

Bayesian Face Recognition Based on Markov Random Field Modeling

Rui Wang, Zhen Lei, Meng Ao, Stan Z. Li*

Center for Biometrics Security Research & National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
{rwang, zlei, mao, szli}@cbsr.ia.ac.cn
<http://www.cbsr.ia.ac.cn>

Abstract. In this paper, a Bayesian method for face recognition is proposed based on Markov Random Fields (MRF) modeling. Constraints on image features as well as contextual relationships between them are explored and encoded into a cost function derived based on a statistical model of MRF. Gabor wavelet coefficients are used as the base features, and relationships between Gabor features at different pixel locations are used to provide higher order contextual constraints. The posterior probability of matching configuration is derived based on MRF modeling. Local search and discriminate analysis are used to evaluate local matches, and a contextual constraint is applied to evaluate mutual matches between local matches. The proposed MRF method provides a new perspective for modeling the face recognition problem. Experiments demonstrate promising results.

Key words: Face Recognition, Markov random field, learning.

1 Introduction

Relations between image features provides important constraints for image analysis and face recognition. A face is seen as an array of pixel values. If we make an image darker or brighter or perform some other nonlinear transformation such as changing the lighting condition, the image is still a pattern of that particular face. However, if the pixels are shuffled, the image no longer contains a face pattern, even if the individual pixel values remain the same. This illustrates that contextual constraints between image features play a more critical role than the pixel values themselves.

Markov random field (MRF) theory provides a convenient and consistent way for modeling context-dependent constraints such as between image pixels and features derived from pixels. This is achieved through characterizing mutual influences among pixels and features using conditional MRF distributions. The practical use of MRF models is largely ascribed to a theorem that states the equivalence between MRFs and Gibbs distributions, as established by [1] and further developed by [2]. MRFs have been widely employed to solve image processing and computer vision problems [3–7].

Recently, researchers have started investigations of face recognition problems from the MRF viewpoint. In [8], Qian and Huang presents a MRF-based method for face

* Stan Z. Li is the corresponding author.

detection, where scale changes are dealt with by using image pyramid and orientation changes are handled by rotating the image pyramid. In [9], Dass and Jain use MRF to model the spatial distribution of pixel values for the face pattern. The MRF modeling is further enhanced by finding the optimal permutation of pixels that maximizes the discriminative power between face and nonface patterns [10]. Huang *et al.* [11] propose a hybrid method for face recognition using MRF and belief propagation. A face image is divided into non-overlapping patches, and the patches in the probe image are matched to those in the gallery images, giving a person ID for each patch. The final decision is made by voting IDs over patches. In [12], straight lines are extracted from a face image. By attaching properties and binary relations to the straight lines, a face is then represented as a graph. A partial matching is used to match two graphs and select the best match.

In this paper, we propose a novel approach for modeling face recognition problems using MRF and Bayesian statistics. Assuming that the face in an image has been properly normalized into a “canonical” frame, the following two issues are addressed: (1) effective representation of a face pattern using image properties and their contextual relations, and (2) matching with the image properties and their contextual relations.

The paper contains the following novel ideas. For issue (1), we explore mutual constraints between the images features in a neighborhood system, where the features used are the Gabor wavelet coefficients. Relationships between Gabor features at different pixel locations are used as contextual constraints. Discriminant analysis is performed on the Gabor features to select most effective ones.

To solve problem (2), MRF is used to constrain matching configurations. The posterior probability of MRF is formulated as the criterion function for the evaluation of MRF configurations for matching between features in the gallery and probe. The best matching is defined as the maximum a posteriori (MAP) solution. The MAP solution is the best match between the two faces in terms not only image features but also relationships between them.

The rest of the paper is organized as follows: Section 2 introduces basic MRF concepts and Gabor wavelets as base features. Section 3 presents the use of MRF modeling to constrain matching configurations and formulates the posterior probability function for face matching. Experimental results are presented in Section 4 and in Section 5, we conclude the paper.

2 Background

2.1 MRF Modeling

The problem of face recognition may be posed as (1) finding correspondence between features in gallery and probe, which can be posed as a labeling problem; (2) computing the score of matching (or distance) between the two, which may be posed as a probability estimation problem, and (3) making decision based on the distances. Here we introduce the notions for labeling problems and how MRF could constrain labeling configurations with contextual relations.

Assume that a face is divided into a number of regions for both probe and gallery faces. Here a region is centered at a pixel location, and some regions may overlap each

other. Index the m regions in the probe by the set of sites $\mathcal{S} = \{1, \dots, m\}$, and L regions in the probe by the set of labels $\mathcal{L} = \{1, \dots, L\}$.

A labeling is then considered as realization $f = \{f_1, \dots, f_m\}$ of a random field denoted $F = \{F_1, \dots, F_m\}$. There, a random variable F_i takes a value f_i in \mathcal{L} , for which f is called a *configuration* of F , corresponding to a realization of the field.

The random variables F may be related to one another contextually via a *neighborhood system* $\mathcal{N} = \{\mathcal{N}_i \mid \forall i \in \mathcal{S}\}$ where \mathcal{N}_i is the set of sites neighboring i . The neighboring relationship can be defined to include sites with the Euclidean distance from i .

A neighborhood system specifies a collection of *cliques*. A clique consists either of a single site $c = \{i\}$, or of a pair of neighboring sites $c = \{i, i'\}$. The collections of single-site and pair-site are denoted by \mathcal{C}_1 and \mathcal{C}_2 , respectively.

F is said to be an MRF on \mathcal{S} with respect to \mathcal{N} if and only if $P(f) > 0$ (positivity), and $P(f_i \mid f_{\mathcal{S}-\{i\}}) = P(f_i \mid f_{\mathcal{N}_i})$ (Markovianity) are satisfied [1], where $\mathcal{S} - \{i\}$ is the set difference, $f_{\mathcal{S}-\{i\}}$ denotes the set of labels at the sites in $\mathcal{S} - \{i\}$ and $f_{\mathcal{N}_i} = \{f_{i'} \mid i' \in \mathcal{N}_i\}$ stands for the set of labels at the sites neighboring i . The Markovianity depicts the local characteristics of F .

The joint probability of a Markov random fields obeys a *Gibbs distribution*, which takes the following form

$$P(f) = Z^{-1} \times e^{-\frac{1}{T}U(f)} \quad (1)$$

where Z is a normalizing constant, T is the *temperature* constant, and $U(f)$ is the *energy function*. The energy

$$U(f) = \sum_{c \in \mathcal{C}} V_c(f) \quad (2)$$

is a sum of *clique potentials* $V_c(f)$ over all possible cliques \mathcal{C} . $P(f)$ measures the probability of the occurrence of a particular configuration, or “pattern”, and is a decreasing function of the energy. A clique potential is defined on the labels on the clique c .

Contextual constraints are encoded in $V_c(f)$ for c containing more than one site. Consider up to two site cliques, $U(f)$ can be expressed as

$$U(f) = \sum_{\{i\} \in \mathcal{C}_1} V_1(f_i) + \sum_{\{i, i'\} \in \mathcal{C}_2} V_2(f_i, f_{i'}) \quad (3)$$

where it is V_2 that encode contextual constraints.

2.2 Gabor Wavelets

In this work, we use Gabor wavelets, instead of the raw face image pixels, as basic face features because Gabor wavelets features exhibit desirable characteristics of spatial locality and orientation selectively, and are optimally localized in the space and frequency domains.

A face is represented Gabor-based vector, derived as a set of convolutions with a family of complex Gabor filters:

$$\psi_{\mu, \nu} = \frac{k_{\mu, \nu}^2}{\sigma^2} \exp(-\frac{k_{\mu, \nu}^2 z^2}{2\sigma^2}) [\exp(ik_{\mu, \nu} z) - \exp(-\frac{\sigma^2}{2})] \quad (4)$$

where μ and ν define the orientation and scale of the Gabor kernels respectively, $z = (x, y)$, and the wave vector $k_{\mu, \nu}$ is defined as follows:

$$k_{\mu, \nu} = k_{\nu} e^{i\phi_{\mu}} \quad (5)$$

where $k_{\nu} = k_{\max}/f^{\nu}$, $k_{\max} = \pi/2$, $f = \sqrt{2}$, $\phi_{\mu} = \pi\mu/8$, with index $j = \mu + 8\nu$. This representation is chosen for its biological relevance and technical properties. The Gabor kernels resemble the receptive field profiles of simple cells in the visual pathway. They are localized in both space and frequency domains and achieve the lower bound of the space-bandwidth product as specified by the uncertainty principle. At each pixel location $X = (u, v)$, the Gabor coefficients are computed as

$$J_j(X) = \int I(X') \psi_j(X - X') d^2 X' \quad (j = 0, 1, 2 \dots 40) \quad (6)$$

where $I(X)$ is the image grey level distribution.

Five scales, indexed by $\mu \in \{0, 1, 2, 3, 4\}$ and eight orientations, indexed by $\nu \in \{0, 1, 2, \dots, 7\}$, are used. Convolving the filters with an image gives 40 complex coefficients, and we can then get 40 Gabor magnitudes as a more salient representation.

The Gabor filtering gives a set of 40 features for each region I of the gallery, denoted as a 40-dim vector $D(I)$. We can similarly compute $d(i)$ for each region i of the probe. These are the original data.

3 MAP-MRF Based Face Recognition

3.1 Formulation

The basic face matching problem is to compare a probe and a gallery by finding the best correspondences or region matches between the two. The regions in the probe that find no good matches in the gallery is labeled NULL (“no-match”) indexed by 0. The problem of matching between probe and gallery using MRF modeling is formulated as labeling the sites in \mathcal{S} in terms of the label set $\mathcal{L}^+ = \{0, 1, \dots, L\}$.

The MAP-MRF matching maximizes the posterior probability, which is defined based on observation data and contextual constraints between sites

$$f^* = \arg \max_f P(f | d) \quad (7)$$

where d is the observation data and in this work corresponds to the set of Gabor features extracted from the probe face image. The posterior probability can be derived from the prior distribution $P(f)$ and the likelihood function $p(d | f)$. Assuming that f is a MRF, its distribution follows a Gibbs distribution of Equ.(1). Assuming that the likelihood distribution is also an exponential function $p(d | f) \propto e^{-U(d | f)}$, then the posterior probability has the following form

$$P(f | d) \propto e^{-U(f | d)} \propto e^{-U(f) - U(d | f)} \quad (8)$$

The MAP estimate is equivalently found by minimizing the posterior energy function

$$f^* = \arg \min_f U(f | d) \quad (9)$$

The following derives $U(f)$ and $U(d | f)$.

The likelihood function $p(d | f)$ has the following characteristics:

1. It is conditioned on pure non-NULL matches $f_i \neq 0$,
2. It is independent of the neighborhood system \mathcal{N} , and
3. It depends on how the model object is observed in the scene which in turn depends on the underlying transformations and noise.

Assuming that the Gabor features in d are invariant under the considered class of transformations, and d of probe is related the corresponding features D of gallery via the observation model

$$d(i) = D(f_i) + e(i) \quad (10)$$

where e is additive independent zero mean Gaussian noise. The Gaussian assumptions may not be accurate but offers an approximation.

Then the likelihood function is a Gibbs distribution with the energy

$$U(d | f) = \sum_{i \in \mathcal{S}, f_i \neq 0} V_1(d(i) | f_i) \quad (11)$$

where the constraint $f_i \neq 0$ restricts the summations to take over the non-NULL matches. The likelihood potentials are

$$V_1(d(i) | f_i) = \begin{cases} \Delta(d(i), D(f_i)) & \text{if } f_i \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

where $\Delta(d(i), D(f_i)) = |d(i) - D(f_i)|_e$ is some distance between the two random vectors respect to the noise e . In practice, we apply LDA (linear discriminant analysis) to find the discriminant subspace of the distribution of the training data, which is done over the whole image for all sites i , and use the Euclidean distance in the LDA subspace to replace the likelihood energy $U(d | f)$.

To model the prior distribution of Equ.(3), the single-site potentials are defined as

$$V_1(f_i) = \begin{cases} v_{10} & \text{if } f_i = 0 \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

where v_{10} is a constant.

Between the probe and gallery, a pair of local matched regions (i, f_i) incurs a displacement vector $D(i, f_i) = [dx(i, f_i), dy(i, f_i)]$. The displacement image can be considered as a “flow map”. The displacement in a neighborhood should be as consistent as possible. This is the smoothness constraint used as the contextual constraint.

We therefore define the pair-site potentials as

$$V_2(f_i, f_{i'}) = \begin{cases} v_{20} & \text{if } f_i = 0 \text{ or } f_{i'} = 0 \\ |D(i, f_i) - D(i', f_{i'})| & \text{otherwise} \end{cases} \quad (14)$$

where v_{20} is a constant and $|D(i, f_i) - D(i', f_{i'})|$ is the flow inconsistency. When the displacement vectors in the neighborhood are similar, the local flow inconsistency is small; otherwise they incur a large penalty. If either f_i or $f_{i'}$ is the NULL, it incurs

a penalty of v_{20} ; otherwise, the penalty is the flow inconsistency. The above clique potentials define the prior energy

$$U(f) = \sum_{i \in \mathcal{S}} V_1(f_i) + \sum_{i \in \mathcal{S}} \sum_{i' \in \mathcal{N}_i} V_2(f_i, f_{i'}) \quad (15)$$

The posterior energy is then obtained as $U(f | d) = U(f) + U(d | f)$.

3.2 Search Algorithm

Finding a MAP-MRF solution is to find the best matched correspondences between regions in the input probe and the gallery, according to both image features and their relationships. This amounts to finding the minimum of $U(f | d)$. While it is a combinatorial optimization problem and the global solution is computationally prohibitive, we present a heuristic local minimization algorithm to find a local solution as follows.

First, we do a local search over a neighborhood of i to minimize the simple Euclidean distance between $d(i)$ and $D(f_i)$. This gives the best matched region f_i^* in terms of the observation $d(i)$. For each i , if the distance is too large, we set $f_i^* = 0$. After this is done for all i , we calculate the Euclidean distance in LDA subspace between the data in d of the probe and D of the gallery. This gives the likelihood energy $U(d | f^*)$ of Equ.(11). After that, the displacements over the neighborhood can be computed. From these, the flow inconsistency can be estimated according to Equ.(14). The final posterior energy $U(f^* | d)$ is evaluated as the cost for the best overall cost.

Supposing there are N gallery faces, then N costs can be obtained for the N galleries: $E(f^{(n)}) = U(f^{(n)} | d)$. Let the lowest cost among the N matches be

$$E_{\min} = \min_{n \in \{1, \dots, N\}} E(f^{(n)}) \quad (16)$$

and the corresponding gallery image be

$$n_{\min} = \arg \min_{n \in \{1, \dots, N\}} E(f^{(n)}) \quad (17)$$

A simplest scheme for face recognition is the following: If the overall minimum energy is small enough, *i.e.* $E_{\min} < E_{Thr}$, the probe is recognized as the person that gallery n_{\min} belongs to, or otherwise it is considered as not in the gallery set. Some more sophisticated schemes are possible by considering all $E(f^{(n)})$ s and the associated person identities.

4 Experiments

The experiments are presented to compare face recognition performance without (non-MRF methods) and with (the proposed MAP-MRF method) using contextual constraint. Six non-MRF algorithms are evaluated: PCA, Fisher face (FLDA) [13], regularized LDA (RLDA) [14], Gabor+PCA, Gabor+FLDA, Gabor+RLDA, which are popular methods and have achieved great success in face recognition. The Gabor+PCA, Gabor+FLDA,

Gabor+RLDA and MAP-MRF methods use a selected set of 4000 Gabor features, for which 100 salient locations (regions) on a face are selected and 40 Gabor magnitude features are computed at each of the locations.

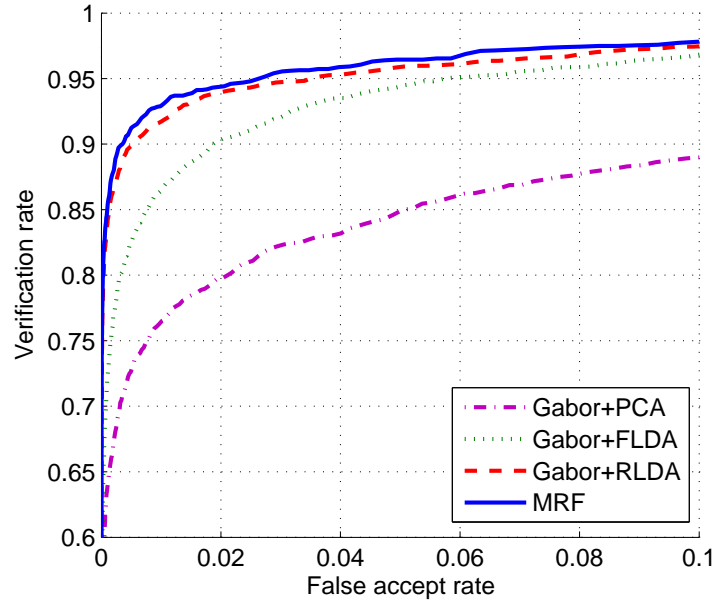
Two face databases, FERET [15] and FRGC ver 2.0 [16] are used for the experiments. All the images are rotated, scaled and cropped to 128×128 according to the provided eye positions succeeded by histogram equalization preprocessing. No further preprocessing is applied. For FERET database, the training set contains 731 images. In test phase, we use the gallery set containing 1196 images from 1196 subjects, and combine four provided probe sets together, totally including 2111 images to compose the probe set. So our test protocol should be more difficult than any of the four original protocols because we consider different factors (expression, illumination, aging etc.) together to evaluate the performance. For FRGC database, we select a subset from query set for experiment 4, which consists of still uncontrolled images including variations of illumination, expression, accessory and blurring. There are 316 subjects, each of which contains at least 10 images. We randomly select 10 images for each subject to get a total $316 \times 10 = 3,160$ images. These images are randomly divided into three sets. The training set consists of 116 persons, with 10 images per person. The left 200 subjects are divided into gallery and probe sets. For each person, 2 images randomly selected from the 10 images compose the gallery set and the left 8 images compose the probe set. The persons in training set are disjoint with those in gallery and probe sets. Fig. 1 illustrates some cropped face examples of FERET and FRGC databases.



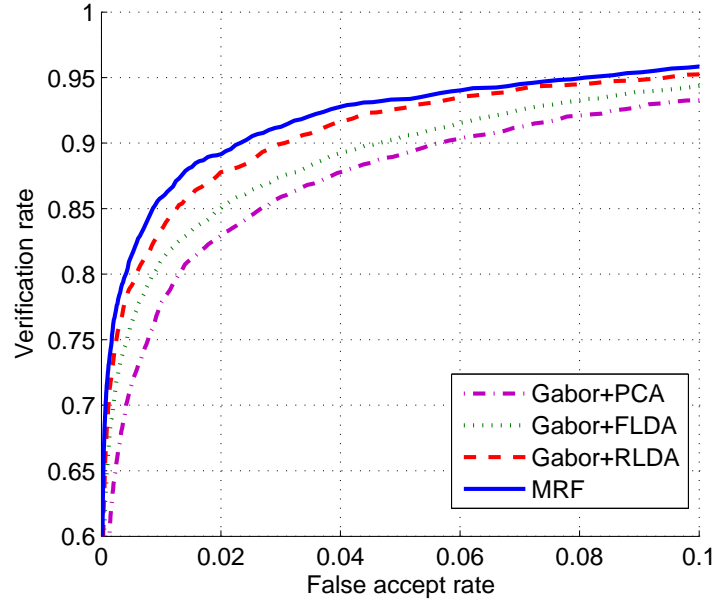
Fig. 1. Face examples of FERET (a) and FRGC (b) databases.

The results are reported in terms of three performance indices: rank-1 recognition rate, verification rate (VR) when the false accept rate (FAR) is 0.001, and equal error rate (EER).

Table 1 and 2 illustrate the results of different methods on FERET and FRGC databases respectively and Fig. 2 plots the corresponding ROC curves. For clarity, we only plot the results based on the Gabor features. In all experiments, the MAP-MRF method achieves the best accuracy. The results suggest that the use of contextual constraints is more robust to expression, aging, mis-alignment etc.



(a)



(b)

Fig. 2. Receiver operating characteristic (ROC) curves of different methods on FERET (a) and FRGC (b) databases.

Table 1. Performance of different methods on FERET database

Method	Rank-1	VR@FAR=0.001	EER
PCA	0.5201	0.4713	0.1268
FLDA	0.6798	0.5585	0.1243
RLDA	0.7688	0.6750	0.0872
Gabor+PCA	0.7144	0.6400	0.1058
Gabor+FLDA	0.7575	0.6296	0.0732
Gabor+RLDA	0.8693	0.8167	0.0543
MAP-MRF	0.8977	0.8541	0.0397

Table 2. Performance of different methods on FRGC database

Method	Rank-1	VR@FAR=0.001	EER
PCA	0.6219	0.2759	0.1686
FLDA	0.7887	0.4172	0.1677
RLDA	0.8625	0.5394	0.1133
Gabor+PCA	0.8581	0.5759	0.0779
Gabor+FLDA	0.8962	0.6169	0.0785
Gabor+RLDA	0.9150	0.6678	0.0666
MAP-MRF	0.9425	0.7250	0.0584

5 Summary and Conclusions

In this paper, we present a MRF modeling method for face recognition. Contextual constraints, which should play an important role in recognition, are encoded in the formulation in the MAP-MRF framework, such that the objective function encodes not only constraints on image features but also relationships between them. An algorithm is provided to find an approximate optimal solution. The results show the advantage of using contextual constraints for face recognition in the MAP-MRF framework, and suggest MAP-MRF as a potential and competitive alternative for robust face recognition.

Acknowledgements. This work was supported by the following funding resources: National Natural Science Foundation Project #60518002, National Science and Technology Support Program Project #2006BAK08B06, National Hi-Tech (863) Program Projects #2006AA01Z192, #2006AA01Z193, and #2008AA01Z124, Chinese Academy of Sciences 100 people project, and AuthenMetric R&D Funds.

References

1. Hammersley, J.M., Clifford, P.: Markov field on finite graphs and lattices. unpublished (1971)
2. Besag, J.: Spatial interaction and the statistical analysis of lattice systems (with discussions). Journal of the Royal Statistical Society, Series B **36** (1974) 192–236

3. Chellappa, R., Kashyap, R.L.: Digital image restoration using spatial interaction models. *IEEE Transactions on Acoustic, Speech and Signal Processing* **30** (1982) 461–472
4. Cross, G.C., Jain, A.K.: Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **5**(1) (January 1983) 25–39
5. Derin, H., Elliott, H., Cristi, R., Geman, D.: Bayes smoothing algorithms for segmentation of binary images modeled by Markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**(6) (November 1984) 707–720
6. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**(6) (November 1984) 721–741
7. Modestino, J.W., Zhang, J.: A Markov random field model-based approach to image interpretation. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (1989) 458–465
8. Qian, R.J., Huang, T.S.: Object detection using hierarchical mrf and map estimation. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (December 1997) 186
9. Dass, S.C., Jain, A.K.: Markov face models. In: *Proceedings of IEEE International Conference on Computer Vision, Vancouver* (July 9–12 2001) 112–116
10. Dass, S.C., Jain, A.K., Lu, X.: Face detection and synthesis using Markov random field models. In: *Proceedings of International Conference on Pattern Recognition, Quebec City* (August 2002) 112–116
11. Huang, R., Pavlovic, V., Metaxas, D.: A hybrid face recognition method using markov random fields. In: *Proceedings of International Conference on Pattern Recognition, Cambridge, UK* (August 2004)
12. Park, B.G., Kyoung-Mu, Lee, S.U.: Face recognition using face-arg matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(12) (2005) 1982–1988
13. Belhumeur, P., Hespanha, J., Kriegman, D.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(7) (1997) 711–720
14. Ye, J., Xiong, T., Li, Q., Janardan, R., Bi, J., Cherkassky, V., Kambhamettu, C.: Efficient model selection for regularized linear discriminant analysis. In: *CIKM*. (2006) 532–539
15. Phillips, P.J., Moon, H., Rizvi, S.A., Rauss, P.J.: The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(10) (2000) 1090–1104
16. Phillips, P.J., Flynn, P.J., Scruggs, W.T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.J.: Overview of the face recognition grand challenge. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. (2005) 947–954