Hierarchical Image Matching for Pose-invariant Face Recognition

Shervin Rahimzadeh Arashloo
S.Rahimzadeharashloo@surrey.ac.uk

Josef Kittler
J.Kittler@surrey.ac.uk

CVSSP
University of Surrey
Guildford, GU2 7XH, UK

Abstract

The paper addresses the problem of face recognition under arbitrary pose. A hierarchical MRF-based image matching method for finding pixel-wise correspondences between facial images viewed from different angles is proposed and used to densely register a pair of facial images. The goodness-of-match between two faces is then measured in terms of the normalized energy of the match which is a combination of both structural differences between faces as well as their texture distinctiveness. The method needs no training on non-frontal images and circumvents the need for geometrical normalization of facial images. It is also robust to moderate scale changes between images. The proposed approach is evaluated on the CMU PIE database and promising results are obtained.

1 Introduction

Face recognition technology compared to most of its alternatives is well suited to security applications in real world conditions as a result of being non-intrusive and operable at a distance. Although impressive performance has been reported in controlled conditions, many problems arising in realistic situations still remain unsolved. The two most challenging problems in realistic scenarios are change of facial image pose and illumination conditions.

A variety of different methods have been proposed to deal with the problem of changes in viewing angle. These works can be roughly categorized into three major groups. The first group can be considered as the methods which make use of multiple images corresponding to different poses stored in the gallery. One of the earliest works in this category is the work of Beymer [2] in which every input image was first aligned with the relevant reference images from the database and then a similarity measure was computed for recognition. Other work by Singh *et al*. [14] is another example of this group in which composite facial images are constructed using semi-profile and frontal views.

The second category are the methods which try to synthesize virtual views either in 2D or 3D. One of the most well known approaches in 2D is the active appearance model by Cootes *et al*. [4]. Another example is the work by Jimenez and Castro [5] in which point distribution models are constructed and eigenvectors responsible for appearance changes due to change of head pose are identified. After correcting the pose, Gabor filtering is used for recognition. In the work by Zhang *et al*. [22] frontal and side-view face images are used to generate virtual views of a subject. Once shape and texture of a subject's face are estimated, virtual images

in different poses are generated. For recognition, holistic methods are used. In 3D, one of the most successful methods is the 3D morphable model proposed by Blanz and Vetter [4]. In parallel with the methods which try to synthesize a novel view of the face, using either one or multiple images from the gallery, there are other methods which try to learn the most discriminant information between subjects across different poses. The work by Kim and Kittler [7] and Kanade and Yamada [6] are good examples of this category. In another work by Kim and Kittler [8] authors have tried to make use of different pose-invariant face recognition experts in a multiple classifier fusion system framework. There are also some relatively robust feature-based methods which can tolerate moderate pose variations such as the work by Pentland *et al*. [11] and by Wiskott *et al*. [21].

The algorithms in the first group fail when only images of one pose (typically frontal) for each subject are available in the database. The main drawback of the methods which try to synthesize novel views is the imperfection of the synthesizing process in addition to the requirement for prior labeling of landmarks which is usually carried out manually. An observation regarding the 3D methods is that although these techniques perform slightly better than the 2D alternatives, they still suffer from unresolved problems. The most important one is that in 3D-based geometric normalization methods, the recovered shape and texture are completely determined by the 3D morphable face model fitted to the query 2D face image which has the capacity to reconstruct only the information captured during statistical learning. As a result, these approaches can not recover atypical features that have not been available in the training set. Moreover, the high computational complexity of 3D methods in comparison with 2D algorithms makes them unsuitable for real-time applications.

One alternative to the synthesis of new views is to compare facial images as they are and rely only on the similarity of common visible parts between the query and the reference images. This process necessitates establishing dense correspondences between the images. With this motivation we propose a coarse-to-fine MRF-based image matching method which can establish pixel-wise correspondences between images and can also tolerate moderate scaling, translation, perspective effects, in-plane and out-of-plane rotation of the head. A pair of images are then compared using their structural differences as well as texture similarities. The underlying idea can be considered similar to the graph-based approaches common in general object recognition settings where the concepts of interest are assumed to be built up from simple neighboring primitives coded as the nodes of a graph and the neighborhood structure and contextual dependencies between primitives are represented as the edges. The recognition is then performed by evaluating the similarity of the two graphs in a matching process. Similar works in the area of face recognition based on this idea include *e.g*. the works in [9, 10, 17, 21].

The paper is organized as follows: In Section 2, the image matching method in [12] while considered as a baseline, is overviewed. In Section 3, a new hierarchical matching method is introduced. The similarity criterion for assessing the quality of a match is presented in Section 4. The results of an experimental evaluation of the method on the CMU PIE database [13] are presented and discussed in Section 5. Section 6 concludes the paper.

2 Image matching

Many computer vision problems can be formulated in an energy minimization framework where the objective function has the following form:

$$E(X|\theta) = \sum_{s \in V} \theta_s(x_s) + \sum_{(s,t) \in \varepsilon} \theta_{st}(x_s, x_t) \tag{1}$$

V corresponds to sites and ε to edges. x_s denotes the label of site $s \in V$. θ defines the parameters of the energy: θ_s denotes unary data penalty functions whereas θ_{st} denotes pairwise potentials. It is worth noting that in this formulation only cliques of size up to two are considered.

The minimum energy in equation (1) corresponds to the maximum probability of the solution on an MRF which is a Gibbs distribution.

2.1 Decomposed model

The method proposed in [12] formulates the image matching as a labeling problem on MRFs. The label set is $L_{reg} = \{(x_{s1}, x_{s2}) | x_{s1}, x_{s2} \in L\}$ where x_{s1} and x_{s2} denote displacements in horizontal and vertical directions. The deformation in horizontal and vertical directions is modeled by two MRFs interacting together. The edge set of this model is comprised of two separate edge sets. The edge potential functions on each of these layers are assumed to be identical while inter-layer edges encode the data term. For the intra-layer edges the following crisp continuity terms are used in [12]:

$$\theta_{st}(x_s, x_t) = \begin{cases} 0, & x_s = x_t, \\ c_r, & |x_s - x_t| = 1, \\ \infty, & |x_s - x_t| > 1. \end{cases} \tag{2}$$

In order to achieve more flexibility in deformation, we replace hard continuity terms by quadratic penalty function:

$$\theta_{st}(x_s, x_t) = c(x_s - x_t)^2 \tag{3}$$

where c is a normalizing constant. In [12], by restricting the neighboring blocks (blocks are of size 4×4) to have relative displacements of no more than one pixel, the scale changes were limited to [.75, 1.25] of the model image size whereas by replacing the hard constraints by a quadratic term a much greater range of scales can be accommodated.

The inter-layer edges encode the data term, *i.e.* the cost of assigning label x_{s1} in layer one and label x_{s2} in layer two to two isomorphic nodes of the graph. The data term has been constructed using *block model*. In the block model, the pixels are grouped into non-overlapping blocks which correspond to nodes of the graph. The data term for the block model is defined as below:

$$\theta_{s1s2}(x_{s1}, x_{s2}) = \frac{1}{\sigma^2} \mathrm{Dis}(I^1_{\vec{s}}, I^2_{\vec{s}+(x_{s1}, x_{s2})}), s^1 \in V^1, s^2 \in V^2 \tag{4}$$

where $I^1_{\vec{s}}$ is a block on image I^1 and the corresponding block on image I^2 is denoted by $I^2_{\vec{s}+(x_{s1}, x_{s2})}$, which is the block with the coordinates $\vec{s} + (x_{s1}, x_{s2})$, where $\vec{s}$ is the vector pointing to the position of block $I^1_{\vec{s}}$. Dis(.,.) is a dissimilarity measure which is defined as the sum of squared differences over the pixels of corresponding blocks. Since edge maps are less affected by unwanted illuminations changes, in order to achieve robustness against changes

in illumination, we use horizontal and vertical edge maps instead of gray scale images. Horizontal and vertical edges are normalized to the range [-1,1] and combined to form the data term. Moreover, in computing the data term, not only the block under consideration but also a neighborhood is used to collect local support from a small neighborhood. One can assign more weight to the block under consideration and lower weights to its neighbors. It has been observed that computing the data term in this way improves the matching accuracy. Thus the data term is computed as:

$$\theta_{s^1 s^2}(x_{s^1}, x_{s^2}) = \frac{1}{\sigma^2} \left[\sum_{\vec{v} \in N(\vec{s})} w_{\vec{v}} \text{Dis}(I_{\vec{v}}^{1h}, I_{\vec{v}+(x_{s^1},x_{s^2})}^{2h}) + \sum_{\vec{v} \in N(\vec{s})} w_{\vec{v}} \text{Dis}(I_{\vec{v}}^{1v}, I_{\vec{v}+(x_{s^1},x_{s^2})}^{2v}) \right],$$
$$s^1 \in v^1, s^2 \in v^2 \qquad (5)$$

where $I_{\vec{v}}^{1h}$ and $I_{\vec{v}+(x_{s^1},x_{s^2})}^{2h}$ denote a block in the horizontal edge map of the first image and its corresponding block in the horizontal edge map of the target image respectively. $I_{\vec{v}}^{1v}$ and $I_{\vec{v}+(x_{s^1},x_{s^2})}^{2v}$ correspond to the same block coordinates but in the vertical edge maps. $w_{\vec{v}}$ stands for the weight we choose for each of the neighbors of the block under consideration. Also, $N(\vec{s})$ denotes a neighborhood of a block with the vector $\vec{s}$ pointing to its center.

It is desirable to rely more on common features of the two images and bypass the atypical features which appear only in one image. This can be achieved by ignoring the weak edges and setting those below a threshold to zero and also by truncating the data term which makes the matching more robust to outliers and occlusions.

3 Hierarchical image matching

The method in [12] groups pixels into non-overlapping blocks and estimates a displacement vector for each block. The advantages of using block model can be stated as follows. First, by using blocks instead of pixels, the number of nodes in the underlying graph reduces which results in reduced computational cost. Secondly, using the block model for computing the data term makes the method more robust against noise. The drawback of this model is the error induced by assuming that all pixels inside a block have similar disparity vectors. One way to obtain disparities for each pixel is to construct the data term using single pixels instead of block model. However, this prohibitively increases the computational burden and also makes the system more susceptible to noise. The robustness against noise is a desirable property here as one would need to match facial images which can potentially be taken with different devices and correspond to different subjects viewed from different angles. The other drawback is the increased probability of the optimization method to get stuck in a local minimum as a result of increased dimensionality of the configuration space. One alternative to this approach is to match images in a multi-scale scheme in which upper levels of the hierarchy give a rough estimate of the motion field and lower levels fine-tune the estimated motion. This can be achieved by employing *relatively* larger (compared to the image size) blocks in the higher levels of the hierarchy and relatively smaller blocks in the lower levels. It should be noted that what differs in different levels is the *relative* size of the block to the image and not the absolute sizes. Thus, in principle instead of low-pass filtering and sub-sampling the image, one can only low-pass filter the image and retain the original size of the image but use different block sizes in different levels. Inspired by this idea which has been previously employed for motion estimation (see *e.g.* [18]), we propose a
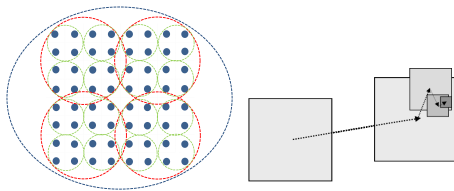
Figure 1: Hierarchical image matching: Left: grouping pixels into clusters in four different scales, Right: disparity estimation for a single node (the darkest block) in four different scales.

novel method which facilitates multi-resolution image matching within the framework of the efficient optimization methods for MRFs. In a method in which higher levels of pyramid are low-pass filtered and **sub-sampled** versions of the original image, fine scales of disparities cannot be detected as a result of sub-sampling. In contrast, by retaining the original image size, even fine disparity information is not lost in the higher levels of the hierarchy. In this approach all images corresponding to different scales are of the same size, but higher levels are more blurred versions of the original image. The efficiency gained in this method comes from the fact that at higher levels, the number of nodes is reduced. Once the finer scales are reached, it is sufficient to search for correspondences in a smaller area since the upper level gives a rough estimate of the disparity vector. Also, the robustness to noise is enhanced by virtue of matching more blurred versions of the images at higher levels. This approach is illustrated in Figure 1. Apart from the speed gain in the hierarchical scheme, comparably better solutions can be obtained as a result of getting closer to the global minimum and reducing the probability of getting stuck in local minima far from the global one. This property comes partly from the fact that at higher levels, larger blocks capture relatively longer range interactions and the rough estimate of the disparities makes the optimization bypass local minima and converging closer to the true global minimum while finer scale fine tune the obtained solution. Typical values for the gaussian filter order (with binomial coefficients), block sizes and disparity search ranges are given in Table 1.

There are two factors that have to be taken into account in image matching. The first one is the uniqueness constraint and the second is the occlusion problem. The uniqueness constraint states that every primitive in one image can at most correspond to one primitive in the other image. This property is not desirable here. In practice, facial images can be in different scales. Because we are interested in finding pixel-wise correspondences, apparently one pixel in the target image with a relatively smaller scale can potentially correspond to more than one pixel in the template image with relatively larger scale. The second issue is the occlusion problem. The problem can be much readily dealt with in applications like stereo in which the disparities are only in one direction and the images are exactly the same except the viewing angle. A good example among other methods which solve for occlusions in a general matching scenario is the work in [19] where sparse image features are matched. The situation in here is more complicated because the images to be matched may correspond to different individuals taken from different viewpoints subject to self-occlusion. Also, instead of sparse image feature based matching, pixel-wise correspondences are sought. As a result, the common approaches used in MRF matching, *e.g.* augmenting the label set with an extra label (null label) representing all content but the content of interest, makes the system very vulnerable to small differences in the appearance and tend not to yield stable results in the

Level	3	2	1	0
Block size	8×8	4×4	2×2	1×1
Maximum disparity	40	20	10	5
Filter order	20	15	5	-

Table 1: Typical values for block size, disparity search range and Gaussian filter order in hierarchical image matching.
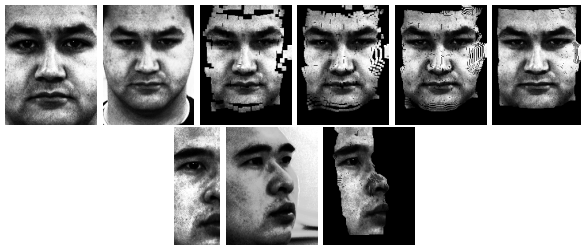

Figure 2: Up row from left to right: template, target and results of matching in four different scales (level 3, 2, 1 and 0), bottom row: result of matching a half face to a non-frontal facial image: from left to right: template, target, matched template.

presence of self-occlusion (a common situation when viewing faces from different angles). Occlusion, in this case can be better dealt with by employing higher order interactions in an MRF (possibly a global constraint on the joint probability of labels at all nodes). There are a number of recent papers proposing possible ways of incorporating higher order interactions in the underlying graph structure, *e.g.* the works in [15, 20?]. The advantage of using higher-order interaction is in tighter relaxations. In spite of the progress in this direction, current methods are not computationally very efficient. Solving the labeling problem with an efficient method in the presence of higher order interactions is left as as a future objective in this work. Instead of directly solving for occlusion during optimization, in the case of pan movement of the head we use only half of the face closer to the sensing device. In the case of tilt movement we make the assumption that the whole face is visible and solve the matching for the whole face. Figure 2 illustrates the results of applying our hierarchical matching method on two example pairs of faces from the CUM-PIE database. In order to decide whether the left half, right half or the full face of the gallery image should be matched to the probe image one may take advantage of a pose estimation method which could only detect whether the subject's head's motion has a pan component to the right or to the left. Although a pose estimation method may speed up the method, it is not strictly needed. The problem can be handled using the assumption that if there is a pan component of the head movement then using half face images would result in lower normalized energy. Hence, one can match full face, left half and also right half of the gallery image to the probe image and compare the normalized energy of the match to determine the pose of the probe image. Although we have not used this exhaustive method (for efficiency reasons) we confirmed its validity during our investigations.

Figure 3: From left to right: initial image, after photometric normalization and after applying LBP.

4 The similarity criterion

Once dense correspondences have been established between a pair of images, their similarity is measured in terms of their textural content as well as structural differences. In the following we first describe how to compare two images using their texture and then how to measure the differences between a pair of facial images using the configurational arrangements of their entities.

4.1 Textural content

Different facial images may be exposed to different lighting conditions during image capture time. Therefore, applying a photometric normalization step before measuring the similarity seems necessary. In [16] an effective photometric normalization method is introduced. The technique is based on a sequence of stages designed to decrease the effects of illumination changes while preserving the fundamental visual content. Figure 3 shows the result of applying the method on a sample face image.

Before comparing the similarity of the textural contents of the two images under consideration, in order to extract texture features, we apply a local binary pattern operator. The LBP is a non-parametric operator defined at a given pixel location (x_o, y_o) as:

$$LBP(x_o, y_o) = \sum_{n=0}^{7} L(i_n - i_o)2^n \tag{6}$$

where i_o is the gray value of the pixel under consideration and i_n is the gray value of its 8 neighboring pixels. Where function $L(x)$ is defined as:

$$L(x) = \begin{cases} 1 & x \geq 0, \\ 0 & x < 0. \end{cases} \tag{7}$$

4.2 Structural information

The structural information of faces can be also exploited to enhance the performance of the system. Apparently, different faces apart from their textures have different structures which can provide discriminatory information in authentication. In [1] factors which unfavorably affect the binary term of the energy for recognition are identified as pose, non-rigidity of the pattern and last but not least, statistical dependencies between residual displacements of neighboring sites as well as the limited cardinality of the potential functions. In order to remove the effect of the rigid motion, the distortion associated with the global spatial transformation was subtracted from distortion vectors thus achieving global spatial transformation

Figure 4: Illustration of pose variation in the CMU PIE database [13].

invariance. In order to take into account non-rigidity of patterns, a number of different exemplars of each class were matched one to another and the average distortion was considered as a class-specific model of deformations. The problems of inherent correlation between residual displacements of neighboring sites and that of the limited cardinality of the cliques defining the potential functions were partly compensated for by modeling these interactions using covariance matrices which convey correlation information between different sites even at a larger range. In this work we estimate covariance matrices for the full face in the case of tilt movement of head and for the half face in the case of pan movement of the head of the subject. The structural differences between a pair of images are hence gauged in terms of the Mahalanobis distance:

$$D_{Mahalanobis}(I_i, J) = (\bar{e}_v - \bar{\mu}_{iv})^t \Sigma_v^{-1} (\bar{e}_v - \bar{\mu}_{iv}) + (\bar{e}_h - \bar{\mu}_{ih})^t \Sigma_h^{-1} (\bar{e}_h - \bar{\mu}_{ih}) \qquad (8)$$

where I_i is a template of class i, $\bar{\mu}_{iv}$ and $\bar{\mu}_{ih}$ are the average distortions for this class in vertical and horizontal directions respectively pursued in a raster scan fashion. $\bar{e}_v$ and $\bar{e}_h$ are the local distortion vectors obtained after matching I_i to J. In order to obtain the local distortions, a projective transformation is fitted to the set of corresponding points and the effect of rigid motion is subtracted from the distortion field. Σ_v^{-1} and Σ_h^{-1} represent inverse covariance matrices capturing the correlation of distortions in vertical and horizontal directions respectively. The final distance measure between a pair of images is finally defined as a weighted sum of their texture differences and structural distance.

5   Experimental setup and results

The proposed method for recognizing rotated faces is tested on a subset of the CMU PIE database [13] consisting of 884 images of 68 subjects viewed from 13 different angles. Nine of thirteen images of each subject are captured roughly at the head height spanning a range of angles from approximately full right profile to full left profile. Two of the remaining images were captured from the corners of the room and the other two were captured from above and below the central camera. In this way images of substantially different poses of each subject are available. Figure 4 shows an example of pose variation in the CMU PIE database. As we want to evaluate the performance of the system subject to viewing angle, we just use images under the same illumination and neutral expression. The frontal views of subjects (pose 27) are considered as gallery images and all the rest (12 different poses) are used as test images.
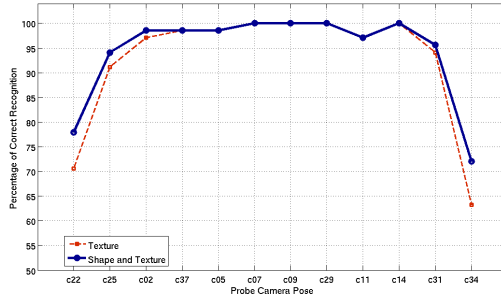
Figure 5: Recognition rates of the proposed approach across different poses using only texture and shape&texture information together.

pose	PDM [■]	AA-LBP [□]	The proposed approach
02	72	95	**98.52**
05	100	100	98.52
07	N/A	100	100
09	N/A	100	100
11	94	100	97.058
14	62	91	**100**
22	N/A	N/A	77.94
25	N/A	89	**94.11**
29	98	100	100
31	N/A	80	**95.58**
34	20	73	72.05
37	89	100	98.52
Average	N/A	93.45	**95.85**

Table 2: Comparison of correct recognition percentage across different poses between the proposed method and the works in [5] and [22].

Figure 5 shows the result of recognition using texture and also shape and texture together. It is worth noting that the obtained recognition rate using only texture information does not need any training, neither on frontal nor on non-frontal images. The improvement using structural information is marginal. One reason for this is the inability of a simple projective transformation to model pose changes of a 3D non-planar object to provide substantially useful discriminatory information. The second reason is that shape and texture information are more or less correlated in the sense that if the method fails to find a good match between images both shape and texture comparison stages get affected. In Table 2 the results of a comparison of the proposed approach with two other methods are presented. It is evident from the table, the proposed approach outperforms the method in [5] in all poses except pose 05. The overall performance of the proposed method is far better than the one in [5]. Also, our method outperforms the state-of-the-art method in [22] in four different poses, especially when there is a large deviation in viewing angle from frontal. Also, the average performance achieved is better than that of the work in [22] (the average performance is computed after excluding pose 22 in order for the result to be comparable with the method in [22]). It should be mentioned that both methods in [5] and [22] take advantage of non-frontal images for training whereas we have only frontal images as gallery images.

6   Conclusions

The problem of face recognition under arbitrary pose was addressed. A hierarchical MRF-based image matching method for establishing pixel-wise correspondences between facial images was proposed and used to densely register a pair of facial images. The quality of the match between two faces was gauged in terms of the normalized energy of the match, capturing both structural differences between faces as well as their texture distinctiveness. Without any training on non-frontal images and also without geometrical normalization of facial images the proposed approach gave promising results comparable and also outperforming (in some poses) the state-of-the-art methods when tested on the CMU PIE database.

References

[1] S. R. Arashloo and J. Kittler. Pose-invariant face matching using mrf energy minimization framework. *Energy Minimization Methods in Computer Vision and Pattern Recognition, 2009. EMMCVPR '09. International Workshop on*, pages 56–69, August 2009.

[2] D.J. Beymer. Face recognition under varying pose. *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, pages 756–761, Jun 1994.

[3] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(9):1063–1074, Sept. 2003.

[4] T.F. Cootes, G.J. Edwards, and C.J. Taylor. Active appearance models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(6):681–685, Jun 2001.

[5] D. Gonzalez-Jimenez and J.L. Alba-Castro. Toward pose-invariant 2-d face recognition through point distribution models and facial symmetry. *Information Forensics and Security, IEEE Transactions on*, 2(3):413–429, Sept. 2007.

[6] T. Kanade and A. Yamada. Multi-subregion based probabilistic approach toward pose-invariant face recognition. *Computational Intelligence in Robotics and Automation, 2003. Proceedings. 2003 IEEE International Symposium on*, 2:954–959 vol.2, July 2003.

[7] Tae-Kyun Kim and J. Kittler. Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(3):318–327, March 2005.

[8] Tae-Kyun Kim and J. Kittler. Design and fusion of pose-invariant face-identification experts. *Circuits and Systems for Video Technology, IEEE Transactions on*, 16(9):1096–1106, Sept. 2006.

[9] C. Kotropoulos, A. Tefas, and I. Pitas. Frontal face authentication using morphological elastic graph matching. *Image Processing, IEEE Transactions on*, 9(4):555–560, Apr 2000.

[10] M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R.P. Wurtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *Computers, IEEE Transactions on*, 42(3):300–311, Mar 1993.

[11] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, pages 84–91, Jun 1994.

[12] A. Shekhovtsov, I. Kovtun, and V. Hlavac. Efficient mrf deformation model for non-rigid image matching. *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–6, June 2007.

[13] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression database. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(12):1615–1618, Dec. 2003.

[14] R. Singh, M. Vatsa, A. Ross, and A. Noore. A mosaicing scheme for pose-invariant face recognition. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 37(5):1212–1225, Oct. 2007.

[15] David Sontag, Talya Meltzer, Amir Globerson, Tommi Jaakkola, and Yair Weiss. Tightening lp relaxations for map using message passing. *UAI*, pages 503–510, 2008.

[16] Xiaoyang Tan and Bill Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Analysis and Modelling of Faces and Gestures*, 4778:168–182, Oct 2007.

[17] A. Tefas, C. Kotropoulos, and I. Pitas. Using support vector machines to enhance the performance of elastic graph matching for frontal face authentication. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(7):735–746, Jul 2001.

[18] A. Murat Tekalp. *Digital video processing*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1995. ISBN 0-13-190075-7.

[19] Lorenzo Torresani, Vladimir Kolmogorov, and Carsten Rother. Feature correspondence via graph matching: Models and global optimization. *Computer Vision - ECCV 2008*, pages 596–609, 2008.

[20] T. Werner. High-arity interactions, polyhedral relaxations, and cutting plane algorithm for soft constraint optimisation (map-mrf). *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8, June 2008.

[21] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):775–779, Jul 1997.

[22] Xiaozheng Zhang, Yongsheng Gao, and Maylor Leung. Recognizing rotated faces from frontal and side views: An approach toward effective use of mugshot databases. *Information Forensics and Security, IEEE Transactions on*, 3(4):684–697, Dec. 2008.