The comparison of Normal Bayes and SVM classifiers in the context of face shape recognition

Adam Schmidt[1]

Poznan University of Technology
Institute of Control and Information Engineering
str. Piotrowo 3a 60-965 Poznan, Poland
adam.schmidt@put.poznan.pl

Summary. In this paper the face recognition system based on the shape information extracted with the Active Shape Model is presented. Three different classification approaches have been used: the Normal Bayes Classifier, the Linear Support Vector Machine (LSVM) with a hard margin and the LSVM with a soft margin. The influence of the shape extraction algorithm parameters on the classification efficiency has been investigated. The experiments were conducted on a set of 3300 images of 100 people which ensures the statistical significance of the obtained results.

1 Introduction

In recent years the face recognition has been among the most actively developed methods of biometric identification. This popularity is caused by two main reasons. Firstly the face appearance is an important clue used by humans in visual identification. Secondly the automatic face recognition is a non-invasive method and can be used without subject's knowledge or permission.

Defining the proper feature set plays a crucial part in the development of an efficient face recognition system. The features used should be discriminating enough to appropriately model the characteristic description of a person. Moreover, it is important to ensure that the selected features can be extracted under varying conditions of the image registration.

The Active Shape Model (ASM) introduced by Cootes et al. [1][2] is a powerful shape extraction method. It is especially well suited to the analysis of complex, flexible objects such as human faces. The success of the ASM encouraged many researchers to further improve this method.

Zhao et al.[3] proposed the Weighted Active Shape Model, which utilized the information on the contour points stability. After fitting to the image contour points were projected to the shape space in a way that minimized the reconstruction error of points with the smallest movement. Additionally the authors introduced an image match measure of the whole shape rather than its particular points. This approach facilitated choosing the final contour.

Zuo and de With[4] noticed that the Active Shapes needed precise initialization. To provide good initial conditions for their algorithm they matched a face template to the gradient image. They also resigned from using the gradient profiles to fit contours to the image. Instead they used a $N \times N$ pixels neighborhood decomposed with the Haar wavelets.

The system developed by Ge et al. [5] strived to improve the robustness of the Active Shapes to the face pose. As the first processing stage two face detectors were used: one detecting the frontal faces and the other detecting face profiles. Their responses were used to estimate the head rotation. The rotation angle was used to decide which of the ten previously trained Point Distribution Models should be used.

Wan et al. [6] observed that different parts of face contour model were differently disturbed by face pose changes and decided to split the face model into two submodels: one representing only the face outline and the second modelling the eyes, eyebrows, nose and mouth. To further increase the robustness of the system three independent models were created for the frontal and profile face views. The genetic algorithm with chromosomes describing both submodels parameters and the similarity transformation was used to fit contours to the image. The fitness function was based on the both submodels match measures and the third component describing the correctness of the submodels relative position. The whole procedure was computationaly expensive and authors admit that the slight improvement in contours quality was payed with a few times longer processing time.

Cristinnace and Cootes[7] used the Haar-like features and boosted classifiers to improve the accuracy of the ASM. They also proposed using boosted regression instead of classification to precisely locate contour points. This method outperformed the classic ASM w.r.t. both the accuracy and the processing time.

However, those authors focused on improving the accuracy of the model, i.e. reducing the difference between manually and automatically marked contours, without giving any thought to possible applications of the method. The goal of this research was to find out if the shape of the face and its features (i.e eyes, mouth, nose and eyebrows) extracted with the Active Shape Model (ASM) contains sufficient information for the successful face recognition. Three different classification methods were used: the Normal Bayes Classifier (NBC)[8][9], the Linear Support Vector Machine[10][11] with a hard margin (LSVM-C) and with a soft margin (LSVM-ν). Moreover, the influence of ASM parameters such as the subspace dimension, the gradient profile length and the type of tangent projection on the efficiency of the recognition system has been investigated.

This paper starts with a short review of the ASM. After that the setup of the experiment and the data used are described. The results are presented in the section 5 and the last section contains the concluding remarks and the future work.

2 Active Shape Model

The ASM is a shape extraction method statistically modelling the distribution of plausible shapes and the appearance of the contour points neighbourhood. It consists of two submodels: the Point Distribution Model (PDM) and the Local Structure Model (LSM).

The PDM is used to represent the variablity of shapes in the training set. The exemplary shapes are aligned to a common coordinate frame (by using the Generalized Procrustes Analysis). After that they can be projected on the space tangent to the mean shape in order reduce the nonlinearities. The tangent projection can be done by either elongating the shape vector (preserving its direction) or by projecting it in the direction parallel to the mean (minimal change). Then the Principal Components Analysis is used to reduce the dimensionality of the model and to supress the noise. The extracted shape is represented as a t-dimensional vector in the subspace defined by the selected Principal Components.

The purpose of the LSM is to model the typical neighborhood of contour points. It is achieved by sampling the gradient along the profiles perpendicular to the contour and passing through the model points. For each point the mean profile and the covariance matrix are estimated according to the training set. The complexity of the model can be set by changing the length of sampled profiles. This length is defined by the parameter k, which corresponds to the number of pixels sampled on each side of the contour. While fitting the shape to the image particular contour points are moved to the positions which minimize the Mahalanobis distance between the sampled profiles and those in the LSM.

The shape is extracted by consecutive moving of the contour points with the LSM and regularizing the shape with the PDM until convergence. To improve the robustness of the method the Gaussian image pyramid is created. The shape is first fitted to the coarsest image. After that it is scaled to the next pyramid level and the fitting procedure is repeated. This multi-resolution framework enables large movements during the initial phases and more subtle shape adjustments on the final pyramid stages.

3 Experiment

The experiments were conducted on a set of 3300 high resolution (2048×1536) images of 100 people (33 per person). For each person 22 images contained near-frontal views of face with neutral expression, the other 11 images featured faces with stronger pose deviations or non-neutral expressions (Figure 1). The images were divided into three sets: Learning Set (LS) - containing a half of the near-frontal images, Testing Set A (TSA) - the other half of near-frontal images and Testing Set B (TSB) - all of the non-standard views.

To facilitate creating of the PDM and LSM models the near-frontal images of 30 people were manually annotated with 166 points forming the face contour model. Three PDMs (differing in the tangent projection method) and ten LSM (with different length of the sampled profiles) were created.

The shapes were extracted for each combination of the tangent projection method (no projection, elongation or parallel projection), the number of principal components used ($t = \{10, 15, \ldots, 50\}$) and the gradient profile length ($k = \{1, 2, \ldots, 10\}$). The 5 level Gaussian pyramid was used and the ASM model was initialized with manually annotated eyes positions.

For each parameters combination shapes extracted from the LS set were used to train three classifiers: the NBC, the LSVM-C and the LSVM-ν. The SVMs were trained using 11-fold cross-validation to find optimal C or ν parameters. Those classifiers were than applied to the shapes from the TSA and TSB sets and the recognition rate (RR) was measured.

4 Results

Figure 2 presents the RR as a function of k and t parameters with no tangent projection used. The maximuml RR for different classifier and tangent projection method pairs are gathered in the Table 1. The maximum RR=98.4% in set TSA was obtained with the LSVM-ν for $k = 8$, $t = 20$ and parallel tangent projection. The best results (RR=66.4%) in the set TSB were obtained with the LSVM-C for $k = 10$, $t = 30$ and projection by elongation.

All classifier types gave similar, close to 100%, results in the set TSA. The RR in the set TSB were severely smaller and the best results of the NBC were about 5% smaller than results of the SVM classifiers.

Increasing the length of sampled profiles up to $k = 6$ significantly improved the RR. Further increase did not provide any noticeable efficiency boost.

The dimensionality of the PDM had strong influence on the RR of the NBC and LSVM-C classifiers. Initial increase (up to $t = 20$ for the TSA and $t = 30$ for the TSB) led to significant increase of the RR. For the $t =< 10, 20 >$ the LSVM-ν classifier performed better than the other two types

Incorporating any type of the tangent space projection in the PDM did not produce any observable effect (Table 1).

5 Conclusions

The presented results showed that the shape of face and its features carries enough information for efficient and reliable face recognition. The influence

Table 1. Maximum recognition rate for different classifiers and tangent projection methods

	TSA			TSB		
Tangent projection	NBC	LSVM-C	LSVM-ν	NBC	LSVM-C	LSVM-ν
None	97.7%	97.7%	98.3%	60.1%	65.8%	66.6%
Elongation	97.1%	97.1%	98.1%	60.3%	66.4%	65.7%
Parallel	97.6%	98.1%	98.4%	61.6%	65.3%	65.7%

of the ASM parameteres on the recognition rate has been assessed and the unimportance of the tangent space projection has been proved.

The high recognition rate obtained by the LSVM classifiers proves that classes form linearly separable clusters. Similar efficiency of the NBC classifier suggests that the decision surfaces can be described by hyperellipsoids.

The significant drop of the RR in the set TSB shows that the proposed system was susceptible to the face pose variations and expression changes. In the real world application the learning sets should contain images taken under different registration conditions to improve the reliability of the system. The shape information could be also used to estimate the face pose and to identify the facial expression in order to facilitate correct image registration.

Although the SVMs outperformed the NBC in the set TSB all classifiers had similar RR in the set TSA. This encourages to use the NBC classifer which is easier to train (in fact training is simply achieved by estimating each class mean shape and covariance). The NBC is also easily scalable - introducing new classes or removing existing does not involve retraining whole classifier. It is also possible to adaptively change the estimates of the mean and covariance to respond to the changes in subjects appearance.

Future work will concentrate on improving the robustness of the system w.r.t the face pose and expression changes. Further plans involve experimental installation of a system in a real world conditions and testing its efficiency in unconstrained environment.

References

1. Cootes T, Cooper D, Taylor C, Graham J (1995) Active shape models - their training and application. Computer Vision and Image Understanding, 61(1):38–59
2. Cootes T, Taylor C (2004) Statistical models of appearance for computer vision. Technical report, University of Manchester, Wolfson Image Analysis Unit, Imaging Science and Biomedical Engineering
3. Zhao M, Li S, Chen C, Bu J (2004), Shape Evaluation for Weighted Active Shape Models. In: Proc. of the Asian Conference on Computer Vision, 1074–1079

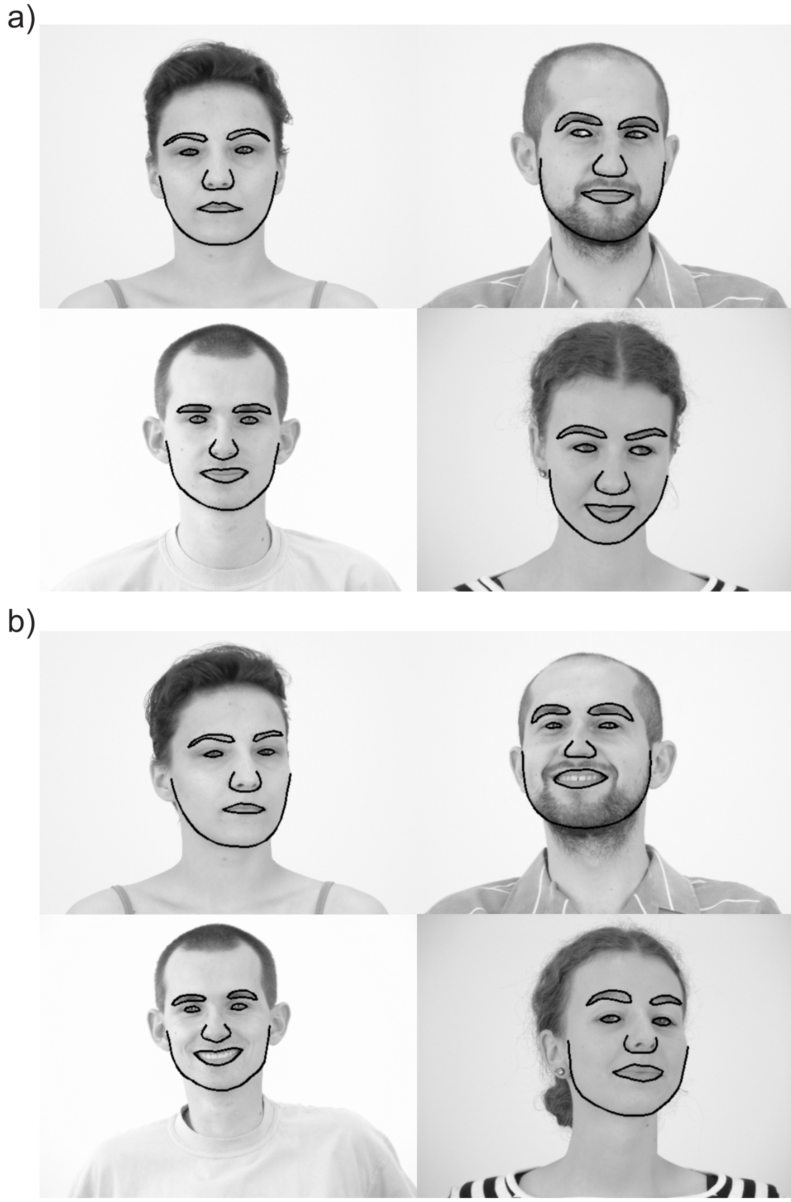

Fig. 1. Examples of the images with shapes extracted with k=6, t=40 and no projection: a) TSA, b): TSB

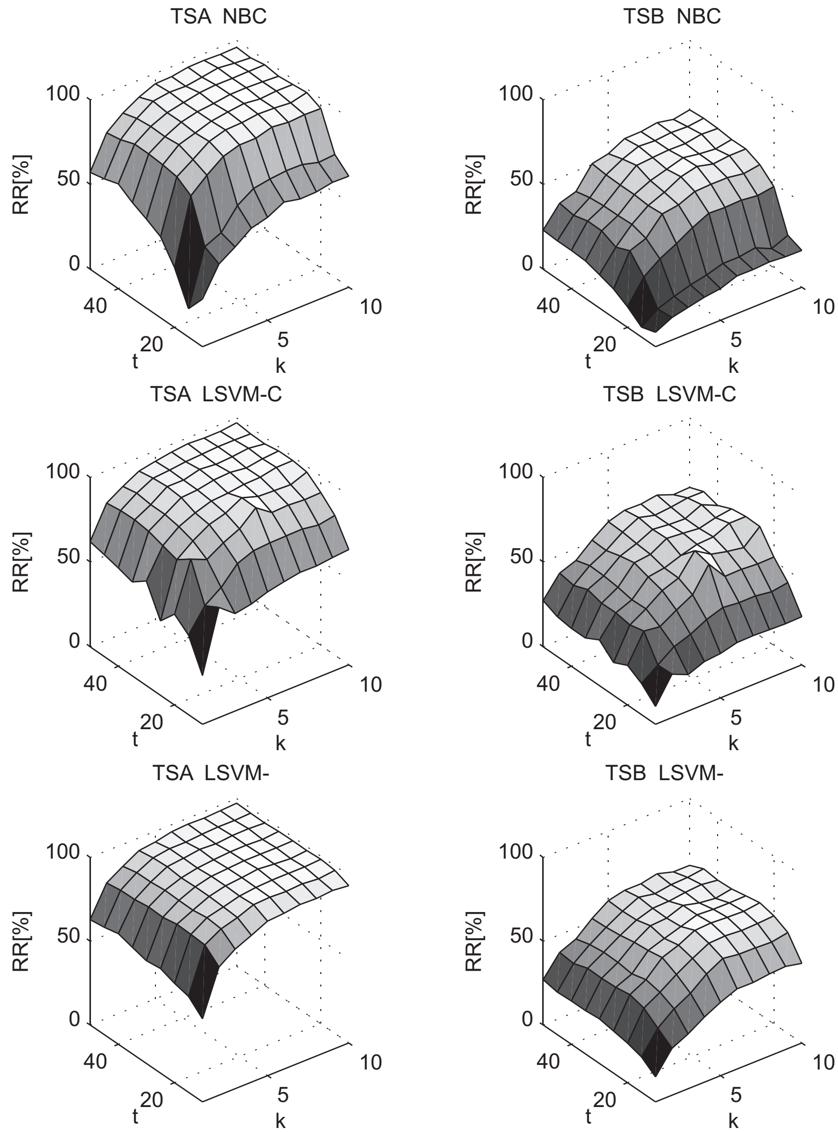

Fig. 2. Recognition rate of the NBC, SVM-C and SVM-ν classifiers, no tangent space projection. Left side: TSA, right side: TSB.

4. Zuo F, de With P (2004) Fast facial feature extraction using a deformable shape model with haar-wavelet based local texture attributes. In: Proc. of ICIP'04, 1425-1428
5. Ge X, Yang J, Zheng Z, Li F (2006) Multi-view based face chin contour extraction. Engineering Applications of Artificial Intelligence, 19:545–555
6. Wan K-W, Lam K-M, Ng K-C (2006) An accurate active shape model for facial feature extraction., Pattern Recognition Letters, 26(15), 2409–2423
7. Cristinacce D, Cootes T (2007) Boosted Regression Active Shape Models. In: Proc. British Machine Vision Conference 2007 2:880–889
8. Theodoridis S, Koutroumbas K (2003) Pattern Recognition. Elsevier/Academic Press, Amsterdam
9. Duda R, Hart P, Stork D (2001) Pattern Classification. John Wiley & Sons, Chichester
10. Vapnik V (2000) The Nature of Statistical Learning Theory,Springer, Berlin
11. Burges C (1998) A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery, 2:121–167