CS 229 Course Project Report:

Comparative study of SVM and ANN in face recognition

NAME: XIAOPENG XU KAUST ID: 129052

Abstract:

Face perception is a very important component of human cognition. Faces are rich in information about individual identity. Face recognition is necessary of many biometric systems[1]. During the past decades, face recognition was also one of the most important and successful applications of machine learning and computer vision[2].

The basic procedure of face recognition contains two steps: face feature selection and classification. PCA and ICA are two very popular feature selection methods, which are widely used in machine learning and signal processing. ANN and SVM are two very popular machine-learning techniques that are widely used. In this study, I use ANN and SVM to do face recognition with PCA and ICA for feature selection. My study shows that SVM gives a better classification after PCA and ICA feature selection than ANN. Also ICA gives a better representation of data features than PCA considering variances of classification accuracy.

Introduction

In recent days, the need of biometric security system is heightened for providing safety and security against terrorist attacks, robbery, etc. [1, 3]. The demand of biometric system has risen due to its strength, efficiency and easy availability. One of the most effective, highly authenticated and easily adaptable biometric security systems is facial feature recognition.

However, face recognition is not easy due to its wide varieties of noise. Basically there are 6 sources of noise, illumination, pose, expressions, ageing, occlusion, and low resolution[3]. The variation in illumination changes the appearance of the face drastically[4]. It is found that the difference between two images of the same person taken under varying illumination is greater than the difference between the images of two different persons under same illumination. Pose variations in an image is also a matter of concern in face recognition[5]. The changes in the posture strike a serious problem for the identification of the input image. This is because the available image in the database may have only the frontal view of the face, which may differ in pose with the input image and so may result in faulty identification. With the increasing age, the appearance of a person also changes, which affect the face recognition system. The occlusion-caused unavailability of the whole input face is also one of the important challenges. This is when some parts of the face are missing, for e.g. when an image is captured from a surveillance camera; the face in the image lacks some parts. This is also possible due to glasses, beard, moustache, scarf, etc. Such a problem can severely affect the classification process. Last but not least, images taken from a surveillance camera generally consists of very small face area and so it is of very low resolution[6]. Such a low-resolution face image consists of very limited information, as most of the details are lost. This can drop down the recognition rate drastically.

Another difficulty in face recognition is caused by its high dimensionality[2]. Normally, an image data is composed of pixels. To process this data, we need to transform it into an array, of which the length equal to the number of pixels. This results in a very big dataset, that is hard to process without some feature selection and extraction methods.

Besides, one difficulty for classification comes from the nature of face recognition -- the classification is a multi-class classification. So traditional binary class classification methods such as SVM are not directly applicable[7]. Also as the number of subjects (persons) in the biometric systems is very large, we have many classes. In order to have a good performance, we need classification methods, which can handle multi-class classification with a huge number of classes.

In this study, I chose Yale face database as my dataset. I used Principal Component Analysis (PCA) and Independent Component Analysis (ICA) for feature selection, and use one-vs-all SVM and ANN for classification. My study shows that SVM gives a better classification after PCA and ICA feature selection than ANN. Also ICA gives a better representation of data features than PCA considering variances of classification accuracy.

Feature Selection

Face feature selection or extraction is an important preprocessing task in biometric systems based on face images. As face data is of high dimension, and contains many noises, face feature selection is necessary to reduce the dimension and remove some noises. Methods of facial feature selection can be divided into 4 groups dependent on the information used: appearance-based, geometry-based, knowledge-based, and 3D Vision-based[1].

In appearance-based approach, values of luminance or gradients in the given area are analyzed. Object features are usually selected by reducing the image data dimensionality. In geometry-based approach, spatial relationships of the face parts are assumed. These relations can be distances, angles, etc., and they are usually described as a graph. In knowledge-based approach, our knowledge about human face is used for face features extraction. This knowledge can be color, symmetry, placement of face parts, such as eyes, mouth, etc. For 3D vision-based approach, 3D model of faces are constructed for face analysis.

In my project, I used PCA and ICA, both of which are from appearance-based approaches, to select features from my dataset.

Principal Component Analysis (PCA)

One of the most popular and simplest image analysis methods is Principal Components Analysis[1, 8-10]. In this algorithm mean vector μ and the covariance matrix of the training set are computed. For the covariance matrix eigenvectors sorted by the corresponding eigenvalues are extracted. Eigenvectors associated with the highest eigenvalues carry most of the object energy. Matrix A is created by choosing k first eigenvectors. Data X representation in the new space is defined by the formula:

$$X' = A^T(X - \mu)$$

This process leads to the dimensionality reduction without loosing most relevant information. Matrix A is usually computed by the SVD decomposition and choosing eigenvectors corresponding to the highest variances if the dimensionality of data is very high, such as image data.

Independent Component Analysis (ICA)

The other method of feature selection method I used in facial features detection is Independent Components Analysis (ICA)[1, 8, 11-13]. Its functionality can be described on the blind source separation. Assuming, that there are n independent scalar sources of the signal x (t) for i = 1…n, where t stands for the time index i, 1 ≤ t ≤ T. For k-dimensional data vector derived from the sensor in the time index, following equation can be given:

$$s(t) = Ax(t)$$

where A is a k by n matrix.

The goal of the ICA is to find independent components derived from the observation s. So for image analysis, we have the assumption that all images from our dataset are composed of independent components or sources, as shown in Figure 1. In my analysis, I use the weights of ICs for each sample as the features in my classification.

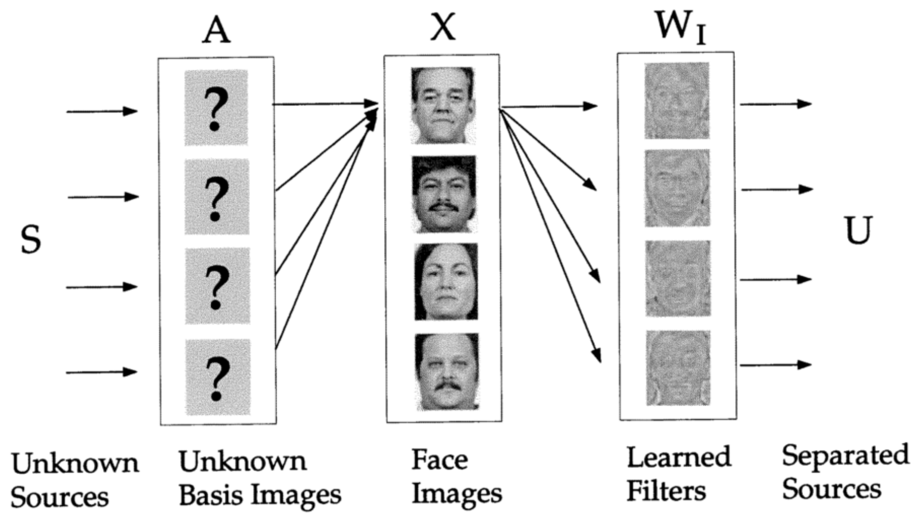

Figure 1. Find independent components from mixed signals

Classification

In classification using the face features, there are four popular classifiers: Support Vector Machine (SVM), Hidden Markov Model (HMM)[14, 15], Artificial Neural Network (ANN), and Self-organizing Map (SOM)[16, 17]. In this project, I used SVM and ANN as classifier to do the classification.

Support Vector Machine (SVM)

One of the most popular classification methods in face recognition is Support Vector Machine (SVM)[1, 7, 8, 10, 18, 19]. For the linear classification of the linearly separable data SVM maximizes margins between two half-spaces given by the equations:

$$H_1: X_i w + b \geq +1 \; for \; y_i = +1,$$
$$H_2: X_i w + b \leq -1 \; for \; y_i = -1$$

where w and b are the parameters of the hyper-plane parallel to and fitted in the middle of the half-spaces H_1 and H_2. X_i corresponds to the i-th support vector and y_i corresponds to its class: +1 or −1. Margin between hyper-planes is equal $m = \frac{2}{\|w\|}$. The name of the method is taken from the fact, that H_1 and H_2 lean on some of the data samples -- support vectors.

However, SVM is typically designed for binary classification. So some modifications are needed to make it applicable for multi-class classification[10]. Usually there are three groups of strategies for multi-class classification using SVM: one-vs-all strategy, pairwise-strategy, and Bayesian SVM. One-vs-all strategy construct a SVM model for each class, and for a test sample, it finds the probability of the sample to belong to each class, and then assign the sample with the corresponding label. Pairwise-strategy usually constructs a binary tree for classification, and decides whether a sample belongs to a class following the tree. Bayesian SVM methods combines Bayesian with SVM and can achieve the same purpose without too much training[19].

In my experiments, I use one-vs-all strategy to implement my SVM as this approach needs only P training and P classification, where P is the number of classes. For other two approaches, the training process is very expensive, though it may give a little lighter expense in testing.

Artificial Neural Network (ANN)

Artificial Neuron Network is another widely used classifier for face recognition[2, 13, 20-23]. ANN is a classifier based on modeling of human decision. It is a multi-element structure processing the data using neurons. Neurons are connected with defined weight given in the training process. In my experiments, I use a 2-layer ANN for my classification – one hidden layer, and one output layer.

Experiments

In my experiment, I use the Yale face database as my dataset. The Yale face database contains 165 gray-scale images in GIF format of 15 individuals. Each image is of 320x243 pixels. There are 11 images for each subject, one per different facial expression or configuration: center-light, with glasses, happy, left-light, with no glasses, normal, right-light, sad, sleepy, surprised, and wink.

In the experiment, I randomly partition the 11 images of each subject into 3 sets, 7 images for training set, 2 for validation set, and 2 for testing set. I run each method (with both feature selection and classification) 10 times and study the mean value, standard deviation, and maximum value of accuracy.

Polynomial SVM with different degree for PCA

To show how the performance changes with the number of Principal Components reduces, I run SVM with different percentage of Principal Components (PCs). As Figure 2 shows, the accuracy I get from SVM decreases with the number of PCs reduces.

To verify which SVM kernel is best suitable for selected features by PCA, I run SVM with different parameters many times. Figure 2 shows accuracy I got with different degrees of polynomial kernel. My result shows that with PCA feature selection, if the PCs are enough, polynomial kernel with degree 1 (that is linear kernel) gives me the best performance, and with the polynomial SVM degree going higher, the classification result get worse.

However, with number of PCs reduces to represent 80% of variance, polynomial SVM gives me a slightly different performance with different degree, degree 3 performs better than degree 2 or degree 1, but degree 5 deteriorates a lot. This may because that with 80% of variances represented, the distribution of the features is not easily separable, so non-linear kernel, with degree 2 and degree 3, shows a better performance. However, with degree 5, there might be an over-fitting for the training data, which cause the deterioration in classification.

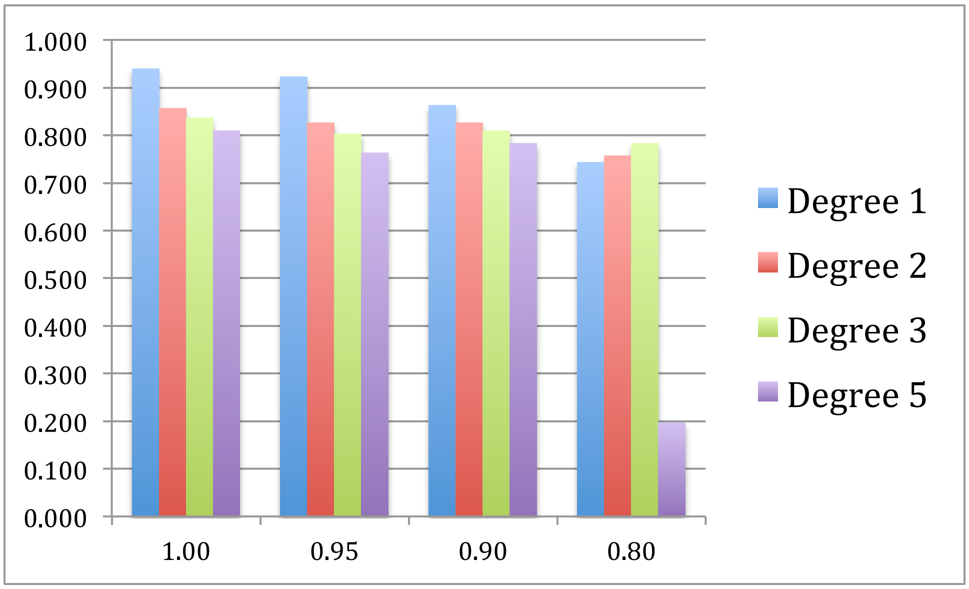

Figure 2. Polynomial SVM kernel with different degree

Comparison of SVM and ANN with PCA

Also I try to compare SVM with different kernels, and also with ANN. I run each kernel 10 times with different number of PCs, which represent 80%, 90%, 95%, and 100% variances.

As shown in figure 3, polynomial kernel of SVM shows nearly same performance as linear because the best degree selected for polynomial is 1, which is just a linear kernel. RBF kernel SVM and ANN shows a poorer performance than that of linear and polynomial. The reason of RBF might be that it over-fits the training data too much that it is not applicable for testing data. For ANN, because it is running on multiple class classification, it may stuck on some local minimum during gradient descent back-propagation process, which makes the prediction varies a lot (from about 0.1 to more than 0.9). Also it may have some over-fitting if the training of the model is too accurate, which make it not applicable to testing data.

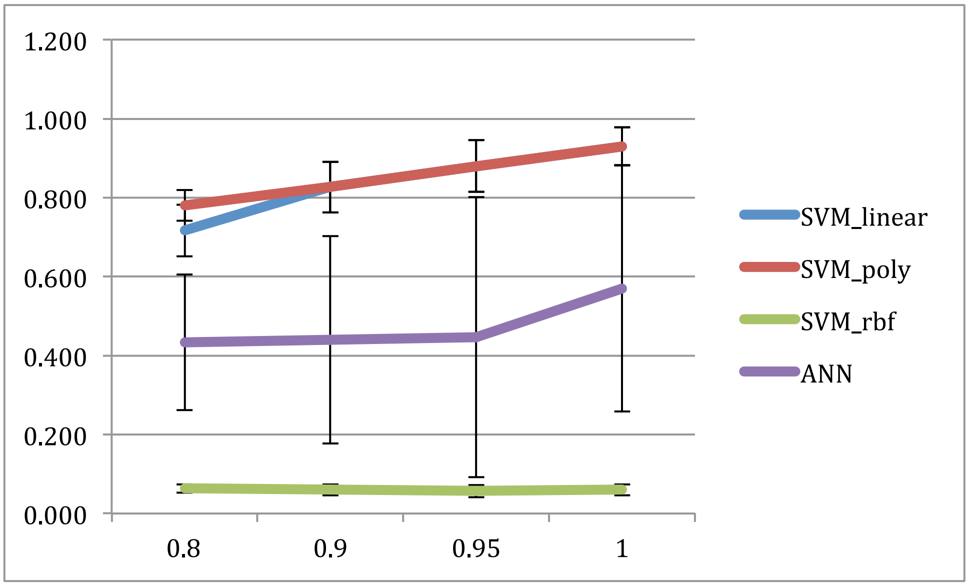

Figure 3. Accuracy of PCA with SVM and ANN as classifier (MEAN)

Comparison of ICA and PCA with SVM and ANN

To run ICA, I first run PCA to reduce the dimensionality, as the dimension of the data is too high which makes it impossible to run ICA directly. Then I use the weights I get for each object with respect to different Independent Components (ICs) as input features for classifiers SVM and ANN.

My results are shown on Figure 4 and Figure 5. For both ICA and PCA, SVM with linear kernel performs better than ANN regarding to both mean (Figure 4) and max (Figure 5). Comparison of ICA and PCA shows that though improve of mean accuracy is not obvious; ICA gives a smaller variance over PCA. This is because the classifier we used can handle complex input features. For PCA, it just selected the variance of the data, some may be useful, but some are not. So this caused big variance of accuracy. For ICA, features are well represented, and this gives a good performance regarding to average accuracy, and also with improvements considering variance.

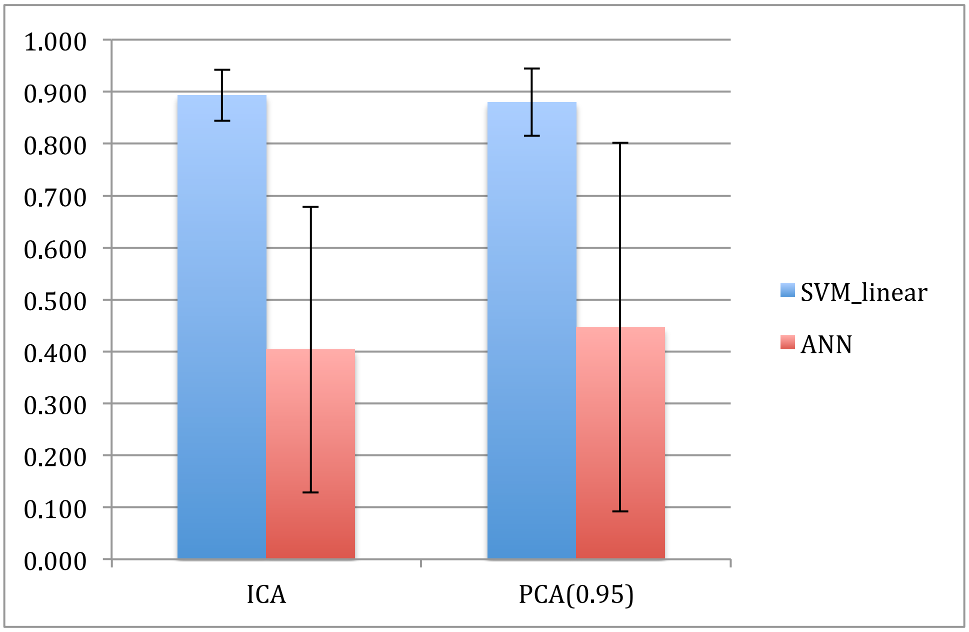

Figure 4. Accuracy of ICA and PCA with SVM and ANN as classifier (MEAN)

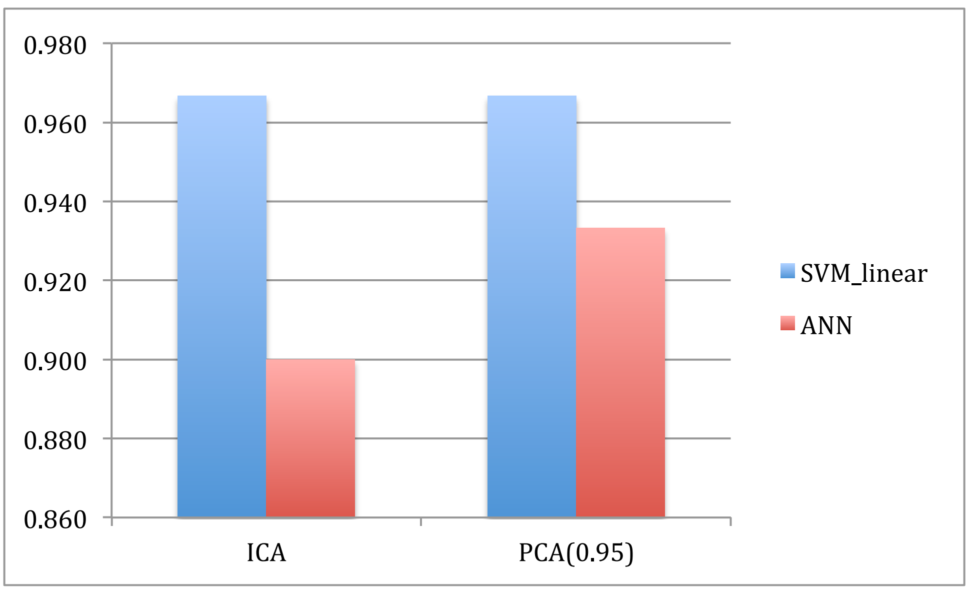

Figure 5. Accuracy of ICA and PCA with SVM and ANN as classifier (MAX)

Conclusion

In this study, I used PCA and ICA for feature selection and SVM and ANN for classification to do face recognition using Yale face database. My experiment shows that after PCA for feature selection, simple classifier gives me a better classification results. Here, classification accuracy I got from SVM with linear kernel is better than that from more complex kernels such as polynomial and RBF. Also, by comparison of ICA and PCA with different classifiers, linear kernel SVM and ANN, I find that ICA gives a better representation of image features than PCA in terms of variance of accuracy in classification. Comparing linear kernel SVM and ANN, linear kernel SVM gives me a better classification results.

Reference:

1. Naruniec, J., *A Survey on Facial Features Detection.* International Journal of Electronics and Telecommunications, 2010. **56**(3).
2. W. ZHAO, R.C., P. J. PHILLIPS, A. ROSENFELD, *Face Recognition_ A Literature Survey.* ACM Computing Surveys,, 2003.
3. Ripal Patel, N.R., Ami Shah, *Comparative Analysis of Face Recognition Approaches: A Survey.* International Journal of Computer Applications, 2012.
4. Li, S.Z., et al., *Illumination invariant face recognition using near-infrared images.* Ieee Transactions on Pattern Analysis and Machine Intelligence, 2007. **29**(4): p. 627-639.
5. Singh, R., et al., *A mosaicing scheme for pose-invariant face recognition.* Ieee Transactions on Systems Man and Cybernetics Part B-Cybernetics, 2007. **37**(5): p. 1212-1225.
6. Zou, W.W.W. and P.C. Yuen, *Very Low Resolution Face Recognition Problem.* Ieee Transactions on Image Processing, 2012. **21**(1): p. 327-340.
7. Phillips, P.J., *Support Vector Machines Applied to Face Recognition.* MIT Press, 1999.
8. Déniz, O., M. Castrillón, and M. Hernández, *Face recognition using independent component analysis and support vector machines.* Pattern Recognition Letters, 2003. **24**(13): p. 2153-2157.
9. Hyeonjoon Moon, P.j.P., *Computational and performance aspects of PCA-based face-recognition algorithms.* Perception, 2001.
10. Md. Omar Faruqe, M.A.M.H., *Face Recognition Using PCA and SVM.* 3rd International Conference on Anti-counterfeiting, Security, and Identification in Communication, 2009.
11. Marian Stewart Bartlett, J.R.M., Terrence J. Sejnowski, *Face Recognition by Independent Component Analysis.* IEEE TRANSACTIONS ON NEURAL NETWORKS, 2002.
12. Comon, P., *Independent component analysis, A new concept?* Signal Processing, 1994.
13. Juha Karhunen, E.O., Liuyue Wang, Jyrki Joutsensalo, *A Class of Neural Networks for Independent Component Analysis.* IEEE TRANSACTIONS ON NEURAL NETWORKS, 1997.

14. Bevilacqua, V., et al., *Pseudo 2D hidden Markov models for face recognition using neural network coefficients.* 2007 IEEE Workshop on Automatic Identification Advanced Technologies, Proceedings, 2007: p. 107-111.
15. Alhadi, F.H., W. Fakhr, and A. Farag, *Hidden Markov Models for face recognition.* Proceedings of the IASTED International Conference on Computational Intelligence, 2005: p. 409-413.
16. Chen, Q., et al., *VQ-based Face Recognition Algorithm Using Code Pattern Classification and Self-Organizing Maps.* Icsp: 2008 9th International Conference on Signal Processing, Vols 1-5, Proceedings, 2008: p. 2060-2065.
17. Li, C.Y., F. Liu, and Y.X. Xie, *Face recognition using self-organizing feature maps and support vector machines.* Iccima 2003: Fifth International Conference on Computational Intelligence and Multimedia Applications, Proceedings, 2003: p. 37-42.
18. Bernd Heisele, Purdy Ho, and T. Poggio, *Face Recognition with Support Vector Machines: Global versus Component-based Approach.* Eighth IEEE International Conference on Computer Vision, 2001.
19. Zhifeng Li, X.T., *Bayesian Face Recognition Using Support Vector Machine and Face Clustering.* IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2004.
20. Leonard0 Franco, A.T., *A Neural Network Facial Expression Recognition System using Unsupervised Local Processing.* Image and Signal Processing and Analysis, 2001.
21. Meng Joo, E., et al., *Face Recognition With Radial Basis Function (RBF) Neural Networks.* IEEE TRANSACTIONS ON NEURAL NETWORKS, 2002.
22. Zhao, Z.-Q., D.-S. Huang, and B.-Y. Sun, *Human face recognition based on multi-features using neural networks committee.* Pattern Recognition Letters, 2004. **25**(12): p. 1351-1358.
23. Melin, P., C. Felix, and O. Castillo, *Face recognition using modular neural networks and the fuzzy Sugeno integral for response integration.* International Journal of Intelligent Systems, 2005. **20**(2): p. 275-291.