Journal of Theoretical and Applied Information Technology
20th June 2016. Vol.88. No.2

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

ACTION BASED FEATURES OF HUMAN ACTIVITY RECOGNITION SYSTEM

[1,3]AHMED KAWTHER HUSSEIN, [1] PUTEH SAAD, [1]RUZELITA NGADIRAN, [2] YAZAN ALJEROUDI, [4] MUHAMMAD SAMER SALLAM

[1] School of Computer and Communication Engineering, University Malaysia Perlis, Perlis, Malaysia
[2]Department of Mechanical Engineering, International Islamic University Malaysia, Selangor, Malaysia
[3]Department of Computer Science College of Education, The University of Al-Mustansiriyah, Baghdad, Iraq
[4]Department of Computer Engineering and Automation, Damascus University

ABSTRACT

Human actions are uncountable and diverse in nature; each action has its own characteristics and nature. Moreover, actions are sometimes different and sometimes very similar. Thus, it is rather challenging to implement a system that is capable of recognizing all human actions. However, the problem of recognition can be made simpler if the system of recognition is built gradually. Firstly, a set of required actions must be selected and then each action must be studied and analyzed to determine the most distinctive features that remain similar if different subjects perform the same action. The concept of action-based features is proposed and validated in this article. The system still has the ability to be extended to recognize more actions, simply by including contextual features of any added action. Experimental results have shown an outperforming performance with 100% accuracy based on the evaluation of UTKinect Action Data Set.

Keywords: *Pattern recognition, Feature extraction, Elm, Neural network*

1. INTRODUCTION

Human activity recognition (HAR) is an emerging research field. Numerous applications require an accurate HAR system; for instances video archiving, surveillance, human computer interfacing and gaming. Different sensors are used for HAR; for example inertial sensing such as gyros and accelerometers, cameras, and other auxiliary sensors such as microphones. We believe that in-depth data provided by RGB-D Kinect are useful in HAR due to the direct sensing of the longitudinal dimension. Moreover, depth data are useful in extracting skeleton joints which carry the most meaningful information about human activities.

The HAR system is referred to as other patterns recognition-based system. Typically, various stages of processing are performed in order to issue the final decision of the system. Data pre-processing, feature extraction and machine learning are the main components of most pattern recognition systems. Most researchers use common signal-based features such as spatial domain and frequency domain features. Some features aim at developing concepts of human action related features. Unfortunately, building a comprehensive feature space of human activities is simply unfeasible. This is because human actions are unlimited. In other words, any system designer has to expect that more features are to be added to the space of recognition. In this article, the concept of action-based features is proposed. The action-based features signify that the features that are dominant in one or two actions may somehow remain passive in other actions. We believe that human's perceive their actions based on such features. For example, when we see periodic movements of hands above the head, we assume that the hand waving action is being performed. By adding the two features of "above head" and "periodic" as the keys to recognize the hand waving action, it therefore separates it from other actions. This concept can be generalized to all other human actions. In other words, it is important to add any action to the system to analyze and find which features are the most dominant in order to add them to the features' space.

Different types of features were used for the application of human activity recognition. Some approaches used the motion-based features (Nguyen et al., 2013; Kliper-Gross et al., 2012) while others used the local descriptor features such

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

as scale-invariant-feature-transform (SIFT) (Scovanner, 2007) and histogram of oriented gradient (HOG) (Das, 2014). Skeleton features extracted from depth image are also used in many approaches (Sung et al., 2012; Xia, 2012).

Other researchers attempted to extract silhouettes features which are used to represent spatial and temporal domain. One of the common spatial-time features that is extracted from silhouette and has been used for video-based human action recognition is the Space-Time Volume feature (Poppe, 2010; Mokhber et al., 2008). This vector of feature has the capability to capture human motion and action by building a vector of silhouette feature that is sequenced in time. Junejo, Junejo, and Al Aghbari (2014) proposed a combination of Symbolic Aggregate appro Ximation (SAX) and time-series representation for the silhouette.

Unfortunately, reaching a required accuracy of human activity recognition requires huge amount of feature data. Therefore, there is a need to select features more effectively. Effectiveness in selecting features means finding out the most discriminative features and reducing the set of feature data in quantity while increasing its power in action classification. Considering that human actions are not limited and any human action recognition system has to have the scalability aspect, it is therefore highly recommended that the features of the system are inserted in a very careful manner in order to assure its effectiveness and capability in discriminating actions. This article proposes an innovative approach of building human activity recognition system gradually by selecting action-based features. The action-based features refer to any action that it is added to the system before its behavior and conducting nature are analyzed carefully. Next, the most discriminative features of the action are modelled mathematically. This approach can guarantee two goals; discriminative features and reduced size of data features. The remaining of the article is organized as following, whereby section II introduces the methodology used while section IV presents the experimental results and section V discusses the conclusion and future work.

2. METHODOLOGY

2.1 Input Data

Actions that are available in the most common data sets are captured through the use of Kinect camera. Kinect sensor provides two types of data; RGB data and depth data. Microsoft SDK of Kinect provides the coordinates of the joints of the human body through the use of skeleton extraction algorithm. The work in this paper depends only on the coordinates of the joints that were extracted from Kinect.

2.2 Joints reduction

A total of 20 joints were extracted from Microsoft SDK. Since each joint consist of three coordinates, a total of 60 values were provided in each of the action frame. Multiplying this number by the number of frames leads to the derivation of a huge quantity of raw data which are available for each action. Therefore, the step of joints reduction was required. The concept of joints reduction was taken from the rigidity of the human body. It is clear that the movement of some adjacent joints is very similar due to rigidity of the body. Therefore, some joints can be ignored without loss in information. In particular, the two feet and two wrists and spine were ignored while other joints are taken into consideration.

2.3 Actions and corresponding features

Walk: In this action, all joints of the body move along translational movement in xy plane in comparison with other actions. In order to detect this translational movement, we have to define the following two terms:

- The center of gravity (CoG) of the upper part of human body is determined by taking the mean value of x coordinate and the mean value of y coordinate (for hip center, left hip, right hip, shoulder center, left shoulder and right shoulder joints).
- Translational component of CoG is defined as the maximum travelling distance of CoG in xy plane after scaling it to 1.3 m.

In actuality, the translational component of CoG of the upper part of the body plays an important distinctive feature of walking action.

2.4 Sit down, Stand up, and Pick up:

These three actions were combined into one group because they are similar in the movements that constitute the actions. However, they are different in terms of the direction and the number of the movements. The most important point here is the CoG of the upper part of the body as described earlier. In the “sit down” action, the human drops the CoG and then leans his back. In the “stand up” action, the human leans his back and then lifts the CoG. In the “pick up” action, the

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

human drops the CoG and then lifts the COG again. Two features are extracted here and scaled to 0.35 m:

$$Start\_Activeness = \frac{zStart - zMin}{0.35} \quad (1)$$

$$End\_Activeness = \frac{zEnd - zMin}{0.35} \quad (2)$$

Where:
Start_activeness, End_activeness: the two features.
zStart: the initial position of COG on z-axis.
zEnd: the final position of COG on z-axis.
zMin: the minimum position of COG on z-axis.

Table.1 explains the two ends of the behavior of these features in each action. As evident, the two ends of the behavior are dependent on the action type.

Table 1: Behavior of "Start Activeness" and "End Activeness" features

Actions	Start_activeness	End_activeness
Sit down	High	Low
Stand up	Low	High
Pick up	High	High

Carry: In this action, human walks while carrying an object between his two hands. This action embeds walking. Thus, the "activeness" feature values approximate to the values of the feature for "walk" action case. Therefore, a new feature has to be computed.

The concept of "closeness" is introduced to this action. Closeness is defined as a measure of distance between two hands and hip center on z-axis. This feature is computed because the distance between the two hands and the hip center in the case of "carry" action is smaller than the distance in the case of "walk" action during the period of action. This can make it a distinctive feature to distinguish between walk and carry.

The following equations explain the method of computing "closeness":

$$Closeness1 = mean(zHip - zLeftHand) \quad (3)$$

$$Closeness2 = mean(zHip - zRightHand) \quad (4)$$

$$Closeness = mean([Closeness1\ Closeness2]) \quad (5)$$

$$Closeness = abs(Closeness)/0.3 \quad (6)$$

Where:
zHip: is the position of hip center on z-axis during the action period.
zLeftHand: is the position of left hand on z-axis during the action period.
zRightHand: is the position of right hand on z-axis during the action period.

2.5 Push, Pull:

These two actions are similar in the movement that constitutes the actions while they are different in the action. The most contributing joint in this action is one of the two hands. It is assumed that the person who is conducting the action is right handed. We take into consideration the two hands and hip center.

$$Tr = dist(RH, Hip, Start) - dist(RH, Hip.End)$$
$$Tl = dist(LH, Hip, Start) - dist(LH, Hip.End)$$
$$Tm = \max(Tr, Tl)$$
$$Ts = \left[\frac{Tr}{Ts}\ \frac{Tl}{Ts}\right] = [T_{rs}\ T_{ls}]$$

Table 2: the behavior of Ts for both Push and Pull actions.

	Push	Pull
Tsr	-1	+1
Tsl	Close to zero	Close to zero

Wave hands: This action is a very distinctive action because it is the only action in which the human lifts his hands above his head. Hence, we define this new feature as "above head". "Above head" is a discrete feature that obtains either 0 if the human does not lift the two hands above head or 1 if the human does. Also, in this action the hands move approximately in similar way. Therefore, if the distance between the right hand and head and the distance between the left hand and head are computed during the action period as well as the correlation between the two resultant signals are computed, the correlation values are thus significantly high. In that case, a new feature is defined as "correlation". In addition to the last two features, the feature "periodicity" is also defined because in this action the two hands perform a periodic movement. Hence, "periodicity" is a discrete feature that obtains either 0 if the two hands do not perform a periodic movement or 1 if they do. In order to detect "periodicity" in the movement, the coordinates system needs to be transformed from the world origin to the hip origin first and then the resultant Cartesian coordinates need to be transformed into spherical coordinates system. Fourier transform of "elevation angle" for both hands are computed and analyzed in order to determine whether or not the two hands perform a periodic movement.

Journal of Theoretical and Applied Information Technology
20th June 2016. Vol.88. No.2

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

Clap hands: In this action, the human lifts his hands above the hip center point, under the head and claps. This thus defines the "highness" feature. "Highness" is a ratio of period of time at which the hands stay above the hip and under the head during the whole action period. Also, in this action, the hands touch each other. As a result, this action achieves the minimum distance between the hands and this thus defines the "touch hands" feature. "Touch hands" is computed by computing the minimum distance between the two hands during the action period and scaling this distance to 0.3 m. Another noteworthy point here is that "touch hands" might result in lower values when the object carried during the "carry" action is small.

After reviewing in details all required features, we summarize them in the following table (Table 2):

Table 2: Summary of features where total number of features is 11

No	Feature Name	Range
1	Translational Component of CoG	[0 1]
2	Start Activeness	[0 1]
3	End Activeness	[0 1]
4	Closeness	[0 1]
5	Left hand distance change	[-1 1]
6	Right hand distance change	[-1 1]
7	Above Head	{0,1}
8	Correlation	[0 1]
9	Periodicity	{0,1}
10	Highness	[0% 100%]
11	Touch hands	[0 1]

2.6 Classification:

Based on the previous section, it is evident that each action is represented by 11 features. The classification process must now take place. In order to decide the right class that each action belongs to, a learning algorithm called Extreme Learning Machine (ELM) for single-hidden layer feed forward neural networks (SLFNs) is used. ELM randomly chooses hidden nodes and analytically determines the output weights of SLFNs. In theory, this algorithm tends to provide good generalization performance at extremely fast learning speed.

3. EXPERIMENTAL RESULTS

3.1 Used data set:

In order to verify our work, the "UTKinect Action Data Set" is used. This data set consists of 10 actions (walk, sit down, stand up, pick up, carry, throw, push, pull, wave hands and clap hands). There are ten subjects in this data set and each subject performs each action twice. Hence, there are a total of samples of actions. Due to an error from the source of data set, the "carry" action of the tenth subject and the second experiment is not readable, resulting in only a total of 199 samples.

3.2 Feature results:

Although we have previously mentioned that the selected set of actions contains nine types of actions, the used data set however contains ten types of actions. Hence, the values of features will be shown for the whole data set. The classification results are divided into two parts; the first part includes the "throw" action while the second part does not. The reason for that is that including throw actions creates high confusion with the push actions due the high similarity between them when the actions were conducted by the subject.

Eleven types of features are discussed and analysed where the "Translational Component of CoG" is the first one to be discussed. The following figure (Figure 3) depicts the values of this feature for the whole data set:

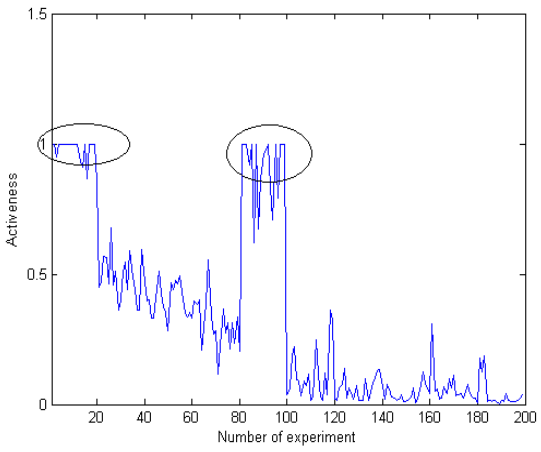

Figure 3: Values Of "Activeness" Feature For The Whole Data Set

Since there are 199 samples of actions, the range of x-axis therefore expands from 1 to 200. Also, samples from 1 to 10 represent the "walk" action for the first experiment while samples from

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

11 to 20 represent the "walk" action for the second experiment. The right arrangement of actions is mentioned in the "used data set" section.

In Figure 3, there are two ellipses that contain the values of "Translational Component of CoG" for the "walk" and "carry" actions. It is clear from the figure that "activeness" obtains the highest values for these two actions.

The next two features are "start activeness" and "end activeness". As discussed, these features are related to three types of actions (sit down, stand up and pick up). Figure 4 clarifies the concept of these features for the "sit down" action:

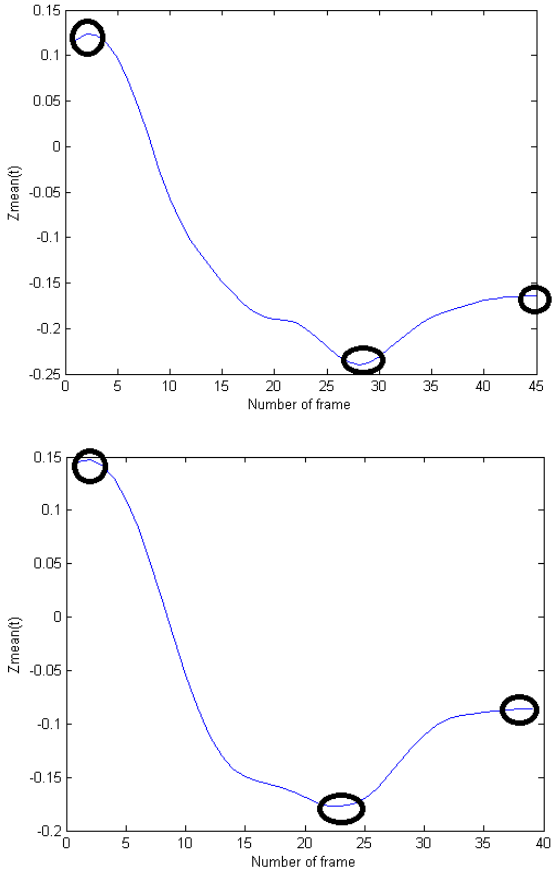

Figure 4 Position Of The Cog Point At Z-Axis During The Action Period For Two Samples Of "Sit Down" Action

In Figure 4, two samples of "sit down" action are displayed. The horizontal x-axis represents the time whereas the vertical y-axis represents the position of CoG point on z-axis. The three ellipses represent the start position, final position and the lowest point on z-axis. Based on this figure and the equations of (1) and (2), it is clear that the "sit down" action achieves **high** "start activeness" and **low** "end activeness".

Figure 5 clarifies the concept of these features for the "stand up" action:

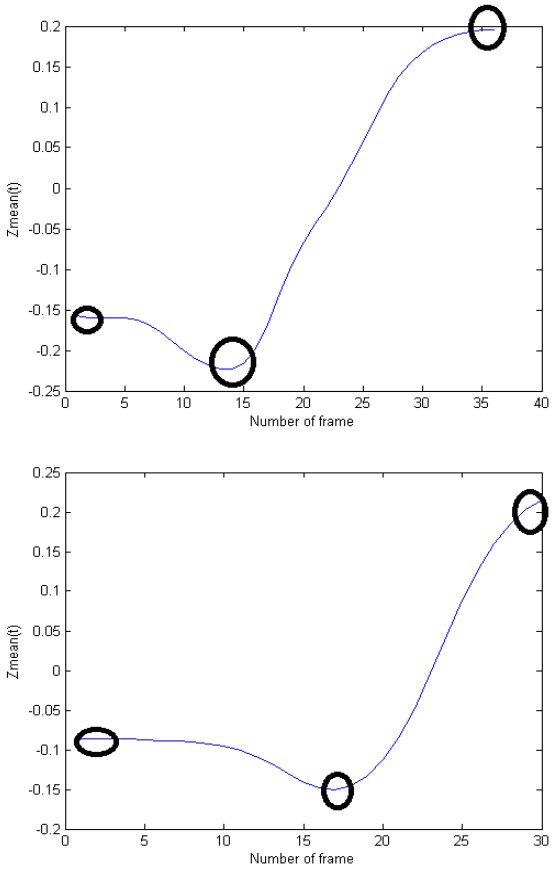

Figure 5: Position Of The Cog Point At Z-Axis During The Action Period For Two Samples Of "Stand Up" Action

In Figure 5, two samples of the "stand up" action are displayed. The horizontal x-axis represents the time whereas the vertical y-axis represents the position of CoG point on z-axis. The three ellipses represent the start position, final position and the lowest point on z-axis. Based on this figure and the equations of (1) and (2), it is obvious that the "stand up" action achieves **low** "start activeness" and **high** "end activeness".

Figure 6 elucidates the concept of these features for the "pick up" action:

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

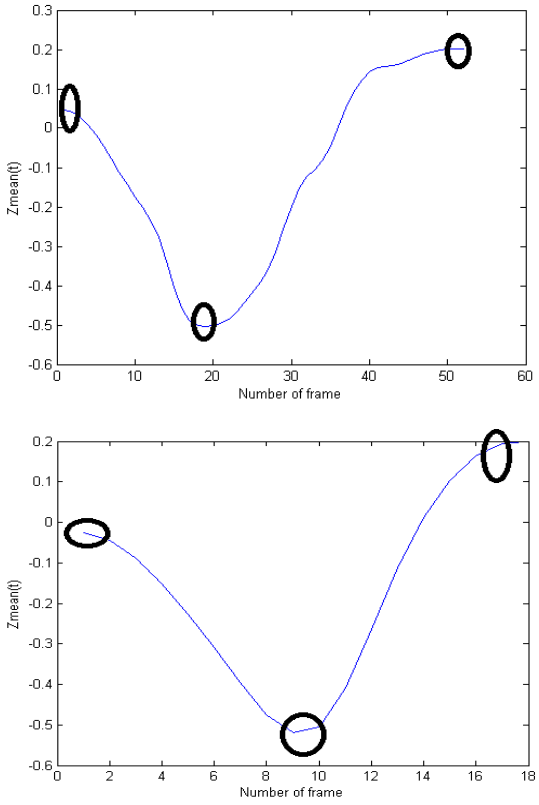

Figure 6: Position Of The Cog Point At Z-Axis During The Action Period For Two Samples Of "Pick Up" Action

In Figure 6, two samples of the "pick up" action are displayed. The horizontal x-axis represents the time whereas the vertical y-axis represents the position of CoG point on z-axis. The three ellipses represent the start position, final position and the lowest point on z-axis. Based on this figure and the equations of (1) and (2), it is evident that the "pick up" action achieves **high** "start activeness" and **high** "end activeness".

The fourth feature is "closeness". The main reason of computing this feature is to discriminate between "walk" and "carry" actions. Figure 7 displays the values of this feature for all samples of both actions:

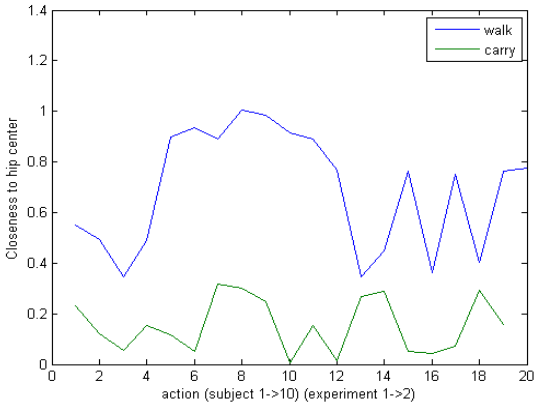

Figure 7: The Values Of "Closeness" Feature For All Samples Of "Walk" And "Carry" Action

As depicted in Figure 7, it is evident that the distance between two hands and hip center always remain smaller in the case of "carry" action than the distance in the case of "walk" action. The next two features are "left hand distance change" and "right hand distance change". Figure 8 illustrates the distance between the right hand and hip center and the distance between the left hand and hip center for a "push" action sample:

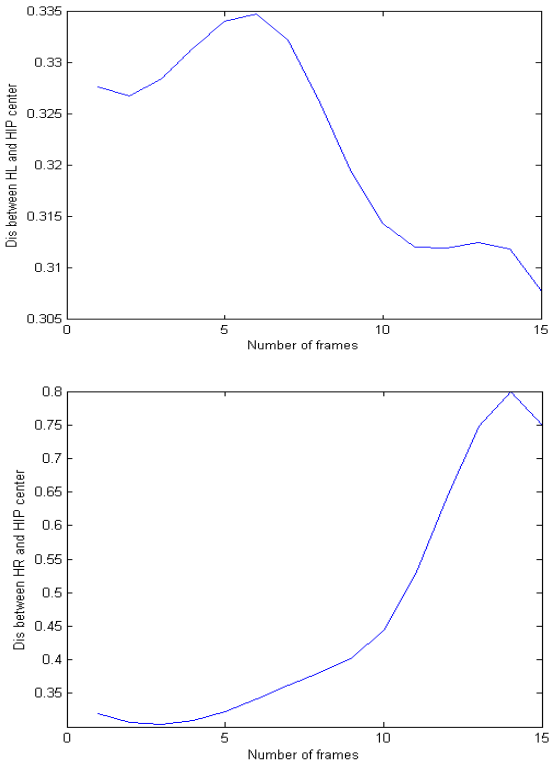

Figure 8: The Distance Between Left Hand And Hip Center During The Action Sample Period; The Distance Between Right Hand And Hip Center During The Action Sample Period

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

Since the right hand is assumed to be the most active joint in the case of "push" and "pull" actions, the figure shows that the left hand is comparatively steady although the right hand moves approximately 0.4m. By reviewing the equations of (7), (8), (9), (10) and (11), the following results in Table 3 are portrayed:

Table.3: Values Of "Left Hand Distance Change" And "Right Hand Distance Change" Features For The Sample Action.

Left hand distance change	Right hand distance change
0.0465	-1

"Above head" is the most distinctive feature for the "wave hand" action. It is obvious that this feature obtains 1 for the "wave hand" action, and 0 for the other actions as illustrated in Figure 9.

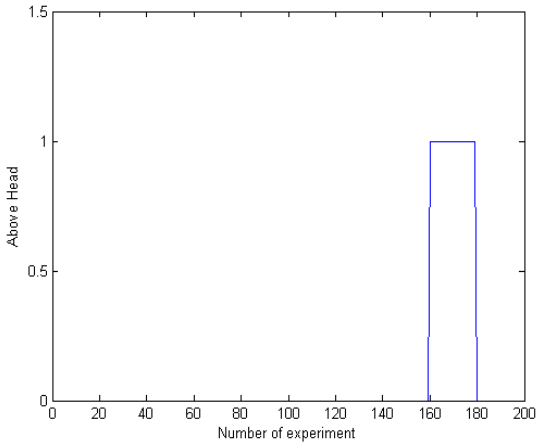

Figure 9: "Above Head" Feature Values For The Whole Data Set

It is evident that all samples of the "wave hands" action obtain the value 1.

The "correlation" feature also obtains higher values for all samples of the "wave hands" actions. Figure 10 denotes the example of the two signals that are required to compute correlation. The example represents a sample from the "wave hand" action:

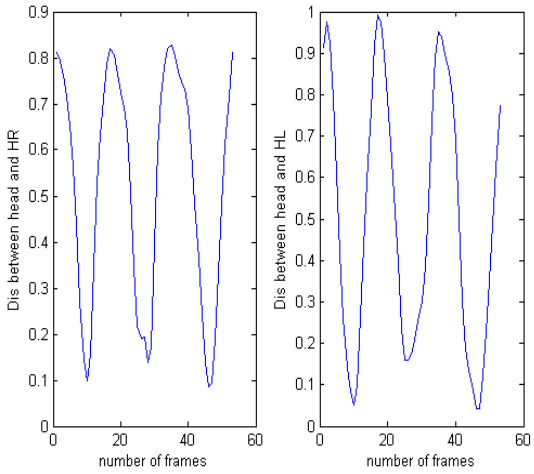

Figure 10: Distance Between The Head And The Right Hand During The "Wave Hand" Action Sample; Distance Between The Head And The Left Hand During The "Wave Hand" Action Sample

The Figure 11 displays the values of "correlation" for the whole dataset:

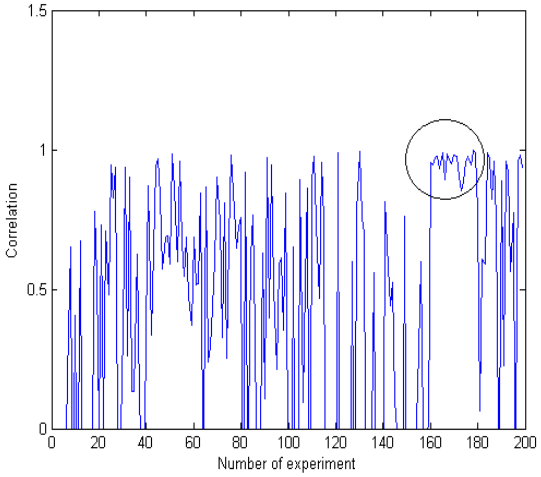

Figure 11: "Correlation" Feature Values For The Whole Data Set

The black ellipse in figure 11 contains the samples of the "wave hands" action.

"Periodicity" feature is also a distinctive feature for the "wave hand" action. Figure 12 displays the values of "periodicity" for the whole dataset:

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

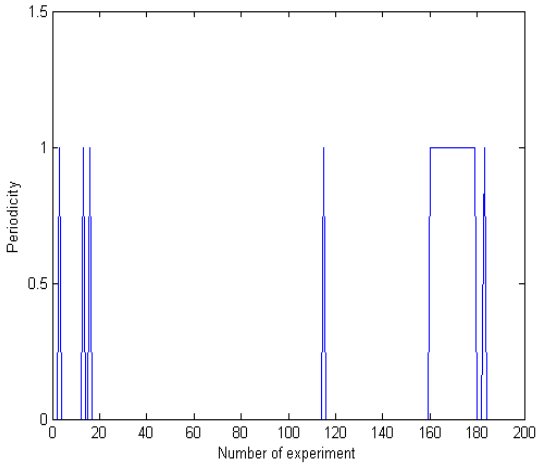

Figure 12 "Periodicity" Feature Values For The Whole Data Set

Based on the figure, it is evident that all samples of the "wave hand" action obtain the value of 1. It is also noted that some samples of the "walk" action obtain the value of 1 for this feature. This is because some humans move their hands back and front in a periodic way when they walk.

"Highness" feature is a very distinctive feature for the "clap hands" action and it obtains high values for all samples of this action as demonstrated in Figure 13.

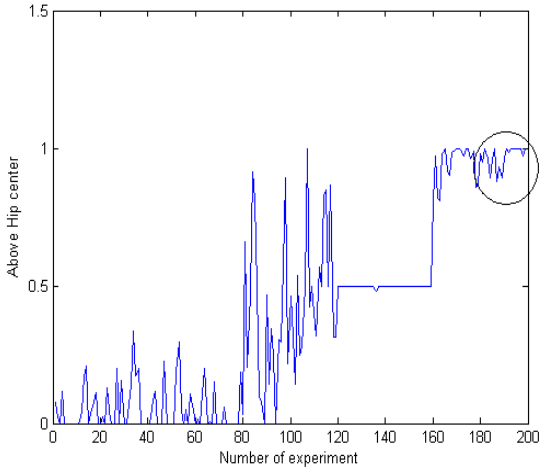

Figure 13: "Highness" Feature Values For The Whole Data Set

The black circle represents the samples of the "clap hands" action.

The "touch hands" feature is a very discriminative feature for both the "carry" and "clap hands" actions as shown in Figure 14:

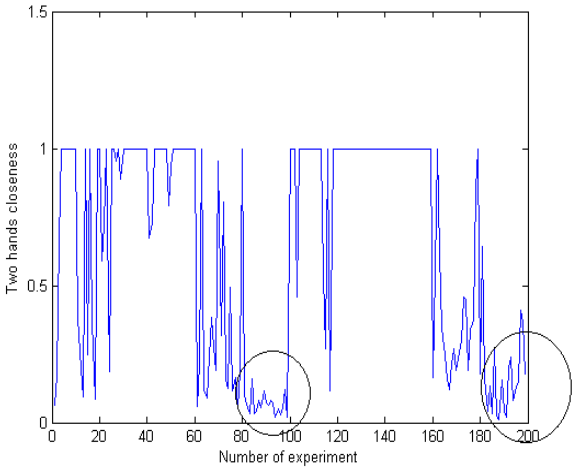

Figure 14 "Touch Hands" Feature Values For The Whole Data Set

It is obvious that all samples of these two actions obtain a lower value for this feature.

3.3 Classification Results

In order to classify the records, ELM is used to train a single hidden layer feed forward neural network with 11 inputs (11 is the number of features) and 18 neurons in the hidden layer. As mentioned, the classification results are divided into two parts:

a- Without including the "throw" action in the training and testing process:
Three scenarios were selected for the training and testing of data. The following Table 4 illustrates these scenarios:

Table 4: Results Of Different Scenarios Classification Without Including The "Throw" Action

Scenario	Training Accuracy	Testing Accuracy
All subjects were used for training and testing where the first experiment was used for training and the second experiment was used for testing	100 %	100 %
Subjects 1 3 5 7 9 were used for training whatever the experiment is. Subjects 2 4 6 8 10 were used for testing whatever the	100 %	100 %

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

experiment is.		
Subjects 1 2 3 4 5 were used for training whatever the experiment is. Subjects 6 7 8 9 10 were used for testing whatever the experiment is.	100 %	100 %

b- Including the "throw" action in the training and testing process:
Three scenarios were selected for the training and testing of data. However, the number of neurons of the hidden layer was changed to 36 in the following table. Table 5 illustrates these scenarios:

Table 5: Results Of Different Scenarios Classification With The Inclusion Of The "Throw" Action

Scenario	**Training Accuracy**	**Testing Accuracy**
All subjects were used for training and testing where the first experiment was used for training and the second experiment was used for testing	97 %	92,93 %
Subjects 1 3 5 7 9 were used for training whatever the experiment is. Subjects 2 4 6 8 10 were used for testing whatever the experiment is.	98 %	92,93%
Subjects 1 2 3 4 5 were used for training whatever the experiment is. Subjects 6 7 8 9 10 were used for testing whatever the experiment is.	99 %	91,92%

The following matrices are the confusion matrices of the first scenario with the inclusion of "throw action" as shown in Table 6:

Table 6: Confusion Matrices Of The First Scenario.

Testing Confusion Matrix									
0.9	0	0	0	0.1	0	0	0	0	0
0	0.9	0	0.1	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0
0	0	0	0	0.9	0.1	0	0	0	0
0	0	0	0	0	0.6	0.1	0.2	0	0.1
0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	1

The following matrices are the confusion matrices of the second scenario with the inclusion of the "throw action" as shown in Table 7:

Table 7: Confusion Matrices Of The Second Scenario.

Testing Confusion Matrix									
0.9	0.1	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0
0	0.1	0	0	0	0.4	0	0.2	0	0.3
0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	1

The following matrices are the confusion matrices of the third scenario with the inclusion of the "throw actions" as shown in Table 8:

Table 8: Confusion Matrices Of The Third Scenario.

Testing Confusion Matrix									
0.9	0	0	0	0	0	0.1	0	0	0
0	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0
0	0.1	0	0.9	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0.4	0.1	0.2	0	0.3
0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	1

As observed from the results, it is evident that the extracted features along with the extreme learning machine play an important role in the performance of the human activity recognition system. The system is capable of performing

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

recognition with an accuracy of 100% for 3 different scenarios without including the "throw" action. The accuracy was more than 91% for all scenarios when the "throw" action was included. This is interpreted by the high similarity between the "throw" and "push" actions which makes distinguishing between the two actions difficult even for human observation. For further validation, the approach used in this study was compared to two other approaches which are HOJ3D (Xia, 2012) and Spatio-temporal feature chain for skeleton-based human action recognition (STFC) (Ding, Liu, Cheng, & Zhang, 2015), See table 9.

Table 9: Comparison With HOJ3D (Xia, 2012) And (STFC) (Ding, Liu, Cheng, & Zhang, 2015).

Method	Walk	Sit	Stand	Pickup	Carry	Throw	Push	Pull	Wave	Clap
HOJ 3D	96 %	91 %	93 %	97 %	97 %	59 %	81 %	92 %	100 %	100 %
STFC	90 %	95 %	95 %	100 %	65 %	90 %	95 %	100 %	100 %	85 %
Ours	90 %	90 %	100 %	100 %	90 %	60 %	100 %	100 %	100 %	100 %

Table 9 proves that our methods outperformed other methods for most actions. The novelty of our approach comes from the fact that this decision making is based on an accurate feature space designed based on the actions itself. Moreover, our approach can be scalable to a wider set of actions because the more features with a discriminative power can be added when the set of actions is extended to other types.

4. FUTURE WORK AND CONCLUSION.

Action-based features are incorporated in the HAR system along with extreme learning machine classification. The system shows 100% accuracy when nine actions were tested and more than 91% accuracy when ten actions were tested with two highly similar actions. Future work is to extend the vector of action-based features by including more features in order to guarantee robustness to highly similar actions and to validate it big data set with numerous number of actions.

REFRENCES:

[1] Das, D. Activity Recognition Using Histogram of Oriented Gradient Pattern History. *IJCSEIT International Journal of Computer Science, Engineering and Information Technology,* vol. 4, No. 4, 2014, pp.23-31.

[2] Ding, W., Liu, K., Cheng, F., & Zhang, J. STFC: Spatio-temporal feature chain for skeleton-based human action recognition. *Journal of Visual Communication and Image Representation*, Vol. 26, 2015, pp. 29-337.

[3] Kliper-Gross, O., Gurovich, Y., Hassner, T., & Wolf, L. Motion Interchange Patterns for Action Recognition in Unconstrained Videos. *Computer Vision – ECCV 2012 Lecture Notes in Computer Science,* 2012, pp.256-269.

[4] Mokhber, A., Achard, C., & Milgram, M. Recognition of human behavior by space-time silhouette characterization. *Pattern Recognition Letters,* Vol. 29, No. 1, 2008, pp.81-89.

[5] Nguyen, T. P., Manzanera, A., & Garrigues, M. Motion Trend Patterns for Action Modelling and Recognition. *Computer Analysis of Images and Patterns Lecture Notes in Computer Science,* 2013, pp.360-367.

[6] Poppe, R. A survey on vision-based human action recognition. *Image and Vision Computing,* Vol. 28, No. 6, 2010, pp.976-990.

[7] Scovanner, P., Ali, S., & Shah, M. A 3-dimensional sift descriptor and its application to action recognition. *Proceedings of the 15th International Conference on Multimedia - MULTIMEDIA '07*. 2007.

[8] Sung, J., Ponce, C., Selman, B., & Saxena, A. Unstructured human activity detection from RGBD images. *2012 IEEE International Conference on Robotics and Automation*. 2012.

[9] Xia, L., Chen, C., & Aggarwal, J. K. View invariant human action recognition using histograms of 3D joints. *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. 2012.