Shape and Motion under Varying Illumination: Unifying Structure from Motion, Photometric Stereo, and Multi-view Stereo*

Li Zhang[1] Brian Curless[1] Aaron Hertzmann[2] Steven M. Seitz[1]
[1]Department of Computer Science and Engineering, University of Washington, Seattle, WA
[2]Department of Computer Science, University of Toronto, Toronto, ON

Abstract

This paper presents an algorithm for computing optical flow, shape, motion, lighting, and albedo from an image sequence of a rigidly-moving Lambertian object under distant illumination. The problem is formulated in a manner that subsumes structure from motion, multi-view stereo, and photometric stereo as special cases. The algorithm utilizes both spatial and temporal intensity variation as cues: the former constrains flow and the latter constrains surface orientation; combining both cues enables dense reconstruction of both textured and texture-less surfaces. The algorithm works by iteratively estimating affine camera parameters, illumination, shape, and albedo in an alternating fashion. Results are demonstrated on videos of hand-held objects moving in front of a fixed light and camera.

1. Introduction

When an object moves in front of a camera, its appearance changes in two fundamental ways: geometrically and photometrically. The former describes how points move in the image, i.e., optical flow. The latter reveals shading variation due to object rotation relative to the viewer and the light source. This paper combines both sources of information to estimate the optical flow, shape, motion, light, and diffuse albedo from a sequence of images.

Traditional shape reconstruction methods recover only a subset of scene properties and assume that either pose or shading is constant over all views. Although allowing both pose and shading to vary appears to complicate the reconstruction problem, we show that in fact it enables estimating flow and shape even in regions with little or no texture, thereby resolving a key ambiguity in prior methods.

This paper generalizes optical flow, photometric stereo, multi-view stereo, and structure from motion techniques under certain assumptions. We assume that objects move rigidly and are observed under orthographic projection; we also assume that surfaces have Lambertian reflectance and are illuminated by fixed distant lighting; furthermore, we assume no shadows, occlusions, or inter-reflections. Despite the fixed lighting, these assumptions imply that the illumination still changes relative to the moving object. We present an iterative algorithm that estimates camera motion, illumination, shape, and albedo in an alternating fashion, using both spatial and temporal shading variations. Our contributions can be interpreted in several different ways:

- **Optical flow with lighting variation.** Optical flow techniques traditionally assume the brightness constancy constraint. We employ a more general constraint allowing brightness to vary along optical flow.
- **Stereo matching with changes in lighting.** Stereo matching usually requires static lighting across all views. We lift this restriction in a principled way.
- **Photometric stereo for moving scenes.** Photometric stereo recovers shape from temporal shading variations, but requires a fixed object and camera. By computing flow under changing illumination, we generalize photometric stereo to moving objects.
- **Dense structure from motion.** Structure from motion recovers 3D positions for a sparse set of feature points. We show that texture-less regions can also be reconstructed, leading to dense surface reconstruction.

In the rest of the paper, we first review previous work and formulate optical flow under varying illumination as a subspace-constrained minimization. We then show how our formulation resolves ambiguities present in previous approaches. Finally, we present a reconstruction algorithm and demonstrate its performance on videos of real objects.

2. Previous work

In this section, we review previous work on motion analysis under temporal brightness variation.

Pentland [15] coined the term *photometric motion* to define the intensity change of a scene point due to object rotation, and proposed an algorithm to recover shape using this cue. Although the algorithm can handle non-Lambertian surfaces, it requires that optical flow be known a priori.

Woodham [22] described a technique for recovering optical flow under controlled illumination. He assumed that the

*This work was supported in part by National Science Foundation grants CCR-0098005 and IIS-0049095, an Office of Naval Research YIP award, the UW Animation Research Labs, and Microsoft Corporation.

object can be imaged two or more times for each pose, each time with different illuminations. Despite the restrictive assumptions, combining constraints from each image resolves the aperture problem, but still fails on uniform regions.

Several tracking techniques have been proposed to model lighting changes using predefined basis images [3, 7]. Other optical flow algorithms [8, 10, 13, 14] modeled lighting changes by introducing more parameters into the standard optical flow equation. Although these methods out-perform standard motion estimation, they require either large windows or global smoothness to regularize flow in low-contrast regions, often over-smoothing the results.

Stereo matching techniques have been extended to handle changes in shading or illumination due to object rotation, e.g., [11, 17, 19]. All of these methods use Lambertian reflectance to constrain matching in multiple images. However, these techniques do not directly compute surface normals or light source directions and therefore ambiguities arise in planar untextured regions.

All known optical flow and stereo algorithms fail to guarantee accurate matches in uniform intensity regions. This paper shows that even though flow is under-constrained in these areas, shape can still be accurately reconstructed by computing surface normals from shading variation over time. Our approach does not assume the lighting or spatial albedo distribution to be known a priori, a key difference from previous work on combining stereo and shape from shading [4, 6, 16].

3. Multi-frame optical flow under varying illumination

In this section, we formulate the optical flow problem under varying illumination using a subspace framework. This framework relates optical flow and intensity changes to surface positions, normals, motion, lighting, and albedo. We begin by describing a general form of optical flow that allows brightness variations.

Optical flow under intensity variation. Optical flow is the trajectory of a scene point in an image sequence. Let $\mathbf{x}_t = [x_t, y_t]^{\mathrm{T}}$ be the trajectory of a scene point $\mathbf{s} \in \mathbf{R}^3$ in an image sequence $I_t(x, y)$. Traditionally, optical flow is computed assuming the *brightness constancy* constraint:

$$I_t(\mathbf{x}_t) = I_0(\mathbf{x}_0). \tag{1}$$

If the motion vector $\mathbf{u}_t = \mathbf{x}_t - \mathbf{x}_0 = [u_t, v_t]^{\mathrm{T}}$ is small, linearizing Eq. (1) results in the *optical flow equation*

$$\nabla {I_t}^{\mathrm{T}} \mathbf{u}_t = I_0 - I_t \tag{2}$$

where $\nabla I_t = [\frac{\partial I_t}{\partial x}, \frac{\partial I_t}{\partial y}]^{\mathrm{T}}$ is the image gradient at $\mathbf{x}_0$ and I_0 and I_t are shorthand notations for image intensities $I_0(\mathbf{x}_0)$ and $I_t(\mathbf{x}_0)$ respectively. Assuming brightness constancy limits the applicability of most optical flow algorithms because the assumption is violated under varying illumination. In fact, the assumption is violated even when the light is static but the object moves relative to the light [15], e.g., a Lambertion object rotating under a directional light.

We now generalize Eq. (1) to describe optical flow under varying illumination. Specifically, we use a scaling variable $\gamma_t = \frac{I_t(\mathbf{x}_t)}{I_0(\mathbf{x}_0)}$ to represent intensity variation as introduced in [10, 14] and write the *generalized brightness constraint* as

$$I_t(\mathbf{x}_t) = \gamma_t I_0(\mathbf{x}_0). \tag{3}$$

Linearizing Eq. (3) results in a *generalized optical flow equation*

$$\nabla {I_t}^{\mathrm{T}} \mathbf{u}_t - \gamma_t I_0 = -I_t. \tag{4}$$

Notice that Eq. (2) constrains $[u_t, v_t]^{\mathrm{T}}$ to lie on a line in the $u - v$ plane and Eq. (4) constrains $[u_t, v_t, \gamma_t]^{\mathrm{T}}$ to lie in a plane in the $u - v - \gamma$ space. However, optical flow can not be computed from either Eq. (2) or Eq. (4) because more than one unknown variable exists in each constraint equation. To address this, we cast the optical flow estimation into a global framework, in which flows of multiple points over multiple frames are estimated together.

Suppose we have $\mathcal{F} + 1$ frames indexed by $t = 0, \cdots, \mathcal{F}$ and $\mathcal{P}$ scene points indexed by $p = 1, \cdots, \mathcal{P}$. We treat frame 0 as a reference frame and let $\mathbf{x}_{t,p} = [x_{t,p}, y_{t,p}]^{\mathrm{T}}$ and $\gamma_{t,p}$ be the position and the intensity scaling variable of scene point $\mathbf{s}_p \in \mathbf{R}^3$ in frame t. Optical flow and intensity variation can be estimated by minimizing the following objective function

$$\Phi(\{\mathbf{x}_{t,p}, \gamma_{t,p}\}) = \sum_{t=1}^{\mathcal{F}} \sum_{p=1}^{\mathcal{P}} \phi(\mathbf{x}_{t,p}, \gamma_{t,p}) \tag{5}$$

where $\phi(\mathbf{x}_{t,p}, \gamma_{t,p}) = (I_t(\mathbf{x}_{t,p}) - \gamma_{t,p} I_0(\mathbf{x}_{0,p}))^2$.

Eq. (5) involves a large number of inter-related variables $\{\mathbf{x}_{t,p}, \gamma_{t,p}\}$ and we constrain these variables by extending Irani's subspace method [9]. Specifically, we propose to impose subspace constraints on both flow trajectories and intensity variations to compute optical flow under lighting variation. We demonstrate that the lighting variation actually improves the flow estimation in low contrast regions. To simplify the problem, we assume a Lambertian object is moving rigidly in front of an orthographic camera, illuminated by a directional light and an ambient light.

Geometric constraints on flow. Following [9], we define constraints on optical flow arising from 3D motion in the scene. Assuming orthographic camera projection, we can relate flow trajectories and surface positions through

$$\begin{bmatrix} x_{t,p} \\ y_{t,p} \end{bmatrix} = \begin{bmatrix} {\mathbf{r}_{\mathrm{x}t}}^{\mathrm{T}} \mathbf{s}_p + o_{\mathrm{x}t} \\ {\mathbf{r}_{\mathrm{y}t}}^{\mathrm{T}} \mathbf{s}_p + o_{\mathrm{y}t} \end{bmatrix} \tag{6}$$

where $\mathbf{r}_{\mathrm{x}t}$ and $\mathbf{r}_{\mathrm{y}t} \in \mathbf{R}^3$ are the x and y camera axes for frame t, and $[o_{\mathrm{x}t}, o_{\mathrm{y}t}]^{\mathrm{T}}$ is the projected object origin in the

Problem	Known	Unknown
Structure from Motion	$\mathbf{X}, \mathbf{Y}$	$\mathbf{R}_\mathrm{x}$,$\mathbf{R}_\mathrm{y}$,$\mathbf{o}_\mathrm{x}$,$\mathbf{o}_\mathrm{y}$,$\mathbf{S}$
Photometric Stereo	Γ, constant $\mathbf{X}$ and $\mathbf{Y}$	$\mathbf{L}$,$\mathbf{N}$
Multi-view Stereo	$\mathbf{R}_\mathrm{x}$,$\mathbf{R}_\mathrm{y}$,$\mathbf{o}_\mathrm{x}$,$\mathbf{o}_\mathrm{y}$,$\Gamma = \mathbf{1}$	$\mathbf{S}$

Table 1. Structure from Motion, Photometric Stereo, and Multi-view Stereo are special cases of Eq. (11).

image plane. Let $[\mathbf{X}]_{t,p} = x_{t,p}$ and $[\mathbf{Y}]_{t,p} = y_{t,p}$[1]. Tomasi and Kanade [20] showed that $\mathbf{X}$ and $\mathbf{Y}$ lie in a three dimensional affine subspace because

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_\mathrm{x} \\ \mathbf{R}_\mathrm{y} \end{bmatrix} \mathbf{S} + \begin{bmatrix} \mathbf{O}_\mathrm{x} \\ \mathbf{O}_\mathrm{y} \end{bmatrix} \quad (7)$$

where $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_\mathcal{P}]$,

$$\mathbf{R}_\mathrm{x} = [\mathbf{r}_{\mathrm{x}1}, \mathbf{r}_{\mathrm{x}2}, \cdots, \mathbf{r}_{\mathrm{x}\mathcal{F}}]^\mathrm{T}, \mathbf{o}_\mathrm{x} = [o_{\mathrm{x}1}, o_{\mathrm{x}2}, \cdots, o_{\mathrm{x}\mathcal{F}}]^\mathrm{T},$$

$$\mathbf{R}_\mathrm{y} = [\mathbf{r}_{\mathrm{y}1}, \mathbf{r}_{\mathrm{y}2}, \cdots, \mathbf{r}_{\mathrm{y}\mathcal{F}}]^\mathrm{T}, \mathbf{o}_\mathrm{y} = [o_{\mathrm{y}1}, o_{\mathrm{y}2}, \cdots, o_{\mathrm{y}\mathcal{F}}]^\mathrm{T},$$

$$\mathbf{O}_\mathrm{x} = [\underbrace{\mathbf{o}_\mathrm{x}, \mathbf{o}_\mathrm{x}, \cdots, \mathbf{o}_\mathrm{x}}_{\mathcal{P}\,\mathtt{columns}}], \mathbf{O}_\mathrm{y} = [\underbrace{\mathbf{o}_\mathrm{y}, \mathbf{o}_\mathrm{y}, \cdots, \mathbf{o}_\mathrm{y}}_{\mathcal{P}\,\mathtt{columns}}].$$

$\begin{bmatrix} \mathbf{R}_\mathrm{x} \\ \mathbf{R}_\mathrm{y} \end{bmatrix}$ and $\begin{bmatrix} \mathbf{o}_\mathrm{x} \\ \mathbf{o}_\mathrm{y} \end{bmatrix}$ form an affine basis for $\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}$.

Photometric constraint on point intensity. We now describe constraints on the intensity variation of scene points. The intensity of scene point $\mathbf{s}_p$ in frame t is given by

$$I_t(\mathbf{x}_{t,p}) = \alpha_p \cdot (l_{\mathtt{a}t} + \mathbf{l}_{\mathtt{d}t}{}^\mathrm{T}\mathbf{n}_p) \quad (8)$$

where α_p and $\mathbf{n}_p$ are the surface albedo and normal vector at $\mathbf{s}_p$, and $l_{\mathtt{a}t}$ and $\mathbf{l}_{\mathtt{d}t}$ are the ambient light and directional light for frame t respectively.[2] We express $l_{\mathtt{a}t}$, $\mathbf{l}_{\mathtt{d}t}$, and $\mathbf{n}_p$ in the object's coordinate system; since we assume the object is rigid, $\mathbf{n}_p$ is constant over time. From Eq. (8), we have

$$\gamma_{t,p} = \frac{I_t(\mathbf{x}_{t,p})}{I_0(\mathbf{x}_{0,p})} = \frac{l_{\mathtt{a}t} + \mathbf{l}_{\mathtt{d}t}{}^\mathrm{T}\mathbf{n}_p}{l_{\mathtt{a}0} + \mathbf{l}_{\mathtt{d}0}{}^\mathrm{T}\mathbf{n}_p} \quad (9)$$

which is dependent on light variation and surface normal but *independent* of surface albedo.

By stacking all $\{\gamma_{t,p}\}$ into an $\mathcal{F}$ by $\mathcal{P}$ matrix Γ with $[\Gamma]_{t,p} = \gamma_{t,p}$, we can factorize Γ as follows

$$\Gamma = \mathbf{LN} \quad (10)$$

where $\mathbf{L} = \begin{bmatrix} l_{\mathtt{a}1}, \cdots, l_{\mathtt{a}\mathcal{F}} \\ \mathbf{l}_{\mathtt{d}1}, \cdots, \mathbf{l}_{\mathtt{d}\mathcal{F}} \end{bmatrix}^\mathrm{T}, \mathbf{N} = \begin{bmatrix} \frac{1}{\beta_1}, \cdots, \frac{1}{\beta_\mathcal{P}} \\ \frac{\mathbf{n}_1}{\beta_1}, \cdots, \frac{\mathbf{n}_\mathcal{P}}{\beta_\mathcal{P}} \end{bmatrix}$, and $\beta_p = l_{\mathtt{a}0} + \mathbf{l}_{\mathtt{d}0}{}^\mathrm{T}\mathbf{n}_p$ is the irradiance at $\mathbf{s}_p$ in the reference frame. Therefore, Γ is spanned by a 4 dimensional linear space and $\mathbf{L}$ is the basis of the subspace.

[1] $[\mathbf{A}]_{i,j} = a_{i,j}$ means "the element of matrix $\mathbf{A}$ at the i'th row and j'th column is $a_{i,j}$"

[2] Basri and Jacobs [1] prove that the right hand size of Eq. (8) is the first-order approximation of the radiance from any Lambertian object under general distant light distribution, where $l_{\mathtt{a}t}$ and $\mathbf{l}_{\mathtt{d}t}$ are interpreted as the mean and the dominant direction of the light distribution respectively.

Subspace-constrained optical flow. We can now formulate multi-point multi-frame optical flow estimation under rigid motion with lighting variation as a subspace-constrained minimization problem:

$$\boxed{\begin{array}{c} \min \Phi(\mathbf{X}, \mathbf{Y}, \Gamma) \\ \mathtt{such\ that} \\ \mathbf{X} = \mathbf{R}_\mathrm{x}\mathbf{S} + \mathbf{O}_\mathrm{x}, \mathbf{Y} = \mathbf{R}_\mathrm{y}\mathbf{S} + \mathbf{O}_\mathrm{y}, \Gamma = \mathbf{LN}. \end{array}} \quad (11)$$

The key observation is that surface positions, normals, motion, and illumination are all coupled together into the same minimization problem. In particular, surface positions and normals are two complementary shape descriptions; the former is constrained by optical flow trajectories and the latter is constrained by intensity variation along these trajectories. By applying subspace constraints to both variables, we are able to densely reconstruct rigidly moving shapes.

As shown in Table 1, our formulation of Eq. (11) subsumes as special cases several traditional vision problems: structure from motion (SFM), photometric stereo (PhS), and multi-view stereo (MVS), which all correspond to assuming some parameters are known and allowing others to vary. In Section 4, we analyze the benefit of solving for all of the parameters together by deriving their estimation uncertainties within our subspace-constrained minimization framework. We begin by introducing a more robust form of the local objective function in Eq. (5) using windows of pixels.

3.1 Window-based flow

The pixel-based local objective function ϕ in Eq. (5) is not robust in practice due to sensor noise, sampling, and quantization. We can define a more robust objective over a small window W_p around $\mathbf{x}_{0,p}$ in the reference frame, over which both flow and surface normal are assumed to be constant. Recall in Eq. (9) that $\gamma_{t,p}$ depends only on lighting and normal, both of which are constant over the window; therefore, $\gamma_{t,p}$ is also constant over the window. The window-based local objective function is then defined as

$$\phi_W(\mathbf{x}_{t,p}, \gamma_{t,p}) = \sum_{\xi \in W_p} (I_t(\mathbf{x}_{t,p} + \xi) - \gamma_{t,p} I_0(\mathbf{x}_{0,p} + \xi))^2. \quad (12)$$

Linearizing the intensity functions in Eq. (12) and minimizing it yields a *generalized Lucas-Kanade* equation:

$$\mathbf{M}_{t,p} \begin{bmatrix} \mathbf{u}_{t,p} \\ \gamma_{t,p} \end{bmatrix} = \mathbf{d}_{t,p} \quad (13)$$

where $\mathbf{M}_{t,p} = \sum_{\xi \in W_p} \left[\begin{array}{cc} \nabla I_t \nabla I_t^{\mathrm{T}} & -I_0 \nabla I_t \\ -I_0 \nabla I_t & I_0^2 \end{array} \right]$

and $\mathbf{d}_{t,p} = \sum_{\xi \in W_p} \left[\begin{array}{c} -I_t \nabla I_t \\ I_0 I_t \end{array} \right]$.

The solution for $[\mathbf{u}^{\mathrm{T}}, \gamma]^{\mathrm{T}}$ is obtained when $\mathbf{M}$ is non-singular. However, $\mathbf{M}$ will be close to singular for any pixel that is not a corner, i.e., for *most pixels*. Consequently, Eq. (13) must be solved with global flow and intensity constraints.

In practice, we achieve better results by defining the local objective function based on an affine motion model within windows around each pixel [18] and generalizing the subspace constraints accordingly. To simplify notation, we use the translational model in the body of this paper, and derive the affine model, used in our implementation, in the appendix.

4. Uncertainties for shape, motion, and light

The subspace-constrained minimization formulation of Eq. (11) involves several sets of unknowns: surface positions, normals, lighting, and motion. In this section, we analyze the uncertainties of these unknowns, revealing the benefits of estimating all the unknowns together instead of treating them in isolation as in previous work.

In particular, we analyze the uncertainties for two subproblems. In the first, we assume known poses and illuminations and estimate surface positions and normals. This case corresponds to the stereo matching problem when the illumination changes from frame to frame. For the second subproblem, we assume known surface positions and normals and estimate poses and illuminations, which corresponds to a camera and lighting calibration problem. In each subproblem, we analyze the uncertainties by deriving the Gauss-Newton approximation of its Hessian matrix with respect to the unknowns.

4.1 Stereo matching with changes in lighting

Traditional stereo matching techniques assume static lighting across views; we now generalize stereo matching to incorporate lighting changes. Formally, given the affine basis $\left[\begin{array}{c} \mathbf{R}_{\mathrm{x}} \\ \mathbf{R}_{\mathrm{y}} \end{array} \right]$ and $\left[\begin{array}{c} \mathbf{o}_{\mathrm{x}} \\ \mathbf{o}_{\mathrm{y}} \end{array} \right]$ for $\left[\begin{array}{c} \mathbf{X} \\ \mathbf{Y} \end{array} \right]$ and the linear basis $\mathbf{L}$ for Γ, we wish to compute $\mathbf{S}$ and $\mathbf{N}$ such that Eq. (11) is minimized.

We first rewrite the generalized Lucas-Kanade equation, Eq. (13), in terms of unknown flow positions $\mathbf{x}_{t,p}$ and brightness scales $\gamma_{t,p}$:

$$\mathbf{M}_{t,p} \left[\begin{array}{c} \mathbf{x}_{t,p} \\ \gamma_{t,p} \end{array} \right] = \mathbf{d}'_{t,p} \tag{14}$$

where $\mathbf{d}'_{t,p} = \mathbf{d}_{t,p} + \mathbf{M}_{t,p}[\mathbf{x}_{0,p}{}^{\mathrm{T}}, 0]^{\mathrm{T}}$.

We then substitute into Eq. (14) the camera pose constraint, Eq. (6), and lighting constraint, Eq. (9),

$$\mathbf{M}_{t,p} \left[\begin{array}{c} \mathbf{r}_{\mathrm{x}t}^{\mathrm{T}} \mathbf{s}_p + o_{\mathrm{x}t} \\ \mathbf{r}_{\mathrm{y}t}^{\mathrm{T}} \mathbf{s}_p + o_{\mathrm{y}t} \\ (l_{\mathrm{a}t} + \mathbf{l}_{\mathrm{d}t}^{\mathrm{T}} \mathbf{n}_p)/\beta_p \end{array} \right] = \mathbf{d}'_{t,p}. \tag{15}$$

Let $\mathbf{l}_t = \left[\begin{array}{c} l_{\mathrm{a}t} \\ \mathbf{l}_{\mathrm{d}t} \end{array} \right]$, $\bar{\mathbf{n}}_p = \left[\begin{array}{c} \frac{1}{\beta_p} \\ \frac{\mathbf{n}_p}{\beta_p} \end{array} \right]$, and $\mathbf{J}_t = \left[\begin{array}{cc} \mathbf{r}_{\mathrm{x}t}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{r}_{\mathrm{y}t}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{l}_t^{\mathrm{T}} \end{array} \right]$; Eq. (15) then becomes

$$\mathbf{M}_{t,p} \mathbf{J}_t \left[\begin{array}{c} \mathbf{s}_p \\ \bar{\mathbf{n}}_p \end{array} \right] = \mathbf{d}''_{t,p} \tag{16}$$

where $\mathbf{d}''_{t,p} = \mathbf{d}'_{t,p} - \mathbf{M}_{t,p}[o_{\mathrm{x}t}, o_{\mathrm{y}t}, 0]^{\mathrm{T}}$.

We finally multiply $\mathbf{J}_t^{\mathrm{T}}$ on both sides of Eq. (16), sum the resulting equations for all frames, and obtain

$$\boxed{\mathbf{Q}_p \left[\begin{array}{c} \mathbf{s}_p \\ \bar{\mathbf{n}}_p \end{array} \right] = \mathbf{w}_p} \tag{17}$$

where $\mathbf{Q}_p = \sum_{t=1}^{\mathcal{F}} \mathbf{J}_t{}^{\mathrm{T}} \mathbf{M}_{t,p} \mathbf{J}_t$ and $\mathbf{w}_p = \sum_{t=1}^{\mathcal{F}} \mathbf{J}_t{}^{\mathrm{T}} \mathbf{d}''_{t,p}$.

Eq. (17) allows us to compute the flow trajectory $\mathbf{x}_{t,p}$ and intensity variation $\gamma_{t,p}$ of point p over multiple frames within the lighting and pose subspaces. $\mathbf{Q}_p$ is the approximated Hessian matrix; inverting $\mathbf{Q}_p$ gives $\mathbf{s}_p$ and $\mathbf{n}_p$.

Analysis. Because $\mathbf{Q}_p$ determines the uncertainty of shape and normal estimation, we now analyze its structure more carefully. We first decompose $\mathbf{M}_{t,p}$ into sub-matrices:

$$\mathbf{M}_{t,p} = \left[\begin{array}{cc} \mathbf{A}_{t,p} & \mathbf{b}_{t,p} \\ \mathbf{b}_{t,p}{}^{\mathrm{T}} & c_p \end{array} \right] \tag{18}$$

where we assume $\mathbf{A}_{t,p} = \left[\begin{array}{cc} \lambda_{1t,p} & 0 \\ 0 & \lambda_{2t,p} \end{array} \right]$ is diagonal without loss of generality,[3] and let $\mathbf{b}_{t,p} = [b_{1t,p}, b_{2t,p}]^{\mathrm{T}}$. Then $\mathbf{Q}_p$ can be shown to have the following structure

$$\mathbf{Q}_p = \left[\begin{array}{cc} \mathbf{R}_{\mathrm{x}}{}^{\mathrm{T}} \Lambda_{1p} \mathbf{R}_{\mathrm{x}} + \mathbf{R}_{\mathrm{y}}{}^{\mathrm{T}} \Lambda_{2p} \mathbf{R}_{\mathrm{y}} & (\mathbf{R}_{\mathrm{x}}{}^{\mathrm{T}} \mathbf{B}_{1p} + \mathbf{R}_{\mathrm{y}}{}^{\mathrm{T}} \mathbf{B}_{2p}) \mathbf{L} \\ \mathbf{L}^{\mathrm{T}} (\mathbf{B}_{1p}{}^{\mathrm{T}} \mathbf{R}_{\mathrm{x}} + \mathbf{B}_{2p}{}^{\mathrm{T}} \mathbf{R}_{\mathrm{y}}) & c_p \mathbf{L}^{\mathrm{T}} \mathbf{L} \end{array} \right] \tag{19}$$

where $\Lambda_{1p} = \mathtt{diag}\{\lambda_{1t,p}\}_t$ is an $\mathcal{F}$ by $\mathcal{F}$ diagonal matrix with $[\Lambda_{1p}]_{t,t} = \lambda_{1t,p}$, and similarly $\Lambda_{2p} = \mathtt{diag}\{\lambda_{2t,p}\}_t$, $\mathbf{B}_{1p} = \mathtt{diag}\{b_{1t,p}\}_t$, $\mathbf{B}_{2p} = \mathtt{diag}\{b_{2t,p}\}_t$.

Notice that the top left submatrix $\mathbf{Q}_{\mathbf{s}p} = \mathbf{R}_{\mathrm{x}}{}^{\mathrm{T}} \Lambda_{1p} \mathbf{R}_{\mathrm{x}} + \mathbf{R}_{\mathrm{y}}{}^{\mathrm{T}} \Lambda_{2p} \mathbf{R}_{\mathrm{y}}$ determines the uncertainty of $\mathbf{s}_p$ if $\mathbf{n}_p$ is given [12]. The bottom right submatrix $\mathbf{Q}_{\mathbf{n}p} = c_p \mathbf{L}^{\mathrm{T}} \mathbf{L}$ determines the uncertainty of $\mathbf{n}_p$ if $\mathbf{s}_p$ is given. On one hand, if the object has enough motion relative to camera, i.e., $\mathbf{R}_{\mathrm{x}}$

[3] In general, $\mathbf{A}_{t,p} = U_{t,p} \cdot \mathtt{diag}\{\lambda_{1t,p}, \lambda_{2t,p}\} \cdot U_{t,p}{}^{\mathrm{T}}$. Defining $[\mathbf{r}_{\mathrm{x}t}', \mathbf{r}_{\mathrm{y}t}'] = [\mathbf{r}_{\mathrm{x}t}, \mathbf{r}_{\mathrm{y}t}] \cdot U_{t,p}$ makes Eq. (19) still valid.

or $\mathbf{R}_y$ is rank 3, $\mathbf{s}_p$ can be recovered if Λ_{1p} or Λ_{2p} is non-zero. As a result, imposing the subspace constraint on optical flow alleviates the *aperture* problem when only one of Λ_{1p} and Λ_{2p} is non-zero. However, low-contrast regions where both Λ_{1p} and Λ_{2p} are nearly zero are still problematic. On the other hand, if the object has enough motion relative to the light, i.e., $\mathbf{L}$ is full rank[4], $\mathbf{n}_p$ can be recovered if $c_p > 0$. Recall that c_p is simply the sum of squared intensity in the window around $\mathbf{x}_{0,p}$ at reference frame 0. Therefore, the surface normal can always be estimated as long as the surface albedo is non-zero. In summary, assuming the scene motion is non-degenerate, we have the following:

- in regions with significant texture, $\mathbf{s}_p$ is computable
- even in texture-less regions, $\mathbf{n}_p$ is computable

These two sources of shape information are thus complementary and can be used together to reconstruct surfaces in both textured and textureless regions.

We should emphasize that in low contrast regions, the surface normals can be accurately estimated in the presence of optical flow errors because small offsets in flow trajectories do not cause large changes in intensity variations along these trajectories. Traditional shape-from-flow methods, e.g., [13], regularize flow and thus often over-smooth the reconstructed shape. Here we argue that optical flow does not have to be strongly regularized in low contrast regions; they can be computed through reconstructed shape integrated from surface normals. We will present an algorithm in Section 5 to combine both flow trajectories and shading variation along these trajectories for shape reconstruction.

4.2 Camera and light calibration

We now consider the subproblem of estimating camera motion $\mathbf{R}_x$, $\mathbf{R}_y$, $\mathbf{o}_x$, $\mathbf{o}_y$ and light $\mathbf{L}$ given the surface positions $\mathbf{S}$ and normals $\mathbf{N}$. Similarly to Section 4.1, we can derive the approximated Hessian matrix $\mathbf{P}_t$ for computing the camera motion and light as:

$$\boxed{\mathbf{P}_t \begin{bmatrix} \mathbf{r}_{\mathrm{x}t} \\ o_{\mathrm{x}t} \\ \mathbf{r}_{\mathrm{y}t} \\ o_{\mathrm{y}t} \\ \mathbf{l}_t \end{bmatrix} = \mathbf{v}_t} \tag{20}$$

where $\mathbf{P}_t = \sum_{p=1}^{\mathcal{P}} \mathbf{K}_p{}^{\mathrm{T}} \mathbf{M}_{t,p} \mathbf{K}_p \quad , \quad \mathbf{v}_t = \sum_{p=1}^{\mathcal{P}} \mathbf{K}_p{}^{\mathrm{T}} \mathbf{d}'_{t,p},$

$$\mathbf{K}_p = \begin{bmatrix} \bar{\mathbf{s}}_p^{\mathrm{T}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{s}}_p^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \bar{\mathbf{n}}_p^{\mathrm{T}} \end{bmatrix}, \quad \bar{\mathbf{s}}_p = \begin{bmatrix} \mathbf{s}_p \\ 1 \end{bmatrix}.$$

[4] Actually, the normal can also be estimated when the ambient term in $\mathbf{L}$ is zero, in which case the rank of $\mathbf{L}$ is only 3.

Under the same assumption that $\mathbf{A}_{t,p}$ is diagonal, $\mathbf{P}_t$ can be shown to have the following structure

$$\mathbf{P}_t = \begin{bmatrix} \bar{\mathbf{S}}\Lambda_{1t}\bar{\mathbf{S}}^{\mathrm{T}} & \mathbf{0} & \bar{\mathbf{S}}\mathbf{B}_{1t}\mathbf{N}^{\mathrm{T}} \\ \mathbf{0} & \bar{\mathbf{S}}\Lambda_{2t}\bar{\mathbf{S}}^{\mathrm{T}} & \bar{\mathbf{S}}\mathbf{B}_{2t}\mathbf{N}^{\mathrm{T}} \\ \mathbf{N}\mathbf{B}_{1t}{}^{\mathrm{T}}\bar{\mathbf{S}}^{\mathrm{T}} & \mathbf{N}\mathbf{B}_{2t}{}^{\mathrm{T}}\bar{\mathbf{S}}^{\mathrm{T}} & \mathbf{N}\mathbf{C}\mathbf{N}^{\mathrm{T}} \end{bmatrix} \tag{21}$$

where $\Lambda_{1t} = \mathtt{diag}\{\lambda_{1t,p}\}_p$ is a $\mathcal{P}$ by $\mathcal{P}$ diagonal matrix with $[\Lambda_{1t}]_{p,p} = \lambda_{1t,p}$, and similarly $\Lambda_{2t} = \mathtt{diag}\{\lambda_{2p,p}\}_p$, $\mathbf{B}_{1t} = \mathtt{diag}\{b_{1t,p}\}_p$, $\mathbf{B}_{2t} = \mathtt{diag}\{b_{2t,p}\}_p$, $\mathbf{C} = \mathtt{diag}\{c_p\}$, and $\bar{\mathbf{S}} = [\bar{\mathbf{s}}_1, \bar{\mathbf{s}}_2, \cdots, \bar{\mathbf{s}}_{\mathcal{P}}]$.

The top left sub-matrix $\mathbf{P}_{\mathtt{m}t} = \begin{bmatrix} \bar{\mathbf{S}}\Lambda_{1t}\bar{\mathbf{S}}^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{S}}\Lambda_{2t}\bar{\mathbf{S}}^{\mathrm{T}} \end{bmatrix}$ determines the uncertainty of camera motion estimation for frame t and is dominated by feature points that have large λ_1 and λ_2.

The bottom right sub-matrix $\mathbf{P}_{\mathtt{l}t} = \mathbf{N}\mathbf{C}\mathbf{N}^{\mathrm{T}}$ determines the uncertainty of light for t and is determined by non-black regions in the images. As more points are used to estimate the light, $\mathbf{N}$ tends to contain more normal variation, and the lighting estimation becomes more certain.

5. Reconstruction algorithm

In this section, we present an iterative algorithm to solve Eq. (11). We begin by computing camera motion and initializing lighting with structure from motion on sparse features. Then, we iterate between solving for the shape and solving for the lighting while fixing other unknowns.

5.1 Solve for $\mathbf{R}_x$, $\mathbf{R}_y$, $\mathbf{o}_x$, $\mathbf{o}_y$, and initialize $\mathbf{L}$

To estimate camera motion, we track feature points using our translation-based generalized Lucas-Kanade equation, Eq. (13), and then apply Tomasi-Kanade factorization to recover $\mathbf{R}_x$, $\mathbf{R}_y$, $\mathbf{o}_x$, $\mathbf{o}_y$. Currently, we select a small number ($\mathcal{M}$) of feature points manually, though automatic methods could also be used [18]. To estimate lighting, we upgrade motion model from translation to affine in feature tracking. In the appendix, we show that the affine motion parameters are also subject to the subspace constraints of camera motion.[5] Affine tracking under these constraints amounts to estimating surface tangents $\frac{\partial \mathbf{s}}{\partial x}$ and $\frac{\partial \mathbf{s}}{\partial y}$ at the feature points. Finally, we compute feature normals from the surface tangents, and estimate the lighting $\mathbf{L}$ using the method to be described in Section 5.3.

5.2 Solve for S and $\{\mathbf{n}_p\}$

Next, we compute the position and normal at each pixel in the reference frame. We begin by solving for $\{\mathbf{s}_p\}$ and $\{\bar{\mathbf{n}}_p\}$ using Eq. (17) subject to the following linear constraint

$$\begin{bmatrix} \mathbf{r}_{\mathrm{x}0}{}^{\mathrm{T}}\mathbf{s}_p + o_{\mathrm{x}0} \\ \mathbf{r}_{\mathrm{y}0}{}^{\mathrm{T}}\mathbf{s}_p + o_{\mathrm{y}0} \end{bmatrix} = \begin{bmatrix} x_{0,p} \\ y_{0,p} \end{bmatrix} \tag{22}$$

[5] We could have used unconstrained affine tracking from the start, but we found that the added degrees of freedom made the tracking less robust.

which forces $\mathbf{s}_p$ to lie along the line of sight through $\mathbf{x}_{0,p}$.[6] As discussed in Section 4, we can expect the normal information to be reasonably good over most pixels, but reconstructed positions will generally be unreliable in textureless regions. Thus, our shape reconstruction relies primarily on normals. Given $\{\bar{\mathbf{n}}_p\}$ for every point, we integrate a depth map $\tilde{z}(x, y)$ by minimizing

$$\Psi(\tilde{z}) = \sum_{x,y} \left(\frac{\partial \tilde{z}(x,y)}{\partial x} + \frac{n_x}{n_z} \right)^2 + \left(\frac{\partial \tilde{z}(x,y)}{\partial y} + \frac{n_y}{n_z} \right)^2 \tag{23}$$

using the conjugate gradient method. In our iterative framework, we improve convergence by initializing the conjugate gradient solver with the depth map from the last iteration.

The depth map $\tilde{z}(x, y)$ obtained from normal integration will not in general correspond to the "true" depth map if the lighting is not accurate. In particular, erroneous lighting gives rise to global distortion of the estimated surface normals and thus global distortion of the reconstructed depth map. This distortion is evident when the surface does not pass through the 3D positions of tracked feature points. To bring the surface closer to these points, we apply a global affine transformation to the depth map:

$$z(x, y) = \mu x + \nu y + \zeta \tilde{z}(x, y) + \eta. \tag{24}$$

For each of the $\mathcal{M}$ feature points $\mathbf{s}_m$, we have both a depth z_m directly computed from Eq. (17), as well as a depth $\tilde{z}_m$ from normal integration in Eq. (23). Thus, using Eq. (24), we can set up a system of $\mathcal{M}$ linear equations and solve for the affine parameters. We then use these parameters to correct the depth map of the reconstructed surface. As shown by Belhumeur et al. [2], we can also use the same parameters to correct normals.

5.3 Solve for $\mathbf{L}$ and $\{\beta_p\}$

After surface positions and normals are computed, we estimate lighting $\mathbf{L}$ and irradiance parameters $\{\beta_p\}$. The index p in this section refers to either sparse feature points or dense flow points. Recall that $\gamma_{t,p} = (l_{\mathtt{a}t} + \mathbf{l}_{\mathtt{d}t}{}^{\mathtt{T}}\mathbf{n}_p)/\beta_p$, which may be rewritten as

$$l_{\mathtt{a}t} + \mathbf{n}_p{}^{\mathtt{T}}\mathbf{l}_{\mathtt{d}t} - \gamma_{t,p}\beta_p = 0. \tag{25}$$

For dense flow, we have $\mathcal{P} \cdot \mathcal{F}$ equations and $4\mathcal{F}$ unknowns for lighting $\{\mathbf{l}_t\}$ and $\mathcal{P}$ unknowns for $\{\beta_p\}$[7]. Recalling the definition of β_p, we have a set of constraints for Eq. (25) in the reference frame:

$$l_{\mathtt{a}0} + \mathbf{n}_p{}^{\mathtt{T}}\mathbf{l}_{\mathtt{d}0} - \beta_p = 0. \tag{26}$$

A least squares solution to Eq. (25) constrained by Eq. (26) is computed using a variant of constrained least squares [5] for homogeneous equations.

In the case that there is no relative motion between the camera and light, the relations $\mathbf{l}_{\mathtt{d}t} = [\mathbf{r}_{\mathtt{x}t}, \mathbf{r}_{\mathtt{y}t}, \mathbf{r}_{\mathtt{z}t}] \cdot \mathbf{l}_{\mathtt{d}0}$ and $l_{\mathtt{a}t} = l_{\mathtt{a}0}$ further constrain the problem and make the solution more robust.

5.4 Implementation

After estimating camera motion and initializing lighting, we solve for shape and lighting in a coarse-to-fine manner using an image pyramid. At each resolution, we iterate twice between the steps described in Section 5.2 and 5.3. In principle, we could also update camera motion in this iterative framework. However, our analysis of Eq. (21) indicates that low contrast points do not improve pose estimation much, and the Tomasi-Kanade factorization already initializes camera motion using a good set of features.

6. Results

Our experimental configuration consists of a single light source and a Basler A301f video camera. We recorded image sequences of handheld objects rotating in front of a fixed camera under static lighting. Figure 1 shows the sample inputs and reconstruction result. If we just solve Eq. (17) for the surface position $\{\mathbf{s}_p\}$, we get a noisy reconstruction (Figure 1e) due to ambiguities in textureless regions. When integrating normals derived from that same equation, we are able to reconstruct a good facsimile of the original shape, as shown by the coarse-to-fine progression (Figure 1f-g). Figure 1c and d show side view renderings, the latter with estimated surface albedo. Figure 2 is an example of a shape containing large planar untextured regions, which confound optical flow and stereo reconstruction algorithms, even those designed to handle brightness changes. Since our method correctly estimates normals without texture, we obtain an accurate reconstruction.

7. Conclusions and future work

We have presented a technique for computing optical flow, shape, motion, lighting, and albedo from a monocular image sequence. The approach combines both geometric (optical flow) and photometric (intensity change) cues to compute dense shape that is accurate even in completely uniform untextured regions.

In order to accomplish our goals, we made a number of assumptions and approximations. For example, our approach is not robust to occlusions, shadows, inter-reflections, or specularity. Further, in Section 5.2, surface positions and normals are computed for each point individually without enforcing their mutual consistency. One direction of future work is to robustly optimize with respect to all unknowns, i.e., solve for a surface whose positions and normals simultaneously satisfy

[6] We do not enforce the quadratic constraint that the L^2 norm of the last three elements of $\bar{\mathbf{n}}_p$ should equal the square of the first element.

[7] Replace $\mathcal{P}$ with $\mathcal{M}$ for the sparse feature case.

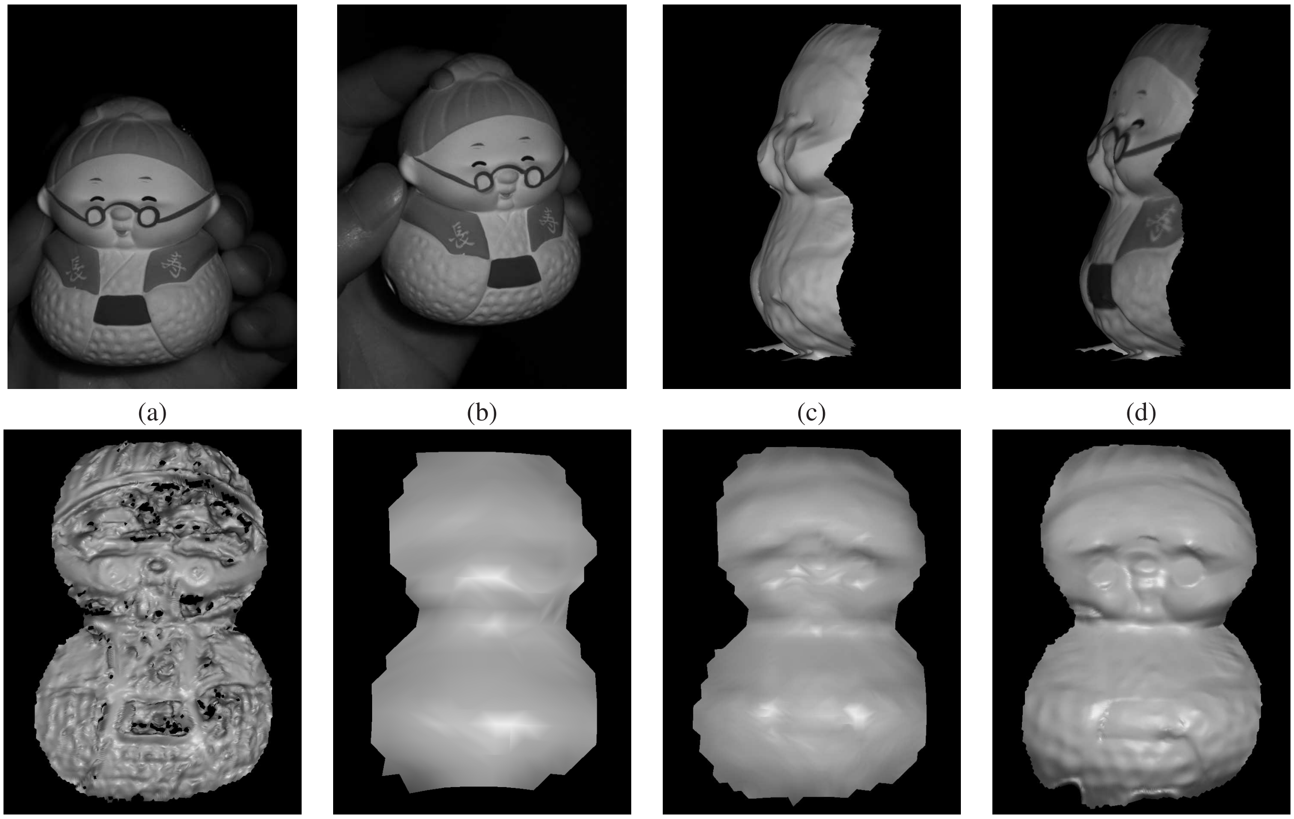
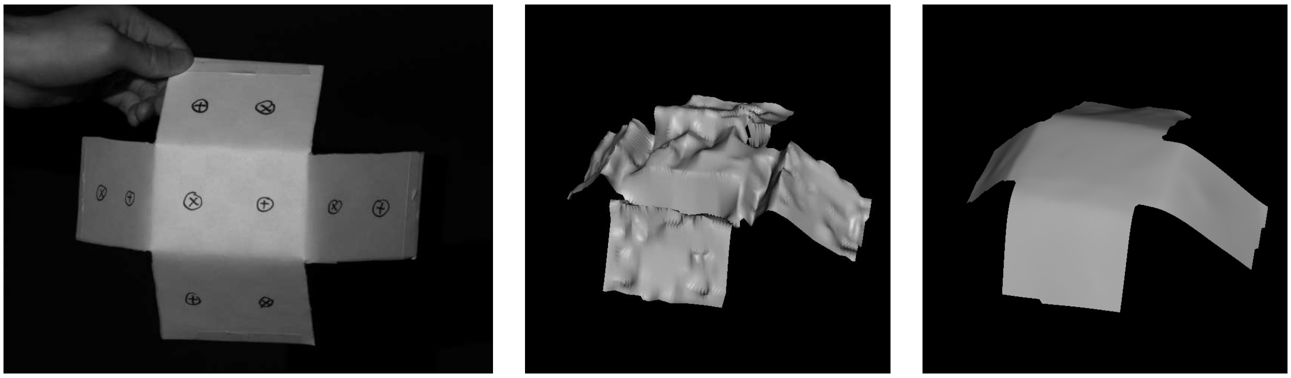


Figure 1. Reconstruction of a figurine. (a) The reference image. (b) Another sample image from a 236 frame sequence. (c) Profile view of the reconstruction. (d) The same view with recovered albedo-map. (e) Shape obtained by solving Eq. (17) without normal integration. (f)-(h) Coarse-to-fine reconstructions using normal integration.



(a) (b) (c)

Figure 2. (a) is an input frame from a 130 frame sequence. (b) is a surface reconstruction by solving Eq. (17) directly instead of by normal integration, (c) is the rendering of the final surface reconstructed with normal integration.

both flow and shading variation constraints. It may also be possible to extend our approach to handle non-rigidly moving scenes, by incorporating recent work on morphable shape bases, e.g., [21].

Appendix

In this appendix, we present the subspace-constrained optical flow with a local objective function $\tilde{\phi}_W$ based on an affine motion model, defined as

$$\tilde{\phi}_W(\mathbf{x}_{t,p}, \gamma_{t,p}) = \sum_{\xi}^{W_p} (I_t(\mathbf{x}_{t,p} + \mathbf{D}_{t,p} \cdot \xi) - \gamma_{t,p} I_0(\mathbf{x}_{0,p} + \xi))^2 \tag{27}$$

where $\mathbf{D}_{t,p} = \begin{bmatrix} e_{t,p} & g_{t,p} \\ f_{t,p} & h_{t,p} \end{bmatrix}$ is the first order approximation for the flow around $\mathbf{x}_{t,p}$. Assuming orthographic camera projection, it follows that

$$\begin{bmatrix} e_{t,p} \\ f_{t,p} \end{bmatrix} = \begin{bmatrix} \mathbf{r_x}_t^{\mathrm{T}} \\ \mathbf{r_y}_t^{\mathrm{T}} \end{bmatrix} \frac{\partial \mathbf{s}_p}{\partial x} \quad \begin{bmatrix} g_{t,p} \\ h_{t,p} \end{bmatrix} = \begin{bmatrix} \mathbf{r_x}_t^{\mathrm{T}} \\ \mathbf{r_y}_t^{\mathrm{T}} \end{bmatrix} \frac{\partial \mathbf{s}_p}{\partial y} \quad . \tag{28}$$

Defining $[\mathbf{E}]_{t,p} = e_{t,p}$, $[\mathbf{F}]_{t,p} = f_{t,p}$, $[\mathbf{G}]_{t,p} = g_{t,p}$, $[\mathbf{H}]_{t,p} = h_{t,p}$, we have

$$\begin{bmatrix} \mathbf{E} \\ \mathbf{F} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{\mathrm{x}} \\ \mathbf{R}_{\mathrm{y}} \end{bmatrix} \frac{\partial \mathbf{S}}{\partial x} \quad \begin{bmatrix} \mathbf{G} \\ \mathbf{H} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{\mathrm{x}} \\ \mathbf{R}_{\mathrm{y}} \end{bmatrix} \frac{\partial \mathbf{S}}{\partial y} \tag{29}$$

where $\frac{\partial \mathbf{S}}{\partial x} = [\frac{\partial \mathbf{s}_1}{\partial x}, \cdots, \frac{\partial \mathbf{s}_{\mathcal{P}}}{\partial x}]$ and $\frac{\partial \mathbf{S}}{\partial y} = [\frac{\partial \mathbf{s}_1}{\partial y}, \cdots, \frac{\partial \mathbf{s}_{\mathcal{P}}}{\partial y}]$. Therefore *the window deformation coefficients are also subject to three dimensional subspace constraints*, and the multi-point multi-frame optical flow problem becomes

$$\boxed{\begin{array}{c} \min \Phi(\mathbf{X}, \mathbf{Y}, \mathbf{E}, \mathbf{F}, \mathbf{G}, \mathbf{H}, \Gamma) \\ \texttt{such that} \\ \mathbf{X} = \mathbf{R}_{\mathrm{x}}\mathbf{S} + \mathbf{O}_{\mathrm{x}}, \mathbf{Y} = \mathbf{R}_{\mathrm{y}}\mathbf{S} + \mathbf{O}_{\mathrm{y}}, \Gamma = \mathbf{LN} \\ \mathbf{E} = \mathbf{R}_{\mathrm{x}}\frac{\partial \mathbf{S}}{\partial x}, \mathbf{F} = \mathbf{R}_{\mathrm{y}}\frac{\partial \mathbf{S}}{\partial x}, \mathbf{G} = \mathbf{R}_{\mathrm{x}}\frac{\partial \mathbf{S}}{\partial y}, \mathbf{H} = \mathbf{R}_{\mathrm{y}}\frac{\partial \mathbf{S}}{\partial y}. \end{array}} \tag{30}$$

References

[1] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. *PAMI*, 25(2):218–233, 2003.

[2] P. N. Belhumeur, D. Kriegman, and A. Yuille. The bas-relief ambiguity. *IJCV*, 35(1):33–44, 1999.

[3] M. J. Black and A. D. Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. *IJCV*, 26(1):63–84, 1998.

[4] P. Fua and Y. G. Leclerc. Object-centered surface reconstruction: Combining multi-image stereo and shading. *IJCV*, 16:35–56, 1995.

[5] Golub, G. H. and Van Loan, C. F. *Matrix Computations*. Johns Hopkins University Press, Baltimore, 3rd edition, 1996.

[6] W. E. L. Grimson. Binocular shading and visual surface reconstruction. *Computer Vision, Graphics, and Image Processing*, 28:19–43, 1984.

[7] G. D. Hager and P. N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *PAMI*, 20(10):1025–1039, 1998.

[8] H. W. Haussecker and D. J. Fleet. Computing optical flow with physical models of brightness variation. *PAMI*, 23(6):661–673, 2001.

[9] M. Irani. Multi-frame optical flow estimation using subspace constraints. In *ICCV*, pages 626–633, 1999.

[10] H. Jin, P. Favaro, and S. Soatto. Real-time feature tracking and outlier rejection with changes in illumination. In *ICCV*, pages 684–689, 2001.

[11] A. Maki, M. Watanabe, and C. Wiles. Geotensity: Combining motion and lighting for 3D surface reconstruction. *IJCV*, 48(2):75–90, 2002.

[12] D. D. Morris and T. Kanade. A unified factorization algorithm for points, line segments and planes with uncertain models. In *ICCV*, pages 696–702, 1998.

[13] N. Mukawa. Estimation of shape, reflection coefficients, and illumination direction from image sequences. In *ICCV*, pages 507–512, 1990.

[14] S. Negahdaripour. Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis. *PAMI*, 20(9):961–979, 1998.

[15] A. Pentland. Photometric motion. *PAMI*, 13(9):879–890, 1991.

[16] D. Samaras, D. Metaxas, P. Fua, and Y. G. Leclerc. Variable albedo surface reconstruction from stereo and shape from shading. In *CVPR*, pages 480–487, 2000.

[17] S. M. Seitz and K. N. Kutulakos. Plenoptic image editing. *IJCV*, 48(2):115–129, 2002.

[18] J. Shi and C. Tomasi. Good features to track. In *CVPR*, pages 593–600, 1994.

[19] D. Simakov and R. Basri. Dense shape reconstruction of a moving object under arbitrary, unknown lighting. In *ICCV*, page to appear, 2003.

[20] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography–a factorization method. *IJCV*, 9(2):137–154, 1992.

[21] L. Torresani, D. B. Yang, E. J. Alexander, and C. Bregler. Tracking and modeling non-rigid objects with rank constraints. In *CVPR*, pages 493–500, 2001.

[22] R. J. Woodham. Multiple light source optical flow. In *ICCV*, pages 42–46, 1990.