Photometric 3D-reconstruction

Jean-Denis DUROU, Yvain QUÉAU, and Jean MÉLOU

IRIT, Toulouse, France

Computational Methods for Inverse Problems in Imaging

Como, July 17, 2018

Outline

1. Introduction
2. Shape-from-shading (SfS)
3. Photometric stereo
4. Multi-view shape-from-shading
5. Conclusion and perspectives

Outline

1 Introduction

2 Shape-from-shading (SfS)

3 Photometric stereo

4 Multi-view shape-from-shading

5 Conclusion and perspectives

3D-scanning $\equiv$ 3D-reconstruction + Color estimation

Some applications of 3D-scanning

- Quality control
- Architecture
- Cultural heritage
- Augmented reality
- Metrology

Different kinds of 3D-reconstruction techniques

- Palpation $\equiv$ Mechanical process
- Kinect V1 $\equiv$ Projection of an infra-red pattern
- Kinect V2 $\equiv$ Time of flight of laser pulses
- Photographic techniques $\equiv$ Shape-from-X

3D-scanning ≡ 3D-reconstruction + Color estimation

Some applications of 3D-scanning

- Quality control
- Architecture
- Cultural heritage
- Augmented reality
- Metrology

Different kinds of 3D-reconstruction techniques

- Palpation ≡ Mechanical process
- Kinect V1 ≡ Projection of an infra-red pattern
- Kinect V2 ≡ Time of flight of laser pulses
- Photographic techniques ≡ Shape-from-X

Main shape-from-X techniques

	Geometric techniques	Photometric techniques
$N = 1$ view $N = 1$ image	Structured light Shape-from-texture Shape-from-contour Shape-from-shadow	Shape-from-shading (SfS)
$N = 1$ view $N > 1$ images	Shape-from-defocus	Photometric stereo
$N > 1$ views $N > 1$ images	Stereoscopy Shape-from-silhouettes Structure-from-motion Multi-view stereo	Multi-view SfS

Main shape-from-X techniques

	Geometric techniques	Photometric techniques
$N = 1$ view $N = 1$ image	Structured light Shape-from-texture Shape-from-contour Shape-from-shadow	Shape-from-shading (SfS)
$N = 1$ view $N > 1$ images	Shape-from-defocus	Photometric stereo
$N > 1$ views $N > 1$ images	Stereoscopy Shape-from-silhouettes Structure-from-motion Multi-view stereo	Multi-view SfS

Geometric techniques vs. Photometric techniques

Geometric techniques

- Goal: match feature points in several images
- Tools: linear algebra
- Pros: simple tools; robust
- Cons: sparse 3D-reconstruction; no color estimation

Photometric techniques

- Goal: explain the color of each pixel in each image
- Tools: variational methods
- Pros: dense 3D-reconstruction; color can be estimated
- Cons: nonlinear models; heavy computations

Outline

1 Introduction

2 Shape-from-shading (SfS)

3 Photometric stereo

4 Multi-view shape-from-shading

5 Conclusion and perspectives

SfS: An intuitive technique to infer 3D-shape

This image suffices to infer that the wall is not perfectly flat

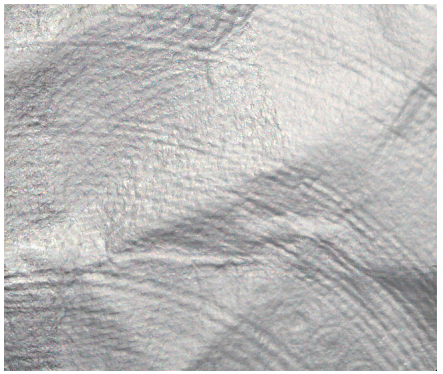
SfS for a Lambertian surface (1/2)

Image

$=$

Albedo

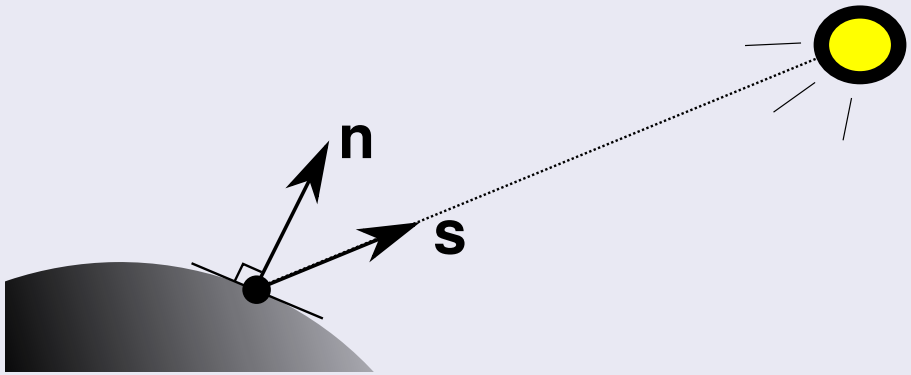
$\times$



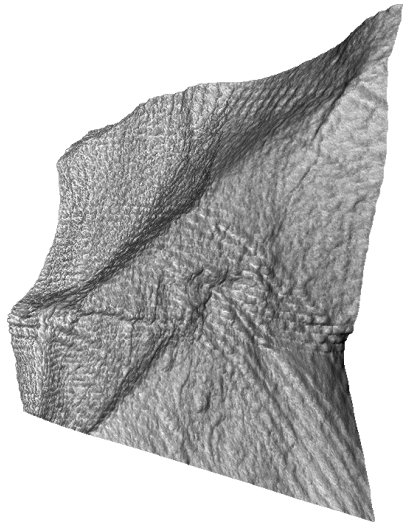
Shading

$$\underbrace{I}_{\text{Image}} = \underbrace{\rho}_{\text{Albedo}} \times \underbrace{\mathbf{s}\cdot\mathbf{n}}_{\text{Shading}}$$

- **s**: lighting vector
- **n**: normal to the surface (3D-shape)



⇓ Shape-from-shading



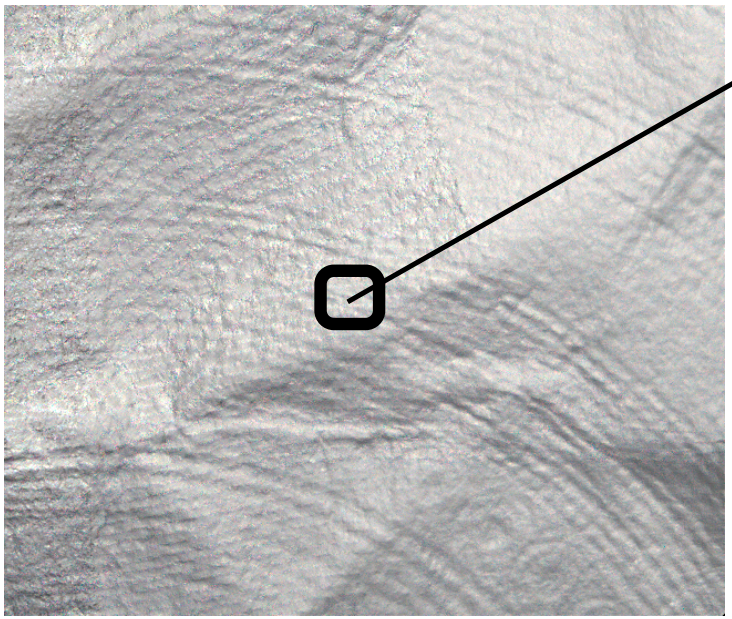
3D-shape

SfS for a Lambertian surface (2/2)

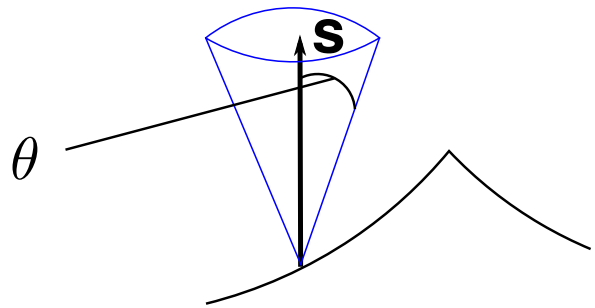
Classical assumptions

- Albedo $\rho \equiv 1$
- Lighting $\mathbf{s} = [0, 0, 1]^\top$

- Data: graylevel I
- Unknown: normal $\mathbf{n}$ such that $\|\mathbf{n}\| = 1$



$$I = \mathbf{s} \cdot \mathbf{n} = \cos\theta$$



The model specifies only one out of the two degrees of freedom of $\mathbf{n}$

$\rightarrow$ Impossible to *locally* estimate the shape

Non-differential approach vs. differential approach

Non-differential approach: unknown normal **n**

$$\min_{\mathbf{n}:\,\Omega\to\mathbb{R}^3} \iint_\Omega [I(u,v) - \mathbf{s}\cdot\mathbf{n}(u,v)]^2\,\mathrm{d}u\mathrm{d}v + \lambda \iint_\Omega \mathcal{R}(\mathbf{n}(u,v))\mathrm{d}u\,\mathrm{d}v$$

- Pros: boundary condition not required; easily discretized
- Cons: approximate solution; **n** integrated afterwards

Differential approach: unknown *depth z*

$$I(u,v) = \mathbf{s}\cdot \underbrace{\frac{1}{\sqrt{\|\nabla z(u,v)\|^2+1}}[-\nabla z(u,v)^\top, 1]^\top}_{\equiv\ \mathbf{n}(u,v)\text{ under orthographic projection}}$$

- Pros: exact solution; direct estimation of the 3D-shape
- Cons: boundary condition required; solution not always realistic

Shape-from-shading: Three major drawbacks

Knowledge of the albedo ρ

- $\rho \equiv 1 \rightarrow$ Arbitrary
- Difficult to get ρ

Knowledge of the lighting vector $\mathbf{s}$

- $\mathbf{s} = [0, 0, 1]^\top \rightarrow$ Arbitrary
- Possible to get $\mathbf{s}$ by calibration

SfS is still ill-posed

- Non-differential approach: Regularization $\rightarrow$ Approximate solution
- Differential approach: Boundary condition $\rightarrow$ Arbitrary

Outline

1 Introduction

2 Shape-from-shading (SfS)

3 Photometric stereo

4 Multi-view shape-from-shading

5 Conclusion and perspectives

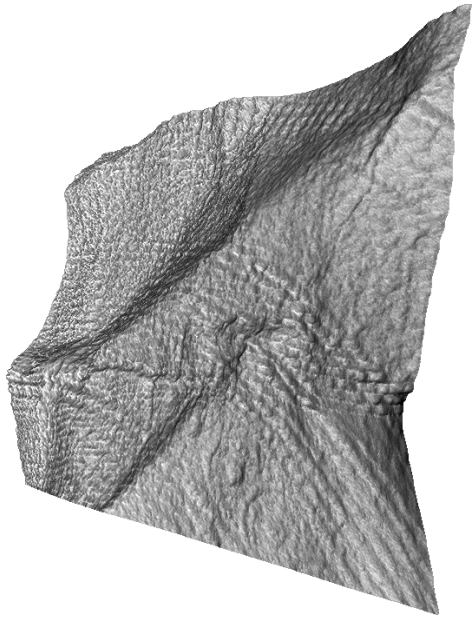
A well-posed extension of SfS

$N > 1$ lighting vectors $\mathbf{s}^1, \ldots \mathbf{s}^N$

$\longrightarrow$

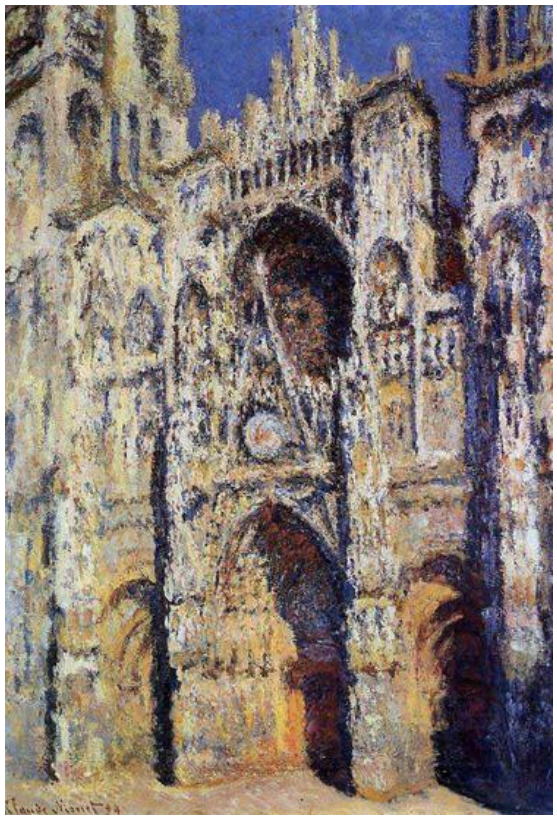
Albedo ρ

$+$



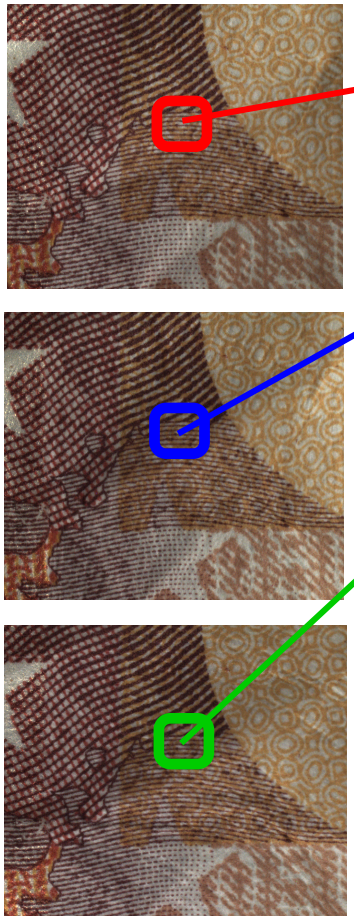
Depth z

Claude Monet: a precursor?



$N \approx 30$ paintings of the cathedral of Rouen (1892-1894)

A linear and local problem [Woodham, 1980]



$$I^1 = \rho\, \mathbf{s}^1 \cdot \mathbf{n} = \mathbf{s}^1 \cdot \mathbf{m}$$
$$I^2 = \rho\, \mathbf{s}^2 \cdot \mathbf{n} = \mathbf{s}^2 \cdot \mathbf{m}$$
$$I^3 = \rho\, \mathbf{s}^3 \cdot \mathbf{n} = \mathbf{s}^3 \cdot \mathbf{m}$$

In each pixel, denoting $\mathbf{m} = \rho\, \mathbf{n}$:

$$I^i = \mathbf{s}^i \cdot \mathbf{m}, \quad i = 1 \ldots N$$

If lighting vectors $\mathbf{s}^i$ are non-coplanar:

- Unique exact solution if $N = 3$
- Unique approximate solution if $N > 3$

Eventually: $\rho = \|\mathbf{m}\|$ and $\mathbf{n} = \frac{\mathbf{m}}{\|\mathbf{m}\|}$

Photometric stereo: Three major advantages vs. SfS

No assumption on the albedo $\rho \rightarrow$ Realistic

- ρ is now an additional unknown
- No other shape-from-X technique can estimate the albedo

Well-posed problem $\rightarrow$ Accurate

- Requires that the N vectors $\mathbf{s}^i$ are known (by calibration)
- Neither regularization, nor boundary condition

Linear and local problem $\rightarrow$ Quick resolution

- Problem to solve in each point: $\min_{\mathbf{m}} \sum_{i=1}^{N} \left(I^i - \mathbf{s}^i \cdot \mathbf{m} \right)^2$
- Integrating the normal field $\mathbf{n}$ into a depth map z is global, yet

First example: $N = 3$ images of a plaster bust

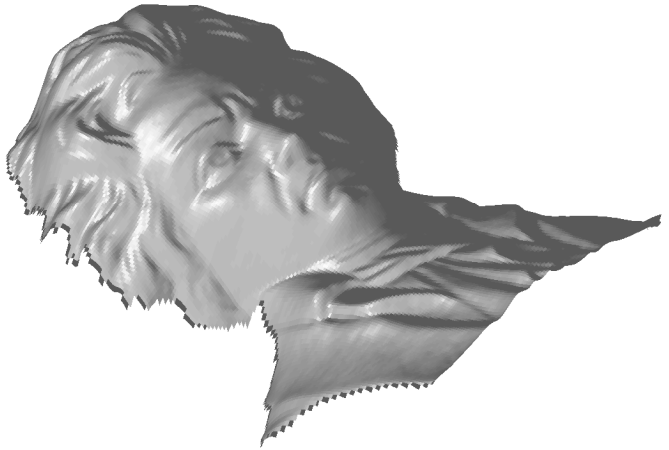
$I^1 = \mathbf{s}^1 \cdot \mathbf{m}$

$I^2 = \mathbf{s}^2 \cdot \mathbf{m}$

$I^3 = \mathbf{s}^3 \cdot \mathbf{m}$

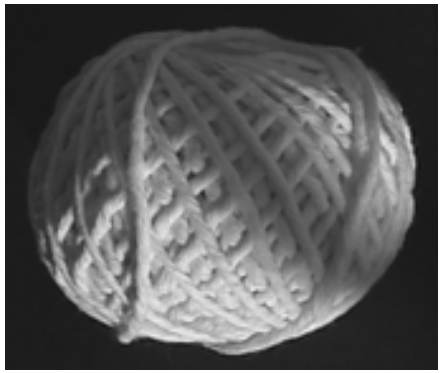
Albedo ρ

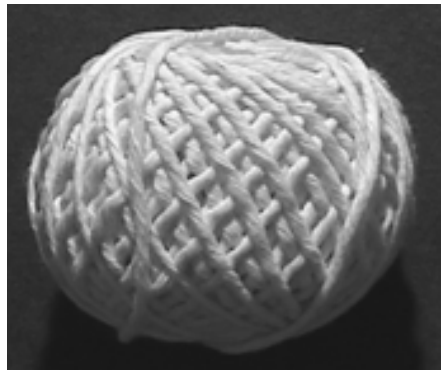
Normal field $\mathbf{n}$



Depth map z

Second example: $N = 2812$ images of a ball of wire





3 images, out of $N = 2812$ [Wu et al., 2006]

Least-squares normal estimation

Robust normal estimation

Historical landmarks

- [Van Diggelen, 1951] First paper on photoclinometry (premise of shape-from-shading)
- [Horn, 1970] First explicit mention of shape-from-shading
- [Woodham, 1980] First paper on photometric stereo
- [Rouy, Tourin, Lions, 1988] First applied mathematics contribution to shape-from-shading (viscosity solutions)
- [Hayakawa, 1994] Renewed interest for photometric stereo (linked to the popularization of digital photography?)
- Since 2000: More papers on photometric stereo than on shape-from-shading
- Since 2010: First convincing applications of photometric stereo

Historical landmarks

- [Van Diggelen, 1951] First paper on photoclinometry (premise of shape-from-shading)
- [Horn, 1970] First explicit mention of shape-from-shading
- [Woodham, 1980] First paper on photometric stereo
- [Rouy, Tourin, Lions, 1988] First applied mathematics contribution to shape-from-shading (viscosity solutions)
- [Hayakawa, 1994] Renewed interest for photometric stereo (linked to the popularization of digital photography?)
- Since 2000: More papers on photometric stereo than on shape-from-shading
- Since 2010: First convincing applications of photometric stereo

Application to 3D-scanning of faces (1/2)

Experimental setup in our lab

Two images of a face, out of $N = 8$

Application to 3D-scanning of faces (2/2)

(bastien.u3d)

CPU time $\approx$ 1 minute

Application to micrometric 3D-scanning (1/2)

Transformation of a dermoscope (skin camera)
into a micrometric 3D-scanner

Application to micrometric 3D-scanning (2/2)

1 euro (Italy)

50 cents (Spain)

1 yuan (China)

3D-reconstructions of metallic coins (accuracy $\approx$ 10 microns)

Outline

1 Introduction

2 Shape-from-shading (SfS)

3 Photometric stereo

4 Multi-view shape-from-shading

5 Conclusion and perspectives

Another well-posed extension of SfS

	Geometric techniques	Photometric techniques
$N = 1$ view $N = 1$ image	Structured light Shape-from-texture Shape-from-contour Shape-from-shadow	Shape-from-shading (SfS)
$N = 1$ view $N > 1$ images	Shape-from-defocus	Photometric stereo
$N > 1$ views $N > 1$ images	Stereoscopy Shape-from-silhouettes Structure-from-motion Multi-view stereo	Multi-view SfS

Multi-view SfS vs. photometric stereo

Photometric stereo $\rightarrow$ Rather inside the lab

- Pro 1: No extrinsic calibration required (single camera pose)
- Pro 2: Estimation of the albedo $\rightarrow$ 3D-scanning
- Pro 3: No matching required
- Con 1: Restricted to depth maps $\rightarrow$ No full 3D-reconstruction
- Con 2: Need to vary the lighting $\rightarrow$ Difficult outside

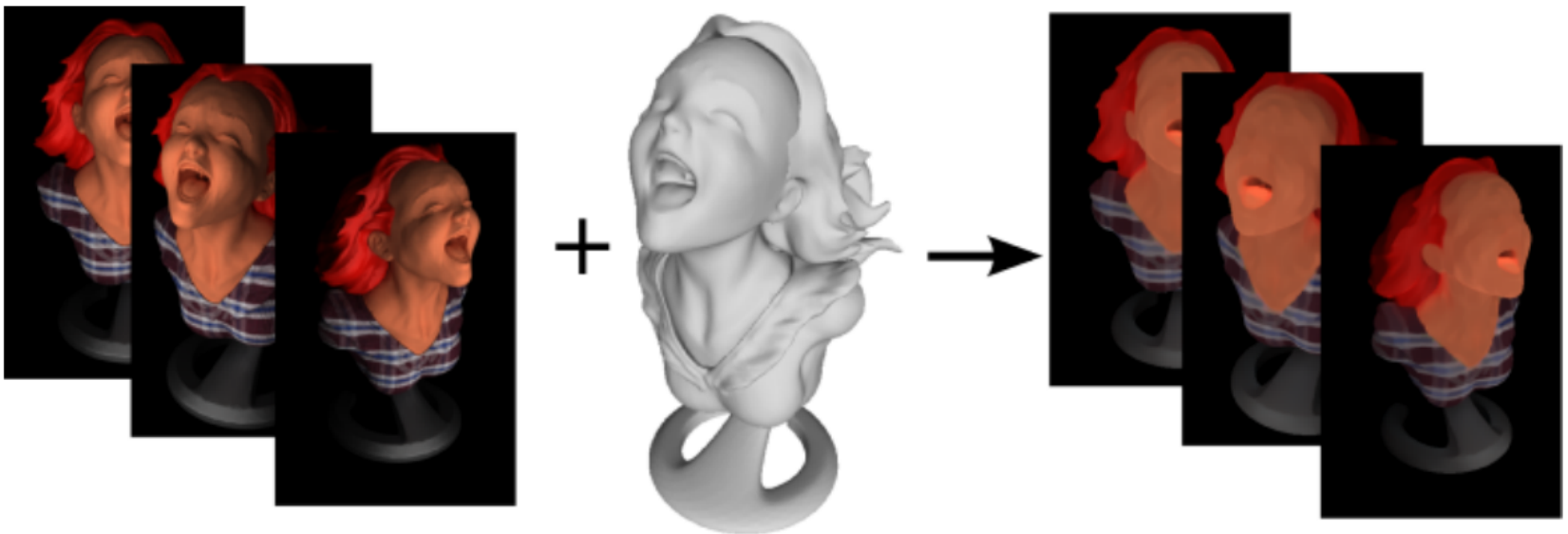
Multi-view SfS $\rightarrow$ Even outside the lab

- Con 1: Extrinsic calibration required ($N > 1$ camera poses)
- Con 2: No estimation of the albedo $\rightarrow$ No 3D-scanning
- Con 3: Matching required
- Pro 1: Not restricted to depth maps $\rightarrow$ Full 3D-reconstruction
- Pro 2: No need to vary the lighting $\rightarrow$ Possible outside

Jean Mélou's PhD thesis (in progress)

3D-scanning for scene relighting applications (augmented reality)

- Collaboration with Mikros Image (Technicolor)
- Part 1: Lighting and albedo estimation from $N > 1$ views + coarse 3D-shape (obtained by depth-sensor or multi-view stereo)
- Part 2: Improvement of the coarse 3D-shape

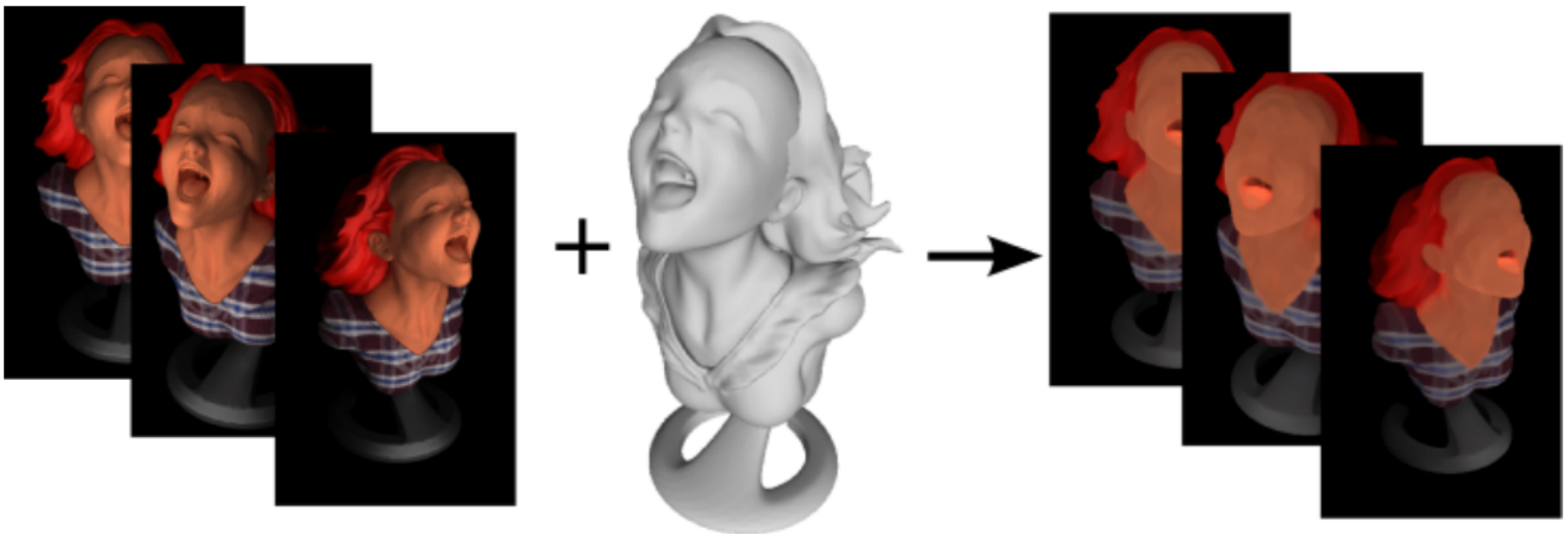


Part 1: Albedo estimation from $N > 1$ views + coarse 3D-shape

Jean Mélou's PhD thesis (in progress)

3D-scanning for scene relighting applications (augmented reality)

- Collaboration with Mikros Image (Technicolor)
- Part 1: Lighting and albedo estimation from $N > 1$ views + coarse 3D-shape (obtained by depth-sensor or multi-view stereo)
- Part 2: Improvement of the coarse 3D-shape $\rightarrow$ Multi-view SfS



Part 1: Albedo estimation from $N > 1$ views + coarse 3D-shape

Multi-view SfS: A practical implementation

Part 2: How to overcome the drawbacks of multi-view SfS

- Con 1: Extrinsic calibration required ($N > 1$ camera poses)
  $\rightarrow$ Use the structure-from-motion technique
- Con 2: No estimation of the albedo $\rightarrow$ No 3D-scanning
  $\rightarrow$ Apply Part 1 of Jean Mélou's PhD thesis
- Con 3: Matching required
  $\rightarrow$ Add an error of reprojection to the model

Model in the case of $N = 2$ images I^1 and I^2

$$\min_z \underbrace{\left\| I^1 - \mathbf{s} \cdot \frac{[-\nabla z^\top, 1]^\top}{\sqrt{\|\nabla z\|^2 + 1}} \right\|^2_{\ell^2(\Omega^1)}}_{\text{SfS error}} + \lambda \underbrace{\left\| I^1 - I^2 \circ h_z \right\|^2_{\ell^2(\Omega^{1,2})}}_{\text{Error of reprojection}}$$

- The homography h_z reprojecting I^1 towards I^2 depends on z
- The set $\Omega^{1,2}$ contains the pixels of Ω^1 which are seen in image I^2

Multi-view SfS: A nonlinear and non-local problem

In comparison with photometric stereo

$$\min_z \left\| I^1 - \mathbf{s} \cdot \frac{[-\nabla z^\top, 1]^\top}{\sqrt{\|\nabla z\|^2 + 1}} \right\|^2_{\ell^2(\Omega^1)} + \lambda \left\| I^1 - I^2 \circ h_z \right\|^2_{\ell^2(\Omega^{1,2})}$$

- The problem is nonlinear in z
- It cannot be solved locally, due to ∇z

Introducing an auxiliary function $\theta : \Omega^1 \to \mathbb{R}^2$

$$\begin{cases} \min\limits_{\theta, z} \left\| I^1 - \mathbf{s} \cdot \frac{[-\theta^\top, 1]^\top}{\sqrt{\|\theta\|^2 + 1}} \right\|^2_{\ell^2(\Omega^1)} + \lambda \left\| I^1 - I^2 \circ h_z \right\|^2_{\ell^2(\Omega^{1,2})} \\ \text{s.t. } \theta = \nabla z \end{cases}$$

Multi-view SfS: A nonlinear and non-local problem

In comparison with photometric stereo

$$\min_z \left\| I^1 - \mathbf{s} \cdot \frac{[-\nabla z^\top, 1]^\top}{\sqrt{\|\nabla z\|^2 + 1}} \right\|^2_{\ell^2(\Omega^1)} + \lambda \left\| I^1 - I^2 \circ h_z \right\|^2_{\ell^2(\Omega^{1,2})}$$

- The problem is nonlinear in z
- It cannot be solved locally, due to ∇z

Introducing an auxiliary function $\theta : \Omega^1 \to \mathbb{R}^2$

$$\begin{cases} \min\limits_{\theta, z} \left\| I^1 - \mathbf{s} \cdot \dfrac{[-\theta^\top, 1]^\top}{\sqrt{\|\theta\|^2 + 1}} \right\|^2_{\ell^2(\Omega^1)} + \lambda \left\| I^1 - I^2 \circ h_z \right\|^2_{\ell^2(\Omega^{1,2})} \\ \text{s.t. } \theta = \nabla z \end{cases}$$

Multi-view SfS: A nonlinear and non-local problem

In comparison with photometric stereo

$$\min_z \left\| I^1 - \mathbf{s} \cdot \frac{[-\nabla z^\top, 1]^\top}{\sqrt{\|\nabla z\|^2 + 1}} \right\|^2_{\ell^2(\Omega^1)} + \lambda \left\| I^1 - I^2 \circ h_z \right\|^2_{\ell^2(\Omega^{1,2})}$$

- The problem is nonlinear in z
- It cannot be solved locally, due to ∇z

Introducing an auxiliary function $\theta : \Omega^1 \to \mathbb{R}^2$

$$\begin{cases} \min\limits_{\theta, z} \left\| I^1 - \mathbf{s} \cdot \dfrac{[-\theta^\top, 1]^\top}{\sqrt{\|\theta\|^2 + 1}} \right\|^2_{\ell^2(\Omega^1)} + \lambda \left\| I^1 - I^2 \circ h_z \right\|^2_{\ell^2(\Omega^{1,2})} \\ \text{s.t. } \theta = \nabla z \end{cases}$$

Multi-view SfS: A nonlinear and non-local problem

In comparison with photometric stereo

$$\min_z \left\| I^1 - \mathbf{s} \cdot \frac{[-\nabla z^\top, 1]^\top}{\sqrt{\|\nabla z\|^2 + 1}} \right\|^2_{\ell^2(\Omega^1)} + \lambda \left\| I^1 - I^2 \circ h_z \right\|^2_{\ell^2(\Omega^{1,2})}$$

- The problem is nonlinear in z
- It cannot be solved locally, due to ∇z

Introducing an auxiliary function $\theta : \Omega^1 \to \mathbb{R}^2$

$$\begin{cases} \min\limits_{\theta, z} \left\| I^1 - \mathbf{s} \cdot \dfrac{[-\theta^\top, 1]^\top}{\sqrt{\|\theta\|^2 + 1}} \right\|^2_{\ell^2(\Omega^1)} + \lambda \left\| I^1 - I^2 \circ h_z \right\|^2_{\ell^2(\Omega^{1,2})} \\ \text{s.t. } \theta = \nabla z \end{cases}$$

Multi-view SfS: Numerical resolution by ADMM

Introducing Lagrange multipliers $\mu : \Omega^1 \to \mathbb{R}^2$ and a parameter $\beta > 0$

$$\min_{\theta,z,\mu} \left\| I^1 - \mathbf{s} \cdot \frac{[-\theta^\top, 1]^\top}{\sqrt{\|\theta\|^2 + 1}} \right\|^2_{\ell^2(\Omega^1)} + \lambda \left\| I^1 - I^2 \circ h_z \right\|^2_{\ell^2(\Omega^{1,2})} + \langle \mu | \nabla z - \theta \rangle + \beta \left\| \nabla z - \theta \right\|^2_{\ell^2(\Omega^1)}$$

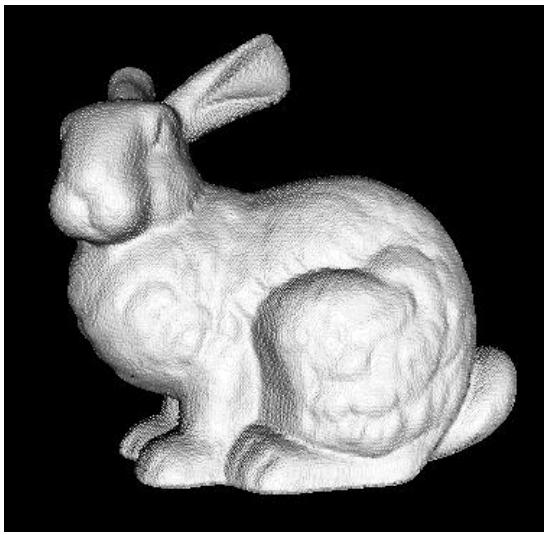
Iterative scheme ADMM (alternating direction method of multipliers)

Repeat until convergence (parameter β is automatically updated):

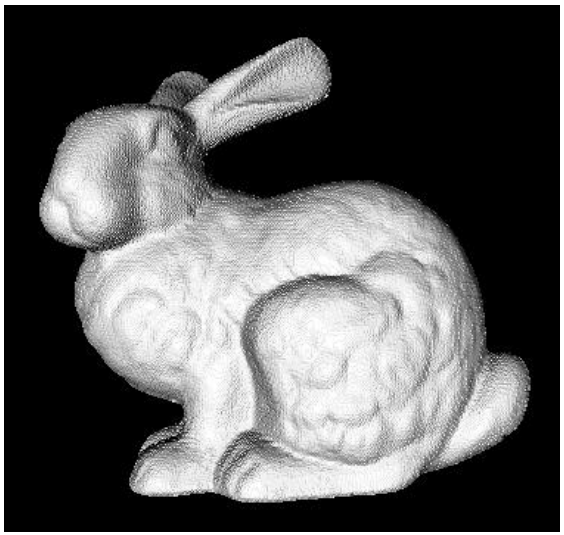
1. $\min_{\theta} \left\| I^1 - \mathbf{s} \cdot \frac{[-\theta^\top, 1]^\top}{\sqrt{\|\theta\|^2 + 1}} \right\|^2_{\ell^2(\Omega^1)} + \langle \mu^{(k)} | \nabla z^{(k)} - \theta \rangle + \beta^{(k)} \left\| \nabla z^{(k)} - \theta \right\|^2_{\ell^2(\Omega^1)}$
2. $\min_{z} \lambda \left\| I^1 - I^2 \circ h_z \right\|^2_{\ell^2(\Omega^{1,2})} + \langle \mu^{(k)} | \nabla z - \theta^{(k+1)} \rangle + \beta^{(k)} \left\| \nabla z - \theta^{(k+1)} \right\|^2_{\ell^2(\Omega^1)}$
3. $\mu^{(k+1)} = \mu^{(k)} + \beta^{(k)} \left(\nabla z^{(k+1)} - \theta^{(k+1)} \right)$

Multi-view SfS: A preliminary result

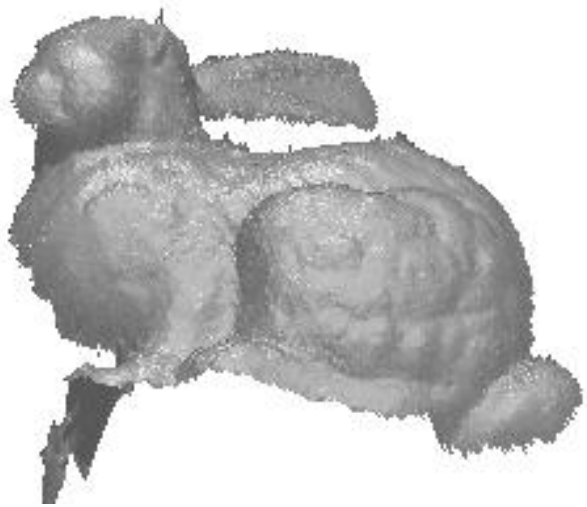
- $N = 2$ synthetic images of size 540×540 ($|\Omega^1| = 63474$ pixels)
- Knowledge of: camera poses, albedo ($\rho \equiv 1$), lighting vector $\mathbf{s}$
- Hyper-parameter $\lambda = 0.1$
- $z^{(0)} \equiv$ fronto-parallel plane
- Result obtained after 84 iterations (92s on an i5 processor)



I^1



I^2



3D-reconstruction

Outline

1 Introduction

2 Shape-from-shading (SfS)

3 Photometric stereo

4 Multi-view shape-from-shading

5 Conclusion and perspectives

Conclusion: Two well-posed extensions of SfS

Both extensions have very different features:

Photometric stereo

- Single camera pose
- $N > 1$ lighting vectors
- Inside the lab
- Linear and local problem

Multi-view SfS

- Single lighting vector
- $N > 1$ camera poses
- Outside the lab
- Nonlinear and non-local problem

Future work on multi-view SfS

Theory: Collaboration with Maurizio Falcone and Silvia Tozza

- Proof of uniqueness by Chambolle (Annales de l'IHP, 1994):
  - Smooth white Lambertian surface
  - Frontal lighting
  - Pure translation of the camera ($N = 2$ views)
  - Depth z known on the boundary
- Extension of the proof:
  - Any direction of the lighting
  - Any pair of camera poses
  - Absence of boundary condition

Practice: Jean Mélou's PhD thesis (in progress)

- Extension to $N > 2$ images $\rightarrow$ Full 3D-reconstruction
- Use regularization or robust estimators in the model $\rightarrow$ Real data
- Towards a complete 3D-scanner?