

Combining Illumination Normalization Methods for Better Face Recognition

Bas Boom, Qian Tao, Luuk Spreeuwers, and Raymond Veldhuis

University of Twente, EEMSC, Signals & Systems,
P.O. box 217, 7500 AE, Enschede, The Netherlands
{b.j.boom,q.tao,l.j.spreuwers,r.n.j.veldhuis}@utwente.nl

Abstract. Face Recognition under uncontrolled illumination conditions is partly an unsolved problem. There are two categories of illumination normalization methods. The first category performs a local preprocessing, where they correct a pixel value based on a local neighborhood in the images. The second category performs a global preprocessing step, where the illumination conditions and the face shape of the entire image are estimated. We use two illumination normalization methods from both categories, namely Local Binary Patterns and Model-based Face Illumination Correction. The preprocessed face images of both methods are individually classified with a face recognition algorithm which gives us two similarity scores for a face image. We combine the similarity scores using score-level fusion, decision-level fusion and hybrid fusion. In our previous work, we show that combining the similarity score of different methods using fusion can improve the performance of biometric systems. We achieved a significant performance improvement in comparison with the individual methods.

1 Introduction

One of the major problems with face recognition under uncontrolled conditions is the illumination variation, which is often larger than the variations between individuals. Using illumination normalization methods, we want to correct the illumination variations in a single face image. In literature, several methods have been proposed to make face images invariant for illumination. These methods can be divided into two categories. The first category normalizes the face image by applying a preprocessing step on the pixel values using information from the local region around that pixel. Examples of these approaches are Histogram Equalization [1] or (Simplified) Local Binary Patterns [2], [3]. These approaches are direct and simple, but fail to model the global illumination conditions. The second category estimates a global physical model of the illumination mechanism and its interaction with the facial surface. In this category falls for instance the Quotient Image [4], Spherical harmonics [5], 3D morphable models [6]. These methods are able to estimate the global illumination condition, but are also more complicated and require training to model the illumination conditions.

In practise, we have observed that both categories of illumination normalization algorithms have their advantages and disadvantages. The methods in the

first category have problems with regions which are not illuminated because of hard shadows. These shadow regions have a large signal to noise ratio which makes the correction for local methods almost impossible. However, these local methods work well on the illuminated parts of the image and on lightly shadowed (soft shadows) areas. The second category is able to reconstruct the parts with hard shadows, using statistical models. But our current implementation of a global method does not model face variations like glasses and expressions. To summarize, the local methods work well on illuminated part of the image, also the parts which are not modelled by the global methods. The global methods are however able to reconstruct parts which contain hard shadows, which is not possible using a local method. By combining methods from both categories, we aim to improve the performance in face recognition under different and uncontrolled illumination conditions.

Combining the two different illumination correction methods can be done on three different levels, namely at the feature level, the score level and the decision level. We will concentrate on the last two levels of fusion because of its simplicity. To achieve this, the preprocessed images are individually classified and the scores are fused using score-level [7], decision-level fusion [8] and a combination of both these methods named hybrid fusion [9].

Our paper is organized as follows, in Section 2 we describe the two illumination correction algorithms. Section 3 explains how we combine these algorithms for face recognition. In Section 4, we show the experiments and results and Section 5 gives the conclusions.

2 Illumination Normalization

For illumination normalization, we use two methods which come from different categories. The method from the first category is the Local Binary Patterns [10] preprocessing, where different papers [2], [3] claim its invariance to illumination conditions. In the second category, we use the illumination correction approach in [11], which is able to correct illumination variations using a single 2D facial image as its input. If we compare this method with [4] or [12], it is more advanced using a 3D shape model and a shadow and reflection model, but in comparison with 3D morphable models [6] it is still computational efficient. In the following subsections, we describe both methods in more detail.

2.1 Local Binary Patterns

Local Binary Patterns method (LBP) is proposed in [10] and are often used as features in face recognition. The standard LBP give the 3×3 -neighbors the value 0 if they are smaller than the center pixel value and 1 otherwise. This result in a 8 bit string, which represents the pattern at the center point. We can also obtain from it a decimal representation between 0 and 255. LBP is a relative measure which makes it invariant against monotonic gray-scale transformations. A big range of illumination changes consist of monotonic gray-scale transformations in large regions in the image.

There are also extensions to the original LBP, which use a bigger radius and more spacing for the neighbors or use a different weighting scheme for the different bits. In this paper, we use the simplest LBP as preprocessing to obtain the illumination invariant images, i.e. based on a 3×3 -neighborhood.

2.2 Model-Based Face Illumination Correction

The method from the second category [11] uses the Lambertian model to correct illumination in face images. It assumes that a face image is illuminated from a single light source at an infinite distance, which gives the following Lambertian equation:

$$b(x) = c(x)\mathbf{n}(x)^T \mathbf{s}i + e(x; \mathbf{s}) \quad (1)$$

In this case, x is the pixel position and $b \in \mathcal{R}$ is the pixel intensity in the image. The pixel intensity is determined by the shape and the light condition. The shape $\mathbf{h}(x) = c(x)\mathbf{n}(x)^T$ is defined by the surface normals $\mathbf{n} \in \mathcal{R}^3$ and the albedo $c \in \mathcal{R}$. The light condition $\mathbf{v} = \mathbf{s}i$ is defined by the normalized vector $\mathbf{s} \in \mathcal{R}^3$ which denotes the light direction and the intensity of the light given by $i \in \mathcal{R}$. We added the error term $e \in \mathcal{R}$, which allows us to handle shadows and reflections, which are not modelled in the Lambertian model. We can also vectorize the face image which gives the following notation: $\mathbf{b} = H\mathbf{v} + \mathbf{e}(\mathbf{s})$. Using an image of M pixels, we have a vectorized face image, $\mathbf{b} \in \mathcal{R}^{M \times 1}$ and a matrix $H \in \mathcal{R}^{M \times 3}$ which contains the face shape and the error term $\mathbf{e}(\mathbf{s}) \in \mathcal{R}^{M \times 1}$. The goal of this method is to correct the illumination in a face image $b(x)$, by estimating the shape $\mathbf{h}(x)$ and light \mathbf{v} . This approach uses a face shape model, so that we are able to estimate a face shape given that we know the light direction. We can also evaluate how well the face shape fit under a certain light direction. Because the light direction is unknown, we calculate multiple face shapes for a set of light direction and evaluate these shapes. Using kernel regression, a sort of weighted interpolation using the evaluations of the face shapes, we can calculate a final shape using the best evaluated shapes. The pseudo-code of this approach is given below:

- Learn a model of the face shape and the error term (offline)
- For a set of different light direction \mathbf{s}_j
 - Calculate a shadow and reflection term $e_j(x; \mathbf{s}_j)$
 - Estimate the light intensity i_j which gives us light condition \mathbf{v}_j
 - Fit the face shape model to the face image which gives us a shape $\mathbf{h}_j(x)$
 - Evaluate the shape $\mathbf{h}_j(x)$ which gives us a distance measure d_j
- Calculate the final shape $\mathbf{h}(x)$ using kernel regression
- Refine the albedo of the final shape $c(x)$ to obtain more details

In this paper, j is the index for a set of different light directions, where we calculate for J light directions error terms $e_j(x; \mathbf{s}_j)$, light conditions \mathbf{v}_j , face shapes $\mathbf{h}_j(x)$ and evaluations of the face shapes. The final shape is used to calculate frontal illumination on the face, giving us an illumination normalized face image. In the next sections, we describe the different steps of the pseudo-code.

Learn the Face Shape and the Error Model. A set of face images and 3D range maps allows us to calculate the face shapes H using a image and a range map (see [11]). Using multiple face shapes, we can determine with Principal Component Analysis (PCA) a mean shape \bar{H} and k variations from this mean shape denoted by $T_k \in \mathcal{R}^{M \times 3}$. This allows us to describe the face shape H as follows $\bar{H} + \sum_{k=1}^K T_k y_k$.

The Lambertian model is unable to deal with shadows and reflections. For this reason, we added the error term $e(x; \mathbf{s})$, which depends on the light direction. We use a 2D face database with labelled light direction to learn the error term for each light directions. We determine the face shape under the frontal illumination conditions. Next, we determine for the same face under different illumination conditions the error with respect to the Lambertian model. We calculate for every pixel position a mean and variance of the error using this dataset. This gives us a face independent error term to model shadows and reflections.

Estimate and Evaluate the Face Shape given a Light Direction. Given the light direction \mathbf{s}_j , we estimate and evaluate a face shape using the face image and our face shape model. We obtain the mean error $e(x; \mathbf{s})$ as described in the previous section using kernel regression for a given light direction (see [12]). We still have to estimate the light intensity i_j and the face shape H_j . By replacing the unknown face shape H_j by the mean shape \bar{H} , we estimate the light intensity i_j using the following equation:

$$i_j = \arg \min_{i_j} \|(\bar{H} \mathbf{s}_j) i_j - (\mathbf{b} - \mathbf{e}_j(\mathbf{s}_j))\|^2 \quad (2)$$

Because this is a overcomplete system, we can use the mean face shape \bar{H} to estimate the light intensity i_j , which still gives a very accurate estimation. This might however normalize the skincolor, because the mean shape contains the mean albedo instead of the user-specific skincolor. The light conditions are now given by $\mathbf{v} = \mathbf{s} i$, so we only have to obtain the face shape H_j . To obtain the face shape H , we substitute $\bar{H} + \sum_{k=1}^K T_k y_k$ into the Lambertian equation as follows:

$$\bar{H} \mathbf{v}_j + \sum_{k=1}^K T_k \mathbf{v}_j y_{j,k} = \mathbf{b} - \mathbf{e}_j(\mathbf{s}_j) \Rightarrow \sum_{k=1}^K T_k \mathbf{v}_j y_{j,k} = \mathbf{b} - \mathbf{e}_j(\mathbf{s}_j) - \bar{H} \mathbf{v}_j \quad (3)$$

We write $T_k \mathbf{v}_j = A_k$ and $\mathbf{b} - \mathbf{e}_j(\mathbf{s}_j) - \bar{H} \mathbf{v}_j = \mathbf{c}$, which allows us to calculate the variations \mathbf{y}_j :

$$\mathbf{y}_j = \arg \min_{\mathbf{y}_j} \|\mathbf{A} \mathbf{y}_j - \mathbf{c}\|^2 \quad (4)$$

To calculate the shape H_j from the parameters \mathbf{y}_j , we use model obtain in section 2.2. This shape allows us to calculate a reconstructed image $\mathbf{b}_j = H_j \mathbf{v}_j + \mathbf{e}_j(\mathbf{s}_j)$. To evaluate the face shape, we use two conditions: first the found variations \mathbf{y}_j to the mean shape should be small and second the reconstructed image \mathbf{b}_j should be similar to the real image \mathbf{b} . This gives the following distance measure:

$$d_j = \sum_{k=1}^K \frac{y_k^2}{\lambda_k} + \frac{(\mathbf{b} - \mathbf{b}_j)^T (\mathbf{b} - \mathbf{b}_j)}{\rho} \quad (5)$$

The calculation for variations which are not model by the shape model (ρ) is described in [11]. Using this distance measure d_j , we can easily evaluate the quality of the found shape for a certain light direction.

Calculate Final Shape Using Kernel Regression. In the previous section, we explained how we obtain a face shape using only the face image \mathbf{b} and the light direction \mathbf{s}_j . We perform this procedure for a set of light directions, where we use a similar grid of light directions as is used in the Yale B database. We calculated the face shape parameters $\{\mathbf{y}_j\}_{j=1}^J$ for each light direction \mathbf{s}_j . We evaluate the shapes which gives us the distance measures $\{d_j\}_{j=1}^J$. Using kernel regression on the obtain parameters \mathbf{y}_j , we calculate the final face shape parameters:

$$\mathbf{y} = \sum_{j=1}^J w_j \mathbf{y}_j / \left(\sum_{j=1}^J w_j \right) \quad w_j = \exp\left[-\frac{1}{2}(d_j/\sigma)^2\right] \quad (6)$$

In the above equation, σ is determined such that 5 percent of the distances lie within $1 \times \sigma$. Using these final face shape parameters, we can calculate the final face shape and estimate the light conditions in the image.

Refinement of Albedo of the Final Shape. The final face shape does not contain all the details present in the original image, which can be explained by the fact that we perform dimension reduction and kernel regression. To recover these details, we recalculate the albedo $c(x)$ of the shape, because the albedo usually contains the details, while the surface normals contain the larger structures of the shape. The albedo can be calculated using a MAP estimate given by the following equation:

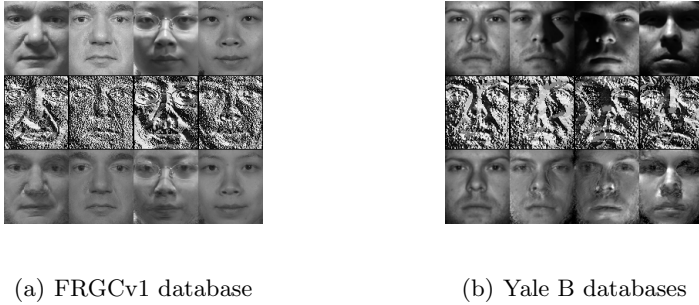
$$c(x)_{MAP} = \arg \max_c P(b(x)|c(x))P(c(x)) \quad (7)$$

$$c(x)_{MAP} = \arg \max_{c(x)} \mathcal{N}(b(x) - (c(x)\mathbf{n}(x))^T \mathbf{s}_i + \mu_e(x, \mathbf{s}), \sigma_e^2(x, \mathbf{s})) \times \mathcal{N}(\mu_c(x), \sigma_c^2(x)) \quad (8)$$

We assume that the probabilities are normal distributions (\mathcal{N}). The mean (μ_e) and variance (σ_e) of the error term are calculated 2D face database with labelled light direction and the mean (μ_c) and variance (σ_c) of the albedo are determined from the set of the face images and 3D range maps used to learn the face shape. We find the minimum by taking the derivative and set it equal to zero. The new albedo c_{MAP} contains more details than the albedo in the final shape.

3 Fusion to Improve Recognition

In the previous section, we proposed two methods to obtain illumination invariant face images. The resulting images of these methods can be seen in Figure 1.



(a) FRGCv1 database

(b) Yale B databases

Fig. 1. Face Images from the Yale B database and FRGCv1 database, upper without correction, middle preprocessed with LBP, lower preprocessed with Model-based Illumination Correction

In this section, we combine these methods to improve the face recognition under different illumination conditions. We train two face classifiers, one with the LBP and one with Model-based Face Illumination correction. The face classifier (log-likelihood ratio after a feature reduction using a PCA and LDA transformation) gives us a similarity score, which we are able to fuse. For fusion, we use the following methods: SUM rule score-level fusion [7], OR rule decision-level fusion [8] and hybrid fusion [9].

In the case of score-level fusion, we can take the joint likelihood ratio, which in our case means that we can sum the scores obtained from the log-likelihood ratios. This gives us the advantage that we do not have to estimate the different density functions or perform a normalization step to the similarity scores. This method of fusion, we denote as SUM rule fusion.

Although theoretically, score-level fusion should achieve the optimal performance, it is not very robust to outliers. For this reason, we also use decision-level fusion with the OR rule to combine the receiver operation characteristics (ROC). The ROC is determined from the similarity scores of the face classifiers and can be obtained by varying the threshold, thus producing a different false reject rate β and false accept rate $\alpha = 1 - p_r$. This specific pair (α, β) is called an operation point, which corresponds to a threshold t in the similarity scores. In the case of fusion, there can be N classifiers and each is characterized by its ROC, $p_{r,i}(\beta_i)$, $i = 1, \dots, N$. By assuming that our classifiers are independent, the final performance of the OR rule can be estimated as $\beta = \prod_{i=1}^N \beta_i$ and $p_r(\beta) = \prod_{i=1}^N p_{r,i}(\beta_i)$. By searching for the optimal operation points, the fusion with the OR rule can be formulated as:

$$\hat{p}_r(\beta) = \max_{\beta_i | \prod_{i=1}^N \beta_i = \beta} \left\{ \prod_{i=1}^N p_{r,i}(\beta_i) \right\} \quad (9)$$

We can prove that the estimated $\hat{p}_r(\beta)$ is never smaller than any of the components $p_{r,i}(\beta)$, $i = 1, \dots, N$ at the same β . Because we do not have the ROC $\hat{p}_r(\beta)$

in analytical form, we estimate a ROC from evaluation data. The ROC $\hat{p}_r(\beta)$ is therefor characterized by discrete values and can be solved numerically [8].

The hybrid fusion is a combination between score-level fusion and decision-level fusion. In hybrid fusion, we first perform the SUM rule score-level fusion to combine the ROCs of both classifiers. The ROCs of both classifiers together with ROC given by the SUM rule are then fused using OR rule decision-level fusion. Using hybrid fusion, we hope to combine the advantages of score-level fusion and the decision-level fusion.

4 Experiments and Results

The purpose of the illumination corrections is to improve the verification rates in face recognition. We performed a recognition experiment on the Yale B databases (the Yale B [13] and extended Yale B [14] database), which contain face images under different labelled illumination conditions created in a laboratory. This experiment tests the ability to make illumination invariant images under all kinds of illumination conditions (also with hard shadows). We also perform an experiment on the Face Recognition Grand Challenge version 1 (FRGCv1) database [15] which contain face images taken under controlled and uncontrolled conditions. This experiment allows us test the ability to correct the uncontrolled illumination conditions and compare them to the controlled images in the gallery.

The Yale B Databases are created to model and test the effects of illumination on face images. In our correction algorithm, we use the Yale B databases to obtain our error model for shadows and reflections. Because we trained our error model on the Yale B databases, we performed a leave one person out experiment for the Model-based Illumination Correction.

In our face recognition experiment, we correct all the face images with both correction methods. We use only the face images with a azimuth and elevation angle below ± 90 degrees in this experiment. For face recognition, we trained on the face images of thirty persons and performed a one-to-one verification experiment on the remaining eight persons, leaving out the face images taken under similar illumination conditions. We repeated this experiment until we compared the face images of all person in the Yale B databases with each other. To train the fusion methods, we used the scores from the Yale B database (10 persons). For testing, we used the scores of Extended Yale B database (29 persons). In Figure 2, we show all the results on the Extended Yale B database.

The experiment which we perform on the Yale B databases is a difficult experiment, because sometimes two images illuminated from opposite positions are compared. We observe from Figure 2(b) that the Model-based Illumination Correction works better at a FAR < 25% than the Local Binary Patterns. By fusing the two methods, we can improve the recognition results significantly. In Figure 2(a), we observe that a diagonal line is probably the best separation between imposter scores and genuine scores. This explains why the SUM rule performs slightly better than the OR rule and hybrid fusion (see Figure 2(b)).

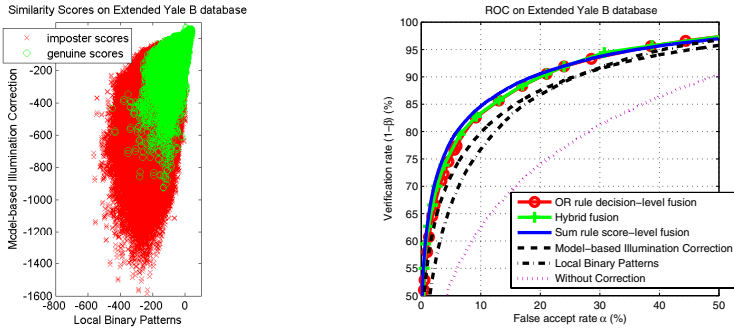


Fig. 2. Score Plot and ROC of the two illumination correction methods on the Extended Yale B database, ROCs also contains the fusion results and the results without illumination correction

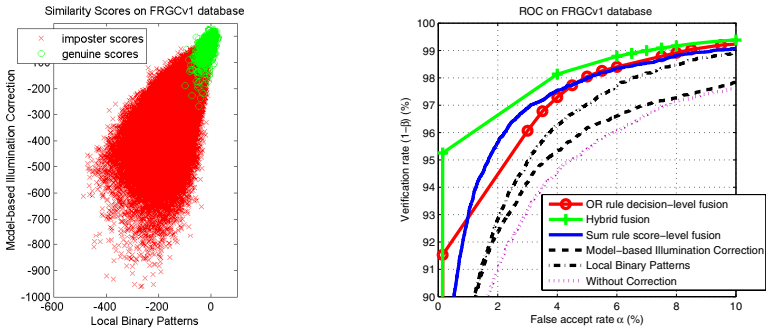


Fig. 3. Score Plot and ROC of the two illumination correction methods on the FRGCv1 database, ROCs also contains the fusion results and the results without illumination correction

The FRGCv1 Database contains frontal face images taken under both controlled and uncontrolled conditions. In our experiment, we randomly divided the uncontrolled and controlled face images into two parts, each containing approximately half of the face images. We used the first halves of both sets to train our face classifiers and the fusion methods, the second half of the controlled images are used for the enrollment of the one user template for every person and the second half of the uncontrolled images are used as probe images. We repeat this experiment 20 times using different random splits of the database to become invariant for statistical fluctuations. This experiment simulates a video surveillance scenario, where we usually have a gallery of high quality images, but the probe images are obtained under uncontrolled conditions. Both our illumination correction algorithms preprocess all the images, also the controlled images. The recognition results are shown in Figure 3.

Although, we have in this experiment less extreme illumination conditions, there are also other challenges in the FRGCv1 database beside illumination, like expressions and out of focus images. From Figure 3(b), we observe that the Local Binary Patterns work better on this database than the Model-based Illumination Correction. The main reason for this difference is that Model-based Illumination Correction has a larger amount of outliers, due to glasses and expressions (see also Figure 3(a) where relatively many genuine scores (circles) have larger negative values for the Model-based Illumination Correction). Using the simple SUM rule to combine the face classifiers already improves the overall recognition results. In figure 3(b), we observe that the recognition results of the OR rule are similar to the sum rule, and the hybrid fusion clearly outperforms the other fusion methods on this database.

5 Conclusions

We combine two different methods to correct for illumination in face images and obtain better results in face recognition. We show that both methods are able to correct for illumination in face images. The Local Binary Patterns method corrects pixel values based on the local neighborhood in the image. This method shows good results on uncontrolled images, but cannot recover large regions with hard shadows. The Model-based Illumination Correction shows that it can deal with these shadow regions, but it has problems in uncontrolled conditions which contain unmodelled effects, like glasses and expressions. Because both methods have different strengths and weaknesses, we combine the illumination normalization methods using fusion. We use three different fusion methods: SUM rule score-level fusion, OR rule decision-level fusion and hybrid fusion. The performance of the simple SUM rule fusion already improves the results significantly and works best for large variations in illumination. The performance of the OR-rule is in both experiments slightly worse than the SUM rule. The hybrid fusion, which tries to combine the advantages of both fusion methods, gives the largest improvement in performance when we correct for uncontrolled illumination conditions occurring in a video surveillance environment.

References

1. Shan, S., Gao, W., Cao, B., Zhao, D.: Illumination normalization for robust face recognition against varying lighting conditions. In: IEEE International Workshop on Analysis and Modeling of Faces and Gestures, AMFG 2003, October 17, 2003, pp. 157–164 (2003)
2. Heusch, G., Rodriguez, Y., Marcel, S.: Local binary patterns as an image preprocessing for face authentication. In: 7th International Conference on Automatic Face and Gesture Recognition, FGR 2006, April 10–12, pp. 6–14 (2006)
3. Tao, Q., Veldhuis, R.N.J.: Illumination normalization based on simplified local binary patterns for a face verification system. In: Biometrics Symposium 2007 at The Biometrics Consortium Conference, Baltimore, Maryland, USA, pp. 1–6. IEEE Computational Intelligence Society, Los Alamitos (2007)

4. Shashua, A., Riklin-Raviv, T.: The quotient image: class-based re-rendering and recognition with varying illuminations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(2), 129–139 (2001)
5. Basri, R., Jacobs, D.: Lambertian reflectance and linear subspaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(2), 218–233 (2003)
6. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In: *SIGGRAPH 1999: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pp. 187–194 (1999)
7. Dass, S.C., Nandakumar, K., Jain, A.K.: A principled approach to score level fusion in multimodal. In: Kanade, T., Jain, A., Ratha, N.K. (eds.) *AVBPA 2005*. LNCS, vol. 3546, pp. 1049–1058. Springer, Heidelberg (2005)
8. Tao, Q., Veldhuis, R.N.J.: Optimal decision fusion for a face verification system. In: Lee, S.-W., Li, S.Z. (eds.) *ICB 2007*. LNCS, vol. 4642, pp. 958–967. Springer, Heidelberg (2007)
9. Tao, Q., Veldhuis, R.: Hybrid fusion for biometrics: Combining score-level and decision-level fusion. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2008*, June 2008, pp. 1–6 (2008)
10. Ojala, T., Pietikainen, M., Harwood, D.: Comparative study of texture measures with classification based on feature distributions. *Pattern Recognition* 29, 51–59 (1996)
11. Boom, B., Spreeuwens, L., Veldhuis, R.: Model-based reconstruction for illumination variation in face images. In: *8th International Conference on Automatic Face and Gesture Recognition (FGR 2008)* (2008)
12. Sim, T., Kanade, T.: Combining models and exemplars for face recognition: An illuminating example. In: *Proc. CVPR Workshop on Models versus Exemplars in Computer Vision* (December 2001)
13. Georghiadis, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(6), 643–660 (2001)
14. Lee, K., Ho, J., Kriegman, D.: Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27(5), 684–698 (2005)
15. Phillips, P., Flynn, P., Scruggs, T., Bowyer, K., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, June 2005, vol. 1, pp. 947–954 (2005)