CLASSIFICATION OF FEATURE SELECTION BASED ON ARTIFICIAL NEURAL NETWORK

Hafizah Husain[1], Aini Hussain[1], Salina Abdul Samad[1],
and Nooritawati Md Tahir[2],
[1] Dept. of Electrical, Electronics and Systems, Faculty of Engineering,
Universiti Kebangsaan Malaysia
43600 Bangi, Selangor DE, MALAYSIA.
[2] Faculty of Electrical Engineering, Universiti Teknologi Mara (UiTM)
40450, Shah Alam Selangor DE, MALAYSIA
Corresponding Author: **aini@eng.ukm.my**

Abstract

Pattern recognition (PR) is the central in a variety of engineering applications. For this reason, it is indeed vital to develop efficient pattern recognition systems that facilitate decision making automatically and reliably. In this study, the implementation of PR system based on computational intelligence approach namely artificial neural network (ANN) is performed subsequent to selection of the best feature vectors. A framework to determine the best eigenvectors which we named as 'eigenpostures' of four main human postures specifically, standing, squatting/sitting, bending and lying based on the rules of thumb of Principal Component Analysis (PCA) has been developed. Accordingly, all three rules of PCA namely the KG-rule, Cumulative Variance and the Scree test suggest retaining only 35 main principal component or 'eigenpostures'. Next, these 'eigenpostures' are statistically analyzed via Analysis of Variance (ANOVA) prior to classification. Thus, the most relevant component of the selected eigenpostures can be determined. Both categories of 'eigenpostures' prior to ANOVA as well as after ANOVA served as inputs to the ANN classifier to verify the effectiveness of feature selection based on statistical analysis. Results attained confirmed that the statistical analysis has enabled us to perform effectively the selection of eigenpostures for classification of four types of human postures.

1. Introduction

Pattern Recognition generally refers to assigning an object to a so far unknown class of objects and identifying an object as a member of the already known class [1]. This field of research has proven that it can solve a broad range of problems related to its use as one of the principal tools in human decision-making tasks, assisting medical practitioners in the diagnosis of diseases, medical waveform classification such as EEG and ECG, computer vision field for face recognition [2], [3] [4], bioinformatics for sequence of DNA or protein analysis, biometric recognition such as face, iris, fingerprint or voice and text/document classification. Some of the best known approaches for PR are syntactic matching, statistical classification, template matching and computational intelligence methods. Such methods are able to perform classification from labeled training data sets as well as to explore structures and classes in unlabelled data. It is well known that one of the main tasks of a PR system is to determine structure in a data set to perform classification over a certain group of elements known as patterns. Patterns are entities characterized by a series of features [1].

The PR system can be divided in three principal stages namely data acquisition, feature extraction and selection followed by classification. In the data acquisition stage, the input data are gathered and converted into a suitable form for machine processing. Feature extraction and feature selection is mainly concerned with the reduction of space dimensionality. In the classification module, two different modes of operation is performed; the learning mode and the decision making mode. In the learning mode, this module is trained to partition the feature space. This means that some parameters in this module are adjusted to produce a correct output over the training samples. In the decision mode, the input patterns are assigned to a specific class based on the parameters that were learned in the previous mode.

One of the issues that require careful attention in a PR system is feature extraction and selection. Feature selection entails the task to select a subset amongst a set of candidate features that performs best under a classification system. This procedure can reduce not only the cost of recognition by reducing the number of features that need to be collected, but in some cases it can also provide better classification accuracy [1].

The structure of this paper is as follows. Section II discusses the methodology, section III consists of results and discussion and finally section IV concludes our findings.

2. Methodology

A. System Overview

Figure 1 depicts an overview of the overall system that outlines the basic structure. It consists of the following steps; pre-processing, feature extraction, feature selection using the rules of thumbs of PCA followed by ANOVA prior to classification. The pre-processing stage extracts the silhouette of a person using the binary image extraction process which consists of background differencing followed by thresholding to obtain a binary mask of the foreground region. In order to remove noise, median filtering and morphological operations are utilized.

Next, the feature extraction component functions by projecting the training images onto a feature space that spans the significant variations among known images. The significant features, which we termed as 'eigenpostures' are the eigenvectors (principal components) of the set of images. Detail description of the eigenpostures approach can be found in [2] [5]. The eigenpostures will undergo the first stage of feature selection process according to the three rules of thumb that will produce the first category of eigenpostures. Next, we will deem further by applying ANOVA to these eigenpostures labelled as category II. In doing so, the most relevant component of the selected eigenpostures for classification can be determined. Both categories of eigenpostures will act as inputs to the ANN classifier.

B. Selection of Eigenpostures

It is a well-known fact that the major goal for using PCA is to replace the *p*-dimensional feature space with a much smaller *m*-dimensional feature space, which nevertheless discards little information. For most empirical data, a large part of the total variance can be sufficiently approximated with the first few principal components only. However, the actual number of principal components needed remains obscure. In the literature, several rules of thumb have been proposed that include the followings:

i) Kaiser Gutman (KG) rule - The KG rule states that any PC with a variance of less than one contain less information than the original variables and is therefore not worth retaining. In other words, the KG-rule retains only those PCs whose variances, i.e. eigenvalues that are ≥ 1. Nevertheless, for large variable spaces *p*, the KG-rule usually retains too many PCs [6] [7].

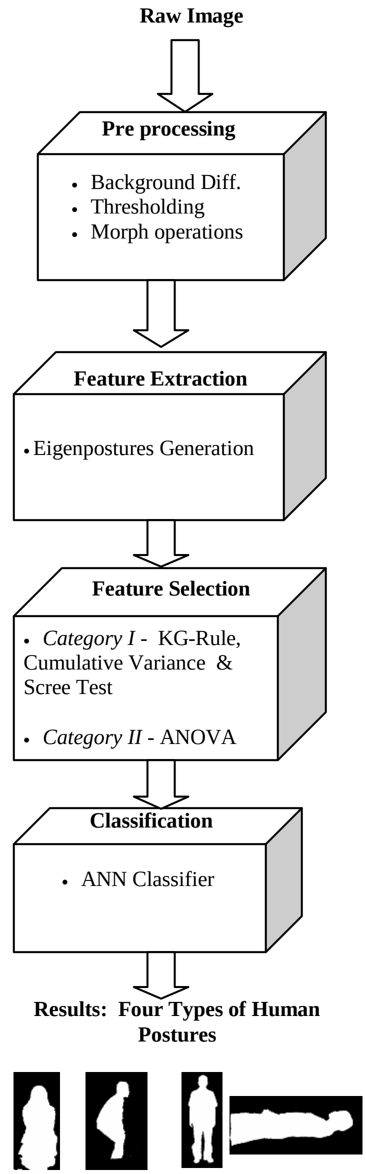

Figure 1: Overview of the overall system

ii) Cumulative Variance - The criterion for choosing m is to select a cumulative variance threshold, t where t is at certain percentage of the total variance that the first m PCs should account for. The required number of PCs is then the smallest value of m for which the chosen percentage is exceeded [7]. From PCA theory, the variance of the i-th PC (eigenvector) is equal to its corresponding eigenvalues λ_i. The total variance T_p can thus be calculated as:

$$T_p = \sum_{i=1}^{p} \lambda_i \quad (1)$$

Since PCs are successively chosen to have the largest possible variance, the obvious definition of the cumulative variance accounted for by the first k PCs is therefore

$$t_k = \frac{1}{T_p} \sum_{i=1}^{k} \lambda_i \quad (2)$$

and m is the smallest value k for which $t_k > t$.

iii) Scree Test - It involves looking at the plot of the eigenvalues λ_i against the factor number k. The Scree Test involves a certain degree of subjectivity since there is no formal numerical cut-off based on the λ_i. The idea behind the Scree Test is that important factors have a large eigenvalue and as such explain a large part of the total variance. If the eigenvalues are plotted, they form a curve heading towards almost 0% variance explained by the last dimension. Thus, the point at which the curve levels-out, sometimes referred to as the 'elbow' indicates the number of useful PCs, which are present in the data [7].

C. Statistical Analysis

ANOVA is a standard technique for measuring the statistical significance of a set of independent variables. It takes a single feature and the class associated with the data samples and measures the significance of the class variables in predicting the means of the feature. The measure that ANOVA produces is the p-value for the feature set. In doing so, the groups that differ significantly are revealed. This will determine the optimized number of eigenpostures that will act as inputs to the ANN for classification of the four main postures.

D. ANN Classifier

ANN is a popular heuristic technique that can deal with complex non linear problem even if the problem is extremely complex to be translated in analytical form. It deals with the training and testing processes before a network can be precisely developed to perform the desired task. The most exhaustive task in ANN is the training process that requires numerous training patterns with informative features or variables. Hence, feature extraction and selection can be utilized to attain the most informative variables that will speed up the convergence process.

3. Experiments and Results

In this study, the aim is to test the validity of eigenpostures that have experienced the feature extraction and both feature selection phases. A collection of 400 images of various human postures constitutes the database to generate the eigenpostures as shown in Figure 2. The various postures include both standing and non-standing positions, bending and lying with the human subjects are either facing front or side with no restriction impose on the type of clothing being worn. Initially, each image has m x n pixels, but eventually reshaped to a column vector of 1 x mn. Then, the eigenvectors and eigenvalues are computed according to [5]. Implementing the three rules mentioned previously, we select the most suitable eigenpostures required as inputs to the classification system. In other words, we select the most relevant eigenvalues or PCs to be retained and used as inputs to the classifier.

Fig. 2: Some of the human shapes images utilized

The three rules mentioned previously are implemented to determine the most apposite eigenpostures required as inputs to the classification system. The outcome is as tabulated in Table 1. Firstly, from the PCA results, applying KG rule that suggests retaining all eigenvalues > 1 results in thirty-five PCs to be considered as significant components. Next, we consider the cumulative variance rule of thumb as our feature selection basis to determine the optimum number of eigenpostures or PCs. From Table 1, the overall cumulative variance of the eigenpostures is shown. As suggested in [7], a threshold t of between 80%-90% can be considered to determine factor number, k. In this case, an 80% criterion would result in k equals 34 as tabulated in Table 1. Finally, the Scree test outcome in Table 1 illustrated the decrease in magnitude for successive eigenvalues implies that the first few principal

TABLE 1
The Significant Eigenpostures Using The Kg Rule, Cummulative Variance And Scree Test

Factor *k*	Eigen value	Cumulative Variance	Scree Test
1	44.37	19.059	19.059
2	23.64	29.289	10.229
3	17.48	36.792	7.5032
4	12.25	42.074	5.2822
5	10.32	46.591	4.5166
6	9.37	50.671	4.08
7	5.77	53.214	2.5429
8	5.64	55.688	2.4741
9	5.21	58.015	2.3272
10	3.89	59.713	1.6985
11	3.79	61.388	1.6749
12	3.17	62.814	1.4258
13	3.07	64.177	1.3629
14	2.90	65.465	1.2879
15	2.64	66.673	1.2079
16	2.60	67.857	1.1841
17	2.48	69.01	1.1528
18	2.39	70.105	1.0947
19	2.19	71.099	0.99471
20	2.02	72.045	0.94541
21	1.97	72.949	0.90395
22	1.77	73.772	0.82319
23	1.71	74.571	0.79945
24	1.67	75.312	0.74046
25	1.63	76.038	0.72617
26	1.57	76.73	0.69188
27	1.42	77.358	0.62865
28	1.35	77.977	0.6185
29	1.28	78.551	0.57428
30	1.26	79.121	0.56953
31	1.20	79.676	0.55543
32	1.14	79.702	0.52608
33	1.11	79.79	0.52042
34	1.06	80.21	0.52036
35	1.01	81.678	0.52025
36	0.984	82.142	0.5017
37	0.982	82.593	0.45103
38	0.981	83.02	0.42684
39	0.981	83.436	0.41549
40	0.980	83.837	0.40115

components can approximate a large part of the original data's variance. In this case, decision to retain the first thirty-five PCs is appropriate and they reasonably represent good approximation of the original data set. These eigenpostures are known as Category I. The Category I eigenpostures will undergo the statistical analysis prior to classification.

Accordingly, we determine the statistical significance of all Category I eigenpostures of the four main postures using ANOVA. In this analysis, null hypothesis will be discarded for p-value near zero and suggests that at least one sample mean is significantly different from the other sample means. Hence, from the ANOVA test, at a significant level of $\alpha = 0.05$, we anticipate that the p-values for eigenpostures 1-9, 11-13, 15-18, and 20-22 are numerically indistinguishable from zero. As a result, the ANOVA test has lucratively reduced the feature vectors to nineteen or 54% of the initial feature extraction quantity and these eigenpostures are known as Category II eigenpostures.

To estimate the classifier generalization error, the training data set was re-sampled using the *k*-fold cross-validation method. A *k*-fold cross-validation divides the training data into *k* subsets. Then, k-1 subsets are used for training and the remaining one subset is used as test data set to predict the classification error. The whole process repeats itself *k* times until each individual subset has been used once [8]. In this study, the classifier performance is estimated using a 5-fold cross-validation in which each posture data were divided equally into five subsets. Therefore, in each fold there will be 80 postures in each subset representing the four posture classes.

As aforementioned, ANN is chosen as our classifier in this study. A three-layer NN with weights adjusted using the Levenberg-Marquardt was trained to determine the relationship between the selected eigenpostures and the respective four posture classes.

The classification result of both categories of eigenpostures via ANN is as illustrated in Table 2. As can be seen, lying postures gained 100% accuracy rate for both categories of eigenpostures. This is due to the nature of lying position that is extremely distinct as compared to the other three postures. As for the bending posture, Category II achieved better recognition accuracy specifically 98% whilst Category I classification rate is only 94%. Further, for sitting posture, both categories achieved equal recognition rate that is 98%. For the standing posture, once again Category II eigenpostures attained perfect classification while category I gained 99%. Overall and as expected, Category II eigenpostures performed the best with an average recognition rate of 99%.

TABLE 2

CONFUSION MATRIX FOR POSTURE RECOGNITION BASED ON ANN

ACTUAL CATEGORY	**PREDICTED CATEGORY**							
	Category I Eigenpostures (35 eigenpostures)				**Category II Eigenpostures (19 eigenpostures)**			
	BEND	**SIT**	**STAND**	**LYING**	**BEND**	**SIT**	**STAND**	**LYING**
BEND	**94**	2	4	0	**98**	0	2	0
SIT	0	**98**	2	0	2	**98**	0	0
STAND	1	0	**99**	0	0	0	**100**	0
LYING	0	0	0	**100**	0		0	**100**

4. Conclusions

In conclusion, a task of classifying four main human postures namely standing, sitting, bending and lying position based on eigenpostures analysis is presented. The initial thirty five feature vectors suggested by the rule of thumbs of PCA namely the KG rule, Scree Test and Cumulative Variance are trimmed down to a new subset of nineteen feature vectors via the ANOVA. This suggests that the eigenspace technique along with PCA rules of thumbs followed by statistical data analysis are an appropriate technique for feature selection for posture recognition, which can lead to a wide variety of applications such as security systems, intruder's alertness, gait analysis and action recognition. The rules of thumb of PCA and ANOVA analysis have facilitated us to achieve the selection of eigenpostures for classification of human postures efficiently.

References

[1] Ani1 K. Jain, Robert P.W. Duin, and Jianchang Mao (2000) Statistical Pattern Recognition: A Review. Proceeding of IEEE Transactions On Pattern Analysis and Machine Intelligence Vol. 22(1): 4-37.

[2] M Turk and A. Pentland. "Eigenfaces for Recognition". *Journal of Cognitive Neuroscience.* 3(1): 71-86, 1991.

[3] M Turk. "A Random Walk through Eigenspace". Special Issue on Machine Vision Applications, *IEICE Transaction Information and System.* E84-D (12): 1586-1595, 2001.

[4] M.E. Algorri and A. Escobar.. "Facial Gesture Recognition for Interactive Applications". *IEEE Proceedings of Fifth Mexican International Conference in Computer Science, Mexico.*185-195, 2004.

[5] Nooritawati Md Tahir, Aini Hussain, Salina Abdul Samad, Hafizah Husain and Mohd Marzuki Mustafa, "Eigenposture for Classification*" in Journal of Applied Sciences* 6(2): 419-424, 2006.

[6] Nooritawati Md Tahir, Aini Hussain, Salina Abdul Samad, Khairul Anuar Ishak and Rosmawati Abdul Halim. "Feature Selection for Classification Using Decision Tree" *in Proceeding of SCORed 2006,* Malaysia. 99-102, 2006.

[7] IT Jolliffe. *Principal Component Analysis*, Springer Series. pp 111-137, 2001.

[8] C. Bishop, *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, 1995.