SOFT MARGIN ADABOOST FOR FACE POSE CLASSIFICATION

Ying Guo, Geoff Poulton, Jiaming Li, Mark Hedley, Rong-yu Qiao

CSIRO Telecommunications and Industrial Physics,
PO Box 76, Epping, NSW 1710, Australia

ABSTRACT

This paper presents a new machine learning method to solve the pose estimation problem. The method was based on Soft Margin AdaBoost (SMA) algorithm. The AdaBoost algorithm has been used with great success as a high-level learning procedure to obtain strong classifiers from weak classifiers, but it tends to overfit in the presence of highly noisy data. Recent studies show that a regularised AdaBoost algorithm, such as SMA, can achieve better results for noisy data. In this paper we propose two new techniques for classifying the image as frontal (face is within $\pm 25^o$) or profile, one based on the original Adaboost algorithm, the other based on SMA. It is shown that SMA based technique is more robust than original AdaBoost based technique, and yields better results. All the techniques were trained and tested on four databases. Experimental results show that the classification error of the SMA method is less than 2% for suitable parameters, regardless of the conditions associated with the face. In addition, the method performs extremely well even when some facial features become partially or wholly occluded.

1. INTRODUCTION

Research in face detection, face recognition and facial expression usually focuses on using frontal view images. However, approximately 75% of faces in normal photographs are non-frontal. Significant improvements in many computer vision algorithms dealing with human faces can be obtained if we can achieve an accurate estimation of the pose of the face, hence pose estimation is an important problem.

CSIRO has developed a real-time face capture and recognition system (SQIS - System for Quick Image Search) [9], which can automatically capture a face in a video stream and verify this against face images stored in a database to inform the operator if a match occurs. It is observed that the system performs better for the frontal images, so a method is required to separate the frontal images from pose images for the SQIS system.

Pose detection is hard because large changes in orientation significantly change the overall appearance of a face. Attempts have been made to use view-based appearance models with a set of view-labelled appearances (e.g. [1]). Gong et al. investigated multi-view pose distribution [3], and further extended SVMs to model the appearance of human faces which undergo nonlinear change across multiple views [8]. They implemented a multi-view face detection and recognition system under a support vector machine framework and achieved good performance on video sequences [5].

Recently, there has been great interest in ensemble methods for learning classifiers, and in particular in boosting algorithms [2], which is a general method for improving the accuracy of a basic learning algorithm. The best known boosting algorithm is **AdaBoost** algorithm. These algorithms have proven surprisingly effective at improving generalisation performance in a wide variety of domains, and for diverse base learners. For instance, Viola and Johns demonstrated that AdaBoost can achieve both good speed and performance in face detection [14]. However, research also showed that AdaBoost often places too much emphasis on misclassified examples,may just be noise. Hence, it can suffer from overfitting, particularly with highly noisy data set. The **Soft Margin AdaBoost** algorithm (SMA) [10] was introduced by using regularisation methods and generalisations of original AdaBoost algorithm to achieve a *soft margin*, which allows mislabeled samples to exist in the training data set.

This paper is directed toward a pose detection system using an SMA based classifier that can classify the frontal images (within $\pm 25^o$) from pose images (greater angles), under different scale, lighting or illumination conditions. The method uses Principal Component Analysis (PCA) to reduce the dimensionality of the training examples. The SMA algorithm is then used to generate a statistical model, which captures variation in the appearance of the facial angle.

All the experiments were evaluated using the CMU PIE database [13], Weizmann database [6], CSIRO Front database and CMU Profile face test set [12]. It is demonstrated that the SMA based technique is able to classify frontal and pose images much better than the original AdaBoost based technique.

The remainder of this paper is organised as follows. We proceed in Section 2 to explain our approach, including the theoretic introduction of original AdaBoost and Soft Margin AdaBoost. In Section 3, we will present some experiment results that we have achieved in pose detection. The conclusions are discussed in Section 4.

2. POSE DETECTION ALGORITHM

2.1. Feature extraction

PCA [7] is performed on the original facial images to reduce the dimensionality. We chose 3003 facial images from the CMU PIE database [13] under 13 different poses to generate an eigen pose space. A mean pose image and set of orthonormal eigen poses are produced. The first 8 eigen poses are shown in Figure 1. Each image can then be projected into a vector $\mathbf{x}$ in the subspace of eigenspace, where $\mathbf{x} \in \mathbb{R}^n$.

According to the pose angle θ_i of the training image, the corresponding label y_i of each vector $\mathbf{x}_i$ is defined as

$$y_i = \begin{cases} +1 & |\theta_i| \leq 25^o \\ -1 & \text{otherwise.} \end{cases}$$

The next task is to generate a decision function $f(\mathbf{x})$ based on a set of m training samples $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)$. The original AdaBoost and the Soft Margin AdaBoost algorithms are applied to solve this binary classification problem.

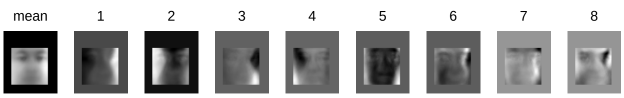

Fig. 1. *Mean face and first 8th eigenposes.*

2.2. AdaBoost and Soft Margin AdaBoost (SMA)

Boosting procedures take a given base learning algorithm, and repeatedly apply it to reweighted versions of the original training data, producing a collection of hypothesises $h_1, \cdots, h_n$ which are then combined in a final aggregate classifier via a weighted linear vote. AdaBoost computes the weight in a particular way, which can force the learner to focus on the "difficult" training examples and pay less attention to those that the most recent hypothesis got right [2].

In binary classification case, we define the *margin* of a sample pair $\mathbf{s}_i = (\mathbf{x}_i, y_i)$ as

$$\gamma(\mathbf{s}_i) := y_i f(\mathbf{x}_i), \text{ for } i = 1, \cdots, m,$$

where $f(\mathbf{x})$ is the final hypothesis. The margin at $\mathbf{s}$ is positive if the correct class label of the sample is predicted. It was shown theoretically and experimentally that AdaBoost is especially effective at increasing the margins of the training examples [11], but the generalisation performance of AdaBoost is not guaranteed. It was shown that AdaBoost does overfit for the noisy cases [4].

Ratsch et al. [10] shows that versions of AdaBoost modified to use regularisation are more robust for noisy data. A regularisation term is introduced to the cost function in AdaBoost. This term represents the "mistrust" to a noisy training sample, and allows it to be misclassified (negative margin) in the training process. The final hypothesis $f(\mathbf{x})$ obtained this way has worse training error but better generalisation performance compared to $f(\mathbf{x})$ of the original AdaBoost algorithm. This new algorithm is called *soft margin AdaBoost*, and the soft margin of a sample $\mathbf{s}_i$ is defined as

$$\tilde{\gamma}(\mathbf{s}_i) := \gamma(\mathbf{s}_i) + \lambda \eta_t(\mathbf{s}_i), \text{ for } i = 1, \cdots, m,$$

where λ is the *regularisation constant* and $\eta_t(\mathbf{s}_i)$ the *regularisation term*. A large value of $\eta(\mathbf{s}_i)$ for some patterns allow for some larger soft margin $\tilde{\gamma}(\mathbf{s}_i)$. Here λ balances the trade-off between goodness-of-fit and simplicity of the hypothesis. In the noisy case, SMA prefers hypotheses which do not rely on only a few samples with smaller values of $\eta(\mathbf{s}_i)$. So by using regularisation method, AdaBoost is not changed for easily classifiable samples, but only for the most difficult ones. The regularisation term $\eta_t(\mathbf{s}_i)$ can be defined as

$$\eta_t(\mathbf{s}_i) = \left(\sum_{r=1}^{t} c_r w_r(\mathbf{s}_i) \right)^2,$$

where w is the sample weight and t the training iteration index (cf. [10] for a detailed description).

3. EXPERIMENTS AND DISCUSSIONS

3.1. Data preparation

The databases used for experiments were collected from 4 databases: the CMU PIE database, the Weizmann database, the CSIRO Front database, and the CMU Profile face database. These contain a total of 41567 faces, of which 22221 are frontal images, and 19346 are pose images. The x-y positions of both eyes were hand-labeled for each image. We compared the performance of original AdaBoost and Soft Margin AdaBoost on a small training sample set ($m = 500$), in order to save the computation time. We also did experiments on a much larger training set ($m = 20783$) for the SMA based technique only, in order to show the best performance this new technique can achieve.

The effect of using different numbers of significant Principal Components (PCs), i.e. the signal components along the principal directions in the EPS, is also observed in the experiments. We will define the number of PCs as the PC-dimension. We tested the PC-dimensions between 10 and

dim	testErr of Ada	testErr of SMA	Mean difference
10	$13.72 \pm 1.34\%$	$11.63 \pm 0.71\%$	2.09%
20	$10.71 \pm 0.81\%$	$7.26 \pm 0.38\%$	3.45%
30	$8.19 \pm 0.63\%$	$6.06 \pm 0.47\%$	2.13%
40	$7.87 \pm 0.92\%$	$5.05 \pm 0.19\%$	2.82%
50	$8.20 \pm 0.69\%$	$5.14 \pm 0.21\%$	3.06%
60	$7.32 \pm 1.04\%$	$4.63 \pm 0.61\%$	2.69%
70	$7.98 \pm 0.85\%$	$4.49 \pm 0.33\%$	3.52%
80	$8.03 \pm 0.74\%$	$4.87 \pm 0.27\%$	3.16%

Table 1. *Performance of pose classifiers on different PC-dimensions (the number of PCA components). The last column shows that the test error of the SMA technique is always at least 2% better than the test error of the AdaBoost technique.*

80 in steps of 10. The training samples were randomly selected from the whole data set, and the rest to test the generalisation performance of the technique. All the experiments were repeated 5 times, and the results were averaged over the 5 repeats. In all experiments, radial basis function (RBF) networks with adaptive centres are used as base classifier.

3.2. Data preprocessing

The face images were normalised for rotation, translation and scale according to eye location. The subwindow of the face is then cropped using the normalised eyes distance. Because the new eye location is fixed without knowledge of the facial pose, the subwindow range is quite different based on different pose. For a large angle pose face, the cropped subwindow cannot include the whole face. Figure 2 shows such face region extraction for 9 poses of the PIE database with the corresponding angle.

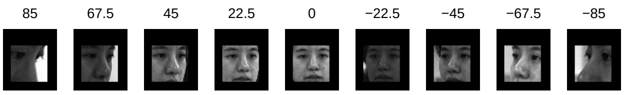

Fig. 2. *Normalised and crop face region of 9 different angles. The pose angle (degrees) is above the image. The subwindows only include eyes and part of nose for the $\pm 85^o$ pose images.*

3.3. Experiment results

3.3.1. Group A

We compared the original AdaBoost and the SMA based techniques in this group of experiments. In Table 1, the average generalisation performance (with standard deviation) over the range of PC-dimensions is given after 800 iterations. Our experiments show that the AdaBoost results are in all cases at least $2 \sim 4\%$ worse than the SMA results. Figure 3 shows one comparison with PC-dimension $n = 30$. The training error of the AdaBoost based technique converges to zero after only five iterations, but the test error clearly shows the overfitting. For the SMA based technique, because of the regularisation term, the training error is not zero in most of the iterations, but the test error keeps decreasing.

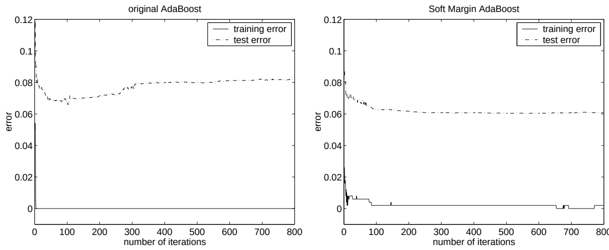

Fig. 3. *Training and test error graphs of original AdaBoost (left) and Soft Margin AdaBoost (right) when training set size $m = 500$, and PC-dimension of the feature vector $n = 30$. The test error of AdaBoost overfits to 8.19% while the test error of SMA converges to 6.06%.*

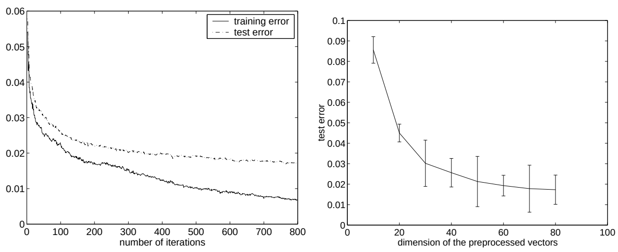

Fig. 4. *Experiment results of SMA. Left: When training set size $m = 20783$, and PC-dimension of the feature vector $n = 80$, the training and test errors of SMA keep decreasing. Right: As the PC-dimension of the feature vector increases, the test error decreases, and converges to a limit at higher dimensions.*

Fig. 5. *Test images correctly classified as profile. Noticing these images include different facial appearance, expression, significant shadows or sun-glasses.*

3.3.2. Group B

In this group, we trained the SMA on more training samples ($m = 20783$), and achieved competitive results on pose estimation problem. Figure 4 (left) shows the training and test errors of the SMA based technique, while the PC-dimension of the feature vector is eighty. As the number of iteration increases, both the training and test errors converge. Especially, the test error after 800 iterations is as low as 1.73%, which proves the good generalisation performance of the SMA based technique.

It is important to point out that the correct detection rate is related to the PC-dimension of the feature vector. Figure 4 (right) shows that the performance in pose detection is better for higher dimensional PCA representation. The test error is only $1.73\% \pm 0.71$ when $n = 80$. On the other hand, low dimensional PCA representation can already provide satisfactory performance, for instance, the correct detection rate is $3.02\% \pm 0.56\%$ when $n = 30$. Figure 5 shows some examples of the correctly classified pose images, which include different scale, lighting or illumination conditions.

4. CONCLUSION

The main strength of the present method is the ability to estimate the pose of the face efficiently by using the soft margin AdaBoost algorithm. The experimental results show that the SMA based technique allows higher training errors to avoid the overfitting problem, and achieves much better generalization performance than the AdaBoost based technique. The experimental results also show that the SMA based technique is very effective for the pose estimation problem. The test error on more than $20,000$ test images can be as low as 1.73%, where the images cover different facial features such as beards, glasses, and a great deal of variability including shape, color, lighting, illumination.

In addition, because the only prior knowledge of the system is the eye locations, the performance is extremely good, even when some facial features such as the nose or the mouth become partially or wholly occluded. For our current interest in improving the performance of our face recognition system (SQIS), as the eye location is already automatically determined, this new pose detection method can be directly incorporated into the SQIS system to improve its peformance.

References

[1] S. Baker, S. Nayar, and H. Murase. Parametric feature detection. *International Journal of Computer Vision*, 27(1):27–50, March 1998.

[2] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Computational Learning Theory: Eurocolt '95*, pages 23–37. Springer-Verlag, 1995.

[3] S. Gong, S. McKenna, and J. Collins. An investigation into face pose distributions. *IEEE Int. Conf. On Face and Gesture Recognition*, pages 265–270, 1996.

[4] A. Grove and D. Schuurmans. Boosting in the limit: Maximizing the margin of learned ensembles. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, pages 692–699, 1998.

[5] Y. Li, S. Gong, and H. Liddell. Support vector regression and classification based multi-view face detection and recognition. *IEEE FGRTs*, pages 300–305, 2000.

[6] Y. Moses, S. Ullman, and S. Edelman. Generalization to novel images in upright and inverted faces. *Perception*, 25:443–462, 1996.

[7] H. Murase and S. K. Nayar. Visual learning and recognition of 3D objects from appearance. *International Journal of Computer Vision*, 14(1):5–24, 1995.

[8] J. Ng and S. Gong. Performing multi-view face detection and pose estimation using a composite support vector machine across the view sphere. *In Proc. IEEE International Workshop on FGRTs*, pages 14–21, 1999.

[9] R. Qiao, J. Lobb, J. Li, and G. Poulton. Trials of the csiro face recognition system in a video surveillance environment. *Proceedings of the Sixth Digital Image Computing Techniques and Application*, pages 246–251, Jan 2002.

[10] G. Rätsch, T. Onoda, and K.-R. Müller. Soft margins for AdaBoost. *Machine Learning*, 42(3):287–320, March 2001.

[11] R. Schapire, Y. Freund, P. L. Bartlett, and W. Sun Lee. Boosting the margin: A new explanation for the effectiveness of voting methods. *Annals of Statistics*, 26(5):1651–1686, 1998.

[12] H. Schneiderman and T. Kanade. A statistical method for 3D object detection applied to faces and cars. *In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1:746–751, 2000.

[13] T. Sim, S. Baker, and M. Bsat. The CMU Pose, Illumination, and Expression (PIE) database. *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, May 2002.

[14] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *IEEE CVPR*, 2001.