Shape Reconstruction by Photometric Stereo with Unknown Lighting

G. Rodriguez
in collab. with R. Dessì, C. Mannu, G. Stocchino, G. Tanda, M. Vanzi

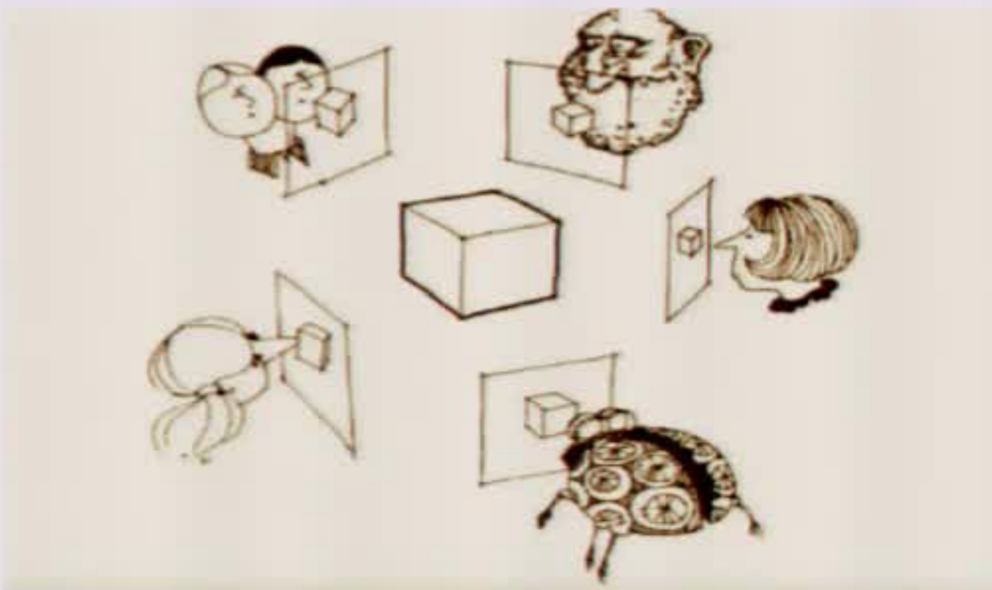
Dipartimento di Matematica e Informatica, Università di Cagliari

SIAM Conference on Applied Linear Algebra 2015
Minisymposium MS14: Recent Advances in Numerical Linear Algebra for Image Processing
Atlanta, U.S.A., October 26–30, 2015

A typical problem in Computer Vision consists of reconstruct the 3D shape of an object, starting from a set of pictures.

There are two main settings:

- Stereo vision or Multivision: a set of pictures is shot as the camera moves around the object; the lighting is static.

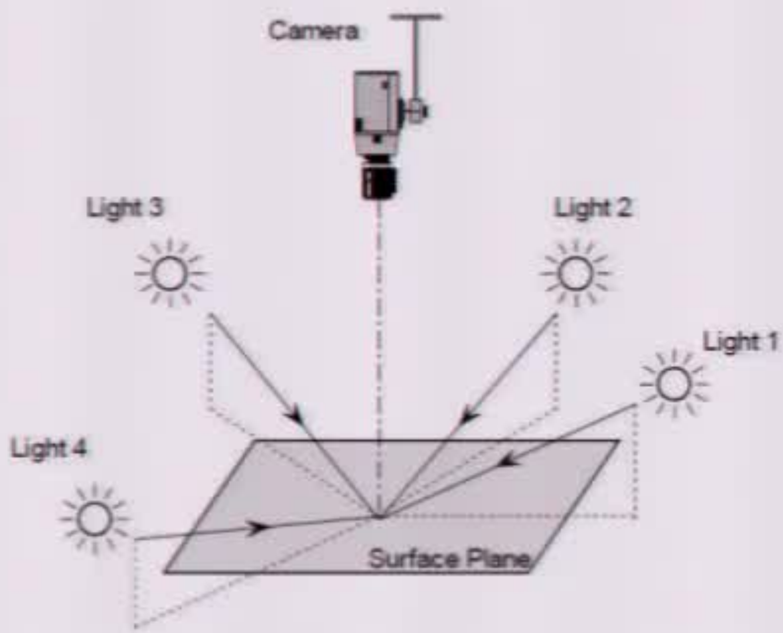


Shape from shading

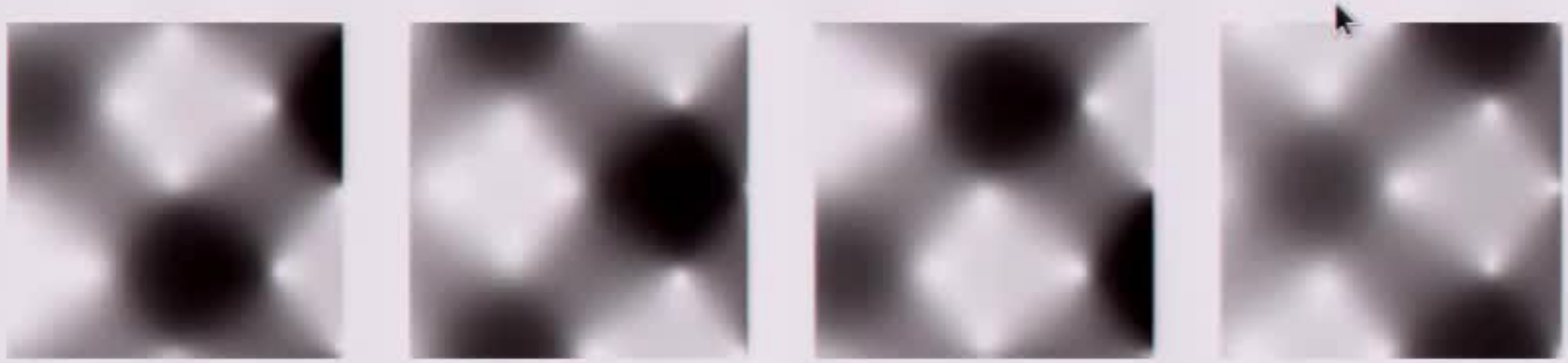
A typical problem in Computer Vision consists of reconstruct the 3D shape of an object, starting from a set of pictures.

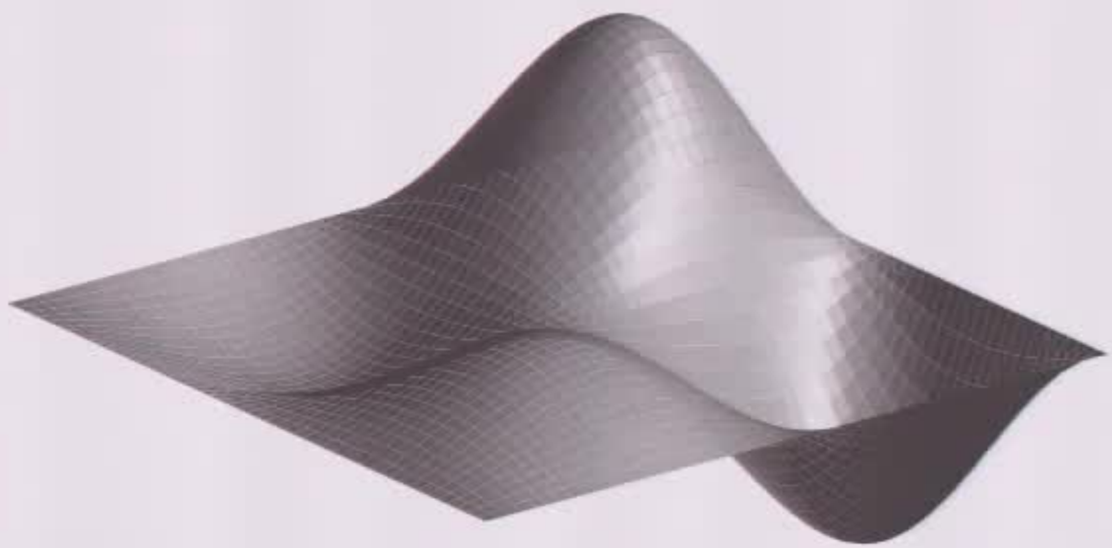
There are two main settings:

- Photometric stereo: the camera and the object are at fixed positions, pictures correspond to different lighting conditions.



Input & Output





Application to Archaeology

Application to Archaeology

Benefits:

- possibility to document any site in full 3D and color by a commercial camera, a tripod, and a hand-positioned flash;
- instrumentation is low cost and easily transportable;
- it allows for simultaneous operation on different findings by a team of researchers;
- data can be processed on site, if fast and reliable algorithms are available;
- these requirement are not met by a laser scanner.

Aim:

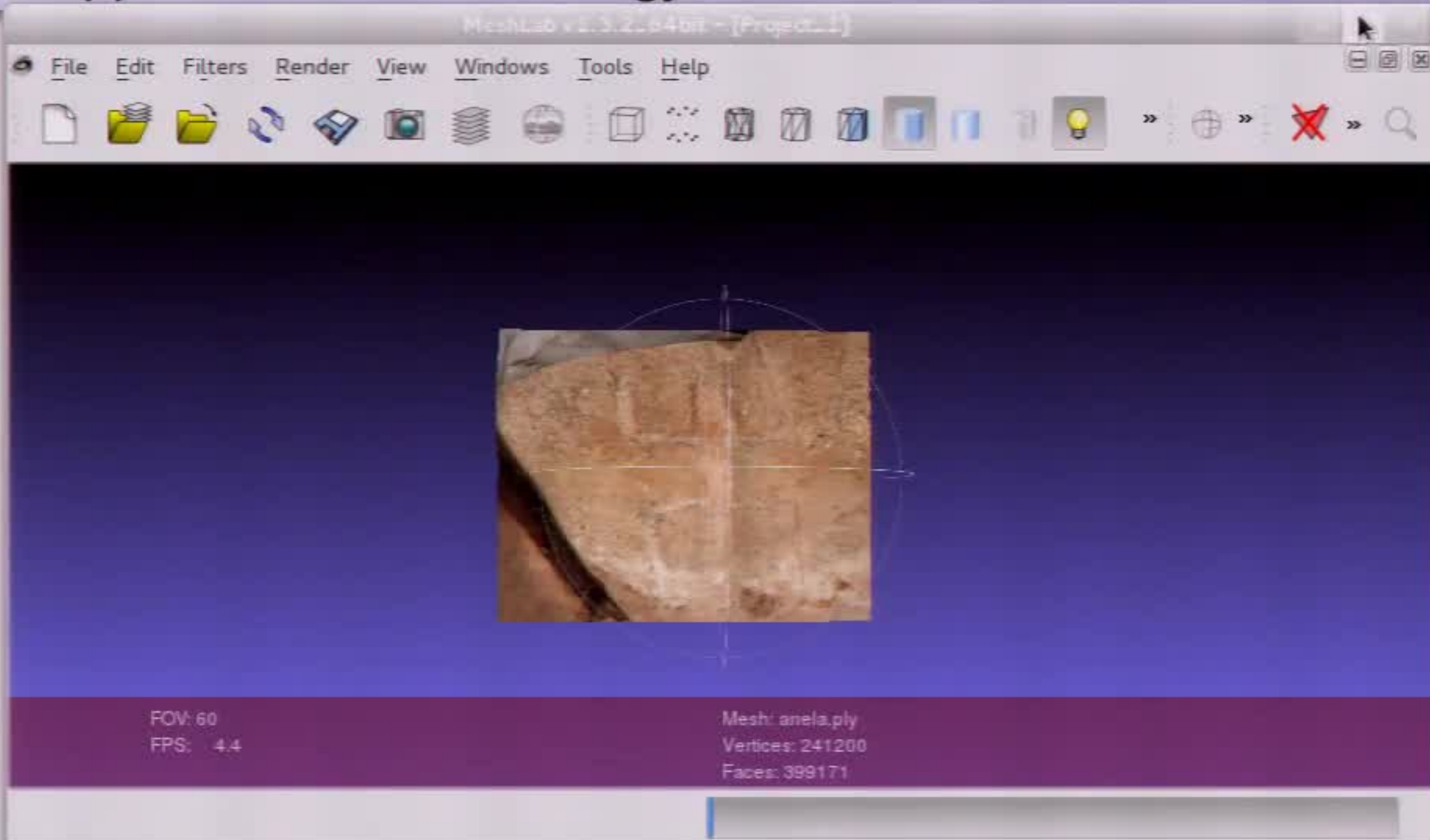
- producing a software capable of real time processing, which does not rely on *hand tuning* of parameters, and whose accuracy has been assessed on reference datasets.

Application to Archaeology

"L'orante" (the praying man)

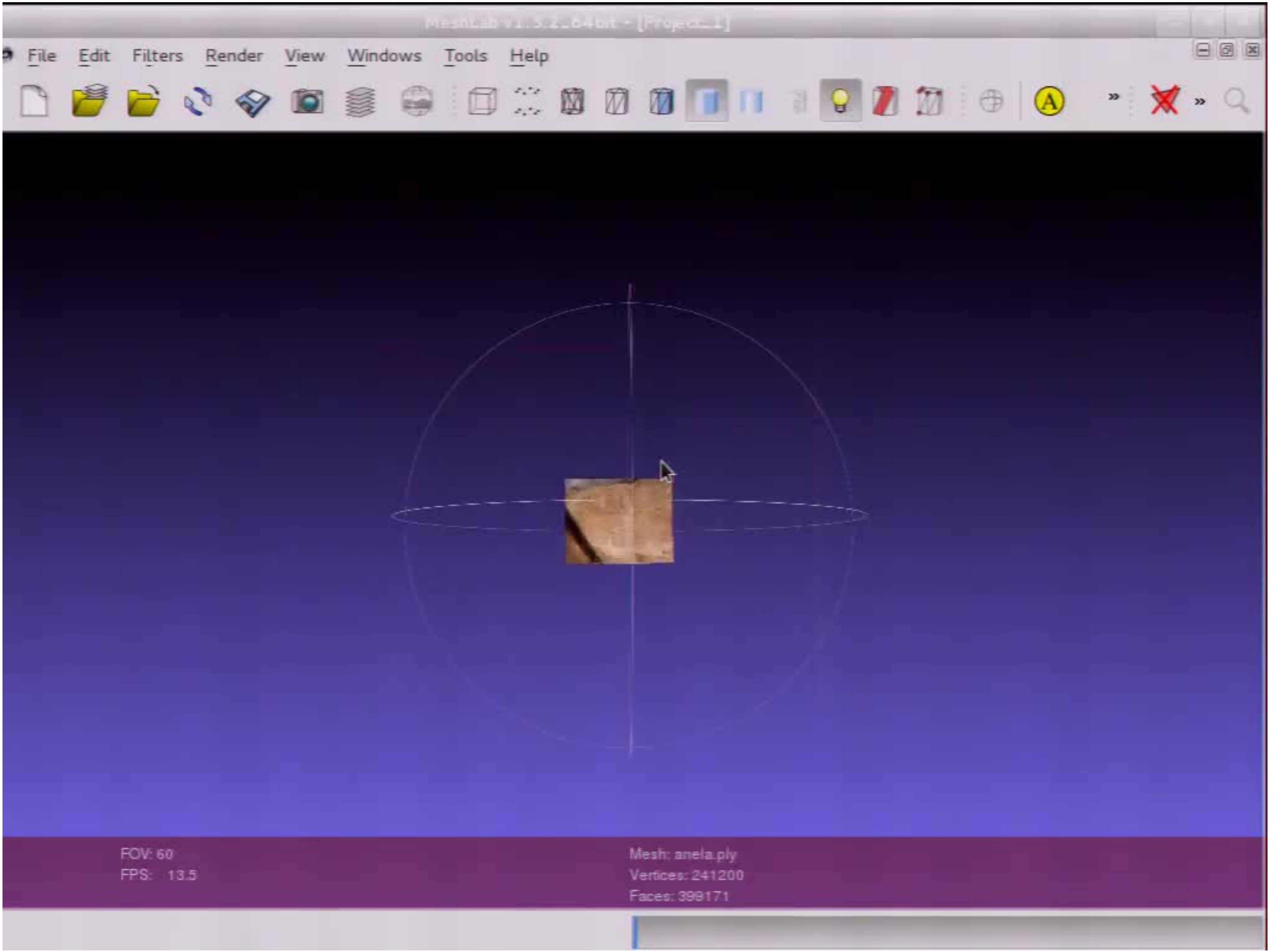
Necropolis (*Domus de Janas*) "Sos Furrighesos", Anela, Sardinia, Italy

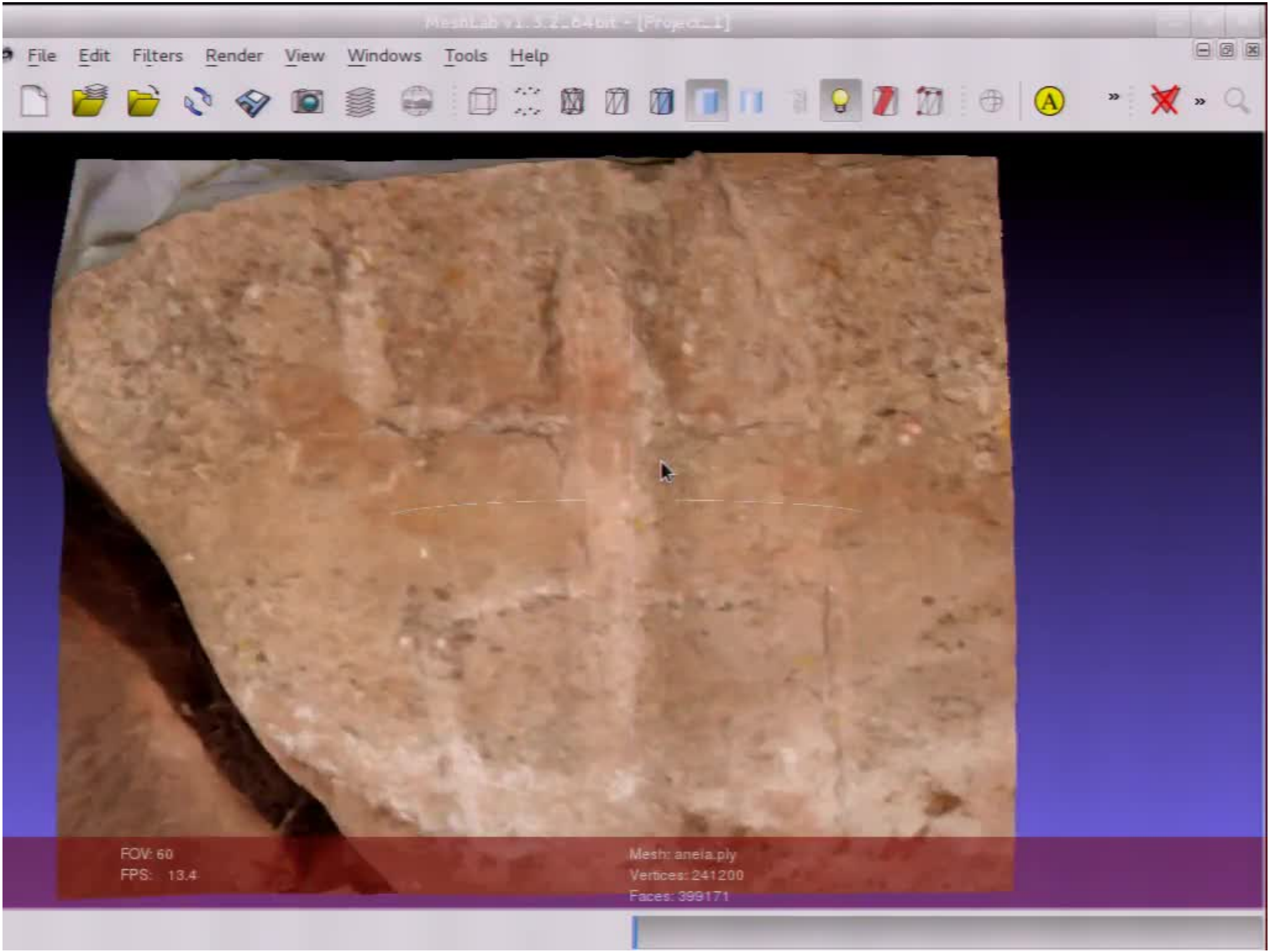
Application to Archaeology

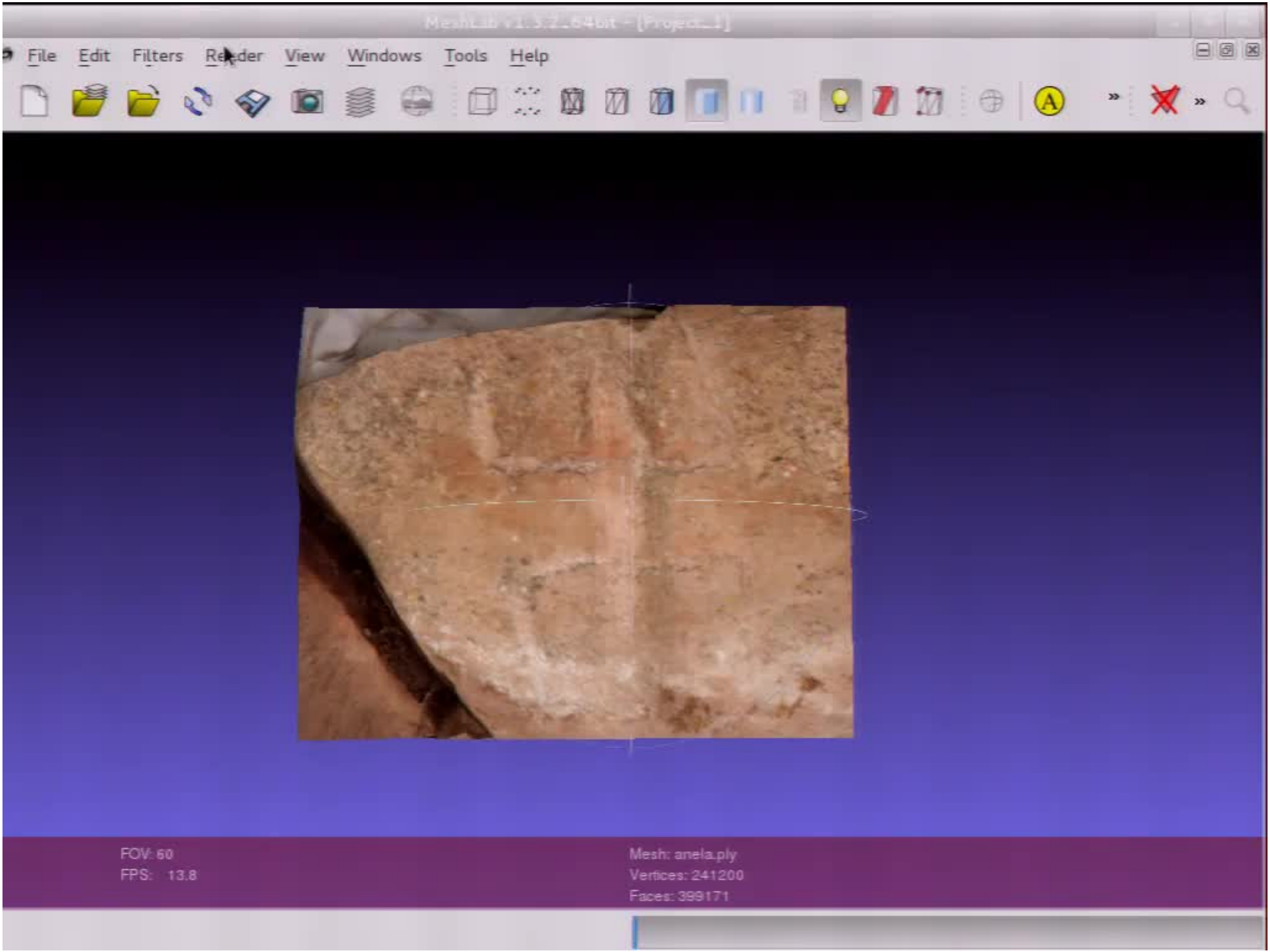


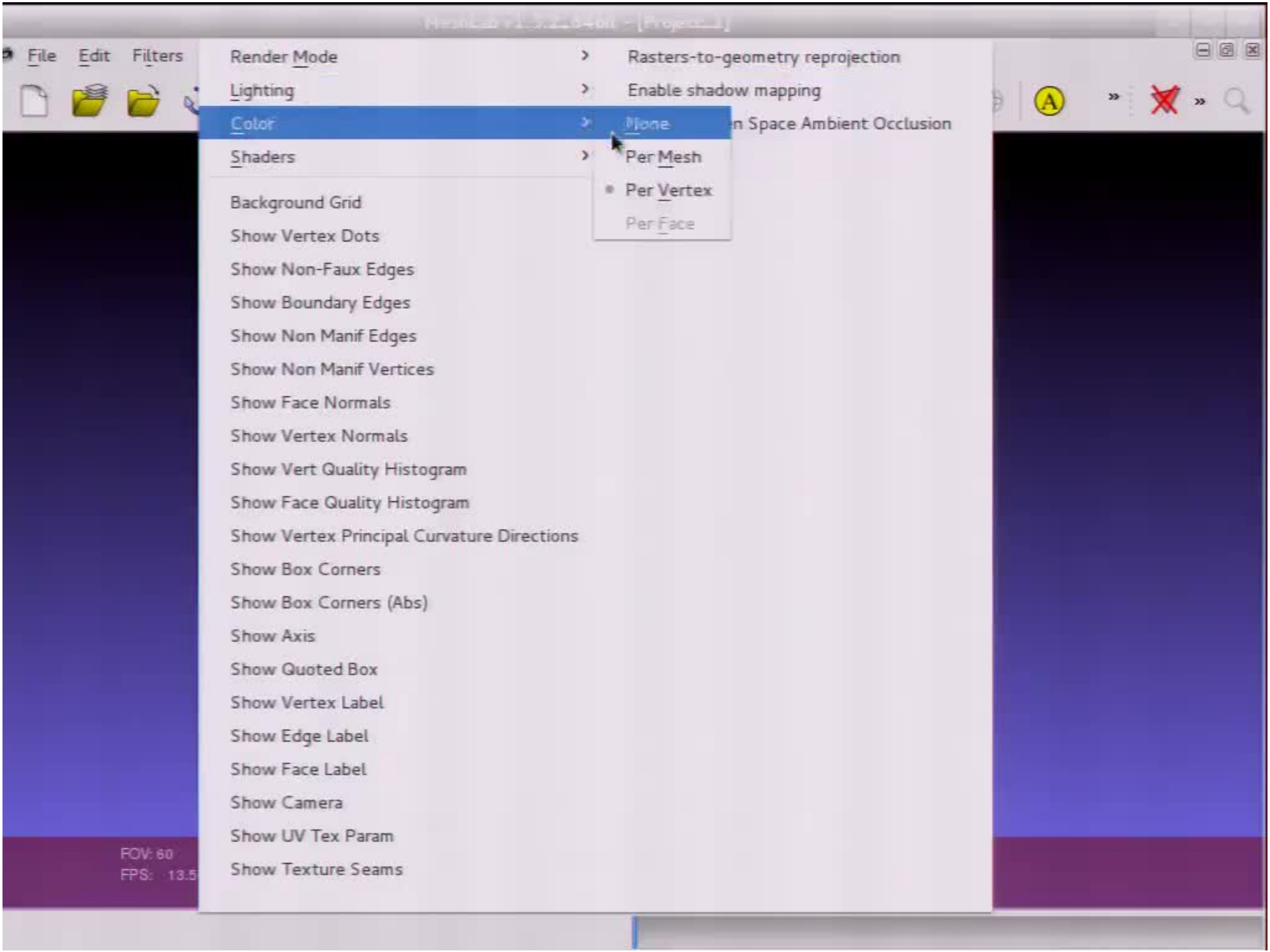
"L'orante" (the praying man)

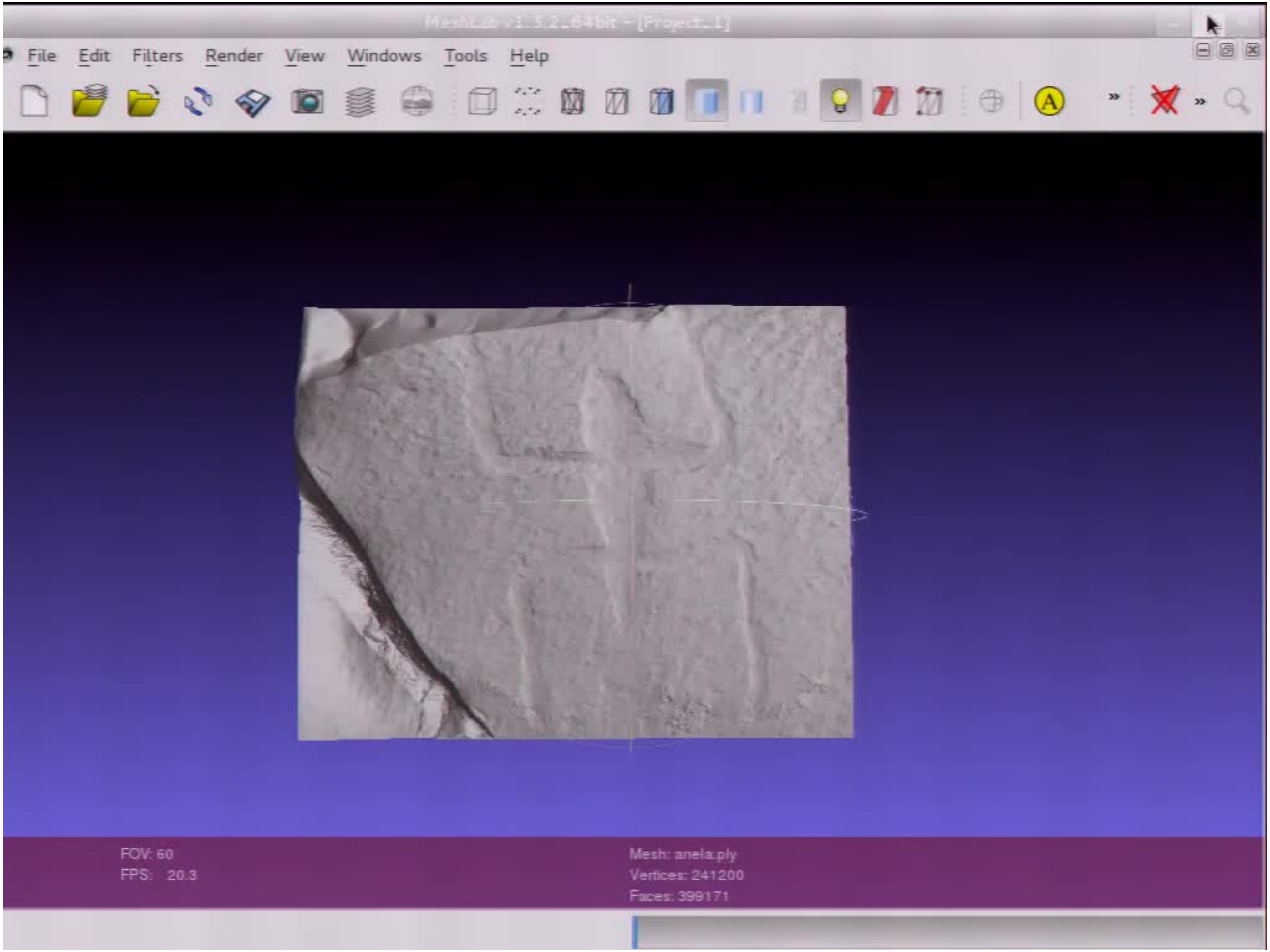
Necropolis (*Domus de Janas*) "Sos Furrighesos", Anela, Sardinia, Italy


File Edit Filters Render View Windows Tools Help
FOV: 60
FPS: 13.5
Mesh: anela.ply
Vertices: 241200
Faces: 399171


File Edit Filters Render View Windows Tools Help
FOV: 60
FPS: 13.4
Mesh: aneia.ply
Vertices: 241200
Faces: 399171


File Edit Filters Render View Windows Tools Help
FOV: 60
FPS: 13.8
Mesh: anela.ply
Vertices: 241200
Faces: 399171


File
Edit
Filters
Render Mode
Lighting
Color
Shaders
Background Grid
Show Vertex Dots
Show Non-Faux Edges
Show Boundary Edges
Show Non Manif Edges
Show Non Manif Vertices
Show Face Normals
Show Vertex Normals
Show Vert Quality Histogram
Show Face Quality Histogram
Show Vertex Principal Curvature Directions
Show Box Corners
Show Box Corners (Abs)
Show Axis
Show Quoted Box
Show Vertex Label
Show Edge Label
Show Face Label
Show Camera
Show UV Tex Param
Show Texture Seams
None
Per Mesh
Per Vertex
Per Face
Rasters-to-geometry reprojection
Enable shadow mapping
n Space Ambient Occlusion
FOV: 60
FPS: 13.5


MeshLab v1.3.2_64bit - [Project_1]
File Edit Filters Render View Windows Tools Help
FOV: 60
FPS: 20.3
Mesh: anela.ply
Vertices: 241200
Faces: 399171

Application to Archaeology

"L'orante" (the praying man)

Necropolis (*Domus de Janas*) "Sos Furrighesos", Anela, Sardinia, Italy

Assumptions

- the surface is Lambertian
- there are no self-obstructions, light reflections, or shades
- the light sources are placed at ∞
- the camera is sufficiently far from the object to avoid perspective distortions

None of these assumptions is perfectly met in practice.

Working assumptions:

- the object is at the origin of a reference system in $\mathbb{R}^3$
- the camera is on the z-axis, aiming at the origin
- q pictures are available, with light sources at $\boldsymbol{\ell}_t$, $t = 1, \ldots, q$
- each digital picture $\mathbf{m}_t$ has resolution $r \times s$, with $p = rs$ pixels
- the pictures are vectorized (pixels in lexicographical ordering)

A nonlinear differential model

Let us assume that the directions $\boldsymbol{\ell}_t$ are known, and write

$$\mathbf{n}(x,y) = \frac{(-u_x, -u_y, 1)^T}{\sqrt{1+\|\nabla u\|^2}}, \qquad \boldsymbol{\ell}_t = \begin{pmatrix} \widetilde{\boldsymbol{\ell}}_t \\ \ell_{3t} \end{pmatrix}.$$

Lambert's law becomes

$$\rho(x,y)\frac{\langle -\nabla u(x,y), \widetilde{\boldsymbol{\ell}}_t\rangle + \ell_{3t}}{\sqrt{1+\|\nabla u(x,y)\|^2}} = I_t(x,y), \qquad t = 1,\ldots,q,$$

i.e., a system of first order nonlinear PDEs of Hamilton–Jacobi type

$$\begin{cases} H_t(x,y,\nabla u(x,y)) = 0, & t = 1,\ldots,q, \\ u(x,y) = g(x,y), & (x,y) \in \partial\Omega. \end{cases}$$

Linearization of the differential model

Following [Mecca, Falcone, SJIS 2013] we substitute

$$\sqrt{1+|\nabla u(x,y)|^2} = \rho(x,y)\frac{\langle -\nabla u(x,y), \widetilde{\boldsymbol{\ell}}_1\rangle + \ell_{31}}{I_1(x,y)}$$

in the equations for $t = 2, \dots, q$, to obtain

$$\begin{aligned}[][\ell_{11} I_t(x,y) - \ell_{1t} I_1(x,y)]\, u_x + [\ell_{21} I_t(x,y) - \ell_{2t} I_1(x,y)]\, u_y \\ = [\ell_{31} I_t(x,y) - \ell_{3t} I_1(x,y)]\,.\end{aligned}$$

This shows that the minimal number of images is 2.

After $u(x,y)$ is computed, the albedo is given by

$$\rho(x,y) = \frac{I_t(x,y)}{\langle \mathbf{n}(x,y), \boldsymbol{\ell}_t\rangle}, \qquad \text{for any } t.$$

Well posedness and practical computation

- conditions for the existence of solutions are discussed in [Kozera, Math. Appl. Comput. 1991];
- [Mecca, Falcone, SJIS 2013] studied the problem under more realistic assumptions;
- the obvious choice for discretization is finite differences;
- the solution does not exist when a portion of the surface is shaded in all pictures;
- the solution may not exist for particular light orientations, when $q = 2$;
- taking $q > 2$ is a good choice for getting a solution and for noise reduction, as it does not make data acquisition significantly harder;
- knowing accurately the light positions ℓ_t is a strong requirement.

Poisson formulation

Since p (number of pixels) is large, to compute R and N from

$$RN^T L = M$$

it is required that $q \geq 3$ and that the ℓ_t vectors are independent.

In this case, the minimal number of images is 3.

Numerically differentiating the first two components of $\mathbf{n}_k$, $k = 1, \ldots, p$, we get an approximation of the Laplacian of $u(x, y)$ on each point of the discretization.

Solving the Poisson equation

$$\Delta u(x, y) = f(x, y),$$

yields the solution [Dessì, Mannu, R, Tanda, Vanzi, DAACH 2015].

Photometric stereo under unknown lighting

The need for accurate information about the relative position of the lights and the object is a severe limitation of the method.

Being able to obtain the lights position from the available set of images opens the possibility of freehand lighting, removing the requirement for accurate positioning of lamps, one of the most difficult issues in practical PS.

Some papers conjecture that the problem can be uniquely solved by 4 images (authors often refer to *4-source photometric stereo*).

[Basri, Jacobs, Int. J. Comput. Vis 2007] propose an approach based on the decomposition of the light intensity, as a function of direction, into linear combinations of spherical harmonics.

The solution is not unique!

From the mathematical point of view, the problem consists of determining the factorization

$$\widetilde{N}^T L = M,$$

where $\widetilde{N} = NR$ is $3 \times p$, L is $3 \times q$ and M is $p \times q$ ($q \geq 3$).

There are infinite solutions: if $(\widetilde{N}, L)$ is a solution, any matrix pair $(A^{-T}\widetilde{N}, AL)$, with $A \in \mathbb{R}^{3\times 3}$ nonsingular, gives another solution.

The solution is not unique!

From the mathematical point of view, the problem consists of determining the factorization

$$\widetilde{N}^T L = M,$$

where $\widetilde{N} = NR$ is $3 \times p$, L is $3 \times q$ and M is $p \times q$ ($q \geq 3$).

There are infinite solutions: if $(\widetilde{N}, L)$ is a solution, any matrix pair $(A^{-T}\widetilde{N}, AL)$, with $A \in \mathbb{R}^{3\times 3}$ nonsingular, gives another solution.

Lemma

The matrices R, N, and L are determined up to a unitary transf., i.e., $(Q\widetilde{N}, QL)$ is a solution for any orthogonal matrix $Q \in \mathbb{R}^{3\times 3}$.

The solution is not unique!

From the mathematical point of view, the problem consists of determining the factorization

$$\widetilde{N}^T L = M,$$

where $\widetilde{N} = NR$ is $3 \times p$, L is $3 \times q$ and M is $p \times q$ $(q \geq 3)$.

Lemma

The matrices R, N, and L are determined up to a unitary transf., i.e., $(Q\widetilde{N}, QL)$ is a solution for any orthogonal matrix $Q \in \mathbb{R}^{3\times 3}$.

- the system object-camera-lights may result rotated in the reconstruction, perhaps with axes inversions
- this constitutes a problem if the object is represented by $z = u(x, y)$

Scheme of the algorithm

1. find the light sources directions, possibly in a rotated system
2. "straighten" the coordinates system
3. integrate the normal field

Some assumptions on the shooting technique are needed.

Determining the lights position

Let the "compact" singular value decomposition of the observations matrix be

$$M = U \Sigma V^T,$$

with $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_q)$, $U \in \mathbb{R}^{p \times q}$, and $V \in \mathbb{R}^{q \times q}$.

In our case, $3 \leq q \ll p$, so standard `svd` may be sufficient. It can be computed by a Lanczos approach [Baglama, Reichel, BIT 2013].

Since it is expected that $\text{rank}(M) = 3$, we set $W = [\sigma_1 \mathbf{u}_1, \sigma_2 \mathbf{u}_2, \sigma_3 \mathbf{u}_3]^T$ and $Z = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]^T$, so that

$$W^T Z \simeq M.$$

This produces the best rank-3 approximation to the data matrix M with respect to the Euclidean norm.

Determining the lights position

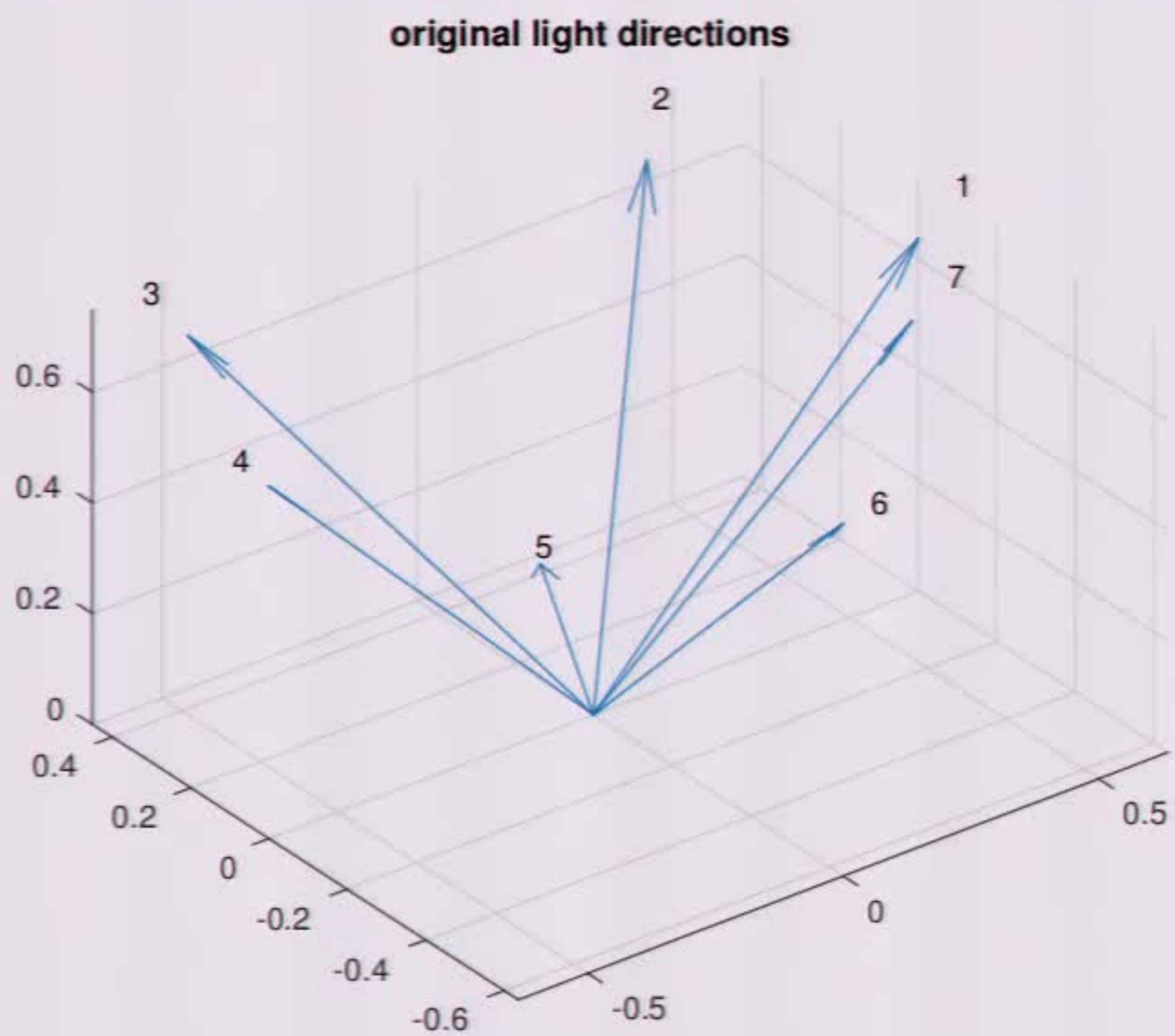
We observed that the initial factorization $W^T Z = M$ gives a good approximation of normal vectors and light directions in a particular situation, that is, when the light vectors span equal angles.

Exploiting the fact that the norms $\|\boldsymbol{\ell}_t\|$ are proportional to the light intensities, which has to be known up to a proportionality constant, it is possible to prove the following result.

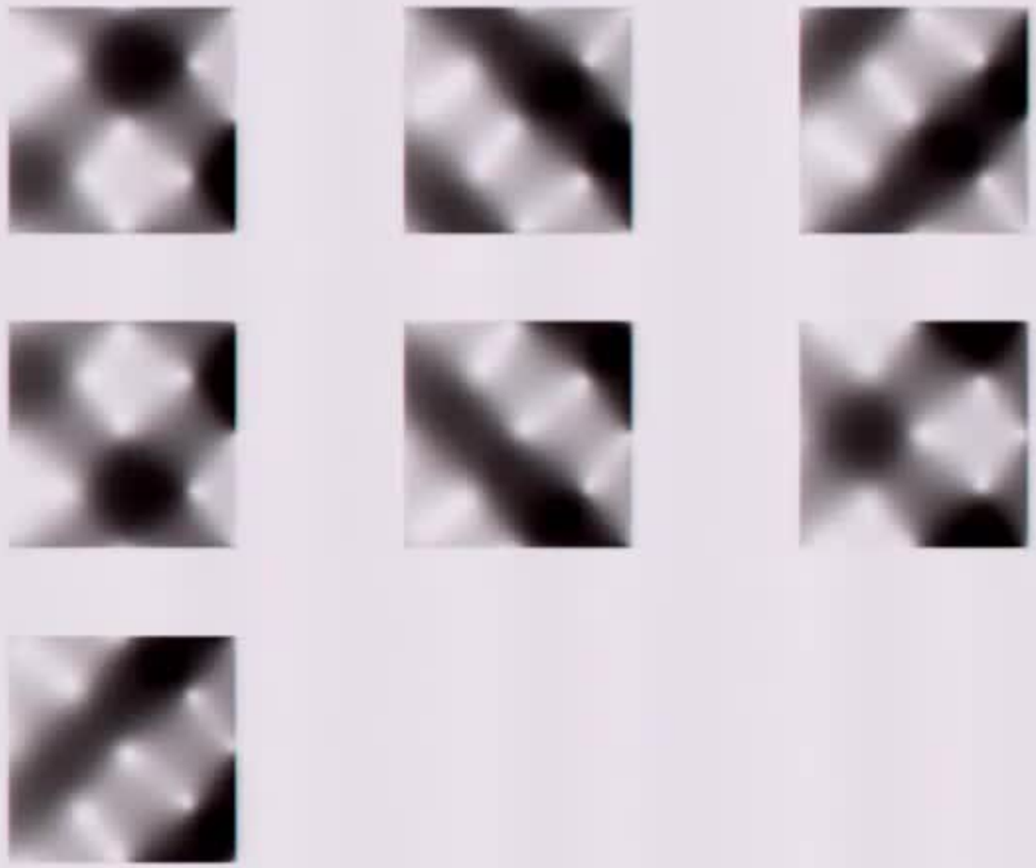
Theorem

The normal vectors and the lights position can be uniquely determined from $RN^T L = M$ up to a unitary transformation, only if at least 6 images taken in different lighting conditions are available.

A numerical experiment

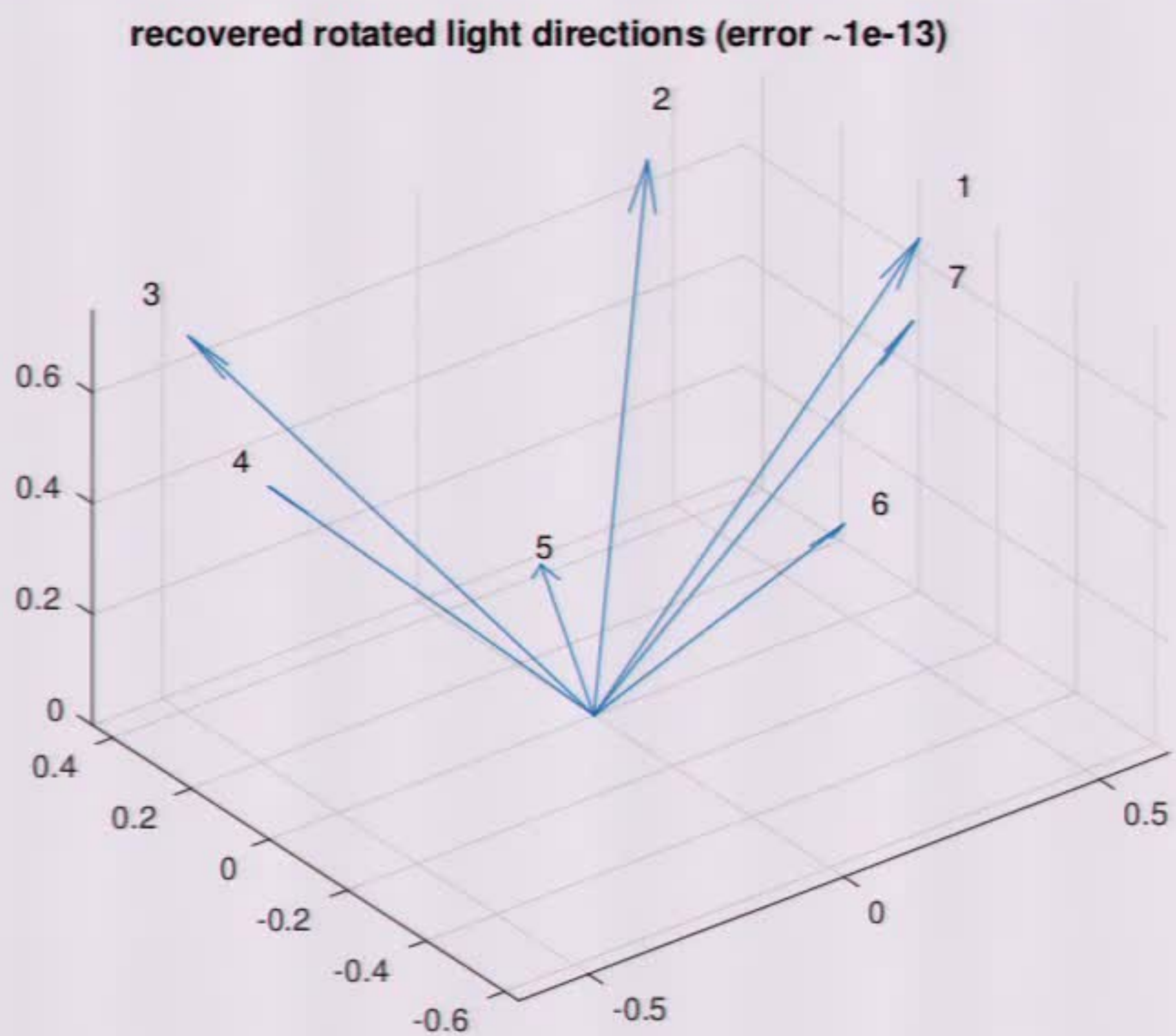


A numerical experiment

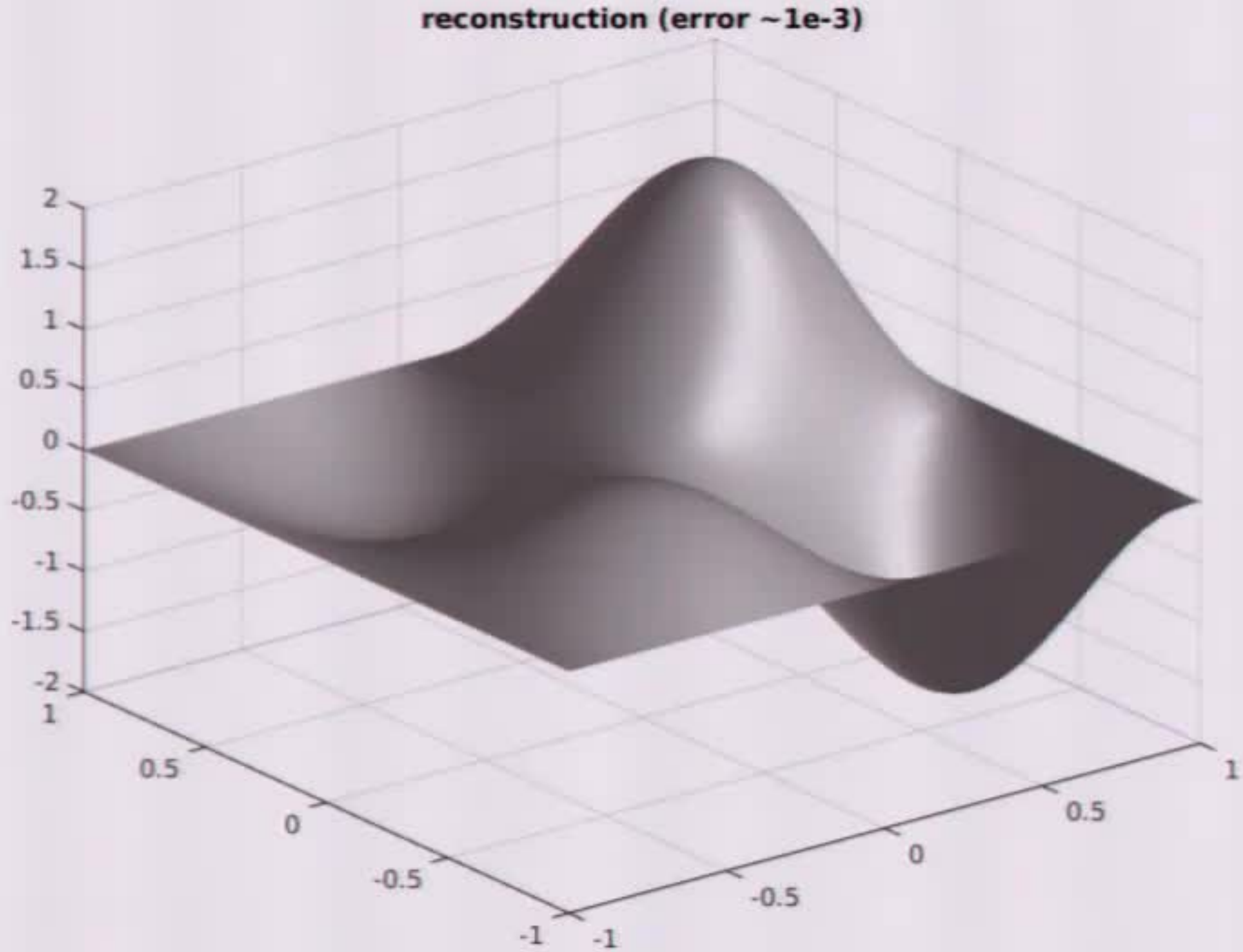


data set (100×100 pixels)

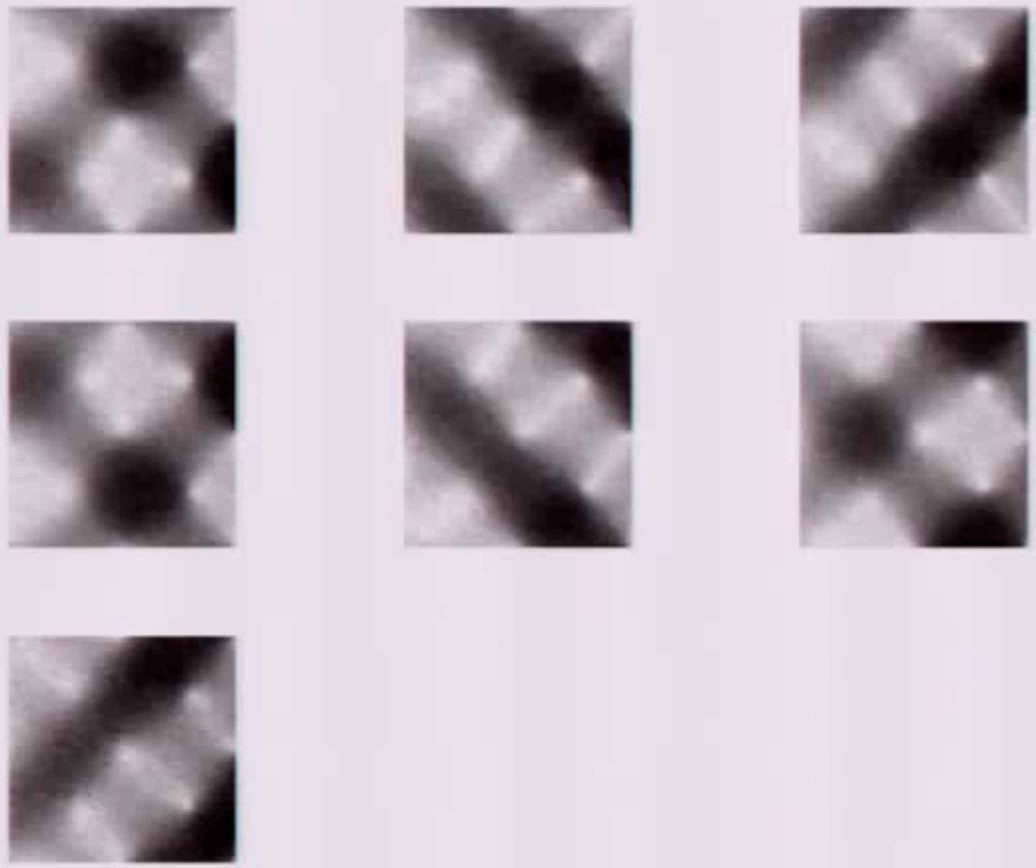
A numerical experiment



A numerical experiment

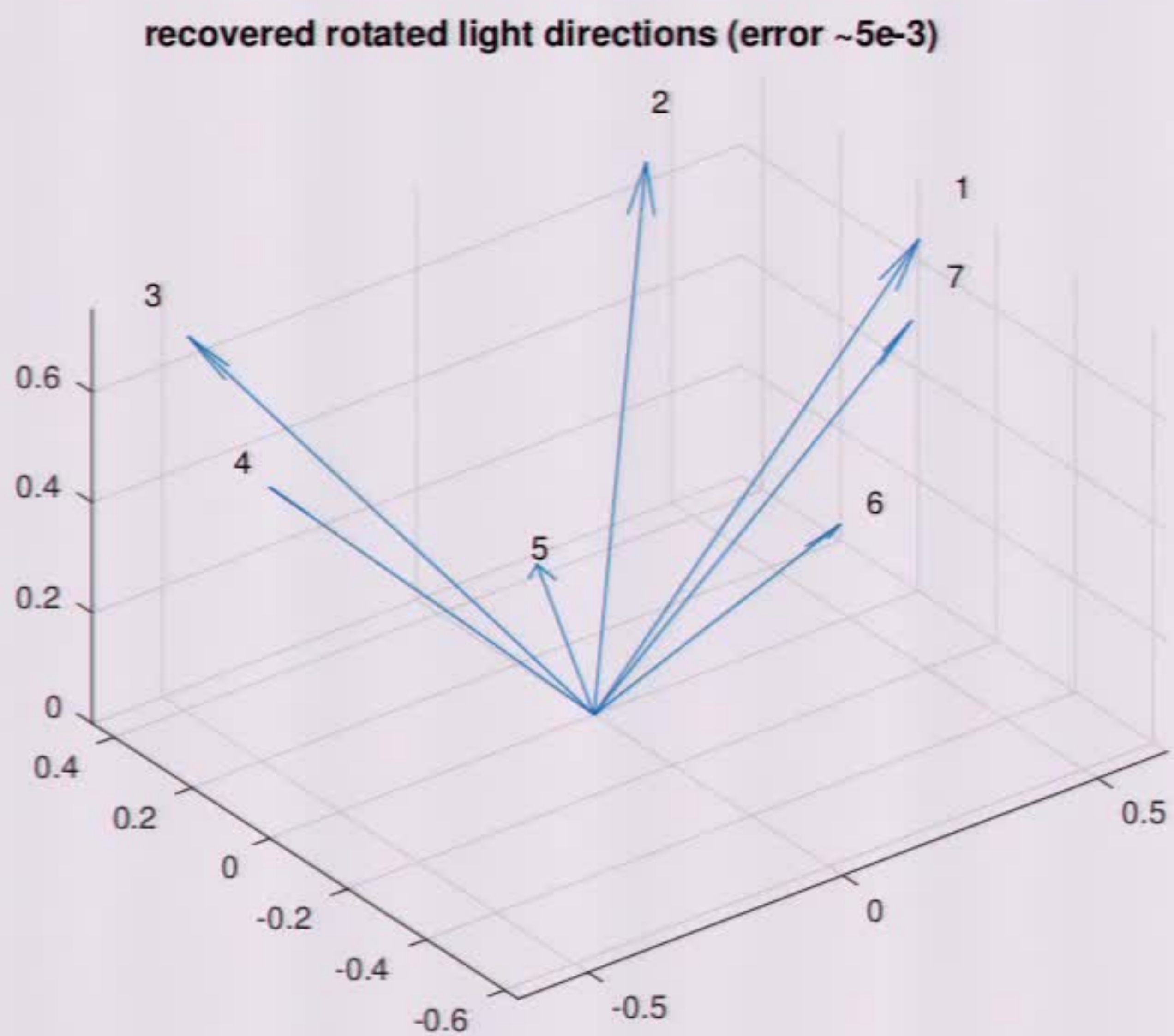


A numerical experiment with 10% white noise on M

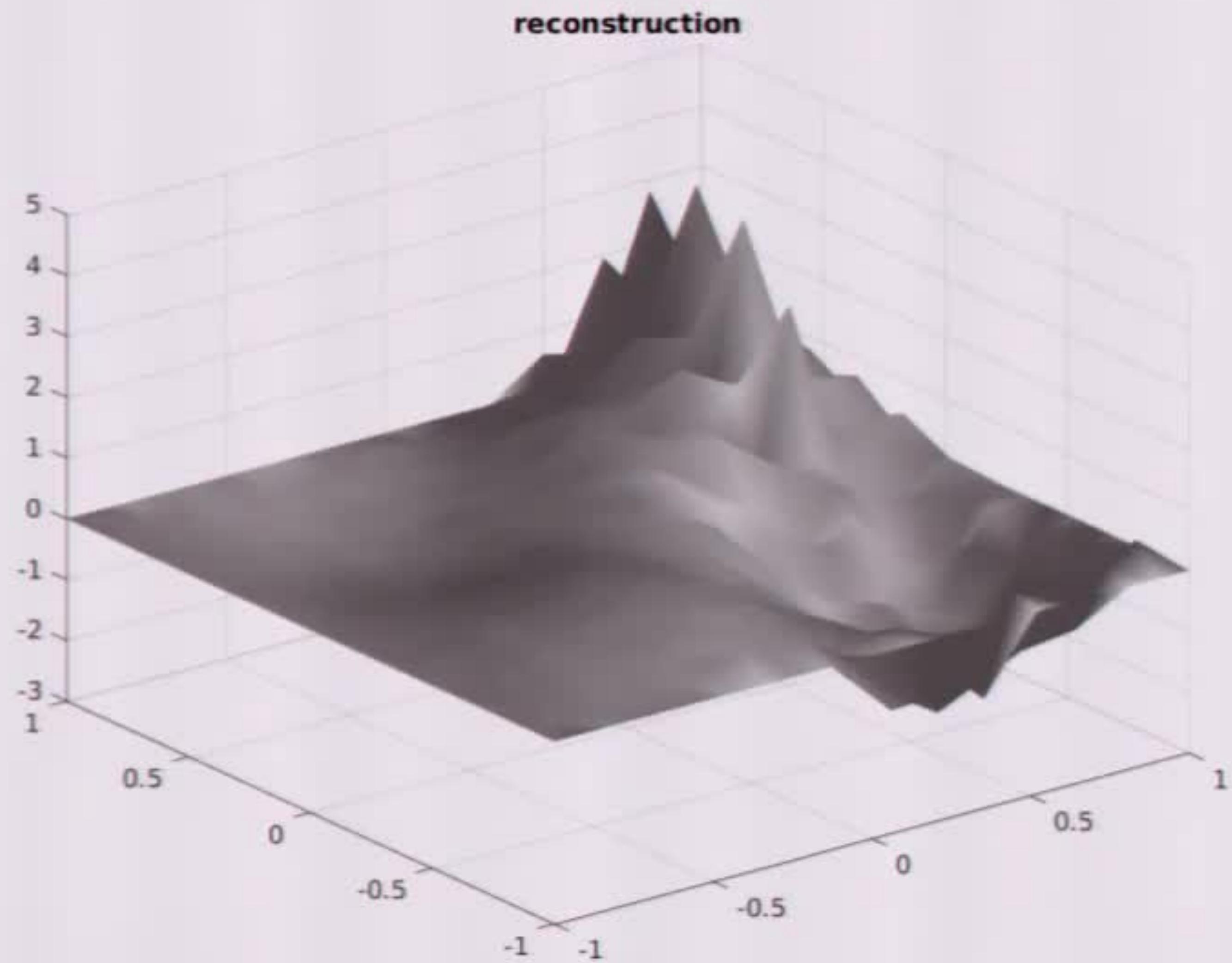


data set with 10% relative noise (100×100 pixels)

A numerical experiment with 10% white noise on M



Solution by H–J with 10% white noise on M



Working in unideal light conditions

Real light sources are often far from being ideal

1. they may be placed close to the surface (especially in narrow locations, like caves or excavation sites)
2. they are differently attenuated at different points of the object, according to distance

A differential model which keeps into account these effects has been studied in [Mecca et al., SJIS 2014].

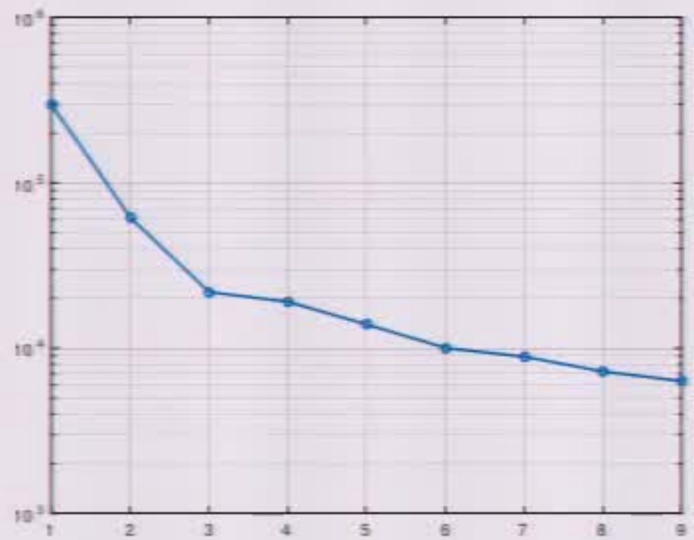
The first problem, causes a spherical deformation in the reconstructed surface, which can be partially "cured" by the SVD, if the light directions are known and the standard model is used.

Working in unideal light conditions

reconstructed surface Z

flattened surface $Z - U_1\Sigma_1 V_1^T$

The *scholar* data set (close lights with different intensities)

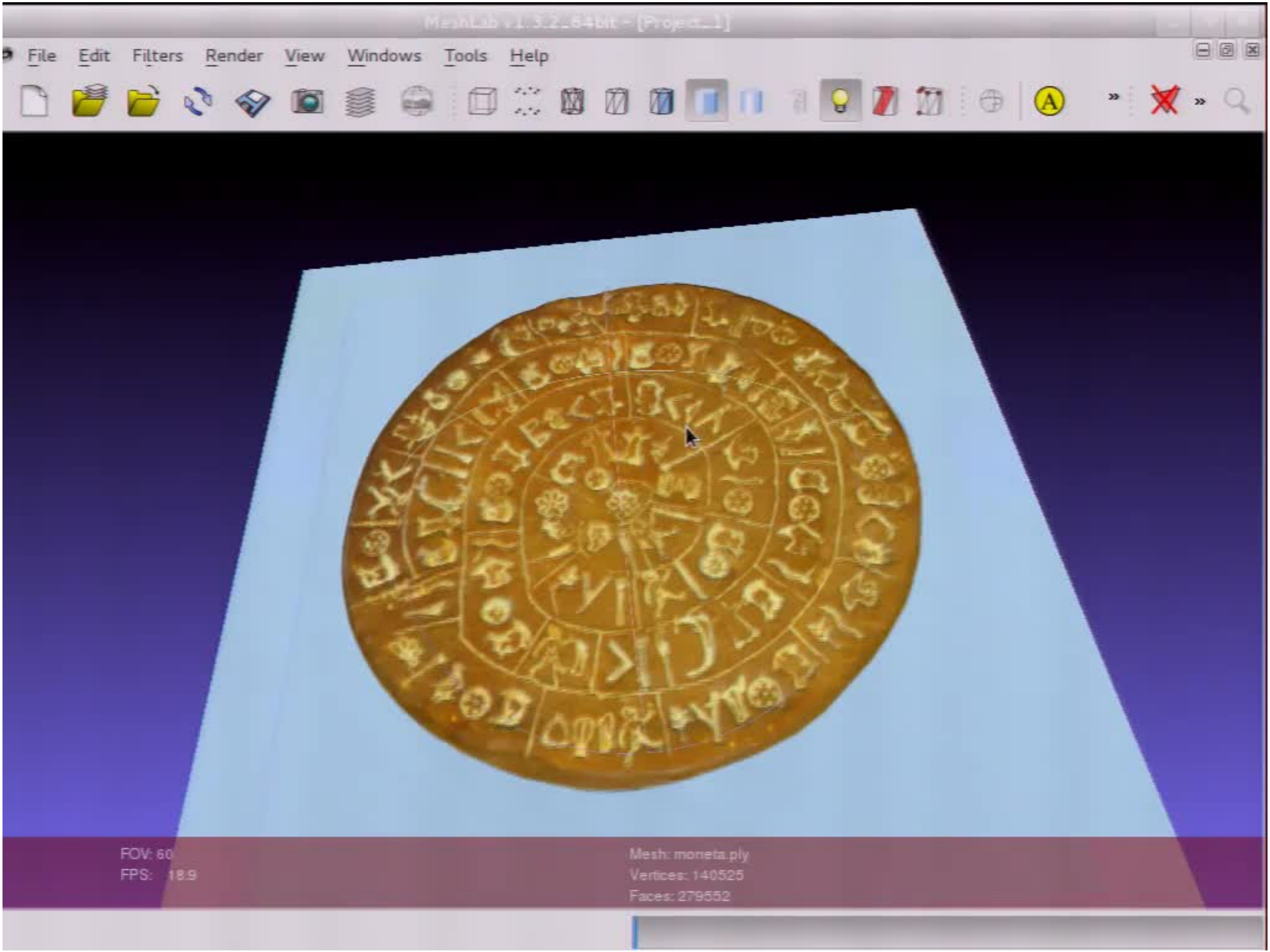


singular values of M

The next steps

- Develop a model for lights at finite distance
- Identify the lights position in the new model

Thank you for your attention!


MeshLab v1.3.2_64bit - [Project_1]
File Edit Filters Render View Windows Tools Help
FOV: 60
FPS: 18.9
Mesh: moneta.ply
Vertices: 140525
Faces: 279552