International Journal of Bio-Science and Bio-Technology
Vol. 3, No. 3, September, 2011

Using Photometric Stereo for Face Recognition

Gary A. Atkinson and Melvyn L. Smith

University of the West of England, Bristol, BS16 1QY, UK
gary.atkinson@uwe.ac.uk

Abstract

This paper aims to review the technique of Photometric Stereo (PS), with specific application to face recognition. PS is a method to rapidly estimate the three-dimensional geometry of a face (or any other Lambertian-like object) using several images with an identical viewpoint but varied illumination directions. The contributions of this paper are to (1) summarise the pros and cons of PS compared to alternative methods; (2) cover the theory of PS, in particular with respect to the related method of shape-from-shading; (3) outline some of the key extensions of PS to help overcome its weaknesses; and (4) discuss an application of PS for a practical and complete face recognition system.

Keywords: *Photometric Stereo (PS), face recognition, face recognition system*

1. Introduction

Of all the biometric modalities researched and deployed to date, automatic face recognition [1] appears to be taking centre-stage at present. This is due to several reasons including its ability to operate without subject contact or interaction (as in fingerprint recognition, for example), the fact that the face of an individual is usually visible and that it requires minimal or no effort from the user. Face recognition is also a natural choice for biometrics applications as it is the most reliable modality used by humans; especially for familiar faces. However, automatic face recognition remains a major challenge for all but the most restricted set-ups: i.e. where all the face images are fully frontal with identical illumination and neutral expression. In an effort to relax these restrictions, many researchers have turned to three-dimensional methods for data capture and recognition.

A great deal of effort has been made in 3D (and 2D+3D fusion) methods for face recognition. An extensive list of approaches and their successes is provided in [2]. The state-of-the-art methods for recognition would be expected to achieve at least 90% recognition on the Face Recognition Grand Challenge (FRGC) version 2 database [3], which contains a total of over 4000 images of over 400 subjects. However, the problem of reliable and convenient 3D data capture for real practical biometric applications is a less studied area.

This paper therefore aims to review one such method for 3D data capture: photometric stereo (PS). This is a technique that captures multiple 2D images of an object (e.g. a face) each with a different light source direction. The changes in pixel intensities at each point are used to deduce surface orientation. The orientations are typically represented by surface normals – vectors located at points on the target object that are oriented orthogonally to the surface at that point. Figure 1 illustrates the photometric stereo paradigm. Figure 2 shows an example of the application of PS to a face, as used for example, in a face recognition system.

The paper first defines PS in a generic (non-biometric) sense, before discussing why it is an attractive option for face recognition. Next, the theory of PS is presented with particular emphasis on the concepts and various ways to view the technique. This is followed by a review of the extensions to PS and an explanation of its use in face recognition. The emphasis of the paper is on the general theory and application of PS, rather than to conduct detailed comparisons of specific methods.

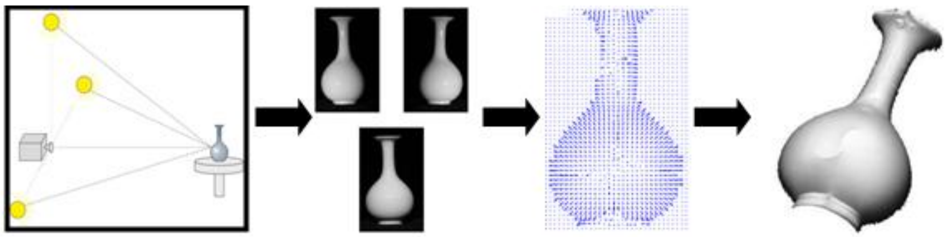

Figure 1: Illustration of a basic photometric stereo capture mechanism. From left: experimental arrangement, raw images, estimated surface normals, recovered depth. The algorithm in [4] was used here to calculate depth from surface normals.

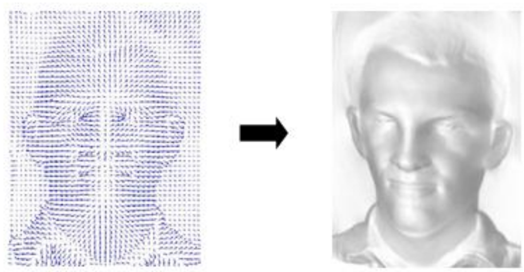

Figure 2: Example of using surface normals from PS for face recognition.

2. Why use photometric stereo?

We can divide this question in to two parts: (1) why 3D in general? and (2) why PS in particular? Figure 3 and Figure 4 illustrate some of the problems of 2D face recognition which can be, at least in part, overcome using 3D methods. Figure 3 shows two images of an individual taken on the same day, using the same camera and camera settings, with the same expression, pose and clothing and no changes to the background. The only change between the two images is the lighting. Nevertheless, the images appear completely different, which would confound most 2D recognition systems. However, since the 3D shape of the face has not changed, 3D methods would be unaffected by this difficulty. To illustrate the point further, Figure 4 shows the same individual, this time with the pose and expression changing, but the light source constant. Again, the images appear completely different. While both 2D and 3D methods suffer in these cases, 3D methods have the advantage that they are better able to correct for such changes [5].

Using 3D data for face recognition therefore allows for pose and illumination correction, which are two commonly cited problems with conventional 2D images. Better recognition rates have also been reported using 3D over 2D data [6], although this is not always replicated [7]. One reason for this may be the representation of the 3D data used in the analysis.

Gökberk et al. [8] performed recognition experiments using numerous 3D representations and they concluded that "...surface normals are better descriptors than the 3D coordinates of the facial points." This is at odds with most research which uses the 3D point coordinates as a starting point. For this reason, PS – which calculates the surface normals directly – is particularly well suited to face recognition.

Figure 3: Illustration of the difficulties caused by changing illumination on 2D face recognition.

Figure 4: Illustration of the difficulties caused by changing pose and expression on 2D face recognition.

One of the main challenges with current 3D recognition technology relates to the means by which the data are collected. A cheap and reliable method to rapidly capture the 3D face shape of an individual is highly desirable. At present, methods tend to be too expensive, of insufficient accuracy, too slow or too computationally demanding for commercial use.

Table 1 shows an informal assessment carried out by the authors of the most common methods for shape reconstruction based on the current literature. The methods described are as follows:

1. Shape-from-shading (SFS) [9]. The shading patterns of a single image are used to deduce geometry.
2. Geometric stereo (GS) [10]. Triangulation between two or more viewpoints is used to calculate depth.
3. Laser triangulation (LT) [11]. Laser light is projected onto the target object and the reflected light pattern is used for triangulation line-by-line.
4. Projected pattern (PP) [12]. Two cameras are used, as in geometric stereo, but an intensity pattern is projected onto the target object to aid triangulation.
5. Photometric stereo (PS) [13]. Two or more images are taken (often in sequence), where the light source direction changes between views.

The methods in Table 1 are measured against six essential criteria for face recognition applications. Note that PS performs well or average in four of these criteria, although there is room for improvement terms of accuracy and practicality. However, the table refers to the most basic form of PS. Later in this paper we present work that goes some way to improve both these shortcomings.

Table 1: Comparison of reconstruction methods. NB. Practicality refers to the ease by which the methods can be commercially deployed. The abbreviations are defined in the list above.

	Cost	Computation	Accuracy	Resolution	Ease of Calibration	Practicality
1. SFS	Good	Average	Poor	Good	Good	Good
2. GS	Good	Good	Variable*	Variable*	Good	Average
3. LT	Average	Good	Average	Average	Good	Poor
4. PP	Poor	Poor	Good	Good	Average	Average
5. PS	Good	Good	Average	Good	Good	Average

* Depends on correspondence density

3. Theory of shape-from-shading and photometric stereo

The mathematics of photometric stereo are covered in detail elsewhere in the literature [13], [14]. Here we concentrate primarily on the concepts and its relation to SFS. To start, we make the following assumptions [14]:

1. No cast/self-shadows, inter-reflections or specularities (mirror-like reflections)
2. Greyscale/linear imaging
3. Distant and uniform light sources
4. Orthographic projection
5. Static surface
6. Lambertian reflectance

We commence by assuming that the surface is reflecting light according to Lambert's Law. This assumes that light is reflected by the material equally in all directions. It can be seen in the literature that this is a reasonably good approximation for human skin [15]. In physics terms, we can state that

$$E = \rho L \cos\theta \tag{1}$$

where E is the emittance (reflected power per unit area) from the surface, ρ is the albedo (ratio of reflected to incident irradiance at normal incidence), L is the irradiance (incident power per unit area) and θ is the angle between the light source vector and the surface normal. For a surface $z = f(x, y)$, the surface normal, $\boldsymbol{n}$, can be written.

$$\boldsymbol{n} = \left[\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}, \text{-}1\right]^{\mathrm{T}} = [p, q, \text{-}1]^{\mathrm{T}} \tag{2}$$

If we let the light source vector be $\boldsymbol{n_s} = [p_s, q_s, -1]^T$, then we can write

$$\boldsymbol{n}.\boldsymbol{n_s} = pp_s + qq_s + 1 = \sqrt{{p_s}^2 + q^2 + 1}\sqrt{{p_s}^2 + {q_s}^2 + 1}\cos\theta \tag{3}$$

Substituting (1) and (3) we have

$$E = \rho L \frac{pp_s + qq_s + 1}{\sqrt{{p_s}^2 + q^2 + 1}\sqrt{{p_s}^2 + {q_s}^2 + 1}} \tag{4}$$

In computer vision, we usually re-write this equation as

$$I = \rho \frac{pp_s + qq_s + 1}{\sqrt{{p_s}^2 + q^2 + 1}\sqrt{{p_s}^2 + {q_s}^2 + 1}} \tag{5}$$

where I is the pixel brightness measured by the camera. Here, we have assumed a linear camera response and have "absorbed" the incident light irradiance and camera response constant into the albedo. For an 8-bit image, this means that both I and ρ fall into the interval $[0,255]$. Typically, we assume that the light source vector $\boldsymbol{n_s}$ is known, meaning that from a single pixel measurement, I, we have one equation with three unknowns (i.e. p, q and ρ).

Figure 5 shows the range of solutions to (5) for given values of ρ and I. For the first case, for example, with a given measurement of I and for a particular (unknown) value of ρ, the solution to the equation lies on a circle in pq-space. In fact, the solution to the equation in general lies on a conic section as illustrated by Figure 6. Here, we see that the solution can be regarded as an intersection between a cone (whose axis lies on the light source vector and whose apex is at the surface point) and the image plane.

A third representation can be seen in Figure 7 where the solution lies on the surface of a sphere. This figure also relates SFS to PS in that the latter technique uses multiple light sources to fully constrain the surface normal.

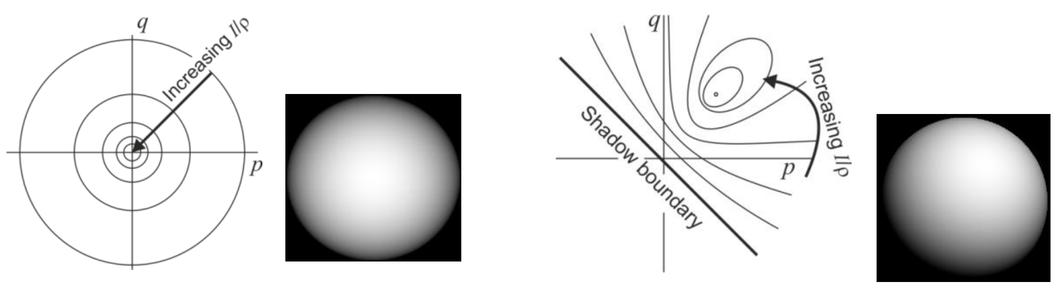

Figure 5: Solutions to (5) for p and q. The left-hand example shows the case where $p_s = q_s = 0$, while for the right-hand case, $p_s = q_s = 0.5$. The rendered images show a Lambertian sphere illuminated under these particular light source directions.

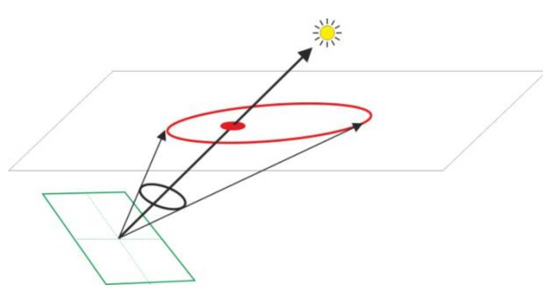

Figure 6: Representation of the solution to (5) as a conic section (in this case an ellipse). The apex of the cone lies at the point on the surface, while its axis lies along the light source vector.

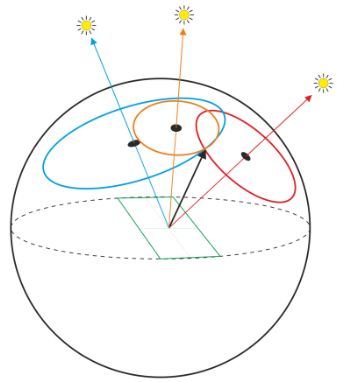

Figure 7: Representing the solution to (5) on a sphere for three different light sources. Use of multiple sequential light sources are able to fully constrain the normal, as in PS.

Usually, PS represents Lambert's Law (1) as a vector equation so that

$$I = \rho \boldsymbol{s}.\boldsymbol{N} = \rho \begin{bmatrix} s_x \\ s_y \\ s_z \end{bmatrix}^T \begin{bmatrix} N_x \\ N_y \\ N_z \end{bmatrix} \tag{6}$$

where $\boldsymbol{N}$ is the *unit* normal vector (as opposed to $\boldsymbol{n}$ which is the surface normal represented by surface derivatives). A three-source PS arrangement can then be written in a matrix equation:

$$\begin{bmatrix} s_x^1 & s_y^1 & s_z^1 \\ s_x^2 & s_y^2 & s_z^2 \\ s_x^3 & s_y^3 & s_z^3 \end{bmatrix}^{-1} \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = \rho \begin{bmatrix} N_x \\ N_y \\ N_z \end{bmatrix} = \begin{bmatrix} m_x \\ m_y \\ m_z \end{bmatrix} \tag{7}$$

Performing substitutions of the above equations yield the following equations for the unknowns in (5):

$$p = -\frac{m_x}{m_z}, \qquad q = -\frac{m_y}{m_z}, \qquad \rho = \sqrt{m_x^2 + m_y^2 + m_z^2} \tag{8}$$

It should be noted at this point that if the three light source vectors are co-planar then (8) becomes insoluble because the light source matrix in (7) becomes non-singular.

4. Advanced Methods in PS

Of the six assumptions listed at the start of Section 3, perhaps the most studied are the assumption that no shadows are present and the assumption of Lambertian reflectance. This section does not aim to present an exhaustive literature review of PS methods aimed to address these issues, but points towards the main approaches of overcoming them.

One means to minimize the effects of shadows and specularities is to introduce a fourth light source (or more). This allows (7) and (8) to be applied four (or more) times: once for each combination of three light sources. This gives four estimates for the surface normal at each point. We can then say that where the discrepancy between these estimates is too great to be accounted for by camera noise [16], [17] then one of the sources should be omitted from the calculation in order to minimize the error between the remaining three.

Another approach that goes some way to reduce shadows, specularities and non-Lambertian reflectance [18] uses the observation that, due to the linear dependence of the light source vectors,

$$a_1\boldsymbol{s_1} + a_2\boldsymbol{s_2} + a_3\boldsymbol{s_3} + a_4\boldsymbol{s_4} = 0 \tag{9}$$

where a_1, a_2, a_3 and a_4 are constants. If we multiply (9) by the albedo and take the scalar product with the surface normal, we have

$$a_1\rho(\boldsymbol{s_1}.\boldsymbol{N}) + a_2\rho(\boldsymbol{s_2}.\boldsymbol{N}) + a_3\rho(\boldsymbol{s_3}.\boldsymbol{N}) + a_4\rho(\boldsymbol{s_4}).\boldsymbol{N} = 0 \tag{10}$$

$$\therefore a_1I_1 + a_2I_2 + a_3I_3 + a_4I_4 = 0 \tag{11}$$

Where condition (11) is satisfied (subject to the confines of camera noise), the pixel is not specular, not in shadow and follows Lambert's Law.

Finally, we mention a method that not only overcomes the non-Lambertian assumption, but also reduces the impact of non-uniform illumination [19]. The method involves imaging the target in the presence of a "gauge" object, which is typically a sphere whose reflectance properties are assumed to be identical to the target object. Using M light sources, intensity vectors are acquired for both the target (scene) and the gauge:

$$\begin{aligned} \boldsymbol{I}^{(\text{scene})} &= \left[I_1^{(\text{scene})} \; I_2^{(\text{scene})} \cdots \; I_M^{(\text{scene})}\right]^T ; \\ \boldsymbol{I}^{(\text{guage})} &= \left[I_1^{(\text{guage})} \; I_2^{(\text{guage})} \cdots \; I_M^{(\text{guage})}\right]^T \end{aligned} \tag{12}$$

The task then is presented as finding a mapping between the normals i on the target with those j on the gauge:

$$\boldsymbol{I}^{(\text{scene})}[i] \mapsto \boldsymbol{I}^{(\text{guage})}[j]: \boldsymbol{N}^{(\text{scene})}[i] = \boldsymbol{N}^{(\text{guage})}[j] \tag{13}$$

The interested reader is referred to the following papers for more approaches to PS in biometrics and in general: [20], [21], [22], [23], [24], [25], [26], [27].

6. The PhotoFace recognition system

Figure 8 shows the hardware from [15], which we call the "PhotoFace" system. The device uses a Field Programmable Graphics Array (FPGA) to synchronize a camera, operating at 200 frames per second, to four separate flash lights. The reconstruction in Figure 2 was captured using this device.

The device in Figure 8 has been used to capture a new face database for use by the international research community [28]. It was placed at the entrance to a busy workplace and employees were simply told to "walk through the archway". The aim was to generate a database of natural poses where participants were not required to pose for the camera or have a particular type of expression. This is in contrast to most existing test databases, such as the well-used FRGC database [3]. Furthermore, PS inherently allows any combination of 2D data, 3D data or surface normals to be used for recognition.

Figure 8: A prototype PS face recognition device.

The PhotoFace database consists of a total of 1,839 sessions of 261 different subjects. Using a range of existing methods for recognition in [28], we easily acquired rank 1 recognition rates of ~85% on this database and equal error rates of 5% for verification (the rate corresponding to equal values of false acceptance and false rejection). Early results suggest that higher recognition rates are possible using novel dimensionality reduction techniques [29] or the ridgelet transform [30]. For example, in a recent adaptation of [29], we acheive over 96% recognition for a subset of 40 frontal images when compressing each face representation to a mere 61 dimensions via a variance analysis of the data. Expanding on the methods in [30], we represent faces in the domain of the ridgelet transform and obtain 100% recognition on galleries of 60 subjects, provided sufficient training data is used. It is intended that these recent developments will appear in the literature shortly.

6. Conclusion

This paper has presented an overview of PS and its application to automatic face recognition. It has been demonstrated that a PS device offers many advantages over other face-shape capture devices in that it is competitive in terms of cost, computational efficiency, resolution capabilities and ease of calibration. The paper showed how its shortcomings in terms of accuracy can be reduced using certain extensions of PS and its practicality can be improved using the high-speed camera set-up of the PhotoFace system. In future work, we hope to improve the system so that near infrared illumination is used so avoid the obtrusive flashing lights. We have already proven the principle of this and indeed shown that the surface reconstructions are marginally more accurate under near infrared illumination compared to visible light [15]. We also aim to demonstrate how the surface normals can be used to aid expression analysis.

References

[1] Zhao, W. and Chellappa, F. (eds.). Face Processing: Advanced Modeling and Methods. *Elsevier.* 2006.

[2] Bowyer, K. W., Chang, K. and Flynn, P. A survey of approaches and challenges in 3D and multimodal 3D + 2D face recognition. *Computer Vision and Image Understanding.* 101:1-15.

[3] Phillips, P. J., Flynn, P. J., Scruggs, T., Bowyer, K. W., Chang, J., Hoffman, K., Marques, J., Min, J. and Worek, W. Overview of the face recognition grand challenge. *Proc. CVPR.* 2005.

[4] Frankot, R. T. and Chellappa, R. A method for enforcing integrability in shape from shading algorithms. *IEEE Trans. Patt. Anal. Mach. Intell.* 1988, 10:439-451.

[5] Colbry, D. and Stockman, G. The 3DID face alignment system for verifying identity. *Image and Vision Computing.* 2009, 27:1121-1133.

[6] Chang, K., Bowyer, K. and Flynn, P. Face recognition using 2D and 3D facial data. *ACM Workshop on Multimodal User Authentication.* 2003, pp. 25-32.

[7] Hüsken, M., Brauckmann, M., Gehlen, S. and Malsburg, C. von der. Strategies and benefits of fusion of 2D and 3D face recognition. *in: IEEE workshop on face recognition grand challenge experiments,.* 2005, p. 174.

[8] Gökberk, B., Irfanoglu, M. O. and Akarun, L. 3D shape-based face representation and feature extraction for face recognition. *Image and Vision Computing.* 2006, 24:857-869.

[9] Zhang, R., Tsai, P. S., Cryer, J. E. and Shah, M. Shape from Shading: A Survey. *IEEE Trans. Patt. Anal. Mach. Intell.* 1999, 21:690-706.

[10] Brown, M. Z. and Hager, G. D. Advances in Computational Stereo. *IEEE Trans. PAtt. Anal. Mach. Intell.* 2003, 25:993-1008.

[11] Polhemus FastScan. [Online] [Cited: 12 6 2011.] http://www.polhemus.com/?page=Scanning_Fastscan.

[12] 3dMD Scanner. [Online] [Cited: 12 6 2011.] http://www.3dmd.com/3dmdface.html.

[13] Woodham, R. J. Photometric Method For Determining Surface Orientation from Multiple Images. *Optical Engineering.* 1980, 19:139-144.

[14] Argyriou, V. and Petrou, M. Photometric Stereo: An Overview. *Advances in Imaging and Electron Physics.* 2009, 159:1-54.

[15] Hansen, M. F., Atkinson, G. A., Smith, L. N. and Smith, M. L. 3D face reconstructions from photometric stereo using near infrared and visible light. *Computer Vision and Image Understanding.* 2010, 114:942-951.

[16] Coleman, E. and Jain, R. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Computer Graphics and Image Processing.* 1982, 18:309-328.

[17] Soloman, F. and Ikeuchi, K. Extracting the shape and roughness of specular lobe objects using four light photometric stereo. *IEEE Trans. Pattern Anal. Mach. Intell.* 1996, 18:449-454.

[18] Barsky, S. and Petrou, M. The 4-Source Photometric Stereo Technique for Three-Dimensional Surfaces in the Presence of Highlights and Shadows. *IEEE Trans. Patt. Anal. Mach. Intell.* 2003, 25:1239-1252.

[19] Saracchini, R. F. V., Stolfi, J. and Leitão, H. C. G. Shape reconstruction using a gauge-based photometric stereo method. *In Proc. CLEI.* 2009.

[20] Hertzmann, A. and Seitz, S. M. Example-Based Photometric Stereo: Shape Reconstruction with General, Varying BRDFs. *IEEE Trans. Patt. Anal. Mach. Intell.* 2005, 27:1254-1264.

[21] Smith, W. A. and Hancock, E. R. Facial Shape-from-shading and Recognition Using Principal Geodesic Analysis and Robust Statistics. *Intl' J. Comp. Vis.* 2008, Vols. 76:71-91.

[22] Hernández, C., Vogiatzis, G. and Cipolla, R. Shadows in three-source photometric stereo. *In proc. ECCV.* 2008, pp. 290-303.

[23] Georghiades, A. S. Recovering 3-D shape and reflectance from a small number of photographs. *In proc. Eurographics Workshop on Rendering.* 2003, 230-240.

[24] Kee, S. C., Lee, K. M. and Lee, S. U. Illumination invariant face recognition using photometric stereo. *IEICE Trans. Inf. Syst. E Ser. D.* 2000, 83:1466-1474.

[25] Zhou, S. K., Chellappa, R. and Jacobs, D. W. Characterization of human faces under illumination variations using rank, integrability, and symmetry constraints. *In proc. ECCV.* 2004, pp. 588-601.

[26] Zhou, S. K., Aggarwal, G., Chellappa, R. and Jacobs, D. W. Appearance characterisation of linear Lambertian objects, generalized photometric stereo, and illumination-invariant face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 2007, 29:230-245.

[27] 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.* 23:643-660.

[28] Zafeiriou, S., Hansen, M. F., Atkinson, G. A., Argyriou, V., Petrou, M., Smith, M. L. and Smith, L. N. The PhotoFace Database. *In. Proc. Biometrics Workshop of CVPR.* 2011, pp. 161-168.

[29] Hansen, M. F. and Atkinson, G. A. Biologically inspired 3D face recognition from surface normals. *In proc. ICEBT.* 2010, pp. 26-34.

[30] Kautkar, S., Koche, R., Keskar, T., Pande, A., Rane, M. and Atkinson, G. A. Face recognition based on ridgelet transforms. *In proc. ICEBT.* 2010, pp. 35-43.