Gabor Feature Selection for Face Recognition using Improved AdaBoost Learning

Linlin Shen[1], Li Bai[1], Daniel Bardsley[1], Yangsheng Wang[2]

[1] School of Computer Science & IT, University of Nottingham, UK
{lls, bai, djb}@cs.nott.ac.uk
[2] Institute of Automaton, Chinese Academy of Sciences, P.R. China
wys@nlpr.ia.ac.cn

Abstract. Though AdaBoost has been widely used for feature selection and classifier learning, many of the selected features, or weak classifiers, are redundant. By incorporating mutual information into AdaBoost, we propose an improved boosting algorithm in this paper. The proposed method fully examines the redundancy between candidate classifiers and selected classifiers. The classifiers thus selected are both accurate and non-redundant. Experimental results show that the strong classifier learned using the proposed algorithm achieves a lower training error rate than AdaBoost. The proposed algorithm has also been applied to select discriminative Gabor features for face recognition. Even with the simple correlation distance measure and 1-NN classifier, the selected Gabor features achieve quite high recognition accuracy on the FERET database, where both expression and illumination variance exists. When only 140 features are used, the selected features achieve as high as 95.5% accuracy, which is about 2.5% higher than that of features selected by AdaBoost.

1   Introduction

Introduced by Freud and Schapire [1], AdaBoost has been successfully applied to object detection [2;3] and face recognition [4]. The essence of AdaBoost is to learn a number of very simple weak classifiers, which are then linearly combined into a single strong classifier. Whilst using only weak classifiers, recognition performance is just slightly better than random guessing. AdaBoost learning minimizes the upper bound on both training and generalization errors [5]. AdaBoost has been applied in applications to select Haar-like features [3] for face detection and recognition [4] as well as proving suitable for Gabor feature selection [6;7] for classification. Since minimizing the classification error rate is the ultimate objective of AdaBoost learning, the weak classifier with smallest weighted error is selected at each iteration. As a result, the learned classifiers are "individually" best and hence the combined strong classifier may not necessarily be optimum choice [8]. For feature selection, classifiers using similar features are more likely to be selected and redundancy will exist among some selected features. Stanz. Li etc. [8] proposed a floating search based algorithm, named FloatBoost, to eliminate these non-effective weak classifiers. A backtracking mechanism is applied to identify the unfavorable weak classifiers in terms of the classification error rate. The learned strong classifier thus consists of fewer weak classifiers and shows improved classification performance. During the learning process, each of the previously selected weak classifiers is individually removed from the combined strong classifier and the results analyzed for improvements in the classification error rate. As a result of this, FloatBoosts computational requirements are approximately five times greater than that of AdaBoost. When the number of features is large (160,000+), as is normal in implementation scenarios utilising Gabor features, the FloatBoost training process could be unmanageable. A boosting algorithm, which is both effective in eliminating the non-effective classifiers and is computationally efficient, is required.

We propose in this paper a novel approach to address this issue. The proposed boosting algorithm uses the idea of mutual information for redundancy elimination. During the learning process, the mutual information between the candidate weak classifier and the selected weak classifiers is examined. As a result, the non-effective classifiers carrying information already captured by the selected feature/classifiers will be excluded. Since the mutual information is checked for only those candidate classifiers with small errors, the additional computation cost above that already required by AdaBoost is very low. In addition, mutual information is statistically calculated, which enables our technique to be classifier and decision method independent. It should be noted that several important distinctions exist between the method presented here and the InfoBoost algorithm [9]. While InfoBoost used the mutual information between the weak classifier and the class label for updating weights and weak classifier selection, our work still uses the classification error for learning, but introduces the idea of using mutual information

between weak classifiers as an additional criteria. A threshold is used such that those classifiers with mutual information outside the given threshold, even with low classification error, are not selected. The experimental results show that the proposed boosting algorithm achieves a lower training error rate with fewer classifiers than other methods. Better performance has also been observed when the selected Gabor features are applied for face recognition.

2 AdaBoost learning

Since the introduction of the original discrete AdaBoost algorithm, more general versions have been proposed. Where RealBoost was proposed to boost weak classifiers with real value output [10], AdaBoost.M1 and AdaBoost.MH [5] address the multi class problem. Here we focus on two class problems only, where weak classifiers take a discrete Boolean value. However, our method could be easily incorporated into the RealBoost or AdaBoost.MH algorithms to solve more general problems.

For the two class problem, a set of N labeled training samples is given as $(x_i, y_i), i = 1,2,..,N$, where $y_i \in \{0,1\}$ is the class label associated with the sample $x_i \in R^n$. A large number of weak classifiers $h(x)$ could be generated to form the classifier learning pool, where $h(x) \in \{0,1\}$. The weak classifier could be very simple, for example, a threshold function on the *kth* coordinate of x in the *n*-dimensional space. The algorithm focuses on difficult training patterns, increasing their representation in successive training sets. Over a number of T rounds, T weak classifiers are selected to form the final strong classifier. In each of the iterations, the space of all possible weak classifiers is searched exhaustively to find the one with the lowest weighted classification error. This error is then used to update the weights such that the wrongly classified samples get their weights increased. The resulting strong classifier is a weighted linear combination of all T selected weak classifiers. See [1] for the listing of the AdaBoost algorithm.

3 Improved AdaBoost learning

3.1 Entropy and Mutual Information (MI)
As a basic concept in information theory, entropy $H(X)$ is used to measure the uncertainty of a random variable (r.v.) X . If X is a discrete r.v., $H(X)$ can be defined as below:

$$H(X) = -\sum_x p(X = x) \lg(p(X = x)) \qquad (1)$$

Mutual information $I(Y;X)$ is a measure of general interdependence between two random variables X and Y :

$$I(Y;X) = H(X) + H(Y) - H(X,Y) \qquad (2)$$

Using Bayes rule on conditional probabilities, Equation 2 can be rewritten as:

$$I(Y;X) = H(X) - H(X | Y) = H(Y) - H(Y | X) \qquad (3)$$

Since $H(Y)$ measures the priori uncertainty of Y and $H(Y | X)$ measures the conditional posteriori uncertainty of Y after X is observed, the mutual information $I(Y;X)$ measures how much the uncertainty of Y is reduced if X has been observed. It can be easily shown that if X and Y are independent, $H(X,Y) = H(X) + H(Y)$ and consequently their mutual information is zero.

The estimation of MI requires the value of marginal distribution $p(X)$, $p(Y)$ and the joint probability distribution $p(Y,X)$. For a r.v. with discrete values, the probability could be estimated by simply counting the number of possible cases and dividing that number with the total number of training samples. For a continuous r.v., its pdf could either be discretized by histogram estimation, or be approximated by Gaussian distribution.

3.2 The Proposed Algorithm

The proposed boosting algorithm incorporates the idea of MI to eliminate those non-effective weak classifiers. Each weak classifier $h(x) \in \{0,1\}$ is now considered as a r.v.. Before a new weak classifier is added, the MI between the new classifier and each of the selected ones is examined to make sure that the information carried by the new classifier has not been captured before. Given stage T where *T-1* weak

classifiers $\{h_{v(1)}, h_{v(2)}, \cdots h_{v(T-1)}\}$ have been selected, the function to measure the MI $R(h_j)$ between a candidate classifier h_j and the selected classifiers can be defined as follows:

$$R(h_j) = \max_t I(h_j, h_{v(t)}), t = 1, 2, \cdots T - 1 \quad (4)$$

1) Input: N training samples $(x_i, y_i), i = 1, 2, .., N$ with m positive $(y_i = 1)$ and l negative $(y_i = 0)$ samples

2) Initialization: weights $w_{1,i} = \begin{cases} 1/2m, & \text{if } i \text{ is a positive sample} \\ 1/2l, & \text{if } i \text{ is a negative sample} \end{cases}$

3) For $t=1, \ldots, T$
    a) Normalize all weights
    b) Classifier selection and redundancy checking:
       For each candidate weak classifier h_j, calculate classification error

$$\varepsilon_j = \sum_i w_{t,i} |h_j(x_i) - y_i|$$

       For (;;)
          Choose $h_{t'}$ with lowest error $\varepsilon_{t'}$ from the candidate classifiers
          Calculate the MI $R(h_{t'})$ according to Eq. (4)
          If $R(h_{t'}) < TMI$
             The classifier found, $h_t = h_{t'}$, $\varepsilon_t = \varepsilon_{t'}$
             go to c)
          Else
             Remove $h_{t'}$ from the candidate list
       End loop

    c) Update weights: $w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$ with $e_i = \begin{cases} 1 : x_i & \text{correctly classified} \\ 0 : & \text{otherwise} \end{cases}$

    and $\beta_t = \varepsilon_t / (1 - \varepsilon_t)$

4) Final strong classifier: $H(x) = \begin{cases} 1 & if \ \sum_{t=1}^T \alpha_t h_t(x) > 1/2 \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases}$ with $\alpha_t = \log(1/\beta_t)$

Figure 1 Details of proposed boosting algorithm

The value of $R(h_j)$ can be directly used to decide whether the new classifier is redundant or not. The value is compared with a pre-defined Threshold Mutual Information (TMI) value, if it is bigger than the TMI, we can deduce that the information carried by the classifier has already been captured. Besides MI, the classification error of the weak classifier is also taken in to consideration, i.e., only those classifiers with small errors are selected. The classifiers thus selected will be both accurate and informative. When all the non-redundant classifiers are combined to form a strong classifier, better performance will be achieved. See Figure 1 for details of the algorithm.

4 Application for Gabor Feature Selection

Motivated by the functional similarity of Gabor filters with the cells in the visual cortex of human and mammalian visual systems, Daugman [11] presented evidence that such visual neurons could optimize the general uncertainty relations for resolution in space, spatial frequency and orientation. From an information theoretic viewpoint, Okajima [12] derived Gabor functions as solutions for a certain mutual-information maximization problem. The work shows that the Gabor-type receptive field can extract the maximum possible information from local image regions. Researchers have also shown that Gabor features, when appropriately designed, are invariant against translation, rotation and scale [13]. Successful applications of Gabor filters in face recognition can be found in the FERET evaluation [14], where the Elastic Bunch Graph Matching method [15] gave the best performance. More recently, results from the face verification competition 2004 [16] also demonstrated the success of Gabor filter based approaches since both of the top two approaches apply Gabor filters for feature extraction. For face recognition applications, the number of Gabor filters used to convolve face images varies between applications, but usually 40 filters (5 scales and 8 orientations) are used [15;17-19]. However, due to the large number of

convolution operations, the computation cost is quite high. Even when a parallel computer system has been used, it was reported in [17] that the convolution of a 128×128 pixel image with 40 Gabor filters took about 7 seconds. For global methods, the dimension of the feature vectors extracted are also incredibly large, e.g., 163,840 for images with a size of 64×64. Similar to the work of Viola and Jones [2], where AdaBoost was used to select Haar-like features for face detection, the task here is to select the most discriminative Gabor features for face recognition.

4.1 Gabor Features and the Personal Difference Space

Give a bank of 40 Gabor filters $\{\varphi_{u,v}(x,y), \ u = 0,...,4, v = 0,...7\}$, image features at different locations, frequency and orientation can be extracted by convolving the image $I(x,y)$ with the filters:

$$O_{u,v}(x,y) = |I(x,y) * \varphi_{u,v}(x,y)| \qquad (5)$$

The resultant Gabor feature set thus consists of the convolution results of an input image $I(x,y)$ with all of the 40 Gabor filters:

$$S = \{O_{u,v}(x,y) : u \in \{0,...,4\}, v \in \{0,...,7\}\} \quad (6)$$

Figure 2 shows the magnitudes of Gabor representations of a face image with 5 scales and 8 orientations. A series of row vectors $\mathbf{O}_{u,v}$ could be converted out of $O_{u,v}(x,y)$ by concatenating its rows or columns, which are then concatenated together to generate a discriminative Gabor feature vector:

$$G(I) = \mathbf{O} = (\mathbf{O}_{0,0} \ \mathbf{O}_{0,1} \cdots \mathbf{O}_{4,7}) \qquad (7)$$

Take an image with size 64×64 as an example, the convolution result will give 64×64×5×8=163,840 features. Since Gabor filter parameters are chosen empirically, it is our belief that a lot of redundant information is included, and therefore a feature selection mechanism should be used to choose the most useful features for classification.
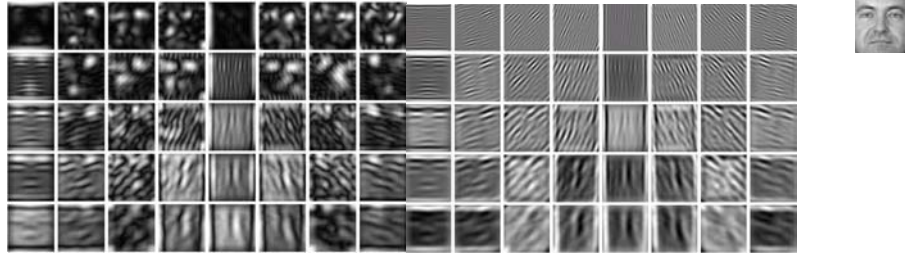

Figure 2 Convolution result - (magnitude and real part) of an image with 40 Gabor filters

To apply the proposed boosting algorithm, the difference space proposed in [20] is used here to convert the face recognition problem into a two class problem. Two classes, dissimilarities between faces of the same person (intra-personal space) and dissimilarities between faces of the different people (extra-personal space) are defined. The two Gabor feature difference sets: *CI* (intra-personal difference) and *CE* (extra-personal difference) can be defined as:

$$CI = \left\{\|G(I_p) - G(I_q)\|, p = q\right\}$$
$$CE = \left\{\|G(I_p) - G(I_q)\|, p \neq q\right\} \qquad (8)$$

where I_p and I_q are the facial images from people p and q respectively, and $G(\cdot)$ is the Gabor feature extraction operation as defined in (7). A set of M training samples in the difference space can now be described as $\{(g_1, y_1), \cdots (g_i, y_i), \cdots, g_M, y_M)\}$, $g_i = [x_1 x_2 \cdots x_n \cdots x_N]$, $y_i \in \{0,1\}$, where y_i is the class label (intra-personal or extra-personal) associated with sample g_i, N is the dimension of extracted Gabor features and $x_n = (g)_n = \left(\|G(I_p) - G(I_q)\|\right)_n = \left(\|\mathbf{O}_p - \mathbf{O}_q\|\right)_n$.

5 Experimental Results

5.1   Datasets
We analyze the performance of our algorithm using a subset of the FERET database, which is a standard test-bed for face recognition technologies [14]. 600 frontal face images corresponding to 200 subjects are

extracted from the database for the experiments - each subject has three images of size 256×384 with 256 gray levels. The images were captured at different photo sessions so that they display variations in illumination and facial expression. Two images of each subject are randomly chosen for training, and the remaining one is used for testing. The following procedures were applied to normalize the face images prior to the experiments:

- The centres of the eyes of each image are manually marked
- Each image is rotated and scaled to align the centre of the eyes
- Each face image is cropped to the size of 64×64 to extract the facial region
- Each cropped face image is normalized to zero mean and unit variance

Figure 3 shows the sample images from the database. The first two rows are the example training images while the third row shows the example test images. We are trying to select the most discriminative Gabor features using the 400 (2 images per subject) training images, which are then applied to recognize the 200 test images. As a result, 200 intra-personal and 1600 extra-personal Gabor feature difference samples are randomly generated. Once a small set of discriminative Gabor features are learned by applying AdaBoost or our proposed algorithm on the training samples, they are used for face recognition.

Figure 3 Sample images used in experiments

5.2   Selected Gabor Features

We first applied AdaBoost on the training samples to select 200 Gabor features for intra-personal and extra-personal difference classification. To show the existence of redundancy among AdaBoost selected features (weak classifiers), the MI $R(h_j)$ for each selected feature is shown in Figure 4a. It can be observed from the figure that some of the features are highly redundant, e.g. the MI of features with numbers 149, 177 and 180 is greater than 0.99. The redundancy among selected features increases with the number of features, it is this undesired redundancy that we aim to eliminate or reduce. We have also shown the MI data for features selected with our proposed algorithm in Figure 4b (with TMI=0.1). Due to the introduction of TMI, all the selected features now show MI values of less than 0.1 and thus we can conclude that the features are informative and non-redundant.
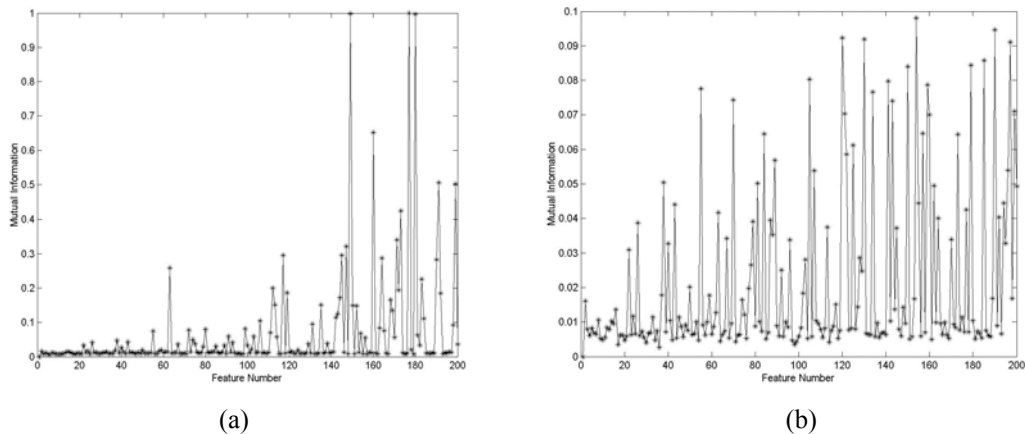

(a)   (b)

Figure 4 MI of features selected by AdaBoost a); Proposed algorithm b)

| (a) | (b) | (c) | (d) | (e) | (f) | (g) |

Figure 5 First five selected Gabor features (a)-(e); and the 200 feature points selected by our algorithm (f); and AdaBoost (g)

Figure 5 (a)-(e) show the first five Gabor features selected by the proposed algorithm, with locations of the first 200 Gabor features selected by our algorithm and AdaBoost shown in (f) and (g) respectively. The features are overlapped on a typical face image in the database. It is interesting to see that most of the selected Gabor features are located around the prominent facial features such as eyebrows, eyes, nose and the chin. This indicates that these regions are more robust against the variations is expression and illumination. This result is agrees with the fact that the eye and eyebrow regions remain relatively stable when a persons expression changes. While almost all of the AdaBoost selected Gabor features are crowded in the eyebrow and eye regions, the features selected by the proposed algorithm are more widely distributed with some of the features located around the nose also being included.

5.3 Recognition Performance Comparison

Once the most discriminant Gabor features are selected, we are now able to apply them to the face recognition problem. In this experiment, 200 Gabor features selected by AdaBoost and by our proposed algorithm (TMI=0.1) are directly used for a similarity comparison, without any further processing. The normalized correlation distance measure and the nearest neighbor classifier are used. Figure 6 shows the recognition performance of Gabor features selected by both AdaBoost and our algorithm on the 200 test images. When 140 features are used, the highest accuracy achieved by AdaBoost and the proposed algorithm, are 93% and 95.5% respectively. Since the MI values for all of the first 60 features are quite small, the proposed algorithm starts by picking up much the same features as AdaBoost. However, once the number of features increases, AdaBoost starts to pick highly redundant features. The improved recognition rate accuracy over AdaBoost caused by the use of features selected using the proposed algorithm proves the usefulness of our techniques in eliminating redundancy.
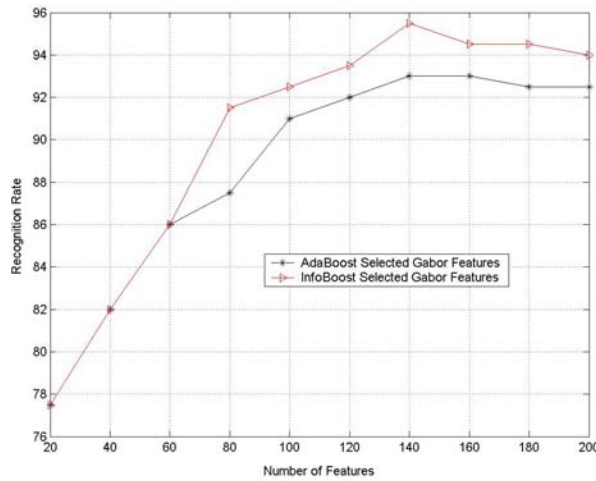

Figure 6   Recognition performance

5.4 Algorithm Complexity and Determination of TMI

Due to the introduction of mutual information, the proposed boosting algorithm required longer training time than that required by AdaBoost. However, the only computation cost added to AdaBoost is the loop to calculate MI values for redundancy checking, see Figure 1 for details. Table 1 shows the Average Number of Loops (ANL) required in each iteration and the corresponding TMI. The table shows that the computation burden added by the introduction of MI is actually very low (ANL is normally less than 10). As a result, the training time required by the proposed algorithm in our experiments is only about 0.1 times greater than that of AdaBoost.

TMI	0.08	0.09	0.10	0.11	0.12
ANL	8.42	8.07	7.25	5.43	3.25

Table 1 ANL for different TMI

As seen from the table, the higher the value of TMI, the less ANL required, i.e. the faster training speed. Actually AdaBoost can be seen as a special case of our algorithm when the value of TMI is set as 1. In this case, the features, or weak classifiers selected by the proposed algorithm will be exactly the same as those chosen by AdaBoost. The value of TMI needs to be selected appropriately to make sure that selected features are both non-redundant and useful for classification. A cross-validation set could be used to determine the TMI for common classification problems. As shown in Figure 4, since the redundancy increases with the number of selected features, an adaptive TMI, which increases with the number of features, might be more suitable.

6 Conclusions

A new and novel boosting algorithm has been proposed in this paper. By examining the mutual information between the candidate classifier and the selected weak classifiers, redundant classifiers can be excluded. The proposed algorithm has also been applied to select discriminant Gabor features for face recognition. The results show the advantage of features selected using our techniques over those learned by AdaBoost, specifically, Gabor features selected using our method are both non-redundant and achieve higher recognition accuracy on the subset of FERET database used for testing. Though the proposed algorithm addresses weak classifiers with discrete values only, it can be easily extended to classifiers with continuous values, where histogram estimation could be used to estimate the probability distribution. Whilst the selected features/classifiers are non-redundant and thus achieve higher accuracy for classification, the computation burden added to AdaBoost is very low. In our experiments, the training time of the proposed algorithm is only about 0.1 times longer than that of AdaBoost.

It should also be noted that though the face images are normalized manually in this paper, the work in [21] proved the robustness of non-redundant features against the mis-alignment of automatic face detection system. We aim to prove the tolerance of our system against the localized in the future. Currently, we are working on designing an adaptive TMI determination algorithm to improve the proposed boosting algorithm further.

Reference

[1] Y. Freund and R. Schapire, "A decision-theoristic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119-139, 2000.

[2] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition* Kauai, Hawaii: 2001.

[3] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," in *Proc. IEEE Conference on Image Processing* 2002, pp. 900-903.

[4] J. Michael and P. Viola, "Face recognition using boosted local features," in *Proc. of International Conference on Computer Vision (ICCV)* 2003.

[5] Y. Freund and R. Schapire, "A short introduction to boosting," *Journal of Japanese Society for Artifical Intelligence*, vol. 14, no. 5, pp. 771-780, 1999.

[6] L. Shen and L. Bai, "AdaBoost Gabor Feature Selection for Classification," in *Proc. of Image and Vision Computing NewZealand* 2004, pp. 77-83.

[7] P. Yang, S. G. Shan, W. Gao, S. Z. Li, and D. Zhang, "Face recognition using ada-boosted gabor features," *Sixth Ieee International Conference on Automatic Face and Gesture Recognition, Proceedings*, pp. 356-361, 2004.

[8] S. Z. Li and Z. Q. Zhang, "FloatBoost learning and statistical face detection," *Ieee Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1112-1123, 2004.

[9] J. Aslam, "Improving Algorithms for Boosting," in *Proceedings of the Thirteenth Annual Conference on Computational Learning Theory* 2000, pp. 200-207.

[10] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Machine Learning*, vol. 37, no. 3, pp. 297-336, 1999.

[11] J. G. Daugman, "Uncertainty Relation for Resolution in Space, Spatial- Frequency, and Orientation Optimized by Two-Dimensional Visual Cortical Filters," *Journal of the Optical Society of America A-Optics Image Science and Vision*, vol. 2, no. 7, pp. 1160-1169, 1985.

[12] K. Okajima, "Two-dimensional Gabor-type receptive field as derived by mutual information maximization," *Neural Networks*, vol. 11, no. 3, pp. 441-447, 1998.

[13] V. Kyrki, J. K. Kamarainen, and H. Kalviainen, "Simple Gabor feature space for invariant object recognition," *Pattern Recognition Letters*, vol. 25, no. 3, pp. 311-318, 2004.

[14] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *Ieee Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090-1104, 2000.

[15] L. Wiskott, J. M. Fellous, N. Kruger, and C. vonderMalsburg, "Face recognition by elastic bunch graph matching," *Ieee Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775-779, 1997.

[16] K. Messer, J. Kittler, M. Sadeghi, M. Hamouz, A. Kostin, F. Cardinaux, S. Marcel, S. Bengio, C. Sanderson, N. Poh, Y. Rondriguez, J. Czyz, L. Vandendorpe, C. McCool, S. Lowther, S. Sridharan, V. Chandran, R. P. Palacios, E. Vidal, L. Bai, L. Shen, Y. Wang, Y. H. Chiang, H. C. Liu, Y. P. Huang, A. Heinrichs, M. Miiller, A. Tewes, C. v. d. Malsburg, R. Wiirtz, Z. G. Wang, F. Xue, Y. Ma, Q. Yang, C. Fang, X. Q. Ding, S. Lucey, R. Goss, and H. Schneiderman, "Face authentication test on the BANCA database," in *Proc. of International Conference on Pattern Recognition* Cambridge, UK: 2004.

[17] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. Vandermalsburg, R. P. Wurtz, and W. Konen, "Distortion invariant object recognition in the Dynamic Link Architecture," *Ieee Transactions on Computers*, vol. 42, no. 3, pp. 300-311, 1993.

[18] C. J. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *Ieee Transactions on Image Processing*, vol. 11, no. 4, pp. 467-476, 2002.

[19] L. Shen and L. Bai, "Face recognition based on Gabor features using kernel methods," in *Proc. of the 6th IEEE Conference on Face and Gesture Recognition* Korea: 2004, pp. 170-175.

[20] P. J. Phillips, "Support vector machines applied to face recognition," in *Proceedings of the 1998 conference on Advances in neural information processing systems II* MIT press, 1999, pp. 803-809.

[21] G. Littlewort, M. S. Bartlett, I. R. Fasel, J. Chenu, and J. R. Movellan, "Analysis of machine learning methods for real-time recognition of expressions from video," 2003.