International Journal of Multimedia and Ubiquitous Engineering
Vol.10, No.10 (2015), pp.211-222
http://dx.doi.org/10.14257/ijmue.2015.10.10.22

Human Action Recognition Using Accumulated Moving Information

Nae-Joung Kwak[1] and Teuk-Seob Song[2*]

*[1] Dept. Communication & Information, Chungbuk National University
Cheongju, 362-763, KOREA*
*[2] Div. Convergence Computer and Media, Mokwon University,
Daejeon, 302-729, KOREA*
[1]knj0125@hanmail.net, [2]teukseob@mokwon.ac.kr

Abstract

This paper proposes a human action recognition algorithm which can be efficiently applied to a real-time intelligent surveillance system. This method models the background, obtains the difference image between input image and the modeled background image, extracts the silhouette of human object from input image, and recognizes human action by using coordinates of object, directions of that and accumulated moving regions of that. The human actions recognized in this study amount to a total of 8 type of actions, which include walking, raising an arm (left, right), raising a leg (left, right), sitting and crouching. The proposed method has been experimented for 8 different movements using 4 people using video input of a webcam and it has shown good results in terms of recognizing human action.

Keywords: *IoT action recognition, moving information, IoT auto detection, IoT human action*

1. Introduction

As a rising interest in human and computer interaction is growing, the research on ubiquitous computing, emotion computing and other forms of research that reflect smart interacting environment have been studied. Also, as advances in scientific technology have created a ubiquitous environment, an intelligent surveillance system is linked to home network to build the intelligent home networking system. Such intelligent home networking system is utilized to monitor infants, seniors living alone and the physically challenged. Studies on human action recognition technologies are being conducted to detect emergencies.

Human action recognition technologies are applied to various fields such as a video surveillance system, human-computer interaction, video indexing and sports video analysis. Human movement recognition, which includes gesture recognition, sign language recognition, gait pattern recognition and action recognition, has received much attention as it became an important field of application in diverse areas. Especially it became so with the increase of senior citizens living alone, rate of crime and violence in the streets, and the need of effective care systems in kindergartens and other forms childcare centers. Therefore automated human action recognition can be applied and used in many different areas. Auto detection of human action is implemented through video surveillance systems and research on smart surveillance systems are currently in progress [1-3].

[*] Corresponding author: Division of Convergence Computer and Media, Mokwon University, Daejeon, Korea 302-729, E-mail: teukseob@mokwon.ac.kr

ISSN: 1975-0080 IJMUE

Video surveillance systems consist of sending input video frames to a computer that analyzes the images to recognize human action patterns. The information gathered through this process can be applied in various different areas [4-5].

Human action recognition follows three steps: preprocessing, action modeling and action recognition. In the preprocessing step, the object of interest is extracted and the background is subtracted [6]. Action modeling is a step to model human actions and get information needed for action recognition. For action modeling, there are two approaches: body structure approach and holistic approach. Body structure approach first models human actions and then gains information needed to recognize actions. Holistic approach is based on whole contour of the human body [7]. Body structure approach is divided into top-down [8] and bottom-up [9] approach. Top-down processing first models body structure and matches input images with the modeling. Bottom-up processing first detects specific body parts from input images and connects them. This approach requires a lot of calculations because it has to detect each part of body and estimate actions.

Holistic approach extracts bodies as objects and models actions by using form, contour, texture, silhouette, location, trajectory and velocity of objects. Since this approach models bodies holistically and requires less calculation than body-part approach, holistic approach is widely used [10]. The features of extracted objects combined with occurring events are used as basic information to recognize human actions. Methods which extract objects of interest from video images and recognize their features and actions have steadily been studied and developed depending on image recognition [11-14]

However, existing methods require studying the characteristics of the human body from many frames and using that data for assessing posture or action. This requires great amount of data and complex learning algorithms.

This paper proposes a method that recognizes human actions, which firstly models the background from input video, extracts main object by using difference input video from background-model, uses accumulated differential image to be obtained from present and previous video frames. The proposed method detects object’s movement from accumulated differential images and the moving region is used to recognize human action.

2. Human Action Recognition Using Accumulated Moving Information Correct

This paper makes the model of background and separates object from background using differential images extracted from modeled background and input video images. Human actions are classifies using accumulated moving information images of the extracted object. Figure 1 show the flowchart of the proposed method.

2.1. Creation of Accumulated Moving Information Image Using Object-Background Segmentation and Moving Region

This paper uses the method proposed in [16] for background modeling. Human action is recognized by detecting the main object, which is a human, gathering accumulated moving information images from accumulated moving information and detecting and tracking the regions where movement exists. The following is the process of obtaining accumulated moving information regions.

1) Obtain segmented object-background Image ($B_i(x,y)$) of modeled between background and input image.

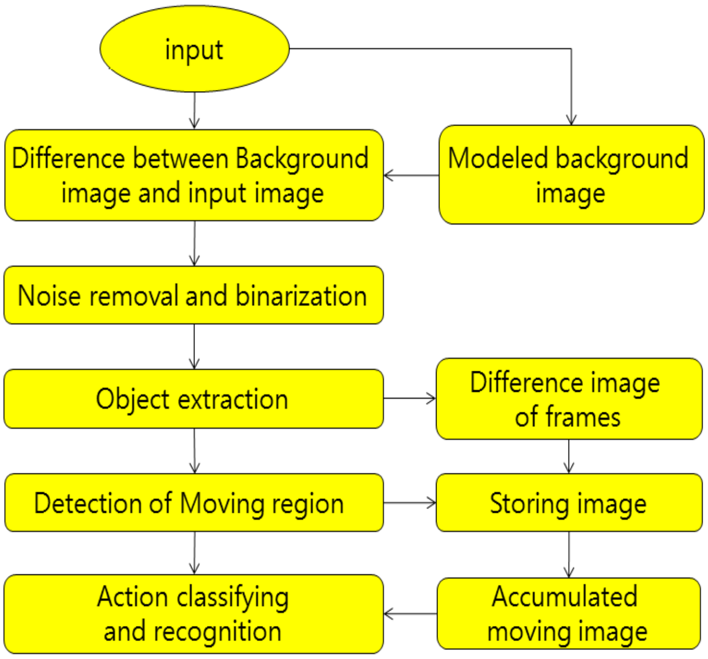

input
Difference between Background image and input image
Modeled background image
Noise removal and binarization
Object extraction
Difference image of frames
Detection of Moving region
Storing image
Action classifying and recognition
Accumulated moving image

Figure.1 The Flowchart of the Proposed Method

2) Obtain moving image ($M_i(x,y)$).

3) Obtain accumulated moving information image ($AM(x,y)$).

Figure 2 shows the process of obtaining accumulated moving information image for human action recognition. First, differential images are extracted from modeled background and each input image frame. The differential image is an image where the object and background have been separated. It will be considered as background if pixel value of input image on three color spaces exists between low threshold value and high threshold value, if not, it will be considered as object. The following is the equation for separating background and object.

$$B(x,y) = \begin{cases} 0 & , \quad Th_l \le I(x,y) \le Th_h \\ 255 & , \quad otherwise \end{cases} \tag{1}$$

Where x, y are the position of input image and object will be binarized as white(255) while background is black(0). Final resulting image is obtained by the addition of three binary planes.

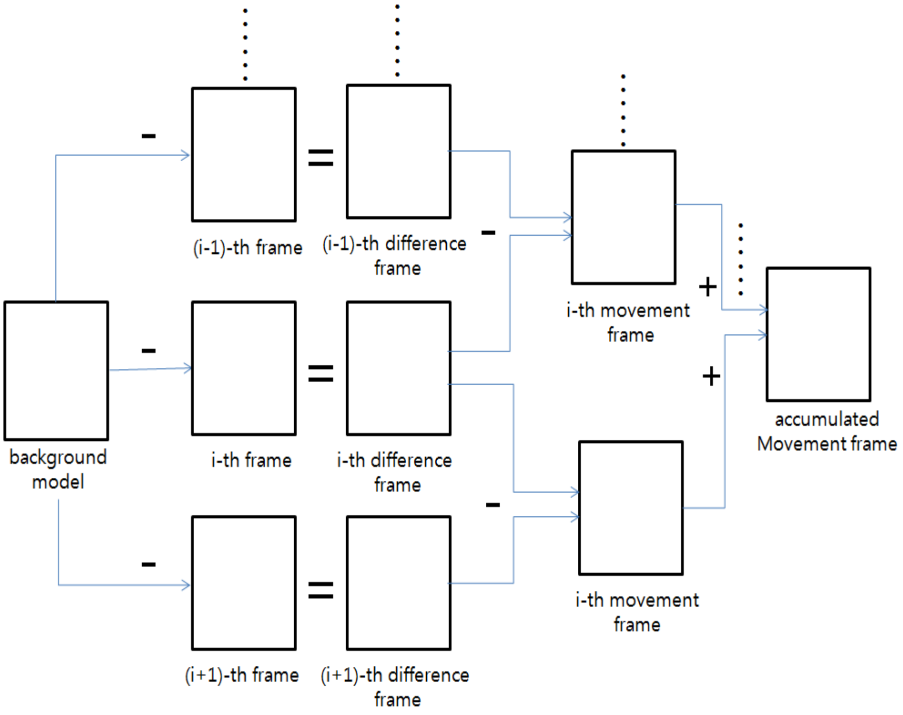


Figure 2. The Creating of Accumulated Moving Information Image

Binary images include various small regions except the object region. Therefore, morphological filter is used to remove these small regions. Therefore, object-background segmented image ($B(x,y)$) is produced by using filling holes by morphological filter and removing small regions below threshold. Only the main object exists within the object-background segmented image ($B(x, y)$) and the difference between two object-background segmented images includes the changing region of the object, which is to say, that it only has the moving region of the object. In this paper, the object's moving region is obtained from the differential image derived from i-1 object-background segmented image ($B_{i-1}(x, y)$) and i object-segmented image ($B_i(x, y)$).

$$D_i(x, y) = \left|B_{i-1}(x, y) - B_i(x, y)\right| \tag{2}$$

Moving image frames obtained through equation (2) includes various moving regions occurring during a given time period and the accumulated moving informations allows us to see where human body moving occurs. Equation (3) is used for obtaining accumulated moving information images.

$$AM(x, y) = AM(x, y) + D_i(x, y) \tag{3}$$

2.2. Action Recognition Using Moving Region

Among several human actions, the paper focuses on recognizing 8 actions which include sitting on a chair, raising one arm (left, right), raising both arms, raising a leg (left, right), crouching and walking. Two parameters are used to recognize human action, two of which are i) horizontal movement of object and ii) accumulated moving information of segmented region. Accumulated moving information images are accumulated moving information of the object. For the purpose of this paper, the moving region of 8 type of human action is detected, classified and numbered to achieve recognition. Figure 3 shows an object being segmented and numbered into a total of 15 regions, which include 3 horizontal moving regions and 5 vertical moving regions [15].

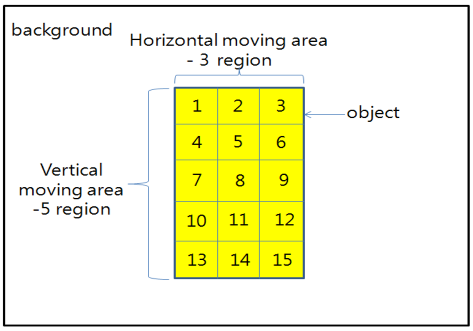
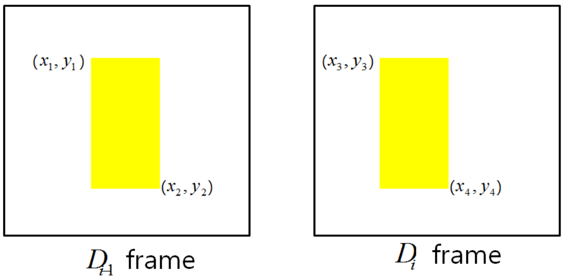


Figure. 3 Number of Moving Region of Object

The proposed method uses the two parameters mentioned above and a set of numbered moving regions is as follows.

(1) First action classification using horizontal moving region



$$\Delta x_1 = x_1 - x_3$$
$$\Delta y_1 = y_1 - y_3$$
$$\Delta x_2 = x_2 - x_4$$
$$\Delta y_2 = y_2 - y_4$$

Figure. 4 Change of Moving Region of Object

Moving region is extracted from $D_{i-1}(x,y)$ and $D_i(x,y)$. The region has horizontal and vertical direction from changes in x and y, extracted from object region.

If changes in Δx_1 and Δx_2 are greater than threshold value and equal in direction, while changes in Δy_1 and Δy_2 are below threshold value, then the action is classified as 'Walking(1)'. If not, the action is classified as other actions which are raising an arm (left, right), raising both arms, raising a leg (left, right), *etc.*

(2) Secondary action classification using numbered moving regions

After first action classification, where an action is Accumulated moving information images are extracted, moving regions are numbered, and action is classified as either action. Table 1 shows the proposed parameters and classification method classified as 1, other actions are classified as action 2~8. Accumulated moving information images are extracted, moving regions are numbered, and action is classified as either action. Table 1 shows the proposed parameters and classification method.

Table 1. 8 Parameter and Action Classification

Action number	Action classification	First classification			Number of moving region	Classification
		Δx_1	Δx_2	Direction of Δx_1 and Δx_2		
1	waking	o	o	same	-	first
2	Raising right arm	O	x	-	1,4,7	second
3	Raising left arm	X	o	-	3,6,9	
4	Raising two arms	O	o	different	1,3,4,6,7,9	
5	Raising right leg	O	x	-	10,13	
6	Raising left leg	X	o	-	12,15	
7	sitting	X	x	-	2,5	
8	crouch	X	x	-	2,5,8	

3. Test and Result

To analyze the performance of the proposed method, test was done indoors which have different backgrounds. The background video and the inputted video were analyzed real time by camera. Test images include total 4 persons and 8 actions for each person. The system was implemented by using Intel CPU 2.0GHz, 1G RAM, visual studio 2008 and Open CV 2.4. The resolution of the input video was 640X480 in 24-bit, which received 15 frames per second. Figure 5 shows the example of the input image.

Figure 6 shows the result to apply the proposed method to an experimental video. Background of figure (a) is used for 'modeling' and (b) is initial input image. (c) is a binary image by the proposed method and (d) is the obtained main object after removing small regions from (c), (e) is the last video frame, (f) is the binary image of (e), (g) is the main object image extracted from (f). (h) shows the moving region of the object using differential video frame obtained from (g) and previous input frames, and (i) is the accumulated moving information image. The results of (i) show the accumulated moving information extracted from the movements of right arm that according to table 1 is classified as action 2, which is raising right arm.

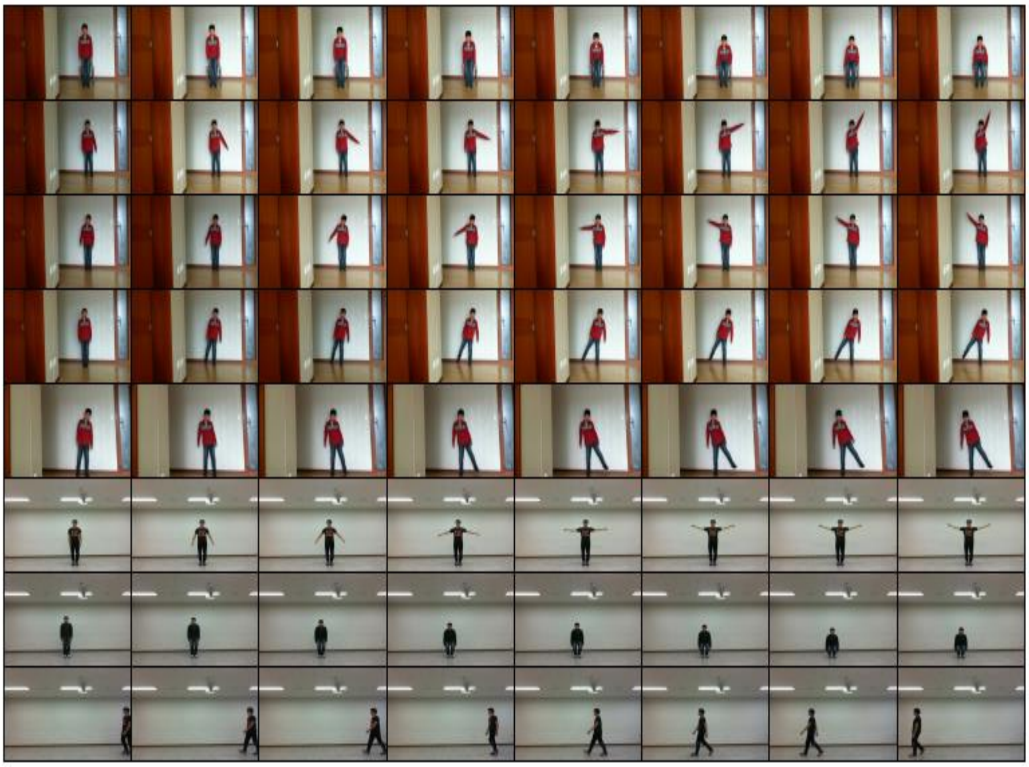
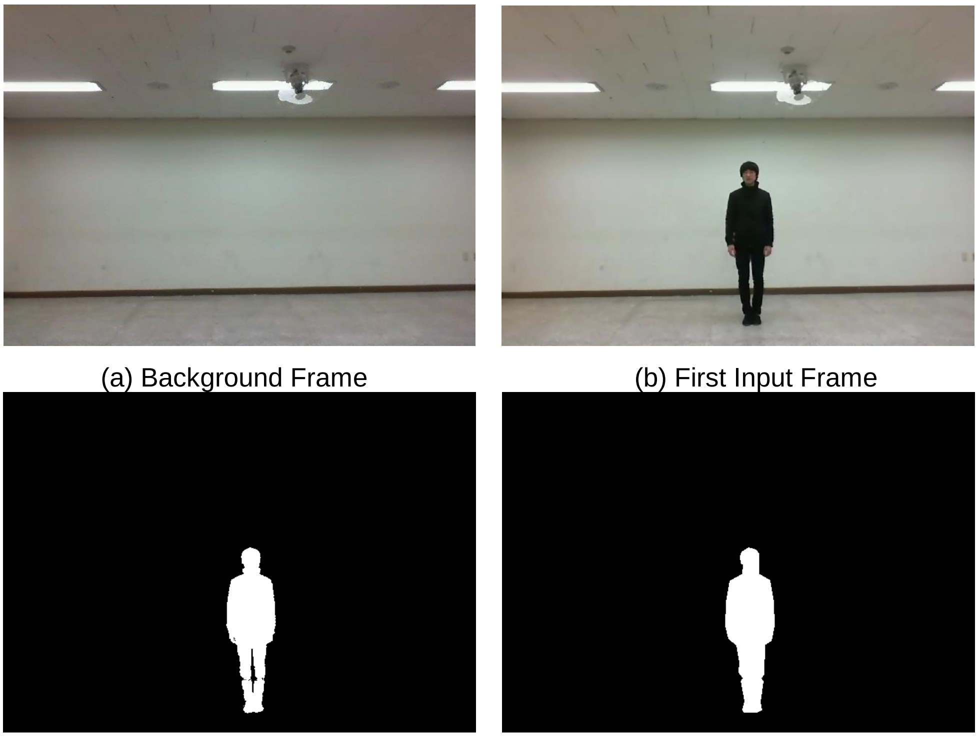


Figure 5. Test Images

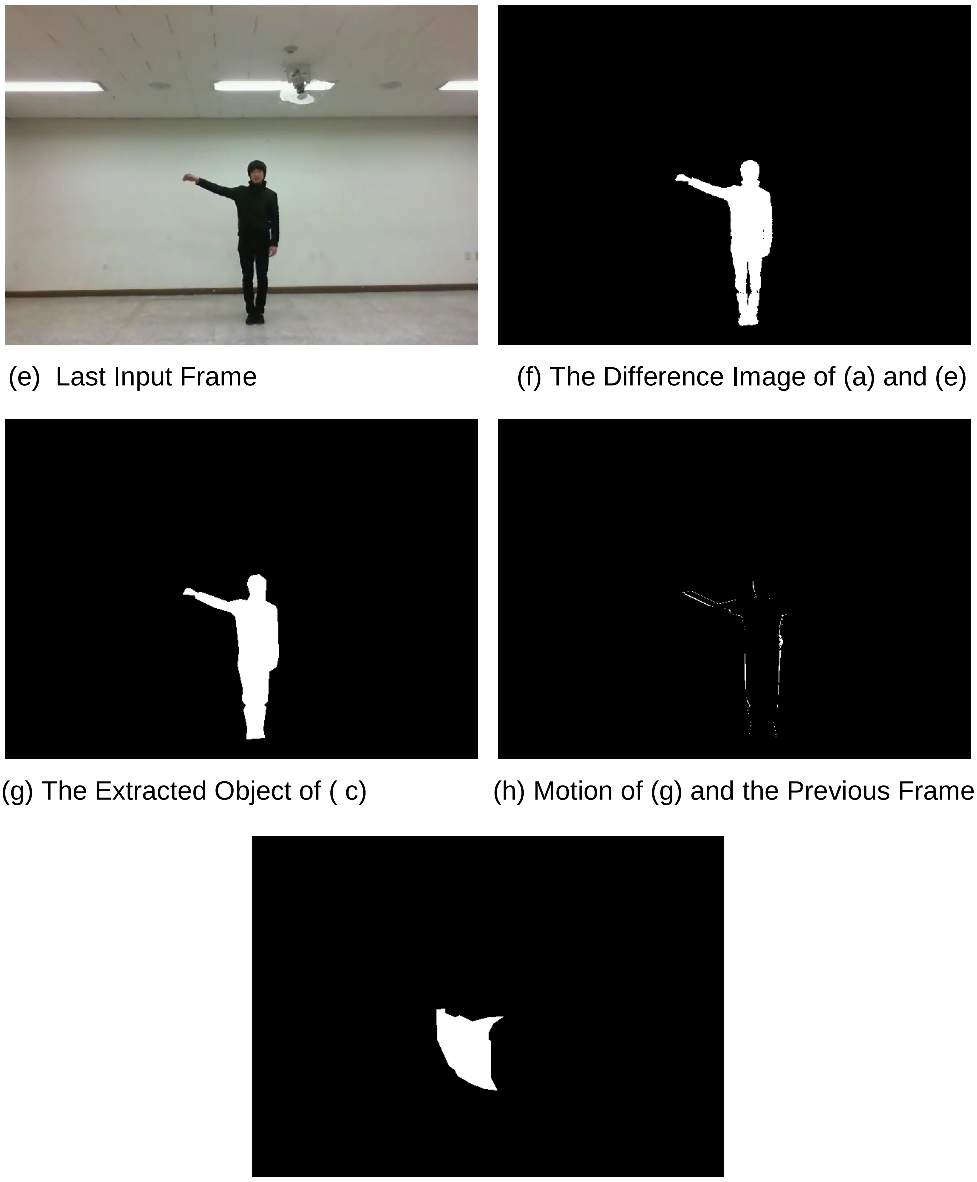


(a) Background Frame

(b) First Input Frame

(c) The Difference Image of (a) and (b)

(d) The Extracted Object of (c)



(e) Last Input Frame

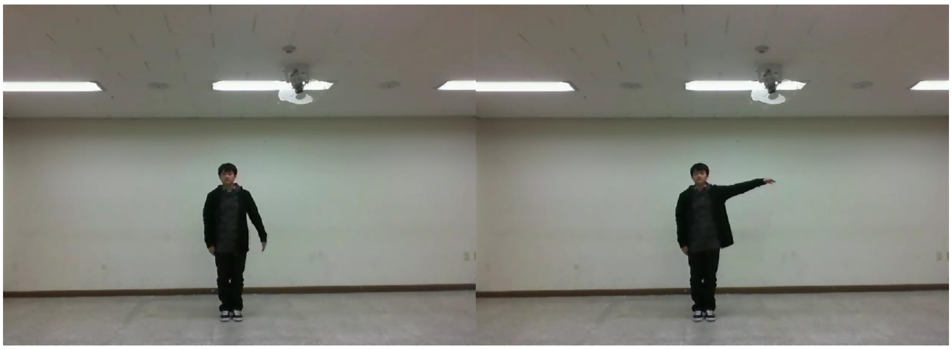
(f) The Difference Image of (a) and (e)

(g) The Extracted Object of (c)

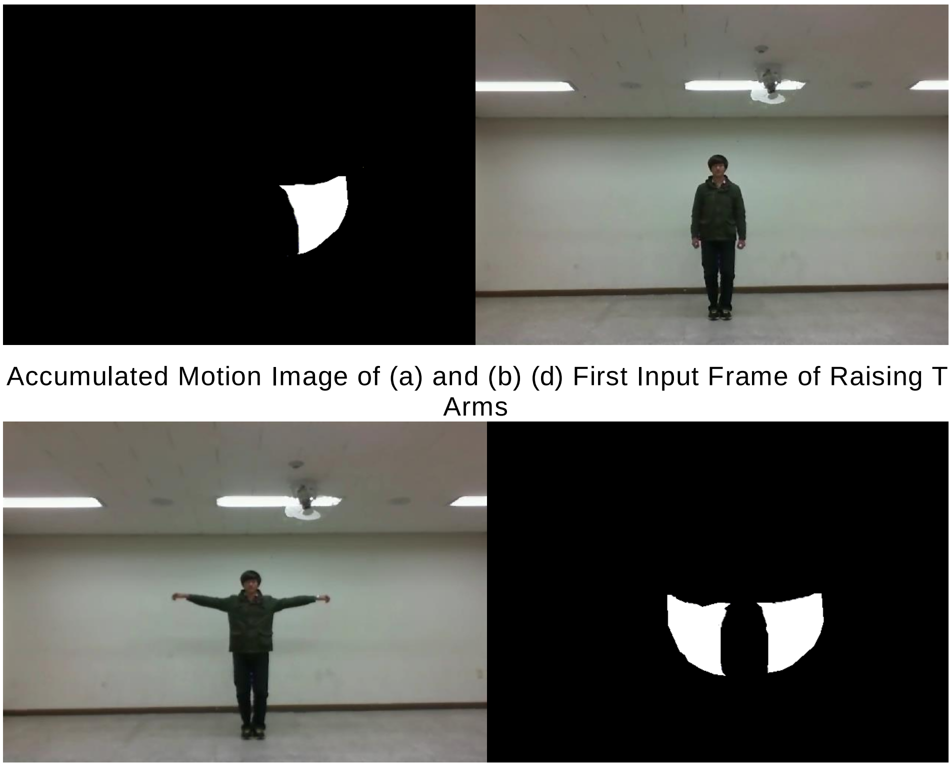
(h) Motion of (g) and the Previous Frame

(i) Accumulated Motion Image

Figure 6. Result Images by the Proposed Method



(a) First Input Frame of Raising Left Arm (b) Last Input Frame of Raising Left Arm



(c) Accumulated Motion Image of (a) and (b) (d) First Input Frame of Raising Two Arms

(e) Last Input Frame of Raising Two Arms (f) Accumulated Motion Image of (d) and (e)

Figure 7. Result Images of Applying the Proposed Method to Raising Left Arm and Two Arms

Figure 7 shows other results to apply the proposed method to an experimental video. (a) is initial input image of raising left arm and (b) is the last video frame of raising right arm. (d) is initial input image of raising two arms. (e) is the last video frame of raising two arms. (c) and (f) are the accumulated moving information image of raising left arm and two arms. The results of (c) and (f) show the accumulated moving information extracted from the movements of left arm and both arms that according to table 1 is classified as action 3 and action 4.

Table 2 shows the result of human action recognition obtained from each 300 frame of 4 people using the proposed method. The proposed method is used to obtain accumulated moving information data using the first 5~10 frames. In the case of 'Walking', results show 100% recognition rate because recognition process doesn’t need the accumulated moving information data. However, the other 7 different type of actions mark a 97~98% recognition rate due to the fact that it requires accumulated moving information extracted from 5~10 frames.

Table 2. The Result of Human Action Recognition

Action	1	2	3	4	Recognition ratio
walking	300	300	300	300	100%
Raising right hand	293	295	294	294	98%
Raising left hand	290	290	290	290	97%
Raising two rams	293	295	294	294	98%

Raising right leg	293	295	294	294	98%
Raising left leg	293	295	294	294	98%
sitting	293	295	294	294	98%
crouch	290	291	290	291	97%

4. Conclusion

This paper proposes an algorithm that recognizes and classifies human actions. The proposed method models background, extracts main object using differential images from input video and modeled background, and recognizes the human action using accumulated moving information of object. The human actions recognized in this study amount to a total of 8 type of actions, which include walking, raising an arm (left, right), raising a leg (left, right), sitting and crouching. Moving regions here have been classified into 15 different regions with focus on the movements of each body part. Human action classification is based on the principle that the difference between two images represents movement of that object. Accumulated moving information images are obtained by adding differential images and are divided into 15 different regions to detect moving region, which is used to recognize human action. The proposed method was experimented using video input of a webcam and it showed good results in terms of detecting and tracking for the purpose of recognizing human action.

The proposed method doesn't require learning or complex algorithms and it can be applied to different applications such as U-health, surveillance, and others. However, 5~10 frames are used to obtain accumulated moving information images, which is a loss, and therefore needs further research for improvement.

Acknowledgements

This study was financially supported by the research year fund of Mokwon University in 2013.

References

[1] R. T. Collins, Afujiyoshi D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, P. Burt and L. Wixson, “A System for Video Surveillance and Monitoring”, Technical Report CMU-RI-TR-00-12, Carnegi Mellin University, **(2000)**.

[2] http://www.ekahau.com/real-time-location-system/blog/2014/08/13/the-internet-of-things-iot-and-indoor-location-tracking-provide-benefits-to-hospitals/, “The Internet of Things (IoT) and Indoor Location Tracking Provide Benefits to Hospitals”, **(2014)**.

[3] Kortuem G., Kawsar F., Fitton D. and Sundramoorthy V., “Smart objects as building blocks for the Internet of things”, IEEE Internet Computing, vol. 14, **(2009)**, pp. 44-51.

[4] G. Gasser, N. Bird, O. Masoud and N. Papanikolopoulos, “Human Activities Monitoring at Bus Stops”, Proceedings of the IEEE International Conference on Robotics & Automation, **(2004)**, pp. 90-95.

[5] J. Tao and Y. P. Tan, “A Probabilistic Reasoning Approach to Closed-Room People Monitoring”, IEEE ISCAS, **(2004)**, pp. II-185-188.

[6] Park J., Tabb A., Kak A., A. C, “Hierarchical Data Structure for Real-Time Background Subtraction”, IEEE International Conf. on Image Processing, **(2006)**.

[7] Aggarwal J. K. and Cai Q., “Human Motion Analysis: A Review”, Computer Vision and Image Understanding, vol. 73, **(1999)**, pp. 428-440.

[8] R. T. Collins, A. J. Lipton, T. Kanade, H. Fujiyoshi, D. Duggins, Y. Tsin, D. Tolliver, N. Enomoto, O. Hasegawa, Urtsasun R., Fleet D. J. and Fua P., “Monocular 3-D tracking of the Golf Swing”, IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, **(2005)**, pp. 932-938.

[9] Ramanan D. and Forsyth D. A., “Finding and Tracking people from the bottom up”, in Proceedings of Computer Vision and Pattern Recognition (CVPR), Madison, Wisconsin, vol. 2, **(2003)**, pp. 467-474.

[10] A. Yilmaz and M. Shah, “Actions Sketch: A Novel Action Representation”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 1, **(2005)**, pp. 984-989.

[11] J. Tao and Y. P. Tan, "A Probabilistic Reasoning Approach to Closed-Room People Monitoring", IEEE ISCAS, **(2004)**, pp. II-185-188.

[12] T. E. de Campos and D. W. Murray, "Regression-based Hand Pose Estimation from Multiple Cameras CVPR", vol. 1, **(2006)**, pp. 782-789.

[13] S. Iwasawa, J. Ohya, K. Takahashi, T. Sakaguchi, S. Kawato, K. Ebihara and S. Morishima, "Real-time 3D extimation of human body postures from triocular images", Proceeding of Workshop on modeling people, **(1999)**, pp. 3-10.

[14] Q. Delamarre and O. Faugeras, "3D articulated models and multi-view tracking with silhouettes", Proc. ICCV, **(1999)**, pp. 716-721.

[15] N. J. Kwak and T. S. Song, "Human Action Recognition Using Segmentation of Accumulated Movement", Proc. SMA, **(2014)**, pp. 52.

[16] K. Kim, T. H. Chalidabhongse, D. Harwood and L. Davis, "Real-time foreground-background segmentation using codebook model", Real-time Imaging, vol. 11, **(2005)**, pp. 167-256.

Authors

Nae-Joung Kwak, She received the B.S. in February 1993 and M.S. in February 1995, PhD in February 2005 from the department of Computer and Communication engineering, Chungbuk National University. She is currently teaches at Mokwon University, Hanbat University, Chungnam National University in Korea. Her research interests include multimedia communication, multimedia signal processing, video surveillance system, and MPEG. She is a member of the Korea Information Science Society, The Korea Contents Association, and the Institute of Electronic Engineers of Korea

Teuk-Seob Song, He received his PhD degree in Computer Science in 2006 and PhD degree in Mathematics from Yonsei University in 2001, respectively. He is currently an assistant professor in the Department of Computer Engineering at Mokwon University in Korea. His research interests include 3D Virtual Environment, Web3D, Annotation Technology and Structured Document Transcoding. He is a member of the Korea Information Science Society, the Korea Information Processing Society, and the Korea Multimedia Society