Dense Non-Rigid Structure-from-Motion and Shading with Unknown Albedos

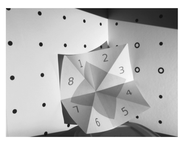
Mathias Gallardo[1], Toby Collins[2] and Adrien Bartoli[1]
[1]EnCoV, Institut Pascal, UMR 6602 CNRS, Université Clermont Auvergne, SIGMA, France [2]IRCAD, Strasbourg, France

Motivations

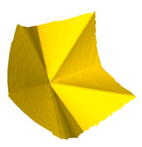
Problem: Non-rigid template-free 3D reconstruction from monocular images with multiple cues
When motion is the only cue used, this is known as Non-rigid Structure-from-Motion (NRSfM)

Current limitations: handling **(i)** poorly-textured surfaces and **(ii)** non-smooth deformations

Reason:
Motion information is insufficient
(feature correspondences)



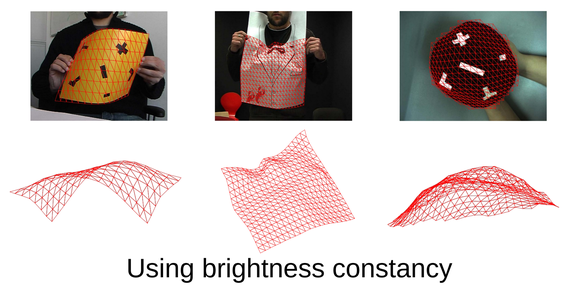
One image of a sequence of 5 images



Ground-truth shape

Best result from current NRSfM method [Parashar *et al.*, 2016]

Closest work:
[Wang *et al.*, 2016] reconstructs **(i)**, but not **(ii)**



Using brightness constancy

Contributions

- Combining shading with motion (NRSfM)
- Estimating simultaneously 3D shapes and spatially-varying surface albedo
- Reconstructing densely poorly-textured surfaces
- Modeling non-smooth deformations
- Working for unorganized and organized sets of images

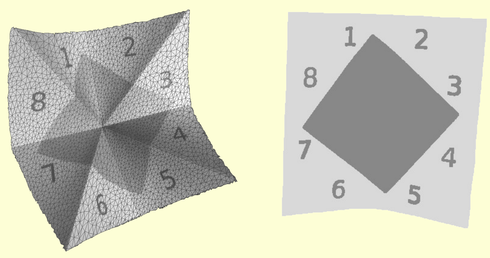
Main challenges:
- High-dimensional deformations spaces ($\mathcal{O}(10^5)$ unknowns)
- Highly non-convex problem
- Dense pixel-level registration (required for shading)

Modeling Assumptions and Inputs

Unknowns

N meshes shapes $\{\mathcal{V}_t\}$

Piecewise constant albedo A



$\mathcal{V}_1$ 1D depth parameterization

Known

Linear camera responses

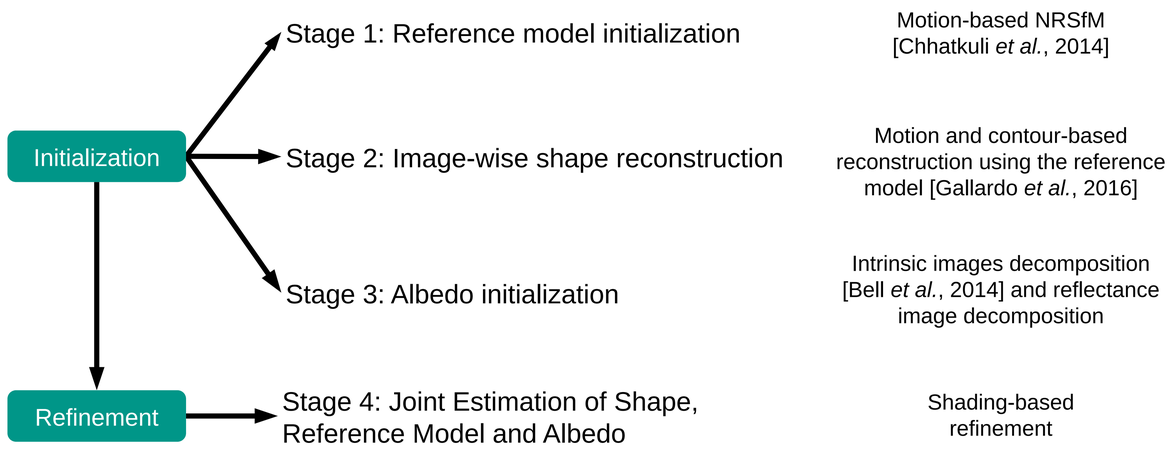
Spherical harmonics light $\mathbf{l}$

Input images $\{\mathcal{I}_t\}$ → Luminance images $\{L_t\}$

Keypoint correspondences between N input images $\{\mathcal{S}_t\}$

Assumptions
- Lambertian reflectance model
- Smooth reference view

Joint Albedo and 3D Shape Estimation Pipeline



Shading-Based Refinement

Energy minimization:

$$C(\mathcal{V}_1,\ldots,\mathcal{V}_N,A) \triangleq \sum_{t=1}^{N} C_{shade}(\mathcal{V}_t,A;L_t,\mathbf{l}) + \lambda_{corresp}C_{corresp}(\mathcal{V}_t;\mathcal{S}_t) + \lambda_{contour}C_{contour}(\mathcal{V}_t;\mathcal{I}_t) + \lambda_{iso}C_{iso}(\mathcal{V}_1,\mathcal{V}_t) + \lambda_{smooth}C_{smooth}(\mathcal{V}_t)$$

- **Improving convergence:** three pyramid levels for shading and contour costs
- **Solving:** Gauss-Newton iterative optimization with back-tracking line-search
- **Scale ambiguity:** at each iteration, fix mean depth to arbitrary positive constant (1.0)
- **Implementation:** several minutes for < 10 images, in Matlab

Evaluation and Experimental Results

Datasets:
- real datasets with sub-millimiter accuracy depth-maps obtained with a structured light system
- datasets of unorganized images (where tracking may not work)

Compared methods: [Torresani *et al.*, 2008], [Varol *et al.*, 2012], [Taylor *et al.*, 2010], [Vicente and Agapito, 2012], [Chhatkuli *et al.*, 2014], [Chhatkuli *et al.*, 2016] and [Parashar *et al.*, 2016]

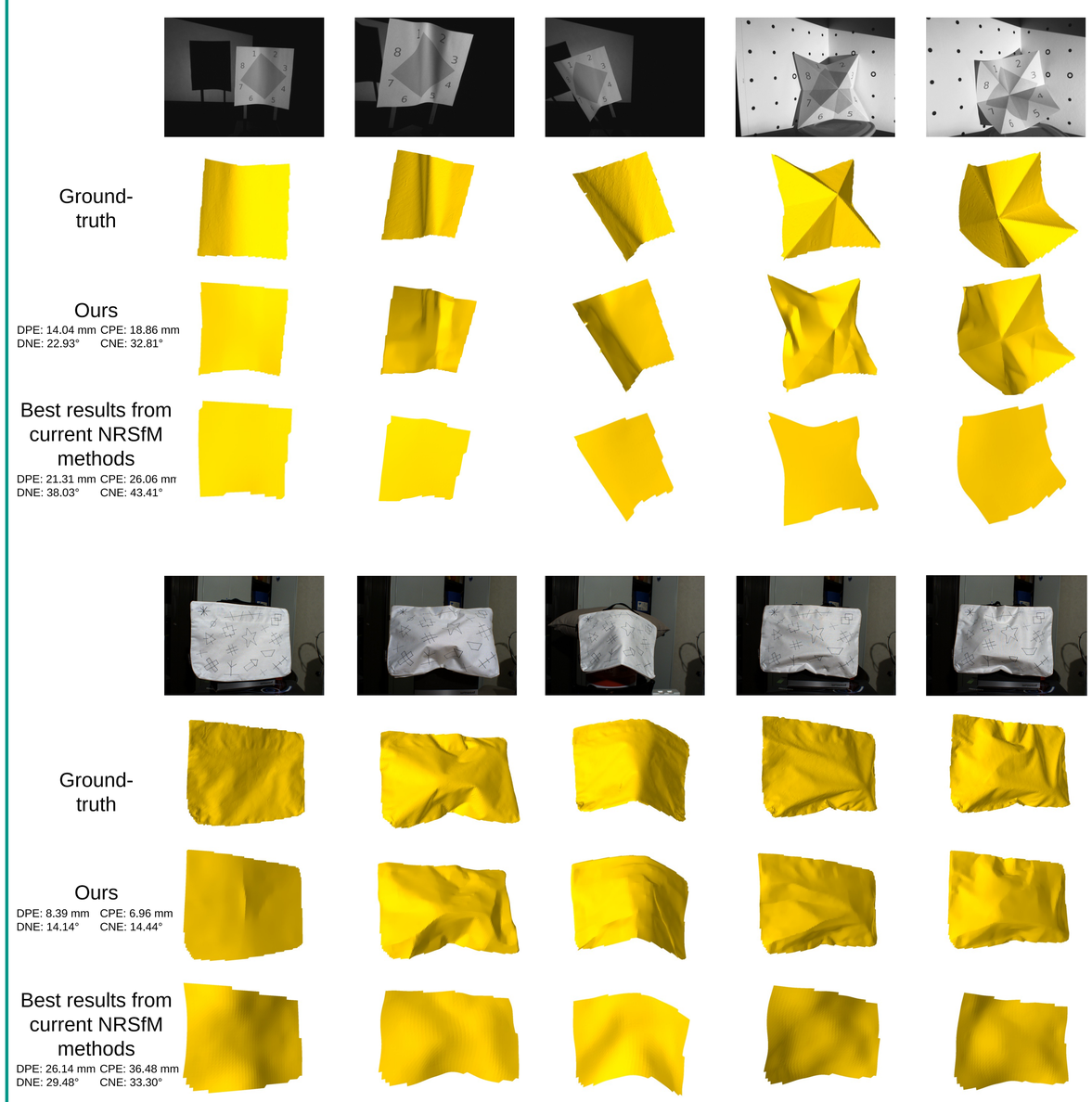
Convergence basin analysis:
- shows that a strong local minimum is close to the ground-truth

Quantitative results:
- our method performs globally better to current NRSfM methods
- shading improves global and creases reconstructions
- boundary helps shading

(more details in the paper)

Qualitative results for sets of 5 input images:

DPE: Dense Position Error (RMSE) in mm
DNE: Dense Normal Error (RMSE) in degrees
CPE: Crease Position Error (RMSE) in mm
CNE: Crease Normal Error (RMSE) in degrees



Conclusion

- Challenging problem (pixel-level registration and dense reconstruction)
- A proof-of-concept for NRSfM and shading with unknown albedos
- Proposed solution: cascaded initialization using motion and boundary information, then full refinement with shading

Perspectives: handling non-smooth reference views, unknown lighting, developing an incremental version to handle larger image sets, reducing shading/creases artefacts.