Journal of Theoretical and Applied Information Technology
31st October 2014. Vol. 68 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

HUMAN ACTIVITIES RECOGNITION VIA FEATURES EXTRACTION FROM SKELETON

[1]GHAZALI SULONG, [2]AMMAR MOHAMMEDALI
[1,2]UTM-IRDA Digital Media Centre (MaGIC-X),Faculty of Computing,
Universiti Teknologi Malaysia,UTM Skudai 81310,
Johor, Malaysia
E-mail:[1] ghazali@utmspce.edu.my, [2]ammar_ncc@yahoo.com

ABSTRACT

Human activities recognition (HAR) enabling the understanding of basic human actions from still images has overriding importance in computer vision and pattern recognition for sundry applications. We propose a novel method for HAR by taking out the skeleton from the image for extracting useful features. This approach comprised of two steps namely (i) an automatic skeletal feature extraction and partitioning into two parts as block that determines angles between terminals and (ii) HAR by using non-linear Support Vector Machine (SVM). The model performance is evaluated using three available challenging datasets such as INRIA, KTH and Willow-action all with seven activities and each possessing eight scenarios. The images are normalized in (64x128) pixels format from digital silhouette via circle algorithm. Our method efficiently achieves a recognition rate as much as 86% with excellent features. The proposed model being highly promising compared to the existing one may contribute towards the development of computer vision architecture.
Keywords— *HAR, Features, Skeleton, Support Vector Machine, Still Image.*

1. INTRODUCTION

In the information technology era the Human Activity Recognition (HAR) play a paramount role in widespread computer applications and vision studies. Generally, HAR is based on still image or video recording. Recently, HAR using video has widely been carried out without much difficulty [1],[2],[3],[4],[5]. However, HAR using still images is not only intricate but pose many challenges [6]. Recognition using video recording is implemented by extracting the picture frames following a process that compares current frame (one movement) with the next or previous frame to predict activities. Still image based recognition is tricky due to the absence of any prior knowledge regarding the features of event and thereby complicated to predict all the important fixed features from the captured image. The main goal of HAR using still images is to develop an automatic analysis scheme for collected data to precisely determine the human behavior in an efficient manner. The appropriate implementation using suitable human computer interfaces significantly depend on such image analyses useful for diversified applications from surveillance to security systems to automation. Despite much effort on HAR using video recording the feature extraction from still image are not widely investigated in the domain of computer vision due to its inherent complexity.

The demand for HAR and categorizing their features from still images using efficient and accurate classifier is never ending. Lately, several methods are employed to extract features from the object in the image for activities recognition in [7] the human body by dividing it into six blocks and then matching each of them with the corresponding part of the body and shapes. Features are extracted from blobs crossing the block. [8] found a new technique consisting of pose extraction from still image using silhouettes extracted by thresholding over the probability maps. Different approaches of HAR with numerous applications are introduced [9]. The systems used to identify the actions in video follow the same procedure of recognition action in still image. Still image recognition is more challenging due to the absence of any prior knowledge and the features have to be extracted from object itself. Due to this reason, recognition of activities from single or still image is somewhat ignored resulting the articulation much harder. In order to address this problem many methods are developed in recent years. Megavannan [10] studied

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

action recognition from depth image taken from single commodity depth camera and proposed Random Occupancy Pattern (ROP) features [11] to develop Depth Motion Maps (DMM) for projected map. DMM is a binary map which is generated by computing and thresholding the difference of two consecutive frames and then summing them for each projective view. Often, the analysis of human activities are difficult not only because of human body is highly articulated but people also tend to wear complex textured clothing which could confuse and hide the important features needed for distinguishing the poses.

In this paper, we attempt to address some important issues related to the features extraction using still image. The object inside the still image has high amount of articulation that requires precise classification and in-depth understanding. The computer vision development using still image by extracting features from the object and classifying them to determine the correct activities is the main concern. A novel method for HAR by taking out the skeleton from the image for extracting useful features is proposed. This approach consists of feature extraction and partitioning using non-linear Support Vector Machine. Simulation is performed on a set of test datasets to evaluate the recognition rate.

The paper is organized as follows. In section 2 the detailed methodology consisting of three major steps such as preprocessing (noise reduction and segmentation), features extraction (based on skeleton and partitioning human body) and skeletonization together with the estimations of joints are described. Section 3 underscores results and discussion. Section 4 concludes the paper. Some suggestions regarding future works are underlined in Section 5.

2. METHODOLOGY

Each system for HAR are achieved using three main stages, firstly preprocessing including noise reduction and segmentation, secondly the feature extraction which is our main concern, and finally the classification via an appropriate classifier [12]HAR from still images captured by camera or scanned by any scanner devices are used as input to the system. These images may contain some noise which needs to be removed before segmentation can take place. The noise reduction properties is decided by the involved neighborhood in the smoothing process that has chosen large enough to reduce it by averaging [13]. Simple and powerful expression for noise reduction yields,

$$f(x,y) = \frac{1}{MN-d}\sum_{(s,t)\in S_{xy}} Id(s,t)$$

where S_{xy} is a neighborhood of size (M×N) associated with a co-ordinate (x,y) in the image I, Id is the remaining gray-level values where $\frac{d}{2}$ smallest and largest detected.

Datasets such as INRIA, KTH and Willow-action from public domain are used for seven activities such as walking, running, jogging, boxing, clapping, waving and clapping with eight scenarios for each activity. The images are normalized as to estimate recognition rate. It is obtained by dividing the number of items correct activities classified over the number of tested image and is g iven by:

$$\text{Recognition rate} = \frac{\text{Number of corrected activities}}{\text{Number of given activities}} \times 100\%$$

2.1 Feature Extraction

Skeleton shaped object is essential for extracting features and the properties from the segment object. One of the aims of using the system proposed is to specify different parts and its position for human body. Each part is separately analyzed for feature extraction. We are interested in three main parts of the body such as head, arms and legs. The present method divides the body into two blocks each one consist of different pose known by the system. The first block includes the arms and the upper half of the body and the second block contains the legs only as shown in Fig.1.

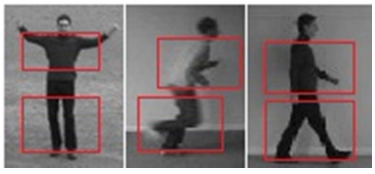

Figure 1: The two partitioning blocks

These blocks are considerably affected with human movements. Analysis of a particular block allows one to comprehend the activities in which some action such as running needs to be incorporated. The features from two blocks or whole body can then be recognized.

Journal of Theoretical and Applied Information Technology
31st October 2014. Vol. 68 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.2 Skeletonization

Skeletonization is a process of skeleton extraction from a digital binary picture. This provides region based shape features and a common preprocessing operation in pattern recognition. In digital space, only an approximated attributes to the "true skeleton" can be extracted. The two main necessary requirements that must be pursued during skeletonization are to retain the topology of the original object and the skeleton must be in the middle of the object.

Following the previous shape extraction procedures from an image (segmentation) the ridge line is calculated from the silhouette shape (object). The central points (critical point) inside the silhouette, especially with terminals parts such as arm and leg are positioned. By applying center algorithm for circle inside the shape as displayed in Fig. 2 the dose algorithm is operated.

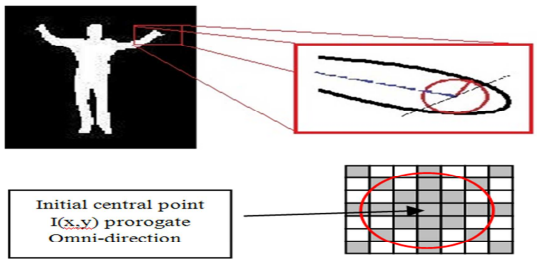

Figure 2: Procedure For Extraction Of Skeleton

The circle moves inside the body and checks the center of specific part via neighborhood pixels along horizontal, vertical and diagonal directions. The point (pixels) in the contour of the circle must belong to the edge of the silhouette shape i.e. along the tangent to the circumference of a circle. This implies that the circle is of variable size based on the trajectory as presented in Fig. 2. The pixels inside the shape grows in every direction alike the seed propagation algorithm [14] till they reach to the edge and then fixes the center of the circle to get shifted in the next position. The pixels are tracked by keeping them along one line and then all shapes are connected together to obtain the entire skeleton.

2.3 Estimation of Joints

The detailed description of the skeleton joints estimation is essential. It is well known that the skeleton of a human body is fixed parts connected together by joints. A standard human skeleton model can be chosen to estimate the joints because for different bodies the proportions of parts are more or less the same [15]. Estimation of joints enables the skeleton construction simplistic by transforming the whole shape as tree so that the calculation of distance between joints and terminals become easier. As illustrated in Fig. 3, the joints are divided into two categories including active and inactive based on their importance and the range of motion. The examples of active joints are elbow, shoulder, knee and waist. Conversely, the inactive joints such as head, neck, wrist, and ankle which are not dynamic within the human activities are beyond the scope of our research.

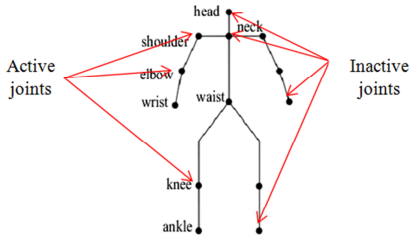

Figure 3: The Standard Skeleton Shape With Joints

2.4 Skeleton Partitioning

Partitioning of the skeleton into two blocks is carried out for extracting useful features from it. The first or the upper block contains shoulder, elbow and wrist joints and second one is the lower part which includes knee and ankle joints. The vertical distance from the head joint to the ground level can be extracted as shown in Fig.4.

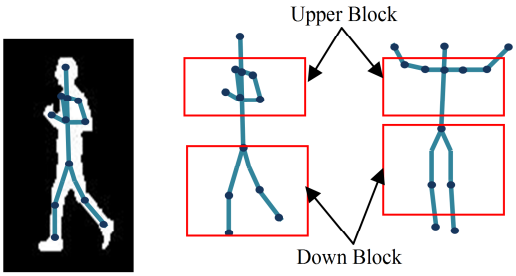

Figure 4: Segmentation And Skeleton Within Two Blocks Consist Of Most Important Features

Some activities due to its intrinsic nature reconstruct the shape of the skeleton and joints to specific structures. For instance, in running activity the arms make angles with elbow joint less than 90^0 (upper block) simultaneously the legs also create angle (lower block) with the vertical distance less than 90^0. Furthermore, for jogging activity the upper

Journal of Theoretical and Applied Information Technology
31st October 2014. Vol. 68 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

block acts same as running and vertical distance become 90^0 with the ground level. In contrast, for walking activity the extracted angles for upper and lower block become more than jogging and running. The same procedure is followed for the rest of the actions. Fig. 5 shows the approach for extracting features from various activities.

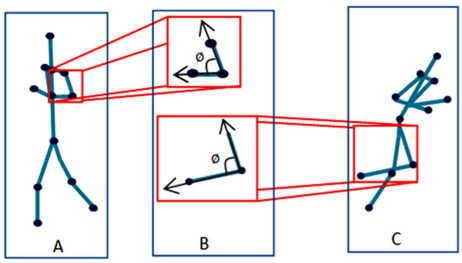

Figure 5: (A) Skeleton For Jogging Activity (B) Angles Extracted From A And C (C) Skeleton For Running Activity

3. RESULTS AND DISCUSSION

The performance of the proposed model is evaluated by applying it to three challenging datasets obtained from public domain. These datasets are INRIA, KTH and Willow-action all with seven activities such as walking, running, jogging, boxing, clapping, waving, and clapping and each activity consists of eight scenarios. The images are normalized as (64x128) pixels. For waving and jogging activities the front sight is considered and for the rest of the activities the side views (left or right) are utilized. These activities are purposely chosen due to their well-admired popularity in still images and datasets.

This dynamic method for feature extraction facilitates the classification of these activities easier and very accurately determines the recognition rate when using non linear SVM classifier. The activities such as running, walking and jogging generate high recognition rate because these features are more accurately extracted from skeleton angles using two blocks. Meanwhile, the boxing and waving activities produce expected less recognition rate than others due to their extraction from one block (upper block). The applicability of our developed approach is limited by online recognition and associated computational time which is somewhat longer than others. Conversely, this method is very accurate for features extraction from still images and is highly satisfactory in terms of richness. The recognition rate as high as 86 % obtained by us is much improved compare to the existing one.

Running	0.81	0.01	0.04	0.03	0.0	0.0	0.12
Walking	0.12	0.90	0.02	0.04	0.02	0.0	0.01
Boxing	0.0	0.02	0.86	0.0	0.13	0.02	0.0
Jogging	0.04	0.14	0.01	0.81	0.0	0.0	0.0
Waving	0.0	0.03	0.0	0.02	0.86	0.14	0.0
Clapping	0.0	0.2	0.0	0.0	0.14	0.87	0.0
Sitting	0.01	0.0	0.2	0.03	0.0	0.0	0.85
	Running	Walking	Boxing	Jogging	Waving	Clapping	Sitting

Figure 6: Confusion Matrix Showing Recognition Rate For Five Activities.

The results for the confusion matrix with supervised method are furnished in Fig. 6. Interestingly, for activities such as walking, running and jogging the system is capable of recognizing the features extracted from two parts of the body (arm and leg) and a high recognition rate is accomplished. In contrast, the activities such as boxing and waving achieves relatively lower recognition rate because the features are extracted from one part of the body (arm).

The training data for given images using dataset INRIA and KTH for five activities in the present classifier produced notable features when compared and summarized in Table 1.

Table 1 Comparison Of Recognition Rate Of The Proposed Method With Others.

Classifier	Dataset Used	Recognition Rate (%)	No. of Activities
LSVM [16]	Weizmann & KTH	70.5	7
Multi SVM+HOG [17]	Willow-action	61.07	6
SVM [18]	Willow-action	60.66	5
BSVM+LDA [19]	Collected & INRIA	85.1	6
Proposed method	INRIA, KTH & Willow-action	86	7

We evaluate the accuracy of our method with seven activities using three different datasets

ISSN: 1992-8645 www.jati

because some activities are available in a particular dataset and some belong to another. Training set with INRIA and Willow dataset are found to be much more computationally time consuming than with KTH because each image is changed to gray scale followed by noise reduction, segmentation and so on. Meanwhile, training with INRIA dataset consisting of 780 images of (64 × 128) pixels format of each activity takes longer time for the machine to learn. A classifier performs a set of steps to define the feature classes or categories. Classifiers are trained by running samples of known classes to identify the unknown one. Finally, by executing these trained classifiers on unknown activities or feature vectors the belongingness of a particular class can be determined.

Firstly, the classes on a set of training content are defined. Then, the classifier uses those classes to analyze other content and performs classification. The classifier performs the features analyses following,

$$D = \{(x_i, y_i) | x_i \in \mathcal{R}, \quad y_i \in \{-1,1\}\}_{i=1}^{n}$$

where each x_i is an n-dimensional real vector and y_i being either +1 or -1 indicate the classes in which the feature x_i belongs. This allows us to determine the best split line (for linear) or curve (for non-linear) that divides the features having $y_i = 1$ (class A) from those of $y_i = -1$ (class B). The best line between the classes can be achieved for linearly separable data.

Every classification machine learning algorithm operates using the training and testing phase. The notion of classification is to create a decision rule $d(x): R^d \rightarrow \{w_1, w_2, \ldots, w_c\}$ to classify d-dimensional observation (point) X into one of the classes. Generally, the rule is built using the observation from a training set.

In the testing (or prediction) phase a given image supported to the system is already learnt by training mode. HAR system captures this image and administers the procedures similar to the training mode including segmentation, feature extraction and so on. Finally, the system compares the features of given image with features extracted from dataset images during training mode to discover the suitable match (or closest counterpart) in recognizing the specific activity as depicted in Fig. 7.

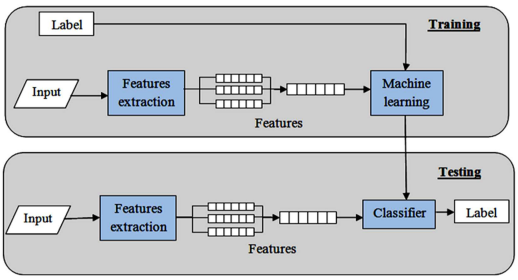

Figure 7: Training And Testing Mode of HAR system.

The number of pores used as features [16] with 408 images and LSVM classifier is limited by features. Meanwhile, the pose of human body (skeleton) with HOG and Multi-SVM classifier [17] and SVM in [18] possess poor features compatibility. Good features with suitable classifier (BOF+LSVM+Inte) are acquired [19]. However, our proposed method stands on the advantage of the previous methods and neglects their disadvantages by using powerful features extraction procedures. Consequently, features are extracted from skeleton by partitioning human body. Different challenging datasets (1220 training images and 420 testing images) those are used in our non-linear SVM classifier make the result more accurate in comparison to those exist in the literature as listed in Table 1.

4. CONCLUSION

We report a method for HAR in which s keleton shape from the object in still image is taken out to extract useful features for various actions. The developed system extract skeleton from digital silhouette by circle algorithm and use angles of arms and legs as a features to non-linear SVM classifier. This two step approach used seven activities including walking, running, jogging, boxing, sitting, waving and clapping where each activity composed of eight scenarios. The model performance corresponding to these activities is evaluated and validated by applying it on three popular datasets (INRIA, KTH and Willow-action) available in the public domain. The excellent features extraction approach using this non-linear SVM classifier achieves a high recognition rate of 86%. Our findings may find potential application towards activities recognition and mimicking in computer aided vision systems. The main goal of HAR system is to reduce the effort of the human by learning machine or computer during support it with information. This allows the computer to predict the human activities to take advantage of prior knowledge. All these are performed by classifying

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

the features which are extracted accurately from human body to enable the system in recognizing them more accurately than before.

5. FUTURE WORKS

Despite our current HAR system is reasonably efficient still there is a need for faster and efficient method. Furthermore, the proposed system is time consuming and unsuitable for on-line applications because it deals with still images. It is worthy to use special HOG features with INRIA dataset where the image conversion into grayscale is not needed. Finally, a versatile system with more than seven activities is worth developing to cover large number of actions for security and industrial applications.

ACKNOWLEDGEMENTS
Ammar is grateful to the Ministry of Science and Technology, Iraq for the study leave and Universiti Teknologi Malaysia for technical assistance.

REFERENCES:

[1] Abdel-Badeeh M. Salem, Adel A. Sewisy, and Usama A. , "A Vertex Chain Code Approach for Image Recognition," *ICGST-GVIP Journal*, vol. 5, Mar. 2005.

[2] A.Bandera, C. Urdiales, and F. Sandoval, "2D object recognition based on curvature functions obtained," *Pattern Recognition Letters*, vol. 20, pp.49-55, Oct.1999.

[3] Y. Du, F. Chen, W. Xu, and W. Zhang," Activity recognition through multi-scale motion detail analysis,". *Science Direct,* vol. 71,pp. 3561–3574, Oct. 2008.

[4] C. Gu, P. Arbel, Y. Lin, K. Yu, and J. Malik," Multi-Component Models for Object Detection," *NEC Labs America*, vol. 13, pp.123-133,Nov.2012.

[5] C. Zhu, and W. Sheng," Motion- and location-based online human daily activity recognition,". *Science Direct*, vol. 7, pp. 156-26, Apr. 2011.

[6] C. Thurau and V. Hlavac," Pose primitive based human action recognition in videos or still images," *IEEE Computer Society Press*, vol. 7,pp. 1-8, Jun. 2008.

[7] S. Min Yoon, A Kuijper," Human action recognition based on skeleton splitting," *Science Direct*, vol. 40,pp. 9848-6855, Aug. 2013.

[8] N. Ikizler, R. Gokberk, S Pehlivan and P. Duygulu," Recognizing Actions from Still Images," *IEEE,* vol. 7, pp. 1-4, Dec. 2008.

[9] J. K. Aggarwal and M. S. Ryoo," Human Activity Analysis: A Review," *ACM Computing Surveys*, vol. 43, pp. 1-16, Apr. 2011.

[10] Megavannan, V. , Agarwal, B. , and Venkatesh Babu, R.," Human action recognition using depth maps," *IEEE*, vol.5, pp. 1-5, Jul.2012.

[11] C. Chen, K. Liu, and N. Kehtarnavaz," Real-time human action recognition based on depth motion maps," *Springer*, vol. 20, Aug. 2013.

[12] F. Luisier, and T. Blu," Image Denoising in Mixed Poisson Gaussian Noise," *IEE*, vol. 20, pp. 165-172, Mar. 2013.

[13] A. Buades, B. Cool, and J.M. Morel," A review of Image Denoising Algorithms, with A new One" *SIAM Journal on Multiscale Modeling and Simulation*, vol. 4, pp. 490-530, Aug. 2005.

[14] T. Horiuchi and S. Hirano," Colorization algorithm for grayscale image by propagating seed pixels" *IEEE*, vol. 1, pp.1-60,Sept. 2003.

[15] J. Ding, Y. Wang, and L Yu," Extraction of human Body Skeleton Based on Silhouette Images" *IEEE*, vol. 1, pp. 71-74, Mar. 2010.

[16] S. Malathi and C. Meena," Fingerprint Matching Based on Pore Centeroid", vol. 1, ICTACT, pp. 220 – 223, May 2011.

[17] W. Yang , Y. Wang , and G. Mori," Recognizing Human Actions from Still Images with Latent Poses," *IEEE*, vol. 1, pp. 2030-2037,Jun. 2010.

[18] V. Delaitre, Josef Sivic, and Ivan Laptev," Learning person-object interactions for action recognition in still images," *NIPS*, vol. 24, Oct. 2011.

[19] Nazli , R. Gokberk , S. Pehlivan , and P. Duygulu," Recognizing Actions from Still Images,", *IEEE* , vol. 1, pp. 1-4, Dec. 2008.