Extracting Face Bump Maps From Video

James Paterson and Andrew Fitzgibbon*
Department of Engineering Science, The University of Oxford

Accurate, realistic and efficient rendering of human faces is of great importance in graphics applications. For high-speed rendering, a low-fidelity 3D model may be effectively enhanced by the use of *bump mapping* techniques which superimpose high-resolution surface normal maps on low-resolution geometry. Traditionally, obtaining high-quality bump maps for human faces has required high-resolution 3D models, or complex studio-bound lighting rigs. In this sketch we describe progress towards the extraction of high-quality bump maps from a single video camera and a static light source.

Our method extends the photometric stereo techniques used by Rushmeier et al [1997]. Their technique requires an array of light sources whose position is known. However, when dealing with non-Lambertian surfaces such as the face, many light sources are required (see also Debevec et al [2000]). Thus the technique imposes a significant setup cost. We aim at a modelling technique which can apply bump maps to a low-resolution morphable model, and which can acquire those bump maps from real people using readily accessible equipment with simple calibration procedures. Paterson and Fitzgibbon [2002] showed how a single light source and camera can be employed, if the object whose bump map is to be recovered is moved relative to the light and camera, but that work was limited to nearly-planar objects. In this work, we show how the technique can be extended to acquire dense surface normal information from the human face using simple equipment.

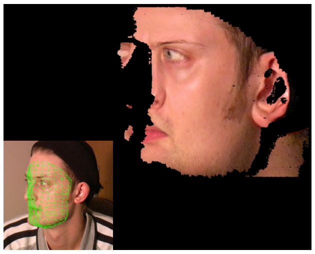
The principle employed is *photometric stereo*: given a set of images $\{I(i,j,t)\}_{t=1}^{T}$ of a static surface captured with a time-varying light source direction $\{\mathbf{l}(t)\}_{t=1}^{T}$, the surface normals $\mathbf{n}(i,j)$ of the static surface can be computed. If we consider a single point on the surface (corresponding to a single pixel (i',j') in the image sequence), the captured sequence is a set of intensities $\{I(i',j',t)\}_{t=1}^{T}$. If the surface is Lambertian, and we know the light-source directions $\mathbf{l}$, then a set of constraints on the normal $\mathbf{n}(i',j')$ is obtained: (quantities in green are known, those in red are unknown.)

$$\left[\mathbf{l}(1) \mid \cdots \mid \mathbf{l}(T)\right]^{\top} \rho\mathbf{n}(i',j') = \left[I(i',j',1),\ldots,I(i',j',T)\right]^{\top}$$

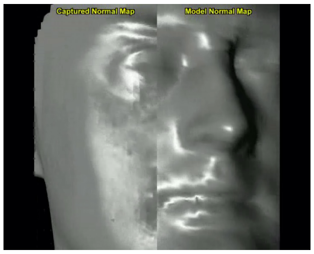
where ρ is the surface reflection coefficient. This system of equations is readily solved for $\rho\mathbf{n}(i,j)$ given at least three images, and for $T > 3$, a least-squares solution is obtained. The method can be applied to non-Lambertian surfaces if some care is taken. In our system, specular highlights and shadows are roughly removed by thresholding samples, with other non-Lambertian effects averaged out by using a large number of images ($T \approx 100$).

Two difficulties have prevented the wide use of photometric stereo for practical bump map extraction. First, it is difficult to know the position of a moving light source without a well-calibrated motion-control rig or robot arm. Second, in order to obtain the direction of a light not at infinity, a rough 3D model of the surface is required, which must be aligned with the images. We solve both of these problems by *moving the head, rather than the light source*. By tracking the head using a morphable model [Paterson and Fitzgibbon 2003], we can transform the moving-head, fixed-light sequence into a fixed-head, moving-light rendering. The tracking algorithm is a nonlinear optimization of the head rotation and translation parameters. The error metric uses the system graphics card to render the model over a captured background frame, and compares the rendered and input images using the mutual information metric. A coarse-to-fine search on each parameter in turn [Ferrari et al. 2003] is robust to the noisy error surface caused by the strong lighting changes in this sequence. Tracking yields the head position in camera coordinates, and the light position is found (once for a given set up) using the mirror-based calibration technique of Paterson and Fitzgibbon [2002]. Thus, for each point in texture coordinates, we obtain a set of constraints on the surface normal, which can then be robustly recovered.

Capture: The subject rotates his head in front of the (fixed) camera and light source.



Tracking: The head position is tracked, and the video image unwrapped into texture coordinates, simulating a moving light source and fixed head.



Output: The recovered normal map (left) is compared with the model normal map (right), as a Lambertian-shaded untextured rendering. The mole on the right cheekbone is correctly recovered.

The "Output" figure above shows the results of the new technique. Fine skin-surface detail is recovered, including a small mole on the subject's cheek. In comparison, the normal map derived from the morphable model is unrealistically smooth. However, there are areas of the recovered map which exhibit gross errors (due to tracking drift and occlusions giving too few samples), which would require manual postprocessing to remove. Future challenges for the method lie in improving the head tracking, which currently requires manual intervention to delete erroneous frames, and dealing in a more principled way with subsurface scattering and other lighting models.

Acknowledgements: We would like to thank Sony Computer Entertainment Europe for their support of this project.

References

Debevec, P., Hawkins, T., Tchou, C., Duiker, H. P., Sarokin, W., and Sagar, M. 2000. Acquiring the reflectance field of a human face. In *Proceedings of ACM SIGGRAPH*, 145–156.

Ferrari, V., Tuytelaars, T., and Gool, L. V. 2003. Wide-baseline multiple-view correspondences. In *Proc. CVPR*, 718–725.

Paterson, J. A., and Fitzgibbon, A. W. 2002. Flexible bump map capture from video. In *Eurographics Short Presentations*, http://www.robots.ox.ac.uk/~jamie/paterson02.html.

Paterson, J. A., and Fitzgibbon, A. W. 2003. 3d head tracking using non-linear optimization. In *Proc. BMVC.*, 609–618.

Rushmeier, H., Taubin, G., and Guéziec, A. 1997. Applying shape from lighting variation to bump map capture. In *Proceedings, Eurographics Workshop on Rendering*, 35–44.

---

*e-mail: {jamie,awf}@robots.ox.ac.uk