3D Assisted 2D Face Recognition: Methodology

J. Kittler, M. Hamouz, J.R. Tena, A. Hilton, J. Illingworth, and M. Ruiz

CVSSP, Surrey University, Guildford, Surrey, GU2 7XH, UK

Abstract. We address the problem of pose and illumination invariance in face recognition and propose to use explicit 3D model and variants of existing algorithms for both pose [Fit01,MSCA04] and illumination normalization [ZS04] prior to applying 2D face recognition algorithm. However, contrary to prior work we will use person specific, rather than general 3D face models. The proposed solution is realistic as for many applications the additional cost of acquiring 3D face images during enrolment of the subjects is acceptable. 3D sensing is not required during normal operation of the face recognition system. The proposed methodology achieves illumination invariance by estimating the illumination sources using the 3D face model. By-product of this process is the recovery of the face skin albedo which can be used as a photometrically normalised face image. Standard face recognition techniques can then be applied to such illumination corrected images.

1 Introduction

Face recognition is considered as one of the major challenges in computer vision. The problem is of interest to the computer vision community because of its importance in personal identity authentication for security applications, in control of physical access to buildings, in personalisation of services, in control of logical access to teleservices, in border control, and for its many other potential uses.

Face recognition differs from a normal task of object recognition in computer vision in the sense that each member of the face category (face of each individual) is considered as a separate entity. In other words, samples of the face population represent different identities. This makes the task of face recognition very difficult as individuals are distinguished by relatively minor differences in face shape and appearance. These minor differences are normally sufficient to discriminate between individuals, provided the reference face image (template) and the probe (test image of unknown identity) are acquired under controlled illumination conditions and in a standard pose, nominally the frontal one. However, in many situations the illumination conditions and the pose are difficult to control and the photometric effect of any illumination and pose changes invariably swamp the subtle differences in shape and texture of face images of two individuals.

One possibility to avoid the pose and illumination problem is to base the face recognition on 3D data rather than 2D. Different 3D sensing technologies

M. Lazo and A. Sanfeliu (Eds.): CIARP 2005, LNCS 3773, pp. 1055–1065, 2005.

have been developed to acquire 3D face images (depth images). These include stripe-based stereo systems (Minolta, Cyberware, etc.), and area-based stereo (3dMD, Surfim, Wicks and Wilson, etc.). The recognition is then accomplished by matching two 3D surfaces, a reference 3D surface (template) and a test 3D face image. This first of all involves 3D surface registration, followed by the extraction of 3D image features and finally decision making. However, the 3D face recognition approach has failed to live up to all expectations. This is partly caused by missing data, inaccuracies in surface registration and most importantly, by not making use of the discriminatory information conveyed by the skin surface texture.

A more promising alternative is to make use of 3D face shape as well as skin texture by simultaneously acquiring 3D and 2D face image. However, this solution requires a relatively expensive sensor system and may not be acceptable in many application scenarios. We propose an approach whereby the recognition, in the operational mode, is based on 2D face images only. However, the recognition process is assisted by a 3D face model. We believe that during the enrolment of a user, it is perfectly feasible to acquire a 3D model of the face using 3D sensing. The 3D model can then be associated with a 2D face template and used for the recognition of 2D test images. Given a 2D image of a face and its 3D model, it is possible to estimate the illumination sources and separate the effect of light and albedo. The albedo image can then be relighted to the same conditions as those used during the user enrolment. The photometrically corrected face image can then provide a better basis for matching.

The above process requires the 3D face model to be registered with the 2D probe image. In principle, once the two types of spatial data sets are registered, the 2D image could also be corrected for pose. However, in this paper we shall assume that the person to be recognised is cooperative and presents himself or herself in the frontal pose. No geometric correction is therefore necessary.

The idea of using a 3D model in conjunction with 2D face data has been explored before. For instance, Zhao and Chellappa [ZC00] investigated a general 3D face model, as well as shape from shading, to improve the recognition performance in an environment with varying illumination. In many respects, our work is similar, but the main difference is that we propose to use client specific 3D face model, rather than a general model. This should lead to a better accuracy in estimating the photometrically corrected image and therefore better performance rates.

In this paper we describe the methodology developed for the proposed 3D assisted 2D face recognition system. In the next section we present a review of the relevant literature. In Section 3.2 we describe the method used for registering 2D face image to a 3D face model. In Section 3.3 we discuss the method of illumination source estimation and the face relighting. Section 4 provides a brief overview of the recognition method, based on Linear Discriminant Analysis. The paper is drawn to conclusion in Section 5.

2 State of the Art

Pose and illumination were identified as major problems in 2D face recognition. Approaches trying to solve these two issues in 2D are bound to have limited performance due to the intrinsic 3D nature of the problem.

Blanz and Vetter [BV03] proposed an algorithm which takes a single image on the input and reconstructs 3D shape and illumination-free texture. Phong's model is used to capture the illumination variance. The model explicitly separates imaging parameters (such as head orientation and illumination) from personal parameters allowing invariant description of the identity of faces. Texture and shape parameters yielding the best fit are used as features. Several distance measures have been evaluated on the FERET and the CMU-PIE databases.

Basri and Jacobs [BJ03] proved that a set of images of a convex Lambertian object obtained under arbitrary illumination can be accurately approximated by a 9D linear space which can be analytically characterized using surface spherical harmonics. Zhang and Samaras [ZS04] used Blanz and Vetter's morphable model together with a spherical harmonic representation for 2D recognition. The method is reported to perform well even when multiple illuminants are present.

Yin and Yourst [YY03] describe their 3D face recognition system which uses 2D data only. The algorithm exploits 3D shape reconstructed from front and profile images of the person using a dynamic mesh. A curvature-based descriptor is computed for each vertex of the mesh. Shape and texture features are then used for matching.

These approaches represent significant steps towards the solution of illumination, pose and expression problems. However there are still several open research problems like full expression invariance, accuracy of the Lambertian model with regard to the specular properties of human skin and stability of the model in presence of glasses, beards and changing hair, etc., that need addressing.

3 Methodology of 3D Assisted Photometric Normalisation

The function of the proposed recognition system can be summarised in the the following steps:

Enrolment:
3D pose normalization $\Longrightarrow$ "3D to 3D" dense registration.

Test:
"3D to 2D" dense registration $\Longrightarrow$ Illumination correction $\Longrightarrow$ 2D recognition.

3.1 3D to 3D Registration

Data coming from the 3D sensor during the enrolment are absolute 3D coordinates relative to sensor's internal coordinate system. Such a coordinate system is typically defined by the calibration chart and is unknown to the user. Assuming

that the coordinate system changes whenever the sensor moves/is recalibrated, data coming from the 3D sensor needs to be registered to a common reference coordinate system. This is necessary for subsequent stages as face surfaces of different people need to be manipulated in the same manner.

The registration process can be divided in the two following stages. First, the surface is pose normalised. This is achieved by a LM-ICP variant of Iterative Closest Point algorithm proposed by Fitzgibbon [Fit01], which assumes a rigid transformation and uses robust matching to improve performance in the presence of outliers. For such purposes a generic face surface template is used and the face surface under consideration is rotated and translated to minimize distance between the two surfaces. Practical experience shows that such registration is only approximate. Due to the inter-personal shape differences, there will always be misregistrations which cannot be expressed by a rigid transformation. However we believe that for the purpose of pose normalization, performance of ICP is adequate.

The second stage in the registration process is a fine registration. This is a necessary step in order to obtain dense correspondences between shapes. Dense correspondences are established for a pair of surfaces by finding for each point in one of them a corresponding point in the other. Constructing dense correspondences facilitates learning inter- and intra-personal shape and texture variability. Most existing algorithms solving this problem exploit the fact that although globally different, face surfaces of different people are locally similar. In other words, the deformation leading from one face surface to another can be locally constrained. This enables an efficient search for similar features. A representative method for finding dense correspondences is the method of Mao et al. [MSCA04]. The algorithm is initiated by 5 manually defined landmarks to establish initial mapping of a generic model onto the surface of an individual. Following global registration, the generic shape is further deformed locally to the input data. The similarity measure is based on a combination of curvature, distance and surface normals. The deformation process is then completed by minimizing the overall energy of the generic model. The movement of generic model vertices is restricted.

An example of the output from this algorithm is depicted in Fig. 1.

Given dense correspondences, intra- and possibly inter-personal shape variations can be analyzed. For one person, the variations are down to changes in expression and possibly aging. With textured meshes, correspondences between textures can be directly derived from 3D surface correspondence. This facilitates efficient texture analysis.

3.2 3D to 2D Registration

Given an image of a person for the purpose of verification, the 3D model has to be first registered with the 2D data. In the verification context, the algorithm proposed by Blanz and Vetter [BV03] suffers from several drawbacks. As it attempts to learn both intra- and inter-personal texture and shape variability simultaneously, the number of optimization parameters is high with too few

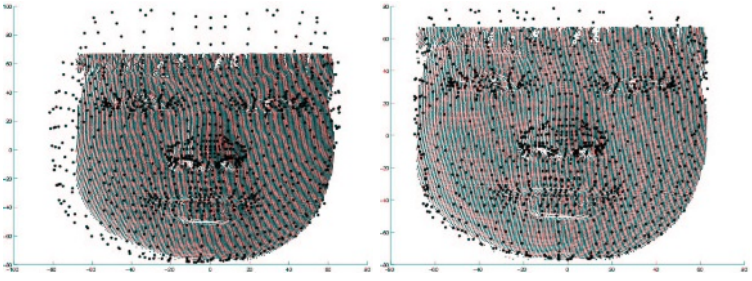

Fig. 1. Examples of registration by Mao et al [MSCA04], initial fit left, final registration right, model surface vertices depicted by dots

constraints and the algorithm often converges to a local minimum. This causes inaccuracies in both the recovered shape and texture and also computational complexity is increased. We believe that using person-specific shape and texture will greatly reduce the number of free parameters in the optimization loop and therefore the accuracy will increase. The shape model will have to capture only the expression variability of the given person and texture model mainly the illumination changes. It is unrealistic to expect that for every person enrolled there will be a huge expression training set available, and thus expression variability will have to be learnt over different people. However, the resulting inaccuracies for the given person will be compensated by introducing person-specific shape in the optimization process. To train such a model, densely registered shape and texture data (described above) are needed. As a part of a score function to be minimized, an illumination factor has to be defined. Recently, Basri and Jacobs [BJ03] have proposed a novel approach for modelling Lambertian reflectance exploiting 3D information.

3.3 Illumination Model Using Spherical Harmonics

Belhumeur and Kriegman proved that the set of images of an object in fixed pose but under all possible illumination conditions is a convex cone (illumination cone) [BK96]. The cone can be well approximated by a low-dimensional subspace for Lambertian objects. Several training images of the object under varying illumination are needed to reconstruct the cone, which makes it impractical. Basri and Jacobs [BJ03] proved that a set of images of a convex Lambertian object under distant lighting lies close to a 9D linear subspace and this subspace can be analytically characterized using surface spherical harmonics. This stems from the observation that a Lambertian surfaces acts as a low-pass filter for the lighting function and therefore the reflectance can be accurately approximated by low-order spherical harmonics. These findings were directly applied for illumination correction in the approach of Zhang and Samaras [ZS04]. As 3D information is needed, Blanz and Vetter's morphable model was used together with a spherical harmonic representation. This method is applied to 2D face

recognition and is reported to perform well even when multiple illuminants are present. Zhang's and Samaras's algorithm for texture recovery is summarized in Alg. 1.

Data : Pixels with corresponding surface normals (shape registered with texture)

Result : Recovered albedo (illumination-free texture)

for *Each pixel of the face* **do**
  Compute the spherical-harmonics basis (9-dimensional vector) using the attached surface normal
end

Iteratively solve equation for $ALBEDO$ and $9D\_LIGHT$:
$INTENSITY = ALBEDO \cdot (BASIS \cdot 9D\_LIGHT)$

where $INTENSITY$ is a vector of gray-scale intensities, $ALBEDO$ is a vector of albedos, $BASIS$ is a [number of pixels] $\times$ 9 matrix representing spherical harmonics basis and $9D\_LIGHT$ is a 9-dimensional illumination vector;

Use $ALBEDO$ as a delit texture;

Algorithm 1. Texture recovery algorithm (for details see [ZS04])

Fig. 2. Original images (top), Recovered albedo (bottom)

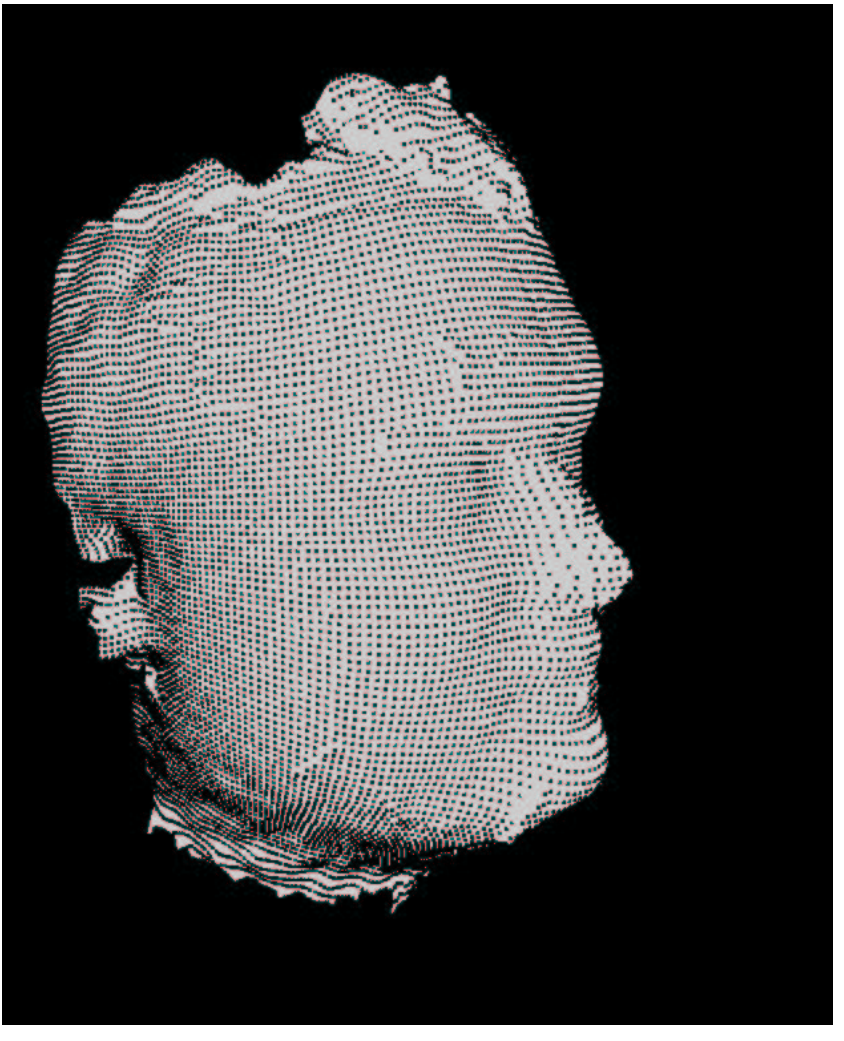

Fig. 3. Face surface produced by the sensor [YHR04]

By using surface normals computed directly from the 3D data instead of the reconstructed normals from the morphable model we believe that recognition accuracy will increase. Fig. 2 shows the example of the recovered albedo using Zhang's and Samaras's algorithm for texture recovery and person specific 3D shape. For each image, 3D shape was acquired using our own active stereo sensor [YHR04]. The shape for the first image is shown in Fig 3. Just to demonstrate the success of the delighting, similarity matrices consisting of correlation coefficients computed directly between the pixels are presented below. The matrix on the left shows the similarity between the original images, the right between delit

images. Ideally, all entries in the similarity matrix should be exactly 1.0 as these are images of the same person. The factors like illumination and misregistration however reduce the score to a number smaller than 1.0. The improvement by the proposed illumination correction is clearly noticeable.

$$\begin{pmatrix} 1.0000 & 0.7847 & 0.5872 \\ 0.7847 & 1.0000 & 0.6989 \\ 0.5872 & 0.6989 & 1.0000 \end{pmatrix} \qquad \begin{pmatrix} 1.0000 & 0.8252 & 0.7424 \\ 0.8252 & 1.0000 & 0.7880 \\ 0.7424 & 0.7880 & 1.0000 \end{pmatrix}$$

4 2D-Based Matching

Once a person-specific morphable model is successfully matched to the input image, densely registered illumination free texture can be obtained. Existing 2D recognition algorithms can be used on such data. As the enrolment data includes full view of the head, even partially non-frontal poses can be used for recognition.

A large variety of face recognition methods have been suggested in the literature [ZCPR03]. However, it is well known that in controlled conditions, which we try to emulate by 3D assisted pose and photometric normalisation of the input image, Linear Discriminant Analysis (LDA) provides an effective pattern representation. Although it is designed to extract only first order discriminatory information, in recent experiments in face based personal identity verification, LDA has been shown to outperform both linear and nonlinear boundary Support Vector Machines [JKLM99]. This may be the consequence of the sparseness of training data in this particular application where only a few gallery images are available in the training set for each client. In such situations only the simplest model, defined in terms of the class mean vector, can be inferred for each client distribution and this is exactly what LDA is able to exploit.

The LDA projection maximises the ratio of between class and within class scatters. Given a set of vectors $x_i, i = 1, \ldots, M$, $x_i \in R^D$, each belonging to one of c classes $\{C_1, C_2, \ldots, C_c\}$, we compute the between-class scatter matrix, S_B,

$$S_B = \sum_{i=1}^{c} (\nu_i - \nu)(\nu_i - \nu)^T \tag{1}$$

and within-class scatter matrix, S_W

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in C_i} (x_k - \nu_i)(x_k - \nu_i)^T \tag{2}$$

where ν is the grand mean and ν_i is the mean of class C_i.

The objective of LDA is to find the transformation matrix, W_{opt}, that maximises the ratio of determinants $\frac{|W^T S_B W|}{|W^T S_W W|}$. W_{opt} is known to be the solution of the following eigenvalue problem [DK82]:

$$S_B W - S_W W \Lambda = 0 \tag{3}$$

where Λ is a diagonal matrix whose elements are the eigenvalues of matrix $S_W^{-1} S_B$. The column vectors $\boldsymbol{w_i}$ $(i = 1, \ldots, c-1)$ of matrix W are referred to as *Fisherfaces.*

In high dimensional problems (e.g. in the case where x_i are images and D is $\approx 10^5$) S_W is almost always singular, since the number of training samples M is much smaller than D. Therefore, an initial dimensionality reduction must be carried out before solving the eigenvalue problem in (3). Commonly, dimensionality reduction is achieved by Principal Component Analysis [TP91]; the first $(M-c)$ eigenprojections are used to represent vectors x_i. The dimensionality reduction also allows S_W and S_B to be efficiently calculated. The optimal linear feature extractor W_{opt} is then defined as:

$$W_{opt} = W_{lda} * W_{pca} \tag{4}$$

where W_{pca} is the PCA projection matrix and W_{lda} is the optimal projection obtained by maximising

$$W_{lda} = arg \max_W \frac{|W^T W_{pca}^T S_W W_{pca} W|}{|W^T W_{pca}^T S_B W_{pca} W|} \tag{5}$$

The LDA axes are known to perform prewhitening of the within class covariances. In other words, the within class covariance matrix becomes an identity matrix. The assumption that each client distribution is Gaussian with mean μ_i and an identity covariance matrix underlies the LDA approach. Under this assumption the optimal metric for face image classification is the Euclidean metric. Accordingly, given a probe image, $\mathbf{x}$, in the LDA space, we can compute a matching score s for the probe and the i-th client mean μ_i as the Euclidean distance between the two vectors, i.e.

$$s_E = \sqrt{(\mathbf{x} - \mu_i)^T (\mathbf{x} - \mu_i)} \tag{6}$$

Alternatively we can match the probe to a model using the normalised correlation as a matching score function. The measure is defined as

$$s_N = \frac{||\mathbf{x}^T \mu_i||}{\sqrt{\mathbf{x}^T \mathbf{x} \mu_i^T \mu_i}} \tag{7}$$

The normalised correlation projects the probe vector onto the mean vector of the claimed client identity, emanating from the origin. It effectively uses just

one dimensional space onto which the test data is projected. The magnitude of projection is normalised by the length of the mean and probe vectors.

Although normalised correlation is very effective, in many situations it has been outperformed by the gradient metric defined as

$$s_o = \frac{||(\mathbf{x} - \mu_i)^T \nabla P(i|\mathbf{x})||}{||\nabla P(i|\mathbf{x})||} \tag{8}$$

where $\nabla P(i|\mathbf{x})$ is the gradient direction for user i defined as

$$\nabla P(i|\mathbf{x}) = \sum_{\substack{j=1 \\ j \neq i}}^{m} p(\mathbf{x}|j)(\mu_j - \mu_i) \tag{9}$$

and $p(\mathbf{x}|j)$ is j^{th} class probability density function assumed to be Gaussian.

Either of these score functions or their combination can be used for final decision making in the LDA space.

5 Discussion and Conclusion

We addressed the problem of pose and illumination invariance in face recognition and propose an approach which makes use of 3D face models in 2D face recognition. The proposed solution is realistic as for many applications the additional cost of acquiring 3D face images during enrolment of the subjects is acceptable. 3D sensing is not required during normal operation of the face recognition system, as the recognition process is based on standard 2D face imaging.

The proposed methodology achieves illumination invariance by estimating the illumination sources using the 3D face model. This involves modelling the effect of illumination using a low order spherical harmonics model. The by-product of the process is the recovery of the face skin albedo which can be used as a photometrically normalised face image, or can be relit to the same lighting conditions as those used during the enrolment. Standard face recognition techniques can then be applied to such illumination corrected images.

The proposed methodology, which is distinguished from existing techniques by deploying user specific, rather than general, 3D face models, was outlined. Simple experiments confirming the benefits of the proposed photometric normalisation were conducted.

Acknowledgements

This work was supported by EPSRC Research Grant GR/S46543/01 with contributions from EU Project Biosecure and COST 275.

References

[BJ03] Ronen Basri and David W. Jacobs. Lambertian Reflectance and Linear Subspaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(2):218–233, 2003.

[BK96] P. N. Belhumeur and D. J. Kriegman. What is the Set of Images of an Object Under All Possible Lighting Conditions. In *Proc. of IEEE Conference of Computer Vision and Pattern Recognition*, pages 270–277, 1996.

[BV03] Volker Blanz and Thomas Vetter. Face Recognition Based on Fitting a 3D Morphable Model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(9):1063–1074, 2003.

[DK82] P. A. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach.* Prentice Hall, 1982.

[Fit01] A. W. Fitzgibbon. Robust Registration of 2D and 3D Point Sets. In *Proceedings of the British Machine Vision Conference*, pages 662–670, 2001.

[JKLM99] K. Jonsson, Josef Kittler, Yongping Li, and Jiri Matas. Support Vector Machines for Face Authentication. In *BMVC*, 1999.

[KHHI05] J. Kittler, A. Hilton, M. Hamouz, and J. Illingworth. 3D Assisted Face Recognition: A Survey of 3D Imaging, Modelling and Recognition Approaches. In *Proc. of IEEE Workshop on Advanced 3D Imaging for Safety and Security, A3DISS 2005 (CD-ROM of the CVPR 2005)*, 2005.

[MSCA04] Z. Mao, J.P. Siebert, W.P. Cockshott, and A. F. Ayoub. Constructing dense correspondences to analyze 3D facial change. In *Proc.of the 17th International Conference on Pattern Recognition, ICPR'04)*, volume 3, pages 144–148, 2004.

[TP91] M. A. Turk and A. P. Pentland. Eigenfaces for Recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[YHR04] I.A. Ypsilos, A. Hilton, and S. Rowe. Video-rate Capture of Dynamic Face Shape and Appearance. In *Proc. 6th Int. Conf. on Automatic Face and Gesture Recognition (FGR 2004)*, pages 117–122, 2004.

[YY03] Lijun Yin and Matt T. Yourst. 3D face recognition based on high-resolution 3D face modeling from frontal and profile views. In *WBMA '03: Proceedings of the 2003 ACM SIGMM workshop on Biometrics methods and applications*, pages 1–8. ACM Press, 2003.

[ZC00] WenYi Zhao and Rama Chellappa. SFS Based View Synthesis for Robust Face Recognition. In *FG '00: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*, page 285, Washington, DC, USA, 2000. IEEE Computer Society.

[ZCPR03] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, 2003.

[ZS04] Lei Zhang and Dimitris Samaras. Pose Invariant Face Recognition under Arbitrary Unknown Lighting using Spherical Harmonics. In *Proc. Biometric Authentication Workshop 2004, (in conjunction with ECCV2004), pp. 10–23*, 2004.