

BRDF estimation for faces from a sparse dataset using a neural network

Mark F. Hansen¹, Gary Atkinson¹ and Melvyn L. Smith¹

Centre for Machine Vision, Bristol Robotics Laboratory, University of the West of England & University of Bristol

Abstract. We present a novel five source near-infrared photometric stereo 3D face capture device. The accuracy of the system is demonstrated by a comparison with ground truth from a commercial 3D scanner. We also use the data from the five captured images to model the Bi-directional Reflectance Distribution Function (BRDF) in order to synthesise images from novel lighting directions. A comparison of these synthetic images created from modelling the BRDF using a three layer neural network, a linear interpolation method and the Lambertian model is given, which shows that the neural network proves to be the most photo-realistic.

1 Introduction

The Bi-directional Reflectance Distribution Function (BRDF) describes the relationship between observed intensity at a point on a surface as a function of the incident and reflected angles between the light source and the observer (Fig. 1). BRDFs are commonly used in Computer Generated Imagery (CGI) to provide photo-realistic rendering as well as being used for solving inverse problems associated with shape recovery. A BRDF completely describes the reflectance behaviour of an object under every possible illumination and observation direction assuming no subsurface light transport exists. A discrete representation of the BRDF therefore involves sampling the space in four dimensions (the two angles each to describe incidence and reflectance). As such, this leads to difficulties both in the practicalities obtaining and using the complete dataset. A BRDF is traditionally measured by using gonio-reflectometers, custom built devices which are expensive [1] and suffer from practical limitations such as angular precision and measurement noise. Some of these limitations can be overcome by employing reflectance models such as Lambertian [2], Phong [3], Torrance-Sparrow [4], Oren-Nayar [5] and more recently a tensor-spline based approach [6], and indeed, these have been used with great success. However, modelling is no substitute for the use of an accurate, image-based BRDF to capture the subtleties of the reflectance properties of an object. In this paper, we present a device which photometrically captures a set of images to create a sparse BRDF, which we then show can realistically model unsampled regions through use of an Artificial Neural Network (ANN).

This paper verifies the capture accuracy of our system and presents progress in obtaining BRDF data from a sparse dataset and using this model to simulate unseen lighting angles photo-realistically via an ANN. The motivation for this work is to show that an accurate BRDF can be modelled from the sparse dataset, and that high speed Near-InfraRed (NIR) capture is suitable for the photometric reconstruction of faces. **We make no claim that the modelled BRDF is state-of-the-art for skin reflectance modelling, for examples of such work please refer to [7], [8].**

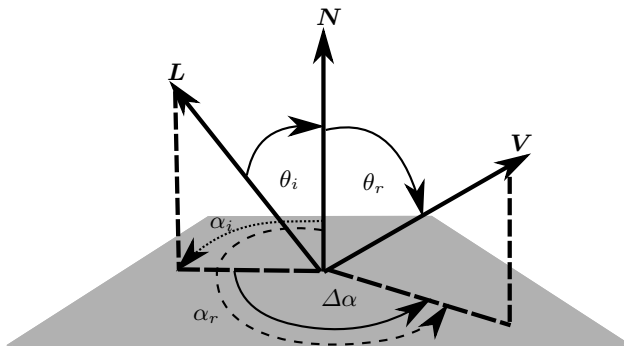


Fig. 1. The four dimensions (zenith incident and reflected angles: θ_i, θ_r and azimuth incident and reflected angles: α_i, α_r) upon which the observed intensity at V (viewer) depends. L is the light source vector, and N is normal to the plane of the reflecting surface. In order to reduce the dimensionality, $\Delta\alpha$, the difference between incident and reflective azimuths is used.

The contributions of this paper are i) to prove the practicality of using NIR for high speed 2.5D data captured in terms of speed and accuracy, compared with a commercial projected pattern scanner and ii) to demonstrate accurate modelling of the BRDF from only five lighting directions via an ANN, which is used to generate photo-realistic images from novel lighting angles.

2 Capture Device

2.1 Hardware

This section details the acquisition device hardware which is based upon the Photoface device presented in [9]. The device, shown in Fig. 2, is designed for practical 3D face geometry capture and recognition. The presence of an individual approaching the device is detected by an ultrasound proximity sensor placed before the archway. This can be seen in Fig. 2(6) 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 to capture six images at a high frame rate: one control image with only ambient illumination and five images each illuminated by one of the NIR light sources in sequence. A captured face is typically 700×850 pixels.

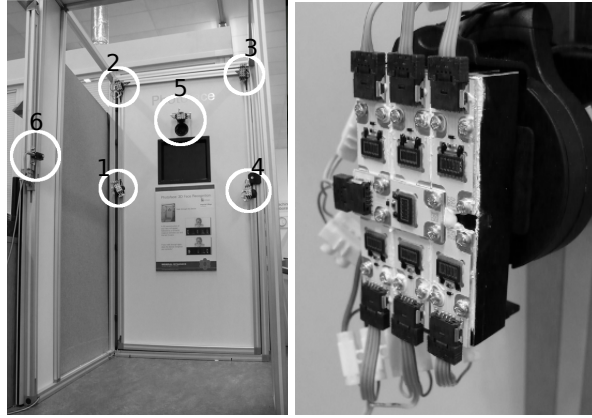


Fig. 2. The NIR geometry capture device (left) and an enlarged image of one of the clusters of LEDs (right). A camera can be seen on the rear panel, above which is located a NIR light source for retro-reflective capture (5). Four other light sources are arranged at evenly spaced angles around the camera (1-4). An ultrasound trigger is located on the left vertical beam of the archway (6).

Note that the ambient lighting is uncontrolled (for the experiments presented in this paper, overhead fluorescent lights are present). The five NIR lamps are made from a cluster of seven high power NIR LEDs arranged in an ‘H’-formation to minimize the emitting area (as can be seen in the right hand side image of Fig 2). The LEDs emit light at $\approx 850\text{nm}$. The light sources and camera are located approximately 1.2m from the head of the subject with four of the light sources arranged at evenly spaced angles, and one placed as close as possible to the camera to capture retro-reflection.

It was found experimentally that for people walking through the device, a minimum frame rate of approximately 150fps was necessary to avoid significant movement between frames. The device currently operates at 210fps, and it should be noted that it is only operating for the period required to capture the six 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.

2.2 Photometric Stereo

The face detection method of Lienhart and Maydt [10] is used to extract the face from the background of the five images. The five intensity images are processed using a MATLAB implementation of a standard Photometric Stereo (PS) method [11, §5.4].

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

$$\begin{bmatrix} I_{1,i} \\ I_{2,i} \\ I_{3,i} \\ I_{4,i} \\ I_{5,i} \end{bmatrix} = \rho_i \begin{bmatrix} \mathbf{L}_1^T \\ \mathbf{L}_2^T \\ \mathbf{L}_3^T \\ \mathbf{L}_4^T \\ \mathbf{L}_5^T \end{bmatrix} \mathbf{n}_i \quad (1)$$

where ρ_i is the reflectance albedo. The intensity values (I) and light source (\mathbf{L}) positions are known, and from these the albedo and surface normal components (\mathbf{n}) can be calculated by solving (1) using a linear least-squares method.

3 BRDF Modelling

Traditionally the generation of a BRDF involves illuminating an object from a large number of directions. Traditional PS illuminates an object using three light source directions, and we extend this to five. However this is still a very sparse amount of information from which to generate accurate reflectance information. We therefore explore the use of a traditional linear interpolation of the data, an ANN and the Lambertian reflectance model (which PS assumes) to model the reflectance information from novel lighting angles in order to see how well they can approximate the actual BRDF.

In order to minimise the number of dimensions, we assume that the surface is isotropic. This allows the use of $\Delta\alpha$ rather than the individual α_i, α_r values i.e. it is the difference between the azimuth angles that affects the reflectance, rather the orientation of that difference. While this assumption may not be perfect for human skin, the trade-off between accuracy and complexity makes it appealing.

3.1 Linear Interpolation of BRDF data

A traditional triangle-based linear interpolation method is used to model the regions between measured points. This method can be expected to work well when the distances between points are not too large and the surface being modelled is relatively uniform and predictable. As the sampled data does not fit this description well, we might expect the results to be poor. Delaunay tessellation is used to fit simplices to the sampled data and these are used to interpolate intensity values for the novel data points given the zenith angle of incidence and reflection (θ_i, θ_r respectively) and the difference in azimuth angle between the incidence and reflection angles ($\Delta\alpha$).

3.2 Neural Network Architecture for BRDF generation

Gargan & Neelamkavil [12] showed that using an ANN provides excellent approximation performance for a dense BRDF generated using a gonio-reflectometer. Experimenting with different numbers of layers (which affects the ability of the network to either generalise or overfit), they concluded that a three-tier feed-forward backpropagation architecture offers the best performance. **The same architecture is used in these experiments, but the novelty is that it is trained with a very sparse dataset instead of the dense BRDF used previously to test whether such good approximation performance is found. Additionally, Gargan & Neelamkavil use an XY parameter space or projected hemispherical space for inputs whereas we use a more lower dimensional co-ordinate space which assumes isotropism.**

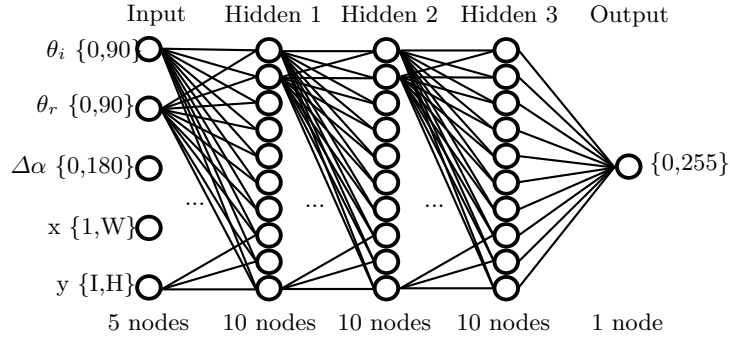


Fig. 3. The architecture of the ANN used to model the BRDF. The inputs θ_i and θ_r are in the range of 0-90 degrees, and $\Delta\alpha$ is in the range 0-180 degrees. x and y give the pixel coordinate and so are in the range of 1 to either the width (W) or height (H). The output is in the range 0-255 for each pixel, so that when all pixels are estimated, a full reflectance image will have been rendered.

The network architecture can be seen in Fig. 3 and was trained using the Levenberg-Marquardt optimisation backpropagation algorithm, taking in the region of 200 epochs to obtain a Mean Square Error (MSE) of 11.58 gray levels and an R value of 0.9598. Using fewer hidden layers generated higher MSEs and lower R-values although training times were faster, using four layers led to very slightly improved results but at a much higher computational cost. 100,000 image locations (approximately 20% of all data) were chosen at random to provide a representative sample of the whole face surface, and for each location, θ_i , θ_r and $\Delta\alpha$ as well as the x, y co-ordinates were used as inputs.

The reason for including the pixel coordinates is an attempt to allow for the different types of reflectance around the face to be captured (i.e. the reflectance of the skin at the nose tip is different to that of the cheeks). In doing so it is possible to correctly model the behaviour of different skin types when the same θ_i , θ_r and $\Delta\alpha$ are provided without having to label different regions of the face as having different skin types. This provides a means of unsupervised learning that will assist in improving the realism of rendered images.

4 Results

This section first presents results showing the reconstruction accuracy under NIR using a commercially available system as ground truth. Then, to assess the potential of the interpolation and ANN BRDF models, we use re-rendered images from the estimated surface normals obtained by PS. We use the BRDF models to generate images from novel lighting angles to see how well the models can generalise. We compare these images with those generated using the Lambertian reflectance model and show that the ANN produces the most photo-realistic images for unseen lighting angles.

4.1 Surface Reconstruction using Near-Infrared Photometric Stereo

Fig. 4 shows a reconstruction using NIR light sources for PS, a reconstruction using a commercially available 3D scanner and a map of ℓ_2 -distances and angular

errors between surface normals at each pixel location. They have been aligned using an Iterative Closest Point (ICP) algorithm¹. It can be seen that PS offers a very similar level of reconstruction to the commercial scanner – the largest differences occur around regions that are hard to integrate e.g. the lateral edges of the nostril. Median ℓ_2 -distance is 0.19 and median angular error (calculated by taking the dot product between corresponding 3dMD and PS vectors) is 11 degrees. These errors appear high, but looking at Fig. 4 (e) (ℓ_2 -distance) and (f) (angular error) it is possible to see that overall errors are low, but that **discrete areas around difficult to integrate areas where cast shadowing occurs (around the nose and lips) as well as the specularities caused by the eyes have extremely high errors.**

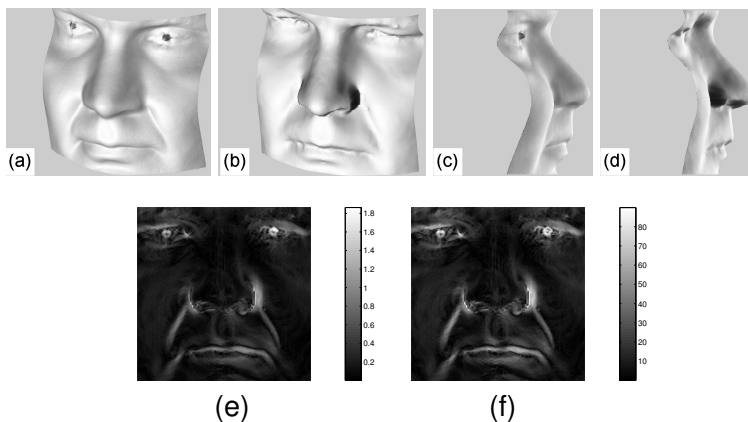


Fig. 4. Reconstructions from the Photoface device (a and c) and 3dMD (b and d) and a map of ℓ_2 -distance (e) and angular error (f).

4.2 Modelling the BRDF using linear interpolation and a neural network, and a comparison with Lambertian reflectance

Fig. 5 shows the results of using novel light source directions (i.e. different to the light source directions used by Photoface that have been used to model the reflectance). The first thing to note is that the images produced using the ANN (top row) show a high degree of realism, whereas the interpolated images shown in the second row are noisy and contain many artefacts, presumably due to the sparseness of the BRDF data. The images produced by assuming a Lambertian reflectance clearly show the lighting directions but again lack any real photo-realism.

5 Discussion

The previous results demonstrated the practicality of using the custom built NIR lamps for PS acquisition. The capture process itself is unobtrusive (most

¹ written by Ajmal Saeed Mian, Computer Science, The University of Western Australia

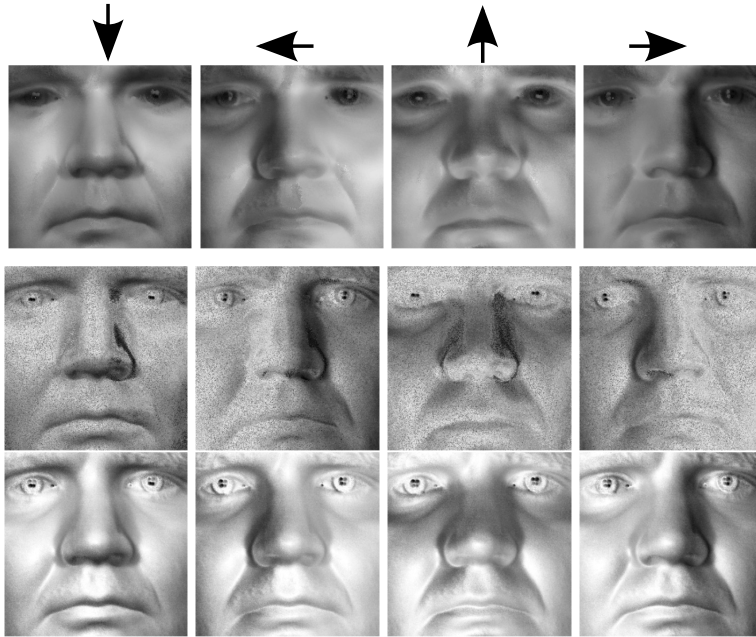


Fig. 5. Synthesised intensity images using estimated surface normals from PS and synthesised light angles (azimuth angles are indicated by arrows. The zenith angle is 15 degrees which is representative of Photoface light sources). Top row: ANN using Photoface surface normals, second row: interpolated Photoface surface normals, bottom row: images generated using the Lambertian reflectance model. A video of the rendering created by the ANN BRDF can be downloaded from www.cems.uwe.ac.uk/~mf-hansen/CAIP13/rerender75.avi

other PS techniques require a sequence of pulsed visible lights), takes only 30ms, and the results generated are accurate and high resolution.

In terms of BRDF modelling, the results show that photo-realistic images can be synthesised by using an ANN to model the BRDF **from a sparse dataset resulting from practical PS acquisition**. It offers more realistic results for novel lighting angles than either a linear interpolation based method or Lambertian model. The ANN offers a compact representation of the BRDF and a fast method of synthesising observed intensities from novel lighting directions.

There are some limitations of using a BRDF for modelling skin reflectance, especially under NIR. The BRDF describes the relationship between incident, reflected angles and observed intensity. However, there will be a certain amount of sub-surface scattering (and this will be increased under NIR which penetrates deeper into the skin) which the BRDF is not designed to capture. Also, the BRDF may deviate from actual values as we have used surface normals estimated by PS, but for purposes such as CGI this is not as important as the perceived realism (e.g. avatar generation). We have shown that photo-realistic results are achievable and future work will aim to overcome the Lambertian assumption by incorporating the BRDF model into normal estimates by iteratively enhancing

the accuracy of the surface normal representations, which in turn can then be used to generate a more accurate BRDF until convergence is reached.

6 Conclusion

We have presented a five source NIR, high speed and resolution 2.5D PS face capture device, which can be used to generate accurate 3D models of human faces. In addition, the five light sources are used to train an ANN to model the individual's BRDF. Using this modelled BRDF, photo-realistic results are attained from novel light source directions. Future work will look at the use of the BRDF to improve the 2.5D estimates by replacing the Lambertian assumption in PS, as well as using it as an additional biometric.

References

1. S. R. Marschner, S. H. Westin, E. P. F. Lafortune, K. E. Torrance, and D. P. Greenberg. Image-based BRDF measurement including human skin. In *Proceedings of the 10th Eurographics Workshop on Rendering*, page 139–152, 1999.
2. Jean-Henri Lambert. *Photometria, sive de Mensura et gradibus luminis, colorum et umbrae*. sumptibus viduae E. Klett, 1760.
3. Bui Tuong Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975.
4. K. E. Torrance and E. M. Sparrow. Theory for off-specular reflection from roughened surfaces. *Journal of the Optical Society of America A*, 57(9):1105–1112, 1967.
5. M. Oren and S. K. Nayar. Generalization of the lambertian model and implications for machine vision. *International Journal of Computer Vision*, 14(3):227–251, 1995.
6. R. Kumar, A. Barmpoutis, A. Banerjee, and B. C. Vemuri. Non-lambertian reflectance modeling and shape recovery of faces using tensor splines. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(3):533–567, 2011.
7. P. Debevec, T. Hawkins, C. Tchou, H. P. Duiker, W. Sarokin, and M. Sagar. Acquiring the reflectance field of a human face. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, page 145–156, 2000.
8. A. Ghosh, T. Hawkins, P. Peers, S. Frederiksen, and P. Debevec. Practical modeling and acquisition of layered facial reflectance. *ACM Transactions on Graphics*, 27(5):1–10, 2008.
9. M. F. Hansen, G. A. Atkinson, L. N. Smith, and M. L. Smith. 3D face reconstructions from photometric stereo using near infrared and visible light. *Computer Vision and Image Understanding*, 114(8):942–951, August 2010.
10. R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. In *IEEE International Conference on Image Processing*, volume 1, pages 900 – 903, 2002.
11. D. A. Forsyth and J. Ponce. *Computer Vision: A modern approach*. Prentice Hall Professional Technical Reference, 2002.
12. D. Gargan and F. Neelamkavil. Approximating reflectance functions using neural networks. In *Rendering Techniques' 98: Proceedings of the Eurographics Workshop in Vienna, Austria, June 29-July 1, 1998*, page 23, 1998.