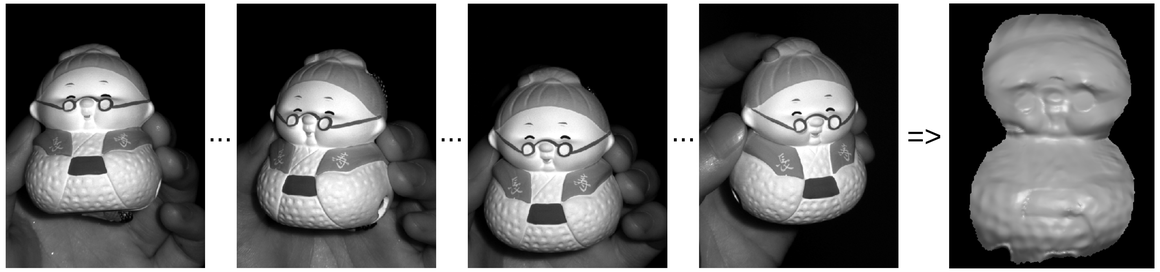
Shape and Motion under Varying Illumination:
Unifying Structure from Motion, Photometric Stereo, and Multi-view Stereo

Li Zhang* **Brian Curless*** **Aaron Hertzmann**** **Steven M. Seitz***
*University of Washington, Seattle, WA, USA
**University of Toronto, Toronto, ON, Canada



Introduction

Goal Dense shape reconstruction of moving objects under varying illumination from a single video.

Example A hand-held figurine rotating in front of a fixed camera under static lighting.

Standard methods

- Structure from Motion —— only feature points
- Multi-view Stereo ———— only constant lighting
- Photometric Stereo ———— only static objects

Our solution

Using both spatial and temporal brightness variations

spatial brightness variation → motion cue → surface position

temporal brightness variation → photometric cue → surface orientation

→ dense surface reconstruction in both textured *and textureless* regions

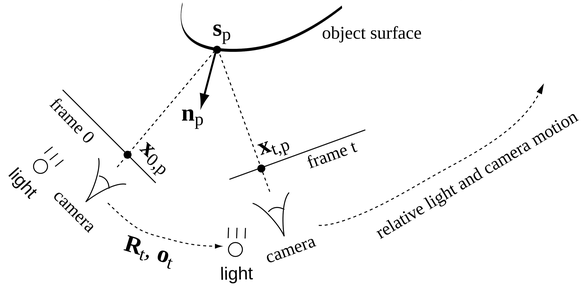
Contributions

- dense structure from motion
- stereo matching under lighting changes
- photometric stereo for moving scenes

Mathematical Formualtion

Optical flow under varying illumination

Assume a Lambertian object moving rigidly in front of an orthographic camera under static distant light.



Let $\mathbf{l}_t$ and b_t be the directional and ambient light at frame t, and $\mathbf{n}_p$ and α_p be the normal and albedo of the point p, then

Image Formation: $I_t(\mathbf{x}_{t,p}) = \alpha_p \cdot (\mathbf{l}_t^T \mathbf{n}_p + b_t)$

Irradiance ratio: $\dfrac{I_t(\mathbf{x}_{t,p})}{I_0(\mathbf{x}_{0,p})} = \dfrac{\mathbf{l}_t^T \mathbf{n}_p + b_t}{\mathbf{l}_0^T \mathbf{n}_p + b_0} = \gamma_{t,p}$

Brightness-varying flow: $I_t(\mathbf{x}_{t,p}) = \gamma_{t,p} \cdot I_0(\mathbf{x}_{0,p})$

Multi-point multi-frame optical flow can be computed by minimizing: $\Phi(\{\mathbf{x}_{t,p}, \gamma_{t,p}\}) = \sum_{t,p} (I_t(\mathbf{x}_{t,p}) - \gamma_{t,p} \cdot I_0(\mathbf{x}_{0,p}))^2$

Constraints on brightness-varying flow

Rank 3 constraint on $\{x_{t,p}, y_{t,p}\}$ [Tomasi&Kanade90]:

$$\mathbf{X} = \begin{bmatrix} x_{1,1} \cdots x_{1,P} \\ \vdots \quad \vdots \quad \vdots \\ x_{T,1} \cdots x_{T,P} \end{bmatrix} = \begin{bmatrix} \vdots \\ \mathbf{r}_{xt}^T \\ \vdots \end{bmatrix} \begin{bmatrix} \cdots \mathbf{s}_p \cdots \end{bmatrix} + \begin{bmatrix} o_{x1} \cdots o_{x1} \\ \vdots \quad \vdots \quad \vdots \\ o_{xT} \cdots o_{xT} \end{bmatrix} = \mathbf{R}_x\mathbf{S} + \mathbf{O}_x$$

$\mathbf{Y} = \mathbf{R}_y\mathbf{S} + \mathbf{O}_y$, similarly.

Rank 4 constraint on $\{\gamma_{t,p}\}$ [Basri&Jacob01]:

$$\Gamma = \begin{bmatrix} \gamma_{1,1} \cdots \gamma_{1,P} \\ \vdots \quad \vdots \quad \vdots \\ \gamma_{T,1} \cdots \gamma_{T,P} \end{bmatrix} = \begin{bmatrix} \vdots \quad \vdots \\ \mathbf{l}_t^T \quad b_t \\ \vdots \quad \vdots \end{bmatrix} \begin{bmatrix} \cdots \dfrac{\mathbf{n}_p}{\beta_p} \cdots \\ \cdots \dfrac{1}{\beta_p} \cdots \end{bmatrix} = \mathbf{L}\tilde{\mathbf{N}}$$

where $\beta_p = \mathbf{l}_0^T \mathbf{n}_p + b_0$.

Consistency constraint on **S** and **N**: $\dfrac{\delta \mathbf{S}}{\delta x} \perp \mathbf{N}$, $\dfrac{\delta \mathbf{S}}{\delta y} \perp \mathbf{N}$.

Multi-point multi-frame brightness-varying optical flow

$$\min \Phi(\mathbf{X}, \mathbf{Y}, \Gamma)$$

$$\text{s.t. } \mathbf{X} = \mathbf{R}_x\mathbf{S}+\mathbf{O}_x,\ \mathbf{Y} = \mathbf{R}_y\mathbf{S}+\mathbf{O}_y,\ \Gamma = \mathbf{L}\mathbf{N},\ \frac{\delta \mathbf{S}}{\delta x} \perp \mathbf{N},\ \frac{\delta \mathbf{S}}{\delta y} \perp \mathbf{N}.$$

Our formulation extends [Irani99] and applies to features, edges, and textureless regions. Our formulation also subsumes Structure from Motion, Multi-view Stereo, and Photometric Stereo as special cases:

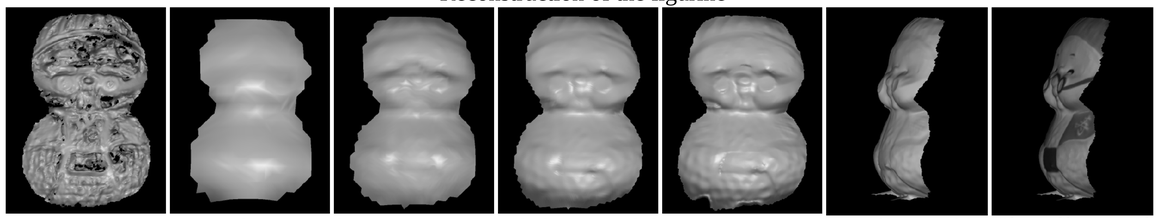
	Known	Unnown
Structure from Motion	**X**, **Y**	$\mathbf{R}_x$, $\mathbf{O}_x$, $\mathbf{R}_y$, $\mathbf{O}_y$, **S**
Multi-view Stereo	$\mathbf{R}_x$, $\mathbf{O}_x$, $\mathbf{R}_y$, $\mathbf{O}_y$, Γ=**1**	**S**
Photometric Stereo	Γ, constant **X**, **Y**	**N**, **L**

Reconstruction algorithm

Initialization Track features, compute $\mathbf{R}_x$, $\mathbf{O}_x$, $\mathbf{R}_y$, $\mathbf{O}_y$, estimate feature normals, and initialize **L**.
Step1 Fix $\mathbf{R}_x$, $\mathbf{O}_x$, $\mathbf{R}_y$, $\mathbf{O}_y$, and **L**, and update **S** and **N**;
Step2 Correct **S** with large uncertainties by integrating **N**;
Step3 Fix $\mathbf{R}_x$, $\mathbf{O}_x$, $\mathbf{R}_y$, $\mathbf{O}_y$, and **S**, and update **L**.
The algorithm is implemented in a coarse-to-fine manner, and two iterations are computed at each level of detail.

Results

Reconstruction of the figurine



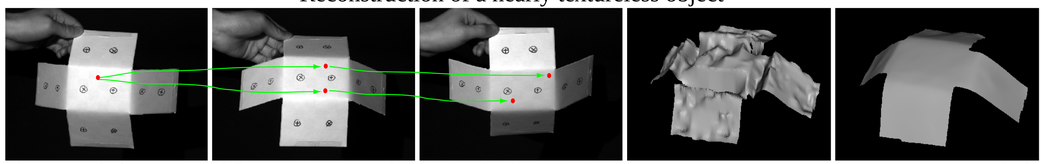
Reconstruction using only motion cue

The coarse-to-fine reconstruction, using both motion and photometric cues

A profile view of the reconstruction

The profile view with recovered albedo map

Reconstruction of a nearly textureless object



Three example frames. Optical flow is ambiguous within constant brightness regions; however, the temporal brightness variations uniquely determine surface orientation.

Reconstruction using only motion cues

Reconstruction using both motion and photometric cues

Conclusion

- Motion cue constrains 3D positions inaccurately for low textured pixels
- Photometric cue reveals normals accurately even for moving scenes, despite the noisy motion estimation
- Combing both cues recovers moving shape densely