The Performance of Two Deformable Shape Models in the Context of the Face Recognition

Adam Schmidt and Andrzej Kasinski

Poznan University of Technology, Institute of Control and Information Engineering,
ul. Piotrowo 3a, 60-965 Poznan, Poland
{Adam.Schmidt,Andrzej.Kasinski}@put.poznan.pl
http://www.cie.put.poznan.pl

Abstract. In this paper we compare the performance of face recognition systems based on two deformable shape models and on three classification approaches. Face contours have been extracted by using two methods: the Active Shapes and the Bayesian Tangent Shapes. The Normal Bayes Classifiers and the Minimum Distance Classifiers (based on the Euclidean and Mahalanobis metrics) have been designed and then compared w.r.t. the face recognition efficiency. The influence of the parameters of the shape extraction algorithms on the efficiency of classifiers has been investigated. The proposed classifiers have been tested both in the controlled conditions and as a part of the automatic face recognition system.

Key words: face recognition, active shapes, normal bayes classifiers

1 Introduction

The face recognition is one of the most actively researched biometric identification tasks. Its popularity is caused mainly by two reasons. Firstly the face appearance is a prominent clue used by humans in the visual identification. Secondly the automatic face recognition is a non-invasive method, and as such can be used even without a person's knowledge or permission.

A human face is a complex 3D body which appearance depends not only on its structure but is also strongly influenced by many external factors (such as lighting, make-up, facial hair etc.). The main difficulty in building an efficient automatic face recognition system is to define the proper features which capture the person's characteristic description and can be extracted under the varying conditions of the image registration.

Our main goal was to test if the shape of the face and its components (eyes, nose, mouth and eyebrows) contains sufficient information for reliable human recognition. Therefore, two different face shape extraction methods have been selected and compared: the Active Shape Model and the Bayesian Tangent Shape Model. To investigate the role of classification scheme on the overall system efficiency three types of classifiers have been tested: the Normal Bayes Classiffier and the Minimum Distance Classifier based on both Euclidean and Mahalanobis metrics. Moreover, the influence of the reduced subspace dimension and of the length of the gradient profiles used on the recognition rates has been asessed.

In this paper we present results of our tests. We start with a review of the currently used deformable shape models and a short presentation of the implemented shape models and classification methods. Than we describe the data used and the environment of the conducted experiments. Finally, we give the results of applying classifiers to manually marked and automatically detected faces.

2 Deformable shape models

Kass et al. [1] were the first to introduce a deformable shape model. The Active Contours represented each curve as an ordered set of points. The state of contours was defined by two energy functionals: the internal energy responsible for the smoothness and continuity of the contour and the external energy attracting the contour to characteristic image features. The contour fitting task was obtained by the minimalization of the energy sum.

The main drawback of the Active Contours lies in the fact that they do not represent any knowledge of the shape of the extracted objects. The Active Shape Model (ASM) introduced by Cootes [2][3][4] was free of this fault. Statistical modelling of both the shape distribution and the appearance of the contour points resulted in a flexible and powerful shape extraction method.

Zhao et al.[5] 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 Witt[6] 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. [7] strived to improve the robustness of the Active Shapes to the face pose variability. 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. [8] 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 occupied with a few times longer processing time.

Zhou et al. reformulated the shape extraction task in the Bayesian framework. In their Bayesian Tangent Shape Model (BTSM) the shape vector was treated as a hidden state and was estimated using the Expectation-Maximization algorithm. They also used an adaptive weighting depending on the contour points stability which improved the model robustness to the noise and partial occlusions.

3 Shape extraction

Two shape models have been used in the experiment: the basic ASM [2][3][4] and the BTSM [9]. They both utilize two submodels: the Point Distribution Model (PDM) and the Local Structure Model (LSM).

The PDM is used to represent the statistical distribution of shapes in the training set. It is created by aligning set of exemplary shapes to a common coordinate frame (by using the Generalized Procrustes Analysis) and then applying the PCA to reduce the dimensionality of the model and to suppress 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.

Both deformable shape models differ in the fitting procedure (Figure 1) and the regularization function used to validate shapes. In case of the ASM the shape is extracted by iterative fitting the model to the image using the LSM and finding shape parameters with PDM. In order to obtain a plausible shape parameters they are regularized by truncation to a specified range. The whole procedure is repeated until convergence.

The extraction procedure of the BTSM is slightly more complicated. At each iteration the shape is calculated as a weighted mean of two shapes: one aquired by fitting points positions to the image, and the second one obtained by applying continuous regularization function to the previous estimate of shape. The weight

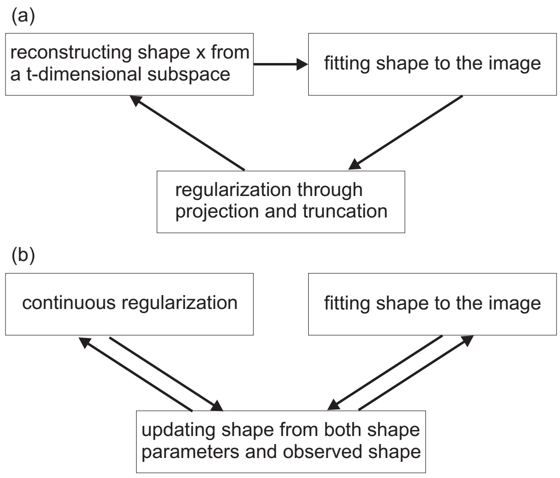

Fig. 1. Fitting procedure scheme: (a) ASM (b) BTSM.

is adaptively changed to reflect the relation between stability of contour points and the specificity of PDM.

The more detailed specification of both algorithms can be found in the cited literature. Figure 2 presents examplary shapes extracted with the ASM and the BTSM.

Fig. 2. Examples of the extracted face shapes: upper row the ASM, lower row the BTSM.

4 Classification

Three classification methods have been used in the experiments: the Normal Bayes Classifier (NBC), the Minimum Distance Classifier using the Mahalanobis metric (MDC-M) and Minimum Distance Classifiers using the Euclidean metric (MDC-E) [10][11]. All of them select the class with the highest *a posteriori* probability according to the *Bayes rule*:

$$P(\omega_m|b) = \frac{P(b|\omega_m)P(\omega_M)}{P(b)} \tag{1}$$

where ω_m is the m-th class and b is the shapes parameter vector. As all classes are considered to be equally probable ($P(\omega_M) = const$) and $P(b)$ is equal for all classes the Eqn. 1 simplifies to:

$$P(\omega_m|b) = const \cdot P(b|\omega_m) \tag{2}$$

The classification methods used differ in the degree of simplyfing assumptions used. The main supposition of NBC is that $P(b|\omega_m)$ can be modeled as a multivariate gaussian:

$$P(b|\omega_m) = \frac{1}{\sqrt{2\pi|C_m|}} \exp(-\frac{1}{2}(b-\bar{b}_m)^T C_m^{-1}(b-\bar{b}_m)) \tag{3}$$

where $\bar{b}_m$ is a mean vector and C_m is the covariance matrix of the class m.

With the assumption that all classes have the same covariance matrix ($C_m = C$) maximizing the *a posteriori* probability is achieved by minimizing the term (Eqn. 3):

$$D_M = (b-\bar{b}_m)^T C_m^{-1}(b-\bar{b}_m) \tag{4}$$

which in fact is the Mahalanobis distance. Classification in MDC-M is achieved by finding the class m with the lowest D_M.

If the covariance matrix C is an identity matrix Eqn. 4 simplifies even further to the Euclidan distance:

$$D_E = (b-\bar{b}_m)^T (b-\bar{b}_m) \tag{5}$$

Decision in MDC-E is based on the Euclidan distance of the new vector b under consideration to the mean vectors $\bar{b}_m$.

5 Experiments

5.1 Data

A new base of almost 10000 images of 100 people was gathered to provide high quality, high resolution color face images. The images were acquired in partially controlled illumination conditions over an uniform background and stored as 2048x1536 pixels JPEG files. Each picture was provided with a manually marked

Fig. 3. Examples of pictures treated as near-frontal face images.

face and eyes ROIs which were stored in the OpenCVStorage YAML files. The main purpose of creating such an extensive database was to ensure the statistical significance of obtained results.

Our research was conducted on the selected subset of 2193 images corresponding to the near-frontal views of faces (Figure 3). Each of the images used was manually marked with 194 point forming a contour model. The contour model is composed of the following subcontours (Figure 4):

- face outline - 41 points
- nose outline - 17 points
- eyes outlines - 20 points each
- eyebrows outlines - 20 points each
- lips outlines (inner and outer) - 28 points each

The proposed system has been developed using the Visual C++ 6.0 and Open Computer Vision Library[12].

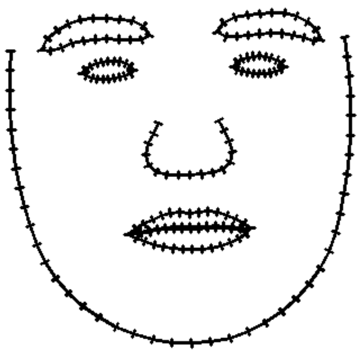

Fig. 4. Contour model - mean shape.

5.2 Experiments setup

To test the influence of the dimensionality of the PDFM (t) on the recognition rate 7 different PDMs with $t = \{10, 15, 20, 25, 30, 35, 40\}$ were created. The fraction of the training set variance modeled by those components is shown in Table 1. In order to investigate the influence of the length of the gradient profile (k) on the system performance 10 LSM were created with $k = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. The PDMs and LSMs were created using face contours of 50 people (half of the base).

Shapes were extracted using every possible combination of the LSM, the PDM and the shape extraction method used, which gives a total of 140 variants. The shape models have been initialized with both manually marked and automatically detected eyes positions. The eyes positions used for the automatic initialization have been detected by using the system proposed by Kasinski and Schmidt in [13][14]. For each variant three distinct sets have been created:

- training set - half of shapes extracted with the manual initialization
- test set I - the other half of shapes extracted with the manual initialization
- test set II - shapes extracted with the automatic initialization

The training sets have been used to create MDC-E, MDC-M and NBC classifiers, which performance has been assessed using the corresponding test sets.

Table 1. The percentage of training set variance modeled by t-dimensional subspace

t	10	15	20	25	30	35	40
%	95.8	97.5	98.5	99.0	99.3	99.5	99.6

5.3 Results

The main goal of this research was to determine the influence of face shape extraction method and its parameters as well as the classifier type on the classification efficiency. As all classes are almost equally numerous we used the True Positive Ratio (*TPR*) as a measure of classification efficiency (Fig. 5 and 6).

The MDC-E has proven ineffective in classifying face shapes regardless of the extraction algorithm used. It has achieved its peak efficiency of $TPR = 58.7\%$ in the test set I for the ASM with $k = 3$ and $t = 40$. This is the result of using PCA for dimensionality reduction. The scale of particular shape vectors is directly connected with the size of corresponding eigenvalues (due to regularization function) which practically renders the Euclidean metric useless.

Using the MDC-M has given significantly better results. Its highest $TPR = 96\%$ has been achieved in test set I for the ASM with $k = 6$ and $t = 40$. The highest TPR in the test set II was equal to 92.1% for the ASM with $k = 6$ and $t = 35$. The high reliability of those classifiers confirms that all classes have similar covariance matrices and the Mahalanobis metric can be succesfully used to discriminate between them. The MDC-M classifies shapes extracted with the ASM better than those extracted with the BTSM, although this difference is smaller in the test set II.

The best classification results have been obtained with the NBC. It has achieved the highest $TPR = 98\%$ in the test set I for the ASM with $k = 6$ and $t = 40$. In the test set II the NBC has had the peak $TPR = 94.3\%$ for the ASM with $k = 6$ and $t = 35$. This shows that the preassumption of modelling the classes probabilities as multivariate gaussians is admissible. The NBC also gives better results with the ASM than with the BTSM, although the difference is smaller then in case of the MDC-M.

The shapes extracted with the ASM have been classified more reliably. It is probably caused by using truncation as a regularization function. This ensures that all shape parameters fall into a certain range, thus the of covariance matrices of different classes are similar and their elements are smaller. The difference between the ASM and the BTSM is less visible while using the NBC. This may lead to the conclusion that classes of shapes extracted with the BTSM have different covariance matrices. As the BTSM has proved to be less reliable than the ASM w.r.t. the classification efficiency while being computationally more complex it is not recommended to use it in a face recognition system.

The length of the gradient profile used strongly influenced the recognition rates. Increasing it to the value of $k = 6$ improved the classifiers reliability but further increase resulted in efficiency drop (especially visible for the BTSM). This can be explained by gradual addition of the information on the contour background. Initially it helps to find the correct contour position, but after a certain point the abundance of information hinders the decision. This impacts the BTSM stronger, as it causes the instability of contour points and reduces the influence of fitting the shape to the image (due to the weighted average).

Increasing the PDM dimensionality initially significantly boosted the classifiers performance (up to $t = 20$). For t greater than 30 the gain is hardly noticeable. This can be explained by the fact, that 30 first principal components already model 99% of the training set variance. The remaining 1% is likely to contain mostly the noise introduced by manual annotation of images. The obtained results also show, that the k and t parameters influence the recognition rates independently and can be adjusted individually.

Initializing deformable shape models with automatically detected eyes positions led to slight performance drop. The main reason of this is the True Negative rate of the eyes detector, which fails to detect almost 3% of eyes.

6 Conclusions

Our computations proved that the MDC-M and NBC successfully classify face shapes extracted with the ASM. The complete processing time of a single image (including face and eyes detection, shape extraction and classification) on a PC with Intel Celeron 2800MHz and 512MB RAM was about 600ms. This clearly

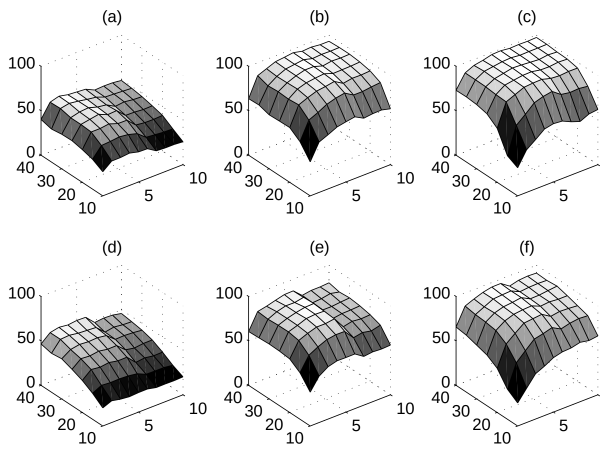

Fig. 5. The *TPR* in test set I as a function of k and t: (a) ASM + MDC-E (b) ASM + MDC-M (c) ASM + NBC (d) BTSM + MDC-E (e) BTSM + MDC-M (f) BTSM + NBC

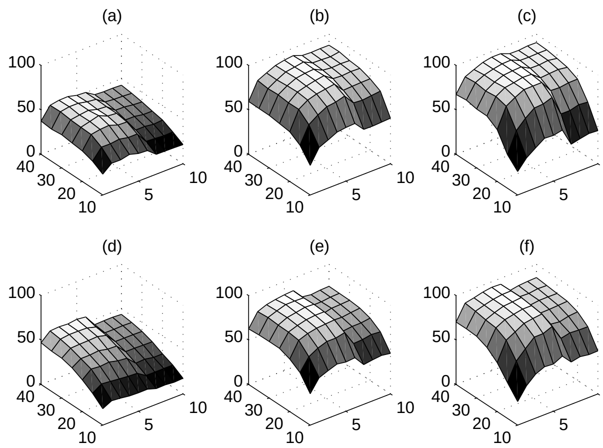

Fig. 6. The *TPR* in test set II as a function of k and t: (a) ASM + MDC-E (b) ASM + MDC-M (c) ASM + NBC (d) BTSM + MDC-E (e) BTSM + MDC-M (f) BTSM + NBC

shows that the proposed method can be used to build fast and reliable face recognition systems.

The choice of the classifier used should be based on the number of available training pictures per class. The MDC-M gives lower *TPR* but it needs fewer training images, as the covariance matrix is assumed to be equal for all classes. Training the NBC requires more images as each class covariance matrix has to be estimated independently, but this type of classifier yields much better results.

Another important property of the NBC is that it facilitates easy modifications of the base. Adding new persons does not require retraining whole classifier but only estimating the covariance and mean vector of the new class. This is especially useful in managing large, dynamically changing databases such as those used in population traffic analyses.

References

1. Kass, M., Witkin, A., Terzopoulos D.: Snakes: Active contour models. International Journal of Computer Vision, 1(4), 321-331, (1987)
2. Cootes, T., Taylor, C.: Statistical models of appearance for computer vision. Technical report, University of Manchester, Wolfson Image Analysis Unit, Imaging Science and Biomedical Engineering, (2004)
3. Cootes, T., Cooper, D., Taylor, C., Graham, J.: Active shape models - their training and application. Computer Vision and Image Understanding, 61(1), 38-59, (1995)
4. Cootes, T.: Image Processing and Analysis, Chapter Model-Based Methods in Analysis of Biomedical Images. Oxford University Press, (2000)
5. Zhao, M., Li, S.Z., Chen, C., Bu, J.:, Shape Evaluation for Weighted Active Shape Models. In Proc. of the Asian Conference on Computer Vision, 1074-1079, (2004)
6. Zuo, F., de With, P.H.N.: Fast facial feature extraction using a deformable shape model with haar-wavelet based local texture attributes. In Proc. of ICIP'04, 1425-1428, (2004)
7. Ge, X., Yang, J., Zheng, Z., Li, F.: Multi-view based face chin contour extraction. Engineering Applications of Artificial Intelligence, 19, 545-555, (2006)
8. Wan, K.-W., Lam, K.-M., Ng, K.-C.: An accurate active shape model for facial feature extraction., Pattern Recognition Letters, 26(15), 2409-2423, (2005)
9. Zhou, Y., Gu, L., Zhang, H.: Bayesian tangent shape model: Estimating shape and pose parameters via bayesian inference. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, (2003)
10. Theodoridis, S., Koutroumbas, K.: Pattern Recognition. Elsevier Academic Press, Second Edition, (2003)
11. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. Wiley, Second Edition, (2001)
12. Open Computer Vision Library. http://www.intel.com/technology/computing/opencv/.
13. Kasinski, A., Schmidt, A.: Computer Recognition Systems 2, volume 45 of Advances in Soft Computing, chapter The Architecture of the Face and Eyes Detection System Based on Cascade Classifiers. Springer Berlin / Heidelberg, (2007)
14. Schmidt, A., Kasinski, A.: Computer Recognition Systems 2, volume 45 of Advances in Soft Computing, chapter The Performance of the Haar Cascade Classifiers Applied to the Face and Eyes Detection. Springer Berlin / Heidelberg, (2007)