Journal of Engineering and Applied Sciences 12 (3): 520-526, 2017
ISSN: 1816-949X

Action Classification on the Berkeley Multimodal Human Action Dataset (MHAD)

[1]Hu Ng, [1]Timothy Tzen Vun Yap, [1]Hau Lee Tong, [1]Chiung Ching Ho,
[2]Lay Kun Tan, [1]Wan Xin Eng, [1]Seng Kuan Yap and [1]Jia Hao Soh
[1]Faculty of Computer and Informatics, Multimedia University,
Jalan Multimedia, Cyberjaya, Malaysia
[2]Sekolah Menengah Seksyen 3, Bandar Kinrara, Malaysia

Abstract: The objective of this study is to classify multimodal human actions of the Berkeley Multimodal Human Action Database (MHAD). Actions from accelerometer and motion capture modals are utilized in this study. Features extracted include statistical measures such as minimum, maximum, mean, median, standard deviation, kurtosis and skewness. Feature extraction level fusion is applied to form a feature vector comprising two modalities. Feature selection is implemented using Particle Swarm Optimization (PSO) Tabu and Ranker. Classification is performed with Support Vector Machine (SVM) Random Forest (RF) k-Nearest Neighbour (k-NN) and Best First Tree (BFT). The classification model that gave the highest accuracy is support vector machine with radial basis function kernel with a Correct Classification Rate (CCR) of 97.6 % for the Accelerometer modal (Acc) 99.8% for the Motion capture system modal (Mocap) and 99.8% for the Fusion Modal (FusioMA). In the feature selection process, ranker selected every single extracted feature (162 features for Acc and 1161 features for Mocap and 1323 features for FusioMA) and produced an average CCR of 97.4%. Comparing with PSO (68 features for Acc, 350 features for Mocap and 412 features for FusioMA) it produced an average CCR of 97.1% and Tabu (54 features for Acc, 199 features for Mocap and 323 features for FusionMA) produced an average CCR of 97.2%. Although, Ranker gave the best result, the difference in the average CCR is not significant. Thus, PSO and Tabu may be more suitable in this case as the reduced feature set can result in computational speedup and reduced complexity. The extracted statistical features are able to produce high accuracy in classification of multimodal human actions. The feature extraction level fusion to combine the two modalities performs better than single modality in the classification.

Key words: Human action classification, multimodal, feature extraction, feature selection, machine learning

INTRODUCTION

Human action or activity is denoted as a regular atomic movement performed by an individual (Lan *et al*., 2010). Human action recognition system is important for surveillance to track the threats of terrorists and for ambient assisted living to support independent living and ageing. In addition, it is can also be used to assist the sick and disabled, for example, monitoring patient activities in home-based rehabilitation.

Human action recognition system is typically divided into two types, sensor-based or vision-based. Sensor-based recognition employs sensors that are attached to human body while vision-based recognition captures and identifies activities on video taken by various cameras. In this research we make use of the action data from accelerometers and motion capture devices that are affixed to the human body. The data is obtained from the Berkeley Multimodal Human Action Database (MHAD) (Ofli *et al*., 2013). From this multimodal human action database from which we intend to develop a model that extract features employing statistical parameters, consisting of minimum, maximum, mean, median, standard deviation, kurtosis and skewness of the motion data. We also employed feature extraction level fusion to form a feature vector comprising two modalities of action data to investigate its effect on the classification performance. The fused feature vector will consist of greater data points in the feature space which is beneficial for the training process.

Selection of features is implemented by applying three feature selectors, specifically Particle Swarm Optimization (PSO) Tabu and Ranker. The output from the selected features is analyzed to find which data points contributed in differentiating various human actions. Lastly the selected features are exploited to classification

Corresponding Author: Hu Ng, Faculty of Computer and Informatics, Multimedia University, Jalan Multimedia, Cyberjaya, Malaysia

process to validate the implementation of the proposed model in identifying human actions by employing four classification techniques such as BFT, RF, SVM and k-NN.

Literature review

Human action recognition systems: The development of a human action recognition system is multifaceted and normally involves four major tasks as listed (Chen *et al*., 2012):

- To find the appropriate instruments to capture subject' behavior
- To find the suitable data analysis techniques for data collection, processing and storage
- To find the appropriate the computational activity models that permits software or agents to perform analysis and exploitation
- To expand reasoning methods to figure out actions from captured data

In general, human action recognition is group into two clusters, namely sensor-based and vision-based. Vision-based Action Recognition System (VARS) recognizes action from a video or image scenes. It requires visual sensing equipments likes single or multiple video cameras to perceive an individual's activity or motion. The cameras captured video progressions or digitized visual data. Then computer vision techniques such as movement tracking, movement segmentation, action extraction feature extraction and structural modeling were applied to analyze visual observations for action recognition. Aggrawal and Park (2004) divided video sequence analysis methods into static and dynamic representations. The static representation analyzed the individual frames and combined them into the video sequences. The dynamic representation analyzed the entire sequences of the captured video (Aggrawal and Park, 2004). Bobick and Davis (2001) applied temporal templates and dynamic matching in time to measure the sensitivity of the action representation.

Sensor-based Action Recognition System (SARS) is dependent on the emerging sensor network technology. The generated sensor data is in time series of state changes and several parameter values. The captured values are then being processed through data fusion, probabilistic and statistical analysis approach for feature extraction. Sensors can be affixed to an individual such as wearable sensors (Lim *et al*., 2015) accelerometers (Ofli *et al*., 2013) and smart phones (Arora *et al*., 2015; Ho *et al*., 2013). Wearable sensors used inertial measurement units and Radio-Frequency Identification (RFID) tags to compute the motion information. The accelerometers attained the action data and conveyed data through the Bluetooth protocol to the acquisition computer where time stamps were employed to each collected frame. Smart phones recorded the gravimetric acceleration experienced by the subjects as they performing actions.

SARS is effective in recognizing physical movements compared with VARS as SARS is not affected by the surrounding environment such as light illumination and the occlusion by humans or objects. Furthermore, it does not require any camera calibration and lighting adjustment with the surrounding environment. SARS is able to perform continuous monitoring of an individual.

In SARS, there are three major ways to extract features from the raw data. The first direction uses standard statistical measure such as maximum, mean, median and standard deviation (Arora *et al*., 2015). The second direction is to employ filter banks such as Fast Fourier Transform (FFT) (Huynh and Schiele, 2005) and wavelets (Najafi *et al*., 2003). The third direction is performed using established feature selection techniques such as Principal Component Analysis (PCA) (Lukowicz *et al*., 2004) and Linear Discriminant Analysis (LDA) (Mantyjarvi *et al*., 2001).

Standard statistical analysis: Standard statistical analysis is being widely used by many researches to extract features from raw data. This is because the raw data collected from sensors or accelerometers is lengthy and often in the format of time series. Lim *et al*. (2015) employed minimum, maximum, median, kurtosis, mean, skewness, standard deviation and absolute deviation of the walking data to extract the gait features in Parkinson's disease analysis. Arora *et al*. (2015) employed mean, standard deviation, median to extract gait motion features from the acceleration data captured by smartphones (Arora *et al*., 2015). In fault diagnosis of the hydraulic brake system by Jegadeeshwaran and Sugumaran (2015) they utilized minimum, standard error, sample variance, kurtosis and skewness to extract the acquired vibration signals from the hydraulic brakes in automobiles (Jegadeeshwaran and Sugumaran, 2015). In the research of finding the similarity of the Inter-Onset-Intervals (IOI) of musical rhythm (Beltran *et al*., 2015) they calculated the rhythms statistical features of the IOI in term of minimum, maximum, range, standard deviation and normalized pair-wise variability index.

Fusion techniques: Fusion is defined as the expansion of multiple types of actions data or schemes of dealing with the intention to improve the performance of the pattern recognition. There are many fusion techniques that can be employed to fuse the features. The most prominent fusion techniques are match score level, feature extraction level and decision level (Ho *et al*., 2015). Match score level fusion generates the feature vectors independently from each modality and then associate them to the enrollment templates. By referring to the likeness of feature vector and the template, each sub modal calculates individual matching score value. Then, these individual scores are joined to find a total score and dispatched to the decision module. This method is widely employed by many research works as it is relatively easy to be implemented. However, the information acquired is normally limited and may cause poorer performance.

Feature extraction level fusion concatenates the feature vectors from their respective modals into a single feature vector. This method is effective if the features joined are in the same type of measurement scale and independent of each other. Normally, feature selection techniques are employed if the joined feature vector is large. Decision level fusion combines the individual authentication decision which has been generated from each modal to compute a final vote. This method is considered inflexible when compared to the other fusion methods due to the availability of limited information from each individual modality.

MATERIALS AND METHODS

This study describes the methodology of the study. Human action data collection and preparation, extraction of features from raw data, selection of extracted features and classification methods are expanded in the subsequent sub-sections.

Multimodal human action dataset: Berkeley Multimodal Human Action Database (MHAD) was developed by Teleimmersion Lab at University of California and Center for Imaging Science, Johns Hopkins University in 2013. MHAD aims to offer researchers a comprehensive platform to build and benchmark new methods across multiple modalities in recognizing human movements and activities (Ofli *et al*., 2013).

MHAD is a controlled multimodal human action dataset that consist of five coordinated and geometrically calibrated human motion data. The dataset consists of five modals which are motion capture, multi-view stereo cameras, Kinect's depth sensors, accelerometers and microphones.

MHAD consists of 11 actions made by 12 subjects (7 male and 5 female) in the age range between 23-30 year and one aged subject. All the subjects performed five repetitions of each action, generating 660 action sequences which made up 82 min of capturing duration. The actions are listed as:

Table 1: List of statistical characteristics

Statistical parameters	Specification
Minimum	Smallest value in a range of values
Maximum	Biggest value in a range of values
Standard deviation	The square root of variance by calculating the difference between each data point relative to the mean
Median	The median value of a range of values
Average	The arithmetic mean of a range of values
Skewness	Calculation of the degree of asymmetry of range of values set around its mean
Kurtosis	The highest value of the normal distribution of a range of values
Absolute Kurtosis	Absolute value of the skewness
Absolute Skewness	Absolute value of the kurtosis

- Jumping in place-the standing subject jumps off the ground with minimal arm movement and return to the original position
- Jumping jacks-the standing subject jumps off the ground with the legs spread out and the arms over the head then returns to a position with the feet together and the arms at the sides
- Bending-the standing subject inclines the body downward with hands up all the way down
- Punching-the subject performs a forceful hit with a fist
- Waving (two hands)-the subject swings the both arms over the head then returns the arms at the sides
- Waving (right hand)-the subject swings the right arm over the head then returns the arm at the side
- Clapping hands-the subject hits the palms of hands together
- Throwing a ball-the subject move out a ball of the hand
- Sit down then stand up: the standing subject sits on a chair and returns to a standing position
- Sit down-the standing subject sits on a chair
- Stand up-the sitting subject rises to a standing position

Features extraction: For this study, two modalities (accelerometer and motion capture system) from MHAD were utilized for the proposed action classification model. For the accelerometer modal, six 30 Hz three-axis wireless sensors were attached to important linkages on the human body for motion capture. The sensors were placed at wrists, hips and ankles. Each sensor provided acceleration data in X, Y and Z axes. For motion capture modal, 43 LED markers acts as the data points that provide 3D body positions in X, Y and Z axes.

To achieve the optimized results, the major statistical parameters of the sample distribution of the raw features from both modals were extracted to develop nine new features. Table 1 shows the nine statistical parameters. Consequently, the total number of 162 ($6\times3\times9$) and 1161 ($42\times3\times9$) features were extracted from Accelerometer modal (Acc) and Motion capture modal (Mocap), respectively.

Normalization and fusion of extracted features: Feature normalization is performed before the extracted features are utilized in classification. It normalizes numerous dimensions of extracted features to ensure that they are regulated and not biased. As the features extracted from accelerometers and motion capture system are not in the same measurement scale, feature normalization is performed. This is to avoid situations where more weights are allocated implicitly to features with bigger values than those with smaller values. Thus, the circumstance of biasing towards a particular feature can be prevented. Linear scaling is utilized in this study to scale the feature values to the range between 0 and 1.

To evaluate feature extraction level fusion on the performance of the classification, a new feature vector is generated through the concatenation of the normalized extracted features from the two modalities (Acc and Mocap).

Features selection techniques: To decrease the dimensionality of extracted features, selection of features is accomplished before proceeding to the classification process. Basically, a good feature selector must meet the two principles as listed below (Peng *et al*., 2005):

- Maximum relevance elected feature ought to correlate strongest to the target variable
- Minimum redundancy elected feature ought to be maximally dissimilar from each other

In this research, PSO, Tabu and Ranker were employed to acquire the features that offered positive contribution to the classification process. There are employed as they have not been utilized on human action classification, although they have been found to research well in pattern recognition.

PSO organizes a population of random potential result, in particular phrased as particles and is flown through the problem feature space (Moraglio *et al*., 2008). PSO seek for optima convincing result or the highest classification rate.

Tabu performs a seeking through the entire of feature subsets (Hedar *et al*., 2006). Tabu avoids local maximums by compliant bad and diverse solutions. It search progression will be stopped when there is no new development in the iterations.

Ranker positions the features according to correlation attribute evaluator (Hall and Holmes, 2003). It assesses an attribute by determining the correlation concerning the feature and the class. It considers each unique feature as the specific importance indicator on a merit principal. The absolute correlation for a nominal feature is reputable by a weighted arithmetic mean.

Classification techniques: Those normalized features that are representation of a subject's motion will be converted into a feature vector. After that, the classifier will construct decisions of which class (actions) of an incoming set of feature vector belongs to. For the performance assessment of our research, four classification techniques were employed, k-NN with Euclidean distance metrics, BFT, SVM and RF were utilized, as they have been widely employed in many classification research.

BFT is a standard decision tree learner contains the internal nodes and terminal nodes (Shi, 2007). Each internal node stands for a selection from a set of alternatives. The terminal node is determined during the classification process. In BFT, the best node is developed first. The best node will be parted adhere to maximum reduction of impurity among all existing nodes for splitting. The final tree is similar when complete developed, only dissimilarity in the ordering of the sequence. Actually, several branches from a complete-developed tree are not replicating the primary data in the domain. The parameters are the smallest number of instances of the terminal nodes (M) and the folds number of internal cross-validation (N). The value of seed (seed) is fixed to one.

The k-NN is considered as the minima parametric classifier to differentiate classes based on the training information in the feature space (Altman, 1992). This technique manipulates the whole collected features to validate its records. k-NN classified an unlabeled class from the data of its k nearest neighbors collected in the memory. Whereby, classes are classified based on the majority poll of nearest neighbors in the training information. For this research, the value of neighbor, k is utilized to calculate a relative measurement.

SVM is based on the non-linearly mapping by the input vectors in a very high dimension feature space (Cortes and Vapnik, 1995). As a result, a linear decision surface is formed together by its ability of the distinguished creation of a network. Three kernels were used in this work, namely Linear (Ln) Polynomial (Poly) and Radial basis function (Rbf) kernels. The parameters for Ln kernel is Cost (C) Poly kernel are C, Gamma (G), Degree (D) and Coefficient (Coef) Rbf kernel are C and G. RF is a classifier that contains of a progression of classification trees (Breiman, 2001). In the forest, all tress have the similar distribution. The final classification of hidden entities is relying on the largest votes across all the trees in the forest. The parameters utilized are number of Attributes (At) number of execution Slots (S) and number of Iterations (I).

Table 2: The parameter values for each classifier

Classifiers	Parameters
BFT	M = 2, N = 4
RF	At =1, S = 2, I = 200
k-NN	k = 1
SVM with Ln kernel	C = 32
SVM with Poly kernel	C = 256 , G = 0, D = 3, Coef = 1.4
SVM with Rbf kernel	C = 256, G = 0.5

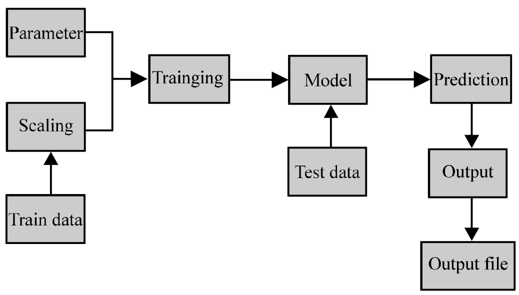

Fig. 1: The flow of the processes in architecture of experiments

Performance evaluation: This research used ten folds cross validation in corroborating the classification models where the generated feature vector were arbitrarily partitioned into ten equal size disjoint subsections. Every subset in turn is employed as the validate test set while the remaining subsets are employed as to training set. The cross-validation development was repetitive for ten rounds where all the features vectors of each disjointed subsection have been regulated into classes during the validation test. Then, the classification rate is computed by mean of the cross validation outputs.

Correct Classification Rate (CCR) is employed as the evaluation measure for the classification. CCR is referred as the percentage of the value of actions being accurately labeled divided by the total actions in the dataset.

Architecture of experiments: The experiments are conducted in two sections: training and testing. Two modals, namely Accelerometers (Acc) and motion capture (Mocap) with 660 instances each from MHAD for were utilized during the training and testing sections (162 and 1161 features from Acc and Mocap, respectively) (Ofli *et al*., 2013).

Experiments were also conducted by merging the extracted features from both modals to generate a new fusion modal (by employing feature extraction level fusion) labeled as FusionMA, consisting of 1323 features (concatenation of 121 and 1161 features). Figure 1 demotivates the flow of the steps involved in the experiments. The training was conducted to find the models for classification by optimizing parameters for each classifier. Imitative from the heuristic results obtained from the training, the parameter values were determined as the models for testing section which are shown in Table 2. In the testing section, the model obtained from the training section is utilized for action classification for the MHAD dataset.

RESULTS AND DISCUSSION

Performance of feature selectors and classifiers: To facilitate the implementation of the proposed model on the MHAD dataset, several experiments were carried out. This part displays and deliberates the results which were intended to measure the finding from classification process. Three features selectors; PSO, Tabu and Ranker were employed to choose the constructive extracted features. Four classification techniques; BFT, RF, k-NN and SVM were utilized to obtain CCR and to authenticate the reliability of the results. Table 3 summarizes the overall classification results for the classifiers in conjunction with the feature selectors for all three modalities. Table 4 shows the average CCR for the three feature selectors for each classifier.

Based on the result in Table 3, it can be observed that SVM outperforms all the other three classification techniques. The classifier that gave the highest accuracy is SVM with Rbf kernel where the average CCRs are 99.7, 97.2 and 99.6% for the three modalities. As expected, the Rbf kernel produced better CCR as it usually superior over Poly kernel and linear kernel (Byun and Lee, 2003). This is because of the generated action feature vectors from this work are not linear; the kernel trick in the Rbf kernel permits the algorithm to adjust the maximum-margin hyper-plane in a transformed feature space (Suykens *et al*., 2002).

In general, it can be proven that by combining the features from two modalities with extracted features level fusion performs better than single modality. This is due to the features extracted from one modality complement the drawbacks of the features extracted from other modality (Ofli *et al*., 2013).

From Table 4 it can be found that the best feature selector is Ranker where all the extracted features were chosen (162 features for Acc, 1161 features for Mocap and 1323 features for FusioMA). Comparing with PSO (68 features for Acc, 350 features for Mocap and 412 features for FusioMA) and Tabu (54 features for Acc, 199 features for Mocap and 323 features for FusionMA) the number of selected features has been reduced significantly, thus reducing the dimensionality of the data. Although Ranker gives the best result, the difference in the average CCR is very insignificant. Thus, PSO and Tabu may be a more suitable feature selector in this case as the reduced

Table 3: Classification results for the three modalities

		CCR (%)		
Classifiers	Feature selectors	Mocap	Acc	Fusion MA
BFT	Ranker	91.5	88.6	93.8
	PSO	92.0	88.2	93.8
	Tabu	92.4	87.4	93.3
RF	Ranker	99.3	97.9	99.3
	PSO	99.3	97.9	99.3
	Tabu	99.3	98.0	99.3
k-NN	Ranker	98.1	96.4	98.6
	PSO	97.6	95.4	96.4
	Tabu	96.0	97.3	96.8
SVM with Ln kernel	Ranker	99.6	97.6	99.6
	PSO	99.4	97.0	99.6
	Tabu	99.4	98.0	99.5
SVM with Poly kernel	Ranker	99.3	97.7	99.6
	PSO	99.6	97.0	99.6
	Tabu	99.4	97.1	99.6
SVM with Rbf kernel	Ranker	99.8	97.0	99.8
	PSO	99.7	97.6	99.6
	Tabu	99.6	97.0	99.6

Table 4: Average CCR (%) for each feature selector

Classifier	Ranker	PSO	Tabu
BFT	91.3	91.3	91.0
RF	98.8	98.8	98.8
k-NN	97.7	96.5	96.7
SVM with Ln kernel	98.9	98.6	99.0
SVM with Poly kernel	98.9	98.7	98.7
SVM with Rbf kernel	98.8	98.9	98.7
Average	97.4	97.1	97.2

Table 5: Highest CCR (%) obtained for each classifier across three modalities

Paramiter	Classifier	Mocap	Acc	Fusion MA
Our approach	BFT	92.4	88.6	93.8
	RF	99.3	98.0	99.3
	k-NN	98.1	97.3	98.6
	SVM with Ln kernel	99.6	98.0	99.6
	SVM with Poly kernel	99.6	97.7	99.6
	SVM with Rbf kernel	99.8	97.6	99.8
In^2	k-NN	75.6	81.8	-
	K-SVM	79.9	85.4	-
	MKL	-	-	97.5

feature set can result in computational speedup and reduced complexity. From the selection of the extracted features by the three feature selectors, few features are discovered to be distinct and distinguishable. For the Acc, all features from the sensors located on the upper body were selected for all actions. The same was also observed for the features from the Mocap markers. A reason attributed to this is that almost all the motions make use of the upper part of the body. Features that were excluded were from the sensors on the wrist and hips (Acc). The same was also observed for the markers placed on the ankles (Mocap). A likely reason for this is that actions such as waving hands, punching, clapping hands and throwing a ball did not involve movement of the lower part of the body.

Comparison of results: Table 5 shows the highest CCR obtained from the classifiers employed in this research across all three modalities (Mocap, Acc and FusionMA) as compared with the baseline results from (Ofli *et al.*, 2013). They performed the classification employing three classifiers; k-NN, SVM with the χ^2 kernel (K-SVM) for both Acc and Mocap and Multiple Kernel Learning (MKL) for FusionMA. It can be found that our approach outperforms the results obtained from their research especially on k-NN and K-SVM. In (Ofli *et al.*, 2013) employed data directly from the sensors and the markers (time series) which has a larger dimension and may not be distinct enough.

CONCLUSION

This study performed a statistical model to extract action features from a multimodal action recognition dataset. Numerous experiments have been carried out to evaluate the performance of the proposed model by utilized difference feature selection techniques and classification tools. The proposed model is proven able to classify 11 actions from 12 subjects with high correct recognition rates. We also proved that combining the features from two modalities with extracted features level fusion performs better than single modality. For future research, we intend to use different normalization methods and different fusion techniques to evaluate the performance of fusion of multimodal human action datasets.

ACKNOWLEDGEMENTS

The researchers would also like to thank Prof. Ofli, Teleimmersion Lab at University of California, Berkeley for offering the database in this research.

REFERENCES

Aggrawal, J.K. and S. Park, 2004. Human motion: Modeling and recognition of actions and interactions. Proceedings of the International Symposium on 3D Data Processing, Visualization and Transmission, Sept. 6-9, IEEE Computer Society, Washington DC., USA., pp: 640-647.

Altman, N.S., 1992. An introduction to kernel and nearest-neighbor nonparametric regression. Am. Statistician, 46: 175-185.

Arora, S., V. Venkataraman, A. Zhan, S. Donohue and K.M. Biglan et al., 2015. Detecting and monitoring the symptoms of Parkinsons disease using smartphones: A pilot study. Parkinsonism Relat. Disord., 21: 650-653.

Beltran, J.F., X. Liu, N. Mohanchandra and G.T. Toussaint, 2015. Measuring musical rhythm similarity: Statistical features versus transformation methods. Intl. J. Pattern Recognit. Artif. Intell., 29: 1-23.

Bobick, A.F. and J.W. Davis, 2001. The recognition of human movement using temporal templates. IEEE Trans. Pattern Anal. Mach. Intel., 23: 257-267.

Breiman, L., 2001. Random forests. Mach. Learn., 45: 5-32.

Byun, H. and S.W. Lee, 2003. A survey on pattern recognition applications of support vector machines. Intl. J. Pattern Recognit. Artif. Intell., 17: 459-486.

Chen, L., J. Hoey, C.D. Nugent, D.J. Cook and Z. Yu, 2012. Sensor-based activity recognition. IEEE Trans. Syst. Man Cybernet. Part C: Applic. Rev., 42: 790-808.

Cortes, C. and V. Vapnik, 1995. Support-vector networks. Mach. Learn., 20: 273-297.

Hall, M.A. and G. Holmes, 2003. Benchmarking attribute selection techniques for discrete class data mining. IEEE Trans. Knowledge Data Eng., 15: 1437-1447.

Hedar, A.R., J. Wang and M. Fukushima, 2006. Tabu search for attribute reduction in rough set theory. Soft Comp. A Fusion Found. Methodol. Appl., 12: 909-918.

Ho, C.C., H. Ng, W.H. Tan, K.W. Ng and H.L. Tong *et al*., 2013. MMU GASPFA: A cots multimodal biometric database. Pattern Recognit. Lett., 34: 2043-2050.

Ho, C.C., M.A. Hussin and H. Ng, 2015. Match score fusion of fingerprint and face biometrics for verification. J. Technol., 77: 93-102.

Huynh, T. and B. Schiele, 2005. Analyzing features for activity recognition. Proceedings of the 2005 Joint Conference on Smart Objects and Ambient Intelligence Innovative Context-Aware Services Usages and Technologies, October 12-14, 2005, ACM, ISBN:1-59593-304-2, pp: 159-163.

Jegadeeshwaran, R. and V. Sugumaran, 2015. Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines. Mech. Syst. Signal Process., 52: 436-446.

Lan, T., Y. Wang, W. Yang and G. Mori, 2010. Beyond actions: Discriminative models for contextual group activities. Adv. Neural Inf. Process. Syst., 1: 1216-1224.

Lim, C.M., H. Ng, T.T.V. Yap and C.C. Ho, 2015. Gait analysis and classification on subjects with parkinsons disease. J. Technol., 77: 1-6.

Lukowicz, P., J.A. Ward, H. Junker, M. Stager and G. Troster et al., 2004. Recognizing workshop activity using body worn microphones and accelerometers. Proceedings of the International Conference on Pervasive Computing, April 21-23, 2004, Springer, Berlin, Germany, ISBN:978-3-540-21835-7, pp: 18-32.

Mantyjarvi, J., J. Himberg and T. Seppanen, 2001. Recognizing human motion with multiple acceleration sensors. Proceedings of the 2001 IEEE International Conference on Systems Man and Cybernetics, October 7-10, 2001, IEEE, Finland, Europe, ISBN:0-7803-7087-2, pp: 747-752.

Moraglio, A., C.D. Chio, J. Togelius and R. Poli, 2008. Geometric particle swarm optimization. J. Artif. Evol. Appl., 2008: 1-14.

Najafi, B., K. Aminian, I.A. Paraschi, F. Loew and C.J. Bula *et al*., 2003. Ambulatory system for human motion analysis using a kinematic sensor: Monitoring of daily physical activity in the elderly. IEEE. Trans. Biomed. Eng., 50: 711-723.

Ofli, F., R. Chaudhry, G. Kurillo, R. Vidal and R. Bajcsy, 2013. Berkeley MHAD: A comprehensive multimodal human action database. Proceedings of the 2013 IEEE Workshop on Applications of Computer Vision (WACV), January 15-17, 2013, IEEE, Berkeley, California, ISBN:978-1-4673-5053-2, pp: 53-60.

Peng, H., F. Long and C. Ding, 2005. Feature selection based on mutual information criteria of max-dependency, max-relevance and min-red undancy. IEEE Trans. Pattern Anal. Mach. Intell., 27: 1226-1238.

Shi, H., 2007. Best-first decision tree learning. Ph.D Thesis, University of Waikato, Hamilton, New Zealand.

Suykens, J.A., D.J. Brabanter, L. Lukas and J. Vandewalle, 2002. Weighted least squares support vector machines: Robustness and sparse approximation. Neurocomputing, 48: 85-105.