Human Activity Recognition Based on Morphological Dilation followed by Watershed Transformation Method

Muhammad Hameed Siddiqi
Ubiquitous Computing Lab, Kyung Hee University, Korea
siddiqi@oslab.khu.ac.kr

Muhammad Fahim
Ubiquitous Computing Lab, Kyung Hee University, Korea
fahim@oslab.khu.ac.kr

Sungyoung Lee
Ubiquitous Computing Lab, Kyung Hee University, Korea
sylee@oslab.khu.ac.kr

Young-Koo Lee
Ubiquitous Computing Lab, Kyung Hee University, Korea
yklee@khu.ac.kr

Abstract—**Efficiency and accuracy are the most important terms for human activity recognition. Most of the existing works have the problem of speed. This paper proposed an efficient algorithm to recognize the activities of the human. There are three stages of this paper, segmentation, feature extraction and recognition. In this paper our contribution is in segmentation stage (based on morphological dilation) and in feature extraction stage (using watershed transformation). The proposed algorithm has been tested on six different types of activities (containing 420 frames). The recognition performance of our method has been compared with the existing method using Principle Component Analysis (PCA) to derive activity features. The results of our proposed method are comparable with the existing work. But in-term of efficiency, our algorithm was much faster than the existing work. The average accuracy and efficiency of the proposed algorithm for recognition was 80.83 % and 302.2 ms respectively.**

Keywords-component; Human activity recognition;computer vision; morphological dilation; watershedding; image segmentation.

I. INTRODUCTION

Activity recognition means the analysis of motions and behaviors of a human from low-level sensors. Automatic human activity analysis is an increasingly active research area of pattern recognition and in computer vision community and is considered an important problem in the field of the computer vision and pattern recognition. Generally there are two types of activities that a human is performing commonly - Low-level activities, which is also called micro-activities (μ activities) e.g. sitting, standing, walking, running, etc. - High-level activities, which is also called macro-activities e.g. watching TV, playing tennis, ridding bus, etc. The general architecture of high-level activities or macro activities is shown in Fig.1.

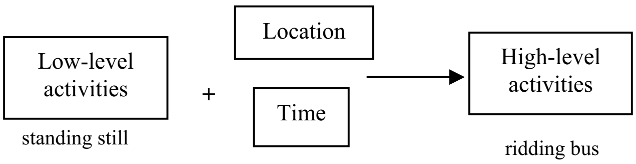

Fig.1. general architectural model of high-level (macro) activities

Some of the existing pattern recognition methods that were used for activity recognition can only utilize labeled activity samples for patterns, although usually large amounts of unlabeled samples exist as they do not need human's labeling effort [1].

In video the analysis of human activities is a monotonically increasingly important research area from surveillance, security, content-based video retrieval, animation and synthesis. There are several challenges at various levels of processing that are hard to solve, vigorousness against errors at low-level processing, view and rate-constant representation at mid-level processing, and semantic representation of human activities at higher-level processing. The author of [2] presented a comprehensive survey to address the abovementioned problems of representation, recognition, and learning of human from video and related applications. In [3] the authors have presented a novel method to address the activity recognition problem. They represented the activities by feature vectors from Independent Component Analysis (ICA) on video frames, and then based on these features they recognized the human activities by using Hidden Markov Model (HMM) classifier.

Many pattern recognition methods have been proposed for the human activity recognition. The important source or origin for activity recognition is the shape information of the frames in the video. The author of [4] proposed an algorithm, which used Independent Component Analysis (ICA) for feature extraction and based on these features, the activity of different types of human were recognized by trained Hidden Markov Models (HMMs).

The objective of this paper was to improve the efficiency and accuracy of recognition of the human activities. The proposed algorithm has been tested on six different types of activities, which gave a reliable accuracy and efficiency of activity recognition

The rest of the paper is organized as: Section II presents some of related work on human activity recognition. Section III presents methodology of the algorithm. The results are discussed in Section IV. And finally the conclusion and some of the future work are presented in Section V.

II. RELATED WORK

The two supervised hierarchical models for activity recognition based on motion words was presented by [5]. In their models, there were two aspects, first a 'visual word' was obtained from a large-scale descriptors from the whole frame, and second was a 'latent topics', which directly corresponded to the different categories of the human activities. P.C. Ribeiro et al [6] addressed the problem of recognizing the short-term human activities (Active, Inactive, Walking, Running, and Fighting) from a video sequence. This method basically focused on three fundamental issues: designing a human activity classifier, performing of feature selection and defining the structure of the classifier. They used, Bayesian for designing a classifier, Gaussian mixture for feature selection with the Bayesian classifier, and adequate to cope with complex data distribution [6].

Motion is the most important source information for the recognition of human activities. S. Ali et al [7] presented the utility of kinematic feature derived from motion information for the purpose of human activity recognition in videos, in which the computed the dominant kinematic modes from the optical flow by performing the Principal Component Analysis (PCA) on each kinematic features, and then nearest neighbor classifier used for the activity recognition.

Activity recognition is a hot topic in context-aware computing. The author of [8] proposed a semi-supervised algorithm for activity recognition. The segmentation process is needed for activity recognition. In [9] the author proposed an algorithm for moving object segmentation based on histogram. The algorithm consisted of YUV color model based on distance formula to differentiate the background from the foreground in moving objects. The author of [10] proposed morphological dilation algorithm for weed leaves category recognition. The simplest approach involves subtracting the current frame from the known background. In [11] the authors have presented a new algorithm that represents each pixel in the frame by a group of cluster, then the incoming pixels is matched against the corresponding cluster group and the backgrounds are classified according to the matched pixels with the clusters.

Some of the existing works [2, 4] have the problem of speed, because in their method they processed multi-dimensional features, which causes of low efficiency. The purpose of this paper was to develop an efficient algorithm for the human activity recognition that was not an autonomous robot system but to investigate a real time machine vision system. The algorithm was tested on six different types of human activities (Side, Jack, Run, Bend, Right hand Wave, and Skip) that gave a reliable accuracy and efficiency. The average accuracy and efficiency of the proposed algorithm for activity recognition is 80.83 % and 302.2 ms respectively. We used Matlab for implementation on Pentium4 desktop equipped with a 2.5 GHz Dual Core processor and 3 GB RAM.

III. MATERIALS AND METHOD

The proposed recognition algorithm consists of video image segmentation (Pre-processing), feature extraction (i.e. watershed transformation), and recognition via the addition of the pixels, which are selected randomly at the feature extraction stage. Fig.9. shows the whole process of the proposed method.

A. Video Image Segmentation (Pre-procesing)

In this stage the activity of the human is divided into number of frames and converted into image format. The color images of the video are converted from RGB to gray and then to binary for fast and easy processing. An image segmentation step is conducted to divide the image into two classes i.e. foreground (activity of the human) and background (empty frame).

Dilation is one of the basic operations in mathematical morphology. The simplest used of dilation is for eliminating irrelevant or hidden details (in term of size) from a binary image and make the object larger that is obtained by merging multiple objects. In dilation first the image has been expanded to binary and then to complete lattices that usually use a structuring element (Mask/Template/Kernel or Window) for expanding the shapes contained in the input image. The size of the structuring element must contain odd number of rows and odd number of columns, i.e. 3 x 3, 5 x 5 or 7 x 7 and so on. Because during the dilation process only the middle pixel interact with the corresponding pixel of the image. The interested pixels of the neighborhood are designed by the structuring element of the dilation that applied the appropriate rule to the pixels in the neighborhood and assigned a value to the corresponding pixel in the output. In dilation process the structuring element moved upon the corresponding image. When the value of the pixel in background is exactly equal to the value of the middle pixel in the structuring element, then the values of those pixels becomes to '1' and in this way the image becomes larger. The structuring element for the proposed algorithm is designed as:

$$\begin{pmatrix} \text{False} & \text{True} & \text{False} \\ \text{True} & \text{True} & \text{True} \\ \text{False} & \text{True} & \text{False} \end{pmatrix}$$

In the dilation process the binary image of the human activity class is dilated by a structuring element i.e. for dilating an image the above structuring element (filter/mask) of size 3x3 move pixel by pixel upon the given image, so in this way the unnecessary details have been removed from an image and the image became more enhanced for fast and easy processing. The general syntax for dilation is given in Eq.1.

$$I \oplus S = U_{s \in S} I_s \qquad (1)$$

The 'I' represents an image while 'S' is used for structuring element. The Graphical User Interface (GUI) was developed in Matlab that showed the original frame, processed frame and the results of the proposed algorithm. The frame resolution was 240 pixel rows by 320 pixel columns.

Fig.2. showed the complete concept of the dilation that has been used for enhancement and noise removing.

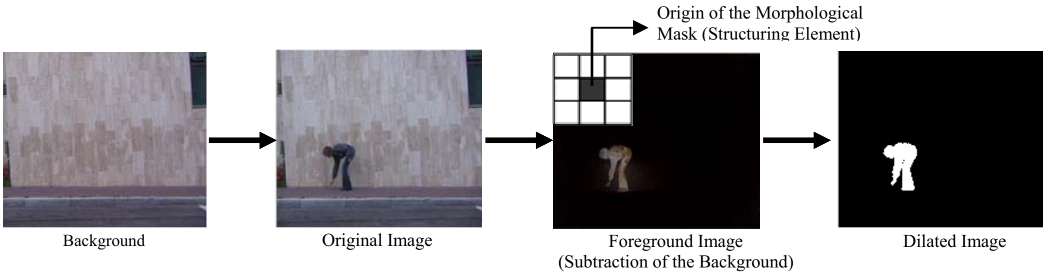

Fig.2. Segmentation Stage (Dilation of the image)

B. *Feature Extraction*

After segmentation process the next stage is feature extraction. In this stage some of the features have been selected randomly by using watershed transform. The watershed transform is used to extract the boundaries of the given image of the high pixel value. The general syntax of the watershed transformation is give an:

Output = watershed (A)

Here 'A' is dilated (segmented) image. The output of the above equation has been shown in Fig.3.

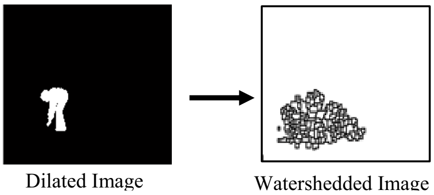

Fig.3. Random value selection by watershed transformation

The concept of watersheding is just like the dam construction which is used to compute a label matrix identifying the watershed regions of the input image A. The values of the output images are integer that are greater than or equal to 0. When the value of the elements is labeled to 0, then those do not belong to a unique watershed region. These are called "watershed pixels". When the elements are labeled 1, then belong to the first watershed region, when the elements are labeled 2, then belong to the second watershed region and so on. By default, watershed uses 8-connected neighborhoods for 2-D inputs. For higher dimensions, watershed uses the connectivity [12] and is given as:

CONNDEF (NDIMS (A), 'MAXIMAL').

The watersheding transformation (that is in geography called 'dam construction') consists of the following steps.

To find the low gray scale pixel in the image. Fig.4 (a).

If we keep a drop of water on that pixel which has a high gray scale and it will move only one direction i.e. towards that pixel which has a low gray scale. Fig.4 (b).

If we keep a drop of water on that pixel which has a high gray scale and it will move all the directions which have low gray scale intensities from that pixel. Fig.4 (c)

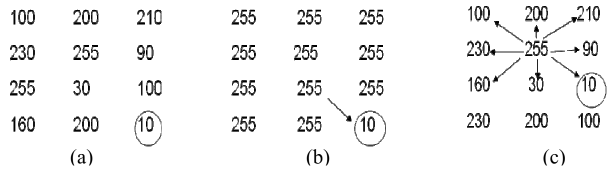

Fig.4. The watershed concept

C. *Recognition (Classification)*

After the selection of random values by watershed transform, all the values were added that used for recognition of the human activity. All the random have been added by using the Eq.2.

$$Sum = \sum_{i=0}^{M} \sum_{j=0}^{N} (i + j) \qquad \textbf{(2)}$$

Where 'M' and 'N' are the number of rows and columns respectively, and 'i' and 'j' indicate the intensity value of a pixel at x and y coordinates. For recognition the activity of the human, all the pixel values of the following frame have been added and then finally the decision has been taken on the behalf of the calculated value shown in Fig.5.

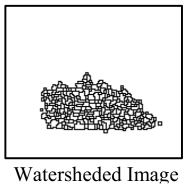

Fig.5. Random value selection by watershed transformation

The proposed algorithm that involves dilation followed by watershed transform required a series of steps to demonstrate how it worked. First, the video Camera captured the video, which was connected to a CPU through parallel ports, and then forwarded it to CPU. Then the CPU processed the video using the proposed algorithm (dilation followed by watershed transform). Through the proposed algorithm the activity of the human can easily be recognized.

IV. RESULTS AND DISCUSSION

In the proposed algorithm six different types of activities (named: bend, jack, run, side, skip, and wave) have been recognized. In each category there are same activities performed by 10 different people. Each activity consists of about 70 frames. So the whole database contains of 420 images. First the video is divided into number of frames and then background is removed, and then the morphological dilation has been applied for noise removing and for image enhancement used to make the algorithm efficient one. Then for feature extraction the watershed transform has been applied, through which some of the random values have been selected. After applying watershed transform all the selected values have been added for activity recognition. For training some of the values that were extracted by watershed transform have been selected from 170 images. The algorithm has been tested by using those values which were selected during the training stage. Fig.6. and 7, and Table I and II show the accuracy of the each category.

TABLE I. ACCURACY RESULSTS OF DIFFIRENT ACTIVITIES USING THE PROPOSED ALGORITHM

Activity Type	Results found Correct (%)
Side	80
Jack	78
Run	75
Bend	90
Right Hand Wave (RHW)	92
Skip	70

TABLE II. EFFICIENCY RESULSTS OF DIFFIRENT ACTIVITIES USING THE PROPOSED ALGORITHM

Activity Type	Results found Correct (ms)
Side	210.5
Jack	250.9
Run	350.8
Bend	400.4
Right Hand Wave (RHW)	390.6
Skip	210.0

In Table I and II the accuracy and efficiency of each category have been determined by applying the proposed algorithm upon the database of 6 different style of activities (consists of 420 images (70 frames of each), from which we determined the accuracy and efficiency of each category by adding the pixels of all the processed database of images. The algorithm gave reliable accuracy and efficiency.

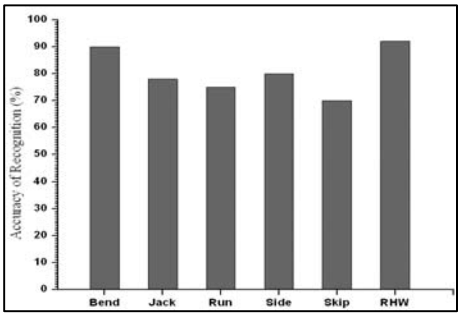

Fig.6 Accuracy of recognition

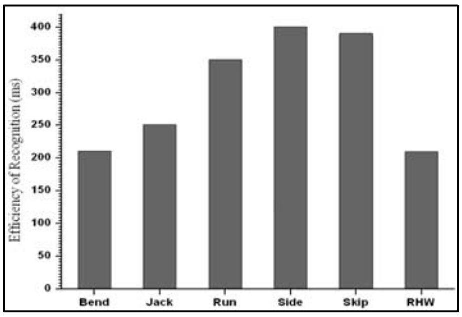

Fig.7 Efficiency of recognition

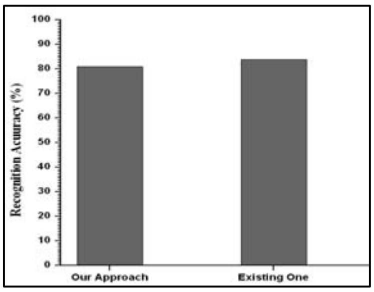

Fig.8. Comparison with existing one [2]

It is obvious from Fig.8 that our approach is comparable to the existing one. But the existing work [2] consisted of three stages, segmentation, feature extraction (using Principal Component Analysis) and recognition (using Hidden Markov Model). PCA and HMM consisted of the following steps:

- Generating the data set.
- Finding mean and subtraction of mean.
- Finding covariance.
- Finding the eigenvectors.
- Deriving the new data set (means finding the projection of the eigenvector).
- Classification using HMM, which is too much complex that needs a huge number of parameter. For example the number of parameters that are evaluated in three-state HMM is 15, and the number of parameters that are evaluated in a four-state HMM with five continues channels are 50 [13].

The existing work also process much parameters, so the processing time for finding all these parameters is very much high. Our proposed method only processes a scalar value (0-dimentional) for classification. So in-terms of efficiency our proposed method is the efficient one (302.2 ms) from the existing one.

V. CONCLUSION

In this paper we have presented an algorithm for human activity recognition based on morphological dilation followed by watershed transform. The proposed consisted of three stages, segmentation, feature extraction, and recognition. Morphological dilation is used for segmentation stage, for noise removing and for image enhancement. Watershed transform has been used for feature extraction stage, which selects random feature from the image. Finally the activity has been recognized by adding the selected features by the watershed transform. The recognition performance of our method has been compared with the existing method using Principle Component Analysis (PCA) to derive activity features, which give comparable results. But in-term of efficiency, our algorithm was much faster than the existing work. The average accuracy 80.83 % and efficiency 302.2 ms of the proposed algorithm for recognition was obtained. The dataset for testing have been taken from [14].

In our method the activity is recognized based on the addition of the selected pixels. So the future work is needed to evaluate our method through other popular classifiers.

ACKNOWLEDGMENT

The author would like to thank Dr. Donghai Guan and Mr. La The Vinh for their contributions in the improvement of algorithm and manuscript and providing materials related to this research.

This research was supported by the MKE (The Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency)" (NIPA-2010-(C1090-1021-0003)).

REFERENCES

[1] D. Guan, Y.-K. Lee, and S.Y. Lee, "Activity Recognition with the Aid of Unlabeled Samples," *Proc. of the 3rd International Conference on Ubiquitous Information Management and Communication (ICUIMC)*, Korea, 2009, pp.720-724.

[2] M.Z. Uddin, J.J. Lee, and T.-S. Kim, "Shape-Based Human Activity Recognition Using Independent Component Analysis and Hidden Markov Model, " *Proc. of 21st International Conference on Industrial, Engineering, and other Applications of Applied Intelligent Systems*, 2008, pp.245-254, Springer-Verlag Berlin Heidelberg.

[3] P. Turaga, R. Chellappa, V.S. Subrahmanian, and O. Odrea, "Machine Recognition of Human Activities: A survey," *IEEE. Trans. Circuits & Systems for Video Tech,* vol. 18, pp.1473-1488, September 2008.

[4] M.Z. Uddin, J.J. Lee, and T.-S. Kim, "Independent shape component-based human activity recognition via Hidden Markov Model," *Jounal of Applied Intelligence*. Springer Netherlands, DOI: 10.1007/s10489-008-0159-2, January 2009.

[5] Y. Wang, and G. Mori, "Human Activity Recognition by Semilatent Topic Models," *IEEE Trans. Pattern Recognition and Machine Intelligence.* vol. 31(10), pp. 1762-1774, October 2009.

[6] P.C. Ribeiro, and J.S. Victor, "Human Activity Recognition from Video: modeling, feature selection and classification architecture," *Proc. of International Workshop on Human Activity Recognition and Modeling* (HAREM), 2005.

[7] S. Ali, and M. Shah, "Human Activity Recognition in Videos Using Kinematic Features and Multiple Instance,". *IEEE Trans. Pattern Recognition and Machine Intelligence,* vol. 32(2), pp. 288-303, Febreuary 2010.

[8] D. Guan, W. Yuan, Y.-K. Lee, A. Gavrilov, and S.Y.. Lee, "Activity Recognition Based on Semi-supervised Learning," *Proc. of the 13th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA)*, Korea, 2007, pp. 469-475.

[9] J. Li, "Moving Objects Segmentation Based on Histogram for Video Surveillance," *Journal of Modern Applied Science,* vol. 3(11), November 2009.

[10] M.H. Siddiqi, I. Ahmad, and S.Bt. Sulaiman, "Weed Recognition Based on Erosion and Dilation Segmentation Algorithm," *Proc. International Conference on Education Technology and Computer (ICETC),* Singapore, 2009, pp. 224-228

[11] D. Butler, S. Sridharan, and V.M. Bove. Jr, "Real-time adaptive background segmentation," *Proc. of the International Conference on Multimedia and Expo (ICME),* USA, 2003, pp.341-344.

[12] M.H. Siddiqi, W. Ahmad, I. Ahmad, "Weed Classification Using Erosion and Watershed Segmentation Algorithm". *International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering (CISSE)*, USA, 2008, pp.366-369.

[13] http://www.cse.unsw.edu.au/~waleed/phd/html/node36.html

[14] http://www.nada.kth.se/cvap/actions/

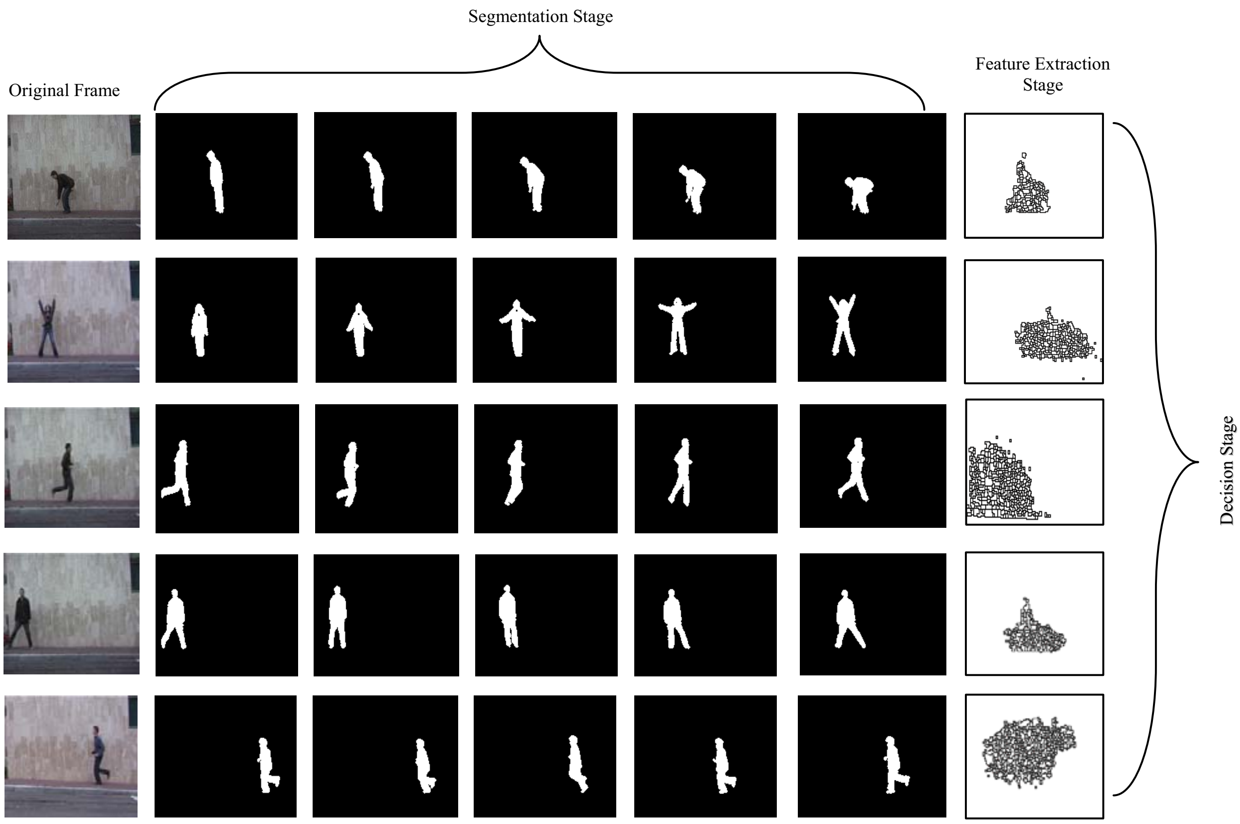

Fig.9. The whole process of the proposed method tested on different activities