Photogeometric Scene Flow for High-Detail Dynamic 3D Reconstruction: Supplementary Material

Paulo F.U. Gotardo[1], Tomas Simon[2], Yaser Sheikh[2], and Iain Matthews[1]

[1]Disney Research [2]Carnegie Mellon University

Figure 1. Application in relighting: a face reconstructed with PGSF was relit in Maya with high-dynamic range environment lighting, subsurface scattering and different intensities of specular reflection that highlight surface detail to different degrees. Additional rendered frames are shown in the supplementary video.

1. Implementation Detail and Runtimes

In our experiments, face segmentation was performed automatically using the subject-independent face tracker of [5]. Temporal smoothing was applied on the detected landmark positions to remove small, high-frequency artifacts due to changing illumination. A simple image thresholding operation was used to enlarge the detected face region to include the full extent of the subject's forehead. Even though the employed face tracker also provides estimated locations of facial landmarks delineating eyes and mouth, these landmarks were not used for reconstruction as they are not required by PGSF. In addition, small detection errors in these landmarks would detract from the quality of our results. However, identity-specific active appearance models [4] can be trained on the recovered geometry and albedo to accurately segment mouth and eyes if necessary.

As a result, the average size of the face region was around 2.7 million pixels in our experiments. Reconstruction runtime was approximately 10 minutes for each 3-frame window on a recent laptop computer with 16GB of memory. The predominant computation time is the estimation of the four optical flows. To reduce runtimes, the minimization of both motion and surface energies in PGSF can benefit from GPU implementations in future work.

Initialization of the unknown optical flows assumes that at a very coarse level, image misalignment is negligible. Thus, as in standard optical flow estimation, image downsampling is initially performed, repeatedly, until the image at the top pyramid level has a dimension smaller than 50 pixels, yielding about 12 pyramid levels in our experiments – this was found to be enough for face capture with a camera framerate of 60Hz. During optimization, to improve the robustness of optical flow estimation against small relighting errors in intermediary surface results, PGSF follows the common approach of using image derivatives as features for image alignment, instead of using RGB values directly [6]. Other additional improvements found in recent optical flow methods (*e.g.*, anisotropic regularization [7]) can be incorporated into PGSF, but were currently left as future work.

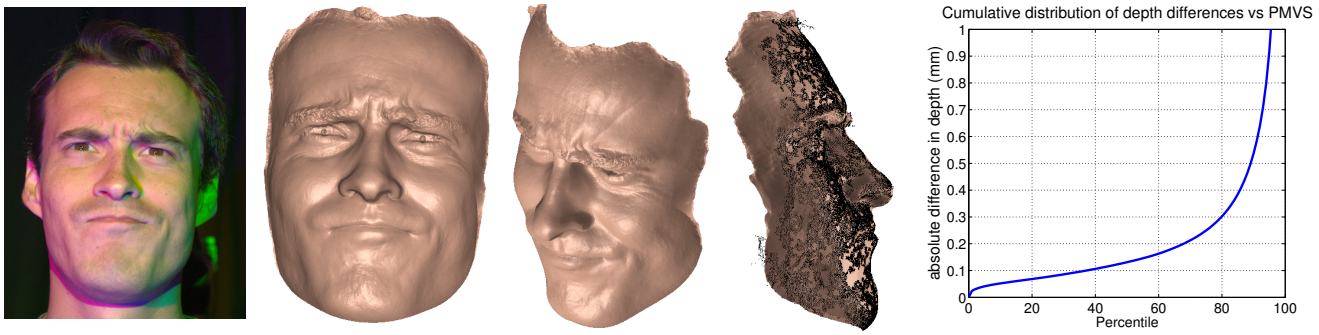

Figure 2. Detailed 3D reconstruction for a real frame in the supplementary video. The face profile view is overlaid with the PMVS point cloud, showing close agreement in recovered depth (median difference of 0.13 mm, 90^{th} percentile at 0.5 mm, over 100 frames).

Surface initialization via orthographic PS only leaves one degree of freedom to be fixed (a translation along the z-axis). We fix this scalar by calibrating cameras so that the origin is at the calibration object (*i.e.*, head position), as suggested in [1]. We have processed thousands of video frames and this simple initialization method has never failed to provide good convergence. As an alternative, triangulation of detected face landmarks could also be used to initialize this single z-translation. We currently do not reuse the results from the previous 3-frame window for initialization, but such an approach can be explored to shorten runtimes.

2. Evaluation against PMVS

Figure 2 shows the 3D reconstruction for a real video frame acquired with the setup above. Since ground-truth geometry is not available for these real images, we validate surface estimates against the popular PMVS algorithm [2]. PMVS is a state-of-the-art MVS method based on patch matching and does not require regularization, providing 3D point clouds instead of dense depthmaps. On a total of 100 video frames, PGSF and PMVS estimate depth consistently within fractions of a millimeter (see plot of depth differences in Fig. 2). However, PMVS triangulates spurious points at highly foreshortened areas and its results also lack the fine detail of PGSF.

3. Face Relighting

Full color albedo is a valuable asset in building realistic models for animation and for post-production. For instance, relighting is a frequent task faced by artists during movie and game production in which previously captured performances have to be adapted to match a certain environment. The application of PGSF in the realistic relighting of captured 3D faces is illustrated in Fig. 1 and in the supplementary video. In this example, the recovered 3D face was rendered in Autodesk Maya with high-dynamic range environment illumination and simulated subsurface scattering. We also show a second example with increased specular intensity, which highlights the fine scale detail in the recovered surface. Animated versions are available in the supplementary video.

Figure 3. Cross-polarized capture of temporal variation in RGB albedo due to changes in blood flow (see supplementary video).

4. Highlights and Polarization Filters

Currently, PGSF relies on the assumption of Lambertian reflectance. Highlights are treated as outliers that must be detected and ignored. To this end, an effective approach is to cross-polarize illumination and cameras, as in [3]. This approach not only removes highlights but also provides better observations of the diffuse RGB color of the inner skin layer just below the skin-air interface, as shown in Fig. 3. However, as discussed in [3], geometric detail is penalized due to more salient effects of subsurface scattering. These issues will be addressed in future work to further improve dynamic 3D reconstruction with PGSF.

References

[1] T. Beeler, B. Bickel, P. Beardsley, B. Sumner, and M. Gross. High-quality single-shot capture of facial geometry. *ACM Trans. Graph.*, 29(4):40:1–40:9, 2010.

[2] Y. Furukawa and J. Ponce. Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. PAMI*, 32(8):1362–1376, 2010.

[3] W.-C. Ma, T. Hawkins, P. Peers, C.-F. Chabert, M. Weiss, and P. Debevec. Rapid acquisition of specular and diffuse normal

maps from polarized spherical gradient illumination. In *Proc. Eurographics*, pages 183–194, 2007.

[4] I. Matthews and S. Baker. Active appearance models revisited. *IJCV*, 60(2):135–164, 2004.

[5] J. Saragih, S. Lucey, and J. Cohn. Face alignment through subspace constrained mean-shifts. In *Proc. IEEE ICCV*, 2009.

[6] D. Sun, S. Roth, and M. Black. Secrets of optical flow estimation and their principles. In *CVPR*, 2010.

[7] H. Zimmer, A. Bruhn, and J. Weickert. Optic flow in harmony. *Int. J. Computer Vision*, 93(3):368–388, 2011.