SciencePG
Science Publishing Group

# Human Activity Recognition Based on Weighted Sum Method and Combination of Feature Extraction Methods

**Nang Hung Van Nguyen[1], Minh Tuan Pham[1], Nho Dai Ung[1], Kanta Tachibana[2]**

[1]Faculty of Information Technology, Danang University of Science and Technology, Danang, Vietnam

[2]Faculty of Informatics, Department of Information System and Applied Mathematecs, Kogakuin University, Tokyo, Japan

**Email address:**
nguyenvan@dut.udn.vn (N. H. V. Nguyen), pmtuan@dut.udn.vn (M. T. Pham), dai.n.ung@gmail.com (N. D. Ung),
kanta@cc.kogakuin.ac.jp (K. Tachibana)

**To cite this article:**
Nang Hung Van Nguyen, Minh Tuan Pham, Nho Dai Ung, Kanta Tachibana. Human Activity Recognition Based on Weighted Sum Method and Combination of Feature Extraction Methods. *International Journal of Intelligent Information Systems*. Vol. 7, No. 1, 2018, pp. 9-14. doi: 10.11648/j.ijiis.20180701.13

**Abstract:** Human Activity Recognition (HAR) is one of the most important areas of computer vision research. The biggest difficulty for HAR system is that the camera could only film in one direction, leading to a shortage of data and low recognition results. This paper focuses on researching and building new models of HAR, including Principal Components Analysis (PCA), Linear discriminant Analysis (LDA) is to reduce the dimensionality and size of data, contributing to high recognition accuracy. First, from the 3D motion data, we conducted a pretreatment and feature extraction of objects. Next, we built a recognition model corresponding to each feature extraction method and we used Support Vector Machine (SVM) model to train. Finally, we used weighted methods to combine the results of the model to train and give the final results. The paper experiment on CMU MOCAP database and the percentage receiving proposed method is higher than that from the previous method.

**Keywords:** Human Activity Recognition, Principal Components Analysis, Linear Discriminant Analysis, Support Vector Machine

## 1. Introduction

Nowaday, Human Activity Recognition [1] has received special attention from researchers around the world. Its applications include intelligent security monitoring systems, health care systems, intelligent transportation systems, and a variety of systems that involve interactions between people and electronic devices such as human computer interfaces [2].

The study in the area of HAR focuses on the recognition of activities from the videos recorded by the regular cameras. The biggest problem of HAR using regular cameras is that the video can only be recorded from one direction, leading to a shortage of data for recognition. To improved the recognition rate for HAR, in recent years, the 3D method using "marker" such as the Motion Capture using stereo camera [3] or using the devices dedicated depth sensor [4] such as the Microsoft Kinect [5] is proposed.

This paper proposed the combination method using many feature extraction methods such as Principal Components

Analysis (PCA) [6] and Multi-class Linear Discriminant Analysis (Multi-class LDA) [7]. The recognition results depend on the feature extraction method. This paper uses the weighted sum method [8] to combine the feature extraction methods to enhance the efficiency of activity recognition.

The paper is described as following. The first section is the introduction of this paper. The second section presents the related researches. The $3^{rd}$ section shows the proposed method. The $4^{th}$ section shows the experiment result. Finally, the $5^{th}$ section summarizes this paper.

## 2. Related Research

### 2.1. Principal Components Analysis

Principal Components Analysis [4] is a statistical algorithm using orthogonal transformation to transform a set of data from a higher dimensional space into new lower dimensional space. This transformation is defined in such a way that the first principal component has the largest

possible variance. Figure 1 illustrates an example of the transformation using PCA.
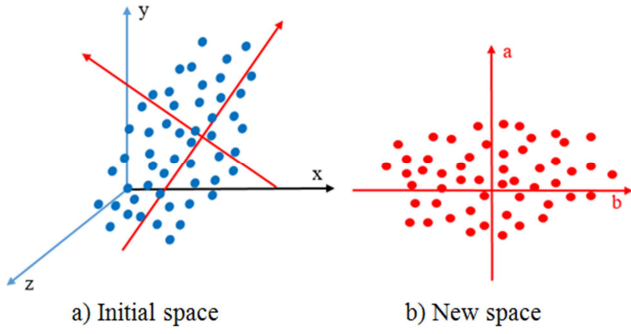


a) Initial space      b) New space

**Figure 1.** *Example of the transformation using PCA.*

Given training set $X = \{x_i | x_i \in R^d\}, i \in \{1 \dots n\}$. Where $x_i$ is a vector in $d$-dimensional space, and $n$ is the number of vectors in the set X.

First, PCA converts centre of all vectors into the original coordinates by:

$$\hat{X} = \{\hat{x}_i\},$$

$$\hat{x}_i = x_i - \mu$$

where μ is the center of all vectors in X, it is calculated by formula:

$$\mu = \frac{1}{n}\sum_{j=1}^{n} x_j$$

Then, PCA method calculates the covariance matrix of the vectors in $\hat{X}$, Gọi $X^* = [\hat{x}_1 \dots \hat{x}_n] \in R^{n \times d}$ is matrix containing all the training vectors.

$$V = X^* X^{*T} \in R^{d \times d}$$

The projection vector $u$ of PCA is solved by the eigen problem:

$$Vu = \lambda u$$

Finally, PCA uses the first $k$ vectors to built the new lower dimensional space. Call matrix $U_{PCA} = [u_1, u_2 \dots u_k] \in R^{d \times k}$. Then the coordinate of all vectors in the new coordinate space is:

$$F = U_{PCA}^T X^* \in R^{n \times k}.$$

## 2.2. Multi-class Linear Discriminant Analysis

Multi-class Linear Discriminant Analysis (Multi-class LDA)[5] is a generalization of standard two-class LDA that can handle arbitrary number of classes. Multi-class LDA finds projection vector $u$ such that all data projected into $u$ have the largest separations.

Given the training data set with label:

$$X = \left\{(x_i, y_i) \middle| x_i \in R^d, y \in \{1 \dots l\}\right\}; i \in \{1 \dots n\}$$

Where $x_i$ is i[th] vector of the training data set in $d$-dimensional space and $y_i$ is a label of $x_i$. $l$ is the number of label kinds and $n$ is the number of vectors in the set X. The separation between these classes is defined by the ratio of the variance between the classes to the variance within the classes as following:

$$\frac{u^T S_b u}{u S_w u}$$

Where $S_b$ is between-class scatter matrix, $S_b$ is calculated by formula:

$$S_b = \sum_{c=1}^{l} n_c (\mu_c - \mu)(\mu_c - \mu)^T \in R^{d \times d}$$

Where, $\mu_c$ is center of class $c$ and $n_c$ is the number of vectors in class $c$. And $\mu$ and $n$ is the mean and the number of all data the training set.

$$\mu_c = \frac{1}{n_c}\sum_{y_i=c} x_i$$

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$S_w$ is within-class scatter matrix, $S_w$ is calculated by:

$$S_b = \sum_{i=1}^{n} (x_i - \mu_{y_i})(x_i - \mu_{y_i})^T \in R^{d \times d}$$

Here, $\mu_{y_i}$ is the center of the class is labeled by $y_i$.

Then, Multi-class LDA determines the projection vector $u$ as following,

$$\hat{u} = \underset{u}{\text{argmax}} \frac{u^T S_b u}{u^T S_w u}$$

We can find vector $w$ by solving the following eigen problem:

$$S_b u = \lambda S_w u$$

## 3. Proposed Method

This paper combine the feature extraction methods and use weighted sum model to improve recognition results. Proposed method includes 3 steps: First, using pre-processing method with CMU data; Next step, use Manual, PCA and LDA method to feature extraction; Finally, proposed weighted sum method for machine learning with Suport Vector Machine (SVM) [9]. Figure 2 shows the proposed model.
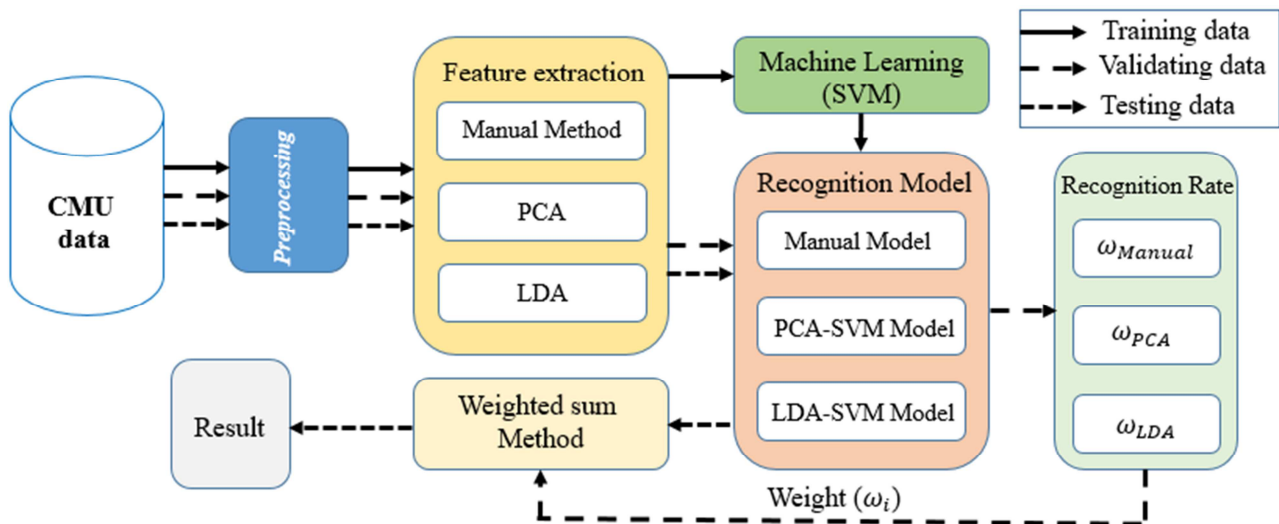
**Figure 2.** *Proposed model.*

### 3.1. Preprocessing

This paper uses skeleton model of human [3]. This step is as followed:

1. Reduce the number of features by reducing the number of the selected bone in 3D human model. This paper references the method proposed by K. Adistambha [10] to select the bones of 3D human model.
2. Normalize monitoring time of all actions. This paper uses the time of shortest actions.

### 3.2. Feature Extraction

The main purpose is to find vectors representing the data with smaller dimensional initial data while ensuring effective human activity recognition. This paper uses 3 methods in feature extraction.

1. Feature extraction proposed by K. Adistambha [10],
2. PCA,
3. Mutli-class LDA.

### 3.3. Weighted Sum Method

Each feature extraction method has the different advantages and disadvantages. Combining these methods can improve the recognition result. This paper uses the recognition rate to calculate the weight in weighted sum model. Call $f_i \in \{1 \dots l\}$ is the label assigned using the $i^{th}$ feature extraction method. $\omega_i$ is the recognition rate of the $i^{th}$ feature extraction method. This paper assigns label $\hat{f}$ as following:

$$\hat{f} = \underset{j}{\operatorname{argmax}}(p_j)$$

where,

$$p_j = \sum_{f_i=j} \omega_i$$

## 4. Experiment Result

This experiment uses the CMU graphics Lab motion capture database [3]. The database includes Acclaim Skeleton File (ASF) structure and Acclaim Motion Capture (AMC) files. This paper uses data consists of 29 bones and represent ASF structure shown as Figure 3. Figure 4 is an example of a 3D image of human bone structure reconstructed from AMC.
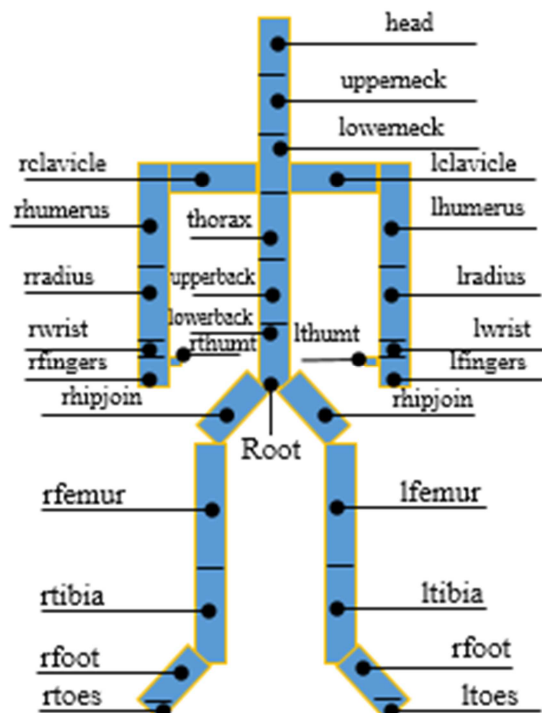


**Figure 3.** *Model skeleton [10].*

This paper uses data formatted by the ASF and AMC:
1. In the ASF file a base pose is defined for the skeleton that is the starting point for the motion data. ASF has information: length, direction, local coordinate frame,

number of Dofs, joint limits and hierarchy, connections of the bone.

2. The AMC file contains the motion data for a skeleton defined by an ASF file. The motion data is given a sample at a time. Each sample consists of a number of lines, a segment per line, containing the data.
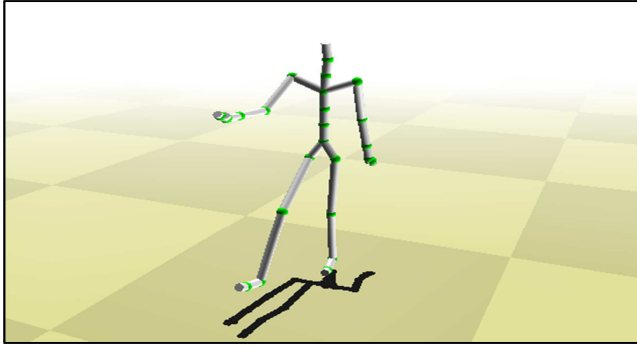


*Figure 4. 3D human reconstructed from AMC.*

This experiment use 4 kinds of the human action. They are running, walking, jumping and dancing. Training data includes 165 action used to build models with PCA, multi-class LDA and Suport Vector Machine (SVM) [9]. Validating data includes 163 actions used to calculate accuracy rates of each feature extraction method. Testing data included 163 actions used to evaluate the proposed method. Table 1 shows the CMU data used in this experiment.

*Table 1. CMU data.*

| Action | run | walk | jump | dance | Total |
|---|---|---|---|---|---|
| Training data | 24 | 75 | 43 | 23 | 165 |
| Validating data | 24 | 75 | 42 | 22 | 163 |
| Testing data | 24 | 75 | 42 | 22 | 163 |
| Total | 72 | 225 | 127 | 67 | 491 |

## 4.1. Recognition Result Using Manual Method

First, this paper uses the manual feature extraction based on K. Adistambha method [10]. This paper defines the feature based on detail groups of bones showed as following.

1  3 bones group: root, lowerback, upperback.
2  4 bones group: group 3 + thorax.
3  7 bones group: group 4 + lowerneck + upperneck, head.
4  13 bones group: group 7 + left and right clavicle, left and right humerus, + left and right femur.
5  23 bones group: group 13 + left and right wrist + left and right hand + left and right foot + left and right radius + left and right tibia.

Experiment in this order, which means starting with the group with the least bone, then adding the bones to produce the new group, finally finding the SVM model with the best results. Table 2 shows the result of human activity recognition based on SVM with manual feature extraction.

*Table 2. Experiment results with the manual method.*

| Number of Bones | run | walk | jump | dance | Accuracy Rate |
|---|---|---|---|---|---|
| 3 | 0.0% | 93.3% | 14.3% | 18.2% | 49.4% |
| 4 | 0.0% | 93.3% | 14.3% | 13.6% | 48.8% |
| 7 | 0.0% | 94.7% | 30.9% | 13.6% | 53.7% |
| 13 | 0.0% | 96.0% | 28.5% | 18.2% | 54.3% |
| 23 | 78.3% | 98.7% | 81.0% | 31.9% | 82.1% |
| 11 | 78.3% | 98.7% | 81.0% | 36.4% | 82.7% |
| 29 (all) | 78.3% | 98.7% | 81.0% | 41.0% | 83.3% |

Experimental results of bone 11 were 82.7% and only lower when 29 bones were 83.3%. It is therefore possible to use a bundle of 11 bones as a method of feature extraction to construct a weighted sum method in the proposed model.
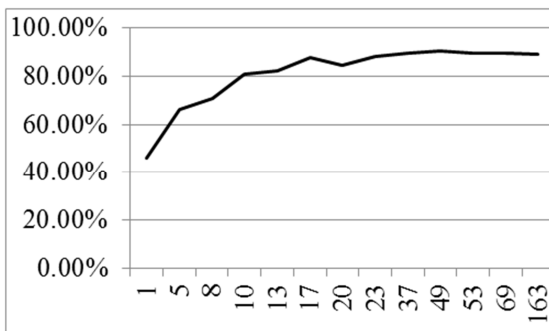
## 4.2. Recognition Result Using PCA



*Figure 5. The accuracy rate using PCA.*

For the PCA method, the parameter to be determined is the number of Eigenvectors (which is the dimension in the new space) so that the recognition model has the highest accuracy. Figure 5 shows the variation of the recognition rate when the number of dimensions changes using the PCA.

As the number increases, the accuracy increases to a threshold value (90.1% with a dimension equal to 49), then begins to decrease and gradually becomes straight when the number is large. The maximum accuracy rate is 90.1% when the feature number is 49. The detail of recognition result is shown as Table 3.

*Table 3. Detail of maximum accuracy using PCA.*

| Action | run | walk | jump | dance |
|---|---|---|---|---|
| run | 78.3% | 17.4% | 4.3% | 0.0% |
| walk | 0.0% | 98.7% | 0.0% | 1.3% |
| jump | 0.0% | 14.3% | 85.7% | 0.0% |
| dance | 0.0% | 9.1% | 9.1% | 81.8% |

### 4.3. Recognition Result Using Multi-class LDA

The multi-class LDA method also needs to determine the number of dimensions of the data extracted so that the recognition model has the highest accuracy.
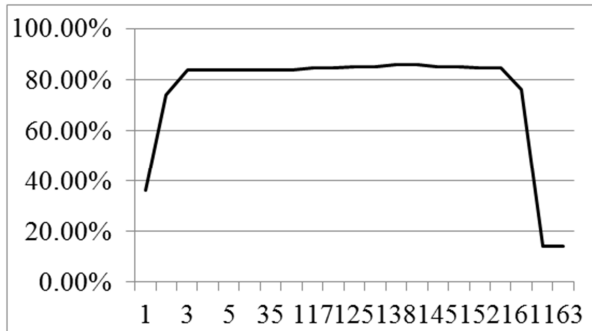


**Figure 6.** *The accuracy rate using multi-class LDA.*

Figure 6 shows the accuracy rate of human activity recognition using multi-class LDA. The maximum accuracy is 86.0% when the feature number is 138. The detail of recognition result is shown as Table 4.

**Table 4.** *Detail of maximum accuracy using multi-class LDA.*

| Accuracy (86.0%) | Run | Walk | Jump | Dance |
|---|---|---|---|---|
| Run | 78.3% | 13.0% | 4.3% | 4.4% |
| Walk | 0.0% | 86.7% | 10.7% | 2.6% |
| Jump | 0.0% | 7.1% | 92.9% | 0.0% |
| Dance | 9.1% | 4.5% | 9.1% | 77.3% |

D. Recognition result using Weighted sum method

This sub-section uses the best result of recognition based on manual method, PCA, mult-class LDA to assign the label of human action based on proposed weighted sum method. The Tabel 5 shows the best performances of each feature extraction method.

**Table 5.** *The best recognition rate of each feature extraction method.*

| Weight | Recognition rate |
|---|---|
| $\omega_{Manual}$ | 82.1% |
| $\omega_{PCA}$ | 90.1% |
| $\omega_{LDA}$ | 86.0% |

**Table 6.** *Results of the proposed model.*

| Accuracy (90.7%) | Run | Walk | Jump | Dance |
|---|---|---|---|---|
| Run | 78.3% | 17.4% | 4.3% | 0.0% |
| Walk | 0.0% | 100% | 0.0% | 0.0% |
| Jump | 0.0% | 14.3% | 85.7% | 0.0% |
| Dance | 0.0% | 9.1% | 9.1% | 81.8% |

Table 6 is a detailed recognition result of the proposed model. It shows that the accuracy rate of the human ativity recognition using proposed weighted sum method is 90.7%.
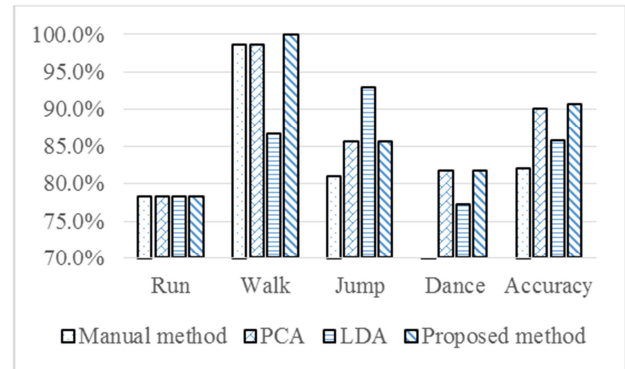


**Figure 7.** *Comparison between conventional methods and proposed method.*

Figure 7 shows the comparison between conventional methods and proposed method. The result of recognition using PCA has higher perfomance when combining with the manual method. And the proposed method has the higher accuracy rate (90.7%). It is better than the method using PCA (90.1%).

## 5. Conclusion

This paper described the related research of feature extraction in human activity recognition area included Principal Components Analysis (PCA) and Multi-class Linear Disciminant Analysis (Multi-class LDA). This paper also described the manual feature extraction method referenced the method proposed by K. Adistambha. Next, this paper showed the proposed method using weighted sum method and the conbination of conventional feature extraction methods. The experiment result showed the accuracy rate of the human activity recognition based on manual feature extraction, PCA and Multi-class LDA. It also showed the perfomance of proposed method. It showed that the perfomance of proposed method is better than conventional methods.

In future work, we combine other feature extraction methods such as kernel feature extraction method to improve the human activity recognition rate.

## References

[1] Aggarwal K, Lu Xia (2014), "Human Activity Recognition from 3D Data-A Review", Pattern Recognition Letters, Elsevier B. V, USA.

[2] M. Debyeche, J. P Haton, and A. Houacine "A New Vector Quantization front-end Process for Discrete HMM Speech Recognition System" International Science Index Vol: 1, No: 6, 2007.

[3] CMU Graphics Lab Motion Capture Database. Carnegie Mellon University, Pennsylvania, United States. Website: http://mocap.cs.cmu.edu/

[4] Ahmad Jalal, Shaharyar Kamal, Daijin Kim (2014), "A Depth Video Sensor-Based Life-Logging Human Activity Recognition System for Elderly Care in Smart Indoor Environments", Sensors-2014.

[5] Kohei Arai, Rosa Andrie Asmara (2013), "3D Skeleton model derived from Kinect Depth Sensor Camera andits application to walking style quality evaluations", *IJARAL – International Journal of Advanced in Artificial Intelligence*.

[6] Jolliffe, I. T. (2002), "Principal Component Analysis", Springer.

[7] Alan J. I. (2012), "Linear Discriminant Analysis", Springer.

[8] Yong Kim, Olivier de Weck "Adaptive Weighted Sum Method for Multiobjective Optimization" 02139, USA

[9] Steinwart, Ingo, Christmann, Andreas (2008), "Support Vector Machines", Springer.

[10] Adistambha K, Ritz C. H, Burnett I. S (2008), "Motion Classification Using Dynamic Time Warping", ICPR 2008, IEEE.

[11] N. H. V. NGUYEN, M. T. PHAM, P. H. DO, "Marker Selection for Human Activity Recognition Using Combination of Conformal Geometric Algebra and Principal Component Regression", SoICT 2016.

[12] N. H. V. NGUYEN, M. T. PHAM, P. H. DO, K. TACHIBANA, "The Fast Gaussian distribution based Adaboost algorithm for Face detection", ISAT-16.