An Illumination Invariant Face Recognition System for Access Control using Video

Ognjen Arandjelović   Roberto Cipolla
Department of Engineering
University of Cambridge
Cambridge, CB2 1PZ, UK
{oa214,cipolla}@eng.cam.ac.uk

Abstract

Illumination and pose invariance are the most challenging aspects of face recognition. In this paper we describe a fully automatic face recognition system that uses video information to achieve illumination and pose robustness. In the proposed method, highly nonlinear manifolds of face motion are approximated using three Gaussian pose clusters. Pose robustness is achieved by comparing the corresponding pose clusters and probabilistically combining the results to derive a measure of similarity between two manifolds. Illumination is normalized on a per-pose basis. Region-based gamma intensity correction is used to correct for coarse illumination changes, while further refinement is achieved by combining a learnt linear manifold of illumination variation with constraints on face pattern distribution, derived from video. Comparative experimental evaluation is presented and the proposed method is shown to greatly outperform state-of-the-art algorithms. Consistent recognition rates of 94-100% are achieved across dramatic changes in illumination.

1   Introduction

Important practical applications of automatic face recognition have made it a very popular research area in the last three decades, see [3, 5, 6, 17] for surveys. Most of the methods developed deal with *single-shot* recognition. In controlled imaging conditions (lighting, pose and/or occlusions) many have demonstrated good (nearly perfect) recognition results [17]. On the other hand, single-shot face recognition in uncontrolled, or loosely controlled conditions still poses a significant challenge [17].

The nature of many practical applications is such that more than a single image of a face is available. In surveillance, for example, the face can be tracked to provide a temporal sequence of a moving face. In access control use of face recognition the user may be assumed to be cooperative and hence can be instructed to move in front of a fixed camera. Regardless of the setup in which multiple images of a face are acquired, it is clear that this abundance of information can be used to achieve greater robustness of face recognition by resolving some of the inherent ambiguities of the single-shot recognition problem.

The organization of this paper is as follows. Section 2 reviews the existing literature on face recognition from video. Section 3 gives the overview of the proposed method. In Section 3.1 the benefits of registration in the proposed framework are explained. Section 3.2 shows how we cluster faces by pose. Section 3.3 introduces the proposed method of illumination normalization. In Section 3.4 it is shown how a unified measure of similarity between face motion manifolds is obtained. Section 4 reports experimental results and compares the proposed method with several competing methods reported in the literature. Finally, Section 5 concludes the paper and discusses promising directions for future research.

2   Related Previous Work

Single-shot face recognition is a well established research area. Algorithms such as Bayesian Eigenfaces [12], Fisherfaces [17], Elastic Bunch Graph Matching [10] or the 3D Morphable Model [4, 13] have

demonstrated good recognition results when illumination and pose variations are not large. However, all existing single-shot methods suffer from the limited ability to generalize to unseen illumination conditions or pose.

Compared to single-shot recognition, face recognition from video is a relatively new area of research. Most of the existing algorithms perform recognition from image sequences, using the temporal component to enforce prior knowledge on likely head movements. In the algorithm of of Zhou *et al.* [18] the joint probability distribution of identity and motion is modelled using sequential importance sampling, yielding the recognition decision by marginalization. In [11] Lee *et al.* approximate face manifolds by a finite number of infinite extent subspaces and use temporal information to robustly estimate the operating part of the manifold.

Fewer methods recognize from manifolds without the associated ordering of images, which is the problem addressed in this paper. Two algorithms worth mentioning are the Mutual Subspace Method (MSM) of Yamaguchi [8] and the Kullback-Leibler divergence method of Shakhnarovich *et al.* [14].

In MSM, infinite extent linear subspaces are used to compactly characterize face sets i.e. the manifolds that they lie on. Two sets are then compared by computing the first three principal angles between corresponding principal component analysis (PCA) subspaces [8]. Varying recognition results were reported using MSM, see [8, 14, 16]. A major limitation of MSM is its simplistic modelling of manifolds of face variation. Their high nonlinearity (see Figure 1(a)) invalidates the assumption that data is well described by a linear subspace. More subtly, the nonlinearity of modelled manifolds means that the PCA subspace estimate is very sensitive to the particular choice of training samples. For example, in the original paper [8] in which face motion videos were used, the estimates are sensitive to the extent of rotation in a particular direction. Finally, MSM does not have a meaningful probabilistic interpretation.

The Kullback-Leibler divergence (KLD) based method [14] is founded on information-theoretic grounds. In the proposed framework, it is assumed that i-th person's face patterns are distributed according to $p_i(\mathbf{x})$. Recognition is then performed by finding $p_j(\mathbf{x})$ that best explains the set of input samples – quantified by the Kullback-Leibler divergence. The key assumption in their work, that makes divergence computation tractable, is that face patterns are normally distributed i.e. $p_i(\mathbf{x}) = \mathcal{N}(\bar{\mathbf{x}}_i, \mathbf{C}_i)$. This is a crude assumption (see Figure 1(a)), which explains the somewhat poor results reported with this method [16]. KLD was also criticized for being asymmetric [1, 9].

More subtly, both approaches have the disadvantage of comparing whole face distributions, which has the implicit assumption that for the same people we expect the training and testing distributions to be similar. This does not have to be the case for confident recognition. Consider the case when in the training video the head motion was from frontal face to the left and in the testing video from frontal face to the right. Clearly, the two manifolds will be different even if the imaging conditions (such as lighting) are unchanged. Still, the intersection of the manifolds in the region of the frontal face provides enough information for confident recognition decision.

Finally, neither of the two methods addresses the issue of changing illumination that inevitably occurs in most practical applications. This is the most challenging problem of automatic face recognition [15].

3 Face Recognition from Motion Manifolds

Video of a face in motion carries information about its 3D shape and albedo. This information can be used either explicitly, by recovering a model of the face (e.g. [4]), or implicitly by modelling manifolds of face pattern variations (e.g. [1]). We employ the latter approach. In our method, manifolds of face variations are modelled using three Gaussian clusters describing small face motion around different head poses. Given two such manifolds, first the pose clusters are determined, then corresponding clusters are compared and finally, the results of the pairwise comparisons are combined to give a unified measure of similarity of the manifolds.

3.1 Registration

Manifolds of faces in motion are complex and nonlinear (see Figure 1(a)), and modelling them using Gaussian clusters becomes increasingly difficult as their intrinsic dimensionality is increased. There-
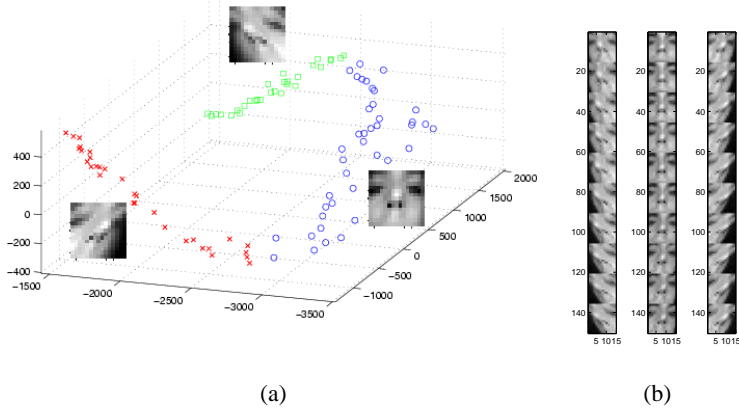
(a)   (b)

Figure 1: A typical face manifold of mainly lateral head rotation around the fronto-parallel face ($\pm30°$) (a). Shown is the projection to the first 3 PCA components. The manifold is smooth, but highly nonlinear. Different pose cluster memberships are marked in different styles with the associated mean images displayed. Example automatically affine registered and cropped faces from the 3 pose clusters can be seen in (b).

fore, it is advantageous to normalize the raw, input frames as much as possible so as to minimize the dimensionality of modelled manifolds. Since reliable methods for facial feature localization have been developed [7], some of the pattern variations are easily removed directly, that is, by recovering transformation parameters from sets of point correspondences. Images can then be registered to have relevant facial features in selected canonical locations.

In our method, 4 characteristic facial points are used for affine registration: the locations of pupils and nostrils (see Figures 6 and 7). Since 4 point correspondences over-determine affine transformation parameters (6), we estimate them in the minimum L_2 error sense.

3.2   Clustering by Pose

In our method, both recognition and illumination normalization are performed on a per pose basis. We describe face motion manifolds using Gaussian clusters corresponding to different head poses.

Inspection of manifolds of registered faces in random motion around the fronto-parallel face shows that they are dominated by the first nonlinear principal component. This principal component corresponds to lateral head rotation, see Figure 1(a). Therefore, the centres of Gaussian clusters used to characterize them should correspond to different yaw angle values. In this work we describe the manifolds using three Gaussian clusters, corresponding to the frontal face orientation, face left and face right.

3.2.1   Finding Pose Clusters

As the extent of lateral rotation, as well as the number of frames corresponding to each cluster, can vary from video to video, a generic clustering algorithm, such as the k-means algorithm, is unsuitable for finding the three Gaussians.

With the prior knowledge of the semantics of clusters, we decide on a single frame membership frame-by-frame. We found that the locations of pupils and nostrils (see Section 3.1) are sufficient to distinguish between the three clusters. On top of its simplicity, this method has the attractive feature of introducing little computational overhead. We define the quantity η as follows:

$$\eta = \frac{1}{2}\frac{x_{reye}+x_{leye}-x_{rnostril}-x_{lnostril}}{x_{reye}-x_{leye}} \tag{1}$$

Quantity η measures the shift of the projection of the centre point of pupils from that of nostrils, see Figure 2(a). As nostrils are further away from the centre of the head (and hence the axis of rotation), the magnitude of η will increase with head yaw deviation from the frontal orientation. The distribution of
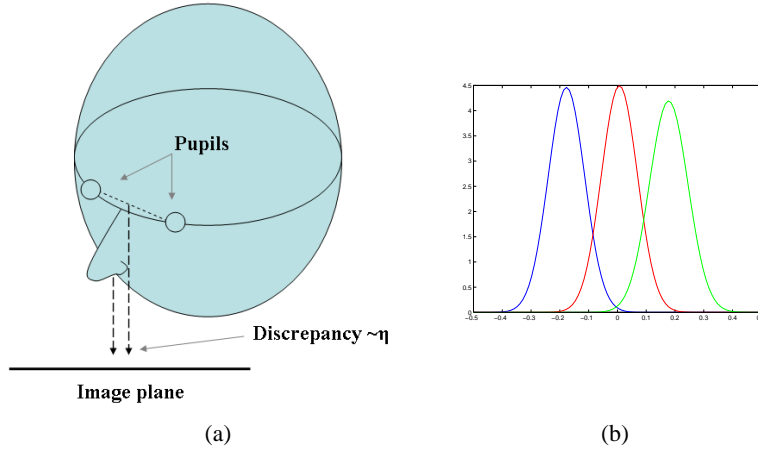
Figure 2: Parallax used to cluster input face images (a). The distributions of η (1) for the three clusters, computed from 200 manually labelled frames is shown in (b). Good separation of clusters is demonstrated.

the value of η for each pose is shown in Figure 2(b). A frame in our method is classified to the maximal likelihood pose. Examples of classified frames can be seen in Figure 1(b).

3.3 Illumination Normalization

Illumination variation of face patterns is extremely complex due to varying texture reflectance properties, face shape and, type and distance of lighting sources. Hence, in such a general setup, it is difficult to learn. However, on the coarse level most of the variation can be described by dominant light direction e.g. 'strong left light'. This is a much easier problem to address and it significantly simplifies the learning of the residual variation. This motivates the two-stage, per-pose illumination normalization employed in the proposed method:

1. Region-based gamma intensity correction, followed by

2. Illumination subspace normalization.

3.3.1 Gamma Intensity Correction

Gamma intensity correction (GIC) compensates for global brightness changes of an image. It transforms image pixel values by exponentiation so as to best match a canonically illuminated image. Formally, given an image I and a canonically illuminated image I_C, the gamma intensity corrected image I^* is defined as follows:

$$\gamma^* = \arg\min_{\gamma} \sum_{x,y} [I(x,y)^{\gamma} - I_C(x,y)]^2 \tag{2}$$

$$I^*(x,y) = I(x,y)^{\gamma^*} \tag{3}$$

This is a nonlinear optimization problem and in our implementation of the proposed method it is solved using the Golden Section search with parabolic interpolation.

In region-based GIC images are divided into regions corresponding to smoothly varying surface normals of the imaged object, and GIC is applied to each of them separately, see Figure 3. An undesirable feature of this method is that it tends to produce artificial intensity discontinuities at region boundaries [15]. This is due to discontinuities in the computed gamma value maps. For this reason, in our method the obtained gamma value maps are Gaussian smoothed before input images are transformed according to them. This almost completely remedies the problem, see Figure 3.
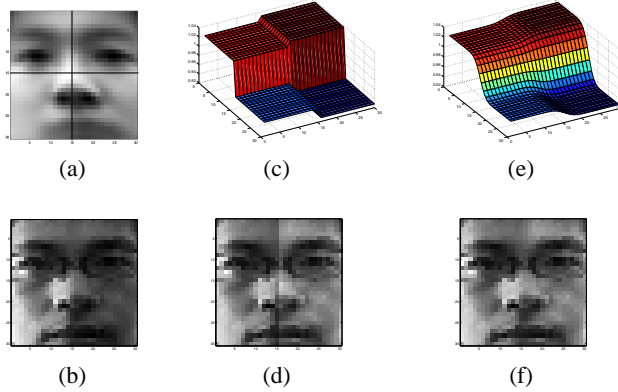
Figure 3: Canonical illumination image and the regions used in region-based GIC (a), original unprocessed face image (b), gamma value map (c), smoothed gamma value map (d), region-based GIC corrected image without smoothing (e), and region-based GIC corrected image with smoothing (f). Notice the artefact at region boundaries in the gamma corrected image (e). The image (f) does not have the same problem. Note that the coarse effects of the strong side lighting in (b) have been greatly removed.

3.3.2 Illumination Subspace Normalization

After region-based GIC is applied to all images, for each of the pose clusters, it is assumed that the lighting variation can be modelled using a linear, *pose illumination subspace*. Given a reference and a novel cluster corresponding to the same pose, each frame of of the novel cluster is normalized for the illumination change. This is done by adding a vector from the pose illumination subspace to it so that its distance from the reference cluster's centre is minimal.

Constructing a pose illumination subspace. We construct a pose illumination subspace by performing PCA on deviations of each persons's images under different illuminations from the same person's mean image for that pose and retain the eigenvectors that explain 90% of data energy. In other words, for each pose, given that $\mathbf{x}_{i,j}^k$ is the k-th frame of person i under illumination j, we perform PCA on data $\mathbf{x}_{i,j}^k - \bar{\mathbf{x}}_i$ (over all i, j and k), where $\bar{\mathbf{x}}_i$ is the person i's mean image.

From the way this subspace is constructed, it can be seen that it explains a lot of variation: changes in illumination due to varying illumination conditions and albedo, as well as some motion (as each pose cluster describes faces for a range of yaw angles). This is especially the case as we do not make the assumption that faces are Lambertian, or that the light sources are point lights at infinity. The significance of this is that a large enough subspace to explain the modelled phenomenon will, undesirably, also be able to explain phenomena not modelled, such as differing identity. For this reason, we use the Mahalanobis distance in the reference cluster's distribution, when computing the illumination subspace correction for each novel frame. This way, prior knowledge from the reference video sequence is used to constrain the expected variation of face patterns in the given illumination conditions. We found that the use of Mahalanobis distance, as opposed to the usual Euclidean distance, in this case achieved better explanation of novel images when the person's identity was the same, and worse when it was different.

Formally, given a reference pose cluster $\{\mathbf{x}_R\}$ and an input frame $\mathbf{x}$, with the same pose, the proposed illumination normalization of $\mathbf{x}$ can be described by the following minimization problem:

$$\mathbf{a}^* = \arg\min_{\mathbf{a}} (\langle \mathbf{x}_R \rangle - \mathbf{x} - \mathbf{B}_I \mathbf{a})^T \mathbf{B}_R \mathbf{C}_R^{-1} \mathbf{B}_R^T (\langle \mathbf{x}_R \rangle - \mathbf{x} - \mathbf{B}_I \mathbf{a}) \tag{4}$$

$$\mathbf{x}^* = \mathbf{x} + \mathbf{B}_I \mathbf{a}^* \tag{5}$$

where $\mathbf{B}_I$ is the pose illumination subspace, $\mathbf{C}_R$ and $\mathbf{B}_R$ the diagonalized covariance matrix and the principal components of $\{\mathbf{x}_R\}$, and $\mathbf{x}^*$ the illumination normalized $\mathbf{x}$. This quadratic minimization problem is solved by differentiation and the minimum is achieved for:
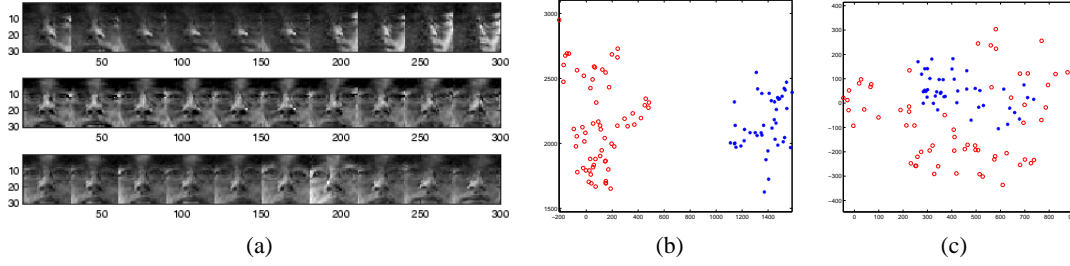
(a)   (b)   (c)

Figure 4: In (a) are respectively shown the original registered and cropped face images from a sequence and the same images after normalization to best match the illumination conditions of the third video sequence. The effects of strong side lighting can be seen to have been removed. Frames from two videos belonging to the same person, before illumination compensation (b), and after the blue one has been re-illuminated (c). Shown are the projections to the first two principal components. Notice that initially the clusters were completely non-overlapping. Illumination normalization has adjusted the location of centre of the blue cluster, but has also contracted it. Now while overlapping, the two sets of patterns are distributed differently.

$$\mathbf{a}^* = \left(\mathbf{B}_I^T \mathbf{B}_R \mathbf{C}_R^{-1} \mathbf{B}_R^T \mathbf{B}_I\right)^{-1} \mathbf{B}_I^T \mathbf{B}_R \mathbf{C}_R^{-1} \mathbf{B}_R^T \left(\langle \mathbf{x}_R \rangle - \mathbf{x}\right) \tag{6}$$

Examples of registered and cropped face images before and after illumination normalization can be seen in Figure 4(a).

Practical considerations. Computation of the optimal value $\mathbf{a}^*$ (6) involves inversion and PCA computation on matrices of size $N \times N$, where N is the number of pixels in a face image. These are both expensive computations. To reduce the computational overhead, we exploit the fact that data modelled is of much lower dimensionality than N. In our implementation of the proposed method, we first perform PCA dimensionality reduction of all pose data by projecting all faces to a *face pose subspace* that explains 95% of face data variation in a specific pose. To additionally speed up the process, we assume that the intrinsic dimensionality of a single pose cluster is 6 (95% of cluster data variability) and that all other variation is due to isotropic Gaussian noise. As finding the largest eigenvalues of a covariance matrix can be done rapidly (e.g. see [2]), we find the 6 largest ones ($\lambda_{1..6}$), with the associated eigenvectors ($\mathbf{v}_{1..6}$) and estimate the rest:

$$\mathbf{C}_R = \mathrm{diag}(\lambda_1 \ldots \lambda_6 \overbrace{\lambda \ldots \lambda}^{N-6}) \tag{7}$$

$$\lambda = \frac{\sum_{i=1}^{6} \lambda_i}{19 \cdot (N-6)} \tag{8}$$

$$\{\mathbf{v}_{7..N}\} = \mathrm{Null}\left(\{\mathbf{v}_{1..6}\}\right) \tag{9}$$

3.4 Comparing Face Distributions

Having normalized one face cluster with respect to illumination, we want to compare it with a corresponding cluster from a different video sequence, under the same lighting. To appreciate the effects that the proposed method of compensating for illumination changes has refer to Figure 4(b,c). An important observation is that the spread of the cluster that is being normalized is reduced. This is the consequence of performing the normalization frame-by-frame, trying to make each as close as possible to the other cluster's centre - a single point. For this reason, 'distance' measures that compare two clusters as pattern distributions, such as the Bhattacharyya distance, the Kullback-Leibler divergence [14] or the Resistor-Average distance [1, 9], are not good choices. As the measure of distance between two clusters we use the simple Euclidean distance between their centres.
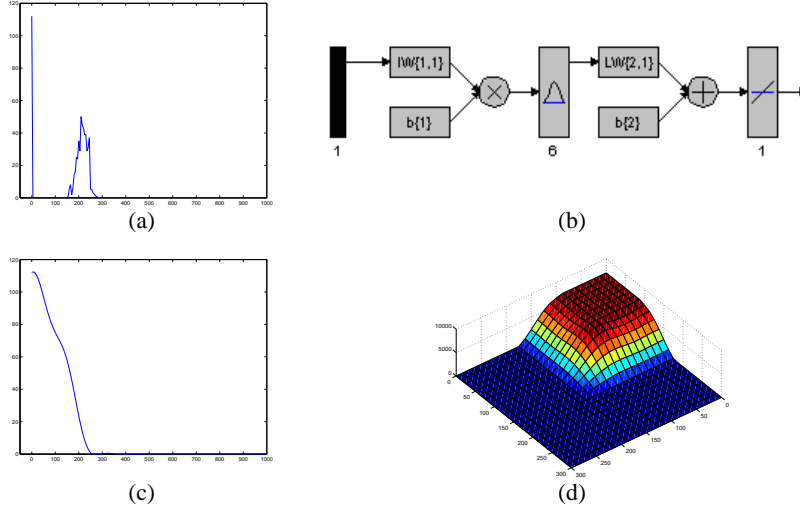
Figure 5: Likelihood ratio corresponding to frontal head pose obtained from the training corpus (a), the RBF network architecture used to interpolate the likelihood ratio (b), the RBF interpolated likelihood ration (c), and the joint interpolated likelihood ratio for pose frontal and left (d). Note that the initial estimate (a) is not monotonically decreasing, while (c) and (d) are.

3.4.1 Integrating Distances Between Pose Clusters

Having computed the three distances $D_{1,2,3}$ between corresponding pose clusters of two manifolds, we want to combine them in a probabilistic manner so that the recognition decision can be made. The decision is made based on the likelihood ratio:

$$\mu = \frac{P(D_{1,2,3}|s)}{P(D_{1,2,3}|\neg s)} \tag{10}$$

where s signifies that the two videos are of the same person. Therefore, we need an estimate of $P(D_{1,2,3}|s)$ and $P(D_{1,2,3}|\neg s)$. To this end, we make the assumption that $P(D_1)$, $P(D_2)$ and $P(D_3)$ are statistically independent. Hence:

$$P(D_{1,2,3}|s) = \prod P(D_i|s) \tag{11}$$
$$P(D_{1,2,3}|\neg s) = \prod P(D_i|\neg s) \tag{12}$$

We learn $P(D_i|s)$ and $P(D_i|\neg s)$ from a labelled, ground truth corpus, in two stages. First, we obtain a Parzen window estimate of intra- and inter- personal pose distances from a small database of videos of faces under varying illumination conditions, see Figure 4(a). It can be seen that the obtained likelihood ratios $\frac{P(D_i|s)}{P(D_i|\neg s)}$ have the salient features of the sought for distributions, but lack some properties that we expect these distributions to have. In particular, we expect them to be monotonically decreasing. The reason why these initial estimates are not is that in the regions with small density of learning corpus samples, the ratio estimates are undefined.

Approximating likelihood ratios using RBF networks. To overcome the problem of non-monotonically decreasing likelihood estimates, we approximate the desired likelihood ratios by training 2-layer RBF networks using carefully selected points from the initial estimates. The points we use are local peaks and the near-zero values for high distances. We obtained good results using 6 neurons in the second layer, with the spread of 60, using the network architecture in Figure 4(b). The results can be seen in Figure 4(c,d).

Figure 6: Frames from typical input video sequences used for evaluation of methods in this paper. The rightmost frame is shown with automatically detected pupils and nostrils, and the region of the face used for recognition. Notice the presence of cast shadows and very varying illumination conditions (different for each frame).

4 Experimental Evaluation

We performed several experiments for the purpose of evaluating our algorithm and comparing its performance with algorithms in the literature. The data sets used in experiments are described in Section 4.1. Algorithms chosen for comparison are:

- The proposed method,

- KLD-based algorithm of Shakhnarovich *et al.* [14],

- Mutual Subspace Method [8],

- Majority vote using Eigenfaces.

In the KLD-based method 85% of data energy was explained by the principal subspace used. In MSM, the dimensionality of PCA subspaces was set to 9 [8]. The first 3 principal angles were used for recognition, as this produced best results in the literature [8]. In the Eigenfaces method, the 22-dimensional principal subspace used explained 90% of data energy.

4.1 Data

Methods in this paper were evaluated using 10 databases with the same 60 individuals, in different illumination conditions for each database. Learning described in Sections 3.2 and 3.3 is performed on randomly selected 20 individuals and 5 lighting conditions. The other 40 individuals in the remaining 5 illumination conditions were used for testing. We emphasize that this makes all learning described completely unbiased as **the evaluation was performed on unseen faces and illumination conditions.**

We performed 25 recognition tests, using each database for training and testing it against all the others. For each person in a database we collected a data set consisting of 20-100 images of a face in random motion (yaw within approximately $\pm 30°$, see Figure 6), sampled from a video at 10fps. Face images were affine registered (Section 3.1) and automatically cropped at approximately mouth and mid-forehead level. Pupils and nostrils were automatically detected using the algorithm described in [7]. Finally, for computational and memory reasons, images were subsampled to 30×30 pixel grayscale images with pixel values normalized to lie in the range $[0,1]$. See Figures 6 and 7. We emphasize that the whole process is automatic – no human intervention is required at any point.

4.2 Results

Recognition results are summarized in Figures 8 and 9. The proposed method significantly outperformed other methods on each training-testing combination, yielding the average recognition rate of 95%. Inspection of failed recognitions of our method suggests that the main problem was significant user motion to and from the camera. For some of the databases used the dominant light sources were relatively close to the user (from $\approx 0.5m$) which invalidated the implicit assumption that illumination conditions were unchanging within a single video sequence. Some examples of very differently illuminated faces within a single sequence can be seen in Figure 7.
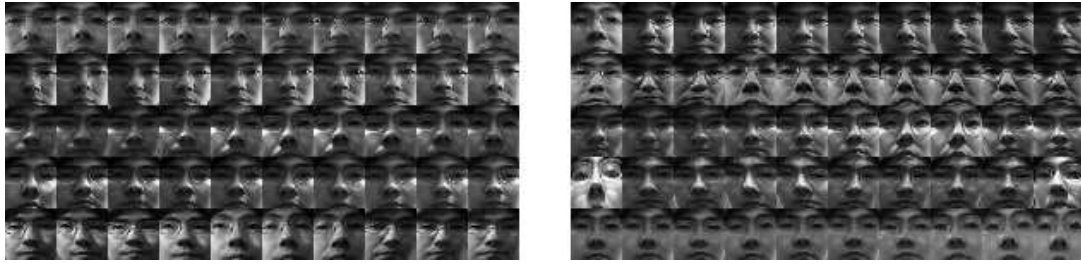
Figure 7: Registered and automatically cropped faces (30×30 pixels) from typical sequences used for the comparison of recognition methods in this paper. All frames are of the same person, in frontal pose, each row corresponding to one of 10 different illumination conditions used for the evaluation. Notice extreme illumination changes.

	Database 1	Database 2	Database 3	Database 4	Database 5	Average	STD
Database 1	100	90	95	95	90	96	4.2
Database 2	95	95	95	95	90	94	2.2
Database 3	95	95	100	95	100	97	2.7
Database 4	95	90	100	100	95	96	4.2
Database 5	100	80	100	95	100	95	8.7

Figure 8: Recognition performance (%) of the proposed method using different training/testing database combinations. Excellent results are demonstrated with little variance of the results with the choice of the training database.

Finally, good separability of intra- and inter- person differences was demonstrated (see Figure 9), thereby showing that the method is suitable for verification purposes. Less than 0.5% of false positive rate is attained for 91% true positive rate.

5 Summary and Conclusions

In this paper we introduced a practical face recognition system from video, robust to illumination changes and pose. In the proposed algorithm recognition is performed by comparing face motion manifolds, described by three Gaussian pose clusters. Compensation for illumination changes is performed on a per-pose basis, first by region-based gamma intensity correction and then using a linear, pose illumination subspace. Normalized pose clusters are compared using the Euclidean distance between their centres. Finally, it is shown how a two-layer RBF network can be trained to estimate the likelihood ratio of two face motion manifolds belonging to the same person. An extensive experimental evaluation and comparison of the proposed method and state-of-the-art algorithms in the literature is presented. It was shown that our method consistently outperformed existing methods, achieving on average recognition rate of 94–100% on 5 databases of 40 people each and extreme lighting variations.

Our future work will concentrate on recognition from face motion manifolds when the extent of

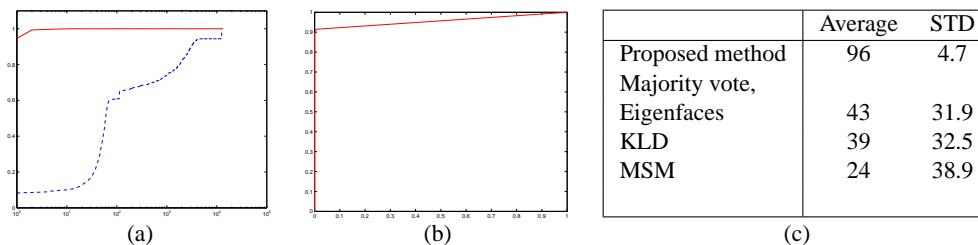|	Average	STD
Proposed method	96	4.7
Majority vote,		
Eigenfaces	43	31.9
KLD	39	32.5
MSM	24	38.9

(a)　　　　　　　　(b)　　　　　　　　(c)

Figure 9: Cumulative distributions of intra-personal (dashed line) and inter-personal (solid line) distances (a). Good separability is demonstrated. The corresponding ROC curve can be seen in (b) – less than 0.5% of false positive rate is attained for 91% true positive rate. Average recognition rates (%) of the compared methods are shown in (c). The performance of the proposed method is by far the best.

motion is even larger and less controlled, as well as other efficient illumination normalization techniques that do not require the assumption of constant illumination for each video sequence.

Acknowledgements

References

[1] O. Arandjelović and R. Cipolla. Face recognition from face motion manifolds using robust kernel resistor-average distance. *IEEE Workshop on Face Processing in Video*, 5:88, June 2004.

[2] J. Baglama, D. Calvetti, and L. Reichel. Iterative methods for the computation of a few eigenvalues of a large symmetric matrix. *BIT*, 36(3):400–440, 1996.

[3] W. A. Barrett. A survey of face recognition algorithms and testing results. *Systems and Computers*, 1:301–305, 1998.

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

[5] R. Chellappa, C. L. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5):705–740, 1995.

[6] T. Fromherz, P. Stucki, and M. Bichsel. A survey of face recognition. *MML Technical Report.*, (97.01), 1997.

[7] K. Fukui and O. Yamaguchi. Facial feature point extraction method based on combination of shape extraction and pattern matching. *Systems and Computers in Japan*, 29(6):2170–2177, 1998.

[8] K. Fukui and O. Yamaguchi. Face recognition using multi-viewpoint patterns for robot vision. *Int'l Symp. of Robotics Research*, 2003.

[9] D. H. Johnson and S. Sinanović. Symmetrizing the Kullback-Leibler distance. *Technical report, Rice University*, 2001.

[10] B. Kepenekci. *Face Recognition Using Gabor Wavelet Transform.* PhD thesis, The Middle East Technical University, 2001.

[11] K. Lee, M. Yang, and D. Kriegman. Video-based face recognition using probabilistic appearance manifolds. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2003.

[12] B. Moghaddam, W. Wahid, and A. Pentland. Beyond eigenfaces - probabilistic matching for face recognition. *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 30–35, 1998.

[13] S. Romdhani, V. Blanz, and T. Vetter. Face identification by fitting a 3D morphable model using linear shape and texture error functions. *In Proc. IEEE European Conference on Computer Vision*, pages 3–19, 2002.

[14] G. Shakhnarovich, J. W. Fisher, and T. Darrel. Face recognition from long-term observations. *In Proc. IEEE European Conference on Computer Vision*, pages 851–868, 2002.

[15] S. Shan, W. Gao, B. Cao, and D. Zhao. Illumination normalization for robust face recognition against varying lighting conditions. *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pages 157–164, 2003.

[16] L. Wolf and A. Shashua. Learning over sets using kernel principal angles. *Journal of Machine Learning Research*, 4(10):913–931, 2003.

[17] W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips. Face recognition: A literature survey. *UMD CFAR Tech. Report CAR-TR-948*, 2000.

[18] S. Zhou, V. Krueger, and R. Chellappa. Probabilistic recognition of human faces from video. *Computer Vision and Image Understanding*, 91(1):214–245, 2003.