Binocular Photometric Stereo

Hao Du[1,3]
duhao@cs.washington.edu

Dan B Goldman[2]
dgoldman@adobe.com

Steven M. Seitz[1,3]
seitz@cs.washington.edu

[1] University of Washington
Seattle, WA, USA

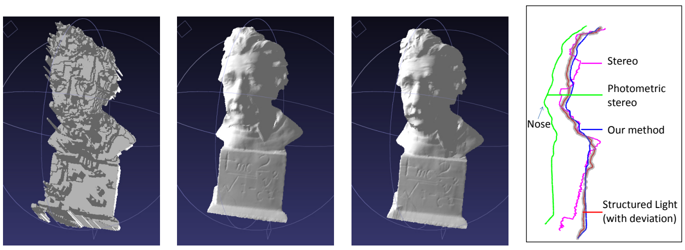
[2] Adobe Systems
Seattle, WA, USA

[3] Google Inc.
USA

Binocular stereo methods [1] yield relatively coarse shape reconstructions (Fig. 1(a)). This lack of geometric detail is intrinsic to the parallax cue and the fact that images are discrete. An additional limitation is that smooth untextured regions are hard to reconstruct. In contrast, *photometric stereo* [3] methods produce beautifully-detailed models (Fig. 1(b)), even in smooth untextured regions, due to their ability to directly estimate *continuous-valued* surface normals. However, photometric stereo lacks *metric shape information*; i.e., it is not possible to compute the absolute depth of scene and the relative depth of discrete objects at different depths.

This paper demonstrates that it is possible to achieve the best of both worlds—fine details and metric depth — by adding a second camera to the traditional photometric stereo setup. Our system's input consists of a stereo sequence (a synchronized pair of image sequences from two cameras) of a fixed object under a sequence of different illumination directions. Such a sequence can be produced, for example, by waving a light source around an object captured from a stereo rig. The output is a *continuous-valued depth map* with metric depths (Fig. 1(c), 1(d)).

We introduce a novel convex formulation for this problem of integrating parallax and shading cues, by casting it into the filter flow framework [2]. Disparity (from stereo) and surface normals (from shading) are non-linearly related and therefore challenging to jointly optimize. Our key insight is that both can be represented by entries of image filter kernels [2], which leads to a global solution via a single linear program.



(a) Stereo (b) PhotometricStereo (c) Our method (d) Cross section

Figure 1: Reconstructions using binocular stereo, photometric stereo and our method.

Assuming a rectified image pair, the parallax is strictly horizontal. Let f be the focal length, b the baseline, and $d^{\mathbf{u}}$ the disparity for pixel $\mathbf{u} = (u,v)$ on one image (i.e., the left image). The 3D position $P^{\mathbf{u}} = (X^{\mathbf{u}}, Y^{\mathbf{u}}, Z^{\mathbf{u}})$ is given by,

$$Z^u = \frac{fb}{d^{\mathbf{u}}} \qquad X^u = \frac{u}{f} Z^{\mathbf{u}} \qquad Y^u = \frac{v}{f} Z^{\mathbf{u}}. \tag{1}$$

The tangents at pixel $\mathbf{u}$ with respect to ∂u is:

$$T_u^{\mathbf{u}} = \frac{\partial P^{\mathbf{u}}}{\partial u} = \Big[-\frac{1}{f}\big(u \frac{\partial Z^{\mathbf{u}}}{\partial u} + Z^{\mathbf{u}}\big) \ ; \ -\frac{1}{f} v \frac{\partial Z^{\mathbf{u}}}{\partial u} \ ; \ \frac{\partial Z^{\mathbf{u}}}{\partial u} \Big], \tag{2}$$

and similarly for $T_v^{\mathbf{u}}$.

Binocular stereo requires brightness consistency between correspondences. Photometric stereo requires surface tangents perpendicular to their normals. These yield a minimization of an objective function (using L1-norm),

$$\| I_l(u,v) - I_r(u + d^{\mathbf{u}}, v) \|_1 \tag{3}$$

$$+\lambda \Big(\ \Big\| T_u^{\mathbf{u}}(P^{\mathbf{u}}) \cdot N^{\mathbf{u}} \Big\|_1 + \Big\| T_v^{\mathbf{u}}(P^{\mathbf{u}}) \cdot N^{\mathbf{u}} \Big\|_1 \ \Big), \tag{4}$$

in which disparities are the variables. Note that there's no closed-form representation for Eq.(3), and Eq.(4) is non-convex.

Using the filter flow formulation on rectified image pairs (Fig.2), each pixel corresponds to a 1D filter $M^{\mathbf{u}}$ that, when applied to the image pixels $I_r^i(u+j, v)$ on the right image, produces the value $I_l^i(u,v)$ matching pixel $\mathbf{u}$ on the left image. That is the *Data Objective*,

$$\sum_{\mathbf{u}} \| I_l^i(u,v) - \sum_j M_j^{\mathbf{u}} I_r^i(u+j, v) \|. \tag{DO}$$

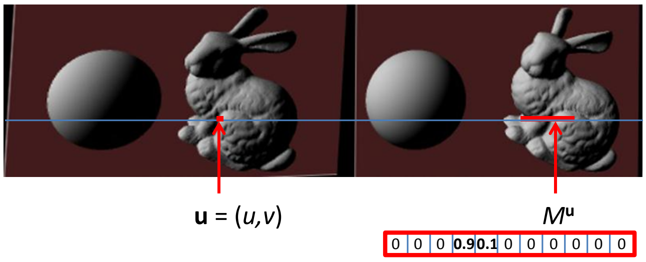


Figure 2: The principle of filter flow for stereo.

Regularizing the filters with *Non-negative* and *Sum-to-one* constraints,

$$M_j^{\mathbf{u}} \geq 0 | \forall j \quad , \quad \sum_j M_j^{\mathbf{u}} = 1, \tag{POS-M,SUM1-M}$$

the centroid of the filter,

$$d^{\mathbf{u}} = \sum_j j M_j^{\mathbf{u}}, \tag{5}$$

can be used to compute the disparity of pixels once filter entries are obtained, e.g., the filter in Fig.2 represents a shift of 1.9 pixels to the left when the energy distribution of filter entries is reaonsably compact.

With filter entries $M_j^{\mathbf{u}}$, we use the following weighted linear sum to approximate the pixel depth in Eqs.(1, 2),

$$\hat{Z}^{\mathbf{u}} = \sum_j M_j^{\mathbf{u}} Z_j^{\mathbf{u}} = \sum_j M_j^{\mathbf{u}} \frac{fb}{j}. \tag{6}$$

With $\hat{Z}^{\mathbf{u}}$, the approximated tangents $\hat{T}_u^{\mathbf{u}}$, $\hat{T}_v^{\mathbf{u}}$ are linear to filter entries, and thus is our *Normal Objective* (NO),

$$\sum_{\mathbf{u}} \Big\| \hat{T}_u^{\mathbf{u}}(P^{\mathbf{u}}) \cdot N^{\mathbf{u}} \Big\|_1 + \Big\| \hat{T}_v^{\mathbf{u}}(P^{\mathbf{u}}) \cdot N^{\mathbf{u}} \Big\|_1. \tag{NO}$$

Substituting the binocular term Eq.(3) with Data Objective (DO) and the photometric term Eq.(4) with Normal Objective (NO), subject to the Non-Negative (POS-M) and Sum-To-One (SUM1-M) constraints, the optimization is convex. A global minimum can be found using linear programming. Once an optimal solution to the entire filter flow is found, we use Eq. (6) to reconstruct the depths and use the projective geometry Eq. (1) to recover the 3D positions.

We demonstrate via simulated and real-world experiments that, utilizing both correspondence and normals with our approach, binocular photometric stereo is able to produce a reconstruction with high quality surface details and metric depth. Full details with analysis about an additional *compactness objective* are presented in the paper.

[1] David A. Forsyth and Jean Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, 2002.

[2] Steven M Seitz and S. Baker. Filter flow. In *ICCV*, 2009.

[3] Robert J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19:139–144, 1980.