ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 1, January 2019

A Hybrid Feature Extraction Approach for Human Action Recognition System based on Skeleton Data

Tin Zar Wint Cho[1], May Thu Win[2]

P.G. Student, Department of Information Science and Technology, University of Technology (Yatanarpon Cyber City), Pyin Oo Lwin, Myamar[1]

Lecturer, Department of Information Science and Technology, University of Technology (Yatanarpon Cyber City), Pyin Oo Lwin, Myamar [2]

ABSTRACT: Human action recognition has potential to impact a wide range of applications from surveillance to human computer interfaces to content based video retrieval. In this paper, the human actions are classified in skeleton data from Kinect sensor and the hybrid feature as joint angles representation and joints distance features are extracted. Then, the proposed system also uses the static k-means algorithm to increase the accuracy rate in which the initial K centroids at the first estimates is statically taken instead of using the non-static (traditional) k-means that it takes the randomized starting centroids at all time. Additionally, to improve the performance, artificial neural network (ANN) is applied to determine the class-labels of the human poses, and discrete Hidden Markov Model (HMM) is also used to correctly identify the human actions based on the sequence of known poses. Experimental results with the public dataset UTKinect show that the proposed approach produces good performance results and accuracy rate from the action-class models.

KEYWORDS: Skeleton joint data, Hybrid Features, Static k-means, Artificial Neural network, Hidden Markov model

I. INTRODUCTION

Human activity recognition is an important field for applications such as surveillance, human-computer interface, content-based video retrieval, etc. Although the advancements in the availability and acquisition of video data have increased, the improvement of automated human activity recognition is still limited. Recently, the rapid development of depth sensors (*e.g.* Microsoft Kinect) provides adequate accuracy of real-time full-body tracking with low cost. This enables us to once again explore the feasibility of skeleton based features for activity recognition.

In 3D skeleton-based action recognition, an action is defined using a collection of time series of 3D joint positions (i.e., 3D trajectories). This representation also depends on the different reference coordinate systems and on biometric differences. Therefore, a variety of coordinate system changes is used. The joint angles between any two connected limbs are considered [2] and an action is represented as a time series of joint angles.

Work such as [3] employs “eigenjoints” which are the most important human pose information for action recognition, and it calculates the position difference of the joints of one frame and the first frame, the joints of two succeeding frames, and all the pairs of joints within one frame. Principle Component Analysis (PCA) is applied to extract “eigenjoints” using the concatenated feature vector and finally, a nearest neighbor-based classifier is also used based on the eigenjoints features for action classification.

The tracked human joint positions in a real time dance classification system are used [4]. In this system, the Principle Component Analysis (PCA) based on the upper-body joint positions is applied to estimate the torso surface,

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 1, January 2019

and the human pose is represented by the angles between the torso surface and the positions of limb joint. To distinguish the temporal structure of actions, it also works Fourier transform over time.

In [5] where a histogram of 3D joints (HOJ3D) is used in which the locations of 11 joints are manually selected to get a compact pose symbol that is invariant to the use of left and right limbs. Linear Discriminant Analysis (LDA) is applied to project the histograms and the K discrete states of the HMM classifier are computed. These states may be considered as elements of a key-poses vocabulary. In general, the required movements contain similar subsets of joints even if each subject may perform the same action with a different style. The activated subsets of joints are only used to distinguish among different action classes.

In [6], body parts are characterized as trajectories of difference vectors between the skeleton joints and a multi-part modeling of the body is adopted. The coordinates of each joint are expressed in a local reference system, and body sub-parts are aligned distinctly and a modified nearest-neighbor classifier is applied to perform action classification by learning the most informative body parts.

Another related representation is proposed in [7], where the body pose is regarded as a continuous and differentiable function of the joint locations over time. Under this statement, it is possible to locally approximate such pose function by using second-order Taylor approximation in a window around the current time step. In this way, the local body pose can be completely represented by means of the current joint locations and differential properties like speed and acceleration of human body joints.

In [8], a human skeleton data is interpreted as a graph and its edge weights based on the distances between all of the joint pairs are computed. Additional edges are also included to link consecutive joints. Later, the entire sequence of skeleton is represented as a spatio-temporal graph. A pyramidal representation based on spectral graph wavelet transform (SGWT) [9] is used to collect joints' trajectories at different scales (in time and space). PCA is also applied to make the high dimensionality representation and the standard SVM is used for action classification task.

Another attempt among the related joints information is represented by using the covariance matrix of a skeleton sequence [10]. In this system, a sequence of joints using a fixed length temporal window is also characterized by means of a covariance matrix that encodes the shape of the joint probability distribution with the set of random variables. A hierarchical representation is implemented by learning the sequential joint dependencies, and finally, action classification is accomplished by linear SVM.

The key difference between the methodology [11] and other systems using HMM is that the emission probabilities are replaced by deep neural networks which contain several feature layers [12]. In [13], only the joints of the left-hand, the right-hand and the pelvis are used to extract relevant features for the actions classification. A convolutional neural network classifier is also applied in which an alternating structure of convolution and subsampling layers is used for classification.

II. PROPOSED SYSTEM

Human action recognition, using visual information in a given image or sequence of images, has been an active area of research in computer vision community. Due to the large diversity of human body, size, appearance, posture, motion, clothing, view angle, camera motion, and illumination changes, besides the complexity of human actions, the task of recognizing human actions is very challenging. A key issue is that which features is more informative for this task.

In this system, the target contribution would be to overcome the above mentioned problems by defining the hybrid features extraction approach and the static k-means algorithm which takes the static centroids at the first time and reduces the random centroids at all time to increase the accuracy of postures selection. Furthermore, a supervised neural network is used to define the labels for each posture and the hidden Markov model is applied to recognize the action as correctly as possible. The fundamental actions (“walking”, “sitting”, “standing”, and “bending”) are recognized in this system. The overview structure of the proposed system is shown in Fig. 1.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 1, January 2019

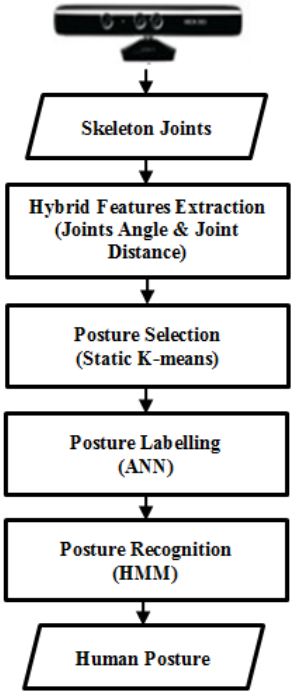

Fig.1. System Design of the Proposed System

A. Hybrid Features Extraction (Joint Angle and Joint Distance)

For the human silhouette-based feature representation and extraction, hybrid feature technique based on joint angles features and joint distance features. To calculate the joint angles features, three joint-pairs - the spine with respect to hip-center, the left-knee with respect to left-hip, and the right-knee with respect to right-hip are considered. The angular feature is defined using eq. (1), eq. (2), and eq. (3).

$$base = \sqrt{(j1(x) - (j2(x))^2 + (j1(y) - j2(y))^2} \quad \text{eq. (1)}$$

$$height = j2\,(z) - j1(z) \quad \text{eq. (2)}$$

$$\mathrm{A\,(f)} = \tan^{-1}\left(\frac{height}{base}\right) \quad \text{eq. (3)}$$

where each j is the joint position of each joint-pair which contains the 3D coordinates. To compute the joint distance features, two joint-pairs are used which are the distance parameters between ankles, and the distance features between head and ankle. These distance features based on the squared Euclidean distance can be formulated using eq. (4).

$$\mathrm{D\,(f)} = \sqrt{(j1(x) - (j2(x))^2 + (j1(y) - j2(y))^2 + (j1(z) - j2(z))^2} \quad \text{eq. (4)}$$

The hybrid feature scheme deals with joint angular information and joint distance representation in overall sequence of each and also provides discriminative features to significantly distinguish each action. This feature is presented as the following eq. (5).

$$\mathrm{F} = (\mathrm{A\,(f)} + \mathrm{D\,(f)}) \quad \text{eq. (5)}$$

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 1, January 2019

where A (f) is the joint-angular features and D (f) is the joint-distance features respectively.

B. Posture Clustering

In this phase, all of the redundant similar poses are reduced to increase the simplification of the postures representation by using the well-known clustering method based on the metric with squared Euclidean distance. Example of a repeated sequence of posture for the "Walking" action is shown in Fig. 2.

In this system, instead of using the non-static (traditional) k-means algorithm that takes the random centroids at the initial estimates at all time, the static k-means technique is applied which takes the static initial centroids to improve the performance. This algorithm processes N feature vectors [f_1, f_2, f_3, . . ., f_N] and groups together into K clusters for each vector [C_1, C_2, C_3, . . . , C_K] in which the centroids of each cluster is presented. Five cluster-identifiers are only used in this system.

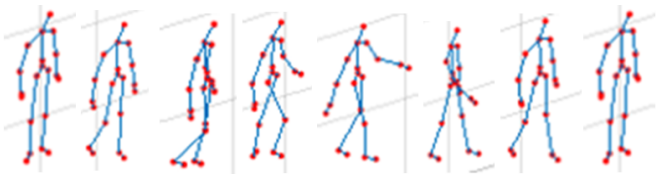

Fig.2. Example of a posture sequence with repetition of the "walking" action

C. Posture Classification

After the sequences of similar pose have been removed using k-means clustering algorithm, the artificial neural network (ANN) is applied which makes the proposed system more intelligent and correctly defines class-labels of each human pose. In the input layer, 20 skeleton joint coordinates are contained, and there are seven classes in the output layer. There are also two hidden layers in which first layer have 40 nodes and second layer have 30 nodes respectively. A flow structure of an ANN applied in this system is shown in Fig. 3.

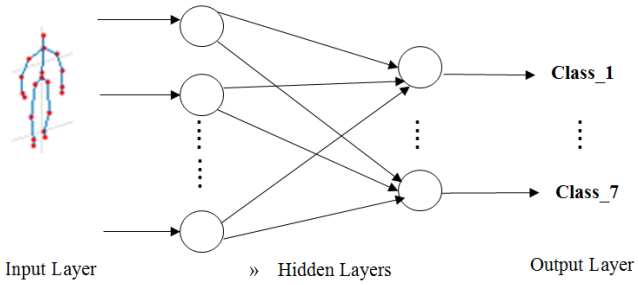

Fig.3. A flow of an ANN in the proposed system

D. Posture Recognition

In the recognition phase, after the postures are labeled by means of ANN, the discrete Hidden Markov Models (HMMs) is also applied to correctly recognize the various occurrences of the posture sequence. The unknown state sequences using HMM based on a known observation sequence are established. The given observed human posture sequences are obtained from the previous steps.

The elements of a Hidden Markov Model (λ) are as follow eq. (6):

$$\lambda = \{N, M, A, B, \pi\} \qquad \text{eq. (6)}$$

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 1, January 2019

where *N*, is the amount of states of the Markov process; *M*, is the amount of discrete output symbols for the Markov process; *A*, is the transition probability matrix between states of the Markov process; $A = \{a_{ij}\}$, from the state S_i to the state S_j , is defined using eq. (7).

$$a_{ij} = P[S_j \mid S_i], \ 1 \le i, j \le n. \qquad \text{eq. (7)}$$

and, *B*, is the emission probability for output symbols per state of the Markov process; $B = \{b_j(k)\}$ is computed using eq. (8).

$$b_j(k) = P[u_k \mid S_j], \ 1 \le j \le n, \ 1 \le k \le R. \qquad \text{eq. (8)}$$

and π , is the probability of starting at a certain state of the Markov process; $\pi = \{\pi_i\}$ is presented as the following eq. (9).

$$\pi_i = P[S^i_{ti}], \ 1 \le i, t \le n \qquad \text{eq. (9)}$$

where π_i is the probability for the initial state of a state sequence that are assumed as the equal probability distribution and n is the number of different states.

Fig. 4 shows the configuration of the HMM for the proposed system. To train and test different actions, the several classes of joint features are applied to the Hidden Markov Models (HMM) to encode each action as a sequence of the postures and build a discrete HMM respectively.

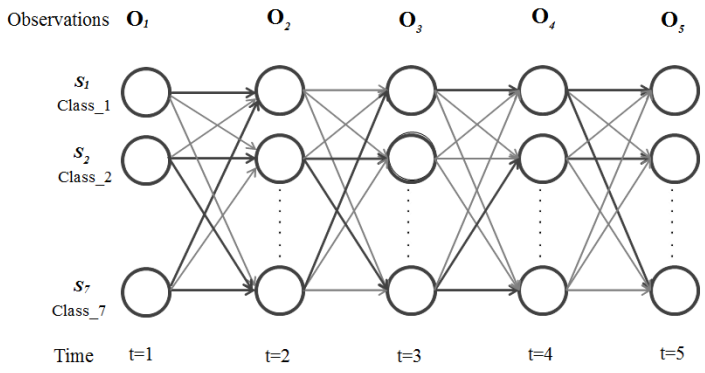

Fig.4. The structure of an HMM used in this system

After the corresponding HMM has been trained on the given posture sequences of each human action, a new posture sequence of an action is tested compared to the training set of HMMs and well recognized based on the largest posterior probability from all of these models. The overview process of the human action recognition system in the training and testing phase are shown in Fig. 5 and Fig. 6 respectively.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 1, January 2019

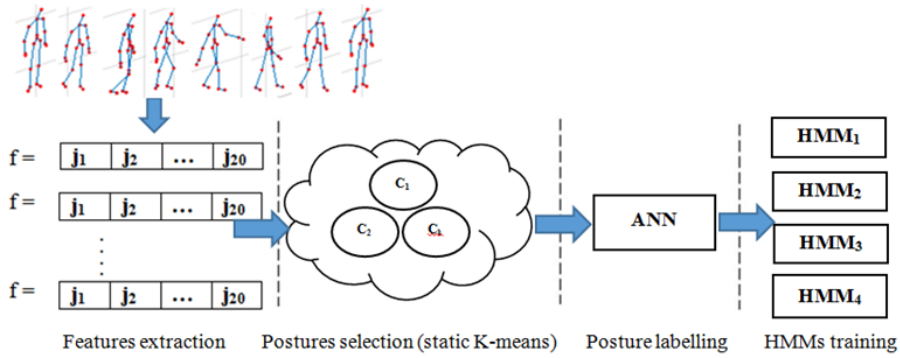

Fig.5. Human Action Recognition Process in Training Phase

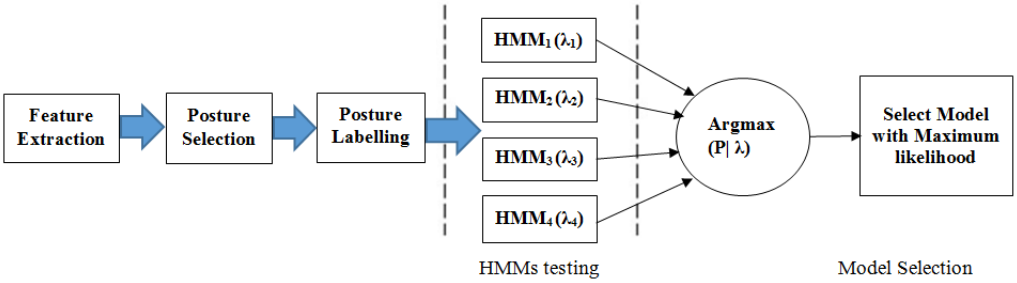

Fig.6. Human Action Recognition Process in Recognition Phase

III. EXPERIMENT

The proposed approach has been evaluated on the public dataset UTKinect-Action3D recorded by the Kinect sensor. There are three channels (skeleton joint positions, color, and depth) in this dataset. It contains ten actions (walk, sit down, stand up, pick up, carry, throw, push, pull, wave hands, clap hands) by ten subjects with two instances. Among them, four actions (*walk, sit down, stand up, and pick up)* are only used that are respective to the system.

From the dataset, skeleton joints data are only used and a hybrid feature which combines the joint angular information and joint distance parameters is used to extract the discriminative features. Then, these features are clustered based on static k-means with the five cluster-identifiers to reduce the similar postures. The class-label of each human pose is determined using artificial neural network, and the corresponding discrete hidden Markov model is built to correctly recognize the posture sequences.

The assessment is performed on the 400 different sequences of human posture that it consists of four actions by ten subjects with two instances and five clusters. In the training set, there are 240 distinct posture sequences that contain four actions by six subjects with two instances and five clusters. In the testing phase, there are 160 different sequences which include four actions by four subjects with two instances and five clusters.

Firstly, the evaluation using this method is performed on the training set. The confusion matrix and experimental results are shown in Table 1 and Table 2. In this experimentation, the correctness rate of all of the actions is high, and the overall precision, recall, and specificity rate is also shown in Table 3.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 1, January 2019

Table 1: Confusion Matrix on the Training set of UTKinect dataset

Type of Action	*Walking*	*Sitting*	*Standing*	*Bending*
Walking	83	-	-	17
Sitting	-	100	-	-
Standing		-	100	-
Bending	-	-	-	100

Table 2: Accuracy rate on the training set of UTKinect dataset

Accuracy (%)			
Walking	Sitting	Standing	Bending
96	100	100	98

Table 3: Precision (%), Recall (%), and Specificity (%) on the Training set of UTKinect dataset

Action	**Precision (%)**	**Recall (%)**	**Specificity (%)**
Walking	100	83	100
Sitting	100	100	100
Standing	100	100	100
Bending	92	100	97

Then, the experiment is also evaluated on the testing set. In this evaluation, the confusion matrix is shown in Table 4, and the Table 5 shows the accuracy rate of each action. From the experimental results, the overall accuracy rate of all of the action is reasonable. The Table 6 also shows the precision, recall, and specificity rate for each action.

Table 4: Confusion Matrix on the Testing set of UTKinect dataset

Type of Action	*Walking*	*Sitting*	*Standing*	*Bending*
Walking	88	-	12	-
Sitting	-	88	-	12
Standing	-	-	100	-
Bending	-	12	-	88

Table 5: Accuracy rate on the testing set of UTKinect dataset

Accuracy (%)			
Walking	Sitting	Standing	Bending
97	94	97	94

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 1, January 2019

Table 6: Precision (%), Recall (%), and Specificity (%) on the Testing set of UTKinect dataset

Action	Precision (%)	Recall (%)	Specificity (%)
Walking	100	88	100
Sitting	88	88	96
Standing	88	100	96
Bending	88	88	96

IV. CONCLUSION

The human action recognition system in skeleton joints data is proposed. The proposed hybrid feature based on the angular information and distance parameters is used and these features are grouped together to reduce the similar human postures using the static k-means algorithm that statically takes only the initial centroids at the first time. Then, the artificial neural network is applied to classify the class-labels of each posture. Furthermore, a sequence of human action is correctly recognized by using the discrete hidden Markov Model (HMM). The evaluation is effectively performed on the public dataset UTKinect by means of these techniques. During experimental results, the proposed system has shown the significant recognition accuracy performance.

REFERENCES

1. F. Oi, R. Chaudhry, G. Kurillo, R. Vidal, R. Bajcsy, Sequence of the most informative joints (SMIJ): A new representation for human skeletal action recognition, Journal of Visual Communication and Image Representation 25 (1) (2014).
2. X. Yang and Y. Tian. EigenJoints-based Action Recognition Using Na¨ıve-Bayes-Nearest-Neighbor. In *CVPR 2012 HAU3D Workshop*, 2012.
3. M. Raptis, D. Kirovski, and H. Hoppe. Real-time classification of dance gestures from skeleton animation. In Proceedings of the SIGGRAPH/Eurographics Symposium on Computer Animation - SCA '11, page 147, New York, USA, 2011. ACM Press.
4. L. Xia, C.-C. Chen, and J. Aggarwal. View invariant human action recognition using histograms of 3d joints. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 20–27. IEEE, 2012.
5. G. Dong, J. Li, Efficient mining of emerging patterns: Discovering trends and differences, in: Proceedings of the Fifth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 1999,
6. C. Meek, D. M. Chickering, D. Heckerman, Autoregressive tree models for time-series analysis, in: Proc. of the Second International SIAM Conference on Data Mining, SIAM, 2002
7. T. Kerola, N. Inoue, K. Shinoda, Spectral graph skeletons for 3D action recognition, in: Proc. of Asian Conference on Computer Vision (ACCV), Springer, 2014
8. D. K. Hammond, P. Vandergheynst, R. Gribonval, Wavelets on graphs via spectral graph theory, Applied and Computational Harmonic Analysis
9. M. E. Hussein, M. Torki, M. A. Gowayyed, M. El-Saban, Human action recognition using a temporal hierarchy of covariance descriptors on 3D joint locations, in: Proc. of International Joint Conference on Arti_cial Intelligence (IJCAI), AAAI Press, 2013
10. X. Yang, Y. Tian, Eigenjoints-based action recognition using Naive-Bayes-Nearest-Neighbor, in: Proc. of Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, 2012.
11. E. P. Ijjina, C. K. Mohan, Human action recognition based on MOCAP information using convolution neural networks
12. J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. CVPR, 2011.