} are

judged their classes employing the *k*-nearest neighbor in the feature space and the recognition rate is reported.

Actually, stages (2) and (3) have a loop structure, since leave-one-out cross validation is performed in the experiment. It is described in a simple way in Figure 3, though.

4. Experimental Results

An experiment on human motion recognition is conducted employing the proposed method. As the motions toward/away-from an observer's view are not available in public motion databases, our own motion videos are employed for the experiment. Six motions performed indoors are chosen. They are WtO (Walk to an Observer), WaO (Walk away from an Observer), FtO (Fall toward an Observer), FaO (Fall away from an Observer), WtL (Walk to the Left) and FtR (Fall to the Right). WtO and FtO are classified as approach motions, whereas WaO and FaO are classified as leave motions. WtL and FtR are the motions toward the left/right normally considered in public database such as [6]. They are included in the experiment for reference. Video sequences of these motions are shown in Figure 4.

Five subjects (male students) act each motion twice. An MHI or an RMHI is made according to if a motion is leave or approach, respectively. This means that an MHI is made for WaO and FaO, whereas an RMHI is made for WtO and FtO. An RMHI is also made for WtL and FtR. Each MHI/RMHI is made using τ successive frames ($\tau = 10, 20, 30, 40$). From a single motion video, 15 MHIs or 15 RMHIs are produced by changing the first image frame, since the start frame of a motion is not very certain. All of these yield 900 Hu moment vectors with each value of τ. Leave-one-out cross validation and the 3-nearest neighbor method are employed for the recognition of the 900 data. (δ is set at 1 in Equations (1) and (2).)

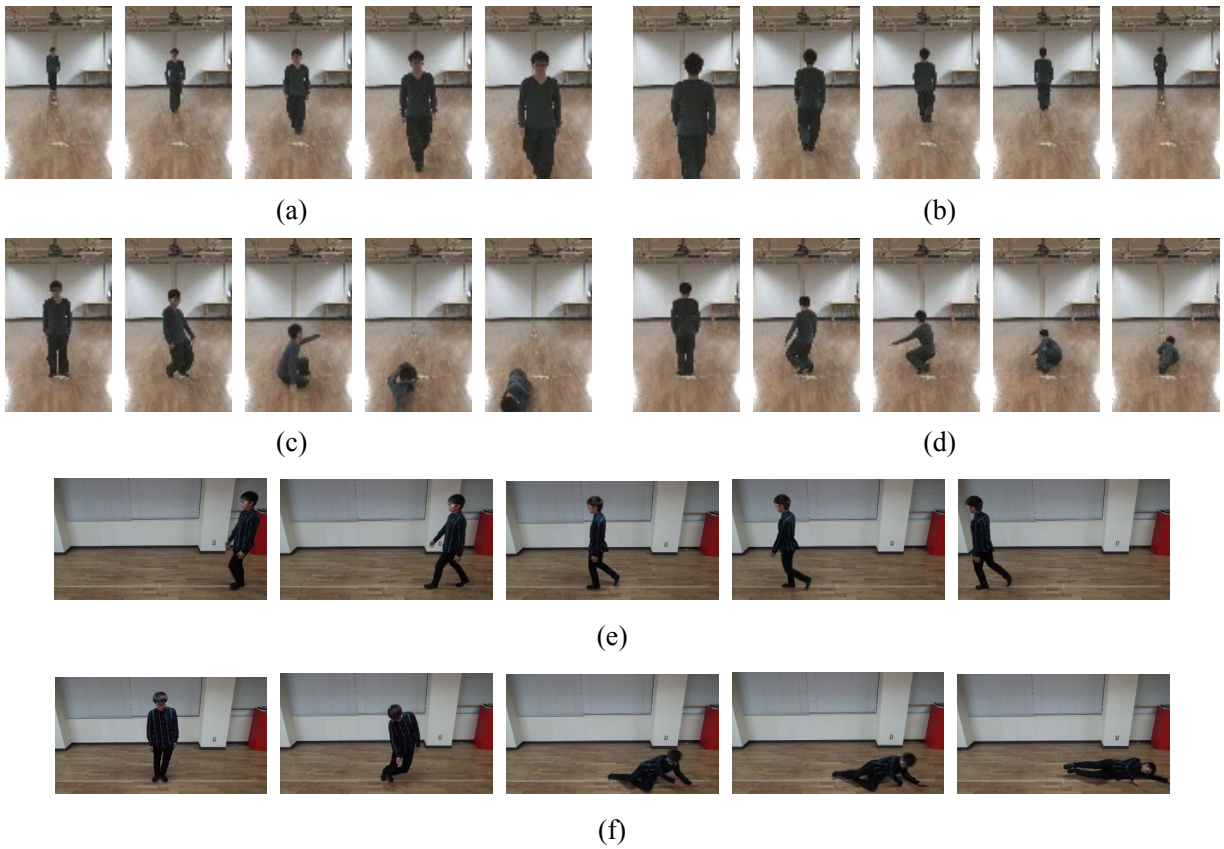


Figure 4. Videos of the employed motions: (a) WtO, (b) WaO, (c) FtO, (d) FaO, (e) WtL, (f) FtR. Time lapse is from the left to the right.

In more detail, two kinds of experiment are performed taking account of how the developed system will be employed practically.

(i) Private use: Leave-one data-out cross validation: 899 data are employed for training and one data for test. This is repeated 900 times. The ratio of the number of

correctly recognized data to 900 is reported as the recognition rate.

(ii) General use: Leave-one set-out cross validation: Four subjects' data sets (containing 720 data) are employed for training and one subject's data set (containing 180 data) for test. It gives a recognition rate of the performance. The procedure is repeated five times and the resultant recognition rates are averaged to yield an overall recognition rate.

Results are shown in Table 1. (i) is the result with private use of the system, whereas (ii) is the result with general use. The recognition rate is almost increased as τ is larger. Case (i) yields better recognition results than case (ii), as expected. Examples of MHIs/RMHIs w.r.t. employed motions are illustrated in Figure 5.

For comparison, two more experiments have been done. One is the experiment where the motions are all described by MHIs, and the other is the experiment in which all the motions are represented by RMHIs. The employed data are the same 900 data when τ is set at 40 as those employed in the main experiment (employing MHIs and RMHIs). The results are given in Table 2. Table 2 indicates the effectiveness of the proposed method.

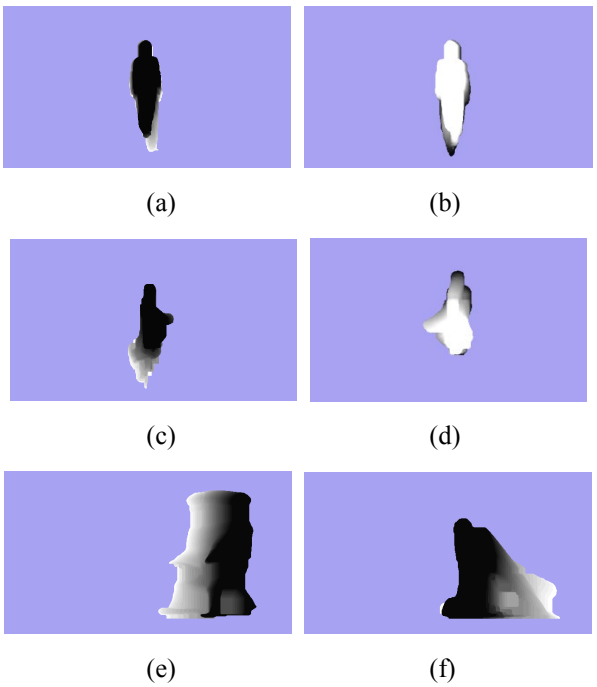


Figure 5. Examples of MHI/RMHI representation employed in the experiment ($\tau = 40$): (a) WtO (RMHI), (b) WaO (MHI), (c) FtO (RMHI), (d) FaO (MHI), (e) WtL (RMHI), (f) FtR (RMHI). Some images are resized for better observation.

Table 1. Recognition rates (%)

τ	10	20	30	40	Average
(i)	88.1	90.2	93.7	94.3	91.6
(ii)	78.2	72.2	83.4	83.7	79.4

Table 2. Comparison of recognition rates (%) among three kinds of motion description when $\tau = 40$

	MHI	RMHI	MHI & RMHI
(i)	80.6	92.3	94.3
(ii)	60.3	79.2	83.7

5. Discussion

The proposed RMHI description was claimed its effectiveness experimentally. It well describes an approach motion to an observer by leaving past motions in the description in a visible way. On the other hand, the original MHI can as well be employed in the description of the motion leaving from an observer. They both contribute to solving the self-occlusion problem in the 2-D motion description methods by a single image.

Let us consider a pedestrian. He walks to this side, but unfortunately he falls on the ground. Then he stands up and continues walking to this side. The RMHI well describes these series of motions. On the other hand, the MHI can describe the reverse motion of the same scenario, i.e., walking away from this side. However, obviously, the approaching motion/person is much more important to an observer than a leaving motion/person, as the approaching person might have some interest in the observer. This supports the importance of the proposed RMHI description.

As for the motion to the left/right, or the motion at a spot, both description methods may equally be employed. However, RMHI description may be employed if an interested action contains a self-occluded approach motion such as raising the left hand to an observer to express hello when walking to the left.

Parameter τ for making an MHI/RMHI depends obviously on the length of a motion or how quick it is performed by an observed person. In the experiment, it was fixed to 10 to 40 and 40 achieved the best result. In a practical situation, however, a motion video may include a person's continuous action which may contain different partial motions of different time duration. In order to

recognize respective motions to understand the whole action, varied values of τ should be employed to a single motion in a parallel way. The results of the partial motion recognition may be arranged to describe the entire action.

The motion recognition strategy proposed in this paper employs both MHI and RMHI. It can be referred to as a hybrid recognition strategy: The both ways of motion representation are necessary for recognizing the motion toward depth direction. However, as mentioned before, an approach motion is more important than a leave motion in the sense that a person approaching to him/her might have an intention to communicate to him/her on something. Or, simply, the observer must be careful to prevent a head-on collision with the approaching person. Therefore, the paper focuses its attention to RMHI.

Human motion recognition can alternatively be done using recent Neural Networks (NN) based on the deep learning. In this study, however, conventional heuristic approach, i.e., explicit motion description and feature definition, is adopted to clarify an automatic recognition process. Although the NN might yield better results in human motion recognition, we believe that it is scientifically important to understand and explain the process of the recognition. It may lead to the understanding of our own brain functions.

6. Conclusion

This paper proposed a method of describing a self-occlusive motion, in particular human approach motion, by an RMHI. The paper proposed a human motion recognition strategy employing the RMHI and the original MHI, and showed its effectiveness experimentally.

Collection of more number of training images needs to be done to raise the recognition rate higher. Considering different kinds of motions is necessary for examining the effectiveness of the proposed method. A strategy for recognizing a series of motions is also to be studied to put the method into practical use.

The proposed method may expand automatically recognizable human motions by offering a way of describing self-occlusive motions. This will lead to practical human daily activity recognition by a robot which will be strongly requested for nursing aged people in near future.

Acknowledgment

This study was supported by JSPS Kakenhi Grant No. 16K01554.

References

[1] Bobick, A. and Davis, J.: The recognition of human movement using temporal templates, IEEE Trans. PAMI, Vol.23, No.3, pp.257-267, 2001.

[2] Meng, H., Pears, N., Freeman, M. and Bailey, C.: Motion history histograms for human action recognition, Embedded Computer Vision, Springer, 2009.

[3] Davis, J. W.: Hierarchical motion history images for recognizing human motion, IEEE Workshop on Detection and Recognition of Events in Video, pp.39-46, 2001.

[4] Oikonomopoulos, A., Patras, I. and Pantic, M.: Kernal-based recognition of human actions using spatiotemporal salient points, Proc. of 2006 Conf. on CVPR Workshop, 2006.

[5] Li, D., Yu, L., He, J., Sun, B. and Ge, F.: Action recognition based on multiple key motion history images, Proc. of 2016 IEEE 13th International Conference on Signal Processing, pp.993-996, 2016.

[6] Chun, Q. and Zhang, E.: Human action recognition based on improved motion history image and deep convolution neural networks, Proc. of the 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, pp.1-5, 2017.

[7] Huang, C. P., Hsieh C. H., Lai, K. T. and Huang, W. Y.: Human motion recognition using histogram of oriented of motion history image, Proc. of 2011 International Conference on Instrumentation, Measurement, Computer, Communication and Control, pp.353-356, 2011.

[8] Ogata, T., Tan, J. K. and Ishikawa, S.: Human motion recognition based on directional motion history images, Proc. of the International Workshop on Advanced Image Technology, pp.857-862, 2007.

[9] Fukumoto, M., Ogata, T., Tan, J. K., Kim, H. and Ishikawa, S.: Human motions representation and recognition by directional motion history images, Artificial Life and Robotics, Vol.13, No.1, pp.326-330, 2008.

[10] Ahsan, S. M. M., Tan, J. K., Kim, H. and Ishikawa,

S.: Human action representation and recognition: An approach to a histogram of spatiotemporal templates, International Journal of Innovative Computing, Information and Control, Vol.11, No.6, pp.1855-1867, 2015.

[11] Ahsan, S. M. M., Tan, J. K., Kim, H. and Ishikawa, S.: Spatiotemporal LBP and shape feature for human activity representation and recognition, International Journal of Innovative Computing, Information and Control, Vol.12, No.1, pp.1-13, 2016.

[12] KTH Dataset: http://www.nada.kth.se/cvap/actions/, 2005.

[13] Weinland, D., Ronfard, R. and Boyer, Y.: Motion history volumes for free viewpoint action recognition, IEEE International Workshop on Modeling People and Human Interaction, 2005.

[14] Yamashita, Y., Tan, J. K. and Ishikawa, S.: Human motion description and recognition under arbitrary motion direction, Proc. of SICE Annual Conf., pp.110-115, 2017.

[15] Lucas, B. D. and Kanade, T.: An iterative image registration technique with an application to stereo vision, Proc. of International Joint Conference on Artificial Intelligence, pp.674-679, 1981.

[16] Fischer, M. A. and Bolles, R. C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM, pp.381-395, 1981.

[17] Hu, M.: Visual pattern recognition by moment invariants, IRE Transactions on Information Theory, Vol.8, No.2, pp.179-187, 1962.

Joo Kooi TAN

She obtained the Ph.D. from Kyushu Institute of Technology. She is presently with Department of Mechanical and Control Engineering in the same university as Associate Professor. Her current research interests include three-dimensional shape/motion recovery, human detection and its motion analysis from videos.

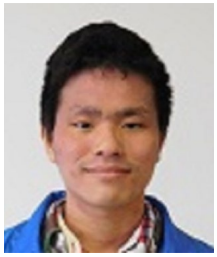
She was awarded SICE Kyushu Branch Young Author's Award in 1999, the AROB Young Author's Award in 2004, Young Author's Award from IPSJ of Kyushu Branch in 2004 and BMFSA Best Paper Award in 2008, 2010, 2013 and 2015. She is a member of IEEE, The Information Processing Society, The Institute of Image Electronics Engineers, and The Biomedical Fuzzy Systems Association of Japan.

Sayaka OKAE

She received B.E. from Kyushu Institute of Technology. Her research includes image processing, and human motion representation.

Youtaro YAMASHITA

He obtained B.E. and M.E. from Kyushu Institute of Technology. His research includes image processing, human motion representation and recognition.



Yuta ONO

He received M.E. from Kyushu Institute of Technology. He is now in the Ph.D. course of the Graduate School of Engineering, Kyushu Institute of Technology. His research includes computer vision, machine learning, and rush-out pedestrian detection from videos.