Chapter 1

Achieving Illumination Invariance using Image Filters

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

1 Introduction

In this chapter we are interested in accurately recognizing human faces in the presence of large and unpredictable illumination changes. Our aim is to do this in a setup realistic for most practical applications, that is, without overly constraining the conditions in which image data is acquired. Specifically, this means that people's motion and head poses are largely uncontrolled, the amount of available training data is limited to a single short sequence per person, and image quality is low.

In conditions such as these, invariance to changing lighting is perhaps the most significant practical challenge for face recognition algorithms. The illumination setup in which recognition is performed is in most cases impractical to control, its physics difficult to accurately model and face appearance differences due to changing illumination are often larger than those differences between individuals [1]. Additionally, the nature of most real-world applications is such that prompt, often real-time system response is needed, demanding appropriately efficient as well as robust matching algorithms.

In this chapter we describe a novel framework for rapid recognition under varying illumination, based on simple image filtering techniques. The framework is very general and we demonstrate that it offers a dramatic performance improvement when used with a wide range of filters and different baseline matching algorithms, without sacrificing their computational efficiency.

1.1 Previous work and its limitations

The choice of representation, that is, the model used to describe a person's face is central to the problem of automatic face recognition. Consider the components of a generic face recognition system schematically shown in Figure 1.

A number of approaches in the literature use relatively complex facial and scene models that explicitly separate extrinsic and intrinsic variables which affect appearance. In most cases, the complexity of these models makes it impossible to compute model parameters as a
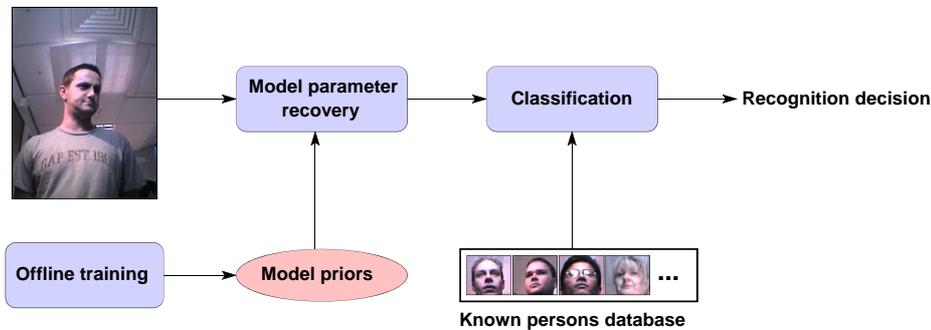
1

Figure 1: *A diagram of the main components of a generic face recognition system. The "Model parameter recovery" and "Classification" stages can be seen as mutually complementary: (i) a complex model that explicitly separates extrinsic and intrinsic appearance variables places most of the workload on the former stage, while the classification of the representation becomes straightforward; in contrast, (ii) simplistic models have to resort to more statistically sophisticated approaches to matching.*

closed-form expression (*"Model parameter recovery"* in Figure 1). Rather, model fitting is performed through an iterative optimization scheme. In the *3D Morphable Model* of Blanz and Vetter [7], for example, the shape and texture of a novel face are recovered through gradient descent by minimizing the discrepancy between the observed and predicted appearance. Similarly, in *Elastic Bunch Graph Matching* [8, 23], gradient descent is used to recover the placements of fiducial features, corresponding to bunch graph nodes and the locations of local texture descriptors. In contrast, the *Generic Shape-Illumination Manifold* method uses a genetic algorithm to perform a manifold-to-manifold mapping that preserves pose.

One of the main limitations of this group of methods arises due to the existence of local minima, of which there are usually many. The key problem is that if the fitted model parameters correspond to a local minimum, classification is performed not merely on noise-contaminated but rather entirely *incorrect* data. An additional unappealing feature of these methods is that it is also not possible to determine if model fitting failed in such a manner.

The alternative approach is to employ a simple face appearance model and put greater emphasis on the classification stage. This general direction has several advantages which make it attractive from a practical standpoint. Firstly, model parameter estimation can now be performed as a closed-form computation, which is not only more efficient, but also void of the issue of fitting failure such that can happen in an iterative optimization scheme. This allows for more powerful statistical classification, thus clearly separating well understood and explicitly modelled stages in the image formation process, and those that are more easily learnt implicitly from training exemplars. This is the methodology followed in this chapter. The sections that follow describe the method in detail, followed by a report of experimental results.
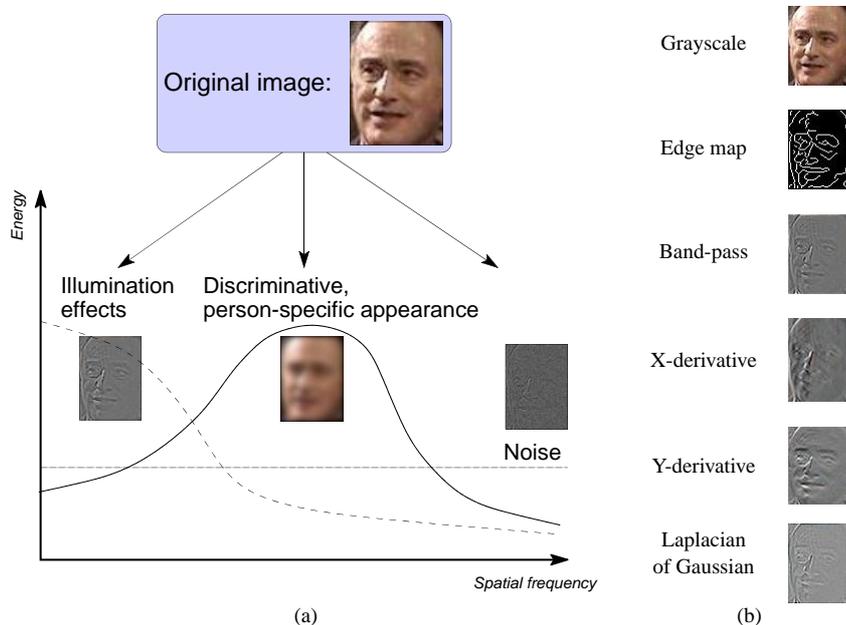
2

Figure 2: *(a) The simplest generative model used for face recognition: images are assumed to consist of the low-frequency band that mainly corresponds to illumination changes, mid-frequency band which contains most of the discriminative, personal information and white noise. (b) The results of several most popular image filters operating under the assumption of the frequency model.*

2 Method details

2.1 Image processing filters

Most relevant to the material presented in this chapter are illumination-normalization methods that can be broadly described as quasi illumination-invariant *image filters*. These include high-pass [5] and locally-scaled high-pass filters [21], directional derivatives [1, 10, 13, 18], Laplacian-of-Gaussian filters [1], region-based gamma intensity correction filters [2, 17] and edge-maps [1], to name a few. These are most commonly based on very simple image formation models, for example modelling illumination as a spatially low-frequency band of the Fourier spectrum and identity-based information as high-frequency [5, 11], see Figure 2. Methods of this group can be applied in a straightforward manner to either single or multiple-image face recognition and are often extremely efficient. However, due to the simplistic nature of the underlying models, in general they do not perform well in the presence of extreme illumination changes.

2.2 Adapting to data acquisition conditions

The framework proposed in this chapter is motivated by our previous research and the findings first published in [3]. Four face recognition algorithms, the *Generic Shape-Illumination* method [3], the *Constrained Mutual Subspace Method* [12], the commercial system *FaceIt* and a *Kullback-Leibler Divergence*-based matching method, were evaluated on a large database using (i) raw greyscale imagery, (ii) high-pass (HP) filtered imagery and (iii) the Self-Quotient Image (QI) representation [21]. Both the high-pass and even further Self Quotient Image representations produced an improvement in recognition for all methods over raw grayscale, as shown in Figure 3, which is consistent with previous findings in the literature [1, 5, 11, 21].
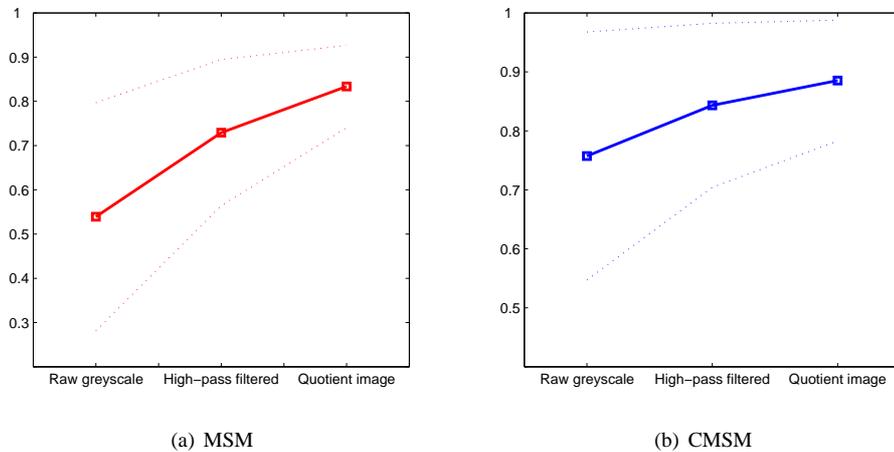

(a) MSM   (b) CMSM

Figure 3: *Performance of the (a) Mutual Subspace Method and the (b) Constrained Mutual Subspace Method using raw greyscale imagery, high-pass (HP) filtered imagery and the Self-Quotient Image (QI), evaluated on over* 1300 *video sequences with extreme illumination, pose and head motion variation (as reported in [3]). Shown are the average performance and* ± *one standard deviation intervals.*

Of importance to this work is that it was also examined in which cases these filters help and how much depending on the data acquisition conditions. It was found that recognition rates using greyscale and either the HP or the QI filter negatively correlated (with $\rho \approx -0.7$), as illustrated in Figure 4. This finding was observed consistently across the result of the four algorithms, all of which employ mutually drastically different underlying models.

This is an interesting result: it means that while on average both representations increase the recognition rate, they actually *worsen* it in "easy" recognition conditions when no normalization is needed. The observed phenomenon is well understood in the context of energy of intrinsic and extrinsic image differences and noise (see [22] for a thorough discussion). Higher than average recognition rates for raw input correspond to small changes in imaging conditions between training and test, and hence lower energy of extrinsic variation. In
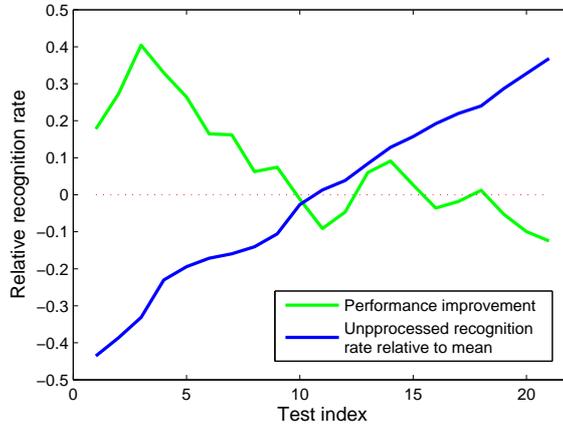
Figure 4: *A plot of the performance improvement with HP and QI filters against the performance of unprocessed, raw imagery across different illumination combinations used in training and test. The tests are shown in the order of increasing raw data performance for easier visualization.*

this case, the two filters decrease the signal-to-noise ratio, worsening the performance, see Figure 5 (a). On the other hand, when the imaging conditions between training and test are very different, normalization of extrinsic variation is the dominant factor and performance is improved, see Figure 5 (b).

This is an important observation: it suggests that the performance of a method that uses either of the representations can be increased further by detecting the difficulty of recognition conditions. In this chapter we propose a novel learning framework to do exactly this.

2.2.1 Adaptive framework

Our goal is to implicitly learn how similar the novel and training (or *gallery*) illumination conditions are, to appropriately emphasize either the raw input guided face comparisons or of its filtered output.

Let $\{\mathcal{X}_1, \ldots, \mathcal{X}_N\}$ be a database of known individuals, $\mathcal{X}$ novel input corresponding to one of the gallery classes and $\rho()$ and $F()$, respectively, a given similarity function and a quasi illumination-invariant filter. We then express the degree of belief η that two face sets $\mathcal{X}$ and $\mathcal{X}_i$ belong to the same person as a weighted combination of similarities between the corresponding unprocessed and filtered image sets:

$$\eta = (1 - \alpha^*)\rho(\mathcal{X}, \mathcal{X}_i) + \alpha^* \rho(F(\mathcal{X}), F(\mathcal{X}_i)) \tag{1}$$

In the light of the previous discussion, we want α^* to be small (closer to 0.0) when novel and the corresponding gallery data have been acquired in similar illuminations, and large

(a) Similar acquisition conditions between sequences

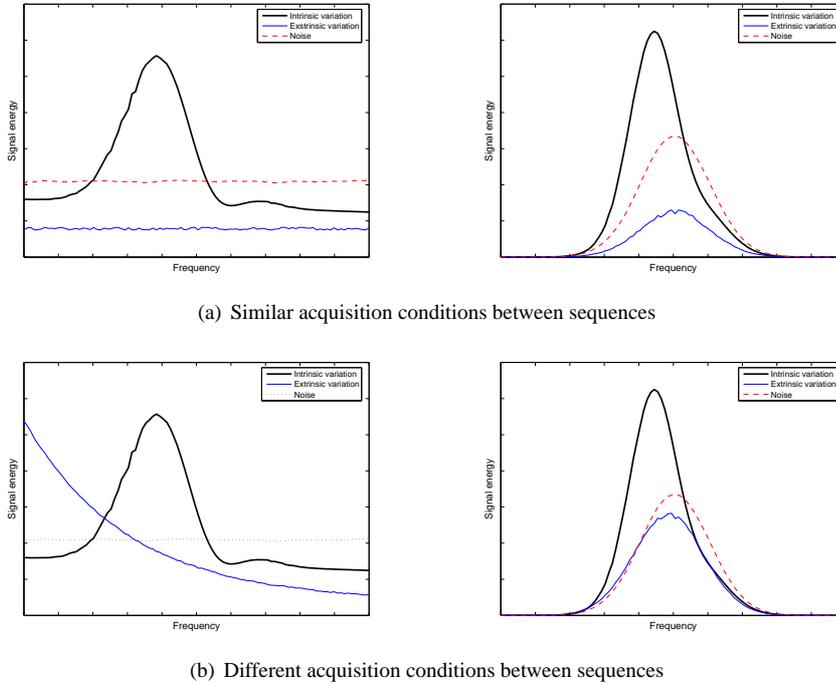(b) Different acquisition conditions between sequences

Figure 5: *A conceptual illustration of the distribution of intrinsic, extrinsic and noise signal energies across frequencies in the cases when training and test data acquisition conditions are (a) similar and (b) different, before (left) and after (right) band-pass filtering.*

(closer to 1.0) when in very different ones. We show that α^* can be learnt as a function:

$$\alpha^* = \alpha^*(\mu), \tag{2}$$

where μ is the *confusion margin* – the difference between the similarities of the two $\mathcal{X}_i$ most similar to $\mathcal{X}$. The value of $\alpha^*(\mu)$ can then be interpreted as statistically the optimal choice of the mixing coefficient α given the confusion margin μ. Formalizing this we can write

$$\alpha^*(\mu) = \arg\max_\alpha p(\alpha|\mu), \tag{3}$$

or, equivalently

$$\alpha^*(\mu) = \arg\max_\alpha \frac{p(\alpha, \mu)}{p(\mu)}. \tag{4}$$

Under the assumption of a uniform prior on the confusion margin, $p(\mu)$

$$p(\alpha|\mu) \propto p(\alpha, \mu), \tag{5}$$

6

and

$$\alpha^*(\mu) = \arg\max_{\alpha} p(\alpha, \mu). \tag{6}$$

2.2.2 Learning the α-function

To learn the α-function $\alpha^*(\mu)$ as defined in (3), we first need an estimate $\hat{p}(\alpha, \mu)$ of the joint probability density $p(\alpha, \mu)$ as per (6). The main difficulty of this problem is of practical nature: in order to obtain an accurate estimate using one of many off-the-shelf density estimation techniques, a prohibitively large training database would be needed to ensure a well sampled distribution of the variable μ. Instead, we propose a heuristic alternative which, we will show, will allow us to do this from a small training corpus of individuals imaged in various illumination conditions. The key idea that makes such a drastic reduction in the amount of training data possible, is to use domain specific knowledge of the properties of $p(\alpha, \mu)$ in the estimation process.

Our algorithm is based on an iterative incremental update of the density, initialized as a uniform density over the domain $\alpha, \mu \in [0, 1]$, see Figure 7. Given a training corpus, we iteratively simulate matching of an "unknown" person against a set of provisional gallery individuals. In each iteration of the algorithm, these are randomly drawn from the offline training database. Since the ground truth identities of all persons in the offline database are known, we can compute the confusion margin $\mu(\alpha)$ for each $\alpha = k\Delta\alpha$, using the inter-personal similarity score defined in (1). Density $\hat{p}(\alpha, \mu)$ is then incremented at each $(k\Delta\alpha, \mu(0))$ proportionally to $\mu(k\Delta\alpha)$ to reflect the goodness of a particular weighting in the simulated recognition.

The proposed offline learning algorithm is summarized in Figure 6 with a typical evolution of $p(\alpha, \mu)$ in Figure 7.

The final stage of the offline learning in our method involves imposing the monotonicity constraint on $\alpha^*(\mu)$ and smoothing of the result, see Figure 8.

3 Empirical evaluation

To test the effectiveness of the described recognition framework, we evaluated its performance on 1662 face motion video sequences from four databases:

Input: training data $D(person, illumination)$,
                filtered data $F(person, illumination)$,
                similarity function ρ,
                filter F.
Output: estimate $\hat{p}(\alpha, \mu)$.

1: Init
   $\hat{p}(\alpha, \mu) = 0,$

2: Iteration
   for all illuminations $i, \ j$ **and** persons p

     3: Initial separation
      $\delta_0 = \min_{q \neq p} \left[\rho(D(p,i), D(q,j)) - \rho(D(p,i), D(p,j)) \right]$

     4: Iteration
      for all $k = 0, \ldots, 1/\Delta\alpha, \ \alpha = k\Delta\alpha$

        5: Separation given α
$$\begin{aligned}
\delta(k\Delta\alpha) = \min_{q \neq p}[&\alpha\rho(F(p,i), F(q,j)) \\
&-\alpha\rho(F(p,i), F(p,j)) \\
&+(1-\alpha)\rho(D(p,i), D(q,j)) \\
&-(1-\alpha)\rho(D(p,i), D(p,j))]
\end{aligned}$$

        6: Update density estimate
         $\hat{p}(k\Delta\alpha, \delta_0) = \hat{p}(k\Delta\alpha, \delta_0) + \delta(k\Delta\alpha)$

7: Smooth the output
   $\hat{p}(\alpha, \mu) = \hat{p}(\alpha, \mu) * \mathbf{G}_{\sigma=0.05}$

8: Normalize to unit integral
   $\hat{p}(\alpha, \mu) = \hat{p}(\alpha, \mu) / \int_\alpha \int_x \hat{p}(\alpha, x) dx d\alpha$

Figure 6: *Offline training algorithm.*

(a) Initialization   (b) Iteration 50   (c) Iteration 100

(d) Iteration 150   (e) Iteration 200   (f) Iteration 250

(g) Iteration 300   (h) Iteration 350   (i) Iteration 400

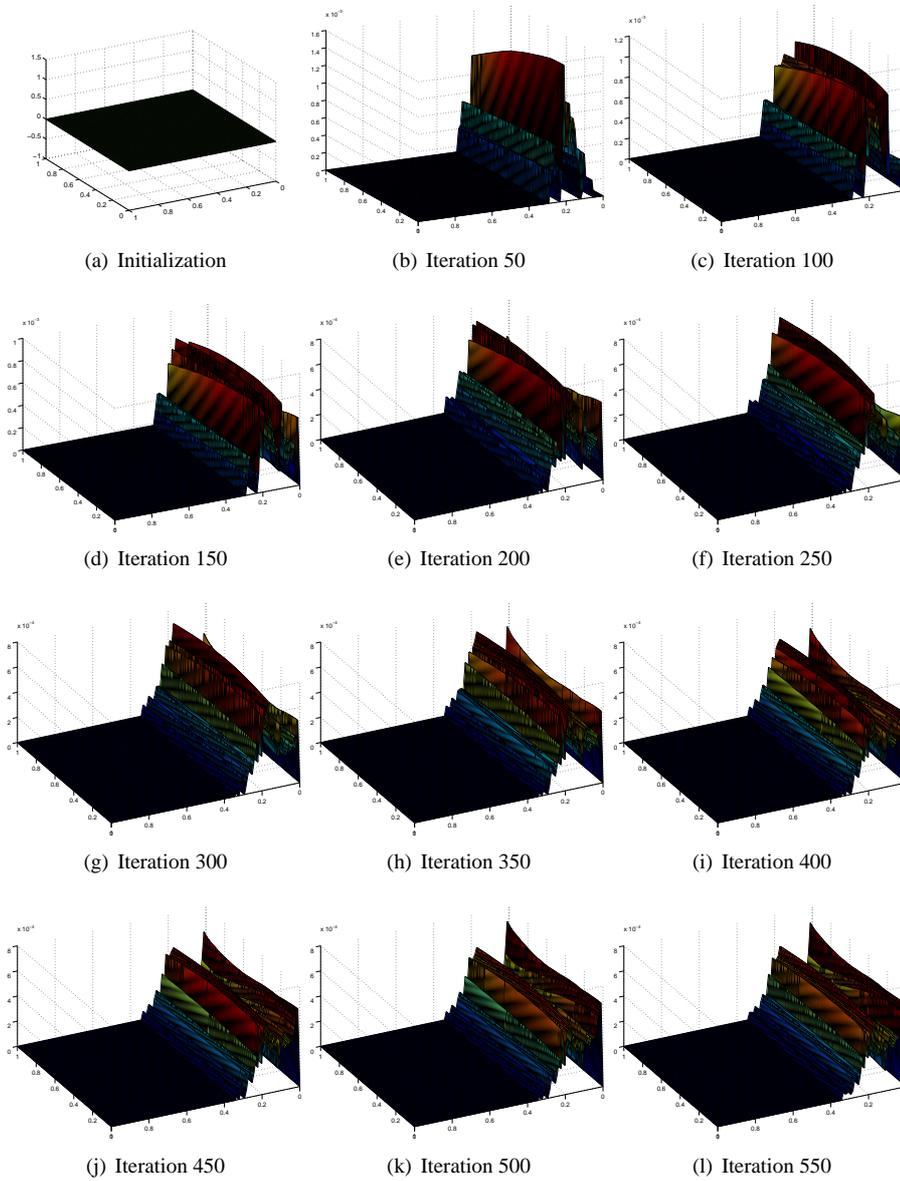(j) Iteration 450   (k) Iteration 500   (l) Iteration 550

Figure 7: *The estimate of the joint density $p(\alpha, \mu)$ through 550 iterations for a band-pass filter used for the evaluation of the proposed framework in Section 3.1.*
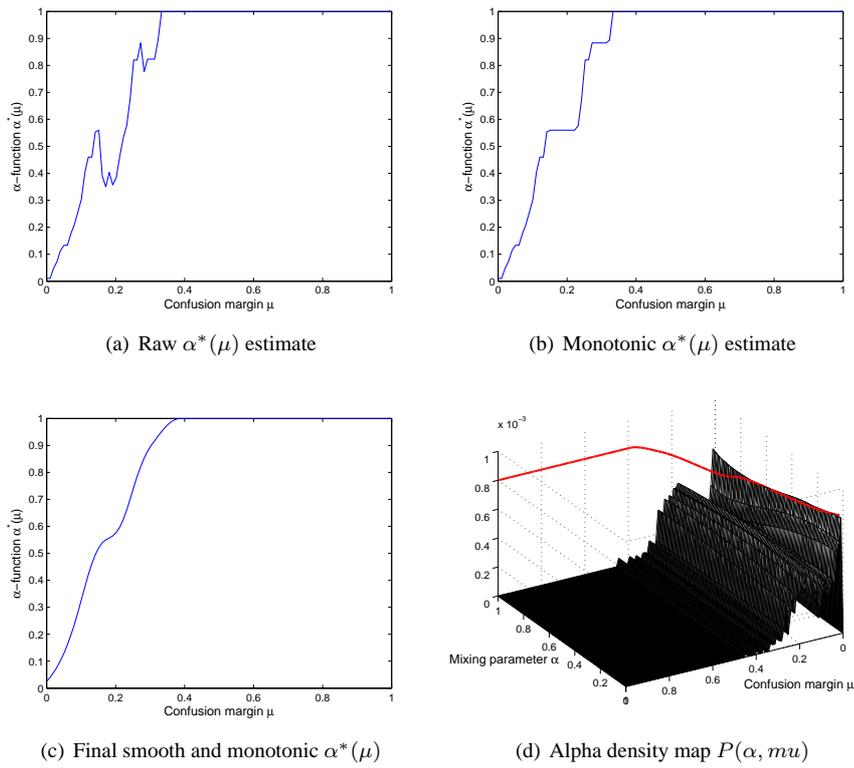
(a) Raw $\alpha^*(\mu)$ estimate

(b) Monotonic $\alpha^*(\mu)$ estimate

(c) Final smooth and monotonic $\alpha^*(\mu)$

(d) Alpha density map $P(\alpha, mu)$

Figure 8: *Typical estimates of the α-function plotted against confusion margin μ. The estimate shown was computed using 40 individuals in 5 illumination conditions for a Gaussian high-pass filter. As expected, α^* assumes low values for small confusion margins and high values for large confusion margins (see (1)).*

CamFace with 100 individuals of varying age and ethnicity, and equally represented genders. For each person in the database we collected 7 video sequences of the person in arbitrary motion (significant translation, yaw and pitch, negligible roll), each in a different illumination setting, see Figure 9 (a) and 10, at 10fps and 320×240 pixel resolution (face size ≈ 60 pixels)[1].

ToshFace kindly provided to us by Toshiba Corp. This database contains 60 individuals of varying age, mostly male Japanese, and 10 sequences per person. Each sequence corresponds to a different illumination setting, at 10fps and 320×240 pixel resolution (face size ≈ 60 pixels), see Figure 9 (b).

Face Video freely available[2] and described in [14]. Briefly, it contains 11 individuals and 2 sequences per person, little variation in illumination, but extreme and uncontrolled variations in pose and motion, acquired at 25fps and 160×120 pixel resolution (face size ≈ 45 pixels), see Figure 9 (c).

Faces96 the most challenging subset of the University of Essex face database, freely available from `http://cswww.essex.ac.uk/mv/allfaces/faces96.html`. It contains 152 individuals, most 18–20 years old and a single 20-frame sequence per person in 196×196 pixel resolution (face size ≈ 80 pixels). The users were asked to approach the camera while performing arbitrary head motion. Although the illumination was kept constant throughout each sequence, there is some variation in the manner in which faces were lit due to the change in the relative position of the user with respect to the lighting sources, see Figure 9 (d).

For each database except *Faces96*, we trained our algorithm using a single sequence per person and tested against a single other sequence per person, acquired in a different session (for *CamFace* and *ToshFace* different sessions correspond to different illumination conditions). Since *Faces96* database contains only a single sequence per person, we used the first frames 1–10 of each for training and frames 11–20 for test. Since each video sequence in this database corresponds to a person walking to the camera, this maximizes the variation in illumination, scale and pose between training and test, thus maximizing the recognition challenge.

Offline training, that is, the estimation of the α-function (see Section 2.2.2) was performed using 40 individuals and 5 illuminations from the *CamFace database*. We emphasize that these were not used as test input for the evaluations reported in the following section.

Data acquisition. The discussion so far focused on recognition using fixed-scale face images. Our system uses a cascaded detector [20] for localization of faces in cluttered images,

---

[1]A thorough description of the University of Cambridge face database with examples of video sequences is available at `http://mi.eng.cam.ac.uk/~oa214/`.

[2]See `http://synapse.vit.iit.nrc.ca/db/video/faces/cvglab`.

(a) Cambridge Face Database

(b) Toshiba Face Database

(c) Face Video Database

(d) Faces 96 Database

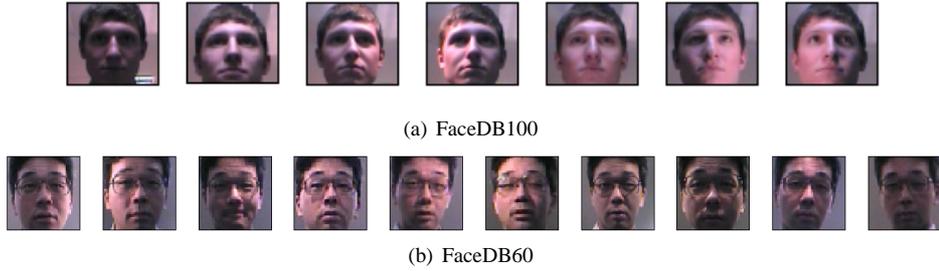Figure 9: *Frames from typical video sequences from the four databases used for evaluation.*

(a) FaceDB100

(b) FaceDB60

Figure 10: *(a) Illuminations 1–7 from database FaceDB100 and (b) illuminations 1–10 from database FaceDB60.*

which are then rescaled to the unform resolution of 50×50 pixels (approximately the average size of detected faces in our data set).

Methods and representations.   The proposed framework was evaluated using the following filters (illustrated in Figure 11):

- Gaussian high-pass filtered images [5, 11] (HP):

$$\mathbf{X}_H = \mathbf{X} - (\mathbf{X} * \mathbf{G}_{\sigma=1.5}), \tag{7}$$

- local intensity-normalized high-pass filtered images – similar to the Self-Quotient Image [21] (QI):

$$\mathbf{X}_Q = \mathbf{X}_H / (\mathbf{X} - \mathbf{X}_H), \tag{8}$$

the division being element-wise,

- distance-transformed edge map [3, 9] (ED):

$$\mathbf{X}_E = \text{DistTrans}(\text{Canny}(\mathbf{X})), \tag{9}$$

- Laplacian-of-Gaussian [1] (LG):

$$\mathbf{X}_L = \mathbf{X} * \nabla \mathbf{G}_{\sigma=3}, \tag{10}$$

and

- directional grey-scale derivatives [1, 10] (DX, DY):

$$\mathbf{X}_x = \mathbf{X} * \frac{\partial}{\partial x} \mathbf{G}_{\sigma_x=6} \tag{11}$$

$$\mathbf{X}_y = \mathbf{X} * \frac{\partial}{\partial y} \mathbf{G}_{\sigma_y=6}. \tag{12}$$
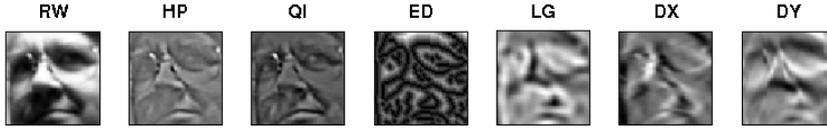
13

Figure 11: *Examples of the evaluated face representations: raw greyscale input (RW), high-pass filtered data (HP), the Quotient Image (QI), distance-transformed edge map (ED), Laplacian-of-Gaussian filtered data (LG) and the two principal axis derivatives (DX and DY).*

For baseline classification, we used two canonical correlations-based [15] methods:

- Constrained MSM (CMSM) [12] used in a state-of-the-art commercial system FacePass[®] [19],

- Mutual Subspace Method (MSM) [12], and

These were chosen as fitting the main premise of the chapter, due to their efficiency, numerical stability and generalization robustness [16]. Specifically, we (i) represent each head motion video sequence as a linear subspace, estimated using PCA from appearance images and (ii) compare two such subspaces by computing the first three canonical correlations between them using the method of Björck and Golub [6], that is, as singular values of the matrix $B_1^T B_2$ where $B_{1,2}$ are orthonormal basis of two linear subspaces.

3.1 Results

To establish baseline performance, we performed recognition with both MSM and CMSM using raw data first. A summary is shown in Table 3.1. As these results illustrate, the *Cam-Face* and *ToshFace* data sets were found to be very challenging, primarily due to extreme variations in illumination. The performance on *Face Video* and *Faces96* databases was significantly better. This can be explained by noting that the first major source of appearance variation present in these sets, the scale, is normalized for in the data extraction stage; the remainder of the appearance variation is dominated by pose changes, to which MSM and CMSM are particularly robust to [4, 16].

Next we evaluated the two methods with each of the 6 filter-based face representations. The recognition results for the *CamFace*, *ToshFace* and *Faces96* databases are shown in blue in Figure 12, while the results on the *Face Video* data set are separately shown in Table 2 for the ease of visualization. Confirming the first premise of this work as well as previous research findings, all of the filters produced an improvement in average recognition rates. Little interaction between method/filter combinations was found, Laplacian-of-Gaussian and the horizontal intensity derivative producing the best results and bringing the best and average recognition errors down to 12% and 9% respectively.
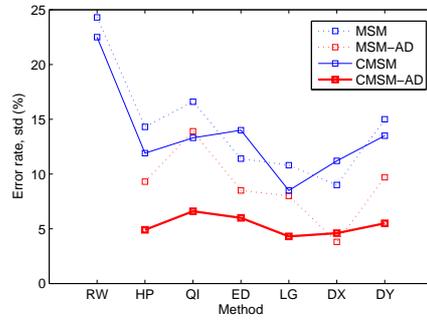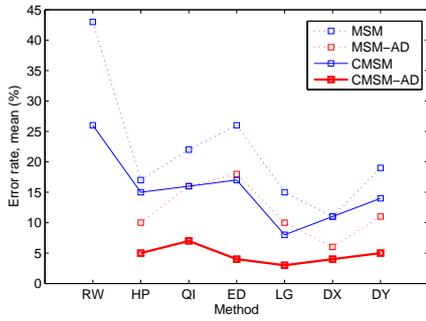
(a) CamFace
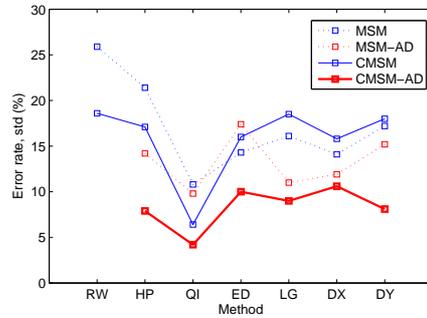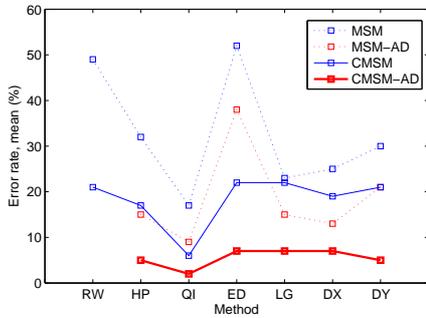
(b) ToshFace
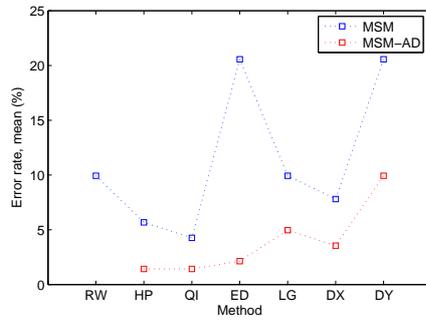
(c) Faces96

Figure 12: *Error rate statistics. The proposed framework (-AD suffix) dramatically improved recognition performance on all method/filter combinations, as witnessed by the reduction in both error rate averages and their standard deviations. The results of CMSM on Faces96 are not shown as it performed perfectly on this data set.*

15

Table 1: *Recognition rates (mean/STD, %).*

	CamFace	**ToshFace**	**FaceVideoDB**	**Faces96**	**Average**
CMSM	73.6 / 22.5	79.3 / 18.6	91.9	100.0	87.8
MSM	58.3 / 24.3	46.6 / 28.3	81.8	90.1	72.7

Table 2: *FaceVideoDB, mean error (%).*

	RW	**HP**	**QI**	**ED**	**LG**	**DX**	**DY**
MSM	0.00	0.00	0.00	0.00	9.09	0.00	0.00
MSM-AD	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CMSM	0.00	9.09	0.00	0.00	0.00	0.00	0.00
CMSM-AD	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Finally, in the last set of experiments, we employed each of the 6 filters in the proposed data-adaptive framework. The recognition results are shown in red in Figure 12 and in Table 2 for the *Face Video* database. The proposed method produced a dramatic performance improvement in the case of all filters, reducing the average recognition error rate to only 3% in the case of CMSM/Laplacian-of-Gaussian combination.This is a very high recognition rate for such unconstrained conditions (see Figure 9), small amount of training data per gallery individual and the degree of illumination, pose and motion pattern variation between different sequences. An improvement in the robustness to illumination changes can also be seen in the significantly reduced standard deviation of the recognition, as shown in Figure 12. Finally, it should be emphasized that the demonstrated improvement is obtained with a negligible increase in the computational cost as all time-demanding learning is performed offline.

4 Conclusions

In this chapter we described a novel framework for automatic face recognition in the presence of varying illumination, primarily applicable to matching face sets or sequences. The framework is based on simple image processing filters that compete with unprocessed greyscale input to yield a single matching score between individuals. By performing all numerically consuming computation offline, our method both (i) retains the matching efficiency of simple image filters, but (ii) with a greatly increased robustness, as all online processing is performed in closed-form. Evaluated on a large, real-world data corpus, the proposed framework was shown to be successful in video-based recognition across a wide range of illumination, pose and face motion pattern changes.

References

[1] Y Adini, Y. Moses, and S. Ullman. Face recognition: The problem of compensating for changes in illumination direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 19(7):721–732, 1997.

[2] O. Arandjelović and R. Cipolla. An illumination invariant face recognition system for access control using video. *In Proc. IAPR British Machine Vision Conference (BMVC)*, pages 537–546, September 2004.

[3] O. Arandjelović and R. Cipolla. Face recognition from video using the generic shape-illumination manifold. *In Proc. European Conference on Computer Vision (ECCV)*, 4:27–40, May 2006.

[4] O. Arandjelović, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell. Face recognition with image sets using manifold density divergence. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1:581–588, June 2005.

[5] O. Arandjelović and A. Zisserman. Automatic face recognition for film character retrieval in feature-length films. *In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1:860–867, June 2005.

[6] Å. Björck and G. H. Golub. Numerical methods for computing angles between linear subspaces. *Mathematics of Computation*, 27(123):579–594, 1973.

[7] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. *In Proc. Conference on Computer Graphics (SIGGRAPH)*, pages 187–194, 1999.

[8] D. S. Bolme. Elastic bunch graph matching. Master's thesis, Colorado State University, 2003.

[9] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 8(6):679–698, 1986.

[10] M. Everingham and A. Zisserman. Automated person identification in video. *In Proc. IEEE International Conference on Image and Video Retrieval (CIVR)*, pages 289–298, 2004.

[11] A. Fitzgibbon and A. Zisserman. On affine invariant clustering and automatic cast listing in movies. *In Proc. European Conference on Computer Vision (ECCV)*, pages 304–320, 2002.

[12] K. Fukui and O. Yamaguchi. Face recognition using multi-viewpoint patterns for robot vision. *International Symposium of Robotics Research*, 2003.

[13] Y. Gao and M. K. H. Leung. Face recognition using line edge map. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(6):764–779, 2002.

[14] D. O. Gorodnichy. Associative neural networks as means for low-resolution video-based recognition. *In Proc. International Joint Conference on Neural Networks*, 2005.

[15] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28:321–372, 1936.

[16] T-K. Kim, O. Arandjelović, and R. Cipolla. Boosted manifold principal angles for image set-based recognition. *Pattern Recognition*, 2006. (to appear).

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

[18] B. Takács. Comparing face images using the modified Hausdorff distance. *Pattern Recognition*, 31(12):1873–1881, 1998.

[19] Toshiba. Facepass. *www.toshiba.co.jp/mmlab/tech/w31e.htm*.

[20] P. Viola and M. Jones. Robust real-time face detection. *International Journal of Computer Vision (IJCV)*, 57(2):137–154, 2004.

[21] H. Wang, S. Z. Li, and Y. Wang. Face recognition under varying lighting conditions using self quotient image. *In Proc. IEEE International Conference on Automatic Face and Gesture Recognition (FGR)*, pages 819–824, 2004.

[22] X. Wang and X. Tang. Unified subspace analysis for face recognition. *In Proc. IEEE International Conference on Computer Vision (ICCV)*, 1:679–686, 2003.

[23] L. Wiskott, J-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, pages 355–396, 1999.