International Journal of Innovative
Computing, Information and Control
Volume **20**, Number **4**, August **2024**

ICIC International ©2024 ISSN 1349-4198
pp. **1093–1103**

HIGH-ACCURACY HUMAN MOTION RECOGNITION INDEPENDENT OF MOTION DIRECTION USING A SINGLE CAMERA

Jing Cao[1] and Yui Tanjo[2,*]

[1]Graduate School of Engineering
[2]Faculty of Engineering
Kyushu Institute of Technology
1-1, Sensui, Tobata, Kitakyushu, Fukuoka 804-8550, Japan
cao.jing644@mail.kyutech.jp; *Corresponding author: tanjo@cntl.kyutech.ac.jp

Received November 2023; revised March 2024

Abstract. *With the rapid development of computer vision research, the technology has received extensive attention and has been widely used. In today's society, the technology of automatic recognition of human movements is particularly important in research related to the development of elderly care systems and crime prevention systems. However, most conventional researches on motion recognition assume that the motion is on the plane perpendicular to the optical axis of the camera. There are only a few studies dealing with the motion in the direction of a camera's optical axis and the reported accuracy is only 83.7%. Therefore, to improve the accuracy while reducing processing time, this paper proposes a TMRIs (Triplet Motion Representation Images) that expresses human motion by extending the traditional method named MHI (Motion History Image), Ex-HOOF (Extended Histogram of Oriented Optical Flow) that includes features of motion direction and velocity information, and the change of features of the area and the center of gravity of the foreground object, to identify the movement toward the optical axis of a camera and hence improve the accuracy of action recognition. The performance and effectiveness of the proposed method are verified by experiments.*
Keywords: Motion recognition, Elderly care, Crime prevention, MHI, Optical flows

1. **Introduction.** In a super-aging society like Japan, elderly people living alone (over 65 years old) account for a large proportion of the elderly population. According to an investigation in 2020 [1], among the total elderly population, approximately 2.31 million men and 4.41 million women, respectively live alone. Social problems, such as solitary death which is concerned by more than 50% of the elderly living alone, are also on the rise. Therefore, there is a growing need for an elderly support/care system to prevent these problems from happening.

There are many care systems for the elderly on the market today. However, most of them require regular visits by specialists or judgments based on the use of home appliances. These approaches can be difficult to handle in emergencies.

On the other hand, according to the 2021 public opinion poll on public security [2], 54.5% of the respondents answered that "domestic public security has deteriorated over the last 10 years". While cameras in public places help people catch criminals and increase arrest rates for robberies, such systems cannot identify and prevent crimes by themselves in advance.

The basic technology for developing a crime prevention and elderly watching system is the automatic recognition of human motion. In conventional researches, there are many

DOI: 10.24507/ijicic.20.04.1093

motion recognition methods using contact sensors [3,4]. The method of detecting feature values using wearable inertial sensors [5] obtained high recognition accuracy. Furthermore, Warunsin et al. [6] proposed a human activity recognition method based on LSTM networks and using the datasets from the triaxial accelerometer and gyroscope sensors provided by MobiAct and WISDM. Most of the database they used for training and testing are standing, walking and jogging. Zhang et al. [7] proposed a method for human motion recognition by analyzing data from a wearable motion capture suit. However, the human motion recognition method using wearable inertia sensors is not suitable for daily use. Moreover, signal data from inertia sensors can only provide motion signals and it does not contain the information on environmental situation compared with a vision sensor which is useful for more precise physical analysis. Waqar et al. [8] proposed a method called a distributed MIMO (Multiple-Input-Multiple-Output) radar system for human activity recognition employing multiple antennas, but it is limited to indoor scenarios. In conventional researches of non-contact sensors, the method using computer vision system is common. Conventional methods include the Flow Vectors method [9], which creates the optical flow distribution of the region of interest, and the MHI method [10], which describes the history of motion as changes in luminance values in one image. However, these methods using a single camera assume that the motion for recognition is on the plane perpendicular to the optical axis of a camera, and the motion toward or leaving the camera is not dealt with because of its self-occlusive nature. Similarly, [11,12] both delved into research on human behavior recognition using deep learning, but they did not focus on motion in the depth direction captured by cameras. In order to solve this problem, an extended 3D-MHI [13] has been proposed. In addition, a method of reconstructing motion three-dimensionally using multiple cameras [14] and a method of recognizing motion trajectory using 3D-MHI (Three-Dimensional Motion History Image) with ELM (Extreme Learning Machine) [15] have also been proposed. However, they had the problem of high computational cost. And the reverse MHI method [16] is proposed to represent and recognize human motion based on inverse representation of actions. However, the inverse representation algorithm depends on the depth direction of movement (from front/back to back/front) and the recognition rate needs to be improved.

In this paper, we propose a novel human motion description and recognition method employing a single camera, which can deal with the motions to and from a camera by expanding the MHI. The method uses TMRIs (Triplet Motion Representation Images) that represents depth information, i.e., approaching or leaving from a camera, by three characteristic images. We also propose Ex-HOOF (Extended Histogram of Oriented Optical Flow) and the center of gravity and area intelligence to represent the speed and direction of motion information. Finally, the proposed method achieved an average motion recognition rate of 96.02%.

2. **Proposed Method.** The proposed method is mainly divided into 6 processes as shown in Figure 1. In the first step, the human region is extracted using a background model that can respond to changes in the background. In the second step, an FoE (Focus of Extension) is detected using the optical flow obtained by tracking the moving object if the motion contains the movement toward an observing camera or away from the camera. Therefore, by detecting the existence of an FoE, we can judge whether or not the motion is in the direction of the optical axis of the camera. The third step is to describe human motion by extending the conventional MHI method, a method which we refer to as TMRIs. The TMRIs are then transformed to Zernike moments [17] to calculate the shape features of TMRIs [18]. In the fourth step, an Ex-HOOF is proposed to extract the feature of a motion's direction and its speed. In the fifth step, the change of motion can be obtained

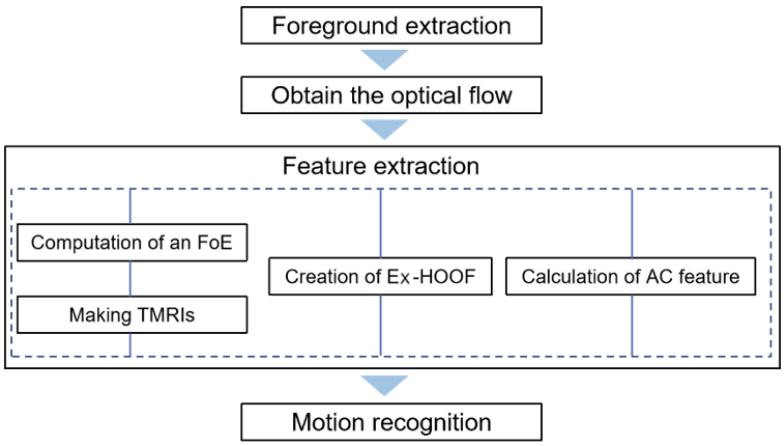


FIGURE 1. Flowchart of the proposed method

by calculating the related characteristics of the area and the center of gravity of the object. For simplicity, we will refer to the feature as AC going forward. Finally, in the sixth step, these features are integrated to perform action recognition through a classifier. Each process is described as follows.

2.1. **Foreground extraction.** In this research, successive background image estimation is performed using GMM (Gaussian Mixture Model) [19] for each pixel that constitutes an image, and human regions are extracted from time-series images. The EM (Expectation-Maximization) algorithm [20], which is an optimization algorithm, is used to estimate each parameter of the initial Gaussian mixture model.

2.2. **Computation of an FoE.** If the line obtained by extending the optical flow of the motion ultimately converges to a specific point, that point is identified as an FoE. To prevent bias in the detection of an FoE, feature points are placed at regular intervals on the contour of the human region. The feature points are tracked using the LK tracker [21] to obtain the optical flow. Then RANSAC is employed to remove the outliers.

In this study, we adopt the method of weighted voting to reduce the possible wrong detection of an FoE due to the deviation of the vote. Weights are assigned to the lines representing the extension of the optical flow. The closer a point is to the center of the line, the greater the assigned weight. Then the voting is performed. If the point with the maximum value of voting results selected from the obtained results is greater than a threshold, it will be considered as an FoE.

2.3. **Description of a motion: TMRIs.** In this paper, we propose a motion description method, called TMRIs, which is an extension of the conventional MHI. It is designed to express human motion, including motion along the camera's optical axis, using three types of motion history images: *newness*, *density*, and *depth*. Among them, *newness* represents the MHI, *density* illustrates the appearance frequency of the foreground in the past τ frames, and *depth* indicates the depth information obtained from the FoE detection results. Each history image is defined by the following equations:

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

$$H_\tau^{den}(x,y,t) = \sum_{i=0}^{\tau} D(x,y,t-i) \tag{2}$$

$$H_\tau^{dep}(x,y,t) = \sum_{i=0}^{\tau} \{N_{layer}(t-i) \times D(x,y,t-i)\} \tag{3}$$

Here $H_\tau(x,y,t)$ is the gray value at (x,y,t), and D is a binary image showing a foreground region. N_{layer} is the number of layers superimposed in the depth direction, and is determined as follows:

$$N_{layer} = \begin{cases} \gamma \times L_{ave} & \text{if } V_{\max} > T_{vote} \text{ or } T_1 < Area_\tau^{ave} < T_2 \\ 1 & \text{otherwise} \end{cases} \tag{4}$$

Here $V_{\max}$ is the maximum value of voting results. T_{vote}, T_1 and T_2 are predetermined thresholds. $Area_\tau^{ave}$ is the average value of the foreground area in the last τ frames. L_{ave} is the average value of the magnitude of the optical flow, and γ is a constant used to determine the value of N_{layer} according to L_{ave}.

The features of TMRIs are described using Zernike moments, which are orthogonal moments with the property of rotation invariance. In this paper, the feature vector of each motion history image described by Zernike moments is defined as $\boldsymbol{s} = (s_1, s_2, \ldots, s_9)$, which is a 9-D vector. Since TMRIs contain three types of motion history images, and each image is described by Zernike moments, the feature vector of TMRIs is a $(3 \times 9 =)$ 27-D vector described as

$$\boldsymbol{V}^{TMRIs} = \left(\boldsymbol{s}^{new}, \boldsymbol{s}^{den}, \boldsymbol{s}^{dep}\right) \tag{5}$$

2.4. **Creation of the Ex-HOOF.** Considering that the TMRIs feature can barely show the speed and direction of the motion, we propose a method called Ex-HOOF [22], an extension of the traditional HOOF [23].

As shown in Figure 2, the directions of optical flow vectors are separated into n directions and the length of the optical flow of each direction is accumulated to form a histogram representing the magnitude of the optical flow for each direction.

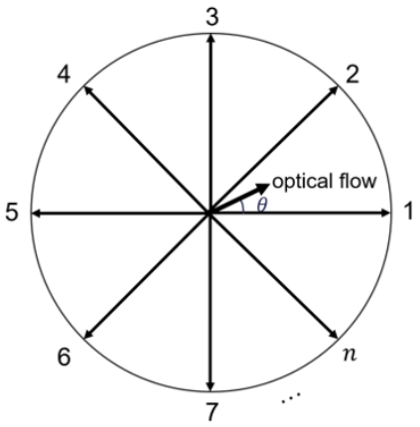


FIGURE 2. The Ex-HOOF

Given an optical flow vector $v = (x,y)^T$, the angle $\theta = \tan^{-1}(y/x)$ and the length $L = \sqrt{x^2+y^2}$ are calculated. If θ is the angle between bin a and bin b $(a, b = 1, 2, \ldots, n)$, $L \times w_a$ and $L \times w_b$ are accumulated to bin a and bin b, respectively, where the weights w_a, w_b are calculated by

$$w_a = \frac{|\theta - 2\pi/n \times (b-1)|}{2\pi/n} \tag{6}$$

$$w_b = \frac{|\theta - 2\pi/n \times (a-1)|}{2\pi/n} \tag{7}$$

As in TMRIs, τ successive frames are also utilized in this section. Optical flow vectors are calculated from every other frame within the τ frames. All these optical flow vectors in the past τ frames are accumulated in a single histogram, and normalization is performed to ensure that the total frequency becomes 1. The feature vector of Ex-HOOF, which has n bins, is described by

$$\boldsymbol{V}^{Ex\text{-}HOOF} = (f_1, f_2, \ldots, f_n) \tag{8}$$

Here, f_i $(i = 1, 2, \ldots, n)$ is the frequency of the ith bin under the condition that $\left\|\boldsymbol{V}^{Ex\text{-}HOOF}\right\| = 1$. The number n is set to 30 in the experiment; therefore, $\boldsymbol{V}^{Ex\text{-}HOOF}$ is a 30-D feature vector.

2.5. **Changes in the area and the center of gravity.** In order to reflect changes in actions and improve the accuracy of action recognition, we introduce features related to the area and the center of gravity of human regions. Among them, the area feature is defined as $\boldsymbol{F}_A = \left(Area_\tau^{ave}, Area_\tau^{sd}\right)$. The average value $Area_\tau^{ave}$ and standard deviation $Area_\tau^{sd}$ of the change of the foreground area can be calculated as follows:

$$Area_\tau^{ave} = \frac{1}{\tau}\sum_{i=0}^{\tau} Area_{comp}^{t-i} \tag{9}$$

$$Area_\tau^{sd} = \sqrt{\frac{1}{\tau}\sum_{i=0}^{\tau}\left(Area_{comp}^{t-i} - Area_\tau^{ave}\right)^2} \tag{10}$$

Here, $Area_{comp}^{t}$ is the changes of areas between the frames and is given by

$$Area_{comp}^{t} = \frac{(Area_t - Area_{t-p})}{Area_t} \tag{11}$$

Here, $Area_t$ is the area of the foreground at time t, and p is a frame interval that automatically adjusts according to the size of the optical flow.

Similarly, the feature of the center of gravity is defined as $\boldsymbol{F}_C = \left(Cx_\tau^{ave}, Cy_\tau^{ave}, Cx_\tau^{sd}, Cy_\tau^{sd}\right)$. C_τ^{ave} means the average value and C_τ^{sd} indicates the standard deviation of the change in the coordinate of the center of gravity. The changes in the center of gravity are defined by the following equations:

$$Cx_\tau^{ave} = \frac{1}{\tau}\sum_{i=0}^{\tau}\left(Cx_t - Cx_{(t-i-p)}\right) \tag{12}$$

$$Cy_\tau^{ave} = \frac{1}{\tau}\sum_{i=0}^{\tau}\left(Cy_t - Cy_{(t-i-p)}\right) \tag{13}$$

$$Cx_\tau^{sd} = \sqrt{\frac{1}{\tau}\sum_{i=0}^{\tau}\left(\left(Cx_t - Cx_{(t-i-p)}\right) - Cx_\tau^{ave}\right)^2} \tag{14}$$

$$Cy_\tau^{sd} = \sqrt{\frac{1}{\tau}\sum_{i=0}^{\tau}\left(\left(Cy_t - Cy_{(t-i-p)}\right) - Cy_\tau^{ave}\right)^2} \tag{15}$$

Here, Cx_t and Cy_t are the x and y coordinates of the center of gravity of the foreground at time t, respectively.

Finally, we get a 6-dimensional feature about the change of the action area and the center of gravity, denoted as $\boldsymbol{V}^{AC}$, defined by

$$\boldsymbol{V}^{AC} = (\boldsymbol{F}_A, \boldsymbol{F}_C) \tag{16}$$

2.6. **Motion recognition.** By integrating the above features, we get a 63 $(27 + 30 + 6)$ dimensional feature vector $\boldsymbol{V}$, as shown below.

$$\boldsymbol{V} = \left(\boldsymbol{V}^{TMRIs}, \boldsymbol{V}^{Ex\text{-}HOOF}, \boldsymbol{V}^{AC}\right) \tag{17}$$

Motion recognition is performed using a k-nearest neighbor classifier in this paper. The class of motion l is denoted by C^l, the jth learning data in C^l by v_j^l, and the input unknown motion is by $\boldsymbol{v}$. Then the procedure for the k-nearest neighbor is given by

$$l^* = maj_l \left\{ \arg_{\boldsymbol{v}_j^l} k \min \left\{ I\left(\boldsymbol{v}, \boldsymbol{v}_j^l\right) \forall l, \forall j, \boldsymbol{v}_j^l \in C^l \right\} \right\} \tag{18}$$

Here l^* represents the recognized class; $maj_l\{S\}$ returns the class that appears the most in set S; $k \min\{T\}$ is the k minimum numbers in the set T; I is the dissimilarity measure defined by $I(\boldsymbol{v}, \boldsymbol{w}) = \|\boldsymbol{v} - \boldsymbol{w}\|$.

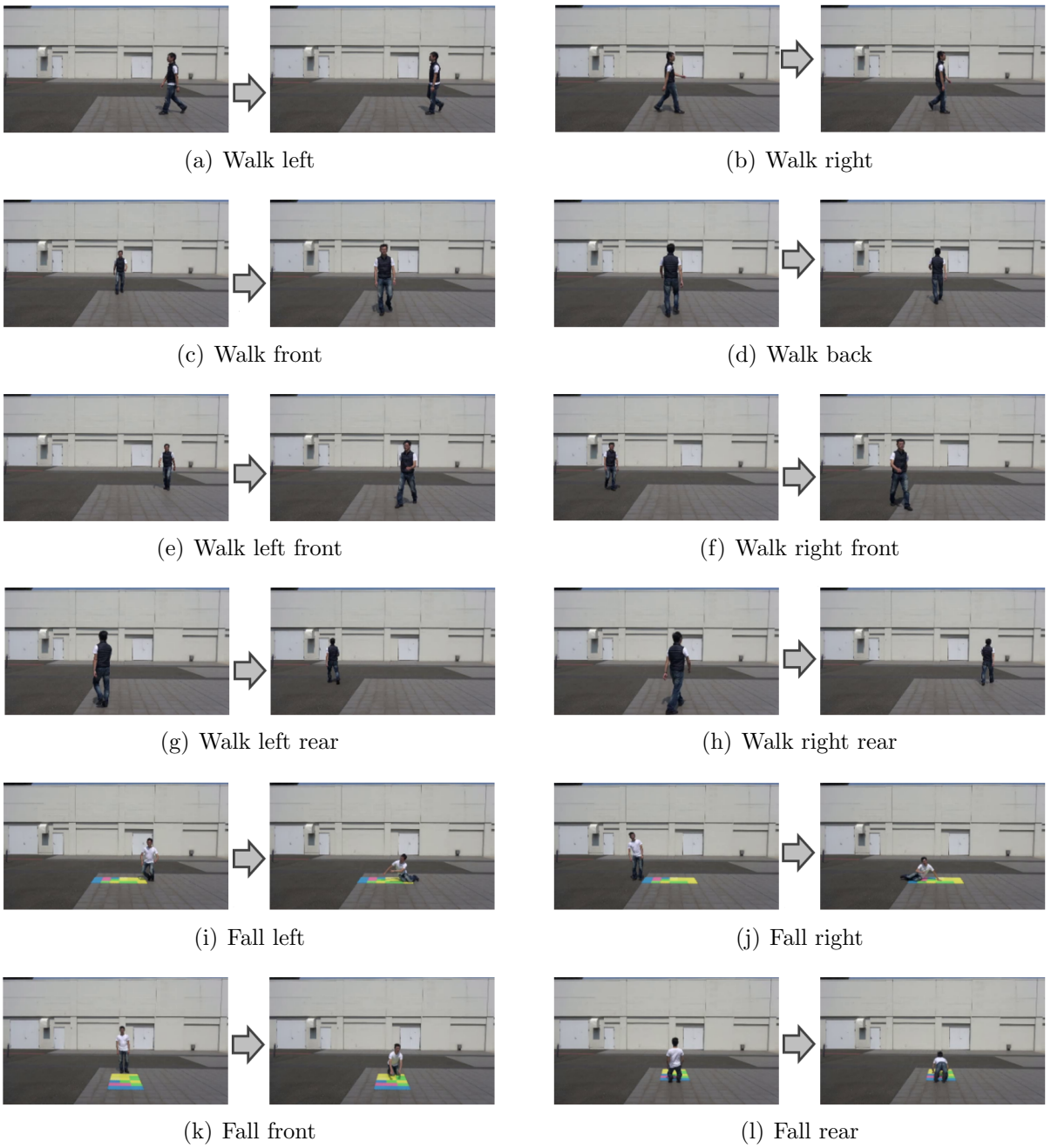
3. **Experimental Results.** In this section, we evaluated the accuracy of the detection of FoE and motion recognition in order to verify the performance of the proposed method. In the experiment, 12 kinds of motions were chosen and acted by 4 people (students) aged 22 to 24. As shown in Figure 3, the motions include walk and fall activities, Figures 3(a)-3(h) show the eight directions of walk activities, i.e., "walk left/right, walk front/back, walk left front/right front/left rear/right rear" whereas Figures 3(i)-3(l) show the four directions of abnormal activities "fall left/right, fall front/rear". Figure 4 shows the examples of TMRIs, with each history image displayed in its respective RGB channel. In Figure 4(c2), it is evident that during the action of walking backward, changes in the depth direction allow for the expression of depth information in the image. Conversely, the action of walking left does not involve movement in the depth direction, resulting in a TMRIs representation with no depth map information. Figure 5 shows examples of TMRIs representation of the motion. To enhance the visibility of the motion, three types of motion history images are superimposed on one image in the figure. Based on MHI, 3D-MHI, and our four types of the proposed method, and multiple experiments by tau values of 15, 20 and 25, and further by considering both recognition accuracy and processing time, we found that the setting parameter $\tau = 20$ yields the best results.

3.1. **Detection of an FoE.** When the maximum count in the voting result exceeds a threshold (set to 20 in this experiment), an FoE is detected. The accuracy of FoE detection is evaluated using the following formula:

$$Accuracy = \frac{F_{TP} + F_{TN}}{F_{ALL}} \times 100[\%] \tag{19}$$

Here, let TP represent the cases where the FoE is detected for the motion in the depth direction, and TN represents the cases where the FoE is not detected for other motions. F_{ALL} is the total number of frames for which detection was performed, and F_{TP} and F_{TN} are the numbers of TP and TN frames, respectively. The average accuracy of the FoE detection for all motions is 69.3%. Specifically, for motions in the depth direction, the accuracy is 76.3% (walk front) and 80.2% (walk back). This indicates that the FoE detection is beneficial for our subsequent action recognition in the depth direction.

3.2. **Motion recognition.** To calculate the recognition rate of the proposed method, 80 sets of feature vectors are chosen from normal daily motion and 40 sets are chosen from each abnormal motion. Consequently, a total of 3200 pairs of vectors are obtained.



(a) Walk left (b) Walk right

(c) Walk front (d) Walk back

(e) Walk left front (f) Walk right front

(g) Walk left rear (h) Walk right rear

(i) Fall left (j) Fall right

(k) Fall front (l) Fall rear

FIGURE 3. Examples of human motion: (a)-(h) Eight directions of walk motions; (i)-(l) four directions of fall motion

Motion recognition is conducted using leave-one-out cross-validation. For the k-nearest neighbor algorithm, k is set to 3. The recognition rate R is defined by

$$R = \frac{N_T}{N_{ALL}} \times 100[\%] \tag{20}$$

Here, N_T represents the total number of correctly recognized features and N_{ALL} represents the total number of features.

Table 1 shows the recognition rate obtained using TMRIs, Ex-HOOF, TMRIs + Ex-HOOF (proposed-1), and the proposed-2 method (TMRIs + Ex-HOOF + AC). Here, the feature of the change of center of gravity and the area is denoted as AC. As shown in

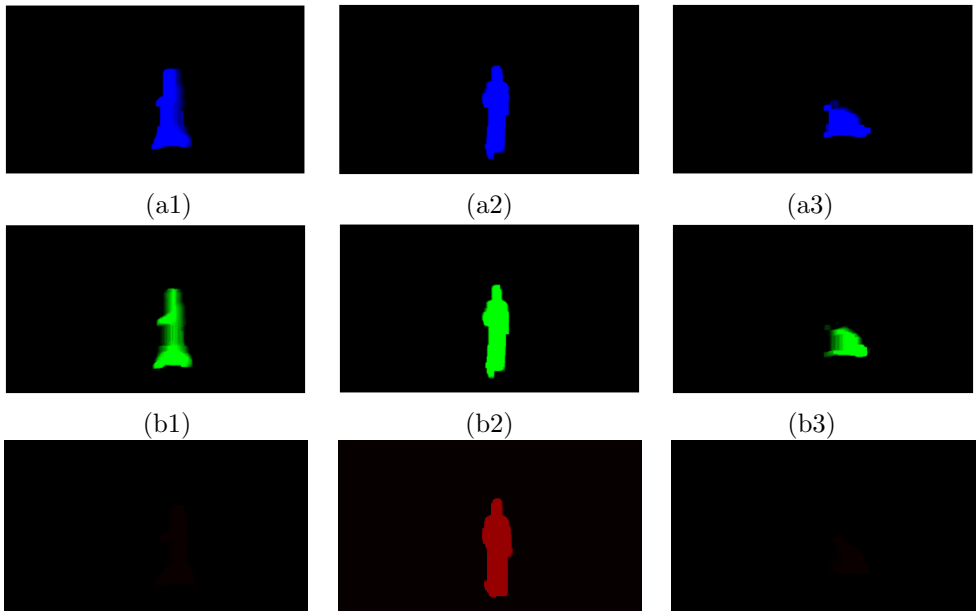


FIGURE 4. TMRIs representation of a human motion: (a) *newness* – B channel; (b) *density* – G channel; (c) *depth* – R channel; (#1) walk left; (#2) walk back; (#3) fall left

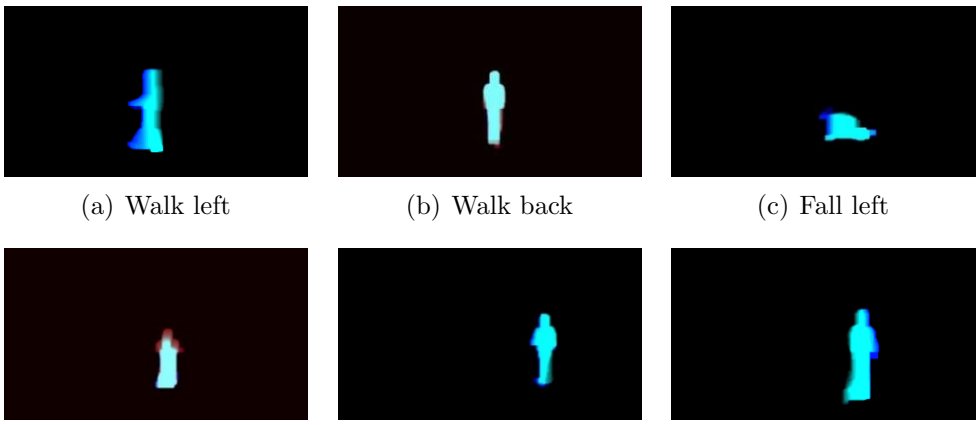


FIGURE 5. TMRIs representation (overlay of RGB images)

the table, the proposed method achieved an average recognition rate of 96.02%. This is because TMRIs contain *depth* that expresses depth information, which is determined by the detection results of the FoE. Ex-HOOF contains the speed and direction information of the movement that TMRIs cannot represent. Additionally, we incorporate detailed features to express falling movements through changes in the area and center of gravity of the movement. Therefore, in the final proposed method, the average accuracy for daily actions is 99.10%, and for fall actions, it is 89.85%. In addition, the average recognition accuracy of actions containing motion in the depth direction is 96.60%, while the average recognition accuracy of other actions is 94.85%. It demonstrates the effectiveness of the proposed approach.

Furthermore, the recognition accuracies obtained using the MHI and reverse MHI methods [16] are 60.3% and 83.7%, respectively, without directional discrimination, whereas

TABLE 1. The recognition rate of the motions

Motion	Recognition rate [%]			
	TMRIs	Ex-HOOF	TMRIs + Ex-HOOF (Proposed-1)	TMRIs + Ex-HOOF + AC (Proposed-2)
Walk left	87.50	44.38	93.13	99.38
Walk right	70.31	93.12	99.38	100.00
Walk front	78.12	75.31	99.38	99.69
Walk back	73.12	97.81	100.00	100.00
Walk left front	83.75	96.88	96.25	99.06
Walk right front	87.81	72.81	94.69	100.00
Walk left rear	83.44	92.19	96.56	98.75
Walk right rear	83.12	94.38	95.31	95.94
Fall left	91.88	63.75	91.88	91.88
Fall right	79.38	71.25	75.63	88.13
Fall front	61.88	44.38	80.63	89.38
Fall rear	69.38	63.12	86.25	90.00
Average	79.14	75.78	94.42	96.02

the average recognition result of 3D-MHI [13] is 96.3%, where two kinds of directions are taken into account such as 'walk left', 'walk back', 'fall left', 'fall back', 'squat on floor', and 'sit on chair', where 'back' means the motion in the depth direction. On the other hand, the method we propose not only achieves motion recognition but also demonstrates high accuracy in directional recognition where 8 kinds of directions are taken into account, that is, left, right, front, back, left front, right front, left rear/back, right rear/back. The overall average of both motion recognition and directions accuracy is 96.02%. Therefore, under the same experimental conditions as this paper, the conventional method cannot achieve the expected results.

Due to differences in experimental equipment configuration, we evaluate the processing time of the proposed method relative to TMRIs processing time. In the experiments based on MHI [10], 3D-MHI [13] and TMRIs, the processing time of MHI, 3D-MHI and TMRIs (using Hu moments for shape feature extraction) is 18.90 ms/frame, 466.08 ms/frame, and 20.12 ms/frame, respectively, resulting in a ratio of approximately 23.17 times faster than 3D-MHI by an OS CPU Core(TM) i7-2600 3.40GHz. Although the reverse MHI method [16] does not report the processing time, considering that it is based on the inversion of MHI, it can be assumed that the processing time is approximately the same as that of MHI. According to our experimental setup, the processing times for TMRIs (using Hu moments [22]) and TMRIs (using Zernike moments) are 37.91 ms/frame and 151.24 ms/frame, respectively, whereas the processing time for the proposed-2 (TMRIs + Ex-HOOF + AC) method is 154.63 ms/frame, resulting in a ratio of approximately 3.01 times faster than 3D-MHI by a computer with CPU Intel(R) Core(TM) i7-10700 2.90GHz. The actual processing time is also influenced by the factors such as generating various number of sub-output results or writing data process.

From this fact, it can be observed that the proposed method in this paper, compared to traditional methods capable of achieving depth-directional action recognition, offers the advantages of faster processing time and higher accuracy.

4. **Conclusions.** In this paper, we propose a human motion recognition method capable of describing and recognizing human motions in the depth direction using a single camera. The distinguishing feature of our method is its ability to handle motions toward the depth direction, setting it apart from existing human motion recognition methods.

For each motion, the proposed method, TMRIs, is applied to express the motion using Equations (1)-(4). Three types of motion history images (*newness*, *density*, *depth*) are created for each of the B, G, and R channels, allowing them to be represented in a single image. Additionally, Ex-HOOF is employed to extract the feature vector representing the direction and speed of each motion. The change in shape and location of the human region is calculated by Equations (9)-(16). The proposed method employs the overall 63-D feature vector $\boldsymbol{V}$ defined in Equation (17), which combines the above three feature vectors. An unknown motion is then recognized using the 3-nearest neighbor classifier.

A recognition experiment involving 12 motions demonstrated the effectiveness of our proposed method. According to the results in Table 1, the proposed method using TMRIs + Ex-HOOF + AC achieved the highest recognition rate across all motions compared to any other method. The average recognition rate reached 96.02% in the experiment.

Although the current action recognition rate is satisfactory, there is still room for improvement. We can enhance the recognition rate through further research on the characteristics of fall actions and by increasing the amount of experimental data.

REFERENCES

[1] Cabinet Office, *2022 White Paper on Aging Society (Overall Version)*, 2022 (in Japanese).
[2] Cabinet Office, *Outline of the 2021 Public Opinion Poll on Security*, 2021 (in Japanese).
[3] M. M. Hassan, M. Z. Uddin and A. Mohamed, A robust human activity recognition system using smartphone sensors and deep learning, *Future Generation Computer Systems*, vol.81, pp.307-313, 2018.
[4] Z. Liu and J. K. Tan, Analysis of human walking posture using a wearable camera, *International Journal of Innovative Computing, Information and Control*, vol.19, no.3, pp.805-819, 2023.
[5] E. M. Tapia, S. S. Intille, W. Haskell, K. Larson, J. Wright, A. King and R. Friedman, Real-time recognition of physical activities and their intensities using wireless accelerometers and heart rate monitor, *Proc. of International Symposium on Wearable Computers*, pp.97-104, 2006.
[6] K. Warunsin, K. Promjiraprawat and O. Chitsobhuk, Human activity recognition using long short-term memory network, *International Journal of Innovative Computing, Information and Control*, vol.19, no.3, pp.973-990, 2023.
[7] Z. Zhang, S. Zhang, Z. Zhao, Y. Liu and Y. Kan, Research on human motion recognition method based on hybrid convolution neural network-hidden Markov model, *2021 3rd International Conference on Robotics and Computer Vision (ICRCV)*, pp.52-56, 2021.
[8] S. Waqar, M. Muaaz and M. Pätzold, Direction-independent human activity recognition using a distributed MIMO radar system and deep learning, *IEEE Sensors Journal*, vol.23, no.20, pp.24916-24929, 2023.
[9] E. L. Andrade, R. B. Fisher and S. Blunsden, Detection of emergency events in crowded scenes, *Proc. of IEEE International Symposium on Imaging for Crime Detection and Prevention*, Hong Kong, China, pp.528-533, 2006.
[10] A. Bobick and J. Davis, The recognition of human movement using temporal templates, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.23, no.3, pp.257-267, 2001.
[11] Y. Wang, S. He, X. Wei and S. A. George, Research on an effective human action recognition model based on 3D CNN, *2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp.1-6, 2022.
[12] T. Shanableh, ViCo-MoCo-DL: Video coding and motion compensation solutions for human activity recognition using deep learning, *IEEE Access*, vol.11, pp.73971-73981, 2023.
[13] Y. Yamashita, J. K. Tan and S. Ishikawa, Human motion description and recognition under arbitrary motion direction, *Proc. of SICE Annual Conference*, pp.110-115, 2017.
[14] J. Michelson and A. Hilton, Simultaneous pose estimation of multiple people using multiple-view cues with hierarchical sampling, *Proc. of British Machine Vision Conference*, pp.1-10, 2003.

[15] Z. Chang, X. Ban, Q. Shen and J. Guo, Research on three-dimensional motion history image model and extreme learning machine for human body movement trajectory recognition, *Mathematical Problems in Engineering*, vol.2, 2015.
[16] J. K. Tan, S. Okae, Y. Yamashita and Y. Ono, A method of describing a self-occlusive motion – A reverse motion history image, *International Journal of Biomedical Soft Computing and Human Sciences*, vol.24, no.1, pp.1-7, 2019.
[17] Z. Zhou, P. Liu, G. Chen and Y. Liu, Moving object detection based on Zernike moments, *2016 5th International Conference on Computer Science and Network Technology*, pp.696-699, 2016.
[18] F. Z. Chelali and A. Djeradi, Zernike moments and histogram of oriented gradient descriptors for face recognition from video sequence, *2014 2nd World Conference on Complex Systems*, pp.686-693, 2014.
[19] C. Stauffer and W. E. L. Grimson, Adaptive background mixture models for real-time tracking, *Proceedings of Conference on Computer Vision and Pattern Recognition*, Fort Collins, CO, USA, vol.2, pp.246-252, 1999.
[20] A. P. Dempster, N. M. Laird and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society*, vol.39, no.1, pp.1-38, 1977.
[21] B. D. Lucas and T. Kanade, An iterative image registration technique with an application to stereo vision, *Proc. of Int. Joint Conf. on Artificial Intelligence*, pp.674-679, 1981.
[22] J. Cao, Y. Yamashita and J. K. Tan, Human motion recognition using TMRIs with extended HOOF, *Journal of Robotics, Networking and Artificial Life*, vol.7, no.4, pp.231-235, 2021.
[23] R. Chaudhry, A. Ravichandran, G. Hager and R. Vidal, Histograms of oriented optical flow and Binet-Cauchy kernels on nonlinear dynamical systems for the recognition of human actions, *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp.1932-1939, 2009.

Author Biography

Jing Cao received her B.E. degree from Yangzhou University of China in 2017 and M.E. degree from the Graduate School of Engineering, Kyushu Institute of Technology, Japan in 2020. She is acquiring the D.E. degree in the same university. Her research interests include computer vision, machine learning and motion recognition.

Yui Tanjo received the Ph.D. degree from Kyushu Institute of Technology. She is currently a professor at the Department of Mechanical and Control Engineering, Kyushu Institute of Technology. Her current research interests include ego-motion analysis by MY VISION, three-dimensional shape/motion recovery, human detection, and its motion analysis from video. She was awarded SICE Kyushu Branch Young Author's Award in 1999, the AROB Young Author's Award in 2004, the Young Author's Award from IPSJ of Kyushu Branch in 2004, and the BMFSA Best Paper Award in 2008, 2010, 2013, and 2015. She is a member of IEEE, The Information Processing Society, and The Institute of Electronics, Information and Communication Engineers of Japan.