Computer Vision and Image Understanding 114 (2010) 942–951

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

3D face reconstructions from photometric stereo using near infrared and visible light

Mark F. Hansen *, Gary A. Atkinson, Lyndon N. Smith, Melvyn L. Smith

Machine Vision Laboratory, University of the West of England, Bristol BS16 1QY, UK

ARTICLE INFO

Article history:
Received 3 August 2009
Accepted 1 March 2010
Available online 7 April 2010

Keywords:
3D reconstruction
Near infrared
Photometric stereo
Skin reflectance
3D face recognition

ABSTRACT

This paper seeks to advance the state-of-the-art in 3D face capture and processing via novel Photometric Stereo (PS) hardware and algorithms. The first contribution is a new high-speed 3D data capture system, which is capable of acquiring four raw images in approximately 20 ms. The results presented in this paper demonstrate the feasibility of deploying the device in commercial settings. We show how the device can operate with either visible light or near infrared (NIR) light. The NIR light sources offer the advantages of being less intrusive and more covert than most existing face recognition methods allow. Furthermore, our experiments show that the accuracy of the reconstructions is also better using NIR light. The paper also presents a modified four-source PS algorithm which enhances the surface normal estimates by assigning a likelihood measure for each pixel being in a shadowed region. This likelihood measure is determined by the discrepancies between measured pixel brightnesses and expected values. Where the likelihood of shadow is high, then one light source is omitted from the computation for that pixel, otherwise a weighted combination of pixels is used to determine the surface normal. This means that the precise shadow boundary is not required by our method. The results section of the paper provides a detailed analysis of the methods presented and a comparison to ground truth. We also analyse the reflectance properties of a small number of skin samples to test the validity of the Lambertian model and point towards potential improvements to our method using the Oren–Nayar model.

1. Introduction

Face recognition is now one of the most active areas of computer vision research. A wide range of different approaches have been proposed for the detection, processing, analysis and recognition of faces within images [1]. A recent trend has been to incorporate 3D information to aid recognition [2]. Unlike for 2D methods, the process of data capture is a complex procedure for 3D methods and may involve expensive and bulky hardware with computationally intensive algorithms.

In this paper, we make significant contributions to 3D face capture and processing by presenting a novel Photometric Stereo (PS) hardware device, a new PS algorithm for mitigating the effects of shadows within images, and a detailed set of experiments to assess the accuracy and practicality of the device. The new variation of PS estimates a field of surface normals by selecting the optimal combination of light sources to reduce the effects of shadow without requiring knowledge of the exact shadow boundaries. This is demonstrated on a novel high-speed practical 3D facial geometry capture device. We have also been successful at facial PS using near infrared (NIR) light. This offers several benefits to existing methods including exploiting skin phenomenology, creating a more covert capture system and making the system less intrusive. Extensive experimental results of these proposed advances are presented, including an analysis of skin reflectance qualities under NIR and visible light in terms of the Lambertian assumption.

In summary, the contributions of this paper are fourfold:

1. The development of 3D data capture hardware suitable for practical face recognition environments.
2. The development of a new algorithm for choosing the optimal light source configuration for each pixel in order to reduce the effects of shadows.
3. Detailed experiments to test the accuracy of the device on a variety of faces under visible and NIR light sources in terms of ground truth reconstructions and the Lambertian assumption.
4. Detailed experiments to assess the validity of the Lambertian assumption and a test to determine any possible improvements that may be possible using the Oren–Nayar reflectance model [3].

The remainder of this section provides an overview of related work and outlines the contributions to the state-of-the-art. Section 2 presents details of the hardware arrangement and image

* Corresponding author.
E-mail addresses: mark.hansen@uwe.ac.uk (M.F. Hansen), gary.atkinson@uwe.ac.uk (G.A. Atkinson), lyndon.smith@uwe.ac.uk (L.N. Smith), melvyn.smith@uwe.ac.uk (M.L. Smith).

1077-3142/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2010.03.001

acquisition process and our method to mitigate the effects of shadows. Detailed experimental results are then provided in Section 3. The implications and potential limitations of the work are discussed in Section 4.

1.1. Related work

The use of 3D information for face recognition has been attracting increasing attention in recent years [2,4,5]. This is due to the ability to overcome certain limitations associated with 2D recognition, e.g. problems of illumination and pose variance. Methods using 3D specific information also allow for representations which offer robustness to facial expression. Such methods include the 3D morphable model of Blanz and Vetter [6] and the geodesic representations of Bronstein et al. [7] and Mpiperis et al. [8]. Frequently, research which directly compares 2D and 3D recognition reports improved success rates for 3D recognition and that the best results occur when 2D and 3D information is fused [2]. As demand for practical face recognition systems is likely to increase, it is important that the most accurate methods are used and that the acquisition devices are both practical and affordable. There are a number of existing ways to capture and reconstruct 3D face information and the benefits and limitations of the most common approaches will now be discussed with the aim of putting our device into context.

Structured light scanning is perhaps the best known approach to generating 3D models of faces. This was used for generating the morphable head model in [6] and also for all the 3D faces used in the FRGC2.0 dataset [9], currently the largest publicly available 3D face database. For face capture, this technique works by scanning the object with a horizontal plane of laser light, capturing the line of light on a sensor and then calculating the location of each point via triangulation. The technique provides potentially very accurate scans; the Minolta Vivid 910 device [10] used for the FRGC2.0 dataset has a quoted accuracy of ±0.10 mm. However, these devices take about 2.5 s to capture the data, during which time the subject could move, thus distorting the reconstruction. It therefore requires a great deal of cooperation from the subject. They are also sensitive to high levels of ambient illumination. For these reasons, and the fact that they are financially costly, laser scanners are currently not suitable for many practical applications. The speed of acquisition can be improved by using striped patterns projected across the whole surface instead of using a scanning line. Distortions in this pattern can then used to calculate the 3D geometry of the surface [11], however accuracy is likely to be compromised and calibration can be time consuming.

The commercially available 3dMD system [12], which is used in this paper to acquire ground truth models, is an example of a projected pattern range finder. This device uses a number cameras to take images of an object from different positions. It uses a projected pattern to solve the correspondence problem between the images. The benefits of this system are its high accuracy (reported as <0.2 mm) and the speed of image acquisition (1.5 ms). However, the processing time is approximately 90 s for a face. This type of system is also expensive and requires a time consuming calibration procedure.

Shape-from-shading (SFS) is a technique for estimating 3D geometry from a single image [13]. Gradients of the surface are estimated from the patterns of intensity changes in an image. However the problem is ill-posed, meaning that there is no guarantee of a unique solution for a given image [14]. The main advantage of SFS is that it does not require any specialist capture apparatus such as laser scanners or projected pattern devices; merely a single ordinary camera. For this reason, finding solutions for the SFS problem are attractive to researchers. One way of overcoming the ill posed problem of SFS is to photograph the object multiple times under different illumination. This technique is known as Photometric Stereo (PS) and was first devised by Woodham [15] who showed that for any Lambertian surface, three differently illuminated images are sufficient to remove the ambiguity associated with a single image. Further details of this method will be given in Section 2.3.

A unique surface normal can be estimated by using three PS images provided that none of the light sources cast a shadow and the surface is Lambertian. In the case of a human face, shadows are frequently cast by features such as the nose. Indeed overcoming the detrimental effects of shadow on PS has been the subject of a number of papers. Smith and Hancock [16] use a statistical model to recover geometry in the presence of shadows. Hernández et al. [17] use two images where shadow is not present to estimate the value of the third where the shadow is present via integration. Coleman and Jain [18] use four light sources to over-determine the surface orientation. If a shadow is present in one image, it can simply be discarded. Similarly, Solomon and Ikeuchi [19] use four sources, but determine shadows and specularities by considerations of anomalies in albedo estimates that cannot be statistically attributed to camera noise. Barsky and Petrou [20] suggest a similar alternative solution to highlights and shadows by using a four source coloured light PS technique.

Georghiades extended PS beyond Lambertian surfaces to incorporate the Torrance and Sparrow model of reflectance [21] and created very accurate reconstructions [22]. However, a large number of images were required for the reconstruction which significantly increases surface computation and image acquisition time. Sun et al. [23] use five lights to handle shadows and specularities on non-Lambertian surfaces and show that a minimum of six lights are required in order fully realise any convex surface using photometric stereo. Using 20 images of a face, Ghosh et al. [24] build up a very detailed model of the skin's reflectance taking into account specular reflection and single, shallow and deep scattering. However, the images are captured over "a few seconds" which makes this approach unsuitable for our needs (i.e. practical applications). Also, their method would add a large amount of complexity for relatively little gain as skin follows Lambert's Law reasonably well (as shown in this paper for example).

Of the vast amount of research into automatic face recognition during the last two decades [1], relatively little work has involved PS. Kee et al. investigate the use of 3-source PS under dark room conditions [25]. They were able to determine the optimal light source arrangement and demonstrate a working recognition system. Zhou, Chellappa and Jacobs apply rank, integrability and symmetry constraints to adapt PS to face-specific applications [26]. Zhou et al. extended a PS approach to unknown light sources [27]. Georghiades et al. show how reconstructions from PS can be used to form a generative model to synthesise images under novel pose and illumination [28].

Comparing point clouds, the shape index, depth maps, profiles and surface normals in terms of face recognition performance, Gökberk et al. [5] concluded that surface normals provide the best features for face recognition. It is surprising therefore, that so few applications to date utilise PS, which inherently generates surface normals. The reason for this is likely to be that the availability and affordability of cameras with high enough frame rates, sensitivity and synchronisation capabilities for PS have only recently reached the market. Such cameras are necessary in commercial and industrial applications to effectively freeze the motion of the person while they may be moving by capturing several images in a short burst.

The majority of past work on PS has been conducted using visible illumination. As explained above, we also consider NIR light in this paper. Studies into the optical properties of skin have shown it to be increasingly reflective in the NIR light band up to wave-

lengths of about 1.1 μm [29]. This suggests that NIR, which is more covert and less intrusive, is a viable alternative to visible light. Furthermore, NIR can be used as a replacement for visible light because its proximity to the visual spectrum means that it is likely to behave in a similar manner on skin. We might expect some fine surface detail to be lost due to sub-surface scattering as reported by Zivanov et al. [30], but this is unlikely to affect overall face shape estimation. In addition to our work, infrared light has been used previously in 2D face recognition to mitigate the negative impact of ambient illumination [31,32] and to aid eye detection algorithms using the "bright eye" effect [33].

1.2. Contributions

The contributions of this paper to the state-of-the-art in 3D face capture and processing are via a novel system of hardware and algorithms. The new PS-based 3D face shape capture device is suitable for practical recognition environments and consists of four illumination sources placed evenly around a high-speed camera, as shown in Fig. 1. Individuals walk through the archway towards the camera located on the back panel and exit through the side. Compared to existing technologies, our device is cheap to build and involves exceptionally short image capture and processing times. The device is also able to operate at high resolution, is robust to ambient illumination and requires only minimal calibration. All images are captured in approximately 20 ms, resulting in only very small misalignment between frames. This allows subjects to be imaged as they casually walk through the archway.

We have tested our device using both visible and NIR illumination sources and found the latter to yield more accurate reconstructions when compared with ground truth. To the best of our knowledge, no published research has looked at using NIR light sources in PS for the purpose of face recognition. These considerations make our method attractive for use in many commercial and industrial settings such as at entrances to high security areas, airport check-in and border control.

For the main algorithmic contribution of this paper, we show how the effects of shadow can be mitigated using a related approach to that of Solomon and Ikeuchi [19] and Barsky and Petrou [20]. The method relies on estimating the likelihood of each pixel being in shadow and weights the contributions of the light sources in the PS computation accordingly. One advantage of our method is that neither the exact shadow boundary nor the camera noise parameters are required.

The final contribution of the paper is a detailed analysis of the quality of reconstructions and the nature of the skin reflectance properties. The device is tested on a variety of subjects and the RMS height errors and ℓ_2-norm errors are presented. Ground truth data is supplied by a 3dMD scanner. A quantitative analysis on the validity of the Lambertian assumption on skin reflectance is then presented. The extent of the discrepancies between the measured skin reflectance and Lambert's Law are demonstrated graphically and shown to be relatively minor for non-grazing angles. Lastly, we show that skin is more Lambertian under NIR illumination, solidifying our earlier claims about the feasibility of NIR as an alternative to visible light. This reflectance analysis also demonstrates the possibilities of improving the reconstructions by incorporating the Oren–Nayar reflectance model into the method.

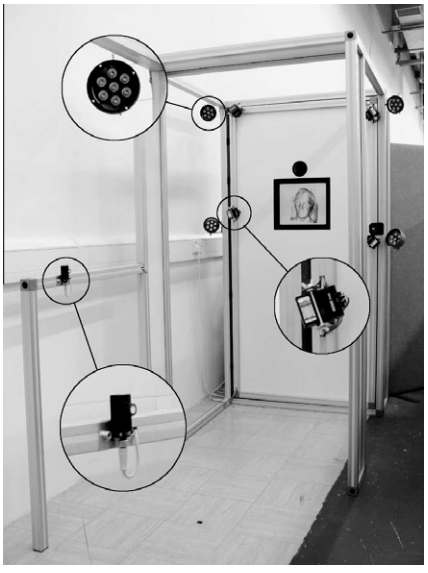

Fig. 1. The geometry capture device. Enlarged areas from top to bottom: a NIR light source, a visible light source and an ultrasound trigger. The camera can be seen on the back panel.

2. Method

This section first outlines the overall PS image acquisition hardware, before moving on to describe the reconstruction process. We also discuss the differences between our use of visible and NIR light sources. The problem of shadowing is then addressed, by presenting a new PS method to automatically select the optimal light source configuration.

2.1. Hardware

This section details the acquisition device hardware. The device, shown in Fig. 1, is designed for practical 3D face geometry capture and recognition. 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 the photograph. The sensor triggers a sequence of high speed synchronised frame grabbing and light source switching.

The aim is then to capture five images at a high frame rate: four images illuminated by the main light sources in sequence and an additional control image with only ambient illumination. Either one image per visible light is captured, or one image per NIR source. Note that the ambient lighting is uncontrolled (for the experiments presented in this paper, overhead fluorescent lights are present). The four visible light sources are low-cost Jessops M100 flashguns (colour temperature 5600 K), while the NIR lights are stripped down X-vision VIS080IR lensed 7-LED clusters, which emit light at ≈850 nm.

It was found experimentally that for people walking through the device, a minimum frame rate of approximately 150 fps was necessary to avoid significant movement between frames. The device currently operates at 200 fps, and it should be noted that it is only operating for the period required to capture the five images. That is, the device is left idle until it is triggered. A monitor is included on the back panel to show the reconstructed face or to display other information.

For visible light, the following sequence of events takes place to capture the five images as an individual passes through the device.

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.
7. Capture control image with ambient lighting only.

All interfacing code is written in NI LabVIEW. The ultrasound sensor is a highly directional Baumer proximity switch. When its beam is broken within a distance of 70 cm, it transmits a signal to an NI PCI-7811 DIO card fitted to a computer. When this signal is received, a trigger is sent to the camera. This is a Basler 504 kc camera with a 55 mm, f5.6 Sigma lens, placed 2 m from the subject. As with many silicon-based sensors, the Basler chip is responsive to both visible and NIR irradiance. The trigger is transmitted to the camera from a frame grabber via Camera Link®. The frame grabber is an NI PCIe-1429, which communicates with the DIO card via a RTSI bus for triggering purposes.

To ensure that the signal has reached the camera, and that the camera has commenced frame capture (i.e. is integrating), a second connection from the camera to the DIO card as added. This connection is TTL-high while the camera is integrating. When the computer receives this signal, the first light source is to be immediately illuminated. A flashgun is discharged by making a short circuit between its input pins. This is achieved here by sending a short pulse from the DIO card to the input pins via a phototransistor opto-isolator IC. This electrically isolates the sensitive DIO card from the high voltages of the flashgun terminals. Finally, the DIO card awaits the falling edge of the camera integration enabled signal before moving on to the next light source.

For NIR light, a slightly different procedure is adopted whereby synchronous TTL signals are sent to the camera and LEDs. This is because the LEDs can be illuminated for the duration of the camera exposure, while the flashguns only last for a small fraction of the exposure. The NIR LEDs are powered independently from the DIO card and interfaced via a simple transistor circuit. As the LEDs are illuminated for only 5 ms, it is possible to overpower them, in order to increase their brightness without causing damage. We therefore apply 20 V across the LEDs, compared to the recommended 12 V.

2.2. Visible and NIR comparison

One possibly negative aspect of the visible light set-up is that the firing of flashguns is obvious to the subject and possibly intrusive to any surrounding people. A possible advantage of NIR is that there may be additional subcutaneous or vascular structures present in the raw images taken under NIR light which may be used to aid recognition. Unfortunately, we found that such features were not visible in the wavelength band considered in this paper, but we aim to study this further in future work. NIR light is also more covert for a face recognition environment and subjects are less inclined to "pose" for the camera, meaning that more neutral expressions are likely. Finally, it is worth noting the advantage that many camera sensors are inherently more sensitive to NIR light.

One disadvantage of NIR illumination is the relative difficulty in obtaining the necessary brightness for the required short exposure times. While the flashguns were easily bright enough with an exposure time of 1 ms, an exposure of 5 ms was needed for the NIR LEDs (i.e. the maximum possible exposure for the given frame rate). Although this was adequate for our experiments, we had to use LED lenses that provided a narrow divergence angle, meaning that the face had to be more precisely positioned to obtain full illumination. For the visible light sources, the images were bright enough even for large diversion angles, removing the need for accurate positioning of apparatus and allowing subjects to pass through the archway without having to consider their exact location with respect to the camera.

To account for ambient illumination, the control image is subtracted from the other four images. These images are then normalised in terms of intensity before reconstruction takes place. This was done by linearly scaling the greylevels of each image so that the mean intensity was equal for each image. A detailed comparison of the resulting reconstructions is presented in Section 3.2.

2.3. Photometric stereo

Fig. 2 shows an example of four raw images of an individual using our prototype operating with the visible light sources. The person was slowly (≈1 m/s) but casually walking through the device. Each image has pixel dimensions of 500 × 400 and there are typically just a few pixel lengths misalignment between the first and last images. The face detection method of Lienhart and Maydt [34] is used to extract the face from the background of the image.

The four intensity images are processed using a MATLAB implementation of a standard PS method [35, Section 5.4]. Denote the general operation of PS by

$$\{\mathbf{n}_i\} = \mathtt{P}(\{I_{1,i}\}, \{I_{2,i}\}, \ldots, \mathbf{L}_1, \mathbf{L}_2, \ldots) \tag{1}$$

where $\{\mathbf{n}_i;\ i = 1,\ldots,N\}$ is the resulting set of surface normals, N is the total number of pixels, $\{I_{k,i};\ i = 1,\ldots,N\}$ is the set of intensities for image k, and $\mathbf{L}_j$ is the jth light source vector. For the bulk of this paper, we use four light sources, resulting in set of surface normals

$$\{\mathbf{n}_i\} = \mathtt{P}(\{I_{1,i}\}, \{I_{2,i}\}, \{I_{3,i}\}, \{I_{4,i}\}, \mathbf{L}_1, \mathbf{L}_2, \mathbf{L}_3, \mathbf{L}_4) \tag{2}$$

The general equation for PS using four sources for pixel i is

$$\begin{bmatrix} I_{1,i} \\ I_{2,i} \\ I_{3,i} \\ I_{4,i} \end{bmatrix} = \rho_i \begin{bmatrix} \mathbf{L}_1^T \\ \mathbf{L}_2^T \\ \mathbf{L}_3^T \\ \mathbf{L}_4^T \end{bmatrix} \mathbf{n}_i \tag{3}$$

where ρ_i is the reflectance albedo. The intensity values and light source positions are known, and from these the albedo and surface normal components can be calculated by solving (3). The resultant dense field of surface normals are then integrated to form height maps using the well-known Frankot and Chellappa method [36]. Fig. 2 shows the resultant reconstruction.

2.4. Optimising light sources

In many cases of PS usage, it is desirable to use all available light sources in the reconstruction in order to maximise robustness. However, where one or more sources do not illuminate the entire surface due to a self/cast shadow, it becomes disadvantageous to

Fig. 2. Left: four raw images. Right: reconstructions using standard PS (Eq. (3)), and the optimal source algorithm (Section 2.4).

use all the sources. In the case of a face, this is most likely to happen around the nose and outer edges of the cheeks, as shown in Fig. 2. In cases where the light source zenith angles (the angles between the viewpoint vector and the light source vectors) are small and the face is looking directly towards the camera, shadows tend not to be too problematic. When this is not the case, then more of the face is in shadow and the reconstructions become distorted. In such cases, it becomes beneficial to omit one or more sources from the PS computation at certain pixels. Assume for now that the faces are frontal, but that the light sources are repositioned to the archway (see Fig. 1) to increase the size of the shadows.

For this paper, we assume that no points on the face are shadowed by more than one source. In other words, each pixel is visible to at least three sources. In practice, a few areas to the sides of the nose and the extreme edges of the cheeks are sometimes shadowed by two sources, but we found that these areas get smoothed over by the subsequent surface integration and therefore have little bearing on the overall reconstructions. Ideally, we would like to determine which points are visible to which sources and only use these lights in the PS computation. However, performing this task precisely is difficult [20] and the resulting fields of surface normals tend to exhibit discontinuities around the estimated shadow boundaries. We therefore propose to adopt the following surface normal for an arbitrary pixel i:

$$\mathbf{n}_{\text{opt},i}(\varepsilon_i) = \varepsilon_i \mathbf{n}_{3,i} + (1-\varepsilon_i)\mathbf{n}_i \tag{4}$$

where ε is a measure of the likelihood of a pixel being in shadow and $\mathbf{n}_3$ is the surface normal estimated from the optimal three light sources. Where $\varepsilon = 0$, the pixel is definitely not in shadow and all four light sources are used. Where $\varepsilon = 1$, the pixel definitely in shadow and only three sources are used. For intermediate values of ε, a mixture of $\mathbf{n}$ and $\mathbf{n}_3$ are used. This has the dual effect that the shadow boundary does not need to be known precisely and that the discontinuities mentioned above become smoothed out.

Methods are therefore needed to determine $\mathbf{n}_3$ (i.e. which are the best three light sources to use) and ε for each pixel. For the former of these, it is adequate to simply use the three light sources that cause the brightest pixel for each point.

Each pixel has four measured intensities, one for each light source. Let us call the brightest intensity I_a, with corresponding light source vector $\mathbf{L}_a$. We shall call the second brightest pixel I_b and so on. We can therefore write our estimates of $\mathbf{n}_3$ as

$$\{\mathbf{n}_3\} = \mathbb{P}(\{I_a\},\{I_b\},\{I_c\},\mathbf{L}_a,\mathbf{L}_b,\mathbf{L}_c) \tag{5}$$

where we are omitting the index suffix, i, for the sake of simplifying the notation.

We know that for a pixel which is in a self shadow, the angle between $\mathbf{L}_d$ and the surface normal is greater than 90°. We can therefore define a self shadow condition as follows:

$$\arccos(\mathbf{L}_d \cdot \mathbf{n}_3) \geqslant 90^\circ \tag{6}$$

where this condition is satisfied, we know that the fourth light source is of no use and so we set $\varepsilon = 1$. Where the condition is not met, then *cast* shadows may or may not be present.

To obtain a suitable value of ε for pixels that do not satisfy condition (6), we compare the value of I_d to the value that which we would expect to measure in the absence of a cast shadow. Call this value I_{ex}. It is possible to estimate this quantity using a combination of $\mathbf{n}_3$, $\mathbf{L}_d$, the albedo found from the brightest three pixels (call this ρ_3), and Lambert's Law:

$$I_{\text{ex}} = \rho_3 \mathbf{L}_d \cdot \mathbf{n}_3 \tag{7}$$

For areas potentially in cast shadow from one source, we use the ratio between I_d and I_{ex} to determine ε. For areas deep in cast shadow, the expectation is that $I_d \ll I_{\text{ex}}$. For these regions we would like $\varepsilon \approx 1$, while for pixels away from shadow, we have $I_d \approx I_{\text{ex}}$, so we require $\varepsilon \approx 0$. Combining this logic with condition (6) we arrive at our final definition of ε:

$$\varepsilon = \begin{cases} 1 & \arccos(\mathbf{L}_d \cdot \mathbf{n}_3) \geqslant 90^\circ \\ \max\left(1 - \frac{I_d}{I_{\text{ex}}}, 0\right) & \text{otherwise} \end{cases} \tag{8}$$

where the "max" is required to deal with points that are not in shadow, but where I_d happens to be slightly greater than I_{ex}.

Fig. 2 shows the surface reconstructions resulting from both standard PS and using the method presented here. The new method was able to more accurately recover the regions of the face that are in shadow. This is especially true for areas of large surface zenith angle, such as the sides of the nose and outer edges of the face. The method proposed here is also able to restore a greater definition in the fine details of the face. Note however, that the method breaks down slightly near the far edges of the cheek, where the region is shadowed by two light sources. The result is that discontinuities in the reconstructed height appear in such regions. Further analysis of the method will be presented in Section 3.2, which also includes a comparison with the Solomon and Ikeuchi method [19].

3. Results

3.1. Basic reconstructions

Fig. 3 shows a series of reconstructions from the method described in Section 2 using visible light. The device was placed at the entrance to a workplace to ensure casual (and thus realistic) usage. The general 3D structures of the faces have clearly been well estimated. Note however, that the spectacles of one of the subjects have been "blended" into the face. This is a combined consequence of the rim of the spectacles being highly specular and the surface being non-integrable for this region of the image [36]. Although, we would ideally be able to estimate the shape of the spectacles accurately, the blending effect can potentially be beneficial to face recognition algorithms because it means that such details have a lesser impact on the overall reconstruction. A set of images and reconstructions using both visible and NIR light sources can be seen in Fig. 4. It is clear that NIR is also capable of providing good estimates of the 3D geometry of the face.

We now compare the accuracy of the face reconstructions against ground truth data. To do this, we scanned eight different faces using a commercial 3dMD projected pattern range finder

Fig. 3. Estimated geometry of three different subjects using visible light sources.

Fig. 4. Example raw images and reconstructions using visible (top) and NIR light sources for four subjects. For these experiments only, the subjects were asked to rest their chin on a support in order to ensure that all subjects are compared to each other in fair conditions.

[12]. The 3dMD models were rescaled so that the distance between tear ducts was the same as in the visible PS reconstruction. All reconstructions were then cropped to 160×200 px regions centred on the nose tip that encompass the eyebrows and mouth. Part of the forehead is omitted by this choice of cropping region as it is frequently occluded by hair and is therefore deemed unreliable for face recognition. An example of the face regions used for comparison can be seen in Fig. 5, which also shows a ground truth reconstruction acquired using a 3dMD scanner. The face regions from visible and NIR light sources are then aligned to ground truth using the Iterative Closest Point (ICP) algorithm [37].

Individual RMS and ℓ_2-norm error results between the reconstructions and ground truth are displayed in Fig. 6. The eight subjects consist of 6 males and 2 females and a mixture of Caucasian and Asian ethnicities. The variations in residual errors and ℓ_2-norm distances between visible and NIR reconstructions are significant according to paired t-tests ($p = 0.05$). This demonstrates that PS using NIR as a light source is a perfectly valid approach and leads to more accurate reconstructions.

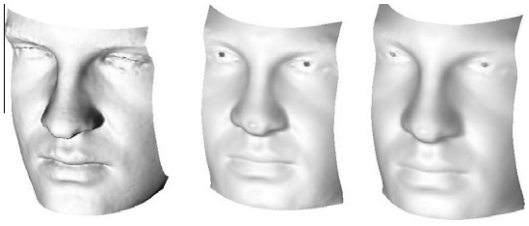

Fig. 5. 3D reconstructions for one subject from a 3dMD scanner (left) which is used as ground truth, PS using visible light sources (middle), and PS using NIR sources (right).

In order to get an indication of the regions where the greatest differences occur between ground truth and PS reconstructions, the residuals and ℓ_2-norm errors at each pixel are plotted in Fig. 7. Typically, the largest variations occur in regions with the highest curvatures, such as eye sockets, nose tips and the sides of the nose.

In attempting to produce the most accurate reconstructions possible via PS, it was found that the estimated surface normals could be enhanced by using normals acquired by re-differentiating the reconstructed height map estimate. It is unclear as to why this should be the case but preliminary analysis indicates that the reason may be due to the imposition of integrability constraints and the fitting of limited basis functions in the Fourier domain [36], as required by our adopted integration method. These factors may cause errant normals to be "smoothed out" leading to a more accurate reconstruction. However, if this method of improving reconstructions is used, a second integration step would be needed thus removing one of the benefits of PS for face recognition: that the surface normals (and hence distinctive differential surface features) are recovered directly. More research is required into this area in order to confirm that the improvements result from the imposed integrability constraints.

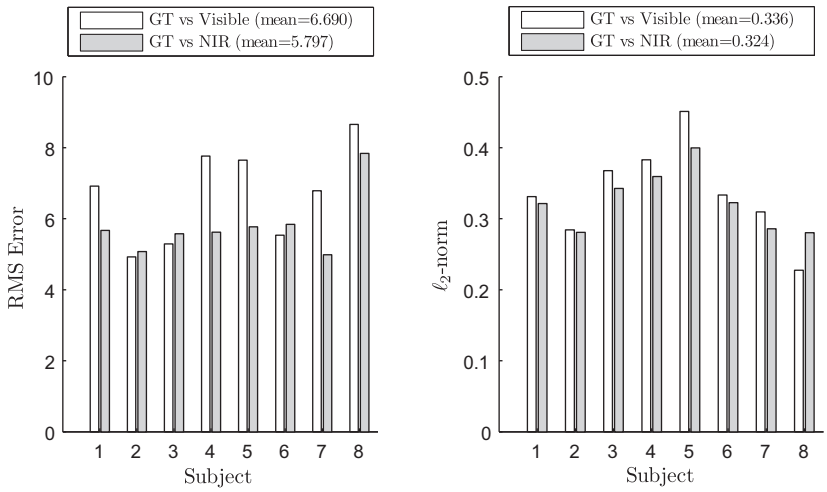

Fig. 6. RMS (left) and ℓ_2-norm (right) errors between Ground Truth (GT) and visible PS and NIR PS for each subject. *NB*: The order of subjects is arbitrary, i.e. there is no significance to the pattern that can be inferred from the ℓ_2-norm errors figure.

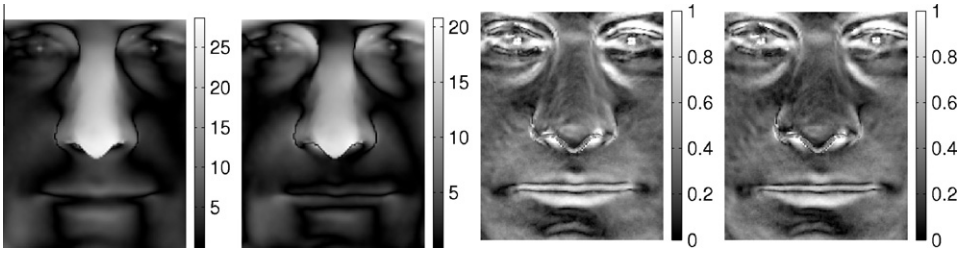

Fig. 7. Representative examples of the residuals and the ℓ_2-norm errors at each pixel. Left to right: residuals for visible and NIR respectively, ℓ_2-norm errors for visible and NIR respectively. Lighter areas represent larger errors.

3.2. Mitigating shadow effects

To demonstrate the improvements in surface normal estimates using the novel technique described in Section 2.4, we have manually selected a square region of the image that contains a strong shadow. Fig. 8 illustrates the surface normals of this region estimated by various methods. Two improvements of the proposed method compared to standard PS are noteworthy. Firstly, slightly finer details of the face are estimated using the new method. Secondly, the area immediately to the left of the nose is badly corrupt in the standard PS estimate, whereas the shadow is barely noticeable for the new method. To quantify the improvement, the ℓ_2-norm error was calculated between the two estimates and ground truth. Using standard PS for the region in Fig. 8, the ℓ_2-norm error was 0.32, while for the new method, the error dropped to 0.30. The difference in error between methods for the entire face is negligible as most regions are not in strong shadows.

Although the difference in ℓ_2-norm error is very small, this could be significant in certain applications as the nose and surrounding regions of the face offer useful biometrics. This area is seldom occluded by headgear/spectacles, varies considerably between individuals [38] and is relatively invariant to expression. Interestingly also, psychological research has shown that the nose is a preferred fixation point for humans attempting face recognition [39]. We should point out however, that although the surface normals and depth are improved through our method, the discontinuities in surface orientation at two-source shadow boundaries may cause reduced Sobolev-norm errors in some cases. In future work, we hope to reduce Sobolev-norm errors by treating two-source shadowed regions in a different manner to one-source shadowed regions.

For comparison, we have also implemented the Solomon and Ikeuchi PS method [19]. Their method is similar to ours in that combinations of three light sources are used to address shadowing/specularity issues. Where the albedo estimates from each combination of sources differs by an amount related to the standard deviation of camera noise at each pixel, σ_i, it is assumed that a shadow or specularity is present. For simplicity here, we assumed that σ_i is constant for all i and estimated the camera noise from 100 images of a planar white surface at close range. The ℓ_2-norm error for the region in Fig. 8 was 0.29 using the calculated value of $\sigma_i = 0.54\,\forall i$. Therefore, our method is comparable to the Solomon and Ikeuchi method in terms of accuracy. However, our method does not require camera noise information in order to attain optimum quality.

The method described in this section can enhance PS shape estimates for images that contain shadows. This means that PS can be used for facial reconstruction with somewhat arbitrarily positioned light sources. For the sake of simplicity, we will assume that the sources are positioned as in Fig. 1 for the rest of the paper and conduct the remainder of our analysis work using standard PS.

3.3. Reflectance analysis

To determine whether Lambert's law is obeyed more strictly under NIR light than visible, we have plotted graphs of I/ρ against θ, the angle between the light source and the normal vector. For a purely Lambertian surface, the relationship between the two should follow a cosine law. The results can be seen in Fig. 9. To generate the graph, values of I, ρ and θ were estimated for each pixel of each image for each of eight faces. The angle θ is calculated for each point of the face from the 3dMD scan data and the known light source vectors. The average values of I/ρ are used for each 1° increment in θ. The line at $\theta = 60^\circ$ indicates a reasonable cut-off point after which data points become too sparse to be significant. The RMS difference between the measured curves and the cosine curve in the range of $0 \leqslant \theta \leqslant 60$ is 0.04 (s.d. 0.11) for NIR light and 0.06 (s.d. 0.12) for visible. For completeness, the RMS difference across the whole curve is 0.11 (s.d. = 0.13) for NIR light and 0.17 (s.d. = 0.12) for visible. The figure demonstrates that skin under NIR light is marginally more Lambertian than under visible light.

Although the data suffers from significant noise levels (as indicated by a standard deviation exceeding 10% of the range for both conditions), the NIR condition has a lower RMS error and is therefore closer to the Lambertian curve than for visible light. This difference is significant given the large numbers of pixels and subjects used in the trials. This represents an average pixel intensity error of 10 greylevels for NIR and 15 for visible light across the image, assuming a maximum of 256 grey level intensities. This supports the hypothesis that skin is more Lambertian under NIR

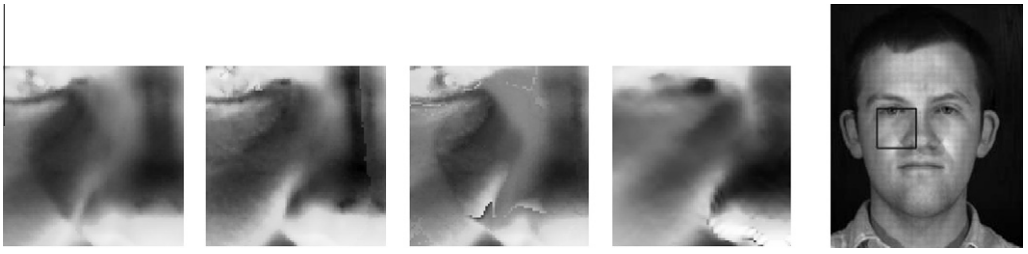

Fig. 8. Image of the vertical component of the surface normal (lighter areas indicate more downward pointing normals) for a region of a face in shadow. From left: estimates using standard four-source PS, using the proposed new algorithm, using the Solomon and Ikeuchi method, using the 3dMD scanner.

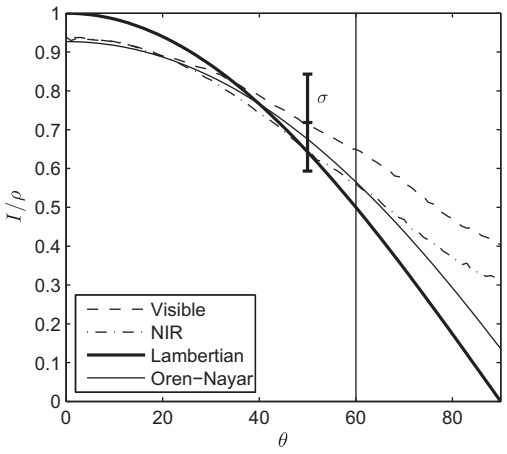

Fig. 9. Mean I/ρ values averaged over eight subjects against θ. To the right of the vertical line at $\theta = 60°$, data were too sparse to be of significance. For reference *one* standard deviation is shown to give an indication of spread.

illumination. We believe that this result is related to the fact that NIR light penetrates more deeply into the skin than visible light [40], which facilitates a more uniform scattering than surface reflection. Note however, that neither the Lambertian model nor the Oren–Nayar model (see below) take account of internal scattering or Fresnel effects. The results in Section 3.1 demonstrate that the more Lambertian behaviour associated with NIR light also leads to more accurate reconstructions.

A more detailed analysis for two individual subjects is shown in Fig. 10 and Table 1. What can be noted immediately is the similarity across the plots. There are small differences in I/ρ caused by different light sources but this appears to have little negative impact on the reconstructions and is likely to be due to environmental effects. The figure suggests that PS using both visible and NIR is robust to different skin types and light intensities. A more thorough analysis of the effects of gender and race on reflectance properties will be the subject of future work.

3.3.1. Comparison to the Oren–Nayar model

We have also compared our reflection measurements to the Oren–Nayar reflectance model [3], as shown in Fig. 9. The Oren–Nayar model represents the reflecting surface as an array of V-shaped groves of random orientation, commonly called "microfacets". The distribution of microfacet orientations is characterised by a roughness parameter and each facet is assumed to act as perfect Lambertian reflector. This model is able to account for the

Table 1
The RMS collective error across all eight reconstructions and for the first two reconstructions shown in Fig. 4 separately. The standard deviations are shown in brackets.

	Visible		NIR	
	RMS, $\theta \leqslant 60°$	RMS, overall	RMS, $\theta \leqslant 60°$	RMS, overall
All faces	0.06 ($\sigma = 0.11$)	0.16 ($\sigma = 0.12$)	0.04 ($\sigma = 0.12$)	0.11 ($\sigma = 0.13$)
Subject 1	0.07 ($\sigma = 0.09$)	0.16 ($\sigma = 0.18$)	0.05 ($\sigma = 0.12$)	0.10 ($\sigma = 0.22$)
Subject 2	0.07 ($\sigma = 0.10$)	0.17 ($\sigma = 0.18$)	0.04 ($\sigma = 0.13$)	0.12 ($\sigma = 0.21$)

common feature of limb-brightening and is itself based on the earlier Torrance–Sparrow model [41] where each microfacet is assumed to be mirror-like.

We have chosen to use the Oren–Nayar model as skin is not a smooth surface (especially on older people) and the model has been shown previously to be successful on a range of materials of varying degrees of roughness [3]. We do not believe that the microscopic structure of skin closely matches the Oren–Nayar model, but are merely demonstrating how alternate methods for reflection may improve our framework in future work. Investigating the various degrees of freedom of the BRDFs is also reserved for future work. Furthermore, there are additional models for skin reflectance which take account of a huge range of physical phenomena [42,43], but these are out of the scope of this paper.

The Oren–Nayar curve in Fig. 9 represents an example intensity profile for reference with a roughness parameter of 0.2. Clearly, this model fits the measured reflectance data significantly more accurately than the Lambertian curve, suggesting that the model could be incorporated into the method in the future. This will however, add significant complexity and computation time to the algorithm. This is because a minimisation method must be implemented in order to recover all the model parameters and to accommodate the increased number of angular degrees of freedom in the model.

4. Discussion

The results presented in Section 3 demonstrate that PS is an effective method for producing 3D facial reconstructions in terms of quality. Our method also requires a relatively short computation time. Using the device with standard PS, LabVIEW interfacing, MatLab processing and a typical modern PC, the time between device trigger and the reconstructed height map was approximately 4 s. Use of the optimised light source algorithm adds a further 3 s of computation time. The construction of the hardware also lends itself well to relatively unobtrusive data capture with a minimum

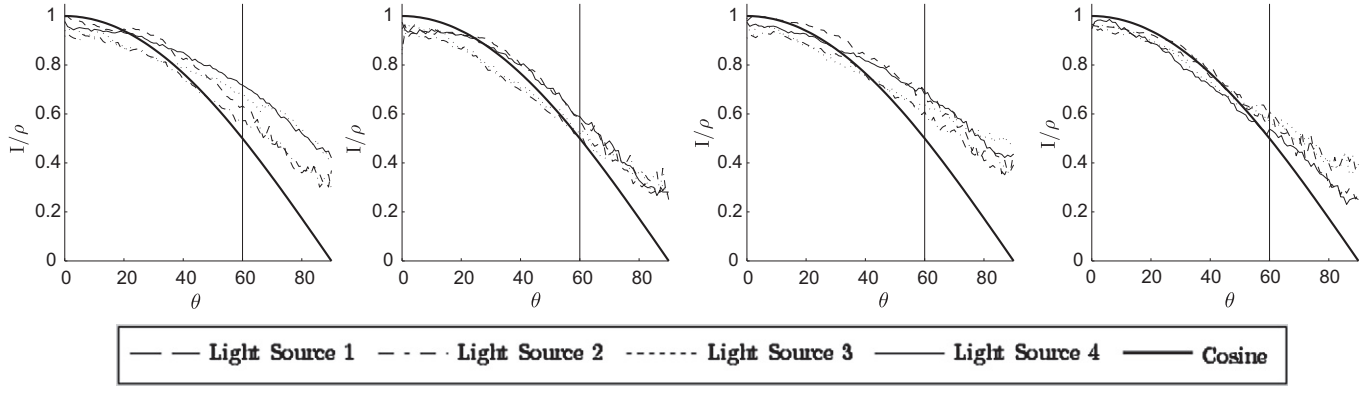

Fig. 10. I/ρ values from individual light sources plotted against θ for the first two reconstructions shown in Fig. 4. Left to right: Subject 1 under visible, Subject 1 under NIR, Subject 2 under visible, Subject 2 under NIR. The light sources are labelled clockwise from the bottom-left in Fig. 1.

amount of effort from the subject. Of particular interest are the following points:

1. The PS technique offers a valid alternative to existing, more expensive and processor intensive, 3D face capture methods.
2. The PS technique is robust to common facial features such as spectacles, makeup and facial hair (see also [44]).
3. NIR light sources produce reconstructions that are more accurate than visible light sources.
4. The optimised light source method described in Section 2.4 permits arbitrary light source arrangements and the presence of shadows.

Our system offers several benefits over commonly used existing laser triangulation and projected pattern 3D shape capture devices:

1. It is significantly cheaper to construct.
2. Acquisition time is shorter than laser triangulation systems.
3. Data processing time is shorter than projected pattern systems.
4. The method is robust to typical ambient illumination conditions.
5. It is very robust against accidental collisions (because it is tolerant to errors in the light source vectors).
6. Very fine details of the face can be reconstructed.
7. Calibration is very quick and simple and only needs to be performed after the initial light source positioning.
8. Due to the optimised light source method, the light sources can be positioned conveniently for different physical environments.
9. Although our system cannot reconstruct hair with high levels of accuracy, it can at least provide some details of its overall shape (see Fig. 3, for example). In contrast, laser triangulation and projected pattern systems usually fail completely with hair.

At present, the 3D reconstructions are not yet as accurate as those from projected pattern range finders. The reconstructions tend to be flatter than their real-world counterparts, with most protrusions understated. They do however provide extremely fine detail of a face such as wrinkles and pores. Even though the reconstructions suffer from a flattening of the features, they would still appear to be viable for recognition purposes (each reconstruction is clearly of a distinct identity) and the additional fine details could potentially be used as supplementary information to aid recognition.

The reconstructions under NIR were shown to be more accurate than those under visible light, but provided no additional 2D texture information. They also diminish the need for flashing lights, making the system less intrusive compared to visible light.

Zivanov et al. [30] offer an alternative argument to ours, stating that shorter wavelength light gives better results. Their justification is that shorter wavelengths undergo less internal scattering and thus provide a crisper, more defined reconstruction. It would appear therefore that a compromise must be reached in deciding between fine detail (using Zivanov's short wavelength suggestion) and overall geometry and covertness (using our NIR method).

4.1. Limitations and future research

One current limitation of the hardware described in this paper is that it does not cope with large deviations of peoples' height. Extremely tall or short people, or wheelchair bound persons would probably trigger the device correctly, but the location of the face could be outside of the field of view of the camera. Two possible solutions for this are (1) to use two cameras and trigger sensors at different heights or (2) to increase the field of view of the camera. The first solution would work by using the most suitable camera depending on which sensor had been triggered. While this is a straightforward solution it would increase the cost of the equipment considerably as the camera is the most expensive piece of apparatus. Increasing the field of view is also straightforward and would provide an adequate solution so long as wide-angle lens distortions did not become evident and that the face remains large enough on the image to provide discriminating information for the later recognition process.

Another improvement which could be made involves detecting the coordinates of the face and adjusting the light source vectors accordingly to improve the accuracy of the PS reconstruction. In the current system the light source unit vectors are calculated from a point at the centre of the camera's field of view and this is used for all reconstructions regardless of where the face is actually located. For this reason, the light source unit vectors are less accurate if the person walking through the device does not locate their face near the centre of the camera's field of view. The exact error caused by this inaccuracy is unknown, but amending the light source angles on a per person basis will improve the surface normal estimates.

5. Conclusion

This paper has brought together a number of advances in state-of-the-art 3D capture and processing technology. We have presented an algorithm for selecting the optimal light sources used for PS reconstruction which has the advantage over similar algorithms of not requiring knowledge of the exact shadow boundary.

The novel 3D facial geometry capture device has proved to be capable of reconstructing 3D models of faces under realistic workplace conditions using both visible and NIR light sources. It is cheaper, more robust and requires less calibration than alternative 3D acquisition devices. Although its reconstructions are less accurate than those of the state-of-the-art commercial 3dMD system, they are suitable for face recognition, which will be the focus of further study. The paper has also shown how human skin is more Lambertian under NIR light which is offered as an explanation for the associated improved accuracy of the reconstructions. A detailed error analysis for these results was also presented.

Acknowledgments

We would like to thank the EPSRC for funding this research and General Dynamics UK Ltd. for assistance in testing and data collection. We would also like to thank Imperial College, London and the Home Office Scientific Development Branch for their ongoing collaboration on this project.

References

[1] W. Zhao, R. Chellappa, Face Processing: Advanced Modeling and Methods, Academic Press, 2006.
[2] K.W. Bowyer, K. Chang, P. Flynn, A survey of approaches and challenges in 3D and multi-modal 3D+ 2D face recognition, Comput. Vis. Image Und. 101 (1) (2006) 1–15.
[3] M. Oren, S.K. Nayar, Generalization of the Lambertian model and implications for machine vision, Int. J. Comput. Vis. 14 (1995) 227–251.
[4] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surv. (2003) 399–458.
[5] B. Gökberk, M.O. İrfanoğlu, L. Akarun, 3D shape-based face representation and feature extraction for face recognition, Image Vis. Comput. 24 (8) (2006) 857–869.
[6] V. Blanz, T. Vetter, Face recognition based on fitting a 3D morphable model, IEEE Trans. Pattern Anal. Mach. Intell. (2003) 1063–1074.
[7] A.M. Bronstein, M.M. Bronstein, R. Kimmel, Three-dimensional face recognition, Int. J. Comput. Vis. (2005) 5–30.
[8] I. Mpiperis, S. Malassiotis, M.G. Strintzis, 3-D face recognition with the geodesic polar representation, IEEE Trans. Inform. Forensic Security (2007) 537–547.
[9] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: Proceedings of the CVPR, vol. 1, 2005.

[10] <www.konicaminolta.com/sensingusa/products/3d/non-contact/vivid910> (accessed 31.03.10).
[11] F. Chen, G.M. Brown, M. Song, Overview of three-dimensional shape measurement using optical methods, Opt. Eng. 39 (2000) 10.
[12] <www.3dmd.com/3dmdface.html> (accessed 31.03.10).
[13] B.K.P. Horn, Shape from shading: a method for obtaining the shape of a smooth opaque object from one view, Ph.D. thesis, MIT, 1970.
[14] P.N. Belhumeur, D.J. Kriegman, A.L. Yuille, The bas-relief ambiguity, Int. J. Comput. Vis. 35 (1999) 33–44.
[15] R.J. Woodham, Photometric method for determining surface orientation from multiple images, Opt. Eng. 19 (1) (1980) 139–144.
[16] W. Smith, E. Hancock, Facial shape-from-shading and recognition using principal geodesic analysis and robust statistics, Int. J. Comput. Vis. 76 (2008) 71–91.
[17] C. Hernández, G. Vogiatzis, R. Cipolla, Shadows in three-source photometric stereo, in: Proceedings of the ECCV, 2008, pp. 290–303.
[18] E.N. Coleman, R. Jain, Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry, Comput. Vis. Image Process. (1982) 309–328.
[19] F. Solomon, K. Ikeuchi, Extracting the shape and roughness of specular lobe objects using four light photometric stereo, IEEE Trans. Pattern Anal. Mach. Intell. 18 (1996) 449–454.
[20] S. Barsky, M. Petrou, The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2003) 1239–1252.
[21] K.E. Torrance, E.M. Sparrow, Theory for off-specular reflection from roughened surfaces, J. Opt. Soc. Am. A 57 (9) (1967) 1105–1112.
[22] A.S. Georghiades, Recovering 3-D shape and reflectance from a small number of photographs, in: Proceedings on Eurographics Workshop on Rendering, Eurographics Association, Leuven, Belgium, 2003, pp. 230–240.
[23] J. Sun, M. Smith, L. Smith, S. Midha, J. Bamber, Object surface recovery using a multi-light photometric stereo technique for non-Lambertian surfaces subject to shadows and specularities, Image Vis. Comput. 25 (7) (2007) 1050–1057.
[24] A. Ghosh, T. Hawkins, P. Peers, S. Frederiksen, P. Debevec, Practical modeling and acquisition of layered facial reflectance, in: International Conference on Computer Graphics and Interactive Techniques, 2008.
[25] S.C. Kee, K.M. Lee, S.U. Lee, Illumination invariant face recognition using photometric stereo, IEICE Trans. Inf. Syst. E Ser. D 83 (7) (2000) 1466–1474.
[26] S.K. Zhou, R. Chellappa, D.W. Jacobs, Characterization of human faces under illumination variations using rank, integrability, and symmetry constraints, Proc. ECCV (2004) 588–601.
[27] S.K. Zhou, G. Aggarwal, R. Chellappa, D.W. Jacobs, Appearance characterization of linear Lambertian objects, generalized photometric stereo, and illumination-invariant face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2) (2007) 230–245.
[28] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. (2001) 643–660.
[29] R.R. Anderson, J.A. Parrish, The optics of human skin, J. Invest. Dermatol. 77 (1) (1981) 13–19.
[30] J. Zivanov, P. Paysan, T. Vetter, Facial normal map capture using four lights – an effective and inexpensive method of capturing the fine scale detail of human faces using four point lights, in: GRAPP, 2009, pp. 13–20.
[31] S.Z. Li, R.F. Chu, S.C. Liao, L. Zhang, Illumination invariant face recognition using near-infrared images, IEEE Trans. Pattern Anal. Mach. Intell. (2007) 627–639.
[32] S.G. Kong, J. Heo, B.R. Abidi, J. Paik, M.A. Abidi, Recent advances in visual and infrared face recognition – a review, Comput. Vis. Image Und. 97 (1) (2005) 103–135.
[33] C.H. Morimoto, D. Koons, A. Amir, M. Flickner, Pupil detection and tracking using multiple light sources, Image Vis. Comput. 18 (4) (2000) 331–335.
[34] R. Lienhart, J. Maydt, An extended set of haar-like features for rapid object detection, in: IEEE ICIP, vol. 1, 2002, pp. 900–903.
[35] D.A. Forsyth, J. Ponce, Computer Vision: A Modern Approach, Prentice Hall Professional Technical Reference, 2002.
[36] R.T. Frankot, R. Chellappa, A method for enforcing integrability in shape from shading algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 10 (4) (1988) 439–451.
[37] P.J. Besl, H.D. McKay, A method for registration of 3-D shapes, IEEE Trans. Pattern Anal. Mach. Intell. 14 (2) (1992) 239–256.
[38] A. Moorhouse, A.N. Evans, G.A. Atkinson, J. Sun, M.L. Smith, The nose on your face may not be so plain: using the nose as a biometric, in: International Conference on Imaging for Crime Detection and Prevention, 2009.
[39] J.H.W. Hsaio, G. Cottrell, Two fixations suffice in face recognition, Psychol. Sci. 19 (2008) 998–1006.
[40] C. Fredembach, N. Barbuscia, S. Süsstrunk, Combining visible and near-infrared images for realistic skin smoothing, in: Proceedings of the IS&T/SID Color Imaging Conference, 2009.
[41] K. Torrance, M. Sparrow, Theory for off-specular reflection from roughened surfaces, J. Opt. Soc. Am. 57 (1967) 1105–1114.
[42] C. Donner, H.W. Jensen, Light diffusion in multi-layered translucent materials, ACM Trans. Graph. 24 (2005) 1032–1039.
[43] L. Li, C.S. Ng, Rendering human skin using a multi-layer reflection model, Int. J. Math. Comput. Simul. 3 (2009) 44–53.
[44] G. Atkinson, M. Smith, Facial feature extraction and change analysis using photometric stereo, in: Proceedings of the IbPRIA, 2009, pp. 96–103.