Class-based Multiple Light Detection: An Application to Faces

Christos-Savvas Bouganis and Mike Brookes
Department of Electrical and Electronic Engineering
Imperial College of Science, Technology and Medicine
University of London
London, SW7 2BT, UK
{ccb98, mike.brookes}@imperial.ac.uk

Abstract

Multiple light detection approaches have limited applicability in real life scenarios due to the need of certain calibration objects. We propose a novel approach, the "class-based" image-based multiple light detection, that relaxes the above assumption to a "class" of calibration objects. We formulate it as follows: Given a set of images of objects belonging to the same class, similar 3D shape and reflectance properties, and illuminated under point light sources, the purpose is to determine the light distribution of a new object of that class. This paper concentrates on the class of human faces. Six algorithms are proposed and their performance is evaluated with real images. Experiments show that a good performance is achieved for up to three lights using a small database of faces.

1 Introduction

The problem of estimating illuminant directions and their intensities arises in a number of areas of computer vision. Early research on illuminant direction detection addressed the case of a single point light source [10, 20]. Recently, the focus has been moved to multiple lights scenarios. Many researchers have proposed algorithms that require a calibration object with specific properties to be part of the scene to detect the light distribution. Some of these methods use a Lambertian sphere [17, 19] and some others use mirror-like objects [4, 11]. The former methods give good results for point light sources, but they cannot estimate complex light distributions. The latter methods can detect complex illumination conditions in the expense of an interaction between the calibration object and the environment. In addition, both methods require interaction with the scene and the insertion of a calibration object at the time of capturing the image. To overcome the latter problem, Marschner and Greenberg [9], and Sato *et. al.* [13] have suggested algorithms that use any object of known geometry and reflectance properties. Even though, these algorithms have reported good performance, and have lifted the restriction of a specific calibration object, in real life scenarios, accurate geometry and Bidirectional Reflectance Distribution Function (BRDF) function of the object under inspection are usually unknown, limiting the applicability of these algorithms.

In this paper, we propose a statistical method that further relaxes the above assumptions. We concentrate on the problem of "class-based" image-based multiple light detection, and more specifically on the class of human faces. The problem is formulated as follows: Given a database containing images of different people under different illumination conditions and an image of a person that does not belong to database, the purpose is to find the light distribution that illuminates the new face. The main advantage of our approach is that it is image-based, thus it does not require a 3D geometric model of the object and an estimation of the BRDF function, but it is based entirely on the statistical properties of images of objects belonging to the same class.

2 Related work

Statistical approaches have been explored by many researchers for single light detection. Pentland [10] is the first who assumed a uniform distribution of object normals to detect the light direction. More recently, some methods have been proposed that use samples of faces under different light conditions and by applying kernel functions, estimate the light direction [2, 14]. However, these approaches work only for one-light scenarios.

Several statistical models for object appearance based on 2D images can be found in literature [16, 3]. However, they concentrate on estimating novel views of an object and not on multiple light detection. Closer to our work is that by Riklin-Ravin and Shashua [12] who proposed a method, the Quotient image, for class-based re-rendering. They propose an algorithm that estimates the object's appearance under novel light distributions. An extension of their approach for different poses can be found in the work of Stoschek [15]. Sim and Kanade [14] modelled the intensity probability of each point in an image of a face, given the light direction. The density functions for each pixel are estimated from the sample images. However, they assume that the density function of each pixel is independent of the others due to the dimension of the joint probability function of the pixels in the image. It is found that the algorithms of Sim and Kanade [14] and Riklin-Ravin and Shashua [12] have a poor performance if the novel light distribution becomes different from the initial one.

To the best of our knowledge, nobody has tried to detect multiple lights using the principle that objects belonging to the same class have similar appearance under the same light conditions. Some approaches [9, 13] for multiple light detection can be easily extended to class-based methods using a generalized 3D head model of a face. The disadvantage is that they require the 3D information to be recovered, and an image synthesis step should be applied, for which the BRDF function is required. In some areas of computer vision such as in face recognition, the BRDF of a face is approximated by a Lambertian model. However, for multiple light detection, specularities give important information for the light distribution and should not be discarded.

Instead, we propose the use of statistical methods to avoid the expensive step of 3D reconstruction of faces and the difficult step of rendering the new face under novel illumination. In this paper we propose and evaluate six algorithms that can be applied to derive a statistical description of the face appearance under novel illumination conditions. Our aim is to take into account the features of the face that maximize the information for the light distribution, and at the same time minimize the variance between images that belong to different people but are illuminated from the same direction. All the proposed methods are based on the fact that an image of a face under multiple light sources can be expressed as the superposition of images that each correspond to a single light direction.

3 Proposed approaches

In the rest of the paper the following notation holds: R denotes the input image under unknown illumination distribution in vectorized form. V_i is a random vector drawn from the class of all possible faces illuminated from direction i. Its instances are denoted by V_i^k, where k corresponds to the k^{th} person in the database, and V_i^k is this person's image in vectorized form.

3.1 Image pre-processing

Initially, an image pre-processing step takes place whose aim is to reduce the variations of the images for different reasons other than the light variation. The images that belong to the same light direction class can exhibit variations due to different albedo, shape and expression. Assuming all the images have been taken under neutral expression, the variations due to different albedo and shape should be minimized. For the shape difference, a mask is estimated that masks out large variations of the intensity in the images that belong to the same light direction class. The images are segmented into blocks and for each block the "between-class" (S_B) and the "within-class" (S_W) scatter matrices are estimated. We reject blocks whose value of the following criterion [5] is below a threshold:

$$S = \mathrm{trace}(S_B)/\mathrm{trace}(S_W) \tag{1}$$

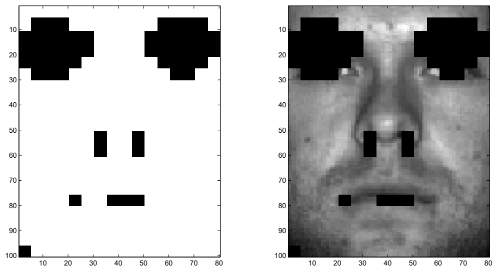
There is a trade-off between the number of retained blocks and the information that the image contains for the light distribution. The more blocks are selected, the more the available information for the light distribution is, but also the larger the variation within the same light direction class. Fig. 1 shows the mask that is applied to the images in our experiments. The resulting masked images are scaled according to the front-illuminated images to minimize the effect of different albedo. All images of a particular person in the database are scaled by the same factor.

3.2 Unknown face

The main idea behind the proposed algorithms is that the effects of multiple lights can be superimposed. Thus, in the general case where we do not have any knowledge about the new face, our goal is to minimize the expected mean square error:

$$T = \min_{a_i} E\left\{(R - \sum_{i=1}^{N} a_i V_i)^T (R - \sum_{i=1}^{N} a_i V_i)\right\} \quad , a_i \geq 0 \tag{2}$$

where R is the new image, V_i is a random vector and corresponds to the light direction i, and N is the number of the possible light direction classes. $E\{\cdot\}$ denotes expectation. After some algebraic manipulation:



(a) (b)

Figure 1: (a) Image mask with 83% of the pixels valid. The parts of the image that are masked out (eyes and eyebrows) have the most variation. Note that the pixels on the part of the nose are valid since the specular reflections there give information about the light distribution. (b) An image of the database with the applied mask.

$$T = \min_{a_i} \left\{ R^T R - 2R^T \sum_{i=1}^{N} a_i E\{V_i\} + \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j E\{V_i^T V_j\} \right\} \quad , a_i \geq 0 \tag{3}$$

3.2.1 Approach 1: Small sample size

In the case where the database is small, estimation of $E\{V_i^T V_j\}$ is not accurate. However, by relaxing the assumptions and assuming that the random variables V_i and V_j are independent we get:

$$E\{V_i^T V_j\} = E\{V_i^T\} E\{V_j\} = E\{V_i\}^T E\{V_j\} \tag{4}$$

Then, (3) can be written in the form:

$$T = \min_{a_i} \left\{ (R - \sum_{i=1}^{N} a_i E\{V_i\})^T (R - \sum_{i=1}^{N} a_i E\{V_i\}) \right\} \quad , a_i \geq 0 \tag{5}$$

For computational efficiency, $E\{V_i\}$ can be estimated in a lower dimensional space by using the KLT transform. The new image R is projected to the same space before the minimization of (5). It should be noted that estimation of the basis of the reduced space is done by subtracting the mean image from the data. However, when the images are projected to that space, the mean image is not subtracted.

3.2.2 Approach 2: Large sample size

The assumption of the random variable independence does not hold in reality, since the image of a person's face under a specific light is correlated to the image of the same person that is illuminated by a different light. In the case of a large database, the accurate estimation of $E\{V_i^T V_j\}$ would be possible. Thus (6) holds, where K is the number of images in each class and V_i^k denotes the image of the k^{th} person in the i^{th} class.

$$E\{V_i^T V_j\} \approx \frac{1}{K}\sum_{k=1}^{K}(V_i^k)^T V_j^k \tag{6}$$

It is assumed that each person in the database has one image in every class. The minimum of (3) is estimated by taking the first derivatives and setting them to zero, with non-negative constraints for the a_i parameters (7).

$$\frac{\partial T}{\partial a_i} = -2R^T E\{V_i\} + 2\sum_{j=1}^{N} a_j E\{V_i^T V_j\} \quad , \forall i \tag{7}$$

Setting (7) to zero, a system of linear equations with non-negative constraints is obtained. It should be also noted that the second partial derivatives are given by (8), because $E\{V_i^T V_j\}$ is always positive in the image space since all the V_i vectors have positive elements.

$$\frac{\partial^2 T}{\partial a_i \partial a_j} = 2E\{V_i^T V_j\} \geq 0 \quad , \forall i, j \tag{8}$$

Thus (7) has only one minimum. Due to the large dimension of the image space, a KLT transform is applied to reduce the space and extract only the necessary information. Again, the reference image is projected to that space before the application of (7). Also (8) still holds. Assuming B is an orthonormal basis for the new space and $\hat{V}_i = B^T V_i$, $\hat{V}_j = B^T V_j$ are the projected vectors, then (9) holds.

$$E\{\hat{V}_i^T \hat{V}_j\} = E\{(B^T V_i)^T B^T V_j\} = E\{V_i^T B B^T V_j\} = E\{V_i^T \widetilde{V}_j\} \approx E\{V_i^T V_j\} \tag{9}$$

where $\widetilde{V}_j$ is the reconstruction of V_j after its projection to the new space. If an adequate number of eigenvectors is used to represent the new space then $\widetilde{V}_j \approx V_j$, yielding (9).

3.2.3 Linear Discriminant approaches

Linear Discriminant approaches have been widely used in face recognition and detection [1]. Instead of finding the projection that contains the most variation (PCA), these approaches find a projection that best separates the classes. More specifically, LDA approaches try to minimize (10), where S_W and S_B are the within-class and between-class scatter matrix respectively [5].

$$J(\mathbf{w}) = \frac{\mathbf{w}^T S_W \mathbf{w}}{\mathbf{w}^T S_B \mathbf{w}} \tag{10}$$

However, due to the large dimension of the image space and the small number of samples, the LDA approach usually overfits the data. Several approaches have been proposed to avoid the overfitting, the most common one being to project the data to a lower dimensional space through PCA and perform the LDA in the new space [1]. Several other variations exist that combine these two steps such as Direct LDA [18] and Enhanced Fisher Linear Discriminant [8] methods. Others seek a compromise between the PCA and the LDA, such as the Principal Discriminant Variate (PDV) [7] approach.

The common LDA approach tries to minimize (10) by first diagonalizing the S_W matrix and removing the null space. However, the null space of S_W may contain useful information if the projection of the S_B matrix is not zero in that direction. In the Direct LDA approach [18], the S_B matrix is diagonalized first. Then the S_W matrix is diagonalized, and only the most discriminating dimensions are selected.

The Enhanced Fisher Linear Discriminant method [8] has been introduced to improve the generalization capabilities of the standard FLD methods. It balances the need for the selected eigenvectors to account for most of the variation of the initial data, and the requirement that the eigenvalues of the "within" scatter matrix are not too small.

The Principal Discriminant Variate (PDV) [7] method seeks a compromise between PCA and Fisher's LDA method, such that the stability of the LDA method is improved without loss in the discrimination power of the algorithm. In the PDV method criterion (11) is maximized, where I is the identity matrix, S_T is the total scatter matrix and $\lambda \in [0,1]$ determines the the balance between the PCA and the LDA criterion.

$$J(\mathbf{w}) = \frac{\mathbf{w}^T[\lambda S_B + (1-\lambda)S_T]\mathbf{w}}{\mathbf{w}^T[\lambda S_T + (1-\lambda)I]\mathbf{w}} \tag{11}$$

In the case where $\lambda = 1$, (11) becomes the LDA criterion, and in the case $\lambda = 0$, (11) becomes the PCA criterion. The optimum value for the λ parameter can be estimated through analysis of the images in the database. Due to the high dimension of the images in the image-space, a KLT approach for dimensionality reduction is applied before the PDV method.

We applied the above mentioned LDA approaches to extract the most discriminating information for multiple light detection from the images. Then, minimization of (5) is applied by substituting the $E\{V_i\}$ term with the mean image of each light direction class in the projected space.

3.3 Known face (front view available)

There are cases where the person under investigation is known and a front-illuminated image of the same person is available. A technique to generate new illuminated views of the same person can be applied. Several approaches can be found in literature for "class-based" re-rendering such as Sim and Kanade's approach [14] and the Quotient image approach [12]. In our case, an approach closer to the Quotient image is followed. Having a database of people illuminated from different directions, a front-illuminated image of a new person can be expressed as a linear combination of the other front-illuminated images in the database. Then, a prediction of the new face under different illumination conditions can be constructed as a linear combination of the images in the database that correspond to the same light direction, using coefficients that are derived from the front-illuminated case. Thus, assuming R_F is the front-illuminated image (in vector form) of the new person, and V_1^k (class 1 corresponds to front-illuminated images) is the front-illuminated image of person k in the database, the coefficients are estimated as (12).

$$c = \left[V_1^1, V_1^2, \ldots, V_1^K\right]^{-1} R_F \tag{12}$$

The new rendering images R_i of the same person under light direction i are estimated as (13).

$$R_i = [V_i^1, V_i^2, \ldots, V_i^K]c \tag{13}$$

Estimating these images, and using them as reference vectors for each light direction, the light distribution can be estimated by minimizing (14).

$$T = \min_{a_i} \left\{ (R - \sum_{i=1}^{N} a_i R_i)^T (R - \sum_{i=1}^{N} a_i R_i) \right\} \quad , a_i \geq 0 \tag{14}$$

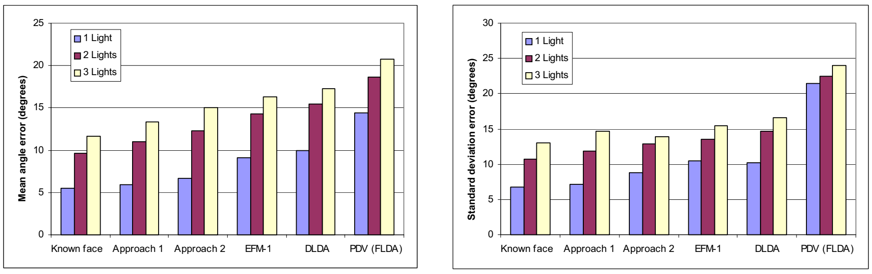
4 Experiments

Experiments with real images are performed to investigate the performance of the above algorithms. The images are taken from the Yale Face Database B [6]. This contains 10 people, each illuminated from 64 different directions. The light directions are the same for each person. In the experiments, two of the people are excluded. The first one is excluded due to not accurate calibration of the lights by [6]. The second one is excluded because it is the only female in the database and a statistical model is thus difficult to derive. The images are cropped around their center of gravity to the size of 200 x 160 pixels, resized to 100 x 80 by using bicubic interpolation, and a sparse correspondence using the coordinated of the eyes, nose, and mouth is applied.

The experiments are performed using the leaving-one-out approach: in each case, all the 64 images of one person are excluded from the database, and the training is performed with the remaining 7 people. 100 experiments are performed for each person and for each light configuration. The algorithms are tested with up to 3 lights. In total 2400 experiments are performed. Only those detected lights with intensity more than a threshold are taken into account. In the experiments, the threshold is set to 0.2. Detected lights are grouped together, if the mean vector of them differs less than 25 degrees from each light vector in the set. It should be noted that the average angle between a light direction in the database and its closest neighbor is 17.5 degrees. The grouping is done by applying a greedy algorithm. In the first and second approaches, the images are projected to a lower dimension space that is defined by the mean images of the light direction classes.

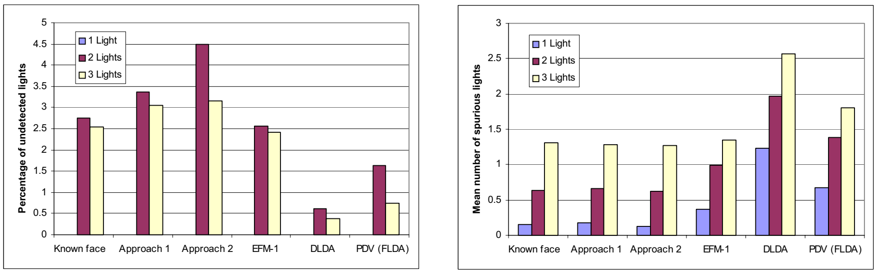
The algorithms are tested for the accuracy of the detected lights (Fig. 2a) and their variations in the results (Fig. 2b). The accuracy is defined as the angle between the detected light source vector and the corresponding true light source vector. Also, the percentage of undetected lights is estimated in all cases (Fig. 3a). Finally, the mean number of spurious lights for each light configuration is estimated (Fig. 3b).

From the experiments it can be concluded that the first approach which assumes random vectors independence gives better results than the approach which takes into account the correlation between the light directions. This is due to the small database; we believe that with a large database the latter method will outperform the former one. The algorithm with the extra information of the frontal view of the face gives the best results as expected. For the PDV method, different values for the λ parameter are tested and the best results are acquired for $\lambda = 0$ i.e. the PCA approach. For comparative reasons, we tested the PDV algorithm with $\lambda = 1$, which is the Fisher LDA approach (Fig. 2). Moreover, it can be concluded that the Enhanced Fisher Linear Discriminant method (EFM-1) gives the best results of all the LDA variations (excluded the PDV method for $\lambda = 0$). It should be noted that all the experiments are automatic without any user interaction.



(a) (b)

Figure 2: (a) Mean angle error in degrees. (b) Standard deviation angle error in degrees.



(a) (b)

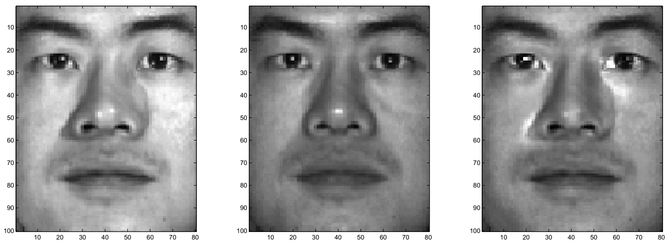
Figure 3: (a) Percentage of undetected lights. (b) Mean number of spurious lights.

5 Illumination compensation

A direct application of the proposed algorithms can be seen in image synthesis and Virtual Reality. The algorithm can be applied in scenes where a frontal view of a face is available in order to detect the light distribution.

Moreover, having estimated the light distribution, it is possible to compensate for changes in illumination and re-render the face under a different illumination condition. This also has an impact to those face recognition applications that have only one image for each person (i.e. front illuminated images only). In this case, any new image has to be normalized to the same light conditions before classification, to reduce the effects of lighting. Several approaches can be found in the literature, the most common ones being histogram equalization and linear fitting.

We propose an approach close to the Quotient image [12] to estimate the normalized face image under a target light distribution. Using the first approach, the algorithm estimates the light directions and their relative strength for a new image. Assuming an ideal class [12] (the 3D shape is the same) and a Lambertian model for the faces, the ratio of two images illuminated with different lights is independent from the albedo. Based on this, the image R of the face of a new person under any illumination can be converted to a front-illuminated image R_F by (15) where $E\{V_1\}$ is the mean front-illuminated image.



(a) (b) (c)

Figure 4: (a) Input image illuminated by 2 lights. (b) Front illuminated image (target image). (c) Illumination compensated image with the proposed approach.

$$R_F = \frac{E\{V_1\}}{\sum\limits_{i:a_i>0} a_i E\{V_i\}} \cdot R \tag{15}$$

The operations in (15) are pixel-wise and performed in the image space. The case where the denominator is near to zero requires special treatment. If the denominator for a pixel is close to zero, it means that there is not enough information available for the algorithm to retrieve the intensity of this pixel under front-illumination. Equation (15) can be also applied for any target illumination distribution by using the preferred distribution instead of $E\{V_1\}$. Fig. 4 shows an example of the initial image, the real front-illuminated image and the illumination compensated for front light illumination.

6 Conclusion and Future Work

In this paper we propose a novel approach for the multiple light source detection problem based only on statistical information from an object's class. The class under investigation is the class of human faces. Six algorithms are proposed and experiments on real images show that our approach can detect up to three lights to within 14 degrees mean error, without any user interaction.

Image synthesis and face recognition are two areas with immediate application of the proposed approach. In image synthesis, our approach can be used for estimating the light directions if a person is present in the scene and faces the camera. For face recognition applications, the illumination compensated images can be used to overcome the problem of illumination changes.

Future work involves investigation of how synthetic images can be used to extend the database in order to increase the accuracy. Moreover, we plan to exploit the impact of our results to face recognition algorithms. Preliminary results show that the application of the proposed algorithm in the illumination compensation step of face recognition algorithms outperforms the histogram equalization method. The improvement in the recognition rate is greater than 10%. Finally, we want to extend our algorithm to faces with different facial expressions and other object classes.

References

[1] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 19(7):711–720, July 1997.

[2] R. Brunelli. Estimation of pose and illuminant direction for face processing. *Image and Vision Computing*, 15(10):741–748, October 1997.

[3] T. Cootes, K. Walker, and C. Taylor. View-based active appearance models. *4th Int. Conf. on Automatic Face and Gesture Recognition*, pages 227–232, 2000.

[4] P.E. Debevec. Rendering synthetic objects into real scenes: Bridge traditional and image-based graphics with global illumination and high dynamic range photography. *SIGGRAPH 98*, pages 189–198, July 1998.

[5] K. Fukunaga. *Introduction to statistical pattern recognition*. Academic Press, second edition edition, 1990.

[6] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23(6):643–660, 2001.

[7] J.-H. Jiang, R. Tsenkova, and Y. Ozaki. Principal discriminant variate method for classification of multicollinear data: Principle and applications. *Analytic Sciences*, 17 Supplement:i471–i474, 2001.

[8] C. Liu and H. Wechsler. Enhanced fisher linear discriminant models for face recognition. *Int. Conf. on Pattern Recognition*, August 1998.

[9] S.R. Marschner and D.P. Greenberg. Inverse lighting for photography. *Fifth Color Imaging Conference, Society for Imaging Science and Technology*, 1997.

[10] A.P. Pentland. Finding the illuminant direction. *J. Opt. Soc. Am.*, 72:448–455, 1982.

[11] M.W. Powell, S. Sarkar, and D. Goldgof. A simple strategy for calibrating the geometry of light sources. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23(9):1022–1027, September 2001.

[12] T. Riklin-Raviv and A. Shashua. The quotient image: class-based re-rendering and recognition with varying illuminations. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23(2):129–131, Feb 2001.

[13] I. Sato, Y. Sato, and K. Ikeuchi. Illumination distribution from shadows. *Computer Vision and Pattern Recognition*, 1:306–312, 1999.

[14] T. Sim and T. Kanade. Illuminating the face. Technical Report CMU-RI-TR-01-31, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, Sept. 2001.

[15] A. Stoschek. Image-based re-rendering of faces for continuous pose and illumination directions. *Computer Vision and Pattern Recognition*, 1(582-587), 2000.

[16] T. Vetter and T. Poggio. Linear object classes and image synthesis from a single example image. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 19(7):733–742, July 1997.

[17] Y. Wang and D. Samaras. Estimation of multiple illuminants from a single image of arbitrary known geometry. *ECCV 2002*, 3:272–288, 2002.

[18] J. Yang, H. Yu, and W. Kunz. An Efficient LDA Algorithm for Face Recognition. *ACM Trans. on Design Automation of Electronic Systems*, 2000.

[19] Y. Zhang and Y.-H. Yang. Illuminant direction determination for multiple light sources. *Computer Vision and Pattern Recognition*, 1:269–276, 2000.

[20] Q. Zheng and R. Chellappa. Estimation of illuminant direction, albedo, and shape from shading. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 13(7):680–702, July 1991.