



ELSEVIER

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Virtual illumination grid for correction of uncontrolled illumination in facial images

B.J. Boom^{*}, L.J. Spreeuwers, R.N.J. Veldhuis

University of Twente, EEMSC, Signals & Systems, P.O. Box 217, 7500 AE, Enschede, The Netherlands

ARTICLE INFO

Keywords:

Virtual illumination grid
Face recognition
Model-based face illumination correction

ABSTRACT

Face recognition under uncontrolled illumination conditions is still considered an unsolved problem. In order to correct for these illumination conditions, we propose a virtual illumination grid (VIG) approach to model the unknown illumination conditions. Furthermore, we use coupled subspace models of both the facial surface and albedo to estimate the face shape. In order to obtain a representation of the face under frontal illumination, we relight the estimated face shape. We show that the frontal illuminated facial images achieve better performance in face recognition. We have performed the challenging Experiment 4 of the FRGCv2 database, which compares uncontrolled probe images to controlled gallery images. Our illumination correction method results in considerably better recognition rates for a number of well-known face recognition methods. By fusing our global illumination correction method with a local illumination correction method, further improvements are achieved.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

One of the major problems of face recognition in uncontrolled conditions is the variation caused by illumination. Our contribution is an illumination correction method that is capable of handling multiple light sources. The purpose is to correct for multiple light sources in a single facial image. Correcting for illumination effects in images taken under uncontrolled illumination conditions is more challenging than the standard experiments for face illumination, which address removing illumination from facial images recorded in laboratory conditions (Yale B database, CMU-PIE database), illuminated with a single light source. In order to correct for uncontrolled illumination conditions, we try to reconstruct the illumination conditions. The requirements of the frontal illuminated facial image are that it removes the illumination variations without introducing artifacts, while preserving the identity information for recognition. The illumination correction method is used as an independent preprocessing method. The advantage is that the generated frontal illuminated facial image can be the input of various face recognition methods and allows us to use a single gallery image in face recognition.

Several methods have been proposed to correct for illumination variations in facial images. We will categorize them based on two criteria. The first criterion is the complexity of the reflectance model. We discriminate between reflectance models that use a

weak assumption about illumination, the Lambertian reflectance model, a reflectance function based on the Lambertian assumption (for example Spherical Harmonics) and advanced reflectance models (Phong, Torrance-Sparrow). The second criterion is the complexity of the face model. There are methods, which use no model of the face, other methods make models of the appearance, some use an implicit model of the 3D surface and texture, while other methods have applied an explicit model of the 3D surface and texture. In Fig. 1, we categorize illumination correction methods based on these criteria and divide them into four groups.

The first group (bottom-left oval of Fig. 1) contains methods that do not need a face model and do not assume an explicit reflectance model. These methods usually perform preprocessing based on the local regions, for example Histogram Equalization [1] or (Simplified) Local Binary Patterns [2,3]. Other methods like Gross et al. [4] and Tan et al. [5] use the local region around the pixel to perform illumination correction based on some properties of the reflectance. The Self Quotient Image [6] uses the local region for correction based on the Lambertian reflectance model without the need of a model of the face.

There are also methods which learn the behaviour of reflectance and so become invariant for illumination in the face. These methods use only the appearance learned for a bootstrap database with different labeled illumination conditions. Note that they do not assume a reflectance model (center-left oval of Fig. 1). Tensorfaces [7] can be trained to handle multiple variations like expression, illumination and pose using multidimensional subspace models. This is extended in [8] with a model which also include 3D shape parameters, but the illumination is still modelled using subspace models. In [9], a subspace model is used

^{*} Corresponding author. Tel.: +31 534892897; fax: +31 534891060.
E-mail address: b.j.boom@ewi.utwente.nl (B.J. Boom).

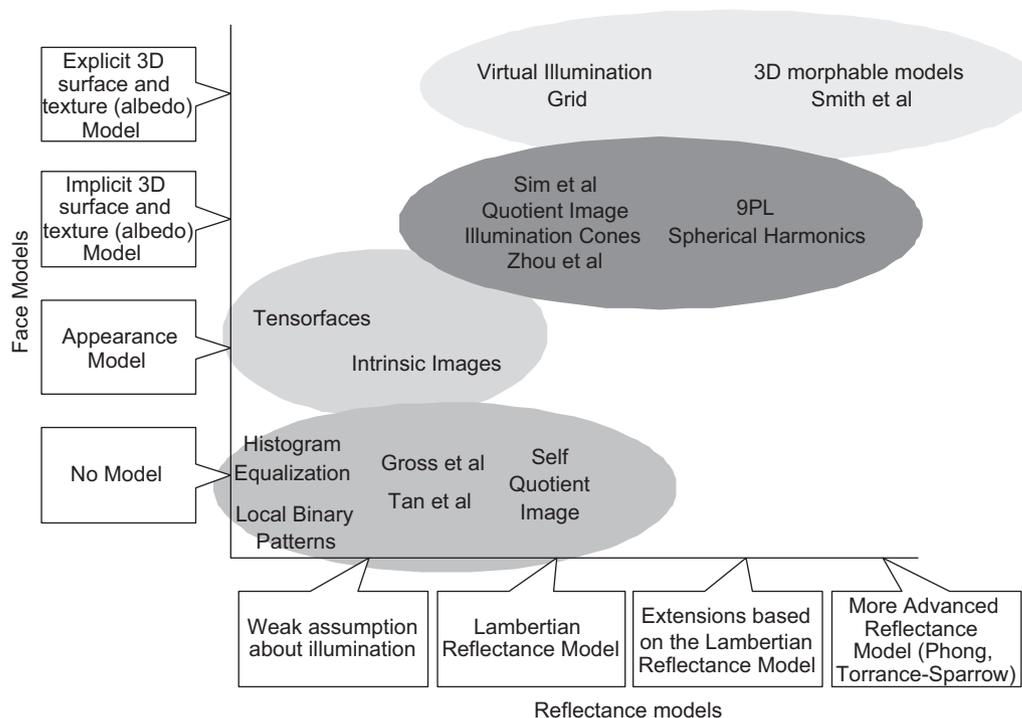


Fig. 1. The categorization (in face and reflectance models) of the illumination correction methods separating four major groups based on two criteria.

to compute Intrinsic Images, separating the image in reflectance and illuminance. Subspace models that learn the behaviour of the reflectance on faces also allow correction for this reflectance. However, these methods usually depend heavily on a bootstrap database with varying illumination conditions. This makes it difficult to predict whether these methods are robust to unseen conditions, which is mostly the case with uncontrolled conditions.

In face illumination correction, many correction methods use both assumptions on the illumination reflectance as well as implicitly taking into account the 3D surface and texture, usually by estimation of surface normals and albedo (center-right oval of Fig. 1). An example is the Quotient Image [10] which estimates illumination using the Lambertian reflectance model. This method computes a quotient image based on the assumption that faces have a similar surface. An estimate of the surface is then obtained from a bootstrap set of faces. In [11], an illumination cone can be determined from three images illuminated with independent light sources. Sim et al. [12] proposed a method based on the Lambertian reflectance model which corrects for illumination in a single facial image. This method uses a large bootstrap database containing many illumination variations in order to correct for these conditions. Spherical Harmonics are proposed in both [13,14] which give an approximation of a 9D linear subspace under all possible Lambertian illuminations. Zhang et al. [15,16] proposed a method to obtain the Spherical Harmonics for a single image illuminated under unknown illumination by using a bootstrap database to model shadows and reflections. In [17], a configuration of nine points of light (9PL) is determined to construct a linear subspace for face recognition. In face recognition, nine gallery images illuminated or rendered under predefined conditions are necessary to perform face recognition, which requires specialized data acquisition of the gallery images. Zhou et al. [18,19] span a linear subspace using object-specific albedo-shape matrices. They find an illumination free identity vector by optimizing both identity vector and illumination conditions to best resemble the input image. This gives them an illumination free identity vector instead of an estimate of the 3D

surface and albedo, from which we render the frontal illuminated facial image. Because Zhou et al.'s method is the closest to ours, we have chosen to point out other differences with Zhou et al.'s method throughout the text.

The last group in Fig. 1 (top-right oval) differs from other methods because it uses 3D face models. The 3D morphable models [20] are among the first to use the 3D information of faces, where PCA models are used for both the 3D shape and texture. Using the Phong reflectance model for illumination, a parameter optimization method is used to render a facial image which is close to the input image. In [21], Smith et al. propose a statistical model for normal maps to accurately estimate the surface and in [22] the same authors use a subspace model of the depth map as a geometrical constraint. In [23], we use a shape model in combination with the Lambertian reflectance model to correct for illumination of a single light source. In [24], we have improved this method by modelling both ambient and diffuse illumination and estimated the depth maps. We observe that most illumination correction methods have difficulties in modelling multiple light sources, especially when they cause some reflectance in shadow areas.

This paper is organized as follows: In Section 2, we describe the illumination correction method that is able to deal with multiple light sources. We first introduce the necessary reflectance and face models which then allows us to calculate a reconstruction of the face shape iteratively. In Section 3, we describe experiments and the results and fuse our global illumination correction with a local illumination correction method. We will discuss the results obtained with our method in Section 4 and finally provide conclusions in Section 5.

2. Method

2.1. Reflectance model

In order to correct for the illumination in a single facial image, we use a reflectance model for the behaviour of illumination.

Because of our focus on uncontrolled illumination conditions, we assume that faces are illuminated by multiple light sources. We use the Lambertian reflectance model, which gives a good approximation of the reflectance behaviour on the surface of faces [11]. The image intensity $b \in \mathcal{R}$ at a certain position $\mathbf{p} = \{x, y\}$ in the image can be described by the following equation:

$$b(\mathbf{p}) = \rho(\mathbf{p}) \sum_I \max(0, \mathbf{n}(\mathbf{p})^T \mathbf{s}_I) \quad (1)$$

where the face shape $\mathbf{h}(\mathbf{p}) = \rho(\mathbf{p})\mathbf{n}(\mathbf{p})^T$ consists of the surface normals $\mathbf{n} \in \mathcal{R}^3$, the albedo of the surface given by $\rho \in \mathcal{R}$ and the max operation allows us to model attached shadows. A normalized vector $\mathbf{s} \in \mathcal{R}^3$ defines the direction of the illumination. The intensity of the light is given by $i \in \mathcal{R}$. Instead of finding multiple light directions in a continuous domain, we use L discrete directions, assuming that a light source in continuous direction can be created using multiple light sources in discrete directions. The Lambertian reflectance model in Eq. (1) in this form cannot model cast shadows on the face surface. There are two kinds of shadows on faces. The first kind is called “attached shadows”. In this case, the Lambertian reflectance model does not hold because the normal is not directly facing the light source. This results in a negative image intensity, which can be easily detected and corrected by replacing the negative value by zero. The second kind of shadow is due to the geometry of the face that blocks the light source, these are called “cast shadows”. These shadows are harder to calculate because we need to perform ray tracing. Shadows can be seen as hard binary decisions. This definition holds with the exception of areas, which contain the transition between light and shadow areas. We propose to model shadows in the Lambertian reflectance model using a weight $e_l(\mathbf{p})$, which is linked to the light direction. This weight is in fact the expectation $e_l \in [0, 1]$ that a shadow occurs at position \mathbf{p} given a certain light direction l :

$$\hat{b}(\mathbf{p}) = \rho(\mathbf{p})\mathbf{n}(\mathbf{p})^T \sum_I \mathbf{s}_I e_l(\mathbf{p}) \quad (2)$$

The illumination conditions for a certain position \mathbf{p} can then be described by $\mathbf{v}(\mathbf{p}) = \sum_I \mathbf{s}_I e_l(\mathbf{p})$. In case of an attached shadow, $e_l = 0$, thus making the max operation in Eq. (1) unnecessary. This user-independent expectation can be used as weight, giving smooth values in the areas which contain the transition between light and shadow, while we have a hard binary decision in area that certainly contain shadows. This expectation is determined from a training set of multiple surfaces, where we calculate for a single surface a binary decision that a shadow occurs at position \mathbf{p} gives a light direction l using a ray tracer. The expectation is obtained by taking the mean over all the binary values. We determine the expectation at all positions \mathbf{p} and for all L light directions in the grid. Apart from the expectation, we also determine the variations $\sigma_l^2(\mathbf{p})$ at every position, which we use for the albedo estimate, described in Section 2.7.

The goal of our correction method is to find the illumination conditions $\mathbf{v}(\mathbf{p})$ and the face shape $\mathbf{h}(\mathbf{p})$ that best explain our input image $b(\mathbf{p})$. This method minimizes the distance between the input image $b(\mathbf{p})$ and an estimate based on the models $\hat{b}(\mathbf{p})$, in our case obtained from Eq. (2). Note that multiple combinations of light conditions and face shapes can result in the same image based on Eq. (2). For this reason, it is necessary to use domain specific knowledge to constrain the shape from shading problem. This domain specific knowledge is enforced using subspace models of the face shape (Section 2.2). The subspace models also allow us to estimate the face shape (surface and albedo). By obtaining the face shape, we can easily compute facial images under frontal illumination by replacing the $\mathbf{v}(\mathbf{p})$ for $\mathbf{v}_{\text{frontal}}(\mathbf{p})$.

Of course, it is also possible for us to illuminate the face with other illumination conditions using our virtual illumination grid.

2.2. Face shape and albedo models

The reflectance of the light depends on the object that is illuminated. Faces are difficult to model, but multiple techniques have been proposed in literature to achieve this. Zhou et al. [19] use a linear subspace of object-specific albedo-shape matrices as illumination free term containing both albedo and surface normals. In VIG, we have chosen to split the illumination free terms in order to create two separate linear subspaces. This has the advantage of allowing us to perform an estimate of the surface (Section 2.6). For the subspace models, we use a vectorized representation of albedo ρ and surface \mathbf{z} (depth map of the face), instead of a notation for every position \mathbf{p} in an image. From a set of surfaces $\{\mathbf{z}_m\}_{m=1}^M$, we obtain the mean surface $\bar{\mathbf{z}}$ and a covariance matrix $\Sigma_{\mathbf{z}}$. This allows us to compute a subspace by solving the eigenvalue problem, obtaining the eigenvalues $\Lambda_{\mathbf{z}}$ and the eigenvectors $\Phi_{\mathbf{z}} = [\phi_1, \dots, \phi_N]^T$ of the covariance matrix $\Sigma_{\mathbf{z}}$:

$$\hat{\mathbf{z}} = \bar{\mathbf{z}} + \Phi_{\mathbf{z}} \mathbf{u}_{\mathbf{z}} \quad (3)$$

Given that we are able to obtain the variations of the surface $\mathbf{u}_{\mathbf{z}}$, we can estimate a surface $\hat{\mathbf{z}}$ using Eq. (3). A similar PCA model is obtained for the albedo. In order to take into account the correlation between depth maps and albedo, which is probably present, we combine both PCA models, creating the following concatenated vector:

$$\mathbf{y}_{\mathbf{t}} = \begin{pmatrix} W_{\mathbf{z}} \mathbf{u}_{\mathbf{z}} \\ \mathbf{u}_{\rho} \end{pmatrix} \quad (4)$$

where $W_{\mathbf{z}}$ is a diagonal weighting matrix, allowing difference in weight between depth maps and albedo. This method of combining both PCA models is similar to the approach used in the Active Appearance Model [25]. Because albedo and depth maps cannot be compared directly, we measure the effects in appearance of changing $\mathbf{u}_{\mathbf{z}}$ and \mathbf{u}_{ρ} using the RMS error. This gives us a relative weight between albedo and depth maps changes. We apply PCA to the vector $\mathbf{y}_{\mathbf{t}}$, giving us a model which contains both depth maps and albedo:

$$\mathbf{y}_{\mathbf{t}} = \Phi_{\mathbf{t}} \mathbf{u}_{\mathbf{t}} \quad (5)$$

In this case, $\Phi_{\mathbf{t}} = [\phi_1, \dots, \phi_K]^T$ are the eigenvectors and $\mathbf{u}_{\mathbf{t}}$ is already zero mean allowing us to control the depth map and shape in the following way:

$$\hat{\mathbf{z}} = \bar{\mathbf{z}} + \Phi_{\mathbf{z}} W_{\mathbf{z}}^{-1} \Phi_{\mathbf{t}} \mathbf{u}_{\mathbf{t}} \quad (6)$$

$$\hat{\rho} = \bar{\rho} + \Phi_{\rho} \Phi_{\mathbf{t}} \mathbf{u}_{\mathbf{t}} \quad (7)$$

where

$$\Phi_{\mathbf{t}} = \begin{pmatrix} \Phi_{\mathbf{t}\mathbf{z}} \\ \Phi_{\mathbf{t}\rho} \end{pmatrix} \quad (8)$$

Several papers already have used separate PCA models for texture and shape in order to correct for illumination [20]. However, they usually do not take into account the correlation between the depth map and the albedo. Using this correlation has the advantage of prohibiting improbable combinations of depth maps and albedo to explain appearances in images.

2.3. Illumination correction method

Given a single image, we want to estimate the face shape and the illumination conditions, using both the Lambertian reflectance model (Eq. (2)) and the PCA models of surface and albedo

(Eqs. (6) and (7)). Of course, we can perform an exhaustive search, as in the 3D morphable models where optimized code can perform such an operation with 4.5 min [26]. Instead, we chose an iterative scheme, to first estimate the illumination conditions and then find the model parameters of the depth map and surface. Using the found depth map and albedo, we can then improve the accuracy of the illumination conditions, which in turn improves the depth map and albedo. We repeat these steps several times (average of 5 iterations). Because the most time consuming computations are linear, this method can be computed much faster than the 3D morphable model. The pseudo-code of our correction method is given below:

- Repeat
 - Estimate illumination conditions (Section 2.4)
 - Estimate crude face shape (Section 2.5)
 - Estimate the surface parameters (Section 2.6)
 - Estimate the albedo parameters (Section 2.7)
- Until convergence (based on evaluation of the obtained
- illumination conditions, surface and albedo (Section 2.8))
 - Refinement of the albedo (Section 2.9)

We will discuss the different components in the following sections.

2.4. Estimation of the illumination conditions

Given the image, we want to estimate the illumination conditions in the image. Because we assume a grid of discrete light directions, we have to determine the intensity of the light for every point in the grid in order to calculate the global illumination conditions. To obtain the illumination conditions, we use an estimate of the face shape $\hat{\mathbf{h}}(\mathbf{p})$. In the first iteration, we use the mean face shape $\bar{\mathbf{h}}(\mathbf{p})$ to obtain the light intensities, while in the next iterations, we use the estimated face shape from the previous iteration. The light intensities are calculated as follows:

$$\mathbf{i} = \arg \min_{\mathbf{i}} \sum_{\mathbf{p}} \left\| \sum_{l=0}^L \hat{\mathbf{h}}(\mathbf{p})^T s_l i_l \mathbf{e}_l(\mathbf{p}) - b(\mathbf{p}) \right\|^2$$

where $i_l \geq 0$ (9)

This can be solved using a constrained linear least square solver where the light intensity cannot be negative. Because Eq. (9) is an overcomplete system, even using a relatively poor estimate of the face shape $\hat{\mathbf{h}}$, still gives acceptable results. The accuracy of the illumination conditions depends also on the configuration of the light sources in the grid. We have experimented with two grids by varying the azimuth and elevation angles by 10° or 20° from -80° to 80° . We have observed that the results of both grids are very similar, while the computations with a grid of 10° take much longer. For this reason, we decide to use the grid of 20° for all experiments. In [17], a configuration of nine points of light (9PL) at predefined locations is used to obtain an illumination free representations. In VIG, however, we want to obtain a reconstruction of the surface and the albedo, for which the large illumination grid is necessary to accurately model cast shadows. With the 9PL methods, accurate modelling of cast shadows is not possible, which will result in artifacts in the reconstruction.

2.5. Estimation of the crude face shape

Given the image and the illumination conditions, the goal is to obtain the face shape. A crude estimate of the face shape is necessary in order to improve the face shape using the linear

subspace models. To obtain the crude estimate, we use the following two assumptions. The first assumption is that the Lambertian reflectance model (Eq. (11)) holds. The second assumption is that the face shape should be similar to the mean face shape. This can be measured by taking a distance between the face shape $\mathbf{h}(\mathbf{p})$ and the mean face shape $\bar{\mathbf{h}}(\mathbf{p})$. In our case, we have taken the Mahalanobis distance between two vectors, where $\Sigma_{\mathbf{h}(\mathbf{p})}$ is the covariance matrix of the face shape at location \mathbf{p} obtain from a training set:

$$\hat{\mathbf{h}}(\mathbf{p}) = \arg \min_{\mathbf{h}(\mathbf{p})} (\mathbf{h}(\mathbf{p}) - \bar{\mathbf{h}}(\mathbf{p}))^T \Sigma_{\mathbf{h}(\mathbf{p})}^{-1} (\mathbf{h}(\mathbf{p}) - \bar{\mathbf{h}}(\mathbf{p})) \quad (10)$$

$$\text{where } b(\mathbf{p}) = \mathbf{h}(\mathbf{p})^T \mathbf{v}(\mathbf{p}) \quad (11)$$

We can minimize the Mahalanobis distance (Eq. (10)) with the Lambertian reflectance model (Eq. (11)) as a constraint using Lagrange multipliers. This gives us a crude estimate of face shape $\hat{\mathbf{h}}$, which we will improve in the following sections.

2.6. Estimation of the surface

Given an estimate of the face shape $\hat{\mathbf{h}}(\mathbf{p})$, we further improve this estimate by applying geometrical constrains. By calculating a depth map from the crudely estimated face shape, we can automatically enforce the geometrical constrains. The PCA model allows us to introduce domain specific information, which ensures convergence of this shape from shading problem. In the method of Zhou et al. [19], integrability and symmetry constrains are used in generalized photometric stereo to recover the shape and albedo, using multiple face images under different illumination conditions. For Zhou et al. to become invariant to illumination in a single image, no geometrical constraints are used, while our method allows us to estimate the surface and thus enforce the integrability constraints even for a single image. We know that the gradient of the surface in x and y direction is equal to $\nabla_x z(\mathbf{p}) = h_x(\mathbf{p})/h_z(\mathbf{p}) = h_{xz}(\mathbf{p})$ and $\nabla_y z(\mathbf{p}) = h_y(\mathbf{p})/h_x(\mathbf{p}) = h_{yz}(\mathbf{p})$. Instead of calculating the depth map directly, we estimate the variations \mathbf{u}_z , using the following equation:

$$\mathbf{u}_z = \arg \min_{\mathbf{u}_z} \|\nabla_x \bar{z} + \nabla_x \Phi \mathbf{u}_z - \hat{h}_{xz}\|^2 + \|\nabla_y \bar{z} + \nabla_y \Phi \mathbf{u}_z - \hat{h}_{yz}\|^2 \quad (12)$$

A similar procedure to obtain the face surface is performed in [22] to find the variations from the surface model. This can be solved using a linear least square solver. The final depth map can be computed using Eq. (3). In our case, we combine the depth map and albedo models, to calculate the final depth map taking into account the correlation between the albedo, see Section 2.2. A practical problem in calculating the depth maps is that vectors which are almost perpendicular to the viewer direction sometimes cause large spikes. In order to deal with this problem, we remove these locations. In this case, we are still able to calculate the variations of the surface because PCA can also be applied on an incomplete set of locations.

2.7. Estimation of the albedo

In the previous section, we estimate the illumination conditions and the surface from which we can obtain the surface normals $\mathbf{n}(\mathbf{p})$. The only remaining unknown in Eq. (2) is the albedo ρ . To calculate the albedo, we can solve Eq. (2) for every pixel value, but we observe that there are sometimes erroneous effects in the albedo caused by the user independent shadow model $e_l(\mathbf{p})$, because the shadow mapping is not precise. To overcome this problem, we use Bayes theorem which allows us to

calculate a MAP estimate for the albedo. The MAP estimate of the albedo at a certain location is defined as follows:

$$P(\rho(\mathbf{p})|b(\mathbf{p})) = \frac{P(b(\mathbf{p})|\rho(\mathbf{p}))P(\rho(\mathbf{p}))}{P(b(\mathbf{p}))} \quad (13)$$

In order to obtain the albedo term, we can maximize the following equation:

$$\begin{aligned} \rho_{noshadow}(\mathbf{p}) &= \operatorname{argmax}_{\rho(\mathbf{p})} P(b(\mathbf{p})|\rho(\mathbf{p}))P(\rho(\mathbf{p})) \\ &= \operatorname{argmax}_{\rho(\mathbf{p})} \mathcal{N}\left(\rho(\mathbf{p}) - \frac{b(\mathbf{p})}{\mu_r(\mathbf{p})}, \sigma_r(\mathbf{p})\right) \times \mathcal{N}(\mu_\rho(\mathbf{p}), \sigma_\rho(\mathbf{p})) \end{aligned} \quad (14)$$

In Eq. (15), we define a mean reflection term $\mu_r(\mathbf{p}) = \sum_i \mathbf{n}(\mathbf{p})^T \mathbf{s}_i i e_i(\mathbf{p})$ and the variations of the reflection term $\sigma_r^2(\mathbf{p}) = \sum_i \mathbf{n}(\mathbf{p})^T \mathbf{s}_i i \sigma_i^2(\mathbf{p})$ (see Section 2.1 for σ_i^2) and assume that the shadow maps are Gaussian distributed. Using the training set from which we calculate the PCA models, we can also easily determine the mean albedo $\mu_\rho(\mathbf{p})$ and standard deviation of the albedo $\sigma_\rho(\mathbf{p})$ at certain locations necessary in Eq. (15). By taking the derivative of the log probabilities, we find a shadow free albedo term $\rho_{noshadow}(\mathbf{p})$. From this shadow free albedo $\rho_{noshadow}$, we computed the variations \mathbf{u}_ρ using a linear least square solver giving us all the subspace model parameters, see Section 2.2.

2.8. Evaluation of the obtained illumination conditions, surface and albedo

In Sections 2.6 and 2.7, we compute both the variations of the surface and albedo. With these variations, we can determine the variation of the combined models, see Eq. (5) and give the estimates for the surface and albedo using Eqs. (6) and (7). Given the estimated albedo and surface, together with the illumination conditions found in Section 2.4, we can reconstruct an image $\hat{\mathbf{b}}$ which should be similar to the original image. This can be measured using the sum of the square differences between the pixel values. It is also interesting to monitor the variations from the PCA model, which shows if overfitting occurs at certain light directions. For this reason, we use an evaluation measure, which is similar to the measure used in [20]:

$$E = \frac{1}{\sigma_b} \|\mathbf{b} - \hat{\mathbf{b}}\|^2 + \sum_{k=1}^K \frac{u_{\mathbf{t}}^2(k)}{\lambda_{\mathbf{t}}(k)} \quad (16)$$

In this case, σ_b controls the relative weight of the distance between the original and reconstructed image, which is the most important factor to minimize, $\lambda_{\mathbf{t}}$ are the eigenvalues of the depth map and albedo model, see Eq. (5). This evaluation measure allows us to determine when the iterative estimation procedure convergence.

2.9. Refinement of the albedo

The albedo $\rho_{noshadow}(\mathbf{p})$, calculated from the MAP estimate and subspace model, misses details (Section 2.7). To recover these details, we perform two steps. In the first step, we determine the albedo using the original facial image to recover the details removed by the subspace models. In the second step, we filter the recovered albedo by using the correlation between the different positions in the images to remove spikes.

The first step is to recover the details in albedo based on the image, where we use the following equation:

$$\rho_{details}(\mathbf{p}) = \frac{b(\mathbf{p})}{\mu_r(\mathbf{p})} \quad (17)$$

In Eq. (17), we assume that both the surface normals and illumination conditions are correctly estimated.

In the second step, we remove erroneous effects in the albedo caused by using a user-independent shadow model, which usually results in spike in areas containing the transition between shadow and light. To suppress the spikes, we learn the relationship between albedo at neighboring locations using correlation. Based on the correlation between neighboring locations, we filter the albedo removing spikes if the correlation between locations expects a different albedo value. If the albedo is similar to the expected albedo value it will hardly change. The correlation between locations is learned using a training set. To explain the filter, we use a different notation, where the location \mathbf{p} will be replaced by the subscript x and y :

$$\hat{\rho}_{x,y}^{x,y+1} = \bar{\rho}_{x,y} + r \frac{\sigma_{x,y}}{\sigma_{x,y+1}} (\rho_{x,y+1} - \bar{\rho}_{x,y+1}) \quad (18)$$

Using statistics, we can predict the value $\hat{\rho}_{x,y}^{x,y+1}$ at location (x,y) using the value at location $(x,y+1)$ given the correlation r between both positions and the means $\bar{\rho}$ and standard deviations σ of the albedo determined from a training set. We perform a similar prediction from all the locations $((x,y+1), (x,y-1), (x+1,y)$ and $(x-1,y)$), which surround (x,y) . In order to compute the final albedo map, we use all surrounding locations in the following equation:

$$\begin{aligned} \rho_{x,y}^{final} + \lambda [(\rho_{x,y}^{final} - \hat{\rho}_{x,y}^{x,y+1}) + (\rho_{x,y}^{final} - \hat{\rho}_{x,y}^{x,y-1}) \\ + (\rho_{x,y}^{final} - \hat{\rho}_{x,y}^{x+1,y}) + (\rho_{x,y}^{final} - \hat{\rho}_{x,y}^{x-1,y})] = \rho_{x,y}^{details} \end{aligned} \quad (19)$$

The final albedo $\rho_{x,y}^{final}$ is estimated by taking into account the correlation between surrounding location with a weight factor of $\lambda = 0.2$ and as an initial estimate we use $\rho_{x,y}^{details}$. To solve Eq. (19) for an entire grid of albedos, we use a multigrid method for boundary value problems (Simultaneous Over-Relaxation) described in [27]. This allows us to determine the final albedo $\rho_{x,y}^{final}$. Using this method, we are able to remove the spikes, but at the same time, we preserve the details in the final albedo $\rho_{x,y}^{final}$.

3. Experiments

3.1. Training VIG

In order to create the PCA models, it is necessary to obtain a dataset from which we can calculate both the depth maps and the albedo. One of the important limitations of this method can be the 3D database used for training. This database has to be sufficient in order to obtain a model which can be used to reconstruct a probe image. We have experimented with two publicly available databases, namely the 3D FRGC training set (Spring 2003 range images) [28] and the 3D Bosphorus Face Database [29]. Both databases contain facial images together with their range images, which give us for each pixel a 3D coordinate. We register the facial images to a common coordinate system, using the landmarks provided by the databases. From the range maps provided with the images, we calculate both depth maps and the surface normals, where we use some simple spike removal and hole filling methods to obtain smooth depth maps. Because the illumination in the images is controlled, it is possible to estimate the illumination conditions from both the surface normals and the appearance in the image. This also allows us to compute the albedo. A disadvantage of the 3D FRGC training set is that the images are overexposed, this makes the albedo estimation less accurate.

In our earlier work, we used the 3D FRGC training set to create our face model. The 3D FRGC training set contains individuals that are also present in Experiment 4 of the FRGCv2 database. A disadvantage of this set, however, is that all faces have a neutral

expression. The 3D Bosphorus Face Database contains all kinds of expressions and other variations and this database has no overlapping individuals with Experiment 4. Still the results that are achieved by training using 3D Bosphorus Face Database are better, mostly because of the presence of expressions. For this reason, we use the 3D Bosphorus Face Database to obtain the PCA models of the depth map and albedo. We have observed that a better training set has a large effect on the face recognition results. Important is that this database contains all kinds of variations (expression, race differences), but at the same time is taken under controlled conditions, like illumination, registration and pose.

3.2. Experimental setup

In order to test the performance of illumination correction methods, we use both the CMU-PIE [30] and the FRGCv2 [28] databases. Most methods for illumination correction are evaluated using databases recorded in a laboratory, with images illuminated using a single light source on a predetermined grid. The CMU-PIE database is such a database, which allows us to compare VIG with other methods under different predetermined illumination conditions. A problem of these kinds of databases is that bootstrap data from the same predetermined light sources are often used, which positively biases the results. In this research, we chose to also evaluate the methods on images taken under uncontrolled conditions. For this reason, we have performed FRGCv2 Experiment 4, where a single gallery image taken under controlled illumination conditions is compared with a probe image taken under uncontrolled illumination conditions. This matches the real-life problem that all illumination correction methods aim to solve. Although one of the biggest challenges is to remove the illumination variations from these images, there are more challenges in face recognition, see [31], that are not addressed in this paper. Examples of problems other than illumination are (small) pose variations, expressions, occlusions due to caps and glasses, which all have negative effects on the final recognition results.

In order to compare our illumination correction method, we have used the PCA-LDA likelihood ratio classifier described in [32] and the kernel correlation filter (KCF) together with the normalized cosine distance [33] for face recognition. The latter method already achieves good performance on the FRGCv2 database, see [33], together with the illumination correction method of Gross et al. [4]. In literature, several illumination correction methods that use local regions show promising results on the FRGCv2 database. For this reason, we use Gross et al. [4] and Tan et al. [5] for comparison. One of the best methods that uses the entire image for illumination corrections is the method of Zhang et al. in [15], which uses the Spherical Harmonics representation. For comparison, we have developed our own implementation of this method. This method is trained on the same database as is used to train our illumination correction methods.

3.3. Face recognition results on CMU-PIE database

We performed a simple experiment on the CMU-PIE [30] database in order to compare our illumination correction method with other methods. In this experiment, the difficult images of CMU-PIE illuminated without ambient light are used. In Fig. 2, we show that our method is able to render the face images with different illumination conditions, although this becomes difficult, if almost half of the face contains shadow (third row of Fig. 2). For the face recognition experiments, the images are corrected to frontal illumination in order to make them comparable. We used

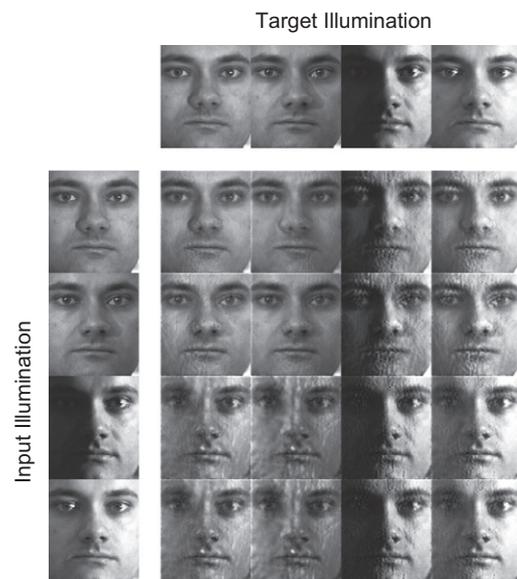


Fig. 2. VIG is able to illuminate faces under different light conditions, the first row contains the recorded images which VIG has to render given the input images on the first columns. The rest of the rows contain the rendered images given the input image.

Table 1

The face recognition results of the PCA-LDA likelihood ratio on the CMU-PIE database in recognition rates.

Uncorrected	Zhang et al.	Gross et al.	Tan et al.	VIG
72.2%	74.5%	86.8%	88.1%	90.8%

20% of the 68 subjects for training and used the frontal illuminated images as gallery images. We repeat this experiment 20 times, randomly assigning subjects to the training and test set. In Table 1, we show the results in face recognition. The illumination correction methods easily improve the recognition results because it mostly contains illumination variations. We observe that VIG performs better than the other illumination correction methods on this face database.

3.4. Face recognition results on FRGCv2 database

In Experiment 4 of the FRGCv2 database, three sets of images are defined: a training set, a target set and a query set. All the images of these sets are corrected using all illumination correction methods. The output of the different illumination correction methods is shown in Fig. 3, where the uncorrected images are in the first row and the corrected images with the methods of, respectively, Gross et al., Zhang et al. and VIG are in the second, third and fourth row. Fig. 3 shows that the illumination correction methods are also used on images taken under controlled conditions (the first two columns), where VIG seems to only change the overall light intensity of the image. The other four images are taken under uncontrolled conditions, which include sometimes difficult illumination conditions as can be seen in the fourth image. We observe that VIG is able to remove most of the cast shadows caused by the nose, especially visible in the fourth image. Furthermore, it is able to correct for the dark areas (right side) in the fourth image. We can, however, not correct for the cast shadows on the cheeks caused by the glasses in the fifth

image, because our model does not include glasses nor the reflections they cause.

In order to evaluate the effects of the illumination corrections methods on the face recognition, we setup the following experiment. After the illumination correction on all images, the face recognition methods are trained using the corrected images in the training set. We perform a one-to-one comparison described by FRGCv2 database, where we compare one image from the target set (controlled illumination) with one image from the query set (uncontrolled illumination). The face recognition methods used for this comparison are the PCA-LDA likelihood ratio and the kernel correlation filter (KCF). The receiver operating characteristics (ROC) of these methods are presented in Figs. 4(a) and (b). We observe from Fig. 4(a) that the virtual illumination grid method clearly performs best in combination with the likelihood ratio, followed by the method of Gross et al. The other global illumination correction method of Zhang et al. performs better than uncorrected images at FAR $\leq 1\%$. Using the KCF, we observe that the difference between VIG and Gross et al. is small.



Fig. 3. The output of the illumination correction methods: the first row contains the original images, the second row shows the output of local preprocessing method of Gross et al., in the third row the images corrected are the global illumination correction method of Zhang et al., the last row shows the images obtained using VIG.

The method of Gross et al. reaching similar results as reported in Fig. 4(b) of the paper on KCF [33], where they performed the same experiment. In the same paper, the KCF achieves even better recognition results, by using the gallery images to train a SVM. We did not perform this experiment, because it deviates for the FRGC protocol. The results of Zhang et al. are in this experiment better than uncorrected images. By comparing Figs. 4(a) and (b) with each other, the likelihood ratio reaches best results in combination with VIG.

3.5. Fusion

In [34], we have fused local and global illumination method in order to improve the performance. In this paper, we fuse VIG with the method of Gross et al. in order to investigate if the combination of both methods improves the performance. To fuse both methods, we use simple z-score normalization on the similarity scores. The ROC curves of this experiment are shown in Fig. 5. The combinations improve the results of the likelihood ratio slightly and the KCF much more. We also combine all scores achieving a much better performance than each of the separate classifiers. Of course, other face recognition methods can be used to even further improve the face recognition. However, here the goal is to show that the combination of VIG and Gross et al. improves the performance of well-known face recognition methods.

4. Discussion

In the previous section, we have shown that VIG improves the face recognition results. In order to provide more inside information about VIG, we discuss the following two issues: The first issue concerns the limitations of our method and we discuss the influence of these limitations on the face recognition results. The second issue is the accuracy of the depth maps in relationship with the reflectance models.

4.1. Limitations

We have shown that VIG performs well for the uncontrolled conditions in Experiment 4, for instance in Fig. 4(a), although fusion with the method of Gross et al. still shows room for

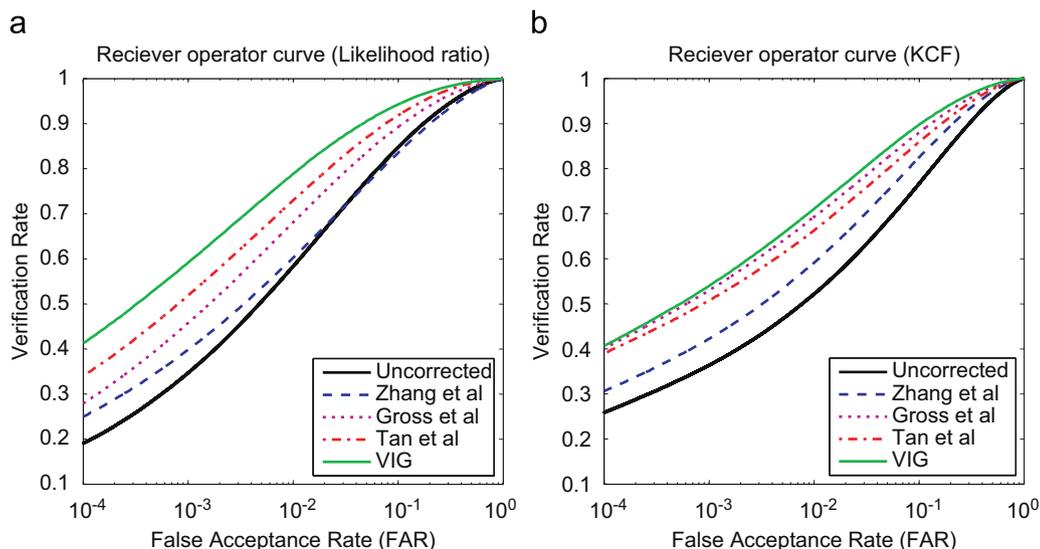


Fig. 4. ROCs for the FRGCv2 database Experiment 4 of the PCA-LDA likelihood ratio (left) and kernel correlation filters (right) for all the illumination correction methods.

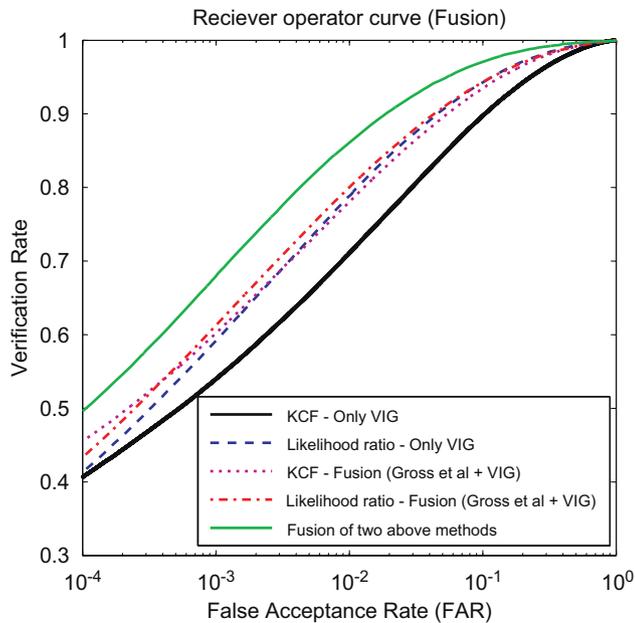


Fig. 5. ROCs for the FRGCv2 database Experiment 4 obtained by fusion of VIG and the method of Gross et al., for each face recognition method separately and also for the combination of all methods.



Fig. 6. Examples of face images containing baseball caps. Because of the cast shadow in the facial images caused by the baseball caps, VIG is unable to determine the correct illumination conditions. The first column shows the entire face, the second column is the region of interest used by VIG, the last column contains the corrected image.

improvement. We observed that the main reason is the use of reflectance and face models, which are not able to include all exceptions. Baseball caps, for instance, cause a cast shadow in the face, making it impossible to estimate the correct illumination conditions using the entire image, as can be observed in Fig. 6. Another well-known problem in face recognition is glasses, which reflect the light in other directions than anticipated by the reflection model, making the estimation of the face shape difficult. In order to solve these problems, a detection method for glasses

and cap can be developed, allowing us to ignore parts of the image in these cases. Although VIG might not always work in case of the mentioned exceptions, the results show a clear contribution of VIG in most other cases. By comparing VIG with the method of Gross et al., which gives us an illumination invariant representation of the face, we observe from the fusion results that both methods make different errors. The illumination invariant representation of Gross et al. is able to deal with for instance caps and glasses, because it uses local assumptions. On the other hand, VIG creates a reconstruction based on more global assumptions which cannot be modelled by Gross et al. The advantage is that the reconstruction will give us more information, like the estimated depth map of the face. This can be used to incorporate 3D and 2D face recognition, which is impossible with illumination invariant representation of Gross et al.

4.2. Accuracy of the depth maps

We estimate the depth maps based on the Lambertian reflectance model and a PCA model of depth maps of faces. From literature, we know that the Lambertian reflectance model is a reasonably good estimate of the reflectance of the skin, but more accurate reflectance models are known. Bidirectional reflectance distribution functions (BRDF) like the Phong and Torrance-Sparrow model probably provide a better explanation. The disadvantage of these models is that they are non-linear, which has large consequences on the computation time. Another possibility is to measure a BRDF for human skin, using 3D face acquisition equipment, although we are not sure that this BRDF holds for different types of light. Because we use the Lambertian reflectance model throughout this paper, we expect that the depth maps are not always accurate in certain regions. We also observe that PCA models have limitations in modelling details. On the other hand, PCA models are able to correct certain mistakes by enforcing domain specific knowledge. From Fig. 7, we observe that the depth maps estimated using VIG fail to recover the sharp contours. Instead, we obtain a smooth version of the surface.

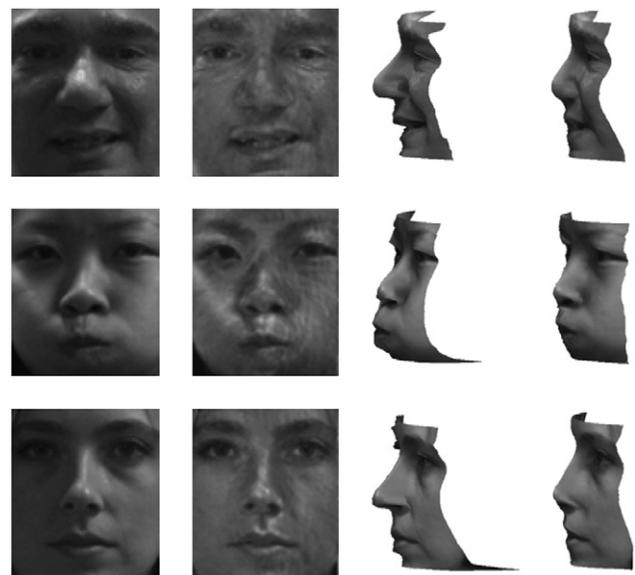


Fig. 7. Comparison of depth maps using the 3D FRGC database: first column contains the original image, second column shows the correction of VIG, third column is the 3D range map acquired with a laser-scanner and in last column our estimate of the depth map is given.

Similar results can be observed in the research presented in [35], where shape from shading is compared with measured profiles. However, Fig. 7 also shows that VIG is to a certain extent able to model expressions, like the round cheeks in the second row. Although, we show that the accuracy of the depth maps is limited, this does not have to affect the face recognition results. The reason that VIG might calculate an incorrect depth map is that our model can be biased to certain explanations. As long as for individuals the same mistakes are made in calculating the depth maps, these mistakes do not have to affect the face recognition results.

5. Conclusions

We present a new method to correct for illumination effects in facial images. Although multiple methods for illumination correction in facial images are described in literature, this method has some advantages with respect to most other methods. First of all, we assume and model multiple light sources using a virtual illumination grid (VIG). This allows us to achieve good recognition results, especially for facial images taken under uncontrolled conditions. Secondly, we use two PCA models, both for the depth maps and the albedo and we couple them, taking into account the correlation between depth map and albedo. Thirdly, we are able to estimate a depth map of the surface, which can be useful for improving face recognition under pose variations or for comparison with 3D face recognition. We test multiple methods on the FRGCv2 database using Experiment 4, where faces are recorded under uncontrolled conditions. Our experiment is different from most other illumination correction methods, which are tested on databases recorded in laboratories. In these cases, the illumination directions are predetermined and they include usually only a single light source, while in truly uncontrolled conditions, both the illumination directions and the number of light sources are usually unknown. We show that VIG is able to improve the results of different face recognition methods significantly under these uncontrolled conditions. Furthermore, we fuse VIG, which is a global illumination correction method, with a local illumination correction by Gross et al. [4], which further improves the recognition results.

Acknowledgement

This research was conducted in the framework of the PV3D project together with the Netherlands Forensic Institute, which was sponsored by the Dutch Ministry of Internal Affairs.

References

- [1] S. Shan, W. Gao, B. Cao, D. Zhao, Illumination normalization for robust face recognition against varying lighting conditions, in: IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 2003, AMFG 2003, 17 October 2003, pp. 157–164.
- [2] G. Heusch, Y. Rodriguez, S. Marcel, Local binary patterns as an image preprocessing for face authentication, in: 7th International Conference on Automatic Face and Gesture Recognition, 2006, FGR 2006, 10–12 April 2006, pp. 9–14.
- [3] Q. Tao, R.N.J. Veldhuis, Illumination normalization based on simplified local binary patterns for a face verification system, in: Biometrics Symposium 2007 at the Biometrics Consortium Conference, IEEE Computational Intelligence Society, Baltimore, MA, USA 2007, pp. 1–7.
- [4] R. Gross, V. Brajovic, An image preprocessing algorithm for illumination invariant face recognition, in: Audio- and Video-Based Biometric Person Authentication, Springer-Verlag 2003, pp. 10–18.
- [5] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, in: Analysis and Modeling of Faces and Gestures, 2007, pp. 168–182.
- [6] H. Wang, S. Li, Y. Wang, J. Zhang, Self quotient image for face recognition, in: International Conference on Image Processing, vol. 2, 2004, pp. 1397–1400.
- [7] M.A.O. Vasilescu, D. Terzopoulos, Multilinear analysis of image ensembles: tensorfaces, in: Computer Vision ECCV 2002, Springer-Verlag 2002, pp. 447–460.
- [8] J. Lee, B. Moghaddam, H. Pfister, R. Machiraju, A bilinear illumination model for robust face recognition, IEEE International Conference on Computer Vision, vol. 2, IEEE Computer Society, Los Alamitos, CA, USA, 2005, pp. 1177–1184.
- [9] C.-P. Chen, C.-S. Chen, Lighting normalization with generic intrinsic illumination subspace for face recognition, IEEE International Conference on Computer Vision, vol. 2, IEEE Computer Society, Los Alamitos, CA, USA, 2005, pp. 1089–1096.
- [10] A. Shashua, T. Riklin-Raviv, The quotient image: class-based re-rendering and recognition with varying illuminations, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2) (2001) 129–139.
- [11] A. Georghiades, P. Belhumeur, D. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 643–660.
- [12] T. Sim, T. Kanade, Combining models and exemplars for face recognition: an illuminating example, in: Proceedings of the CVPR Workshop on Models versus Exemplars in Computer Vision, 2001.
- [13] R. Basri, D. Jacobs, Lambertian reflectance and linear subspaces, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2) (2003) 218–233.
- [14] R. Ramamoorthi, P. Hanrahan, On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object, J. Opt. Soc. Am. A 18 (10) (2001) 2448–2459.
- [15] L. Zhang, D. Samaras, Face recognition under variable lighting using harmonic image exemplars, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2003, pp. 19–25.
- [16] L. Zhang, D. Samaras, Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics, IEEE Trans. Pattern Anal. Mach. Intell. 28 (3) (2006) 351–363.
- [17] K.-C. Lee, J. Ho, D.J. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 684–698.
- [18] S.K. Zhou, R. Chellappa, D.W. Jacobs, Characterization of human faces under illumination variations using rank, integrability, and symmetry constraints, in: Computer Vision—ECCV 2004, 2004, pp. 588–601.
- [19] S. Zhou, G. Aggarwal, R. Chellappa, D. Jacobs, Appearance characterization of linear Lambertian objects, generalized photometric stereo, and illumination-invariant face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2) (2007) 230–245.
- [20] V. Blanz, T. Vetter, A morphable model for the synthesis of 3d faces, in: SIGGRAPH '99: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, 1999, pp. 187–194.
- [21] W. Smith, E. Hancock, Recovering facial shape using a statistical model of surface normal direction, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 1914–1930.
- [22] W. Smith, E. Hancock, Recovering face shape and reflectance properties from single images, in: 8th IEEE International Conference on Automatic Face and Gesture Recognition, 2008, FG '08, 2008, pp. 1–8.
- [23] B. Boom, L. Spreeuwers, R. Veldhuis, Model-based reconstruction for illumination variation in face images, in: 8th IEEE International Conference on Automatic Face and Gesture Recognition, 2008, FG '08, 2008, pp. 1–6.
- [24] B.J. Boom, L.J. Spreeuwers, R.N.J. Veldhuis, Model-based illumination correction for face images in uncontrolled scenarios, in: Computer Analysis of Images and Patterns 2009, Munster, 2009.
- [25] T. Cootes, G. Edwards, C. Taylor, Active appearance models, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 681–685.
- [26] V. Blanz, Face recognition based on a 3d morphable model, in: 7th International Conference on Automatic Face and Gesture Recognition, FGR '06, IEEE Computer Society, Washington, DC, USA 2006, pp. 617–624.
- [27] W. Press, S. Teukolsky, W. Vetterling, B. Flannery, Numerical Recipes in C, second ed., Cambridge University Press, Cambridge, UK, 1992.
- [28] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, CVPR 2005, vol. 1, 2005, pp. 947–954.
- [29] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çelikütan, B. Gökberk, B. Sankur, L. Akarun, Bosphorus database for 3d face analysis, in: Biometrics and Identity Management: First European Workshop, BIOID 2008, 2008, pp. 47–56.
- [30] T. Sim, S. Baker, M. Bsat, The cmu pose, illumination, and expression (pie) database of human faces, Technical Report CMU-RI-TR-01-02, Robotics Institute, Pittsburgh, PA, January 2001.
- [31] A.M. Martínez, Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class, IEEE Trans. Pattern Anal. Mach. Intell. 24 (6) (2002) 748–763.
- [32] R. Veldhuis, A. Bazen, W. Booi, A. Hendrikse, Hand-geometry recognition based on contour parameters, in: Proceedings of SPIE Biometric Technology for Human Identification II, Orlando, FL, USA, 2005, pp. 344–353.
- [33] M. Savvides, R. Abiantun, J. Heo, S. Park, C. Xie, B. Vijayakumar, Partial and holistic face recognition on FRGC-II data using support vector machine, in: Conference on Computer Vision and Pattern Recognition Workshop, 2006, CVPRW '06, 2006, pp. 48–48.

- [34] B.J. Boom, Q. Tao, L.J. Spreeuwers, R.N.J. Veldhuis, Combining illumination normalization methods for better face recognition, in: The 3rd IAPR/IEEE International Conference on Biometrics, Alghero, Italy, 2009.
- [35] G.A. Atkinson, M.L. Smith, L.N. Smith, A.R. Farooq, Facial geometry estimation using photometric stereo and profile views, in: The 3rd IAPR/IEEE International Conference on Biometrics, Alghero, Italy, 2009, pp. 1–11.

Bas Boom is a PhD student at the Signals and Systems Group of the University of Twente, in the field of biometrics. He received his MSc in Computer Science from Free University of Amsterdam. His research topic is face recognition in video surveillance environments.

Luuk Spreeuwers received both his MSc and PhD at the University of Twente. He worked as postdoc at ITC and at the Hungarian Academy of Sciences. He was a senior researcher Mindmaker, Ltd. He worked at University Medical Center in Utrecht and is now assistant professor at the University of Twente.

Raymond Veldhuis graduated from the University of Twente. He was a researcher at Philips Research, Eindhoven. He received the PhD degree from Nijmegen University. He worked at the IPO (Institute of Perception Research) Eindhoven. He is now an associate professor at the University of Twente.