

# Photometric 3D-reconstruction

Yvain QUÉAU

CNRS researcher

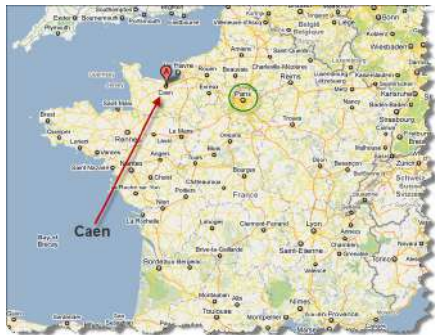
GREYC laboratory  
Normandie Univ, UNICAEN, ENSICAEN, CNRS  
Caen, France

`yvain.queau@ensicaen.fr`

QCAV 2021  
May 12th, 2021

# About me

- 2015: PhD in computer vision in Toulouse
- Since 2018: CNRS researcher in Caen



# Aim of the talk

How to obtain highly detailed 3D-models of the world ?



Photometric 3D-reconstruction of the Bayeux Tapestry  
(11th century, wool strings on linen)

# Aim of the talk

How to obtain highly detailed 3D-models of the world ?



Photometric 3D-reconstruction of a 50 euro cent coin

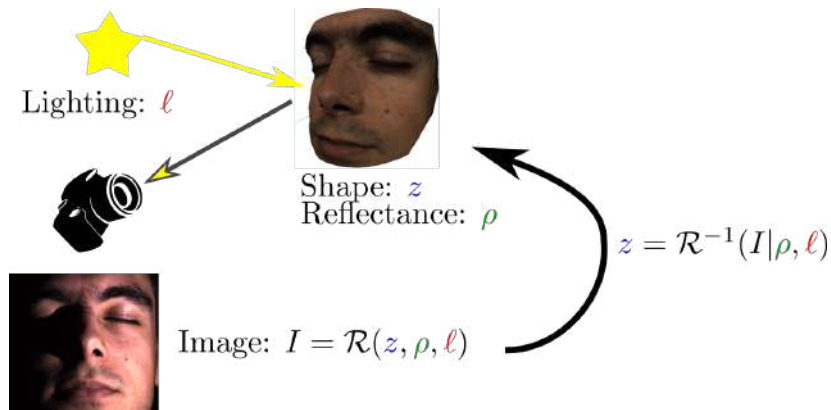
# Outline

- 1 Introduction to photometric 3D-reconstruction
- 2 Shape-from-shading
- 3 Photometric stereo
- 4 The Inclusive Museum Guide project

# Outline

- 1 Introduction to photometric 3D-reconstruction
- 2 Shape-from-shading
- 3 Photometric stereo
- 4 The Inclusive Museum Guide project

# Photometric 3D-reconstruction as an inverse problem



Goal: invert the radiance function  $\mathcal{R}$  to recover the shape  $z$  (and, possibly, lighting  $\ell$  and reflectance  $\rho$ ) of the pictured surface

# Context: shape-from-X

	Geometric techniques	Photometric techniques
Single image	Structured light Shape-from-shadows Shape-from-contours Shape-from-texture Shape-from-template	<b>Shape-from-shading</b>
Multi-images	Structure-from-motion Stereopsis Shape-from-silhouettes Shape-from-focus	<b>Photometric stereo</b> Shape-from-polarisation

*Geometric techniques aim at identifying and analysing features. This presentation rather focuses on two photometric techniques (shape-from-shading and photometric stereo), which aim at inverting a physics-based image formation model.*



# Geometric techniques

Mature workflows based on structure-from-motion and multi-view stereo are available, e.g. Alice Vision's Meshroom:



## Limitations:

- May miss thin structures
- Only estimate shape, not reflectance

# 3D-scanning

3D-scanning = estimation of shape *and* color



Image



Reflectance



(a)



(b)



(c)



3D-model



(d)



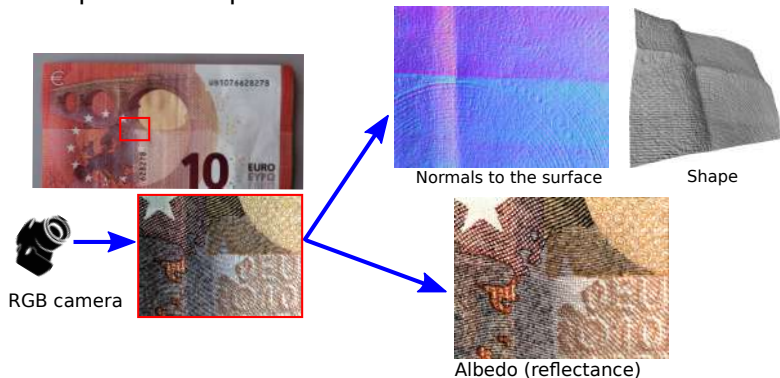
(e)

- Geometric techniques (e.g. SfM, MVS) only recover shape (no color estimation)
- Photometric techniques (e.g., shape-from-shading, photometric stereo), which are based on **inverting the image formation model**, recover both ( $\approx$  *inverse rendering*)

# Photometric techniques

Top-performing for the recovery of very thin structures

Example: close-up on a 10 euros banknote



**Photometric 3D-reconstruction =**

Shape analysis through *luminous quantities measurement* (photo-metry), by reverting the image formation process

# Photometric techniques

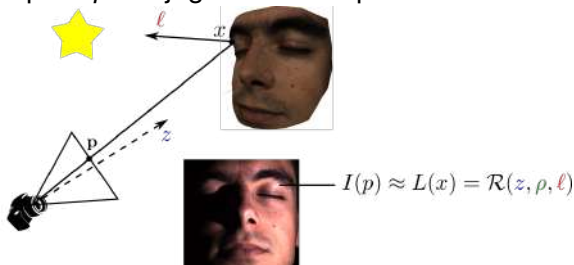
Top-performing for the recovery of very thin structures

Example: close-up on a 5 euros banknote



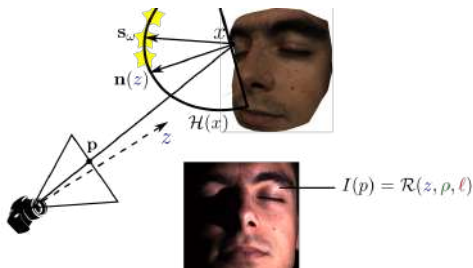
# Digital cameras as luminance measurement devices

Consider a pixel  $p$  conjugate to a 3D-point  $x$  on the surface:



- The measured brightness (irradiance)  $I(p)$  is proportional (up to thermal noise, saturation, etc.) to the scene's luminance  $L(x)$
- The scene's luminance  $L(x)$  is a function  $\mathcal{R}$  of shape ( $z$ ), reflectance ( $\rho$ ) and lighting ( $l$ )

# Forward image formation model



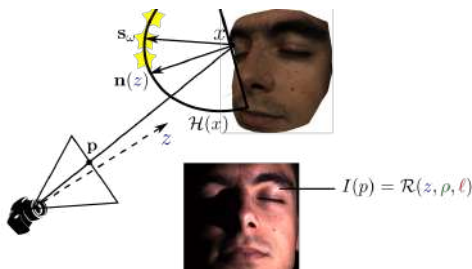
Considering all incident light directions  $\omega \in \mathcal{H}(x) \subset \mathbb{S}^2$  yields the celebrated *rendering equation*:

$$I(p) = \mathcal{R}(z, \rho, \ell) = \int_{\omega \in \mathcal{H}(x)} \rho \phi_\omega \max\{0, \underbrace{\mathbf{s}_\omega \cdot \mathbf{n}(z)}_{\text{shading}}\} d\omega$$

with  $\mathbf{n}$  the surface normal at  $x$ , which characterizes the shape  $z$

3D-reconstruction  $\approx$  Inverse rendering:  $z = \mathcal{R}^{-1}(I)$

# General inverse problem



$$I(p) = \mathcal{R}(z, \rho, \ell) = \int_{\omega \in \mathcal{H}(x)} \rho \phi_{\omega} \max\{0, \mathbf{s}_{\omega} \cdot \mathbf{n}(z)\} d\omega$$

→ Inverting this image formation model i.e., computing  $\mathcal{R}^{-1}$ , allows the estimation of the surface shape (characterized here by the normals  $\mathbf{n}(z)$ ) and its material (characterized here by the reflectance  $\rho$ ).

# Simplifying assumptions

In general, we need to simplify the forward model

$$I(\rho) = \mathcal{R}(z, \rho, \ell) = \int_{\omega \in \mathcal{H}(x)} \rho \phi_{\omega} \max\{0, \mathbf{s}_{\omega} \cdot \mathbf{n}(z)\} d\omega$$

because of:

- the integral over all lighting directions
- the complexity of the function  $\rho$  (BRDF)

For simplicity, we often (but not always) assume that:


- There is a single, infinitely far away light source  $\rightarrow$  integrand is a Dirac and lighting is characterized by a vector  $\ell \in \mathbb{R}^3$
- The surface is assumed Lambertian  $\rightarrow$  reflectance is simply a map  $\rho : \Omega \rightarrow \mathbb{R}$  of *albedo* values

$$I(\rho) = \mathcal{R}(z, \rho, \ell) = \rho \max\{0, \ell \cdot \mathbf{n}(z)\}$$



# Simplified image formation model

Extending to RGB images, we obtain the following simplified image formation model:



The diagram illustrates the simplified image formation model. It shows three images of a white cat lying down, arranged horizontally. The first image on the left is the RGB image, showing the cat with some colored markings. The second image in the middle is the Albedo map, which is a grayscale representation of the cat's surface color, with the same colored markings. The third image on the right is the Shading map, which is a grayscale representation of the lighting and geometry interaction, showing the cat's form and shadows. The images are separated by an equals sign (=) and a dot operator (⊙).

RGB image  
 $I : \Omega \rightarrow \mathbb{R}^3$

Albedo  
 $\rho : \Omega \rightarrow \mathbb{R}^3$

Shading  
 $S(\mathbf{z}, \ell) : \Omega \rightarrow \mathbb{R}$

where albedo (Lambertian reflectance)  $\equiv$  color, and shading  $\equiv$  lighting-geometry interaction.

Remark: this is the classic model used in “intrinsic image decomposition”

Credit: MIT Intrinsic Images Dataset

# Simplified image formation model

Shading  $S(\mathbf{z}, \ell)$  represents lighting-geometry interaction:



=



⊙



Shading  
 $S(\mathbf{z}, \ell)$



=



⊙



Lighting  
 $\ell \in \mathbb{S}^2$

·



Normals  
 $\mathbf{n}(\mathbf{z}) : \Omega \rightarrow \mathbb{S}^2$

RGB image  
 $\mathbf{I} : \Omega \rightarrow \mathbb{R}^3$

Albedo  
 $\rho : \Omega \rightarrow \mathbb{R}^3$

where the surface normal  $\mathbf{n}$  locally characterizes the surface's shape. *Shape-from-shading* aims at inverting this forward model, assuming known albedo and lighting

# Outline

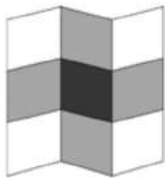
- 1 Introduction to photometric 3D-reconstruction
- 2 Shape-from-shading**
- 3 Photometric stereo
- 4 The Inclusive Museum Guide project

# Shape-from-shading: a classic ill-posed problem

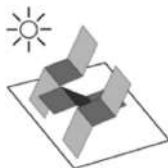
Given an image  $\mathbf{I} : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}^m$ , *Shape-from-Shading* (SfS) consists in inverting the image irradiance equation

$$\mathbf{I} = \mathcal{R}(z, \rho, \ell)$$

with  $\mathcal{R}$  a *radiance* function depending on the unknown depth  $z : \Omega \rightarrow \mathbb{R}$ , surface reflectance  $\rho : \Omega \rightarrow \mathbb{R}^m$ , and incident lighting  $\ell : \Omega \rightarrow \mathbb{S}^2$ .



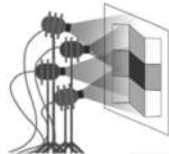
RGB image:  
 $\mathbf{I}$



Sculptor's  
explanation:  
 $z$



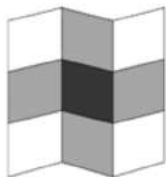
Painter's  
explanation:  
 $\rho$



Gaffer's  
explanation:  
 $\ell$

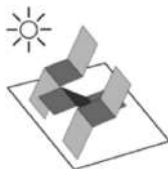
*Adelson and Pentland's workshop metaphor*

# Simplifying assumptions



RGB image:

$I$



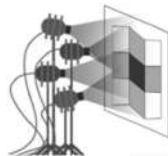
Sculptor's  
explanation:

$z$



Painter's  
explanation:

$\rho$



Gaffer's  
explanation:

$\ell$

It is impossible to tell **reflectance** from **shape** and **lighting** !

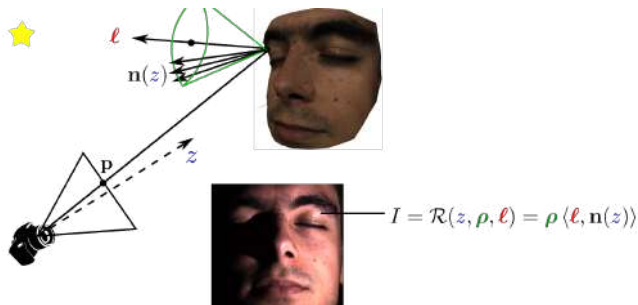
→ Lighting and reflectance are usually assumed to be known

# Illustration of ill-posedness

Even with known **reflectance**  $\rho$  and **lighting**  $\ell$ , shape estimation by SfS is an ill-posed inverse problem (Horn, 1970).

Example: two solutions of  $\mathbf{I} = \mathcal{R}(z, \rho, \ell)$  with  $\mathbf{I} := \text{Lena}$ , **white reflectance** ( $\rho \equiv 1$ ) and **frontal lighting** ( $\ell \equiv [0, 0, -1]^T$ ):

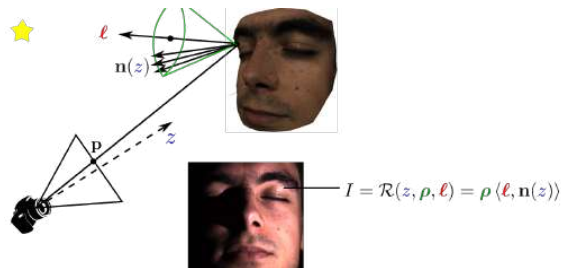
# Local ill-posedness in shape-from-shading



Even if  $\rho$  and  $\ell$  are known, the normal cannot be estimated directly by inverting the forward model, because infinitely many normals, lying on a cone, are equally admissible

→ A regularization mechanism is necessary. The most natural way to do so consists in considering a differential version of the forward model, assuming the surface is differentiable almost everywhere

# Differential image formation model



Assuming orthographic projection, the scene's surface is characterized by a depth map  $z : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$  such that

$$\mathbf{n}(z) = \frac{1}{\sqrt{|\nabla z|^2 + 1}} \begin{bmatrix} \nabla z \\ -1 \end{bmatrix}$$

And thus we get the following nonlinear PDE in  $z$ :

$$I = \mathcal{R}(z, \rho, \ell) = \frac{\rho}{\sqrt{|\nabla z|^2 + 1}} \left\langle \ell, \begin{bmatrix} \nabla z \\ -1 \end{bmatrix} \right\rangle$$



# Eikonal model

An important instance of the differential model

$$I = \mathcal{R}(z, \rho, \ell) = \frac{\rho}{\sqrt{|\nabla z|^2 + 1}} \left\langle \ell, \begin{bmatrix} \nabla z \\ -1 \end{bmatrix} \right\rangle$$

is that of a white surface ( $\rho \equiv 1$ ) under frontal lighting ( $\ell = [0, 0, -1]^\top$ ). This yields the celebrated *eikonal equation*:

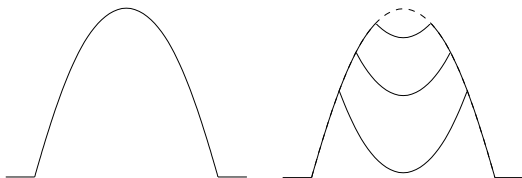
$$|\nabla z| = \sqrt{\frac{1}{\rho^2} - 1}$$

# The concave/convex ambiguity

The eikonal equation

$$|\nabla z| = \sqrt{\frac{1}{f^2} - 1}$$

characterizes the gradient's magnitude, but not its sign: this is the concave/convex ambiguity:



*Left: example of the 1D-surface  $z = 1 - x^2$ . Right: Under vertical lighting, three other solutions (amongst an infinity), which are differentiable almost everywhere, satisfy the same eikonal equation.*

# The concave/convex ambiguity

Humans solve the concave/convex ambiguity intuitively because we learnt that the world is mostly “convex”



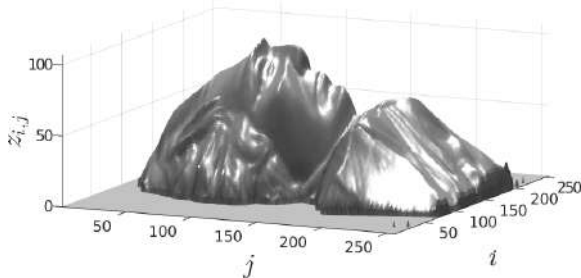
*A volcano in Hawaii (copyright Whitman Richards), looking like a crater if the image is rotated by  $\pi$ .*

Among all the solutions to the eikonal equation, it is thus reasonable to look for the “maximum” one

# Eikonal shape-from-shading

Example of maximal viscosity solution estimate, on a real-world dataset:

3D-reconstruction at iteration 512



Results are reasonable, yet they could be improved by adding regularization (e.g., to ensure smoothness, or to enforce adequation to a low-resolution prior) → **Variational methods**

# Application to RGB-D sensing



Depth  
image



Shape



RGB image

## Depth image has

- noise and quantization
- missing areas
- **coarse resolution**

## RGB image has

- less noise and quantization
- no missing area
- **high resolution**

Goal:

Combine data to get **high-resolution shape**

# Application to RGB-D sensing



Depth  
image



Shape



RGB image



High-resolution  
shape

## Depth image has

- noise and quantization
- missing areas
- **coarse resolution**

## RGB image has

- less noise and quantization
- no missing area
- **high resolution**

Goal:

Combine data to get **high-resolution shape**

# Joint depth super-resolution and shape-from-shading

Given a low-resolution depth map  $z_0 : \Omega_{LR} \rightarrow \mathbb{R}$ , *depth super-resolution* consists in inverting a forward downsampling model

$$z_0 = Dz$$

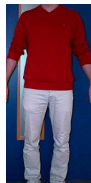
with  $z : \Omega_{HR} \rightarrow \mathbb{R}$  the (unknown) high-resolution depth, and  $D$  a downsampling operator

- $D$  is not invertible: there are infinitely many ways to interpolate between two points (“high-frequency” ambiguity)
- $\rightarrow$  Among the possible interpolations, we will choose the one which is consistent with the image formation model
- The low-resolution depth will serve as prior to solve the “low-frequency” ambiguities of SfS (e.g., concave/convex)

# Joint depth super-resolution and shape-from-shading

In real-world:

- albedo is rarely uniform
- lighting is rarely frontal



→ Simultaneous estimation of high-resolution shape, *piecewise-constant* reflectance and *spherical harmonics* lighting within an energy minimization framework



# Variational formulation

$$\min_{\substack{\mathbf{z}: \Omega_{HR} \rightarrow \mathbb{R} \\ \rho: \Omega_{HR} \rightarrow \mathbb{R}^3 \\ \ell \in \mathbb{R}^4}} \underbrace{\left\| \mathbf{I} - \rho < \ell, \begin{bmatrix} \mathbf{n}(\mathbf{z}) \\ 1 \end{bmatrix} > \right\|_{\ell_2(\Omega_{HR})}^2}_{\text{Shape-from-shading}} + \underbrace{\mu \|\mathbf{z}_0 - D\mathbf{z}\|_{\ell_2(\Omega_{LR})}^2}_{\text{super-resolution}} + \nu P_1(\mathbf{z}) + \lambda P_2(\rho)$$

Here:

- $P_1$  is a minimal surface regularization term on shape:

$$P_1(\mathbf{z}) = \|\mathrm{d}A(\mathbf{z})\|_{\ell_1(\Omega_{HR})}$$

- $P_2$  is a Potts regularization term on albedo:

$$P_2(\rho) = \|\nabla \rho\|_{\ell_0(\Omega_{HR})} = \sum_{\mathbf{p} \in \Omega_{HR}} \begin{cases} 0, & \text{if } |\nabla \rho(\mathbf{p})|_F = 0 \\ 1, & \text{otherwise} \end{cases}$$

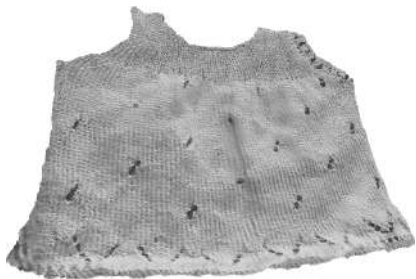
# Some results



Input depth



Input RGB



Depth estimate

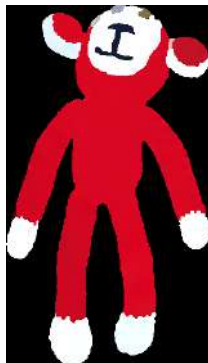
# Some results



Input RGB



Input  
depth



Albedo  
estimate



Depth estimate

# Some results



Input RGB



Input depth



Albedo estimate



Depth estimate

# Some results



Input RGB



Albedo estimate



Depth image



Depth estimate

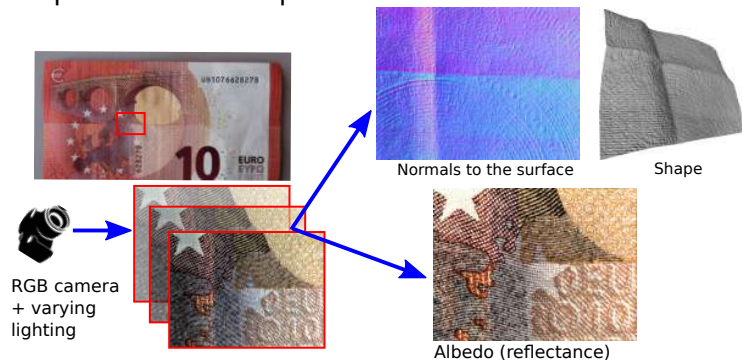
# Outline

- 1 Introduction to photometric 3D-reconstruction
- 2 Shape-from-shading
- 3 Photometric stereo**
- 4 The Inclusive Museum Guide project

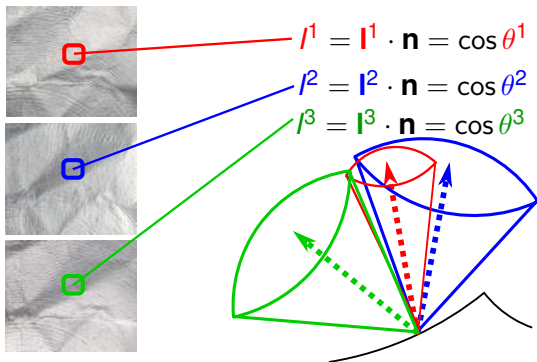
# Photometric stereo

When no prior knowledge is available, it is hopeless to achieve a reasonable 3D-reconstruction based on SfS

Photometric stereo (Woodham, 1978) is an extension of SfS which considers multiple images of the surface, taken from the **same viewing angle** but under **varying lighting**. It can unambiguously recover geometry, and it is the only shape-from-X technique which is able to **estimate reflectance**



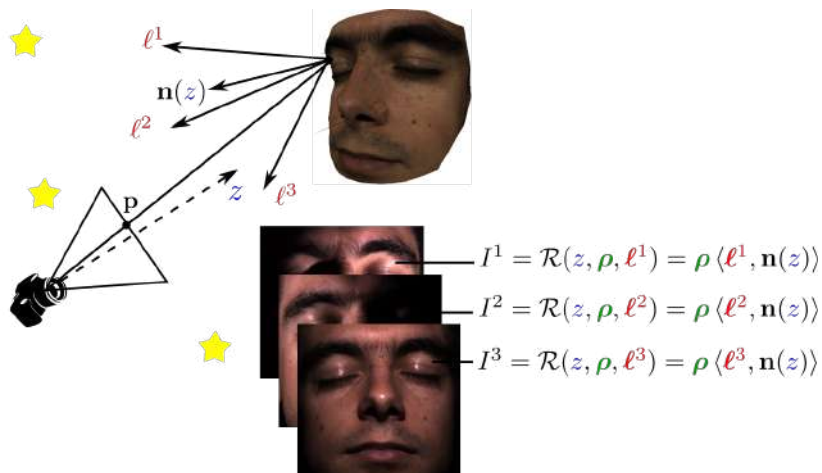
# Well-posedness of photometric stereo



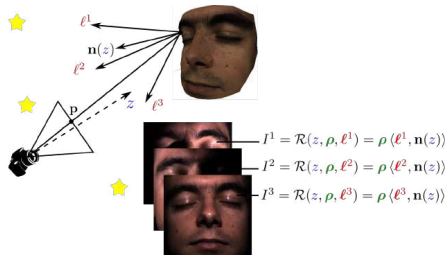
- $m = 1$  image (SfS), known albedo: infinitely many possible unit-length normals in each point
- $m = 2$  images, known albedo: up to two possible normals
- $m \geq 3$ : unique approximate solution - problem is over-constrained (thus albedo can be estimated !)



# Linear photometric stereo model



# Local estimation of shape and reflectance



Linear system in  $\mathbf{m}(\rho, \mathbf{z}) := \rho \mathbf{n}(z) \in \mathbb{R}^3$ :

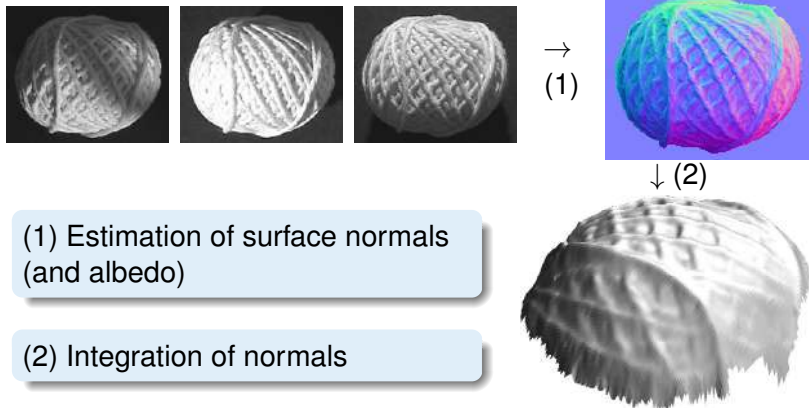
$$\begin{bmatrix} I^1 \\ I^2 \\ \vdots \\ I^m \end{bmatrix} = \begin{bmatrix} \ell^{1T} \\ \ell^{2T} \\ \vdots \\ \ell^{mT} \end{bmatrix} \mathbf{m}(\rho, \mathbf{z})$$

- If: a)  $m \geq 3$ , b) lighting vectors  $\ell^i$  are known, and c) they are non-coplanar,
- Vector  $\mathbf{m}(\rho, \mathbf{z})$  can be estimated in each pixel,
- Then we deduce the normal **and albedo** by

$$\rho = |\mathbf{m}(\rho, \mathbf{z})|$$
$$\mathbf{n}(z) = \frac{\mathbf{m}(\rho, \mathbf{z})}{|\mathbf{m}(\rho, \mathbf{z})|}$$

- Eventually, normals  $\mathbf{n}(z)$  are *integrated* into depth  $z$

# Two-stage resolution of photometric stereo



Such a pipeline can be extended to unknown or non-distant lighting, non-Lambertian reflectance, perspective camera, etc.

# Outliers in photometric stereo

Real-world objects rarely meet the linear Lambertian model:

- self-shadows
- cast-shadows
- inter-reflections
- kinks and discontinuities
- slight camera displacements...



Ball



Bear



Cat



Pot1



Reading

*The "Diligent" image dataset*

# Handling outliers

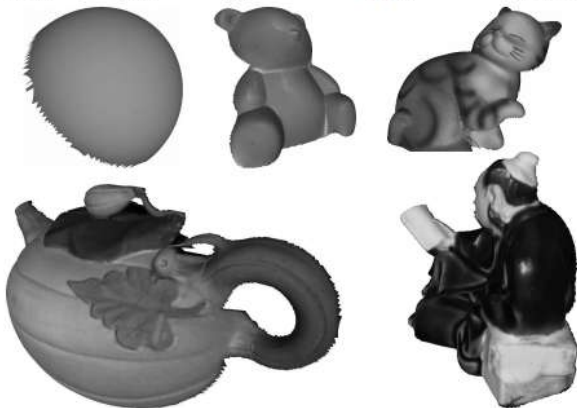
Outliers to the Lambertian model can be taken into account either by:

- 1) Considering a more evolved reflectance model, e.g., Cook-Torrance (yields numerical difficulties)
- 2) Training on a large dataset (current state-of-the-art)
- 3) Adopting a robust solving strategy (very efficient if outliers are “sparse”). E.g.:

$$\min_{z, \rho} \sum_{i=1}^m \iint_{x \in \Omega} \Phi \left( \left| \rho(x) \ell^i \cdot \left[ \nabla z(x)^\top, -1 \right]^\top - I^i(x) \right| \right) dx$$

with  $\phi$  a robust estimator (1-norm, Cauchy, etc.)

# 3D-reconstruction of the Diligent dataset



# 3D-reconstruction of metallic coins



1 euro (Italy)



50 cents (Spain)



1 yuan (China)



3D-reconstructions

# 3D-reconstructions of the human skin



# 3D-reconstruction of faces



(a)



(b)



(c)



(d)

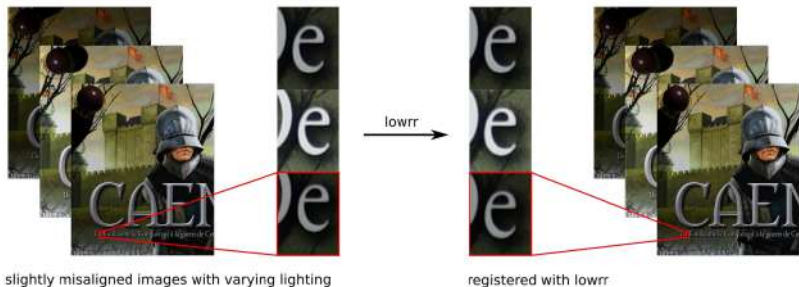


(e)

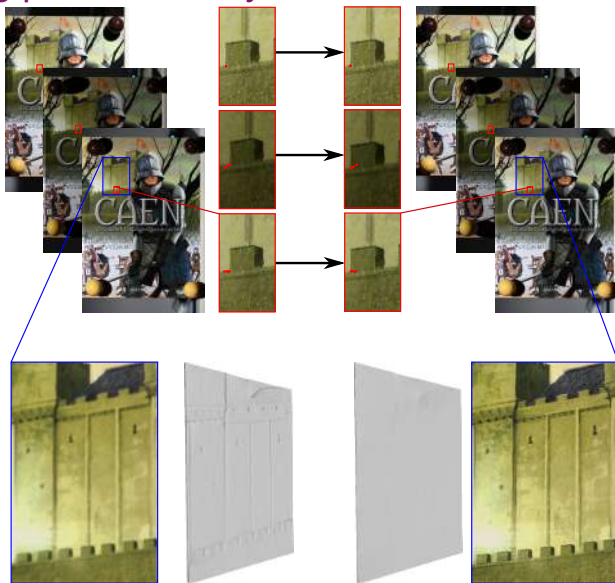
# Reaching pixel accuracy

Since the camera is still, we can get one depth and albedo estimate per pixel  $\rightarrow$  the only limitation is the camera resolution – in theory

Yet, triggering the camera may induce slight displacements (up to a few pixels) and cause geometry hallucination.  $\Rightarrow$  Images should be registered using, e.g., low-rank :

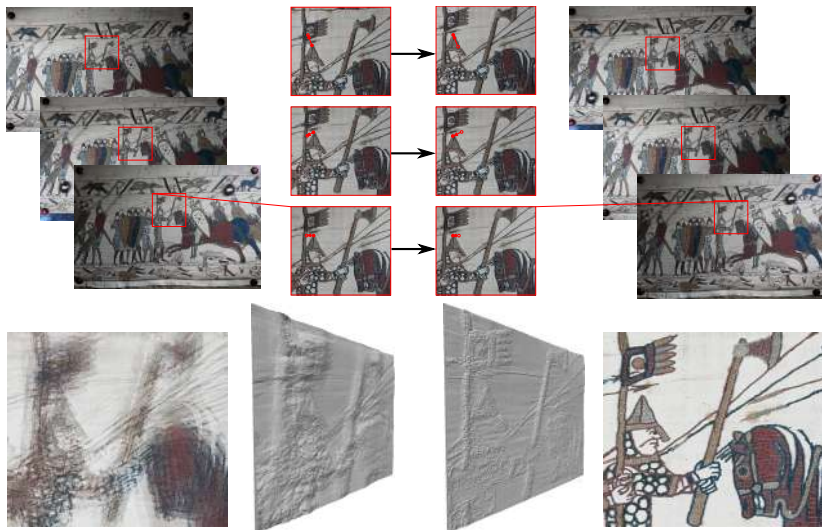


# Reaching pixel accuracy



*Effect of image registration on 3D-reconstruction by photometric stereo*

# Reaching pixel accuracy



*3D-reconstruction of the Bayeux Tapestry from a hand-held image sequence*

# Outline

- 1 Introduction to photometric 3D-reconstruction
- 2 Shape-from-shading
- 3 Photometric stereo
- 4 The Inclusive Museum Guide project**

# Context of the project

- Goal: Make visual artworks accessible to visually-impaired people (“inclusive” approach)
- Funding: Normandy region and ANR projects “Inclusive Museum Guide” (2019-2024)
- Partners: Caen univ., Rouen univ., CNRS, Bayeux museum, Royal Holloway univ., Westminster univ.



# Bayeux tapestry



70m-long medieval wool and linen *embroidery* telling the conquest of England by William, Duke of Normandy, in 1066

## Bayeux tapestry: Edward



Story starts with the death of old king Edward...



## Bayeux tapestry: William



Edward had chosen his cousin William for his succession on the British throne

# Bayeux tapestry: Harold



Yet Edward's brother-in-law Harold took the throne

# Bayeux tapestry: the Northmen armada



William's armada crossed the Channel

# Bayeux tapestry: the battle of Hastings



... and William defeated Harold during the battle of Hastings, hence becoming King of England

# The IMG project



Question: how can we make visually-impaired people “feel” such artifacts, knowing that

- they cannot see it
- they cannot touch it (the tapestry is protected by glass)

→ Re-think the way we access visual artworks:  
audio-descriptions, navigation tools, **tactile representations**,  
etc.

# Towards a tactile representation of visual artworks



Goal: extract micro-geometry of the scene (“texture”),  
in view of 3D-printing

Challenges: no direct access (behind glass), fragility to light  
exposure, thinness of the wool strings...

# Campaign preparation: handling the glass

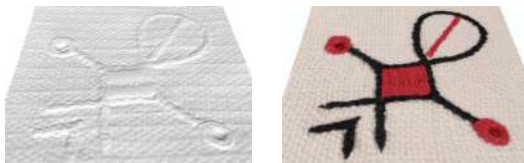


Proof of concept experiment by M. Pizenberg during lockdown # 1: self-made embroidery behind a glass, with stuck spheres for calibrating a moving flash light source

# Campaign preparation: handling the glass



Captured images



3D-reconstruction

Fine-scale structures are revealed. Robust photometric stereo algorithms are able to handle reflections and shadows induced by the glass



## Campaign preparation: limiting light

The tapestry is fragile, so we need to limit the luminous flux we will project onto it → very short camera exposure time of  $1/200s$ , and use of a synchronized flash light source



## Campaign preparation: measuring light

To get access to the tapestry, we had to obtain agreement from authorities, announcing how much light would be projected → need to measure the luminous flux emitted by the flash, by calibrating the camera response and using a lux-meter



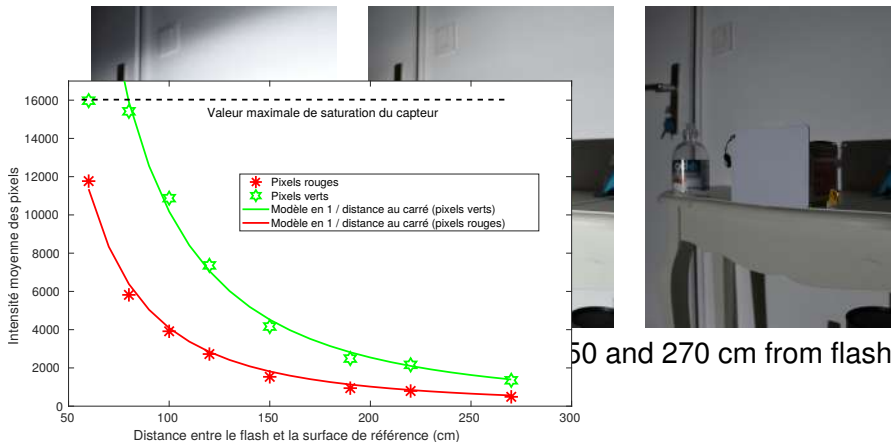
Instant light intensity measured by a lux-meter, to verify the sensor linearity

## Campaign preparation: measuring light



Images of a Lambertian plane at 60, 150 and 270 cm from flash

# Campaign preparation: measuring light



→ total surfacic energy at 1 m, for 12 shots per scene: 312 lux.s  
( $\approx$  6 sec. of the already in-place lighting system)

⇒ acceptable for authorities

# Data acquisition

Selection of 12 parts of the tapestry for digitization, and in-situ image acquisition in January 2021:

Acquisition of the death of Harold sequence

# The death of Harold sequence

Acquired data and 3D-reconstruction:



# The death of Harold sequence

Acquired data and 3D-reconstruction:



# The death of Harold sequence

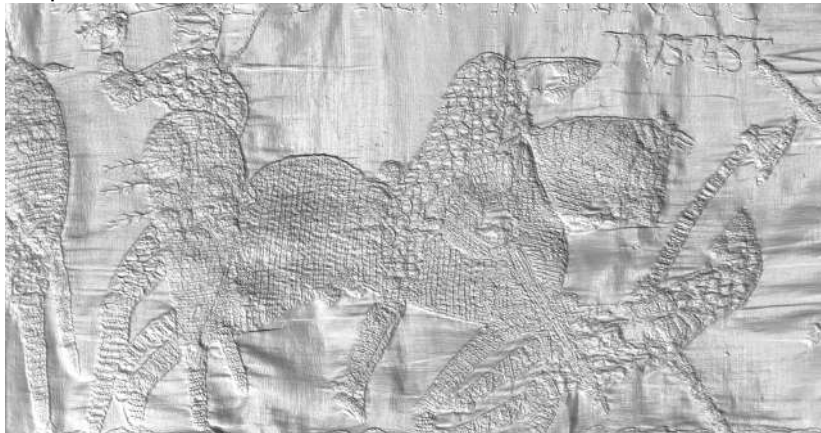
Acquired data and 3D-reconstruction:



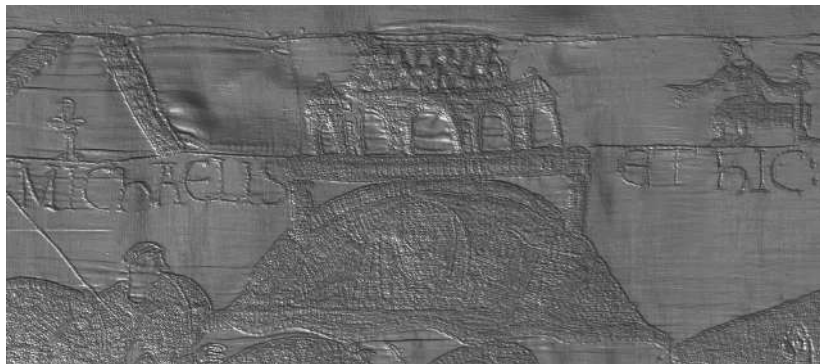


# The death of Harold sequence

Acquired data and 3D-reconstruction:



# The Mont St-Michel sequence





Now ready for 3D-printing :)

## References

Durou, J.-D., Falcone, M., Quéau, Y., and Tozza, S. *Advances in Photometric 3D-Reconstruction*. Springer, 2020.

Haefner, B., Peng, S., Verma, A., Quéau, Y., and Cremers, D. "Photometric Depth Super-Resolution". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* (2020).

Quéau, Y., Wu, T., Durou, J.-D., Lauze, F., and Cremers, D. "A Non-Convex Variational Approach to Photometric Stereo under Inaccurate Lighting". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017.

Pizenberg, M., Quéau, Y., and Elmoataz, A. "Low-rank registration of images captured under unknown, varying lighting". In: *International Conference on Scale Space and Variational Methods in Computer Vision (SSVM)*. 2021.

Thank you for your attention !

ご清聴ありがとうございました