IDIAP RESEARCH REPORT

Efficient Diffusion-based Illumination Normalization for Face Verification

Guillaume Heusch [a] Fabien Cardinaux [a]
Sébastien Marcel [a]

IDIAP–RR 05-46

August 2005

[a] IDIAP Research Institute

IDIAP Research Institute **www.idiap.ch**
Rue du Simplon 4 P.O. Box 592 1920 Martigny – Switzerland
Tel: +41 27 721 77 11 Fax: +41 27 721 77 12 Email: info@idiap.ch

IDIAP Research Report 05-46

Efficient Diffusion-based Illumination Normalization for Face Verification

Guillaume Heusch Fabien Cardinaux Sébastien Marcel

August 2005

Abstract. In this paper, the problem of face verification across illumination is addressed. In order to cope with different lighting conditions, a preprocessing step is applied to the face image so as to make it independent on the illumination conditions. The illuminant invariant representation of the image is obtained by first applying an anisotropic diffusion process to the original image (Gross and Brajovic, 2003). Hence, it implies the numerical resolution of an elliptic partial differential equation on a large grid: the image. So, a comparison is performed on two methods to resolve such diffusion problems, namely the Gauss-Seidel relaxation and the Multigrid V-cycle. The preprocessing algorithm with its different resolution schemes is applied prior to the task of face verification. Experiments conducted on the challenging BANCA database showed a significant improvement in terms of face verification error rate, while staying computationally efficient.

1 Introduction

Illumination has been identified as one of the major factor affecting the performance of face verification systems. Actually, Adini et al. (1997) showed that image variation due to lighting changes is more significant than that due to different personal identities. They also empirically noticed that no classical image features are sufficient to compensate for illumination changes. Later, Chen et al. (2000) formally proved that there are no functions of an image that are discriminative illuminant invariants. Existing algorithms to cope with illumination variations can be divided into two categories. The aim of the first category is to *model* the subspace of the image space described by varying the light source only. For doing so, those methods require training images of the same object, viewed under slightly different lighting conditions. Georghiades et al. (2001) developed a powerful tool to perform face recognition across illumination. Based on their previous result stating that the set of images of an object seen under arbitrary lighting conditions is a low-dimensional subspace of the image space (Belhumeur and Kriegman, 1998), they proposed a method to approximate this subspace by building a so-called *illumination cone*. Recognition is then performed by measuring the distance between the probe image and the pre-stored illumination cone representations. Ho et al. (2003) extended this algorithm by proposing a specific metric to efficiently cluster illumination cones.
There exists several other approaches based on the idea that the illumination subspace can be well approximated by a low-dimensional subspace of the image space. Among them, one can mention the spherical harmonic representation (Basri and Jacobs, 2001), and also the simpler subspace construction using linear combination of training images (Wang et al., 2003). These methods are generally computationally expensive and require training images of the same object under various lighting conditions.
The second approach used to deal with illumination changes is to apply a preprocessing step prior to the face verification task itself. A well-known example of such enhancement algorithm is histogram equalization. The main advantage of preprocessing compared to model-based approaches is that it is completely independent on the classifier, so that no training phase regarding the illumination is required: the goal is to turn the original image into a better suited representation for recognition purpose. Hence, training images of the same object under different lighting conditions are not needed anymore.

For the purpose of face verification, a comparison of photometric normalization algorithms was made by Short et al. (2004) and showed the superiority of the diffusion-based model proposed by Gross and Brajovic (2003). Their study was perform using a face verification system based on linear discriminant analysis (LDA), which belongs to so-called holistic classifiers: the whole face image is used as raw input to the system. On the other hand, feature-based classifiers first derives local features as representation of the face region before proceeding with the classification. Compared to holistic approaches, feature-based methods are generally more robust to change in illumination and to imperfect face localization (Zhao et al., 2003).
We thus propose to study the most promising normalization technique on a system based on local features and Gaussian Mixture Models (GMM) (Cardinaux et al., 2003, 2004). Moreover, no practical implementation of the diffusion-based algorithm (Gross and Brajovic, 2003) was proposed and since it involves the resolution of partial differential equations, there exist numerous methods to numerically resolve such problems. Hence, we will investigate two different resolution schemes, in terms of both face verification performance and computational efficiency. Performances of the preprocessing algorithms are assessed on the BANCA database (Bailly-Baillire et al., 2003).

This paper is organized as follows: the next section describes the algorithm proposed by Gross and Brajovic (2003). Section 3 details the different numerical resolution schemes, from the relaxation to the multigrid V-cycle and section 4 presents the face verification system. Section 5 describes the conducted experiments as well as the obtained results. Finally, a conclusion is drawn in section 6.

2 Illumination Normalization

By conducting visual experiments, Land and McCann (1971) derived a computational framework of how the eye perceive the reflectance of a scene: the reflectance value of a particular point $R(x, y)$, can be obtained by comparing the point's intensity with many other ones in the scene.
A digital image $I(x, y)$ can be regarded as the product between the illuminance $L(x, y)$ and the reflectance $R(x, y)$ (Land and McCann, 1971), hence an illumination free representation of the image can be obtained by dividing the original image by the illuminance term:

$$I(x, y) = R(x, y) \cdot L(x, y) \quad \Rightarrow \quad R(x, y) = \frac{I(x, y)}{L(x, y)} \tag{1}$$

The difficulty lies in estimating the illuminance term $L(x, y)$. A practical implementation of this so-called Retinex algorithm is due to Jobson et al. (1997): the illuminance field is obtained by applying a Gaussian filter to the original image. Hence, the illuminance can be estimated as a *smooth version* of the original image.

Gross & Brajovic Algorithm

Based on the previous result that the illuminance can be derived as a smooth version of the original image, Gross and Brajovic (2003) proposed to estimate $L(x, y)$ by applying an anisotropic diffusion process to the input image. Unlike isotropic diffusion (which is equivalent to gaussian filtering), meaningful informations in the image, represented by edges, are not discarded. The diffusion process is implemented by minimizing the following variational model:

$$E(L) = \iint_\Omega \rho(x, y)(L(x, y) - I(x, y))^2 dxdy + \lambda \iint_\Omega (L_x^2 + L_y^2) dxdy \tag{2}$$

The first term forces the luminance function to be close to the original image and the second term adds a smoothness constraint on $L(x, y)$. λ controls the relative importance of this constraint (e.g. the amount of blurring introduced in the image) and $\rho(x, y)$ are the anisotropic diffusion coefficients. Note that if $\rho(x, y) = 1 \quad \forall x, y$ then isotropic diffusion is performed. The solution for $L(x, y)$ is found thanks to the discrete version of the Euler-Lagrange equation associated to equation (2):

$$\begin{aligned} L_{i,j} \quad & +\lambda \Big[\tfrac{1}{\rho_{i,j-1}}(L_{i,j} - L_{i,j-1}) + \tfrac{1}{\rho_{i,j+1}}(L_{i,j} - L_{i,j+1}) \\ & + \tfrac{1}{\rho_{i-1,j}}(L_{i,j} - L_{i-1,j}) + \tfrac{1}{\rho_{i+1,j}}(L_{i,j} - L_{i+1,j}) \Big] \quad = I_{i,j} \end{aligned} \tag{3}$$

where the anisotropic diffusion coefficients $\rho_{i,j}$ are defined as the reciprocal of the Weber contrast measure: $\rho_{i,j} = |I_i - I_j| / \min(I_i, I_j)$. Equation (3) actually represents a boundary value problem, and can be viewed as a large sparse linear system: $\mathbf{A} \cdot L = I$, where $\mathbf{A}$ is called the *differential operator* (Press et al., 1988).

3 Numerical Resolution

Since the matrix $\mathbf{A}$ is of size NxN, N being the number of pixel in the image, it cannot be directly inverted without huge computational load and memory requirements. In this section, we present two methods to approximate the solution of large sparse linear systems induced by the discretization of elliptic partial differential equations, namely the relaxation and the multigrid V-cycle.

3.1 Relaxation

Consider the linear problem: $Ax = b$ and suppose that we have an arbitrary initial estimate of the solution x^0. This linear iterative method recursively corrects the initial estimate according to:

$$x^{k+1} = x^k + B(b - Ax^k) \tag{4}$$

where B will be an approximate of A^{-1}. Note that if $B = A^{-1}$ only one iteration is sufficient to resolve the problem. A popular choice for B is the so-called Gauss-Seidel relaxation: let us write $A = D - L - U$ where D is the diagonal of A, $-L$ and $-U$ are the lower and the upper part of A respectively. In Gauss-Seidel relaxation, B is defined as $B = (D - L)^{-1}$. Consequently, the result of the iterative process (4) depends on the ordering of processed elements. In our experiments, the chequer-board ordering was used (Figure 1) since it is best suited for typical second-order elliptic equations (Press et al., 1988). The Gauss-Seidel relaxation then becomes a two-stage process: first, odd points are updated (equation (5)) and then even points (equation (6)).

$$a_{i,i}x_{i,i}^{k+1} = b_{i,i} - \left(a_{i-1,i}x_{i-1,i}^{k} + a_{i,i-1}x_{i,i-1}^{k} + a_{i,i+1}x_{i,i+1}^{k} + a_{i+1,i}x_{i+1,i}^{k}\right) \tag{5}$$

$$a_{i,i}x_{i,i}^{k+1} = b_{i,i} - \left(a_{i-1,i}x_{i-1,i}^{k+1} + a_{i,i-1}x_{i,i-1}^{k+1} + a_{i,i+1}x_{i,i+1}^{k+1} + a_{i+1,i}x_{i+1,i}^{k+1}\right) \tag{6}$$

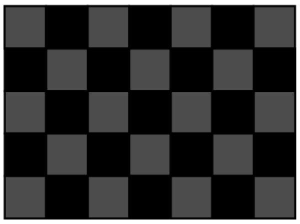

Figure 1: Checker-board ordering of the Gauss-Seidel relaxation: all the gray points are processed before updating the black ones.

3.2 Towards Multigrid

3.2.1 Coarse Grid Correction

Although the convergence of the Gauss-Seidel relaxation is guaranteed, the major drawback of this procedure is that the convergence speed depends on the accuracy of the initial estimate. So the goal of the coarse-grid procedure is to improve an initial guess of the solution. Let us define x_0^i as the initial estimate on the finest grid 0, we also introduce the residual, defined as $r_0^i = A_0x_0^i - b_0$, which is a measure of the error between the estimate and the solution. In order to improve the initial estimate x_0^i, the residual is projected onto a coarser, thus smaller grid 1. On this grid, the correction c_1^i can be computed thanks to the *residual equation* $A_1c_1^i = r_1^i$, solved by direct resolution. The correction corresponds to the quantity to add to the estimate so as to yield the true solution on grid 1. Since the solution has to be found on the finest grid, the correction c_1^i is then projected back to the fine grid and finally added to the initial estimate: $x_0^i \leftarrow x_0^i + c_0^i$. In order to go from one grid to another, transfer functions have to be defined: as suggested in Press et al. (1988), the *full-weighting* scheme for the fine-to-coarse function ($\mathcal{R}$) and the bilinear interpolation for the coarse-to-fine function ($\mathcal{P}$) were used. The coarse-grid correction is summarized below:

1. Compute the residual on the fine grid: $r_0^i = A_0x_0^i - b_0$
2. Restrict the residual to the coarse grid: $r_1^i = \mathcal{R}r_0^i$

3. Solve the residual equation: $A_1 c_1^i = -r_1^i$ (direct resolution)
4. Interpolate the correction c_1^i to the fine grid: $c_0^i = \mathcal{P} c_1^i$
5. Add the correction to the estimate: $x_0^i \leftarrow x_0^i + c_0^i$

3.2.2 V-cycle

The multigrid V-cycle algorithm is a particular implementation of the multigrid framework, which is known to be an optimal solver for large and sparse linear systems induced by partial differential equations (Briggs et al., 2000). Considering the coarse grid correction, the residual equation is assumed to be solved by direct matrix inversion. Although this resolution is faster since the grid is smaller, the computational load may remain important. The multigrid V-cycle algorithm goes one step further: the idea is to solve the residual equation by applying, once again, another coarse-grid correction. This process is then repeated until the coarser grid is small enough, for the residual equation to be solved without significant computational load. The complete V-cycle algorithm is summarized below (Briggs et al., 2000):

1. Apply $\nu_1 > 0$ relaxation steps to the initial estimate: $\hat{x}_{l_0}$
2. Compute the residual on the fine grid: $r_{l_0} = A_{l_0} \hat{x}_{l_0} - b_{l_0}$
3. Restrict the residual to the next coarser grid: $r_{l_1} = \mathcal{R} r_{l_0}$
4. Solve $A_{l_1} c_{l_1} = -r_{l_1}$ by a **recursive call** to the V-cycle algorithm
   - until the grid is small enough to perform efficient matrix inversion
5. Interpolate the coarse grid solution c_{l_1} to the next finer grid: $c_{l_0} = \mathcal{P} c_{l_1}$
6. Add the correction to the estimate: $\hat{x}_{l_0} \leftarrow \hat{x}_{l_0} + c_{l_0}$
7. Apply $\nu_2 > 0$ relaxation steps to the corrected estimate: $\hat{x}_{l_0}$

4 Face Verification

In order to compare the performance of the diffusion-based preprocessing algorithms and their different numerical resolution schemes, we investigated the effect of the lighting normalization on the popular task of face verification. Face verification consists in giving an opinion, based on a face image, on whether an individual claiming its identity is the *true claimant* or an *impostor*. In our experiments, this decision is made based on a system using local features and Gaussian Mixtures Models (GMM) (Cardinaux et al., 2003, 2004).

Let us consider a client C described by its parameters θ_C, and a generic face model $\overline{C}$ with parameters $\theta_{\overline{C}}$. The generic face model is also known as the *world model* and is used to simulate an arbitrary impostor. The decision on whether the client is the true claimant or not is given by:

$$\Theta(X) = \log P(X|\theta_C) - \log P(X|\theta_{\overline{C}}) \tag{7}$$

where $P(X|\theta_C)$ denotes the likelihood that the set of features vectors comes from the true claimant and $P(X|\theta_{\overline{C}})$ is the likelihood that the set of feature vectors comes from an impostor. These likelihoods are estimated with:

$$P(X|\theta) = \prod_{t=1}^{N_v} P(\vec{x}_t|\theta) \tag{8}$$

where the probability distribution $P(\vec{x}_t|\theta)$ is modelled as a sum of gaussian distributions. In order to derive the client-specific parameter set θ_C, the observation X is derived from the face image on a block-by-block basis. Each block is decomposed using 2D-DCT basis functions, and the final feature vector for a block is built according to the DCT$mod2$ scheme (Sanderson and Paliwal, 2002).

5 Experiments and Results

5.1 Database and Protocols

The challenging BANCA database (Bailly-Baillire et al., 2003), especially designed for biometric authentication, is used to assess the performances of the normalization algorithms. In our experiments, we consider the English corpus of the database, containing 52 individuals (26 males and 26 females), equally divided into two groups used for hyper-parameter selection ($g1$) and test ($g2$). Recordings were made according to three distinct scenarios: controlled, degraded and adverse.

(a) (b) (c)

Figure 2: The different scenarios: controlled (a), degraded (b) and adverse (c).

The BANCA protocol also defines various training and testing policies. Here we focus on four particular conditions, that is: Matched controlled (Mc), Unmatched adverse (Ua), Unmatched degraded (Ud) and the Pooled test (P). In each protocols, the training phase is performed using images taken from the controlled conditions. The system is then tested on images acquired under various conditions, depending on the protocol: controlled (for Mc), adverse (for Ua), degraded (for Ud) and the pooled test (P) does the test on images from the three different conditions (see Figure 2). This procedure makes sense since it corresponds to the realistic and challenging scenario of capturing images for the gallery set in good conditions, whereas probe images are more likely to be acquired in various conditions.

Before being processed by the classifier (for both training and testing), faces windows are manually located, resized to 64x80 pixels, converted to grayscale and finally normalized with respect to illumination conditions. Moreover, histogram equalization is applied on the normalized image. Accuracy of the verification system is measured using the Half Total Error Rate (HTER), which combine the False Acceptance Rate (FAR) and the False Rejection Rate (FRR):

$$HTER = \frac{FAR + FRR}{2} \tag{9}$$

Since the FAR and the FRR are related to each other (reducing the FAR will increase the FRR and vice versa). The HTER is computed at the Equal Error Rate (EER), where FAR = FRR.

5.2 Results

Table 1 shows the accuracy of the face verification systems in terms of Half Total Error Rate (HTER) on the test group $g2$. For comparison purposes, results obtained with no preprocessing, simple histogram equalization and isotropic diffusion (i.e. $\rho(x, y) = 1 \quad \forall\, x, y$) are also presented. Resolution schemes are denoted by *relax* for the Gauss-Seidel relaxation and by *v-cycle* for the Multigrid V-cycle resolution. For each normalization algorithms, the hyper-parameter λ remains the same throughout the different protocols. It was empirically selected, using the set $g1$ so as to minimize HTER on the pooled test (P), since it include images in all available conditions (controlled, degraded and adverse).

Normalization algorithm	Mc [%]	Ua [%]	Ud [%]	P [%]
None	3.686	20.192	28.205	24.733
Histogram equalization	7.372	20.032	22.917	19.818
Isotropic diffusion: *relax*	4.808	20.192	**11.538**	14.850
Isotropic diffusion: *v-cycle*	**1.923**	16.186	11.699	15.224
Anisotropic diffusion: *relax*	7.692	18.109	14.583	19.391
Anisotropic diffusion: *v-cycle*	4.487	**15.705**	12.019	**12.981**
Comparison				
P2D-HMM Cardinaux et al. (2004)	3.4	12.7	15.4	16.4
LDA + Weber diffusion Short et al. (2004)	4.2	8.6	20.7	13.4

Table 1: Verification Performances on the BANCA database.

In the match controlled protocol (Mc), training and testing images come from the controlled scenario and most of the normalization algorithms do not improve the verification rate. This result is not surprising since there are no changes in terms of illumination conditions between the gallery and the probe images.
The unmatched protocols (Ua and Ud) allow us to verify that the various algorithms and proposed resolution schemes are indeed efficient to tackle illumination changes. Actually, with these protocols, every algorithms improves the accuracy of the system. The best results in the adverse conditions (Ua), achieved with anisotropic diffusion resolved by multigrid V-cycle reduces the error rate from 20.2% to 15.7%. Moreover, in the degraded conditions (Ud), the error rate decreases from 28.2% to 11.5% thanks to the isotropic diffusion solved by relaxation. The pooled test P indicates which strategy performs the best on all conditions: probe images may be captured from controlled, adverse or degraded conditions. In our case, the anisotropic diffusion-based normalization with the V-cycle resolution obtains the lowest error rate. Concerning the resolution scheme, it is better for both investigated algorithms to use the multigrid V-cycle, since it gives, in almost every configuration, the best verification rate.
Note that thanks to this preprocessing step, the lowest HTER can be compared to state-of-the-art algorithms on the same database. Actually, applying the lighting normalization scheme prior to GMM-based verification yields better results than more sophisticated HMM-based approaches reported in Cardinaux et al. (2004). Results are also comparable with those obtained in Short et al. (2004), where the anisotropic diffusion-based normalization is applied prior to a LDA-based classifier.

Since face verification may be a time critical task, it is also interesting to consider the computational load of each normalization approach. The most efficient is, as one could have expected, the multigrid V-cycle: it takes 40 ms to process a face window (64x80 pixels) using Weber contrast coefficients on a P4 running at 3.2 Ghz. Relaxation is more dependent on its intrinsic parameter: the number of relaxation sweeps. However, the computational load of the relaxation-based resolution of the isotropic diffusion remains acceptable for real-time application (200ms for 20 steps). On the other hand, relaxation applied with anisotropic coefficients is the most computationally expensive.

6 Conclusion

In this article, we proposed two distinct numerical resolution schemes for one of the state-of-the-art preprocessing algorithms for illumination invariant face recognition (Gross and Brajovic, 2003). We also investigated the effect of the different resolution schemes on the same problem but with isotropic coefficients. Experiments conducted on the BANCA database showed significant improvement in terms of verification rate. The best results are obtained with the combination Weber diffusion and

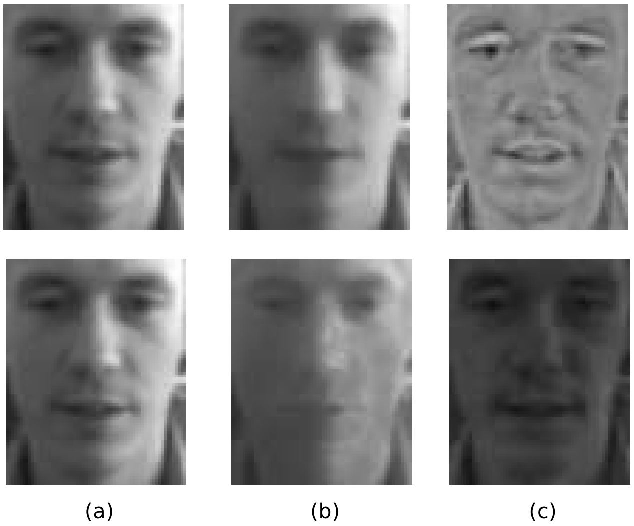

Figure 3: Results of the anisotropic diffusion-based normalization algorithm: (a) input face image (b) estimated luminance field (c) estimated reflectance. The first line show results obtained with relaxation whereas images of the second line are obtained with the multigrid V-cycle resolution

V-cycle resolution: the error rate is reduced by a factor of 34% compared to simple histogram equalization. The error rate on the realistic Pooled Test P is comparable with other state-of-the-art face verification systems on the same database. Moreover, the multigrid V-cycle implementation allows the preprocessing to be performed in real time since a face is processed in 40ms only.

7 Acknowledgments

This work was performed within the AMI Training Programme (*www.amiproject.org*). All the algorithms described in this article were implemented within the framework of the Torch machine learning library (Collobert et al., 2002) and its vision package[1], and are freely available to the research community for benchmarking purposes, evaluations and improvements.

References

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, volume 19, Issue 7, 1997.

E. Bailly-Baillire *et al.*: *The BANCA Database and Evaluation Protocol*, International Conference on Audio- and Video-Based Biometric Person Authentication, 2003.

P. Belhumeur and D. Kriegman: *What Is The Set of Images of an Object under All Possible Illumination Conditions*, International Journal of Computer Vision, volume 28, Issue 3, 1998.

R. Basri and D. Jacobs: *Lambertian Reflectance and Linear Subspaces*, in IEEE International Conference on Computer Vision, 2001.

W. L. Briggs, V. E. Henson and S.F. McCormick: *A Multigrid Tutorial*, SIAM publications, 2000.

[1] www.idiap.ch/∼marcel/en/torch3/torch3vision.php

F.Cardinaux, C. Sanderson and S. Marcel: *Comparison of MLP and GMM Classifiers for Face Verification on XM2VTS*, International Conference on Audio- and Video-Based Biometric Person Authentication, 2003.

F. Cardinaux, C. Sanderson and S. Bengio: *Face Verification Using Adapted Generative Models*, in IEEE International Conference on Automatic Face and Gesture Recognition, 2004.

H. Chen, P. Belhumeur and D. Jacobs: *In search of Illumination Invariants*, IEEE International Conference on Computer Vision and Pattern Recognition, 2000.

R. Collobert, S. Bengio, and J. Mariéthoz. Torch: a modular machine learning software library. Technical Report IDIAP-RR 02-46, IDIAP, 2002.

A. Georghiades, P. Belhumeur and D. Kriegman: *From Few to Many: Illumination Cone Models for Face Recognition Under Variable Lighting and Pose*, IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 23, Issue 6, 2001.

R. Gross and V. Brajovic: *An Image Preprocessing Algorithm for Illumination Invariant Face Recognition*, International Conference on Audio- and Video-Based Biometric Person Authentication, 2003.

J. Ho, M-H Yang, J. Lim, K-C Lee and D. Kriegman: *Clustering Appearances of Objects Under Varying Illumination Conditions*, IEEE International Conference on Computer Vision and Pattern Recognition, 2003.

D. Jobson, Z. Rahmann and G. Woodell: *A Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observations of Scenes*, IEEE Transanctions on Image Processing, volume 6, Issue 7, 1997.

E.H. Land and J.J. McCann: *Lightness and Retinex Theory*, Journal of the Optical Society of America , volume 61, n1, 1971.

W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling *Numerical Recipes in C*, Cambridge University Press, 1988.

C. Sanderson and K. K. Paliwal: *Polynomial Features for Robust Face Authentication*. IEEE International Conference on Image Processing, 2002.

J. Short, J. Kittler and K. Messer: *A Comparison of Photometric Normalisation Algorithms for Face Verification*, IEEE International Conference on Automatic Face and Gesture Recognition, 2004.

H. Wang, S. Li and Y. Wang and W. Zhang: *Illumination Modeling and Normalization for Face Recognition*, in International Workshop on Analysis and Modeling of Face and Gesture, 2003.

W. Zhao, R. Chellapa, P.J. Phillips and A. Rosenfeld: *Face Recognition: A Literature Survey*, ACM Computing Surveys (CSUR), Volume 35 , Issue 4, 2003.