Statistical multiple light source detection

C.-S. Bouganis and M. Brookes

Abstract: Multiple light source detection has many applications in image synthesis and augmented reality. Current techniques can provide accurate results but have limited applicability in real-life scenarios where interaction with the scene is not possible. The authors provide a statistical framework for multiple light source detection that relies on the common features of objects belonging to a particular class and illustrate it using the class of human faces. Experiments with real data demonstrate that a light distribution with up to three light sources can be detected within 13° mean error. Application of the proposed framework to the problem of 3D reconstruction from multiple images under arbitrary lighting demonstrates the effectiveness of the framework compared with current techniques.

1 Introduction

Multiple light source detection, the process of determining the light distribution around an object of interest, is an active research field in computer vision that attracts many researchers. Information about the light distribution around an object of interest can be used to reduce the ambiguities concerning the 3D shape of the same object or of surrounding objects, and to permit the seamless insertion of artificial or real objects in the scene by illuminating them under the same light conditions. Applications can be found in image synthesis and augmented reality. Current techniques for estimating the illumination distribution require a mirror-like object to be inserted in the scene [1] or a specific object to be part of the scene [2]. The above methods provide good results, but for real-life scenarios their applicability is limited. Recently, researchers have focused on estimating the light distribution of a scene using an object that is likely to be part of the scene. In the case where a human face is in the scene Tsumura *et al.* [3] used the human eye as a light probe to estimate the positions of point light sources in the scene; Nishino and Nayar [4] present a methodology that extracts a dense illumination distribution from an image of an eye.

In this work, the limitation of a specific object to be part of the scene for multiple light detection is removed and it is only required that an object of a specific class be part of the scene. This has a negative impact on the accuracy of the light distribution estimation and limits the complexity of the light distributions that can be estimated. However, it allows the algorithm to be applied to more general real-life scenarios than the algorithms that require a specific object to be part of the scene. The class under investigation is the class of human faces. We present a statistical model for the estimation of the light distribution in the scene given a facial image of an unknown person. Experiments demonstrate the good accuracy that can be achieved in the light distribution estimation when up to three lights illuminate the face under investigation.

The paper is organised as follows. Section 2 describes previous work on the problem of multiple light detection. Section 3 introduces the notation that is used in the paper. The proposed framework for multiple light detection is described in Section 4. Section 5 evaluates the proposed method for multiple light detection where its application on the 3D reconstruction problem is described in Section 6. Finally, Section 7 concludes the paper.

2 Related work

Statistical approaches have been explored by many researchers for single light detection. Pentland [5] is the first who made the assumption that the change in the surface normal is distributed isotropically to detect the light direction. More recently, methods have been proposed that use samples of faces under different light conditions and by applying kernel functions, estimate the light direction [6, 7]. However, these approaches work only for a single illumination light.

Several statistical models for object appearance based on 2D images can be found in the literature [7–10]. However, they concentrate on estimating novel views of an object and not on the multiple light detection problem.

To the best of authors' knowledge, nobody has tried to detect multiple lights exploiting the principle that objects belonging to the same class have similar appearance under the same light conditions. Some approaches [11, 12] for multiple light detection can be extended to class-based methods using a generalised 3D head model of a face. The disadvantage is that the 3D shape of the face has to be recovered and an image synthesis step must be performed, for which the BRDF function is required. Accurate retrieval of the 3D geometry of many faces is a tedious and costly job requiring special equipment [13]. The BRDF of a face has been approximated by a Lambertian reflectance model in face recognition applications, leading to good results. However, for multiple light detection, specularities give important information for the light distribution and should not be discarded as in the case under a Lambertian model. This implies that more complex reflectance models, like the Phong [14] and Torrance–Sparrow [15] reflectance models should be used even though they

doi:10.1049/iet-cvi:20065001

Paper first received 24th November 2006 and in revised form 23rd March 2007

The authors are with the Department of Electrical and Electronic Engineering, Imperial College London, South Kensington Campus, London SW7 2AZ, UK

E-mail: ccb98@ic.ac.uk

increase the computational complexity of the algorithm. Recently, Kemelmacher and Basri [16] proposed an algorithm that estimates the light distribution and the 3D model of a face using a single image and a generic 3D model of a face. The proposed algorithm is based on the spherical harmonics formulation for the reflectance function which assumes a Lambertian reflectance model and ignores the effect of cast shadows and inter-reflections. The authors demonstrate that their algorithm can detect a single point light source with a mean error of 11.3° and standard deviation of 6.2°. It should be noted that their algorithm finds the light distribution that illuminates the given face in terms of spherical harmonics approximation rather than as the actual point light sources. Zhang and Samaras [17] propose an algorithm for face recognition under variable lighting using harmonic image exemplars. Their algorithm estimates the light distribution around a novel face in terms of spherical harmonics, but it is based on the statistical information gathered from 3D models of faces. Debevec *et al*. [18] estimated the BRDF of a human face from a small set of viewpoints under dense sampling of incident illumination directions using a light stage. Their algorithm gives good results but acquiring the images is a costly and tedious operation.

This work proposes instead the use of statistical methods to avoid the expensive steps of 3D reconstruction of faces and of rendering them under novel illumination. We propose and evaluate five algorithms that can be applied to derive a statistical description of the human face appearance under novel illumination conditions. The aim is to take into account the features of the face that maximise the information for the light distribution, and at the same time minimise the variance between images that belong to different people but are illuminated by the same light distribution. The given problem can be seen as the converse of the face recognition problem under variable lighting [19]. In the proposed framework, we consider algorithms based on the Karhunen–Loeve transform (KLT) and on linear discriminant analysis (LDA).

In [20, 21], the authors propose the illumination cone as a representation for modelling the appearance of objects under different illumination conditions. In their paper, the authors are focused on the class of human faces, and under the assumptions that a human face can be modelled using a Lambertian reflectance model and can be approximated as a convex object, they show that the complete set of images of a person's face lies inside a convex cone in the R^n space, where n denotes the number of pixels in the image. They have also shown in [22] that any image in the cone can be expressed as the linear combination of a specific set of images, which they call extreme rays. These images are the result of illuminating the face under specific point lights at infinity. The current work is based on the observation that the appearance of different people under the same light conditions is more similar than the appearance of the same person under different light conditions. This implies that the distance between the points in the R^n space of images of people illuminated under the same light conditions is smaller than the distance between points that correspond to images from the same person illuminated under different light conditions. This indicates that the illumination cones of different people have the points that correspond to images with similar light conditions close to each other. The common geometric general characteristics of human faces enhances the similarity of these images. The proposed algorithms that are based on the KLT transform can be considered as building an illumination cone of the average face. Beyond that, the proposed framework does not make any assumptions on the reflectance properties of the object or does not require the objects to be convex. It is relied solely on statistical information in order to retrieve the light distribution around the object of interest.

3 Notation

This section introduces the notation that is used in the rest of the paper. A database with faces from M people is assumed to be available. The coordinates of the eyes, nose and the centre of the mouth are manually identified. All images are roughly aligned, scaled and cropped using these coordinates. The space of the possible light directions is discretised by selecting C light directions. Each person in the database has C images, where in each one, the person is illuminated by a single light source. The image of a person m that is illuminated by the light source c is denoted by $X_{m,c}$ and is in vectorised form. The images are grouped into C different sets according to the light direction.

4 Multiple light estimation

4.1 Preprocessing

In order to estimate the light distribution around a face, the variations in the appearance of the face due to any reason other than the light variation should be minimised. The images that belong to the same light direction 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 are remained to be minimised.

To reduce the effect of light intensities, camera gain, and varying albedo between people, the images are normalised using a single scale factor per person. The use of a single scale factor per person avoids distorting the relative brightness between images of the same person illuminated under different light directions. For this reason, the images in the database are scaled according to a reference person. In the current implementation, the person that minimises the sum of variations in brightness between the images that belong to the same class (light direction) along all light directions is selected as the reference person.

To minimise the variations due to shape difference between different people, a mask is estimated that masks out those parts of the image that give poor discrimination between light directions. The images are segmented into square blocks and for each block i the between-light-source $\mathbf{S}^i_\mathbf{B}$ (1) and the within-light-source $\mathbf{S}^i_\mathbf{W}$ (2) scatter matrices are estimated, where the superscript i is omitted below for clarity. $\mathbf{b}_{m,c}$ denotes the block of the image of person m that is illuminated by light direction c in vectorised form. $\bar{\mathbf{b}}_c$and $\bar{\mathbf{b}}$ are given by $\bar{\mathbf{b}}_c = (1/M)\sum_{m=1}^{M}\mathbf{b}_{m,c}$ and $\bar{\mathbf{b}} = (1/C)\sum_{c=1}^{C}\bar{\mathbf{b}}_c$.

$$\mathbf{S}_\mathbf{B} = \sum_{c=1}^{C}(\bar{\mathbf{b}}_c - \bar{\mathbf{b}})(\bar{\mathbf{b}}_c - \bar{\mathbf{b}})^\mathrm{T} \tag{1}$$

$$\mathbf{S}_\mathbf{W} = \sum_{c=1}^{C}\sum_{m=1}^{M}(\mathbf{b}_{m,c} - \bar{\mathbf{b}}_c)(\mathbf{b}_{m,c} - \bar{\mathbf{b}}_c)^\mathrm{T} \tag{2}$$

From [23], the criterion $J = \mathrm{trace}(\mathbf{S_B})/\mathrm{trace}(\mathbf{S_W})$ can be used to discriminate between blocks that exhibit small intensity variation among images with the same light source imposing at the same time a restriction on these blocks to give adequate information about the light distribution, and blocks that do not have this property. The size of the block is selected to be 5 × 5 pixels for image size 100 × 80, which contains enough information about intensity variability and at the same time allows for fine masking of the image features. The desired image blocks are the ones that give high values to the above criterion. Fig. 1 illustrates the selected mask. It should be noted that the YALE Database B [22] was used to determine the illustrated mask, where the subjects do not wear glasses or have beards or moustaches. However, as it is demonstrated in the evaluation section, the proposed framework is robust and copes with cases where the subject wears glasses or has a beard or a moustache.

4.2 Statistical framework

The underlying idea of the proposed framework is that when a face is illuminated by multiple lights, the final effect can be modelled as the superposition of the effects of single point light sources [24]. Note that any extended light sources can be approximated as a set of point light sources [25]. We denote by V_c the random vector that represents all possible images of faces in vectorised form, that are illuminated by the light direction c. That is, V_c is a vector of random variables, each of which models the intensity of a pixel in the image. The aim is to find a linear combination of these vectors, imposing non-negative constraints to the coefficients, that can approximate an input image Y that is illuminated by multiple lights. The coefficients of the linear combination give the relative strengths of the light sources. The aim is to minimise the expected mean square error of the reconstruction which is given by

$$\operatorname{argmin}_{a_1,\, a_2,\, \dots,\, a_C} E\left\{\left(Y - \sum_{c=1}^{C} a_c V_c\right)^{\mathrm{T}} \left(Y - \sum_{c=1}^{C} a_c V_c\right)\right\}, \quad a_c \geq 0 \tag{3}$$

where $E\{\cdot\}$ denotes the expectation. After some algebraic manipulation (3) becomes

$$\operatorname{argmin}_{a_1,a_2,\dots,a_C} \left\{ \begin{array}{l} Y^{\mathrm{T}}Y - 2Y^{\mathrm{T}} \sum_{c=1}^{C} a_c E\{V_c\} \\ + \sum_{c=1}^{C} \sum_{j=1}^{C} a_c a_j E\{V_c^{\mathrm{T}} V_j\} \end{array} \right\}, \quad a_c \geq 0 \tag{4}$$

The special case where the identity of the person under investigation is known and a front illuminated image of him/her is available is considered in [26]. In this work, we focus on the case where there is no available information about the person under investigation. The presented framework can be easily extended to estimate the ambient illumination in the scene by adding an extra random vector V_a which models the appearance of faces under ambient light illumination conditions.

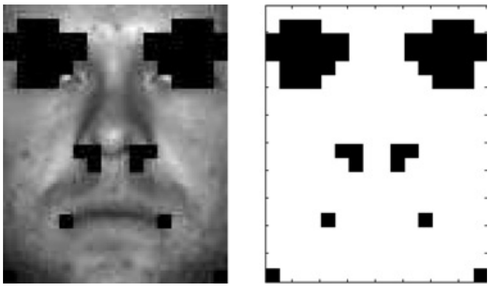

Fig. 1 *The selected mask*

Parts of the image with the highest information about the person's identity have been masked out. However, even though the top part of the nose is a feature with adequate variability between people, the specularities that it exhibits give adequate information about the light distribution and should not be masked out

By minimising (4), the relative intensities, a_c, of the used set of light directions, C, are obtained. From this set, only the lights with intensity exceeding a certain threshold T_{I} are taken into account. Then, the detected lights are grouped together, forming groups where the mean vector of the group differs less than a predefined threshold T_{G} from each light vector belonging in the same group, permitting detection of a light source with direction that does not belong in the original set of light directions. The thresholds T_{I} and T_{G} trade the accuracy of light detection with the number of detected and spurious lights.

Five approaches are described below: two using the (KLT) and three using variants of LDA.

4.2.1 KLT-L: large sample size: The light distribution around the face under inspection is estimated from (4) using a constraint minimisation method. The first partial derivatives of (4) are given in (5) and the second partial derivatives in (6).

$$\frac{\partial F}{\partial a_c} = -2Y^{\mathrm{T}} E\{V_c\} + 2\sum_{j=1}^{C} a_j E\{V_c^{\mathrm{T}} V_j\} \tag{5}$$

$$\frac{\partial^2 F}{\partial a_c \partial a_j} = 2E\{V_c^{\mathrm{T}} V_j\} \tag{6}$$

The second partial derivatives are always non-negative, because $E\{V_c^{\mathrm{T}} V_j\}$ is non-negative in the image space since all the V_i vectors have non-negative elements which are the intensities of the pixels. The second derivative can be equal to zero only in the case where one of the lights c or j do not illuminate the face at all or in the case where they illuminate non-overlapping regions of the face. Thus, the minimisation of (4) is a convex problem which has only one global minimum.

Owing to the large dimension of the image space, a KLT transform [27] is applied to reduce the space. The basis of the space is defined using the mean image $E\{V_c\}$ of each class c. The image with unknown illumination Y is projected to the same space before the minimisation of (4). Estimation of the basis of the reduced space is performed using data from which the mean image has been subtracted in order to find the directions with the maximum variance leading to the most compact representative sub-space. However, it should be noted that when the data are projected to that space the mean image is not subtracted since this would add extra terms in (4).

Assuming $\mathbf{B}$ is an orthonormal basis for the new space and $\tilde{V}_i = \mathbf{B}^{\mathrm{T}} V_c$, $\tilde{V}_j = \mathbf{B}^{\mathrm{T}} V_j$ are the projected vectors, then $E\{\tilde{V}_c^{\mathrm{T}} \tilde{V}_j\} = E\{\tilde{V}_c^{\mathrm{T}}(\mathbf{B}\mathbf{B}^{\mathrm{T}} V_j)\}$. We note that the second derivative is still positive after projection. The expression $\mathbf{B}\mathbf{B}^{\mathrm{T}} V_j$ is the reconstructed V_j image after it is projected to the new subspace and recovered. The expected mean square error between the $\mathbf{B}\mathbf{B}^{\mathrm{T}} V_j$ and V_j is given by the sum of squares of the eigenvalues of the eigenvectors that are not taken into account in the basis of the space. If adequate eigenvectors are taken into account, the expected mean square error is small and $\mathbf{B}\mathbf{B}^{\mathrm{T}} V_j \simeq V_j$. Thus, $E\{\tilde{V}_c^{\mathrm{T}} \tilde{V}_j\} \simeq E\{V_c^{\mathrm{T}} V_j\}$, which means that (6) remains non-negative in the reduced space, implying that the problem remains convex after the dimensionality reduction.

4.2.2 KLT-S: small sample size: In the case where the database is small, estimation of $E\{V_c^{\mathrm{T}} V_j\}$ in (4) is not accurate because of the high dimensionality of the V_c vector. However, by relaxing the assumptions and assuming that the appearance of a person when it is illuminated by two different light sources is independent, then the random variables V_c and V_j are independent. Moreover, it is assumed that $E\{V_c^{\mathrm{T}}V_c\} = E\{V_c\}^{\mathrm{T}}E\{V_c\}$ which clearly does not hold in reality. However, this assumption is necessary for the case of small sample size because of the fact that $E\{V_c^{\mathrm{T}}V_c\}$ cannot be estimated accurately enough. Then, (4) can be written in the form

$$\operatorname{argmin}_{a_1, a_2, \ldots, a_C} \left\{\left(Y - \sum_{c=1}^{C} a_c E\{V_c\}\right)^{\mathrm{T}}\left(Y - \sum_{c=1}^{C} a_c E\{V_c\}\right)\right\}, \quad a_c \geq 0 \tag{7}$$

and the $E\{V_c\}$ term can be approximated from the database of faces as $E\{V_c\} \simeq (1/M)\sum_{m=1}^{M} X_{\mathrm{m,c}}$.

For computational efficiency and performance enhancement, (7) is minimised in a lower dimensional space by applying the KLT transform.

4.2.3 Linear discriminant approaches: PDV, EFM-1, DLDA:

Linear discriminant analysis (LDA) provides an alternative to the KLT for selecting a subspace in which to minimise (7). LDA has been used in face recognition problem as a method that minimises the variability in the appearance of a person and at the same time maximises the variability in the appearance between different people [19]. The aim in this work is to estimate the features of the images that enhance the effects of different light positions and suppress variations of the images due to different people. We have evaluated three methods: the principal discriminant variate (PDV) [28], the enhanced Fisher linear discriminant (EFM-1) [29] and the direct LDA (DLDA) [30].

4.2.4 Principal discriminant variate (PDV):

In the case where the number of variables is larger than the number of samples, or where the variables suffer from multicollinearity, that is where there is a strong determinist relationship among some variables, most of the current discriminant methods fail. Jiang *et al.* [28] proposed the algorithm in order to address the problem of data multicollinearity. They argue that one of the main drawback of the Fisher's LDA is that in the case of multi-collinear data it tends to overfit the data, leading to meaningless discrimination when new data are encountered. Thus, they proposed to integrate the principal component analysis algorithm (PCA) and Fisher LDA (FLDA) in order to avoid the problem of over-fitting.

The PDV method is a compromise between the PCA and the FLDA. The PCA algorithm is responsible for inserting stability into the system, where the FPDA algorithm is responsible for the discrimination capabilities of the system. The criterion that is maximised is given by (8), where $\mathbf{S_T} = \mathbf{S_B} + \mathbf{S_W}$, and $\mathbf{I}$ is the identity matrix which has the same size as the $\boldsymbol{S_T}$.

$$Q = \frac{\mathbf{W}^{\mathrm{T}}[\lambda\mathbf{S_B} + (1-\lambda)\mathbf{S_T}]\mathbf{W}}{\mathbf{W}^{\mathrm{T}}[\lambda\mathbf{S_T} + (1-\lambda)\mathbf{I}]\mathbf{W}} \tag{8}$$

The parameter λ is responsible for controlling the influence between the two algorithms. It takes values from the range [0, 1] and includes as special cases the PCA when $\lambda = 0$ and the FLDA when $\lambda = 1$. The optimal value of the parameter λ is recovered through cross-validation [28]. Details about the procedure for computing the vectors $\mathbf{w}$ are given in [28].

4.2.5 Enhanced Fisher linear discriminant (EFM-1):

Liu and Wechsler [29] proposed the EFLD algorithm in order to improve the generalisation abilities of the FLDA transform. In order to reduce the dimensionality of the data, the authors apply first the PCA and then they proceed with their FLD type algorithm. The main idea behind the EFM-1 algorithm is to balance the number of principal components that are kept from the first stage of the PCA algorithm against the requirement that the eigenvalues of the within-scatter matrix in the reduced space are not too small.

This is based on the fact that the standard FLD procedure derives the discriminatory variables through simultaneously diagonalisation of $\mathbf{S_W}$ and $\mathbf{S_B}$ matrices. This is achieved by first whitening the within-scatter matrix, and then applying the transformation to the between-scatter matrix to diagonalise it. Thus, in the case where the within-scatter matrix encodes noise information, which implies that the matrix has very small eigenvalues, the FLD will fit the noise too, reducing the generalisation capabilities of the process. By reducing the number of the principal components that are taken into account from the PCA step, the noise is removed from the within-scatter matrix and hence the performance of the algorithm is enhanced.

4.2.6 Direct LDA:

Yang *et al.* [30] have proposed the DLDA algorithm that has better discriminatory power than current FLDA algorithms. Their argument is based on the fact this most FLDA algorithms apply first a PCA step, in order to reduce the dimensionality of the data, and then the FLDA algorithm for discrimination. By doing this, there is a possibility of losing discrimination power due to the PCA step. They propose the DLDA algorithm which performs the PCA step and the LDA step simultaneously.

The motivation behind their algorithm is that the PCA step removes the null spaces from both $\mathbf{S_W}$ and $\mathbf{S_B}$ matrices and this can lead to reduction of discriminatory information. If the projection of the $\mathbf{S_B}$ matrix to the null space of the within-scatter matrix $\mathbf{S_W}$ is not zero, then this null space has the most discriminatory power and should not be discarded. Thus, the null space of the $\mathbf{S_W}$ should not be discarded as it may contain useful information. However, the null space of $\mathbf{S_B}$ can be discarded [30]. Thus, in evaluating the discriminatory space, the algorithm starts by diagonalising the $\mathbf{S_B}$ matrix first, and then proceeds to diagonalise the $\mathbf{S_W}$ matrix.

Each of the aforementioned LDA approaches have been applied to extract the most discriminating information for multiple light detection from the images. The random vector V_C in (7) is replaced with the random vector that represents the projections of the image of faces that are illuminated by the light direction c and lies in the subspace that is determined by the LDA algorithms. The light distribution is estimated from (7) assuming independence between the effects of different light directions and also $E\{V_c^{\mathrm{T}}V_c\} = E\{V_c\}^{\mathrm{T}}E\{V_c\}$.

5 Performance evaluation

Experiments using real data are performed to evaluate the performance of the proposed algorithms. The Yale Face Database [22], which contains ten subjects illuminated by 64 light directions, and the PIE database [31], which contains 68 subjects illuminated by 21 light directions, are used to evaluate the performance of the algorithms. The images are cropped to remove the background and scaled to 100 × 80 pixels. Whenever the assessment of an algorithm parameter is required, it is performed using the

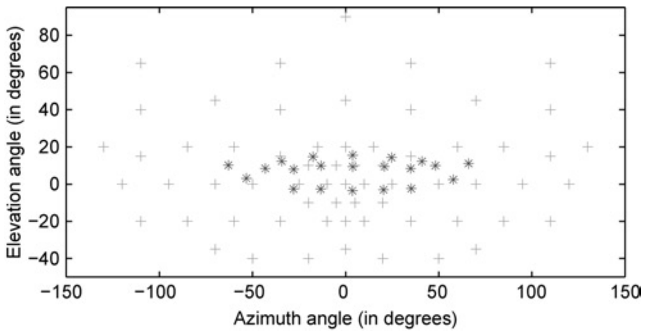

Fig. 2 *Positions of the light probes in YALE database (+) and PIE database (*)*

‘leaving-one-out’ approach [32] in which, all the images of the test subject under consideration are excluded from the database and the training is performed with the remaining subjects. The images of the faces are aligned using the coordinates of the eyes, nose and mouth by applying linear transformations. It should be noted that the light source positions in the YALE database are different from the light source positions in the PIE database. Figure Fig. 2 illustrates the positions of the light probes in the YALE and PIE databases. The figure demonstrates that the YALE database has more sparsely spaced light sources covering more of the illumination direction space than the PIE database, whereas the PIE database contains light sources placed around the centre of the viewer. Fig. 3 illustrates three of the images used for testing. In the first image, the person is illuminated by a single light, in the second image the person is illuminated by two lights, where in the third image there are three light sources in the scene.

5.1 Single light detection

The performance of the algorithms is first evaluated for the detection of a single light. The PIE database is used for training and evaluation purposes. In addition to the five algorithms proposed earlier, two reference algorithms from the literature are implemented. The first of these is the single light detection algorithm by Sim and Kanade [7]. The authors treated the light detection as a regression problem and used smooth kernel functions to detect the illumination in an image in the case of a single light. Using M training images X_m, where each one is illuminated by a single light direction s_m, they recover the light direction s in a new image Y using kernel regression as in (9),

$$s = \frac{\sum_{m=1}^{M} w_m s_m}{\sum_{m=1}^{M} w_m} \tag{9}$$

where

$$w_m = \exp\left\{-\frac{1}{2}\left(\frac{\|Y - X_m\|_2}{\sigma_m}\right)^2\right\}$$

The parameter σ_m controls the extent of the influence of each light direction s_m. In our implementation, we used $\sigma_1 = \sigma_2 = \cdots = \sigma_M$, and their value was fixed to the one that gives the best results using the PIE database with the ‘leaving-one-out’ approach.

The second reference algorithm uses a generic 3D model of a face in order to detect the illumination distribution in the case of a single light. Under this framework, the light source can be estimated by solving an over-determined system of equations as in (10), where $\mathbf{N}$ denotes the matrix with the normals for each point in the image scaled by the albedo, and Y is the input image in a vector form.

$$s = (\mathbf{N}^{\mathrm{T}}\mathbf{N})^{-1}\mathbf{N}^{\mathrm{T}}Y \tag{10}$$

Kee *et al.* [33] suggessted to use only certain parts of the face and especially the ones concentrated around the nose area. They demonstrated that by using the information around that area only, a better estimation of the light direction is achieved. However, since we are also interested in light detection estimation under extreme conditions, where the region around the nose may not be illuminated at all, the information from the whole face is taken into account. The generic 3D model of a face is estimated using the PIE database and by applying standard photometric stereo techniques.

Regarding the proposed algorithms, the PIE database is used for the training and the evaluation phases, using the leaving-one-out technique. We chose to use the same database for training and evaluation because of the considerable different sampling of the illumination direction space between the two databases. The threshold for the grouping of the lights sources, T_G, depends on the training database and is set to 7°, which is kept constant through all the experiments. Table 1 summarises the obtained results from the two reference algorithms and the five proposed algorithms. The table illustrates the achieved mean detection error and the standard deviation. The results demonstrate that the KLT-S, KLT-L and EFM-1 algorithms give results that

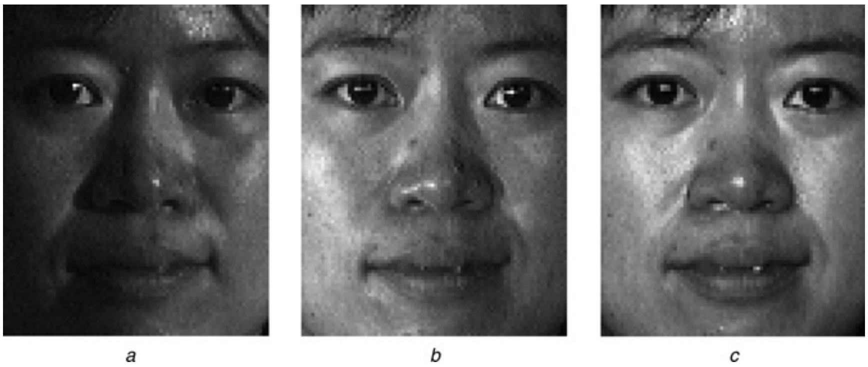

Fig. 3 *Three images used for the evaluation of the algorithms’ performance*

a Person illuminated by a single light source
b Person illuminated by two light sources
c Person illuminated by three light sources

Table 1: Single light source detection results using PIE database

	Mean	Standard deviation
Sim and Kanade [7]	5.50°	5.67°
Kee *et al.* [33]	8.47°	8.02°
KLT-S	9.61°	6.40°
KLT-L	11.39°	7.90°
PDV	13.37°	7.95°
EFM-1	10.89°	6.79°
DLDA	15.60°	13.11°

are close to that of the reference algorithms without using the prior knowledge that there is only a single light source. The DLDA algorithm gives the worst results.

5.2 Multiple light detection

The proposed algorithms are evaluated under multiple light illumination using light configurations with up to three lights. In total, 3000 experiments are performed. The true light intensities vary between 0.5 and 1.0. The minimum angle between any two lights that are used for testing is set to 30° and 15° for the YALE and PIE databases, respectively. The threshold for the grouping of the lights sources, T_G, depends on the training database and is set to 15° for the YALE database and to 7° for the PIE database. Both parameters are kept constant through all the experiments. The performance of the algorithms is illustrated for different values of the T_I threshold.

The performance of the proposed algorithms is compared against a reference algorithm that uses a generic 3D model of a face and spherical harmonics in order to render the face under different illumination conditions. In [34], the authors demonstrate that the reflectance functions of Lambertian objects produced by distant isotropic lights lie close to a 9D space. They show that in the case where an object is illuminated by M distant light sources, the image of the object can be expressed as

$$Y = \sum_{i=1}^{M} a_i \sum_{n=0}^{\infty} \sum_{m=-n}^{n} h_{nm}(\theta_i, \phi_i) b_{nm} \tag{11}$$

where a_i is the intensity of the light i, with direction (θ_i, ϕ_i), $h_{nm}(\theta_i, \phi_i)$ is a surface spherical harmonic function at point (θ_i, ϕ_i) and b_{nm} denotes the harmonic images of the object. θ denotes the polar coordinate with $\theta \in [0, \pi]$ and ϕ denotes the azimuthal coordinate with $\phi \in [0, 2\pi]$. The surface spherical harmonics are a set of functions that form an orthonormal basis on the surface of a sphere, denoted by h_{nm} for $n = 0, 1, 2, \ldots$ and $-n \le m \le n$ and given by (12), where P_{nm} are the Legendre polynomials.

$$h_{nm}(\theta, \phi) = \sqrt{\frac{(2n+1)}{4\pi}\frac{(n-|m|)!}{(n+|m|)!}} P_{nm}(\cos\theta) e^{im\phi} \tag{12}$$

The b_{nm} is called a harmonic image and is constructed by the product of the albedo with the harmonic reflectance r_{nm}. The harmonic reflectances are scaled versions of the surface harmonic functions, $r_{nm} = \sqrt{(4\pi/(2n+1))}k_n h_{nm}$, where k_n is the expression of the Lambertian kernel in harmonic series defined in (13).

$$k_n = \begin{cases} \dfrac{\sqrt{\pi}}{2} & n = 0 \\ \sqrt{\dfrac{\pi}{3}} & n = 1 \\ (-1)^{n/2+1} \dfrac{\sqrt{(2n+1)\pi}}{2^n (n-1)(n+2)} \dfrac{n!}{\frac{n}{2}!\frac{n}{2}!} & n \ge 2, \text{ even} \\ 0 & n \ge 2, \text{ odd} \end{cases} \tag{13}$$

By imposing non-negativity constraints in the parameters a_i, (11) can be used for multiple light detection. A similar method has been presented in [20], except that the images are rendered in the image space. The main difference is that the authors in [20] retrieve the 3D model of a face and render it under different illumination conditions in order to retrieve its illumination cone. It should be noted that the spherical harmonic approach does not model specularities or cast shadows, and that the approximation of a distant light source degrades as the angle between the normal of the object and the light direction becomes smaller [34]. Following the above framework and using the first nine spherical harmonics, a generic 3D model of a face was constructed and rendered using the same light sources as in the reference database. The linear system of equations (11) was used to detect multiple lights in the scene given a new image of a face. The minimisation of $\|Y - \sum_{i=1}^{C} a_i \sum_{n=0}^{2} \sum_{m=-n}^{n} h_{nm}(\theta_i, \phi_i) b_{nm}\|$ subject to $a_i \ge 0$ is performed in a C-dimensional subspace using the KLT algorithm, constituting the reference algorithm. Fig. 4 illustrates the two generic 3D models that are used in the experiments. Fig. 4*a* corresponds to the 3D model acquired from the YALE database, using standard photometric stereo techniques to calculate the normals of the model and the height derivation is based on [35]. The acquired 3D model using the PIE database is illustrated in Fig. 4*b*.

The first row of Fig. 5 depicts the achieved rendering of a face using a 4D spherical harmonics representation, the generic 3D model of a face and an average albedo map. For a direct comparison with real images, a set of real images acquired under the same light conditions are illustrated in the second row. It is clear that the spherical harmonic model can capture the general appearance of a face due to light variation, but it fails when cast shadows are dominant (last column). The spherical harmonics-based algorithm among with other illumination cone-based algorithms from the literature, have as target to approximate the appearance of a face under different light conditions using a low dimensional space. In these algorithms, specularities and cast shadows are not taken into account, since their nonlinearities demand a high-dimensional space. However, these features give information about the light distribution around the object of interest. In contrast, the proposed statistical framework uses the specularities and cast shadows in order to retrieve the light distribution.

Three sets of experiments are performed. In all sets of experiments, the leaving-one-out approach is employed, where all the images of the subject under consideration are excluded from the training phase of the algorithms. This approach allows us to evaluate the performance of the presented algorithms in the case of new subjects that are not included in the database. In the first set of experiments, the PIE database is used for training and evaluation of the algorithms. The graphs in Fig. 6 show the performance of the six algorithms as the threshold T_I is varied.

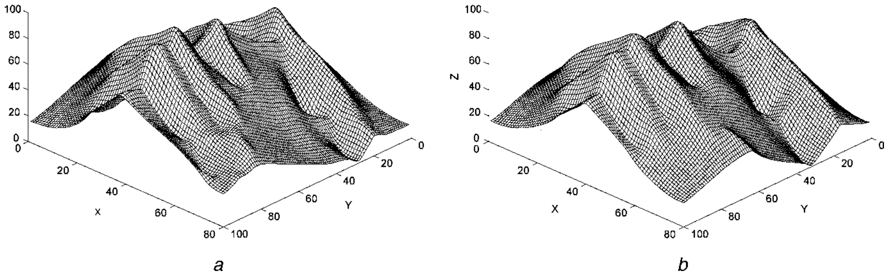

Fig. 4 *Generic 3D models of a face*

a Acquired from the YALE database
b Acquired from the PIE database

The smaller the value of T_I is, the more sensitive the algorithms become to detect a light. Thus, small values of T_I make the algorithms more prone to detect lights that are not actually in the scene, spurious lights. The larger the value of T_I is, the more robust the algorithms become to noise. Thus, this makes the algorithms more prone to miss a real light from detection, increasing the number of undetected lights. The three rows correspond to experiments with one, two and three lights, respectively. The first column shows the trade off between mean angular accuracy and the number of spurious light detections and the second column shows the trade off between mean angular accuracy and the number of undetected lights. The number of spurious lights is defined as the number of excess lights that the algorithms have detected relative to the real number of lights. The mean angular accuracy is calculated using the real lights and the corresponding detected lights. In the case where the number of detected lights is larger than the number of the real lights, a case of spurious lights, the mean angular accuracy is calculated using the subset of the detected lights that minimises the mean angular accuracy metric. In the case where the number of detected lights is less than the number of the real lights, a case of undetected lights, the mean angular accuracy is calculated using the subset of the real lights that minimises the mean angular accuracy metric. The results demonstrate that the KLT-S, KLT-L and EFM-1 algorithms give the best results, achieving accuracy in the detection around 12°, while at the same time keep the number of spurious and undetected lights low. The KLT-S algorithm gives slightly better results than the KLT-L and EFM-1 algorithms. The next best performing algorithms are the PDV and the DLDA, which have similar performance. In the evaluation of the PDV algorithm, the parameter λ is set to 10^{-4}, which gives the best results using the leaving-one-out approach. The algorithm that is based on spherical harmonics achieves the worst performance, achieving an accuracy in the detection in the range $16^{\circ}-20^{\circ}$. It should be noted that the mask that was used in all the experiments (Fig. 1) was determined by using the YALE database, which does not contain any subject that wears glasses or has a beard or a moustache. The performance of the proposed algorithms on the PIE database, which contains a large number of subjects that have the above features, demonstrates the robustness of the proposed framework. Moreover, the results demonstrate that the performance of the algorithms, apart from the algorithm that is based on the spherical harmonics, is not sensitive to the value of T_I parameter. The T_I parameter takes values from the interval [0.4, 1], where each point in the graphs corresponds to an increase of the parameter by 0.05. Varying the parameter, the accuracy in the detection remains almost constant, where the mean number of spurious lights and the mean number of undetected lights vary slowly.

A second set of experiments is performed, where the YALE database is used in the training phase and the PIE database is used for the evaluation phase. This scenario is more difficult, since the light directions in the two databases

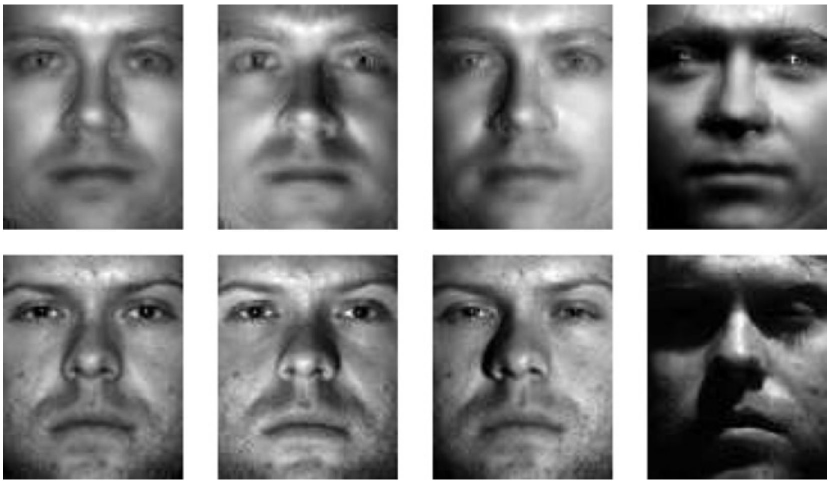

Fig. 5 *4D spherical harmonics representation compared with real images*

First row illustrates a rendering of a face using a 4D spherical harmonics representation, the generic 3D model of a face and an average albedo map. Second row depicts the real images of a face illuminated under the same light conditions. Spherical harmonics formulation can capture the general appearance of the face, but fails in the case where the cast shadows are dominant

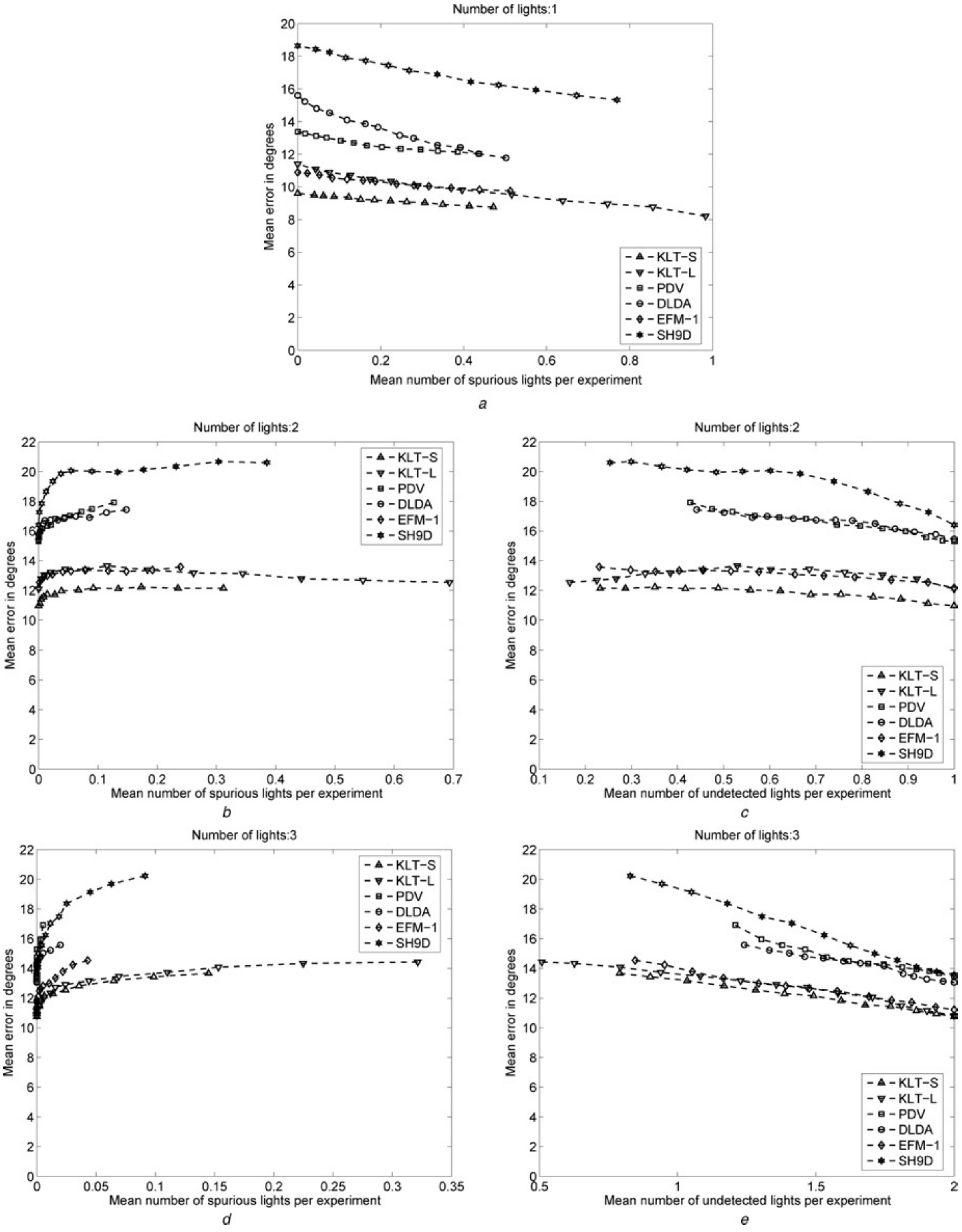

Fig. 6 *Results obtained using the PIE database for training and evaluation*

a b, d Performance of the algorithms regarding the accuracy of the detection against the number of spurious lights detected by the algorithm
c, *e* Performance of the algorithms regarding the accuracy in the detection and the number of undetected lights
a Case of a single light
b, *c* Case of two lights
d, *e* Case of three lights in the scene

differ greatly. The YALE database has more sparsely spaced light sources, where the PIE database contains light sources placed around the centre of the viewer. Thus, it is expected the obtained results to be inferior from the previous test set because of the coarser sampling of the illumination space in the YALE database (training). Using this test configuration, we can evaluate the performance of the algorithms in the case where the new subject does not exist in the database, and the light sources that illuminate the subject are not aligned with any of the light

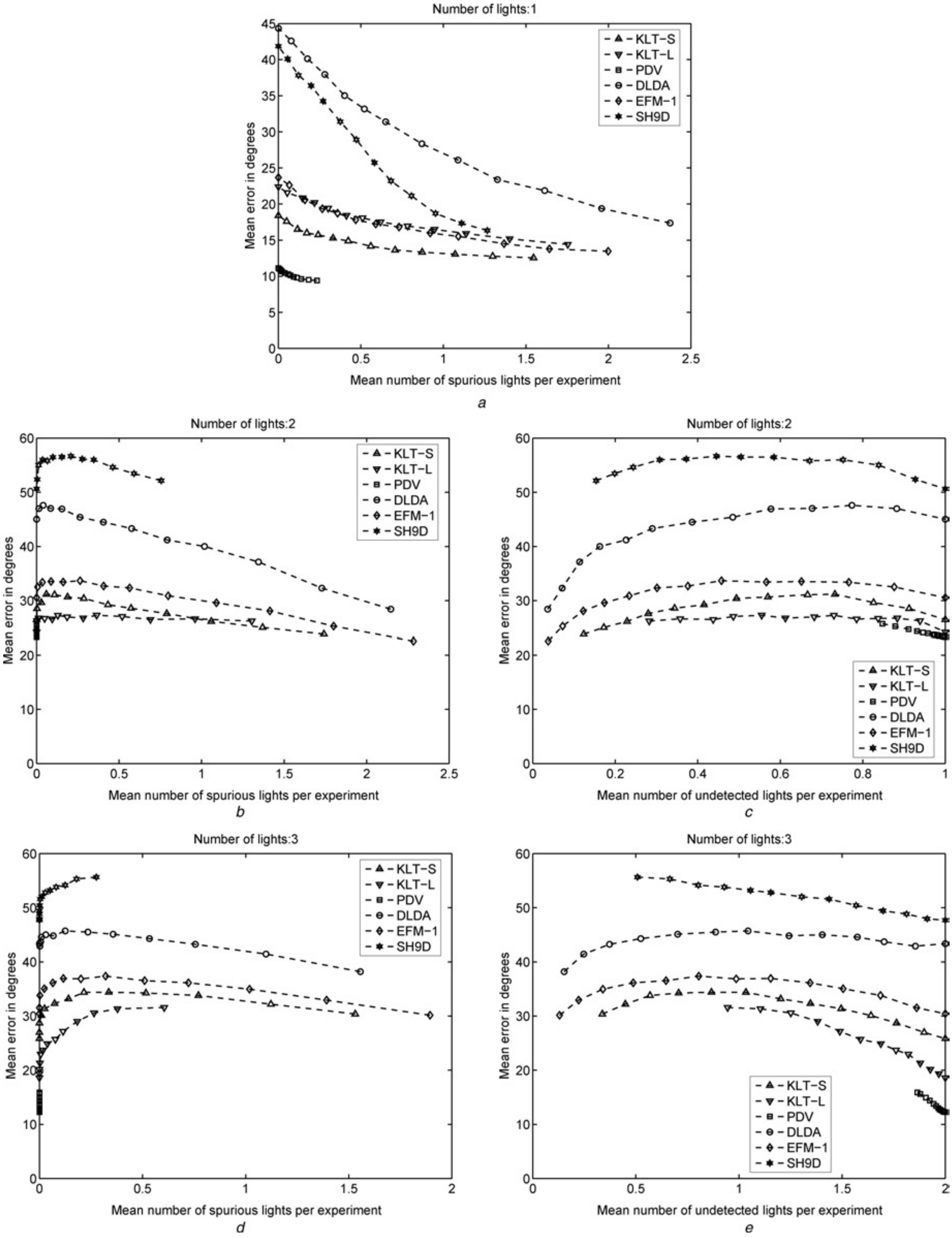

Fig. 7 *Results obtained using the YALE database for training and the PIE database for evaluation*

a, *b, d* Performance of the algorithms regarding the accuracy of the detection against the number of spurious lights detected by the algorithm
c, *e* Performance of the algorithms regarding the accuracy in the detection and the number of undetected lights
a Case of a single light
b, *c* Case of two lights
d, *e* Case of three lights in the scene

sources that exist in the database. Fig. 7 illustrates the obtained results with the same format as Fig. 6. The results demonstrate that the performance of the algorithms has been reduced except for the PDV algorithm (λ is set to 10^{-5}). The performance ranking of the remaining algorithms has remained the same. It should be noted that the output obtained by the algorithm based on the spherical harmonics is meaningless since the error in the light detection is around 50°. This is due to the fact that the spherical harmonics do not capture the appearance of the object when is illuminated under extreme light conditions, as is the case with the YALE database, where the effect of cast shadows is more dominant.

A final experiment is performed in order to investigate the effects of applying the mask of Fig. 1 to the performance of the algorithms. The experiment uses the YALE database for training and the PIE database for evaluation since this allows the investigation of the algorithms under realistic

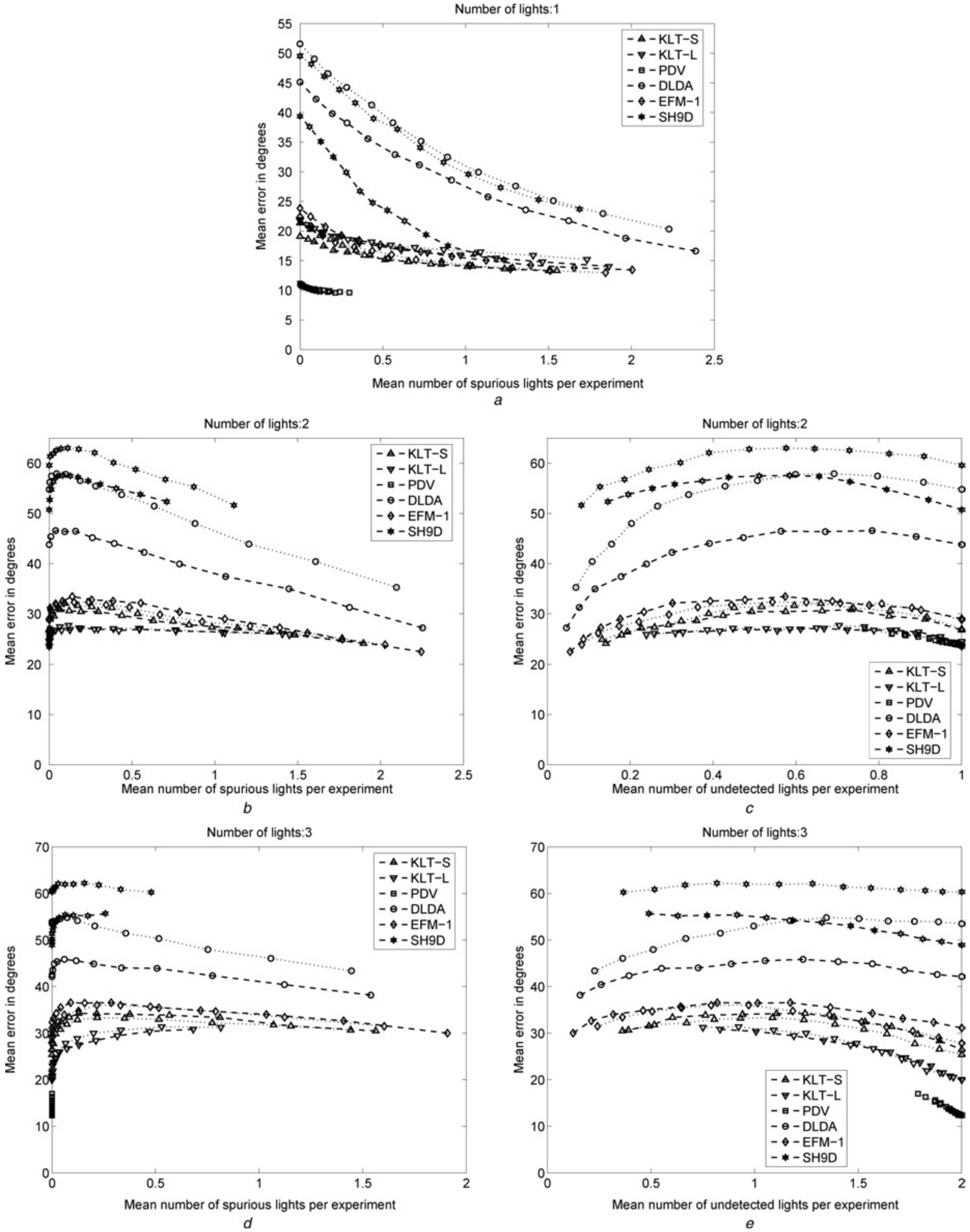

Fig. 8 *Results obtained using the YALE database for training and the PIE database for evaluation when the mask is applied (dashed lines) and when is not (dotted lines)*

a b, d Performance of the algorithms regarding the accuracy of the detection against the number of spurious lights detected by the algorithm
c, e Performance of the algorithms regarding the accuracy in the detection and the number of undetected lights
a Case of a single light
b, c Case of two lights
d, e Case of three lights in the scene

situations. Fig. 8 illustrates the obtained results with the same format as Fig. 6. The dashed lines correspond to the case where the mask has not been used, where the dotted lines correspond to the case where the mask has been applied to the images. The results demonstrate that the performance of the algorithms has been overall enhanced when the mask is applied. Only the performance of the EFM-1 algorithm degrades slightly (around one degree). Moreover, the results show that the gain that is achieved by applying the mask varies with the algorithm. The spherical harmonic-based algorithm and the DLDA algorithm are the ones that are affected the most by the mask application. The application of the mask in the remaining algorithms, apart of the EFM-1 algorithm, result in an enhancement of their performance by one to two degrees.

It should be noted that the performance of the proposed framework depends on the reflectance properties and geometric characteristics of the object class under investigation. In the case where the object class exhibits high specular reflectance characteristics and heavily cast shadows, the light configuration that illuminates the object is unique and is easy to find. Conversely, convex Lambertian objects make the multiple light source detection problem more difficult. The uniqueness of the solution in the latter case depends on the light configuration as has been reported in [2, 36].

6 Surface normal estimation under arbitrary lighting

This section demonstrates an application of the proposed framework in the problem of surface normal estimation under arbitrary lighting. It introduces two algorithms for surface normal recovery that are based on the proposed framework for multiple light detection and compares their performance with existing state-of-the-art algorithms.

Initial research for the estimation of the surface normals of an object was concentrated in the case of a single light illumination using only one image of the object [7, 37]. In this work, we focus in the case where the object is illuminated under arbitrary lighting and more than four images of the object are available in order to estimate the albedo and the normal at each point on the object's surface. Moreover, the relative position of the camera to the object is kept the same in all images, but the lighting varies. Basri and coworkers [34, 38] demonstrated that by using the spherical harmonics representation, it is possible to retrieve the 3D shape of an object up to a 4 × 4 linear ambiguity under arbitrary lighting, known as Lorentz transformation. Their framework is based on the assumption that the first four spherical harmonics capture most of the appearance of the object. Using the fact that the above space of spherical harmonics contains the albedo and a scaled version of the normals by the albedo, the authors propose a method based on singular value decomposition to retrieve the albedo and the normals of the object. It should be noted that the proposed framework by Schechner *et al.* [24] cannot be applied since the acquired images do not necessarily have common light directions.

Our proposed framework can also be used for shape recovery from many images under arbitrary lighting. Two algorithms are proposed: the first algorithm is based on the photometric stereo technique, whereas the second algorithm is based on spherical harmonics. The two algorithms differ in how they apply the estimated information about the light directions in the scene.

The photometric stereo-based algorithm first applies the KLT-L algorithm to retrieve the lights in the scene. Then, using a statistical model regarding which part of the face is illuminated under which light direction, the albedo and the normals of the object are recovered. The spherical harmonics-based algorithm also recovers the light distribution in the scene using the KLT-L algorithm. It then projects the recovered light directions in the first four spherical harmonics and by using (11), the normals of the object are recovered.

Experiments targeting the 3D reconstruction of faces are performed in order to assess the performance of the two proposed algorithms and the reference algorithm by Basri and cowerkers [34, 38]. It should be noted that their method recovers the normals of the object up to a linear 4 × 4 transformation, which blends the albedo and the normals. In order to retrieve the true normals, information about the albedo and normals of six points is required. They suggest a method that imposes a non-convex constraint and requires information about the normals and the albedo of only two points in the object in order to retrieve the true surface normals. The use of such constraint makes the whole formulation a non-convex optimisation problem, which means that a global solution is hard to find. In our implementation, we tried to find the global solution by initiating the optimisation process using different starting points. In contrast, our framework always finds the global solution. In the results, two numbers are reported. The first one corresponds to the case where the non-convex constraint is used simultaneously with information about four random points, two more points than actually needed in order to add robustness, in the object, and the second one refers to the case where the constraint is omitted and information about ten points, four more points in order to add robustness, is used allowing the determination of the 4 × 4 linear transformation. It should be noted that the results reported follow the application of a 3D rotation to the retrieved normals that minimises the mean square error between them and the actual normals of the object. Table 2 illustrates the obtained results. The numbers in round brackets correspond to the second case. The results show that the proposed algorithms produce considerably better results than the Basri *et al.* algorithm in the case where not enough information about the object is available and the non-convex constraint is applied. In the case where enough information about the object is available, the Basri *et al.* algorithm gives similar results to the proposed algorithm that is based on the light detection and photometric stereo. However, it should be noted that the second case is rarely applicable in reality. Moreover, the results demonstrate that in the case of the photometric stereo-based algorithm, the more lights present in the image, the less accurate the 3D reconstruction of the object becomes. We believe that this is due to the uncertainty about which point of the object is illuminated by which light sources in the scene, but further investigation is required. The algorithms based on the spherical harmonics representation do not exhibit a similar behaviour.

Fig. 9 depicts the obtained height maps for the first test configuration of Table 2. Fig. 9*a* is the ground truth obtained using photometric stereo techniques, Fig. 9*b* is the ground truth superimposed by the albedo of the face, where Figs. 9*c–h* depict the obtained results. The first column corresponds to the case where the non-convex constraint is used simultaneously with information regarding four random points of the object, whereas the second column corresponds to the case where the constraint is omitted and ten points are

Table 2: Reconstruction results under different light configurations

Configuration description	Algorithms		
	Basri *et al.*	Light detection (photometric)	Light detection (spherical harmonics)
Five images (one light per image)	53.21° (7.28°)	10.57° (8.33°)	30.16° (28.00°)
Four images (two lights per image)	51.77° (7.52°)	14.66° (12.92°)	31.74° (28.83°)
Four images (four to six lights per image)	60.39° (14.15°)	24.97° (21.20°)	20.96° (18.89°)
Four images (one to seven lights per image)	47.29° (12.37°)	28.85° (19.01°)	24.82° (24.58°)

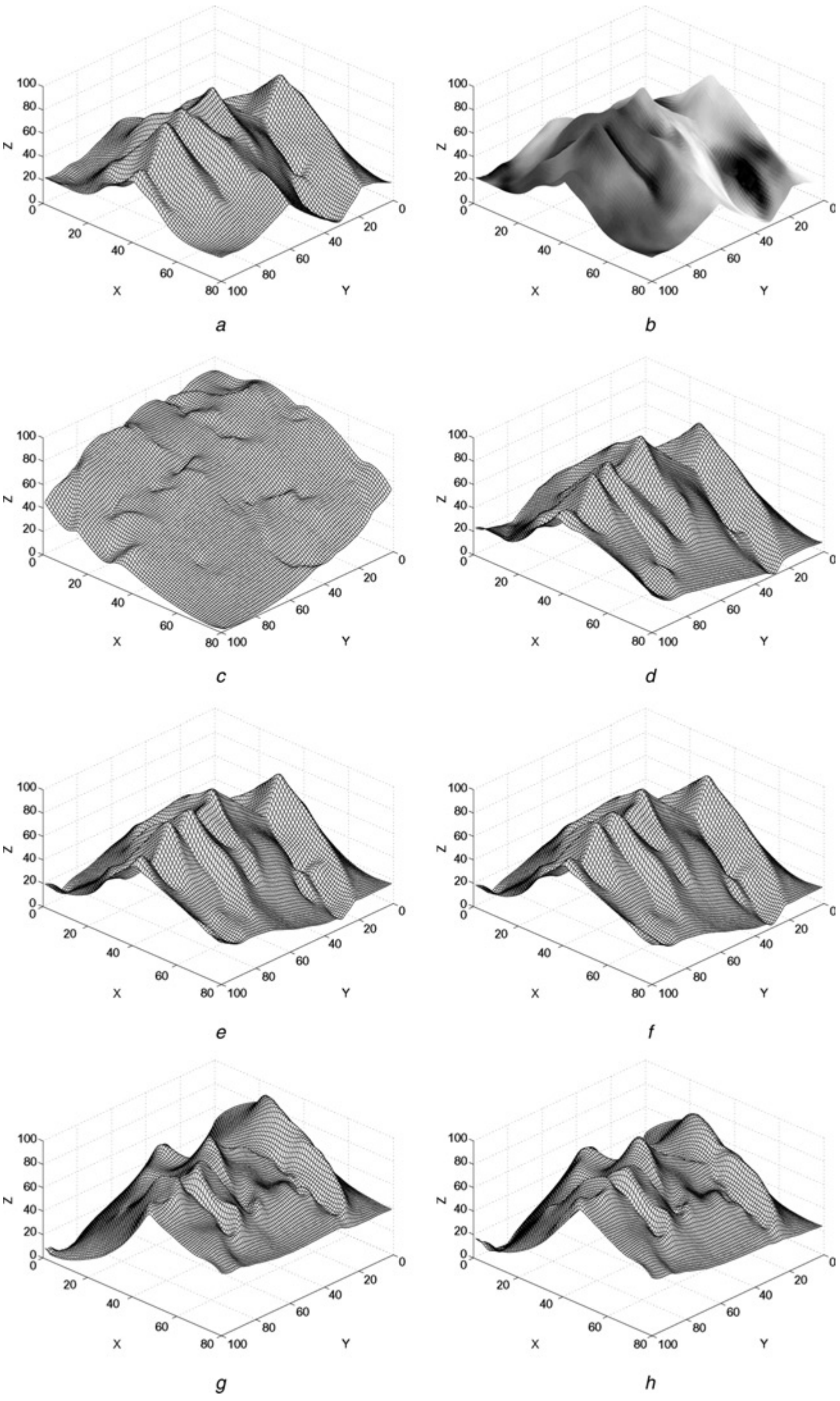

Fig. 9 *Estimated height maps comparison between the proposed and Basri et al. algorithms*

a Height map of the face using standard photometric stereo techniques and assuming that the light direction is known
b Same height map superimposed by the texture of the face
c, d Results for Barsi *et al.* algorithm [38]
e, f Light detection using photometric stereo
g, h Light detection using the spherical harmonics formulation
First column corresponds to the case where the non-convex constraint is used simultaneously with information regarding four random points of the object and the second column corresponds to the case where the constraint is ommited and ten points are used allowing the determination of the 4 × 4 linear transformation

used allowing the determination of the 4×4 linear transformation. Fig. 9*c* and *d* correspond to the Barsi *et al.* algorithm [38], Figs. 9*e* and *f* to light detection using photometric stereo, and Figs. 9*g* and *h* to the light detection using the spherical harmonics formulation. The results illustrate the effectiveness of the proposed framework even when there is a limited information about the object's surface, compared with the current algorithm by Barsi *et al.* [38] which fails to estimate the 3D model of the face. In all the cases, the height map was calculated from the surface normals using the algorithm in [35].

7 Conclusions

The paper presents 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. The proposed framework is based on the observation that different human faces appear very similar when illuminated by the same light configuration, which gives adequate information for recovery of the illumination distribution. Five algorithms have been proposed and experiments on real images show that the KLT-L algorithm can detect up to three lights to within 13° mean error, without any user interaction. The proposed framework can be extended to different pose by considering an appropriate set of training images. Moreover, application of the proposed framework to the problem of 3D reconstruction from multiple images under arbitrary lighting demonstrates one of the possible application areas of the framework.

8 References

1 Debevec, P.E.: 'Rendering synthetic objects into real scenes: Bridge traditional and image-based graphics with global illumination and high dynamic range photography'. SIGGRAPH 98, July 998, pp. 189–198

2 Bouganis, C.-S., and Brookes, M.: 'Mulitple light source detection', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2004, **26**, (4), pp. 509–514

3 Tsumura, N., Dang, M.N., Makino, T., and Miyake, Y.: 'Estimating the directions to light sources using images of eye for reconstructing 3D human face'. IS&T/SIDs 11 Color Imaging Conf., 2003, pp. 77–81

4 Nishino, K., and Nayar, S.K.: 'Eyes for relighting'. SIGGRAPH, 2004, vol. 23, pp. 704–711

5 Pentland, A.P.: 'Finding the illuminant direction', *J. Opt. Soc. Am.*, 1982, **72**, pp. 448–455

6 Brunelli, R.: 'Estimation of pose and illuminant direction for face processing', *Image Vis. Comput.*, 1997, **15**, (10), pp. 741–748

7 Sim, T., and Kanade, T.: 'Illuminating the face'. Tech. Rep. CMU-RI-TR-01-31. Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, September 2001

8 Vetter, T., and Poggio, T.: 'Linear object classes and image synthesis from a single example image', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, (7), pp. 733–742

9 Cootes, T., Walker, K., and Taylor, C.: 'View-based active appearance models'. 4th Int. Conf. on Automatic Face and Gesture Recognition, 2000, pp. 227–232

10 Riklin-Raviv, T., and Shashua, A.: 'The quotient image: class-based re-rendering and recognition with varying illuminations', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, (2), pp. 129–131

11 Marchener, S.R., and Greenberg, D.P.: 'Inverse lighting for photography'. 5th Color Imaging Conf., Society for Imaging Science and Technology, 1997

12 Sato, I., Sato, Y., and Ikeuchi, K.: 'Illumination distribution from brightness in shadows: adaptive estimation of illumination distribution with unknown reflectance properties in shadow regions'. Int. Conf. Computer Vision, 1995, vol. 2, pp. 875–882

13 Cyberware, available at: http://www.cyberware.com/(1999)

14 Phong, B.T.: 'Illumination for computer generated images', *Commun. ACM*, 1975, **18**, pp. 311–317

15 Torrance, K.E., and Sparrow, E.M.: 'Theory for off-specular reflection from roughened surface', *J. Opt. Soc. Am.*, 1967, **57**, pp. 1105–1114

16 Kemelmacher, I., and Basri, R.: 'Molding face shapes by example'. ECCV'06, May 2006, vol. 1, pp. 277–288

17 Zhang, L., and Samaras, D.: 'Face recognition under variable lighting using harmonic image exemplars'. CVPR, 2003, vol. I, pp. 19–25

18 Debevec, P., Hawkins, T., Tchou, C., Duiker, H.P., Sarokin, W., and Sagar, M.: 'Acquiring the reflectance field of a human face'. SIGGRAPH 00, 2000, pp. 145–156

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

20 Georghiades, A.S., Kriegman, D.J., and Belhumeur, P.N.: 'Illumination cones for recognition under variable lighting: faces'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1998, pp. 52–58

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

22 Belhumeur, P.N., and Kriegman, D.J.: 'What is the set of images of an object under all possible lighting conditions?'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1996, pp. 270–277

23 Fukunaga, K.: 'Introduction to statistical pattern recognition' (Academic Press, 1990, 2nd edn.)

24 Schemer, Y.Y., Nayar, S.K., and Belhemeur, P.N.: 'A theory of multiplexed illumination'. 9th IEEE Inter. Conf. on Computer Vision (ICCV'03), 2003, vol. 2, pp. 808–815

25 Lee, K.C., Ho, J., and Kriegman, D.: 'Nine points of light: acquiring subspaces for faces recognition under variable lighting'. Proc IEEE Conf. Computer Vision and Pattern Recognition, 2001

26 Bouganis, C.-S., and Brookes, M.: 'Class-based multiple light detection: an application to faces'. British Machine Vision Conf., 2003, vol. 1, pp. 113–122

27 Jolliffe, I.T.: 'Principle component analysis' (Springer-Verlag, 1986)

28 Jiang, J.-H., Tsenkova, R., and Ozaki, Y.: 'Principal discriminant variate method for classification of multicollinear data: principle and applications', *Anal. Sci.*, 2001, **17**, (suppl.), pp. i471–i474

29 Liu, C., and Wechsler, H.: 'Enhanced Fisher linear discriminant models for face recongnition'. Int. Conf. on Pattern Recognition, August 1998

30 Yang, J., Yu, H., and Kunz, W.: 'An Effcient LDA algorithm for face recognition'. 6th Int. Conf. on Control, Automation, Robotics and Vision (ICARCV 2000), 2000

31 Sim, T., Basker, S., and Bast, M.: 'The CMU pose, illumination, and expression (PIE) database'. Proc. 5th Int. Conf. on Automatic Face Gesture Recognition, 2002

32 Duda, R., and Hart, P.: 'Pattern classification and scene analysis' (Wiley, New York, 1973)

33 Kee, S.C., Lee, K.M., and Lee, S.U.: 'Illumination invariant face recognition using photometric stereo', *IEICE Trans. Inf. Syst.*, 2000, **E83-D**, (7), pp. 1466–1474

34 Basri, R., and Jacobs, D.W.: 'Lambertian reflectance and linear subspaces', *IEEE Trans Pattern Anal. Mach. Intell.*, 2003, **25**, (2), pp. 218–233

35 Kovesi, P.: 'Shapelets correlated with surface normals produce surfaces'. IEEE Int. Conf. on Computer Vision, 2005, vol. 2, pp. 994–1001

36 Shashua, A.: 'On photometric issues in 3d visual recognition from a single 2d image', *Int. J. Comput. Vis.*, 1997, **21**, pp. 99–122

37 Zhao, W.Y., and Chellappa, R.: 'Symmetric shape-from-shading using self-ratio image', *Int. J. Comput. Vis.*, 2001, **45**, (1), pp. 55–75

38 Basri, R., Jacobs, D., and Kemelmacher, I.: 'Photometric stereo with general, unknown lighting', *Int. J. Comput. Vis.*, 2007, **72**, (3), pp. 239–257