Human Action Classification Using Movement of Joints and Decision Tree*

Nae Joung Kwak[1], 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 classify human actions. The proposed method extracts objects, finds joints' information from the objects, tracts the movement of joints and classifies human action using changes in moving information of joints and decision tree. The proposed method designs a decision tree and combines the defined conditions with a conditional expression of tree node, which enables effective action classification. The proposed method is tested from a video of a webcam about six actions of four persons. As for action classification, the proposed method works well and effectively.

Keywords: silhouette, object tracking, auto detection, joint, action classification

1 Introduction

A smart visual surveillance system is defined as a system to classify information of images from a camera in real time and classify objects through object extraction, tracing and image analysis[1]. Since it classifies patterns of human actions and collects a variety of information, it can be applied to various fields [2]. Likewise, classification of human's actions can make by detecting and analyzing how body parts change with the passage of time.

The method [3,4] extracting target objects from input images of a video camera has been studied and developed in terms of image classification. However, the existing methods need to learn specific features of body information to classify or estimate actions. Therefore, they need a lot of learning data and complex learning algorithm. This paper proposes a new method to classify actions after image data analysis based on collected information. The method extracts an object and joints of a body from input images of a camera. Actions are classified through following three steps: 1) analyzing movement of joints, 2) setting action classification 3) making a decision tree.

Chapter 2 describes about object extraction, joints detection, action classification. Chapter 3 evaluates the presented method and chapter 4 makes a conclusion.

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).
Correspondence should be address to Teuk-Seob Song, teuksob@mokwon.ac.kr

2. Human Action Classification Using Movement of Joints and Decision Tree

2.1 Objects Extraction and joints detection

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 classifies direction of motion vector to classify human actions.

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 [5] in 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(black)." The alternative is considered "Object (white)." 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 ($B(x,y)$).

The proposed method extracts silhouette and joints from an extracted object with the proposed [6]. After tracking contours with a contour tracking system offered by OpenCV, the proposed method defines a body model and its sixteen joints of [6].

2.2 Action Classification Using Movement of Joints and Decision Tree

Among various actions, the proposed method classifies six ones such as sitting, raising hand (right and left) and both hands, lifting one leg (right and left). Above all, joints are detected and then classified based on whether there is movement of joints and whether there is a change in X and Y coordinates.

Six 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", 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 classification: 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 classification despite motion to *.

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.

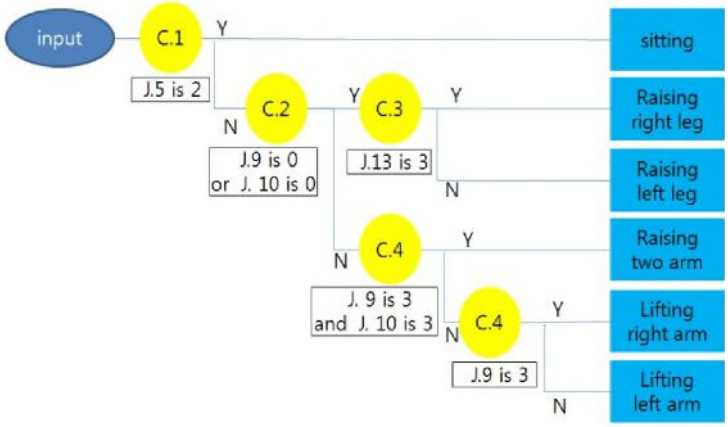
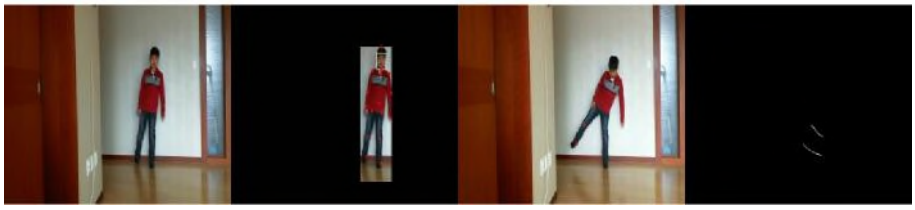


Fig. 1. The classification process of the proposed decision tree

Figure 1 shows the classification process of the proposed decision tree. C.X is a x-th condition for classification and the condition in a box below C.X is the reference for action classification. In the sentence " J.N is k", N is the number of joint and k is one among 0 ~ 3 defined above as change value in X and Y coordinates at each joint. 'Y' is the case of satisfaction of condition C.X while 'N' is the case of the alternative.

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 six 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.



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

Fig. 2. The results of activity classification (lifting a leg)

Figure 2 shows each step of the experiment. (a) is the original image (b) show the result of extracted face and extracted joints(c) shows the result of joints of movement

area of right leg after tracking the motion of lifting right leg. (d) is the result of tracking movement of each joint from initial image (a) to final joint tracking image (c).

4. Conclusion

This paper presents an algorithm recognizing human actions using information of body joints. From a video of a webcam, objects are extracted and then joints of the objects are detected. By analyzing extracted the movement of extracted joints, the proposed method finds parameters to classify human actions and defines the conditions of classifying nodes of a decision tree based on the result.
Actions classified in accordance with defined conditions are stored as a result value of a decision tree. The proposed method classifies six defined actions and gets good results in terms of action classification.

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. G. Gasser, N. Bird, O. Masoud, and N. Papanikolopoulos : Human Activities Monitoring at Bus Stops.: Proceddings of the IEEE International Conference on Robotics & Automation, pp. 90-95 (2004).
3. D.H. Wilson, A.C.Long, and C. Atkeson : A Context-Aware Classification Survey for Data Collection Using Ubiquitous Sensors in the Home. : In Proceedings of CHI 2005: Late Breaking Results, pp. 1865-1868, Portland (2005).
4. E. Murphy-Chutorian and M. Trivedi: Head Pose Estimation in Computer Vision: A survey. : IEEE Trans. on Pattern Anaysis and Machine Intelligence, vol. 31, no. 4, pp. 607626 (2009).
5. K.Kim, T.H. Chalidabhongse, D.Harwood, and L. Davis : Real-Time Foreground-Background Segmentation using Codebook Model.: Real-time Imaging 11, pp.167-256 (2005).
6. N.J.Kwak, T.S.Song: Joint Tracking and Transmission System for Simulating Motion of the Human Body on Android Smart Phone.: IJCA, vol. 4, no. 4, pp. 81-89 (2011)