

A New Method for Human Posture Recognition Using Principal Component Analysis and Artificial Neural Network

M. Z. Uddin¹, M. A. Yousuf^{2*}

¹Dept. of Computer Education, Sungkyunkwan University, 53, Myungryun-dong, 3-ga, Chongno-gu, Seoul, 110-745, Republic of Korea

²Institute of Information Technology, Jahangirnagar University, Savar, Dhaka-1342, Bangladesh

Received 14 July 2014, accepted in final revised form 7 May 2015

Abstract

The recognition of human posture from images is currently a very active area of research in computer vision. This paper presents a novel recognition method to determine a human posture is of walking or sitting using Principal Component Analysis (PCA) and Artificial Neural Network (ANN). In this paper, two types of learning are used to recognize the human posture. One is unsupervised and another is supervised learning. We have used PCA for unsupervised learning and ANN for supervised learning. To evaluate the performance of the proposed method, we have considered four types of human posture; walking, sitting, right leg up-down and left leg up-down. The experimental results on the human action of walking, sitting, right leg up-down and left leg up-down database show that our approach produces accurate recognition.

Keywords: Human action recognition; Principal Component Analysis (PCA); Feature extraction; Artificial Neural Network (ANN).

© 2015 JSR Publications. ISSN: 2070-0237 (Print); 2070-0245 (Online). All rights reserved.

doi: <http://dx.doi.org/10.3329/jsr.v7i3.19527>

J. Sci. Res. 7 (3), 11-19 (2015)

1. Introduction

Recognition of human actions from images is very important area in computer vision. Recently, Human Activity Recognition (HAR) has got considerable interests from many researchers in distinguished research areas due to its potential use in proactive computing, which is a technology to proactively anticipate peoples' necessity in situations such as health-care and take requisite actions [1-3]. The recognition of posture is one step in the global process of analyzing human behavior [4]. Several human action recognition methods have been proposed in the past decades. A detailed survey can be found

* Corresponding author: usuf672@yahoo.com

elsewhere [5]. Action recognition by using shape is performed by several researchers. For example, [6,7], authors used 2D and 3D shape features for recognizing actions. Cohen *et al.* [6], presented view independent 3D shape description for classifying and identifying human posture using support vector machines. The edge data from Canny edge detector represented the shape, and each actions defined the key frames [7]. Masoud *et al.* used the Principal Component Analysis (PCA) for obtaining the features and recognition is performed using distance measure [8]. PCA is a way of identifying patterns in data, and expressing the data in such a way that highlight their similarities and differences [9,10]. PCA is a useful statistical technique that has found application in fields such as face detection [11], face recognition [11,12], and gesture classification [13], and is a common technique for finding patterns in data of high dimension. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analyzing data [14]. PCA is one of several statistical tools available for reducing the dimensionality of a data set [15]. The extracted feature vectors in the reduced space are used to train the ANN Classifier. Basically, binary silhouettes of various activities are commonly used to represent different human activities [16]. Shift-invariant sparse coding was used to perform unsupervised learning to train an auto encoder network [17]. In a work, authors have presented “Action Bank”, a high-level representation of video which is comprised of many individual action detectors sampled broadly in semantic space as well as view point space [18].

In this work, PCA is used to extract prominent activity features and Artificial Neural Network (ANN) is used for classification. For feature extraction, we are given some data x and the cost function to be minimized, that can be any function of the data x and the network's output, f . We are given a set of example pairs (x,y) , $x \in X$ and $y \in Y$ and the aim is to find a function $f: X \rightarrow Y$ in the allowed class of functions that matches the examples.

After extracting features and applying PCA to reduce the feature vector dimensionality, an ANN is trained on a set of given actions to recognize human activity. An ANN, usually called neural network (NN), is a mathematical model or computational model that is inspired by the structure and/or functional aspects of biological neural networks. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. ANN is simple mathematical model defining a function $f: X \rightarrow Y$ or a distribution over X or both X and Y .

In this paper, four types of human actions like walking, sitting, right leg up-down and left leg up-down have been considered. The proposed human action recognition method consists of three major steps, human detection and tracking, feature extraction, and training or classification. The first step is to detect and track a human figure in an image sequence. Background subtraction is a simple method for human detection and tracking. For each image sequence, a background subtraction algorithm and a simple correspondence procedure are used for segmenting human action. The second step is feature extraction from each frame and to normalize the image size. PCA is then applied to reduce the dimensionality of input feature space. The third step is to recognize human actions using ANN in the lower-dimensional eigenspace.

This paper is organized as follows: Section 2 presents the method to recognize human postures which include the extraction of region of interest (ROI), feature extraction using PCA and classification using ANN. Experimental results of the proposed approach is explained in section 3. Discussions and conclusions are presented in Section 4.

2. Experimental

2.1. Extraction of Region of Interest (ROI)

We assume that each image includes only one person and performs a single activity like walking or sitting. The color value of each background pixel is modeled by using a single Gaussian distribution. The Gaussian probability density function of a pixel to be in foreground is

$$p(J(x,y)=k) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(k-I(x,y))^2}{2\sigma^2}\right) \quad (1)$$

The steps of ROI extraction are as follows:

- a. Compute the probability of each pixel.
- b. Choose a threshold, T .
- c. Classify all pixels as foreground when $P(J(x,y)=k) < T$. T is chosen by hand that seems to produce pleasing results.
- d. Convert the ROI to binary.
- e. Find the region from the whole image.

Fig. 1. shows the result of extraction of ROI from the image.



Fig. 1. Extraction of region of interest (ROI). (a) Background (b) Human in foreground and (c) After background subtraction.

2.2. Feature extraction using PCA

The fundamental approach is to compute the eigenvectors of the covariance data matrix Q and then approximation is done using the linear combination of top eigenvectors. The

covariance matrix of the sample training image vectors and the principle components of the covariance matrix can be calculated respectively as:

$$Q = \frac{1}{T} \sum_{i=1}^T (x_i x_i^T) \quad (2)$$

$$E^T Q E = A \quad (3)$$

Where E represents the matrix of orthonormal eigenvectors and A diagonal matrix of the eigenvalues. E reflects the original coordinate system onto the eigenvectors where the eigenvectors corresponding to the largest eigenvalue indicates the axis of the largest variance and the next largest one is the orthogonal axis of the largest one indicating second largest variance and so on. Fig. 2 shows the representation of new axis of two principal components.

Usually, the eigenvalues that are about zero values carry negligible variance and hence can be excluded. So, several m eigenvectors corresponding to the largest eigenvalues can be used to define the subspace. Thus the full dimensional image vectors can be easily represented in reduced dimension.

The steps of recognition of walking and sitting patterns using PCA are as follows:

- a. Convert every image as a single column vector and concatenate horizontally.
- b. Subtract the mean. For PCA to work properly, we have to subtract the mean from each of the data dimensions. The mean subtracted is the average across each dimension.
- c. Calculate the covariance matrix.
- d. Calculate the eigenvectors and eigenvalues from the matrix. Since the covariance matrix is square, we can calculate the eigenvectors and eigenvalues for this matrix.
- e. Plot the values and choose top components to reduce dimension.

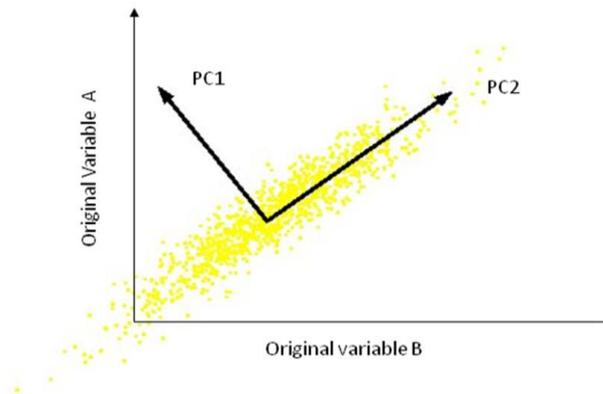


Fig. 2. Representation of new axis of two principal components.

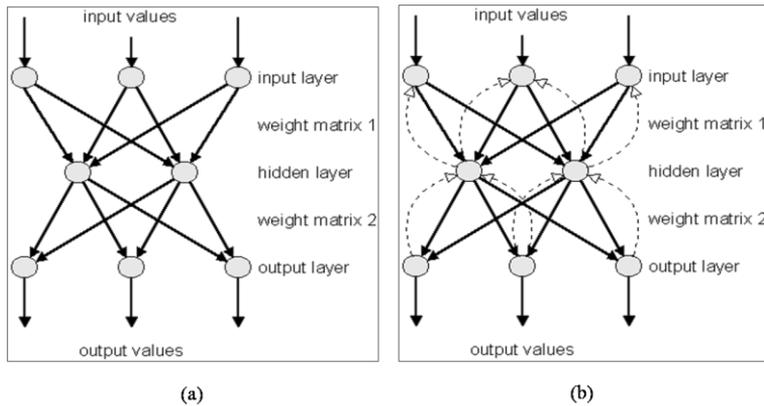


Fig. 3. (a) Multi-layer-perceptron structure; (b) Back propagation net structure.

2.3. Classification using ANN

After feature extraction and dimensions reduction, we used ANN for classification. Classification is a pattern recognition (PR) problem of assigning an object to a class. Thus, the output of the PR system is an integer label. The task of the classifier is to partition the feature space into class-labeled decision regions. Multilayer Perceptron (MLP) is chosen as our classifier in this study. A MLP is an ANN which is an extremely popular and widely documented architecture. It is a good tool for classification purposes. It can approximate almost any regularity between its input and output. The ANN weights are adjusted by supervised training procedure called back propagation. Back propagation is a kind of gradient descent method, which searches an acceptable local minimum in the neural network weight space in order to achieve minimal error. MLP comprises an input, output and one or more hidden layers. With the exception of the input layer, all layers compute their output with a weighted output formula, an optional bias and an activation function. For this study, only a single hidden layer was used. Fig. 3 shows the multi layer perceptron and back propagation net structure.

3. Results

The proposed method is evaluated by using the human action database containing different types of human actions like walking, sitting, right leg up-down and left leg up-down activities. In Fig. 4a, the training data sets of 40 binary images of walking and sitting are shown. These are the couple of sequences of generalized of ROIs for walking and sitting.

To apply feature extraction, every normalized ROI image is represented as a row vector in a raster scan fashion where the dimension of the vector is equal to the number of pixels in the entire image. In this paper, every image is converted to (2500X1) vector. After applying PCA to the Image database the total feature space is (2500X2500). Image database is shown in Fig. 4b. Only 10 features are taken which represent the total database

because the eigenvalues of other features are about to zero. So our final feature space is (2500X10) matrix. Fig. 5 shows the top 10 eigenvalues after applying PCA.

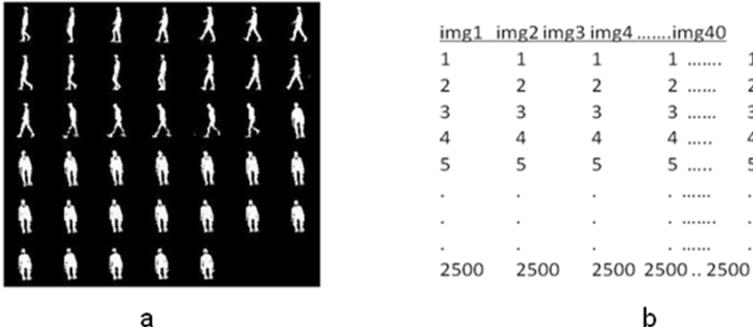


Fig. 4. a) Training Images (Walking and Sitting); b) Image dataset (Walking and Sitting) in matrix form.

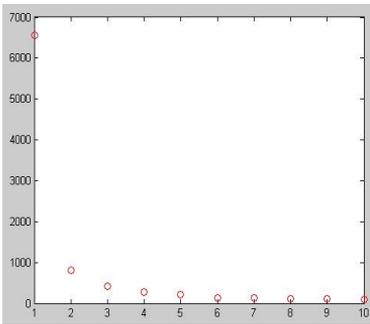


Fig. 5. Top 10 eigenvalues after applying PCA.

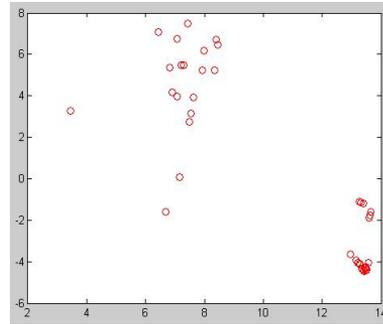


Fig. 6. Projection of 40 images (Training) into 2 principal component.

Only 10 feature vectors from total PCA feature space have been taken where every vector is a co-ordinate. So 10 dimensions are available to project the images i.e. every image will be represented as a single point in the 10 dimensions.

The projection formula is $\text{Projection} = (\text{Feature Space}^T \times \text{Image vector})$
 So for image1, the projection $= (2500 \times 10)^T \times (2500 \times 1)$
 $= (10 \times 2500) \times (2500 \times 1)$
 $= (10 \times 1)$

So every image has 10 co-ordinate values. For other images the projection can be calculated in the same way.

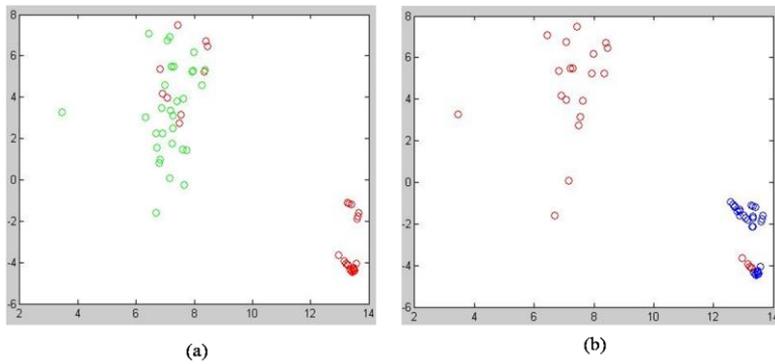


Fig. 7. (a) Projection of 20 images of sitting (testing) into 2 principal components (green points), (b) Projection of 20 images of walking (testing) into 2 principal components (blue points).

Table 1. Achieved testing output for training images (walking).

Sl.	Output 1	Output 2
1.	1.00	0.00
2.	1.00	0.00
3.	1.00	0.00
4.	1.00	0.00
5.	1.00	0.00
6.	1.00	0.00
7.	1.00	0.00
8.	0.98	0.00
9.	1.00	0.00
10.	1.00	0.00
11.	0.99	0.00
12.	1.00	0.00
13.	1.00	0.00
14.	1.00	0.00
15.	0.99	0.00
16.	1.00	0.00
17.	1.00	0.00
18.	1.00	0.00
19.	1.00	0.00
20.	1.00	0.00

Table 2. Achieved testing output for training images (sitting).

Sl.	Output 1	Output 2
1.	0.00	1.00
2.	0.00	0.99
3.	0.00	1.00
4.	0.00	0.99
5.	0.00	1.00
6.	0.00	1.00
7.	0.00	1.00
8.	0.00	0.99
9.	0.00	1.00
10.	0.00	0.99
11.	0.00	1.00
12.	0.00	0.99
13.	0.00	1.00
14.	0.00	1.00
15.	0.00	0.99
16.	0.00	1.00
17.	0.00	0.99
18.	0.00	1.00
19.	0.00	1.00
20.	0.00	0.99

Now, we see the values in 2-dimension space because, 10 dimensional representation is not possible. If every projection is taken as a pattern then we have 40 patterns. Fig. 6 represents the projection of all 40 images (training) into 2 dimensions (principal component) from 10 dimensions. Fig. 7 represents the projection of 20 images of sitting (testing) and walking (testing) into 2 principal components. We have taken every PCA projection of images as training patterns. So the training patterns are of size of 10 values

for each image. If there are 40 training images then we have 40 patterns. Back propagation method is used. We have 10 input neurons. Only 2 output neurons to take decision of two activities. The data are trained by two hidden layers. We have given the test patterns and specified the output for every pattern. Values of 2 output neurons for every walking images are specified as (1 0). Values of 2 output neurons for every sitting images are specified as (0 1). Tables 1 and 2 represent the achieved output result for the training images of walking and sitting respectively.

Experiment for two more human activities like right leg up-down and left leg up-down was also done. Our proposed system model using PCA for feature extraction and MLP ANN for classification resulted in an average recognition rate of 91.875% for four activities. Our proposed method was compared with the conventional human activity recognition approach based on Naive Bayes, AdaBoost and k-NN classifier, shown in Tables 3-5 respectively, where the maximum mean recognition rate was 77.375% using k-NN classifier. Table 6 shows the recognition result using our proposed method. It turns out from these experimental results that the proposed approach (recognition rate: 91.875%) shows the superior recognition performance over other approaches.

Table 3. Experimental result using Naïve classifier.

Activity	Recognition Rate (%)	Mean (%)
Walking	67	
Sitting	69.50	67.50
Right leg up-down	65.50	
Left leg up-down	68	

Table 4. Experimental result using AdaBoost classifier.

Activity	Recognition Rate (%)	Mean (%)
Walking	71	
Sitting	73.50	71.75
Right leg up-down	72	
Left leg up-down	70.50	

Table 5. Experimental result using k-NN classifier.

Activity	Recognition Rate (%)	Mean (%)
Walking	76	
Sitting	78	77.375
Right leg up-down	80.50	
Left leg up-down	75	

Table 6. Experimental result using proposed method.

Activity	Recognition Rate (%)	Mean (%)
Walking	89	
Sitting	93.50	91.875
Right leg up-down	94	
Left leg up-down	91	

4. Discussions and Conclusion

A method for human body posture recognition based on PCA and ANN is presented. As can be seen from the experimental results, proposed method can be used for human posture recognition with high degree of accuracy. In future, it is possible to use multiple frames for recognition at a time, combine motion features with shape features for classification and motion features of the ROI can be calculated by optical flow of the consecutive frames.

References

1. M. Kokkinos, N. D. Doulamis, and A. D. Doulamis, *Int. J. Adv. Robotic Syst.* **10**, 1 (2013). <http://dx.doi.org/10.5772/54049>
2. A. Rehman and T. Saba, *Int. Arab J. Inform. Technol. (IAJIT)*, 126 (2013).
3. S. S. Yun, M. T. Choi., M. Kim and J. B. Song, *Int. J. Adv. Robotic Syst.* **9**, 1 (2012). <http://dx.doi.org/10.5772/50648>
4. K. Fukunaga, *Introduction to Statistical Pattern Recognition* (Academic Press, 1990).
5. D. M. Gavriila, *Comput. Vision Image Understand.* **73**, 82 (1999). <http://dx.doi.org/10.1006/cviu.1998.0716>
6. I. Cohen and H. Li, in *Inference of Human Postures by Classification of 3D Human Body Shape - Proc. of IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures* (2003) pp. 74-81.
7. S. Carlsson and J. Sullivan, in *Action Recognition by Shape Matching to Key Frames - Proc. of IEEE CS Workshop on Models versus Exemplars in Computer Vision* (Florida, USA, 2002) pp. 263-270.
8. O. Masoud and N. Papanikolopoulos, *Recognizing Human Activities - Proc. of IEEE Conference on Advanced Video and Signal Based Surveillance* (Florida, USA, 2003) pp. 157-162. <http://dx.doi.org/10.1109/AVSS.2003.1217916>
9. L. I. Smith, *A tutorial on Principal Components Analysis* (2002). http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_component.pdf
10. O. Masoud, *Image Vision Comput.* **21**, 729 (2003). [http://dx.doi.org/10.1016/S0262-8856\(03\)00068-4](http://dx.doi.org/10.1016/S0262-8856(03)00068-4)
11. M. Turk and A. Pentland, *J. Cognitive Neurosci.* **3**, 71 (1991). <http://dx.doi.org/10.1162/jocn.1991.3.1.71>
12. M. Turk, *IEICE Transaction Information and System.* **E84-D**, 1586 (2001).
13. M. E. Algorri, and A. Escobar, in *Facial Gesture Recognition for Interactive Applications - IEEE Proc. of Fifth Mexican Int. Conf. in Computer Science* (Mexico, 2004) pp. 185-195.
14. C. Schuldt, I. Laptev, and B. Caputo, in *Recognizing Human Actions: A Local SVM Approach - Proc. 17th Int. Conf. Pattern Recognition* (2004) pp. 32-36. <http://dx.doi.org/10.1109/ICPR.2004.1334462>
15. K. Bouzned, M. Ramdani, N. Zermi, K. Mendaci, *J. e-Technol.* **4**, 138 (2013).
16. A. Jalal, M. Z. Uddin, J. T. Kim, and T. S. Kim, *Indoor Built Environ.* **21**, 184 (2012). <http://dx.doi.org/10.1177/1420326X11423163>
17. C. Vollmer, H. -M. Gross, and J. P. Eggert. in *Learning Features for Activity Recognition with Shift-invariant Sparse Coding - Int. Conf. on Artificial Neural Networks* (Springer, 2013), pp. 367-374.
18. S. Sadanand and J. J. Corso, in *Action Bank: A High-Level Representation of Activity in Video - Proc. of IEEE Conf. on Computer Vision and Pattern Recognition* (2012) pp. 1234-1241.