The Photoface Database

Stefanos Zafeiriou[†], Mark Hansen[*], Gary Atkinson[*], Vasileios Argyriou[⋆],
Maria Petrou[‡,•], Melvyn Smith[*] and Lyndon Smith[*]
† Department of Computing, Imperial College London,
180 Queen's Gate, London SW7 2AZ UK.
* Machine Vision Lab, Faculty of Environment and Technology, University of the West of England,
Frenchay Campus, Bristol BS16 1QY UK.
⋆ Faculty of Computing, Information Systems and Mathematics, Kingston University London,
River House, 53-57 High Street, Kingston upon Thames, Surrey KT1 1LQ UK.
‡ Department of Electrical and Electronic Engineering, Imperial College London,
Exhibition Road, South Kensington Campus, London SW7 2AZ UK.
• Informatics and Telematics Institute, Centre of Research & Technology - Hellas
6th km Xarilaou - Thermi, Thessaloniki 57001 Greece
{s.zafeiriou,maria.petrou}@imperial.ac.uk, vasileios.argyriou@kinston.ac.uk
{mark.hansen,gary.atkinson,melvyn.smith,lyndon.smith}@uwe.ac.uk. [*]

Abstract

In this paper we present a new database suitable for both 2D and 3D face recognition based on photometric stereo, the so-called Photoface database. The Photoface database was collected using a custom-made four-source photometric stereo device that could be easily deployed in commercial settings. Unlike other publicly available databases the level of cooperation between subjects and the capture mechanism was minimal. The proposed device may also be used, to capture 3D expressive faces. Apart from the description of the device and the Photoface database, we present experiments from baseline face recognition and verification algorithms using albedo, normals and the recovered depth maps. Finally, we have conducted experiments in order to demonstrate how different methods in the pipeline of photometric stereo (i.e. normal field computation and depth map reconstruction methods) affect recognition/verification performance.

1. Introduction

Face recognition researchers have been collecting databases of face images for several decades now [15, Chapter 13]. While some databases can be regarded as superior to others, each of them is designed to test different aspects of recognition and have their own strengths and weaknesses. One of the largest databases available is the FERET database [17]. This has a total of 1199 subjects with up to 20 poses, two expressions and two light source directions. The FERET database was originally acquired using a 35mm camera. Others on the other hand, for example the widely used CMU PIE database [19] or the Harvard RL database [11], concentrate more on varying the capture conditions such as pose and illumination.

The PIE database is one of the most extensively researched. This is due to the fact that the faces are captured under highly controlled conditions involving 13 cameras and 21 light sources. The Yale B database [8] offers similar advantages to the PIE database except with an even larger number of lighting conditions (64), but just using ten subjects. Nine poses were considered per subject. The original Yale database [4] was designed to consider facial expressions, with six types being imaged for 15 subjects. Finally, the extended Yale B database was published. It contains 28 subjects with 9 different poses and 64 illumination conditions [14].

Even though the PIE [19], Yale [8] and extended Yale [14] databases provide facial samples taken under different illumination directions, they contain very few persons. More recently, the CMU Multi-PIE database [10] has been constructed with the aim of extending the image sets to in-

*This work was supported by the EPSRC project EP/E028659/1 Face Recognition using Photometric Stereo. This work has been conducted while Stefanos Zafeiriou was with Department of Electrical and Electronic Engineering, Imperial College London

clude a larger number of subjects (337) and to capture faces taken in four different recording sessions. This database was recorded under controlled laboratory conditions, as with the others mentioned above.

Recent trends in face recognition research to incorporate three-dimensional information into the recognition process lead to the collection of databases with 3D facial samples. This trend is due to the fact that a large number of viewing conditions adversely affect the 2D appearance of a face image but not the 3D appearance. This was the motivation for the FRGC2.0 database [16], which consists of a multipartition 2D and 3D database including a validation set of 4007 scans of 466 subjects. A Minolta Vivid 900/910 series laser range finder [24] was used for data capture.

This paper describes the construction of a new type of database of faces to aid research into face recognition. As explained, there is a growing number of related databases available for public research use. Each of these is designed to test different aspects of recognition, such as expression or illumination invariance. These databases have been built under meticulously designed and calibrated physical arrangements. The purpose of the new database described in this paper, however, is to capture a large number of faces in a more industrial setting. Most existing 3D capture devices (e.g. [23, 24]) are both financially and computationally expensive which can be highly inhibiting for commercial application. By contrast, we use a four-source high-speed capture arrangement, which permits the use of photometric stereo methods [22] to recover the 3D information with minimal computational expense. Furthermore, the device is significantly financially cheaper than most other 3D capture mechanisms.

Two important features of the database are that (1) the level of cooperation between subjects and the capture mechanism is minimal, and that (2) the capture process was carried out in a realistic commercial setting. For this reason, we placed the device near the entrance of a busy workplace and gave all of the volunteer subjects the sole instruction to "walk through the archway". This arrangement accurately simulates one of the ultimate goals for access-control face recognition, where there is no interaction required between the subjects and the technology. The database therefore offers an ideal testbed for face recognition algorithms designed for real world applications. As photometric stereo can be applied to the four images to calculate the 3D structure of the face, the database also allows for both 2D, 3D and hybrid algorithms to be evaluated.

In addition to describing the device and the database, we also present baseline experiments on the Photoface database applying baseline face recognition/verification techniques on albedo, depth and normal images. The focus of the conducted experiments is neither to compare various 2D/3D face recognition and verification methods nor to demonstrate that fusion of information of 3D and 2D data increase the recognition performance [5],[9]. The aim of the experiments conducted in this paper is: 1) to demonstrate how different methods in the pipeline of photometric stereo (i.e. normal field computation and depth map reconstruction methods) affect recognition/verification performance, and 2) to verify that a similar conclusion to [5] can be drawn for the modalities derived from photometric stereo methods. We applied three different photometric stereo methods in order to compute the normal field and the albedo image and five different integration methods that compute the height map from the normal field. To the best of the authors' knowledge this is the first experiment on a real-world photometric stereo database which also explores the effect of the use of different methods in the processing pipeline. In summary, the contributions of this paper are:

- The presentation of the first realistic commercial acquisition arrangement for the collection of 2D/3D facial samples using Photometric Stereo (PS).
- The presentation of the first facial image database collected under such a setting.
- The demonstration of how different methods in the pipeline of PS affect recognition/verification performance via a detailed set of recognition/verification experiments using a range of algorithms.

2. Capturing Device and Database Collection

The Photoface database was collected using a custom-made four-source PS device. Unlike previous constructions, our aim was to capture the data using a hardware that could be easily deployed in commercial settings. The setup is as follows: individuals walk through the archway towards the camera located on the back panel and exit through the side (Fig. 1). This arrangement makes the device suitable for usage at building entrances, high security areas, airports etc. The presence of an individual is detected by an ultrasound proximity sensor placed before the archway. This can be seen in Fig. 1 on the horizontal beam towards the left-hand side of it.

The hardware equipment used to create the entire system was the following:

- Camera: Basler 504kc with Camera Link interface operating at 200fps, 1ms exposure time, placed approximately at a distance of 2m from the head of the subject.
- Lens: 55mm, $f5.6$ Sigma lens.
- Light sources: low cost Jessops M100 flashguns, approximately at a distance of 75cm from the head of the subject.

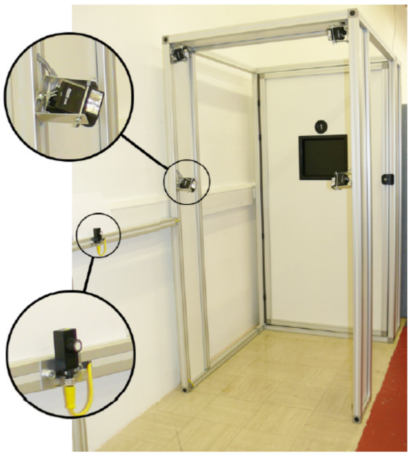


Figure 1. The image capturing device. One of the light sources and an ultrasound trigger are shown on the left. The camera is located at the back panel.

- Device trigger: Baumer highly directional ultrasound proximity switch. Range 70cm.
- Hardware IO card (for receiving and distributing triggers): NI PCI-7811 DIO.
- Frame grabber: NI PCIe-1429.
- Interfacing code: NI LabVIEW (the reconstruction and recognition algorithms were written in MATLAB).

The device also contains a monitor (as can be seen in Fig. 1) that provides instructions and indicates whether or not an individual was recognised, in the case of a recognition scenario, or whether an identity claim was accepted or rejected, in the case of a verification scenario.

The device captures one image of the face for each light source in a total of approximately 20ms. This time, also chosen for our experiments, was regarded as an adequately short period of time in which the inter-frame motion is no greater than a few pixels. The only case in which the performance of the system is expected to deteriorate significantly, is when a person runs passing through the device, due to the large inter-frame motion observed. For each person passing through the device, the following sequence of events takes place to capture the four images:

1. Await signal from ultrasound sensor.
2. Send trigger to camera.
3. Await integration enabled signal from camera.
4. Discharge first flashgun.
5. Await end of integration enabled signal.
6. Repeat from step 2 for the remaining light sources.

Figure 2. Four raw input images.

Figure 2 shows an example of four raw images of an individual (the resolution of the captured images were 1280×1024).

The capturing device was placed at the entrance of a busy workplace for a period of four months. Volunteer employees casually passed through the booth at regular intervals throughout this period. No instructions were given, other than to instructing them to walk through the archway looking at the camera or monitor. Thus, the volunteers typically passed through the device on their way in and out of the building. This arrangement is of great importance as:

1. It means that the capturing conditions were realistic for a real-world example. This is in contrast with existing face databases such as the widely used CMU-PIE database [10] or the FRGC database [16].
2. The whole setup was non-invasive, thus being suitable for any recognition algorithms developed for immediate commercial use.

3. Statistics of the Database

The Photoface database was collected in a period of four months (February 2008 to June 2008). It consists of a total of 1,839 sessions of 261 subjects and a total of 7,356 images. Some individuals used the device only once, while some others walked through it more than 20 times. The majority of people in the database are men (227 men over 34 women). The vast majority of the individuals are Caucasians (257 persons). Since there was no supervision, most of the captured faces in the database display an expression (for example more than 600 smiles and more than 200 surprises, open mouth, scream like expression etc. were recorded).

98 people walked through the device only once. For 126 of the 163 subjects that used the device more than once, the sessions were collected over a period of more than a week's

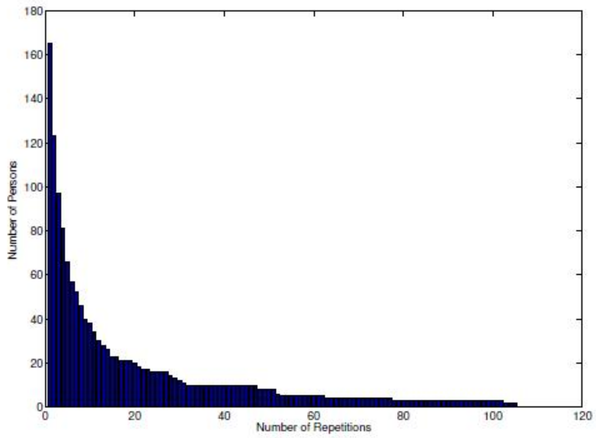


Figure 3. Four raw input images.

interval. For the majority of those (90 people), this interval was greater than one month. A histogram corresponding to the number of subject recordings by the device is depicted in Figure 3.

4. Photometric Stereo (PS) and Surface Reconstruction Methods

In this Section we summarise the standard PS method [6, §5.4], which we implemented using both three and four light sources. For these experiments we used an implementation of the standard PS method of Woodham [22]. We have mainly concentrated on a four-source version of the technique, although we have also compared our results with methods using three light sources. For the latter, we omitted the upper-right source in Fig. 1 from the computation. In order to do so, we examined a few reconstructions using various combinations of sources, resulting in the conclusion that deleting this specific light source was the most safe choice, as the performance was neither enhanced nor decreased, in comparison with the removal of one of the remaining three sources.

The standard PS method we used, assumes three or more greyscale images of a Lambertian object and constructs the following matrix equation from Lambert's Law for each pixel $\mathbf{x} = [x, y]$:

$$[\mathbf{I}_1(\mathbf{x})\ \mathbf{I}_2(\mathbf{x}) \cdots \mathbf{I}_N(\mathbf{x})]^T = \rho(\mathbf{x}) \left[\mathbf{l}_1^T\ \mathbf{l}_2^T \cdots \mathbf{l}_N^T\right]^T \mathbf{n}(\mathbf{x}) \tag{1}$$

where $\mathbf{I}_m(\mathbf{x})$ is the mth measured pixel brightness, $\mathbf{l}_m$ is the mth light source vector, N is the number of light sources, $\rho(\mathbf{x})$ is the reflectance albedo and $\mathbf{n}(\mathbf{x})$ is the surface unit normal. Examples of the raw images under 4-lights can be seen in Figure 2. The intensity values and light source positions are known a-priori and from these the albedo and surface normal components can be calculated by solving (1). (example of the computed albedo and surface using the 4-lights PS are shown in Figures 4 In our experiments, apart from the above mentioned PS method, we also applied the PS method proposed in [3].

In the following, we briefly review the problem of reconstructing a surface from the surface normals. In order to compute the shape of the surface, we need to obtain the depth map. This suggests representing the surface as $(\mathbf{x}, f(\mathbf{x}))$, so that the normal is a function of $\mathbf{x}$:

$$\tilde{\mathbf{n}}(\mathbf{x}) = \frac{1}{\sqrt{1 + \frac{\partial f}{\partial x}^2 + \frac{\partial f}{\partial y}^2}} \left(-\frac{\partial f}{\partial x}, -\frac{\partial f}{\partial y}, 1\right)^T \tag{2}$$

To recover the depth map, we need to determine $f(\mathbf{x})$ from the computed values of the unit normal.

Let us assume that the computed value of the unit normal at some point $\mathbf{x}$ is $\mathbf{n}(\mathbf{x}) = [a(\mathbf{x}), b(\mathbf{x}), c(\mathbf{x})]$, as calculated by (1). Then

$$\frac{\partial f}{\partial x} = \frac{a(\mathbf{x})}{c(\mathbf{x})} \quad \frac{\partial f}{\partial y} = \frac{b(\mathbf{x})}{c(\mathbf{x})}. \tag{3}$$

Here, we also perform another check on the data set. Let the images $\mathbf{P}(\mathbf{x}) = [\frac{a(\mathbf{x})}{c(\mathbf{x})}]$ and $\mathbf{Q}(\mathbf{x}) = [\frac{b(\mathbf{x})}{c(\mathbf{x})}]$. Because

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} \tag{4}$$

we expect

$$A(\mathbf{x}) \equiv \frac{\partial\left(\mathbf{P}(\mathbf{x})\right)}{\partial y} - \frac{\partial\left(\mathbf{Q}(\mathbf{x})\right)}{\partial x} \tag{5}$$

to be small (close to zero) at each point $\mathbf{x}$.

Assuming that the partial derivatives satisfy the above condition, we can reconstruct the surface up to some constant error in depth. The partial derivatives give the change in surface height with a small step in either the x or the y direction. This means that we can get the surface by summing these changes in height along some path. In particular, we have

$$f(\mathbf{x}) = \oint_C \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) \cdot \vec{dl} + c \tag{6}$$

where C is a curve starting at some fixed point and ending at $\mathbf{x}$, $\vec{dl}$ is the infinitesimal element along the curve and c is a constant of integration, which represents the unknown height of the surface at the starting point. All methods proposed for surface reconstruction solve the above problem with similar results. In our experiments we applied the surface reconstructions described in [7, 20, 8, 1, 2]. A recent discussion regarding the accuracy of different algorithms for face reconstruction from normals can be found [12].

Figure 4. The reconstructed surface with the computed albedo.

5. Face recognition/verification using Albedo and Depth Images

In this Section we outline the baseline methods used for feature extraction from albedo and depth images for face recognition and verification. The family of methods that we applied extract features using linear projections (also referred to as subspace methods). This family includes Principal Component Analysis (PCA) (the so-called Eigenfaces), Nonnegative Matrix Factorization (NMF) etc. In our experiments NMF [13] produced the best recognition and verification results. In subspace methods, like NMF, the facial images are lexicographically scanned in order to form vectors Let M be the number of samples in the image database $\mathcal{U} = \{\mathbf{u}_1, \mathbf{u}_2, .., \mathbf{u}_M\}$ where $\mathbf{u}_i \in \Re^n$ is a database's image. A linear transformation of the original n-dimensional space onto a subspace with m-dimensions ($m \ll n$) is a matrix $\mathbf{W}^T \in \Re^{m \times n}$. The new feature vectors $\mathbf{y}_k \in \Re^m$ are given by:

$$\mathbf{y}_k = \mathbf{W}^T(\mathbf{u}_k - \bar{\mathbf{u}}), \quad k \in \{1, 2, \ldots, M\} \tag{7}$$

where $\bar{\mathbf{u}} \in \Re^n$ is the mean image of all samples. Classification is performed using a simple distance measure and a nearest neighbour classifier using the normalized correlation.

5.1. Face Recognition using Normalfaces

In this paper we use a face recognition method based on the orientation of the normals. The baseline method is a very easy to implement method that is based on a novel representation of faces, the so-called Normalfaces. For an image $\mathbf{I}$ using the computed $\mathbf{P}$ and $\mathbf{Q}$ from PS we compute:

$$\mathbf{\Phi}(\mathbf{x}) = \text{atan}\frac{\mathbf{Q}(\mathbf{x})}{\mathbf{P}(\mathbf{x})} \tag{8}$$

which is an image that contains the normal orientations. We measure the orientations in the interval $\in [-\frac{\pi}{2}, \frac{\pi}{2}]$. For two images $\mathbf{\Phi}_1(\mathbf{x})$ and $\mathbf{\Phi}_2(\mathbf{x})$ we use the following dissimilarity measure:

$$d(\mathbf{\Phi}_1(\mathbf{x}), \mathbf{\Phi}_2(\mathbf{x})) = 1 - \frac{1}{NM\pi} \sum_{i=1}^{N \times M} |\mathbf{\Phi}_1(\mathbf{x}_i) - \mathbf{\Phi}_2(\mathbf{x}_i)|. \tag{9}$$

The above dissimilarity measure is then used in order to extract features using metric multidimensional scaling. Classification is performed using the normalized correlation in the new space.

6. Baseline Experiments

6.1. Recognition Experiments

We used the subset of images taken with more than a week's interval (126 people). For the majority of them (90 people) the interval was greater than one month. We assessed the recognition performance of all three modalities (i.e., albedo image, normals and height maps). Moreover, we experimented using fusing strategies. For the experiments presented here we tested using two setups:

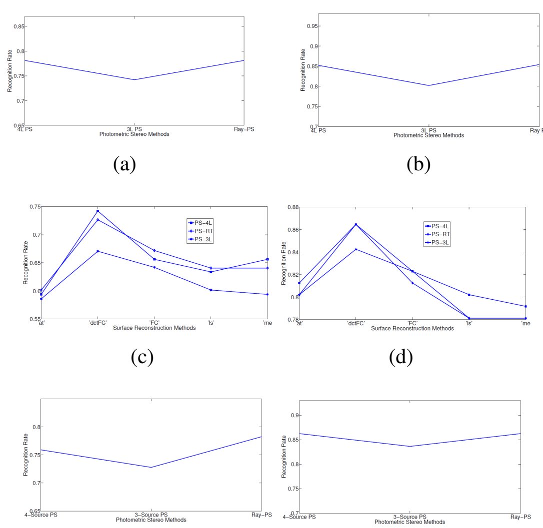
- In the first one, a very challenging experimental procedure was followed, exploiting only one grayscale albedo image, the surface normals derived from the application of PS, and the depth image derived from the integration of the normal field. Similarly, one grayscale albedo image, one set of normals and a height map was used for testing. Most of the training and testing images display a different facial expression. One sample face recognition is among the most challenging face recognition scenarios with various applications [21]. In [17], a face recognition scenario was designed based only on one sample per person for training. Similar recognition/verification experiments were also described for the FRGC database [16]. A similar scenario was tested in [5].
- In the second, two samples for training and one for testing were used. In our database we have 96 persons with three or more samples per person. The testing image for all 96 subjects was the same one used in the one sample experimental setup. This realization was implemented in order to test whether or not recognition using two samples of the same modality is better than fusing information across different modalities.

6.1.1 Face recognition from Albedo Images

Four source, three source and ray trace-based PS methods were employed for albedo computation. These methods are abbreviated as 4L-PS, 3L-PS and RAY-PS, respectively. The recognition rates using one albedo image for training and one for testing for all the tested PS methods are depicted in Figure 5 (a). As it can be seen the recognition rate is affected by the PS method applied and noticeably better recognition performance is achieved by PS methods that use all four illuminants. The best recognition rate was equal to 78%.

For the case of the two samples experiment, we used a decision fusion strategy similar to [5]. That is, we combined the matching scores for each person across the two samples of 2D albedo images and ranked the subjects based on the combined scores. Scores from each modality are linearly normalized to the range of $[0, 100]$ before combining. We explored various confidence-weighted versions of the sum, product and minimum rules. Among the fusion rules that we tested, the sum rule provided the best performance overall. The recognition rates for the two samples experiments is depicted in Figure 5 (b). As it can be seen, the use of more than two samples increases the recognition performance. Moreover, the methods which use all four illuminants achieved better recognition rates than those using only three. The best recognition rate was equal to 85%.

It is worth noting here that when using only the 96 persons of the second experiment in the first experiment the recognition rate was also about 78%, as well.



(a) (b) (c) (d) (e) (f)

Figure 5. Experiments using, (a) one albedo image for training and one for testing; (b) two albedo images for training and one for testing; (c) one depth image for training and one for testing; (d) two depth images for training and one for testing; (e) one Normalface for training and one for testing; (f) two Normalfaces for training and one for testing .

6.1.2 Face recognition from Depth and Normalface Images

We applied five different methods for surface reconstruction from the normal field. For the reconstruction methods we use the following abbreviations: 1) 'at' for the method in [1] 2) 'dctFC' for the DCT Frankot-Chellappa method [8], 3) 'FC' for the original Frank-Chellappa method [7], 4) 'ls' for the least square solution of the poison equation [20] and 5) 'me' for the reconstruction based on M-estimator [2]. The recognition rates for the one sample experiment and for all reconstruction and PS methods are plotted in Figure 5 (c). The best recognition result it was equal to 74%. As can be seen, PS and reconstruction methods greatly affect the recognition performance. More precisely, four source PS methods always achieve better recognition results. Moreover, the depth maps that were produced by dctFC constantly outperformed the performance of the depth maps produced by all other reconstruction methods.

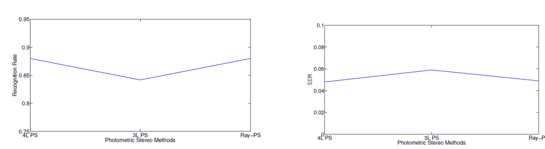
Experiments using two samples for training and one sample for testing were conducted in a similar manner as the ones for the albedo images. These results are depicted in Figures 5 (d). The best recognition result was equal to 86%.

The experiments using NormalFace for all tested PS methods are depicted in Figures 5 (e) and 5 (f) for one sample and two samples recognition, respectively.

6.1.3 Fusion 2D and 3D

Multimodal decision fusion is performed by combining the match scores for each person across the modalities of 2D albedo and depth image and ranking the subjects based on the combined scores in a similar manner as in the two samples experiments. The sum rule provided the best performance. We performed fusion only on depth images derived from the DCT-FC method. Fusion of intensity and geometry information was conducted only on the subset of persons that have more than 2 samples available in order to be directly comparable with the single modality two samples experiments. The recognition results from multimodal fusion using various PS methods are summarized in Figure 6 (a). The best recognition result was equal to 85%.

A summary of the best recognition results for the single modalities and multimodal fusion is given in Tables 1 and 2.



(a) (b)

Figure 6. Multimodal fusion results (a) for recognition, (b) for verification.

6.2. Verification Experiments

A person verification system should decide whether an identity claim is valid or not. The performance of face verification systems is measured in terms of the False Rejection Rate (FRR) achieved at a fixed False Acceptance Rate (FAR) [26, 25]. There is a trade-off between FAR and FRR. This trade-off between the FAR and FRR can create a Receiver Operating Characteristic (ROC) curve, where FRR is plotted as a function of FAR. The performance of a verification system is often quoted by a particular operating point of the ROC curve where FAR=FRR. This operating point is called Equal Error Rate (EER). The performance of the algorithms is quoted for the Equal Error Rate (EER) which is the scalar figure of merit that is often used to judge the performance of a verification algorithm. Verification experiments were conducted in the same database as well. The verification protocol was similar to the one defined in the FERET verification protocol in [18]. The probe (or client set) was defined by the 126 persons as in the recognition experiments. The first image is used for training while the second is used for testing client claims. The remaining 135 people in the database, with one image per person, are considered to be impostors.

In the second experiment we used two images from the 96 subjects for training while the third is used for testing client claims. The other 135 persons were used for impostor claims.

6.2.1 Face Verification using Albedo, Depth and Normalface Images

The EERs for various PS methods for the one sample experiment are depicted in Figure 7 (a). The verification results for the two samples experiment and for various PS methods are depicted in Figure 7 (b).

The EERs for various PS and surface reconstruction methods for the one sample experiment are depicted in Figure 7 (c). The verification results for the two samples experiment and for various PS and surface reconstruction methods are depicted in Figure 7 (d).

The EERs for various PS methods for the one sample experiment are depicted in Figure 7 (e). The verification results for the two samples experiment and for various PS methods are depicted in Figure 7 (f).

6.2.2 Multimodal Fusion

Multimodal decision fusion was performed exactly as in the recognition experiments case, by combining the match scores for each person across the modalities of the 2D albedo image and depth map and ranking the subjects based on the combined scores. The fusion results for verification using various PS methods are summarized in Figure 6 (b).

Table 1. A summary of the best percentage of recognition (PR) for all the conducted experiments across different modalities.

Perc. of Recognition (PR%)					
One Sample			Two Samples		
Albedo	Depth	Normal	Albedo	Depth	Normal
78	74	78	85	86	86

Table 2. A summary of the best PR for fusion across different samples and modalities.

Perc. of Recognition (PR%)			
Sample Fusion			Modality Fusion
Albedo	Depth	Normal	Albedo + Depth
85	86	86	85

Table 3. A summary of the best percentage of EER for all the conducted experiments across different modalities.

Verification (EER%)					
One sample			Two samples		
Albedo	Depth	Normal	Albedo	Depth	Normal
7.4	10.5	9.1	5.2	5.7	5.2

A summary of the best verification results for the single modalitities and multimodal fusion is given in Tables 3 and 4.

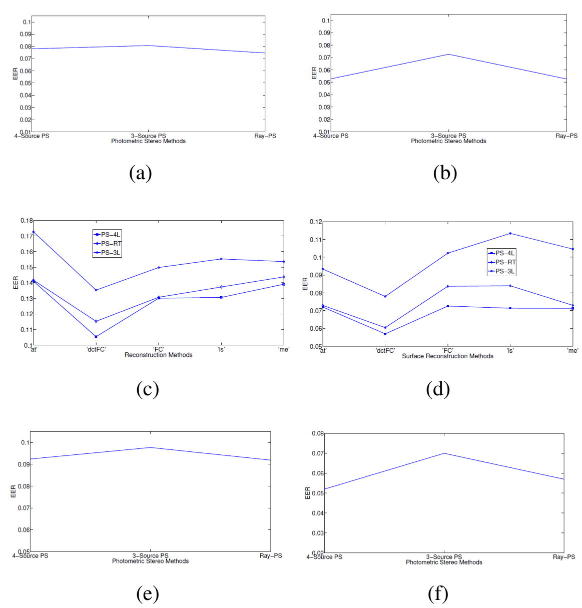


Figure 7. Verification experiments (a) using one albedo image for training; (b) using two albedo images for training; (c) using one depth image for training; (d) using two depth images for training; (e) using one Normalface image for training; (f) using two Normalfaces for training.

Table 4. A summary of the best percentage of EER for fusion across different samples and modalities.

Verification (EER%)			
Sample Fusion			Modality Fusion
Albedo	Depth	Normal	Albedo + Depth
5.2	5.7	5.2	5.2

7. Discussion and Concluding Remarks

In this paper, we presented a new database collected using a real life commercial setting based on photometric stereo. We presented the first experiments which demonstrate how different methods in the pipeline of photometric stereo affect the recognition performance and concluded the following: (1) Four source photometric stereo methods produce facial samples (albedo, normals) that achieve constantly better recognition and verification performance regardless of the reconstruction method applied. (2) The reconstruction methods greatly affect the recognition and verification performance. The method which constantly produces the best recognition/verification performance proved to be the one proposed in [8].

Moreover, we have verified most of the findings of [5]: (1) In most cases the best recognition and verification results of recovered albedo, normals and the reconstructed depth maps achieve approximately the same results, in some cases the recovered albedo produces better results. (2) Fusion of albedo and reconstructed surfaces produce significantly better results than using only the albedo or the depth images. (3) Fusion of two albedo images in the same way that we fused the results of albedo and depth map gave approximately the same recognition and verification results. Details on how the database can be provided to researchers are provided in http://Photoface.iti.gr/ or
http://www.uwe.ac.uk/research/Photoface.

References

[1] A. Agrawal, R. Chellappa, and R. Raskar. An algebraic approach to surface reconstruction from gradient fields. In *ICCV*, pages 174–181, 2005. 172, 174

[2] A. Agrawal, R. Raskar, and R. Chellappa. What is the range of surface reconstructions from a gradient field? *ECCV*, 3951:578, 2006. 172, 174

[3] V. Argyriou and M. Petrou. Recursive photometric stereo when multiple shadows and highlights are present. In *CVPR*, pages 1–6, 2008. 172

[4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE T-PAMI*, 19:711–720, 1997. 169

[5] K. Chang, K. Bowyer, and P. Flynn. An evaluation of multimodal 2D+ 3D face biometrics. *IEEE T-PAMI*, pages 619–624, 2005. 170, 173, 174, 176

[6] D. A. Forsyth and J. Ponce. *Computer Vision, A Modern Approach*. Prentice Hall, Upper Saddle River, NJ, 2003. 172

[7] R. T. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE T-PAMI*, 10:439–451, 1988. 172, 174

[8] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE T-PAMI*, 23:643–660, 2001. 169, 172, 174, 176

[9] B. Gokberk, H. Dutagaci, A. Ulas, L. Akarun, and B. Sankur. Representation plurality and fusion for 3-D face recognition. *IEEE T- SMCB*, 38(1):155–173, 2008. 170

[10] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *IVC*, 28:807–813, 2010. 169, 171

[11] P. W. Hallinan, G. G. Gordon, A. L. Yuille, P. Giblin, and D. Mumford. *Two- and Three- Dimensional Patterns of the Face*. A K Peters Ltd., 1999. 169

[12] M. Hansen, G. Atkinson, L. Smith, and M. Smith. 3D face reconstructions from photometric stereo using near infrared and visible light. *CVIU*, 114(8):942–951, 2010. 172

[13] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. In *NIPS*, pages 556–562, 2000. 173

[14] K. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE T-PAMI*, 27(5):684–698, 2005. 169

[15] S. Li and A. Jain. *Handbook of Face Recognition*. Springer-Verlag, 2005. 169

[16] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *In Proc. CVPR*, pages 947–954, 2005. 170, 171, 173

[17] P. J. Phillips, H. Wechsler, and P. Rauss. The feret database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16:295–306, 1998. 169, 173

[18] S. Rizvi, P. Phillips, and H. Moon. The FERET verification testing protocol for face recognition algorithms. *FG98*, pages 48–53, 1998. 175

[19] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression database. *IEEE T-PAMI*, 25:1615–1618, 2003. 169

[20] T. Simchony, R. Chellappa, and M. Shao. Direct analytical methods for solving Poisson equations in computer vision problems. *IEEE T-PAMI*, 12(5):435–446, 1990. 172, 174

[21] X. Tan, S. Chen, Z. Zhou, and F. Zhang. Face recognition from a single image per person: A survey. *Pat. Rec.*, 39(9):1725–1745, 2006. 173

[22] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Opt. Eng.*, 19:139–144, 1980. 170, 172

[23] www.3dmd.com/3dmdface.html. [Accessed 13 March 2010]. 170

[24] www.konicaminolta.com/sensingusa/products/3d/non contact/vivid910. [Accessed 15 Dec 2009]. 170

[25] S. Zafeiriou, A. Tefas, and I. Pitas. Exploiting discriminant information in elastic graph matching. In *ICIP*, volume 3, pages III–768, 2005. 175

[26] S. Zafeiriou, A. Tefas, and I. Pitas. Learning discriminant person-specific facial models using expandable graphs. *IEEE T-IFS*, 2(1):55–68, 2007. 175