Android-based Human Action Recognition Alarm Service using Action Recognition Parameter and Decision Tree

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

[1]*Genetel systems.Inc, Dae-jeon, Korea*
[2] *Mokwon University, Doandong, Seogu, Dae-jeon, Korea*

knj0125@hanmail.net, teuksob@mokwon.ac.kr

Abstract

This paper proposes a method to analyze human actions and provide alarm service through a smart-phone. The proposed method extracts objects using camera input images and modeling background. This paper defines a action recognition parameter as an extracted object' histogram and its joints' information about human action recognition and analyzes changes in defined parameter to determine conditions for recognizing human actions. Furthermore, in order to systemically classify actions, the proposed method designs a decision tree and combines the defined conditions with a conditional expression of tree node, which enables effective action recognition and classification. Recognized actions are transferred to a smart phone as a simple value and the transferred result is displayed on the phone.

The proposed method is tested from a video of a webcam about nine actions of four persons. As for action recognition and alarm service of a smart phone, the proposed method works well and effectively.

Keywords: *silhouette, object tracking, auto detection, joint, mobile device*

1. Introduction

Switching from analog technology to digital, a video surveillance system has integrated IP based Network technology such as computers and software storage products as well as device technology and developed together with them. Recently, it has many applications such as mechanics, robots, games and toys, crime prevention, home network, parking management, and facilities of schools, the handicapped and seniors [1]. As its application has expanded, the surveillance system has expanded into "smart visual surveillance system" and drawn much attention. A smart visual surveillance system is defined as a system to analyze information of images from a camera in real time and recognize objects through object extraction, tracing and image analysis. By determining and analyzing whether an object causes an event relevant to physical surveillance policy, this technology provides administrators with information or stores data and events to maximize effectiveness of preventing and searching the similar occasions. Since it analyzes patterns of human actions and collects a variety of information, it can be applied to various fields [2, 3]. Likewise, classification of human's actions can make by detecting and analyzing how body parts change with the passage of time.

The method [3-7] extracting target objects from input images of a video camera has been studied and developed in terms of image recognition. However, the existing methods need to

Correspondence should be addressed to T.S. Song, teukseob@mokwon.ac.kr

learn specific features of body information to recognize or estimate actions. Therefore, they need a lot of learning data and complex learning algorithm.

This paper proposes a new method to recognize actions after image data analysis based on collected information and to transfer the result to a smart phone for alarm service. This method extracts an object and joints of a body from input images of a camera. Actions are recognized through following four steps: 1) analyzing histogram of an object and movement of joints, 2) setting action recognition parameter 3) defining conditions for recognition and 4) making a decision tree. Recognized actions are transferred to a smart phone as a simple value and the phone restores the value and lets users know the movement of the target object.

Chapter 2 describes about object extraction, joints detection, action recognition, transfer of recognized actions and smart-phone's alarm service. Chapter 3 evaluates the presented method and chapter 4 makes a conclusion.

2. Human Action Recognition and Alarm Service using Decision Tree

The proposed method extracts an object from input images of a single camera by using background image. Then, it extracts its joints automatically, tracks the extracted joints related to action patterns, and analyzes direction of motion vector to recognize human actions. Figure 1 is a flowchart of the presented method.

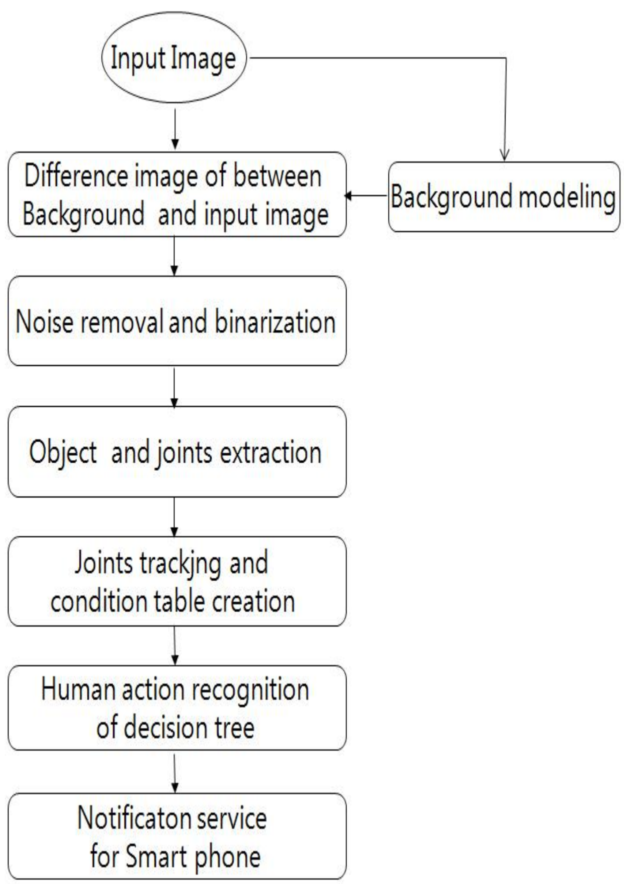


Figure 1. The Flowchart of the Proposed Algorithm

2.1. Extracting Background Modeling and Objects

Extracting objects, which preprocesses for extracting body joints, is very important to decide the precision of the next step. Also, to separate object from background is needed for effectively extracting objects. The proposed method separates objects from background by using threshold value after modeling background. Background modeling and image binarization uses the method of the proposed [8]. If a pixel value is greater than a certain value, the method consists of designing a new colors set of background and makes color models as box- shape in a 3-dimensional color space. The color models can use color space such as RGB, YUV, HSV and the proposed method used RGB color space.

Firstly, the proposed method gets difference image between modeling background and each input frame. The difference image is an image where background and objects are separated and if an input image has pixel value between lower and upper threshold value, it is considered "Background." The alternative is considered "Object." The following is the formula for extracting "Object" from "Background."

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

Where, is the coordinates of an input image, and "Object" and "Background" are binarized into white (255) and black (0), respectively. The final image is produced by combining the three binary planes.

The binary final image has noise and the small areas except for "Object" area. Therefore, the proposed method removes the areas and noise with morphological filter and use threshold value of area size to remove everything but an object, which makes an "Object and Background separation" image $B(x,y)$.

2.2. Extracting Silhouette and Joints

The proposed method extracts silhouette and joints from an extracted object with the proposed [9]. After tracking contours with a contour tracking system offered by OpenCV, the proposed method defines a body model and its sixteen joints of [9].

Joints are extracted by using proportion of each body part of [9] based on width and height of a face. The length of face is produced by OpenCV's face detector. The measurements of a body are different from person to person and so it is very difficult to detect exact location of joints. Therefore, the proposed method extracts initial joints by using ratio of body measurements. Then, corner points are computed using Harrison corner detector from body silhouette in order to complement the location of joints of each person and final joints are extracted by choosing the closest corner points to initial joints.

2.3. Action Pattern Recognition using Action Recognition Parameters and Decision Tree

Among various actions, the proposed method recognizes nine ones such as sitting on a chair, raising hand (right and left) and both hands, lifting one leg (right and left), crouching, walking horizontally and walking forward. On the nine actions, "Horizontal Walking" means that an object moves horizontally. "Horizontal Walking" is detected first with an object's histogram before detecting joints while as for the other eight, joints are detected and then classified based on whether there is movement of joints and whether there is a change in X and Y coordinate.

2.3.1. Parameters to Detect Actions: The proposed method defines parameters as histogram projection on X-coordinate of detected objects and six essential joints among sixteen joints.

Of the nine actions, “Horizontal Walking” is detected by using histogram projection on X-coordinate. Figure 2 is the result of histogram of “Horizontal Walking” and “Forward Walking” after projecting objects on X-coordinate. The histogram of “Forward Walking” shows near symmetry with respect to the center red line while that of “Horizontal Walking” does not. Thus, when histogram projection on X-axis is asymmetric, it is categorized into “Horizontal Walking.”

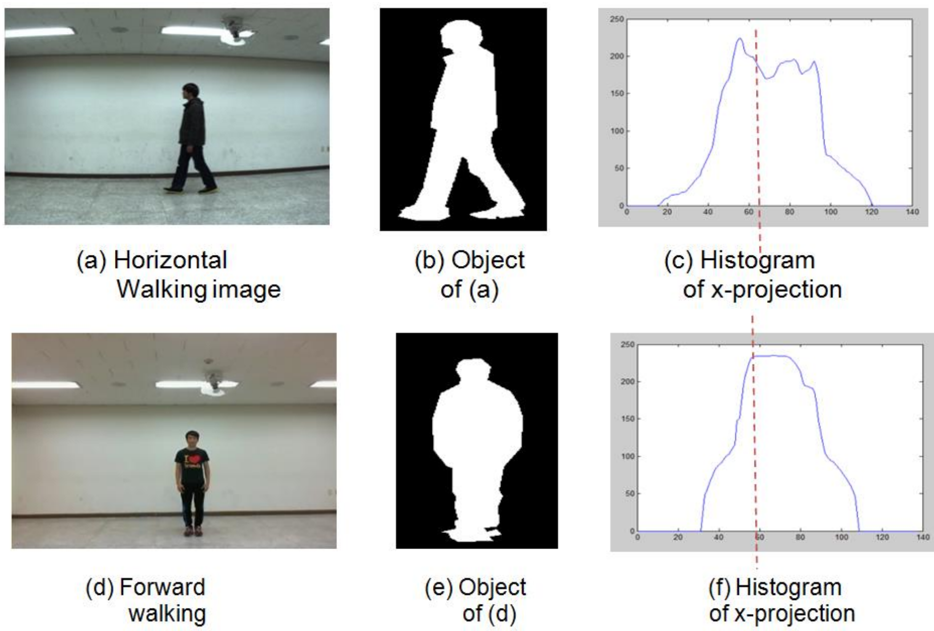


Figure 2. Histogram of x-horizontal Projection

After classifying “Horizontal Walking”, the other eight actions are classified based on whether there is movement of joints. It requires only a few of sixteen joints to detect each action and their movement and direction are different. If you raise left hand, at Joint 10, X-coordinate value increases and Y-coordinate one decreases. In terms of “Sitting on chair”, Y values of Joint 5, 6, 7 and 8 can change while X values of Joint 13 and 14 can change. That is, there are some joints whose values change along with changed motion. Furthermore, sometimes, either of X or Y value changes and sometimes both of them change. Therefore, parameter to detect actions can be defined based on coordinates of joints which are essentially moved by each motion (essential joint). The proposed method defines six essential joints among sixteen. There are four categories as for action recognition: When there is no change in both X and Y coordinates at each joint, it is defined as 0, a change in x refers to 1, a change in Y to 2, a change in both X and Y to 3, and no significant effect on action recognition despite motion to *. According to this rule, Table 1 shows the reference for classification of human action recognition.

Table 1. Classification Reference for Human Action Recognition

	Histogram	Joints						
		5	9	10	13	14	15	16
Horizontal walking	3	0	0	0	0	0	0	0
Forward walking	0	1	*	*	*	*	1 or 2 or 3	1 or 2 or 3
Raising right arm	0	0	3	0	0	0	0	0
Raising left arm	0	0	0	3	0	0	0	0
Raising two arms	0	0	3	3	0	0	0	0
Raising right leg	0	0	0	0	3	0	3	0
Raising left leg	0	0	0	0	0	3	0	3
Sitting	0	2	*	*	1	1	0	0
Crouch	0	2	*	*	2 or 3	2 or 3	0	0

The proposed method uses block matching to detect movement of joints and changes in X and Y coordinates. Search block for tracking movement is 20x20 for each joint and search window is ±5 vertically and horizontally

2.3.2. Human Action Decision Tree Depending on Decision Condition: Figure 3 shows the decision tree based on reference of Table 1. ‘Y’ is the case of satisfaction of condition C.X while ‘N’ is the case of the alternative. C.X is a x-th condition for classification shown in Table 2.

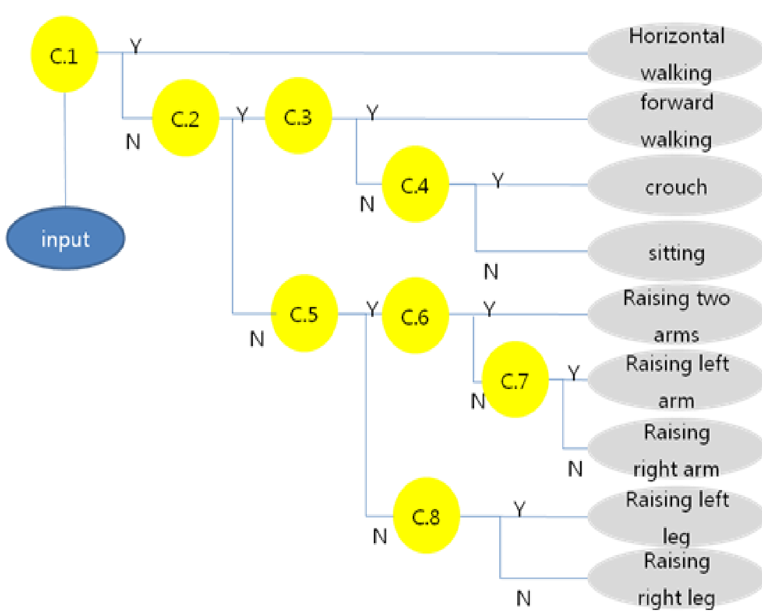


Figure 3. The Proposed Decision Tree

Table 2. Condition Number C.X

Condition No.	Condition for classification	Condition No.	Condition for classification
C.1	Histogram's value is '3'	C.5	Joint '9' or '10' 's value is not '0'
C.2	Joint 5's value is not '0'	C.6	Joint '9' and '10' 's values are '3'
C.3	Joint 5's value is '1'	C.7	Joint '10' 's value is '3'
C.4	Joint 13's value is '2' or '3'	C.8	Joint '13''s value is '3'

Table 3. Human Action and Transmission Number

Human action	Horizontal walking	Forward walking	Raising right arm	Rasing left arm	Rasing two arms	Lifting right leg	Lifting left leg	Sitting	Crouch
Transmissio n number	0	1	2	3	4	5	6	7	8

2.4. Transferring Information on Human Action

The information of human actions classified by a decision tree can be provided through a smart phone with alarm. Users can check the result of action recognition by pushing a button. The nine actions are transferred to a smart phone as a number from zero to eight and the phone displays it as a string related to the actions. Table 3 shows human's actions and the transferred numbers.

3. Experiment and Result Analysis

The proposed method is evaluated by using background and input images with a camera in real time. Test images have nine actions of four persons at background of two different indoor. Intel CPU 2.0GHz, 1G RAM, Visual Studio 2008 and OpenCV 2.1 are used. The resolution of the image is 640x480 24bit and at 15 frames per second. To confirm the result transferred to smart-phone, it is programmed to operate in HVGA skin, SD card 32Mbyte of Eclipse and Android-based Google APIs level 8.

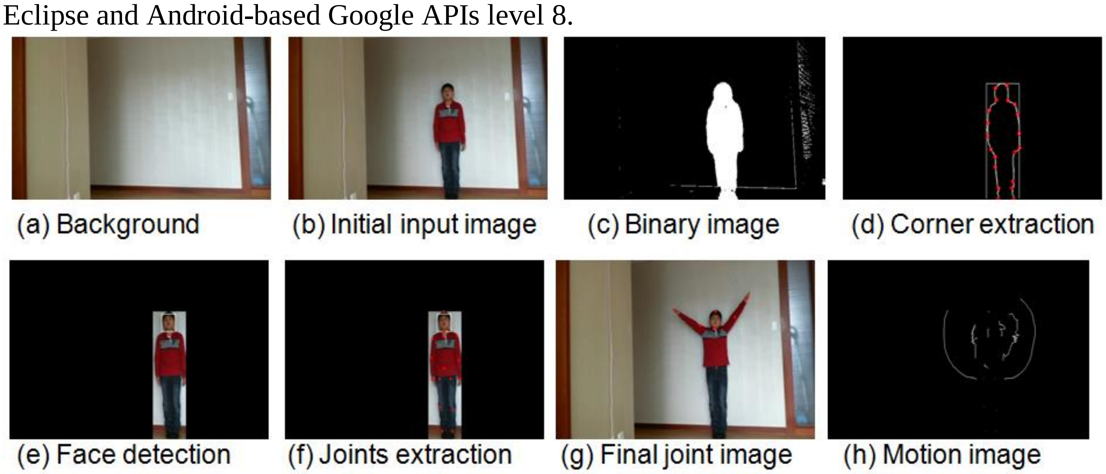
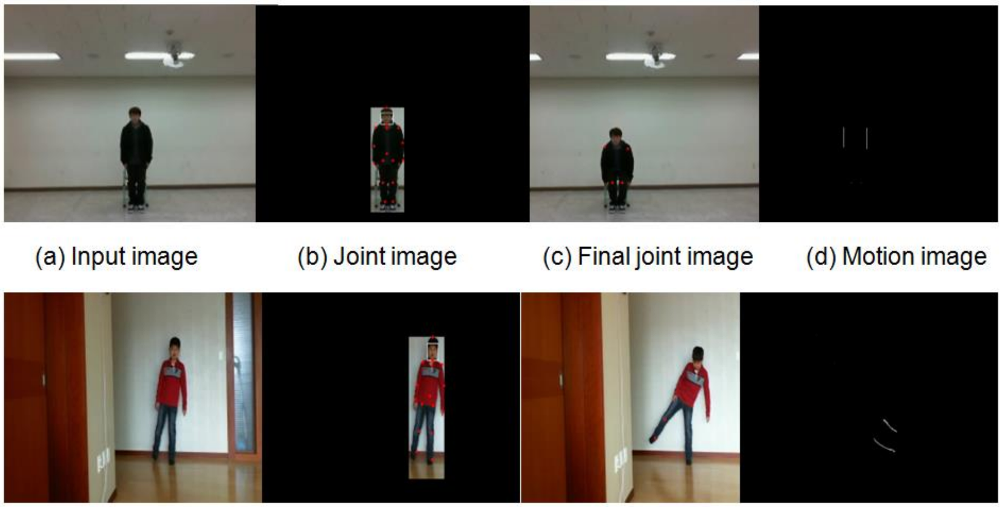


Figure 3. Each Stage Image by the Proposed Method



(a) Input image (b) Joint image (c) Final joint image (d) Motion image

(e) Input image (f) Joint image (g) Final joint Image (h) Motion Image

Figure 4. The Results of Activity Recognition (Sitting and Lifting a Leg)

Figure 3 shows each step of the experiment. (a) is the model of the background and (b) is an initial image. (c) is a binary image of (b) by the proposed method. (d) is the result to extract its silhouette and corner points after extracting a main object from the image (c). (e) shows extracted face and (f) is extracted joints. (g) shows the result of face's joints and joints of movement area(both hands) after tracking the motion of raising both hands. (h) is the result of tracking movement of each joint from initial image (b) to final joint tracking image (g). In the (h), active movement of both hands and arms can be seen, compared with other joints.

Figure 4 shows the result of "Sitting on chair" and "Lifting one leg." It shows that the proposed method traces well changing joints of each action as well as a change in motion vector.

Figure 5 shows the displayed result of recognized actions of Figure 4 on a smart phone. The transferred information is the number of recognized action. For "Sitting on chair", 7 is sent and for "Lifting Right Leg", 5. These numbers are translated into a string to be displayed on a smart phone.

(a) Result of Figure 4(c) (b) Result of Figure 4(g)

Figure 5. The Result of Smart Phone of Human Recognition by the Proposed Method

Table 4 is the result of analyzing recognition rate of each action. Each input image use 300 frames. The proposed method recognized every action but it has some partial errors about "Lifting One Leg". This error can be generated because of shades surrounding legs.

Table 4. Recognition Ratio of Human Action by the Proposed Method

Action	1	2	3	4	Recognition ratio
Horizontal walking	300	300	300	300	100%
Forward walking	300	300	300	300	100%
Raising right hand	300	300	300	300	100%
Raising left hand	300	300	300	300	100%
Raising two rams	300	300	300	300	100%
Lifting right leg	295	295	295	294	98%
Lifting left leg	293	295	294	294	98%
sitting	300	300	300	300	100%
crouch	300	300	300	300	100%

4. Conclusion

This paper presents an algorithm recognizing human actions by using histogram of an object and information of body joints. From a video of a webcam, objects are extracted and then joints of the objects are detected. By analyzing extracted object’ s histogram and the movement of extracted joints, the proposed method finds parameters to recognize human actions and defines the conditions of classifying nodes of a decision tree based on the result.

Actions recognized in accordance with defined conditions are stored as a result value of a decision tree and can be transferred to a smart phone when users require. The proposed method recognizes nine defined actions and gets good results in terms of action recognition and alarm service.

The proposed method can have various applications such as U-health and surveillance camera in a ubiquitous environment. However, the shades by lights can be yield to lower the recognition rate and so further research is required to solve the problem. In addition, there are needed further researches about classification of actions of multiple objects in a complex background.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology(2012-0002852).

References

[1] R. T. Collins, A. Fujiyoshi, 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] G. Gasser, N. Bird, O. Masoud and N. Papanikolopoulos, “Human Activities Monotoring at Bus Stops”, Proceedings of the IEEE International Conference on Robotics & Automation, **(2004)**.

[3] Ji Tao and Yap-Peng Tan, “A Probabilistic Reasoning Approach to Closed-Room People Monitoring”, IEEE ISCAS, **(2004)**.

[4] D. H. Wilson, A. C.Long and C. Atkeson, “A Context-Aware Recognition Survey for Data Collection Using Ubiquitous Sensors in the Home”, Proceedings of CHI 2005, Late Breaking Results, Portland, **(2005)**.

[5] E. Murphy-Chutorian and M. Trivedi, “Head Pose Estimation in Computer Vision: A survey”, IEEE Trans. on Pattern Anaysis and Machine Intelligence, vol. 4, no. 31, **(2009)**.

[6] T. Horptasert, I. Haritaoglu, C. Wren, D. Harwood, L. Davis and A. Pentland, “Real time 3D motion capture”, Processings of Workshop on perceptual user interface, **(1998)**.

[7] Pengfei Zhu and Paul M.Chrlian, “On Critical Point Detection of Digital Shapes”, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 8, no. 17, **(1995)**.
[8] K. Kim, T. H. Chalidabhongse, D. Harwood and L. Davis, “Real-Time Foreground-Background Segmentation using Codebook Model. Real-time Imaging 11, **(2005)**.
[9] N. J. Kwak and T. S. Song, “Automatic Detecting and Tracking Algorithm of Joint of Human Body using Human Ratio”, JKCA, vol. 4, no. 1, **(2011)**.

Authors

Nae-Joung Kwak received her B.S. in February 1993 and M.S. in February 1995, Ph.D in February 2005 from the department of Computer and Communication engineering, Chungbuk National University. She was a contract professor at Mokwon University in korea from 2005 to 2008. She has taught at Mokwon University, Hanbat University, Chungnam National University, and Chungbuk National University in Korea since 2009. She currently works as a senior researcher at Genetel Systems. Inc, in Korea. Also she is a visiting researcher of Electronic and Telecommunication Research Institute 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 received his Ph.D. degree in computer science in 2006 and Ph.D. degree in Mathematics from Yonsei University in 2001, respectively. He is currently an associate professor in the Department of Computer Science at Mokwon University in Korea. His research interests include 3D Virtual Environment, Semantic Navigation, XML Document, 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.