Image Processing and Computer Graphics

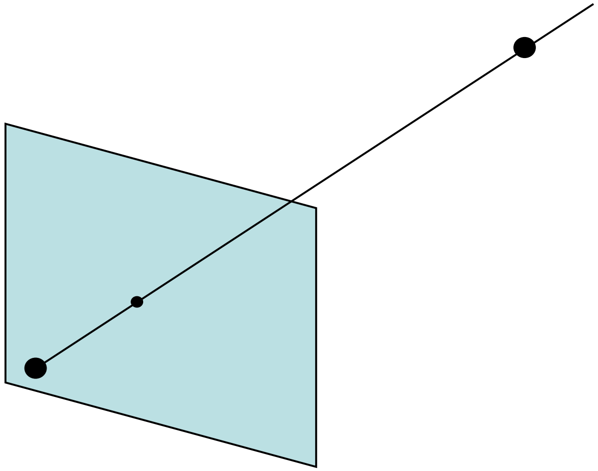
Image Processing

Class 9

Shape from X

Depth reconstruction

- Another important goal in computer vision is to reconstruct the 3D structure of a scene.
- The missing cue is the depth of an observed point. This depth information has been lost by the projection of the scene to the image plane.
- Any point along the projection ray that passes the camera center and the image point could have caused the observation.



- Thus from the position of a single image point we cannot determine the coordinates of the corresponding 3D point.
- We need additional information either from further images or from a-priori knowledge about the scene.

Shape from X

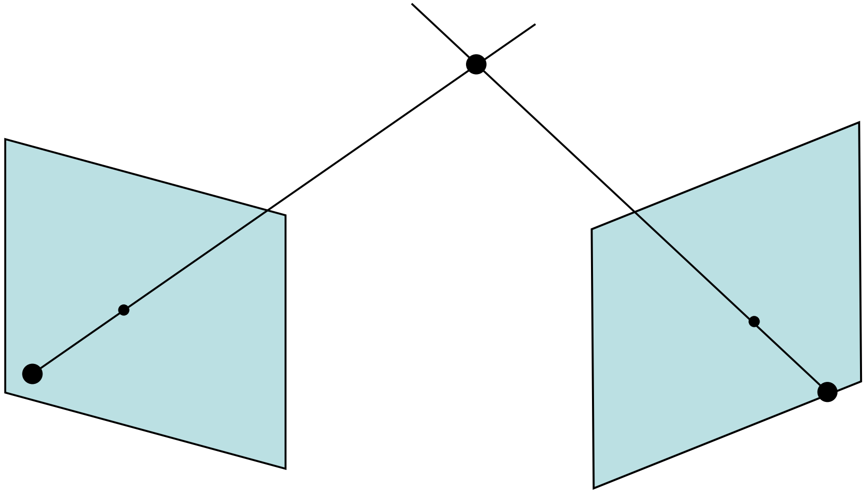
- There are many ways to reconstruct the depth of a surface point. Often these techniques are called **shape from X**, where X stands for the cue that helps reconstruct the depth.

- Some of these cues are rich of information others not so much.

- Most prominent is the reconstruction of depth from binocular images.

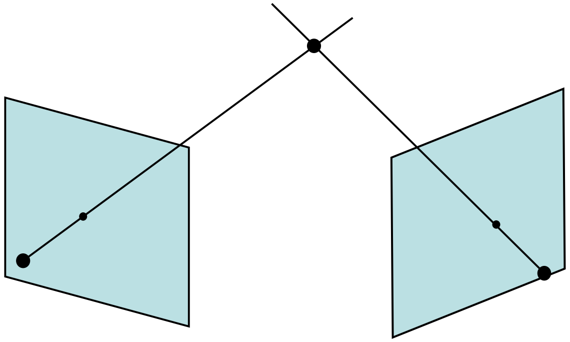
- Knowing the corresponding image point in the second image and the camera center of the second camera yields a unique 3D point.

- The camera center of the second camera <u>must not</u> coincide with the one of the first camera.



Stereo reconstruction

- For reconstructing the 3D point, we need the two corresponding image points. Finding this pair is the key challenge in stereo reconstruction → **disparity estimation**
- With the corresponding points known, we must reconstruct the two projection rays. The cross section tells us the sought 3D point.
- For reconstructing the projection rays, we need to know the projection properties of the cameras.
- These properties are comprised in the projection matrix $\mathbf{P}$
- Finding $\mathbf{P}$ is called **camera calibration**.



Multiview reconstruction

- We can add further cameras to reconstruct surface points in the scene.

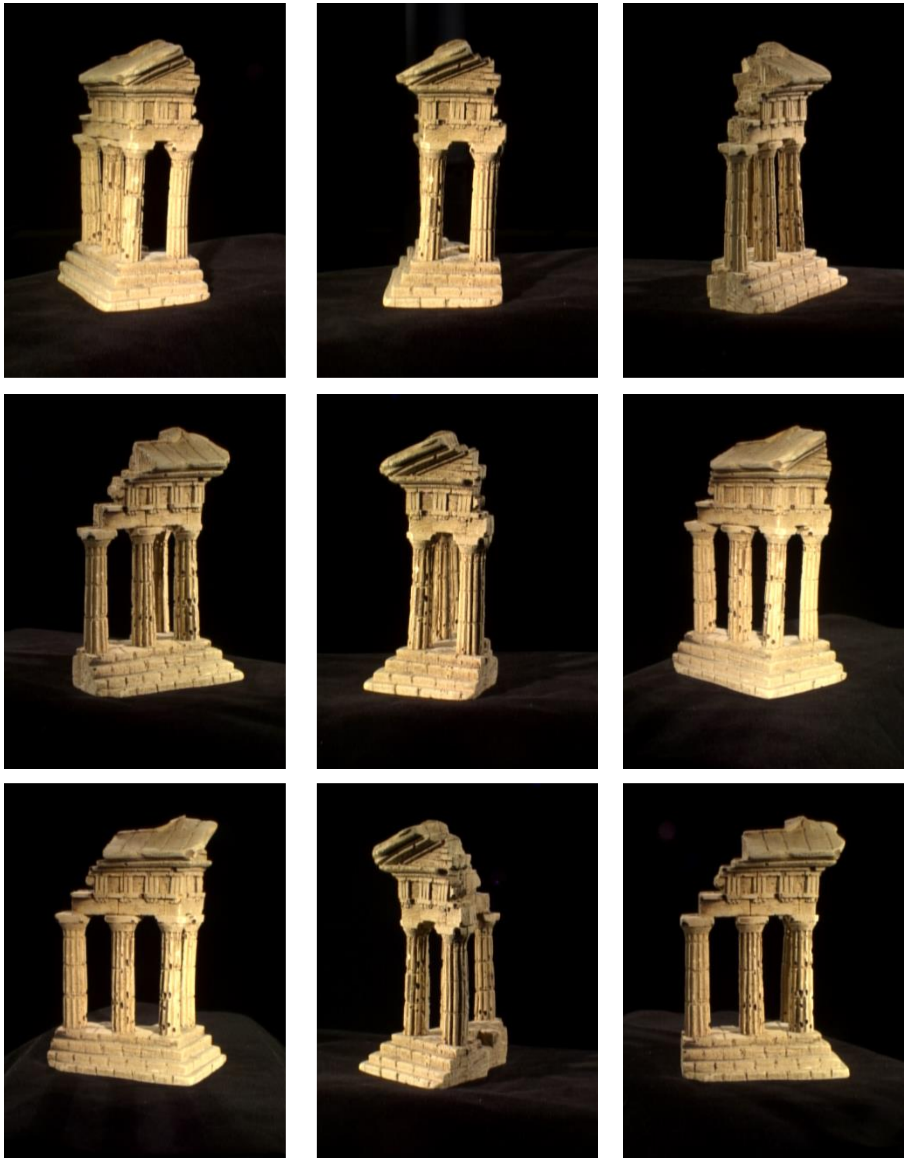
- Geometrically, only two cameras are necessary to reconstruct the 3D point.

- In practice, we have several challenges:
  - We must **find corresponding points**. Correspondences must be expected to be noisy and some can be even completely wrong.
  - For **occluded points** we cannot establish a correspondence.
  - The **accuracy** of the correspondences is limited by the pixel grid.

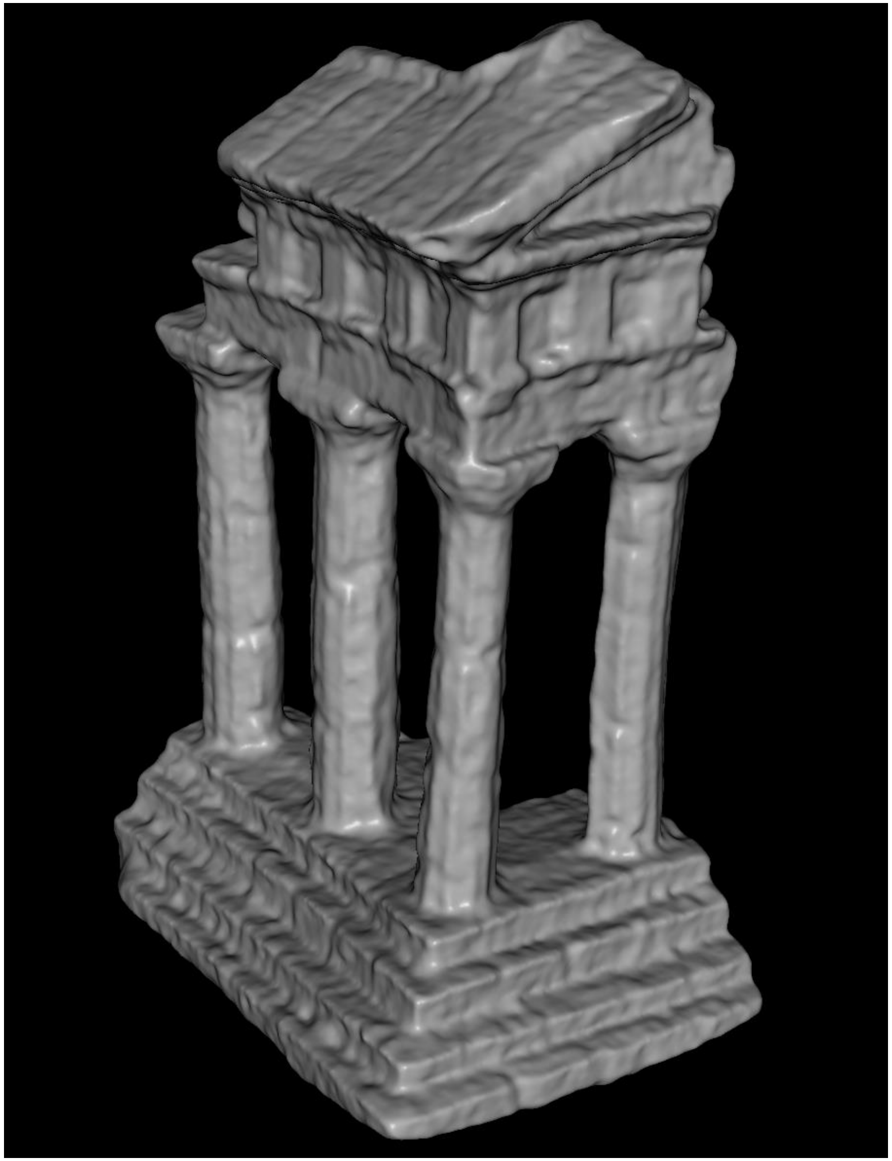
- With more cameras, part of the noise in the correspondences is averaged out and the total accuracy of the depth estimates is increased.

- Drawback: multiple cameras are more difficult to calibrate.

Multi-view reconstruction example



Multiple views



Reconstructed surface

Structure from motion

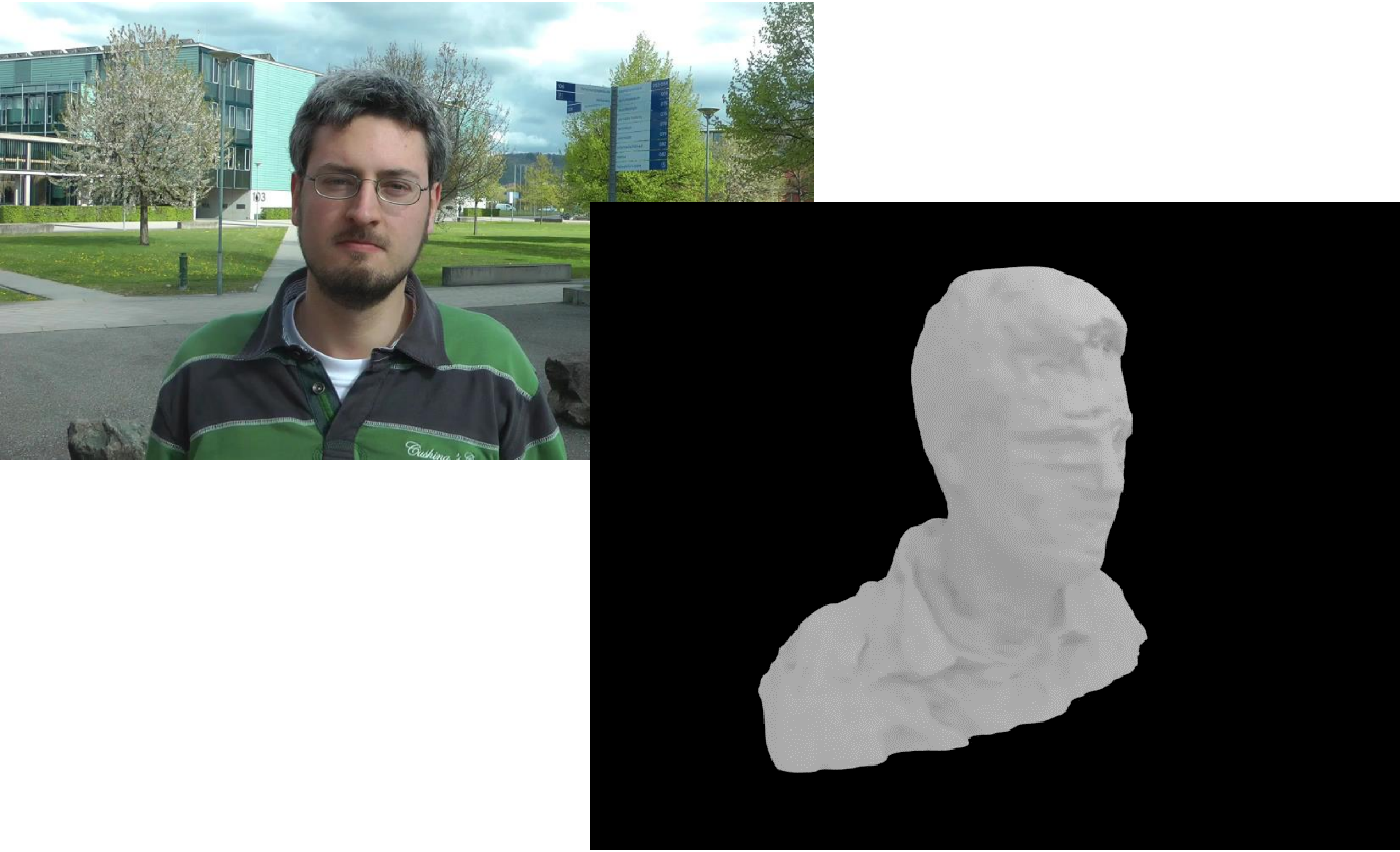
- Instead of multiple static cameras that simultaneously observe a 3D point, we can also consider a single moving camera that observes a static 3D point.

- Advantages:
  - We obtain plenty of images from a single camera.
  - With a reasonable frame rate the displacements in successive views are small → easier correspondence search and smaller occlusion areas

- Problems:
  - We cannot use a fixed, calibrated camera setup. We must estimate the camera motion (**egomotion**), <u>together</u> with the structure.
  - The scene must be <u>static</u>. Otherwise we get ambiguities between the depth, the camera motion, and the motion in the scene.

- Sometimes the camera motion is approximately known (due to some other sensors, e.g, in robotics). This can simplify the task.

Structure from motion example



Decomposition of the projection matrix

- For structure from motion, we must factorize the projection matrix:

$$\mathbf{P} = \mathbf{K}\mathbf{M}$$

- The first part is the **camera calibration matrix**

$$\mathbf{K} = \begin{pmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{pmatrix}$$

  which contains the camera’s **internal parameters**.

- The second part is the pose of the camera relative to a world coordinate system

$$\mathbf{M} = (\mathbf{R}|\mathbf{t})$$

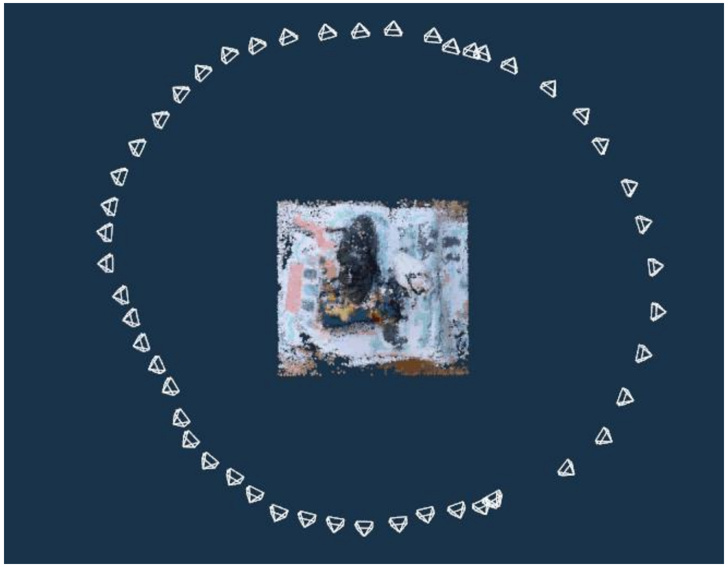
  consisting of a rotation matrix and a translation vector, each having 3 degrees of freedom. These are the **external parameters**.

Structure from motion: how it works

- First the camera’s **internal parameters** are calibrated (details in Computer Vision)
- We then only need to calibrate the **external parameters** online. This is the translation and rotation of the camera. This egomotion can be estimated from point correspondences (details in Computer Vision)
- At this point it is assumed that the internal parameters stay fixed.
- In case of autofocus, this is no longer the case. Also some of the internal parameters can be estimated online (**autocalibration**), but this decreases the robustness of the estimation.
- With the camera being fully calibrated (we know $\mathbf{P} = \mathbf{KM}$), we can estimate the 3D structure of the scene from point correspondences.

Structure from motion: loop closing

- In robotics, a related problem is known as **self-localization and mapping (SLAM)** → leads to the same type of optimization problem
- Localization errors accumulate over time.
- When the camera returns to an earlier position and if localization errors are small enough, the image can be matched to the earlier image. This is called **loop closing**.
- This matching is a special case of object recognition (instance recognition)
- Loop closing is the only way to correct the accumulated errors (**drift**).

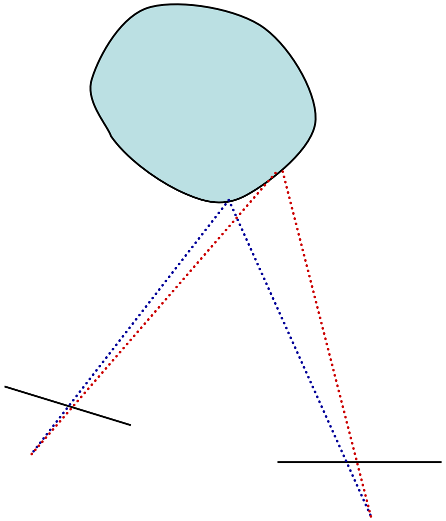


Iterative correction after loop closing

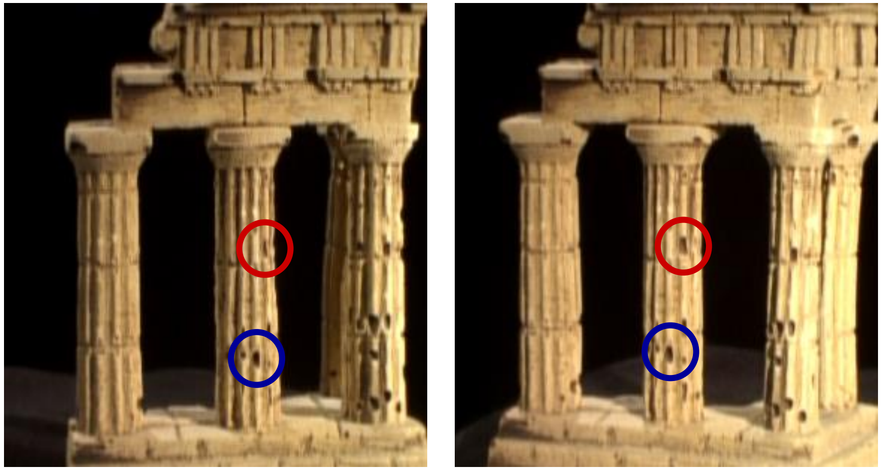
Author: Benjamin Ummenhofer

Shape from silhouette

- Most shape reconstruction methods are based on point correspondences between images from different viewing directions.



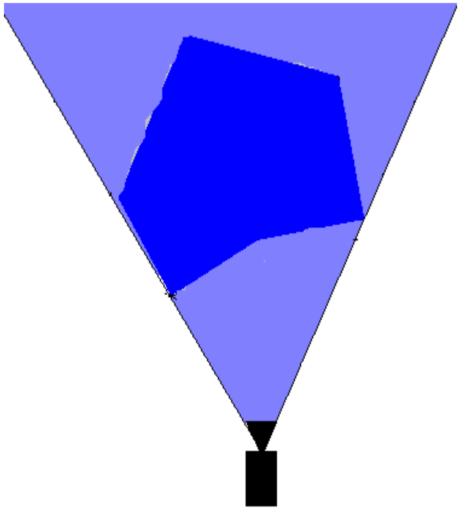
- Correspondences are preferably established for points whose projection rays hit the surface in nearly orthogonal direction.
- If the projection ray is nearly tangential to the surface in one camera, structures on the surface are severely deformed in the image and do not allow for robust matching anymore.

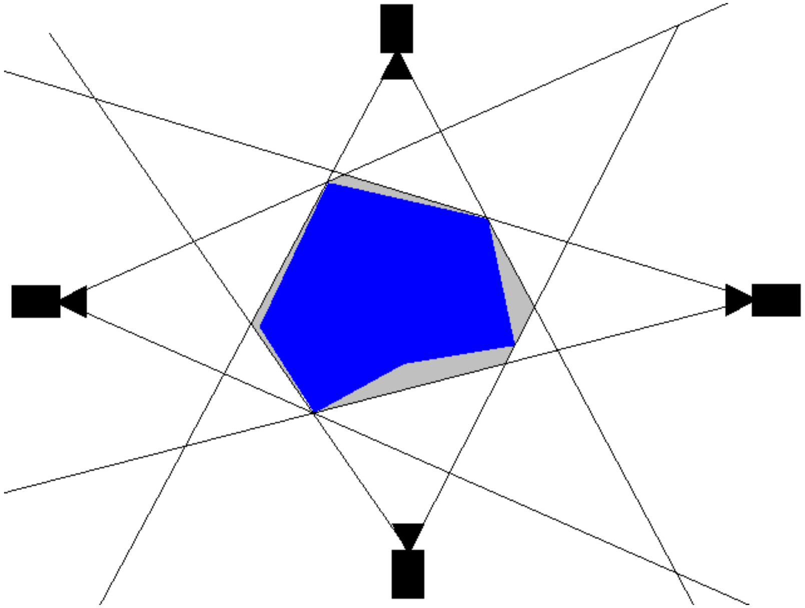


- **Shape from silhouette** is kind of the contrary approach: shape is inferred from points where the surface is tangential to the viewing direction.

Shape from silhouette

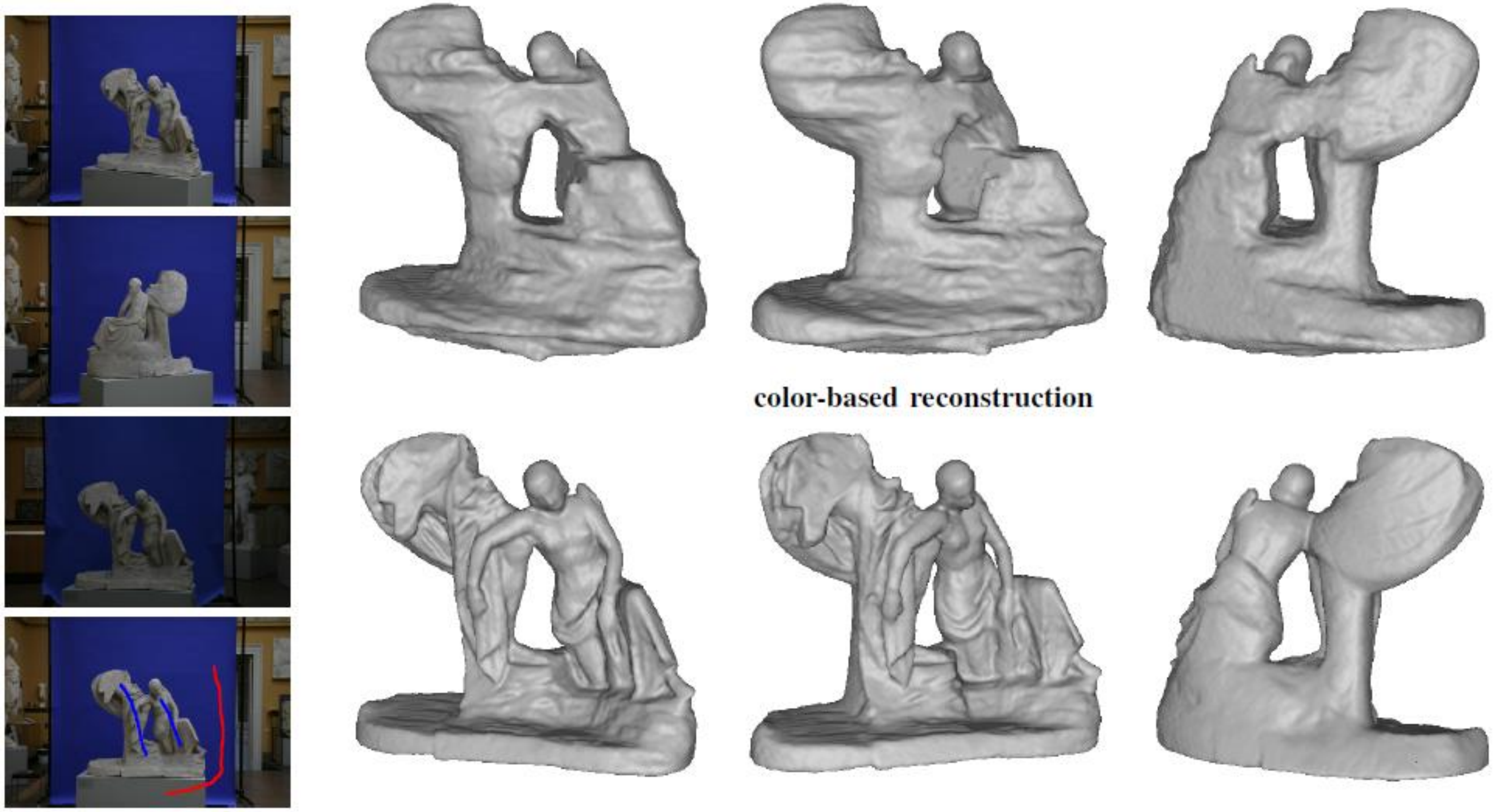
- The silhouette in the image restricts the volume that can be occupied by the observed object.
- Points along projection rays that see background cannot belong to the object volume.
- Points along the projection rays that see foreground may or may not belong to the object volume.
- With additional cameras, the volume that can be occupied by the object is successively reduced.
- The remaining volume is the maximal volume that is consistent with all silhouettes. It is called the **visual hull**.





Author: Keith Yerex

No intrusions



color-based reconstruction

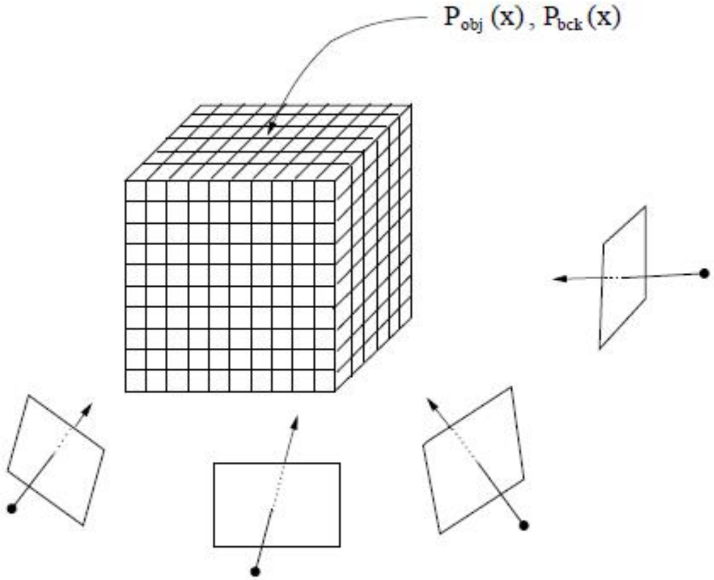
stereo-refined reconstruction

Author: Kalin Kolev

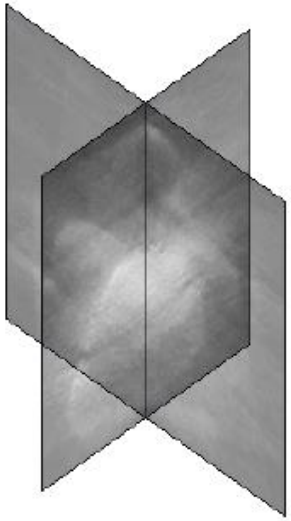
- Intrusions do not show in any of the silhouettes → considered as object

- Shape from silhouette mainly helps find good initializations for other reconstruction methods, e.g., stereo.

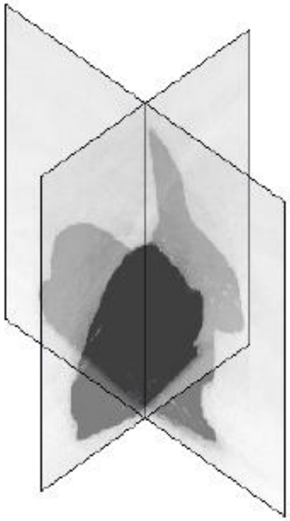
Shape from silhouette as 3D segmentation



probabilistic volume intersection



foreground probability map P_{obj}



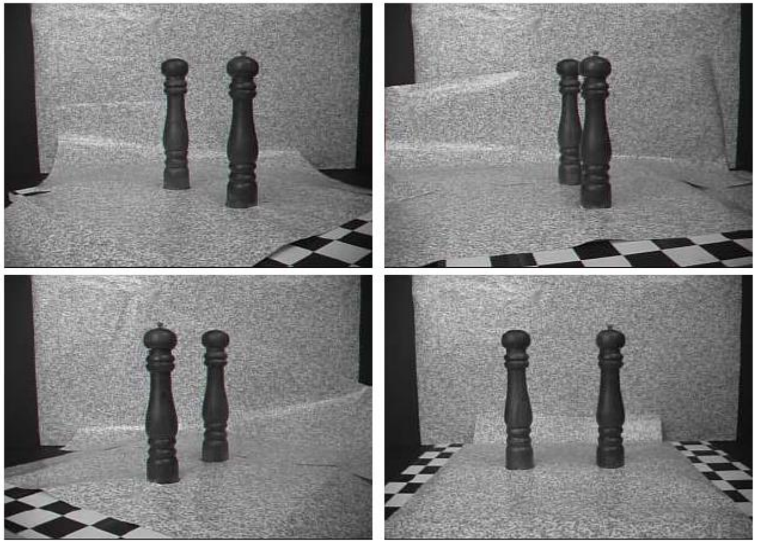
background probability map P_{bck}

- Finding the silhouette in the image (segmentation) and reconstructing the surface can also be considered as a single 3D segmentation problem.

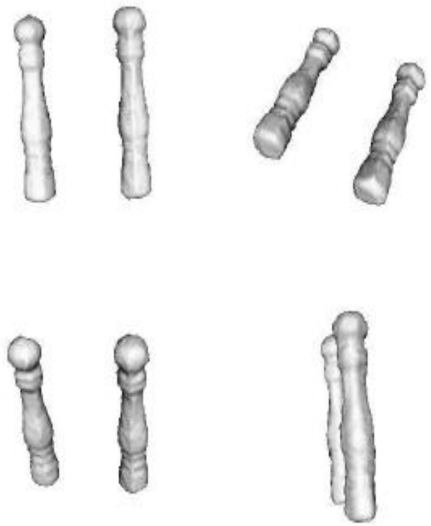
- The task is to find a surface that consistently separates all images into a foreground and a background region.

Examples

Authors: A. Yezzi, S. Soatto

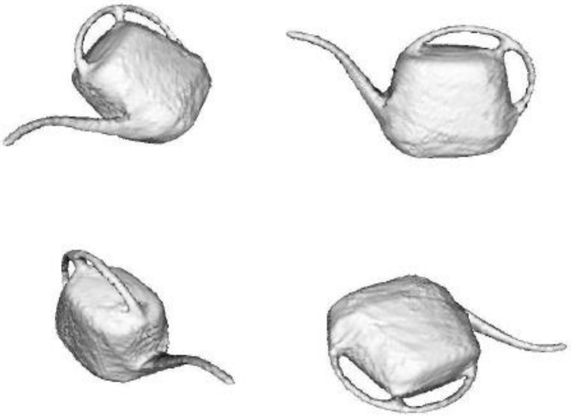


4 of 22 input images

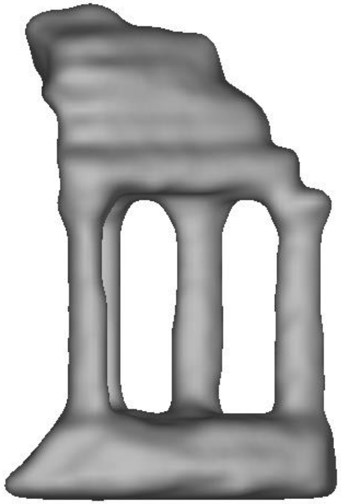


Result

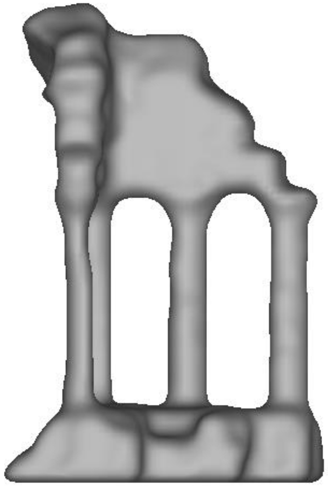
Input image



Result



Shape from silhouette



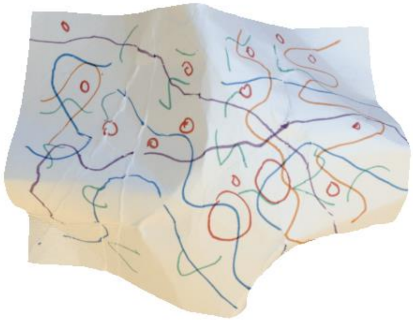
Multi-view stereo

Author: Kalin Kolev

Shape from shading: general case

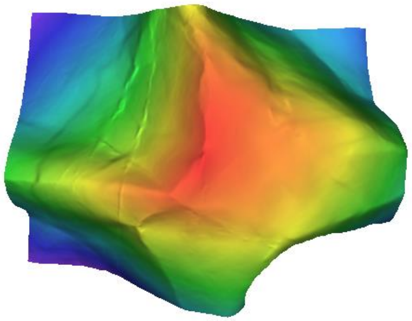
Author: Jon Barron

Given:

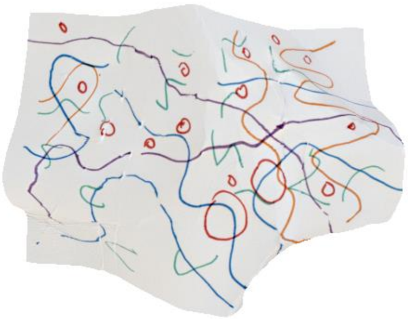


a single image

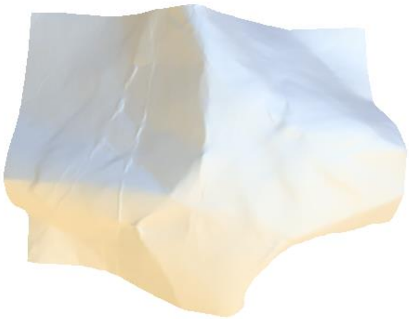
Sought:



Shape



Albedo



Scattering

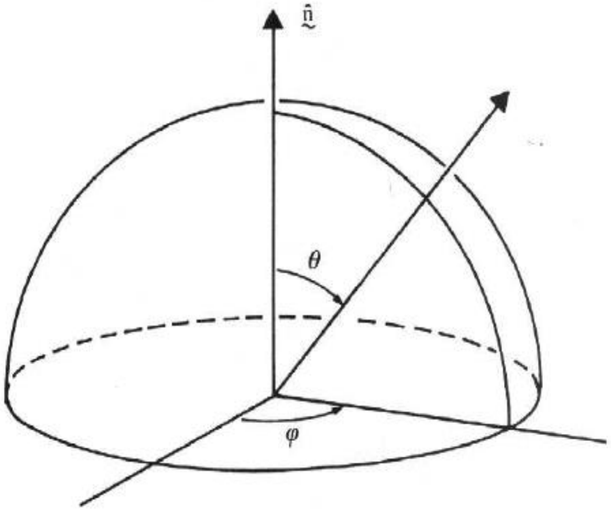
Illumination

Reflectance

Inverse rendering: from the observed image, reconstruct shape, reflectance, and illumination

→ severely under-constrained problem, requires assumptions

Bidirectional reflectance distribution function (BRDF)



Author: B. Horn

- Consider a single light source and a surface element.

- Assume the direction of the light source is given in polar coordinates (θ, ϕ) and the incoming energy is $E(\theta, \phi)$.

- The light emitted by the surface in direction (θ_e, ϕ_e) is given by $L(\theta_e, \phi_e)$.

- The function that describes the reflectance properties of the material is called the **bidirectional reflectance distribution function (BRDF)**:

$$L(\theta_e, \phi_e) = f(\theta, \phi, \theta_e, \phi_e) E(\theta, \phi)$$

- In some more detailed models, the BRDF can have more than four parameters.

Lambertian reflectance

- A **Lambertian surface** emits the incoming light uniformly in all directions, i.e., the **radiance** only depends on the angle between the surface normal $\mathbf{n}$ and the light source direction $\mathbf{s}$:

$$I(x,y) = R(X,Y,Z) = \rho \mathbf{s}^\top \mathbf{n}$$

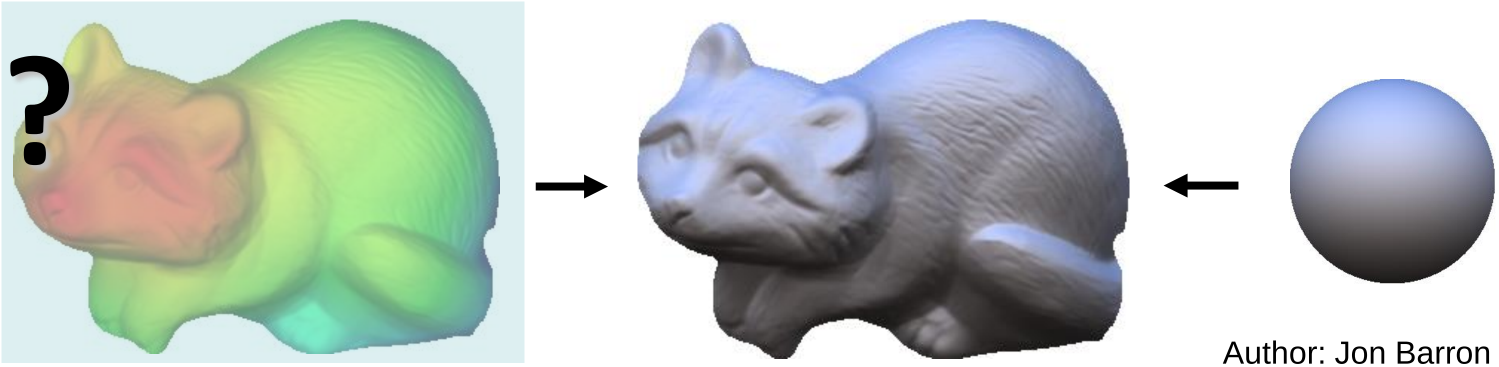
- Corresponds to rough materials (e.g. stone, textiles).

- Most of these materials also absorb some of the incoming light. The non-absorbed fraction of the incoming light is called **albedo** ρ.

- Lambertian reflectance model often used in computer vision, e.g., in stereo matching

- Assuming constant albedo ρ and given light source direction $\mathbf{s}$ leads to a basic shape from shading approach

Basic shape from shading approach (Horn 1970)



Author: Jon Barron

Lambertian reflectance; illumination and albedo are known

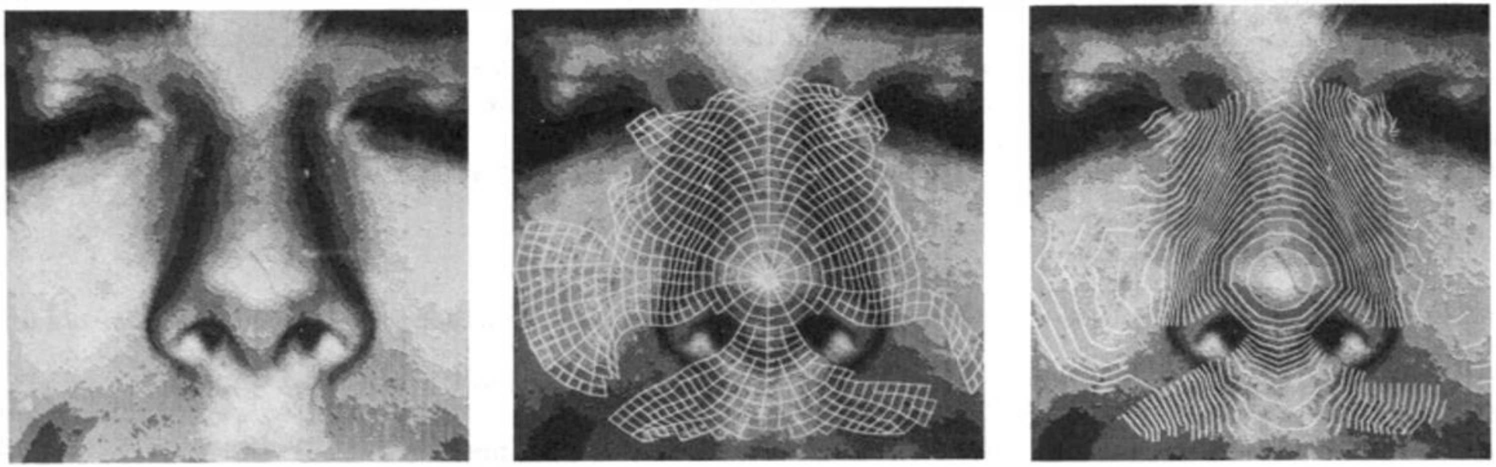


Figure 11-7. The shape-from-shading method is applied here to the recovery of the shape of a nose. The first picture shows the (crudely quantized) gray-level image available to the program. The second picture shows the base characteristics superimposed, while the third shows a contour map computed from the elevations found along the characteristic curves.

B. K. P. Horn. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. Technical report, MIT, 1970.

Basic shape from shading approach (Horn 1970)

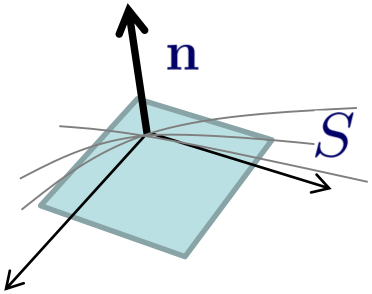
- The sought depth map $Z(x,y)$ can be related to the surface normal direction $\mathbf{n}$ → estimate $\mathbf{n}$

- The derivatives of the surface $S(x,y) = (x,y,Z(x,y))$ with respect to x and y yield two vector fields that span the tangential plane:

$$\partial_x S = (1,0,p)^\top, \qquad \partial_y S = (0,1,q)^\top$$

- The cross product of the two vector fields yields the normal field:

$$\begin{pmatrix} 1 \\ 0 \\ p \end{pmatrix} \times \begin{pmatrix} 0 \\ 1 \\ q \end{pmatrix} = \begin{pmatrix} -p \\ -q \\ 1 \end{pmatrix} \qquad \mathbf{n} = \frac{1}{\sqrt{p^2+q^2+1}} \begin{pmatrix} -p \\ -q \\ 1 \end{pmatrix}$$



- Plug into Lambertian radiance function:

$$I(x,y) = \frac{\rho}{\sqrt{p^2+q^2+1}} \mathbf{s}^\top \begin{pmatrix} -p \\ -q \\ 1 \end{pmatrix}$$

→ like in optical flow: one equation, two unknowns

Variational shape from shading

- Assuming a smooth surface, leads to a variational approach (Ikeuchi-Horn 1981):

$$E(p,q) = \int (I - R(p,q))^2 + \underset{\substack{\uparrow \\ \boxed{\text{Tuning parameter}}}}{\alpha}(|\nabla p|^2 + |\nabla q|^2)\, dx dy$$

with

$$R(p,q) = \frac{\rho}{\sqrt{p^2+q^2+1}}\, \mathbf{s}^\top \begin{pmatrix} -p \\ -q \\ 1 \end{pmatrix} \qquad \boxed{\text{Lambertian reflectance equation}}$$

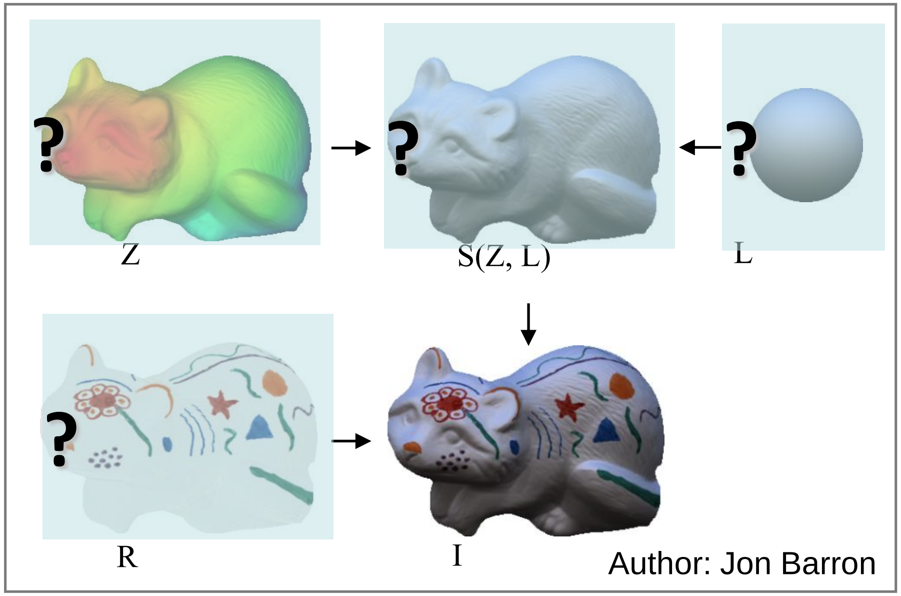
- After deriving the Euler-Lagrange equations, a minimizer for p and q can be found by gradient descent.

- This energy is non-convex and can have multiple local minima. Coarse-to-fine methods can be used as a heuristic to find the global minimum.

SIRFS: Shape, Illumination and Reflectance From Shading

- Recent framework by Barron and Malik considers more components of inverse rendering

$$\underset{Z,R,L}{\text{minimize}} \quad g(R) + f(Z) + h(L)$$
$$\text{subject to} \quad I = R + S(Z, L)$$



Author: Jon Barron

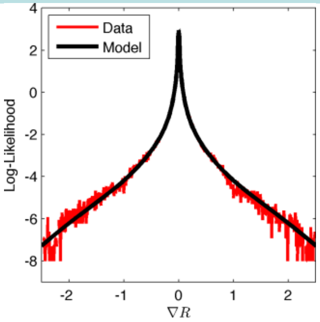
with appropriate cost functions

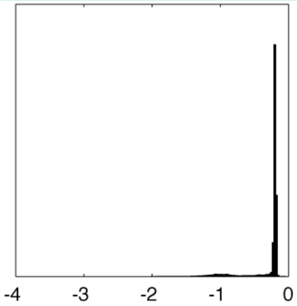
g, f, h

- Includes multiple sources of prior knowledge

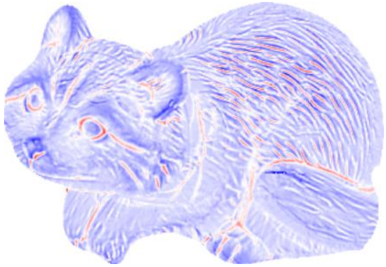
Priors learned from training data (four among many)

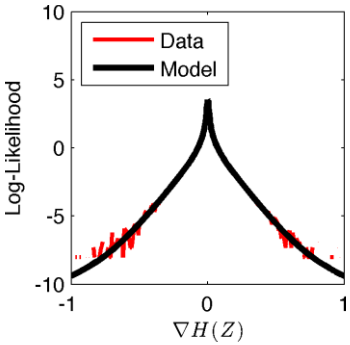
1. Reflectance is mostly smooth and the histogram is sparse



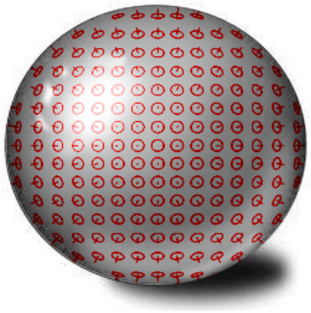


2. Shape is mostly smooth

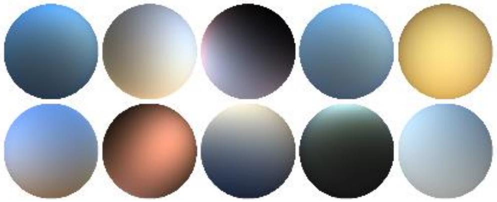




3. Isotropy of shapes and the fact that an observed surface is more likely to point towards the observer

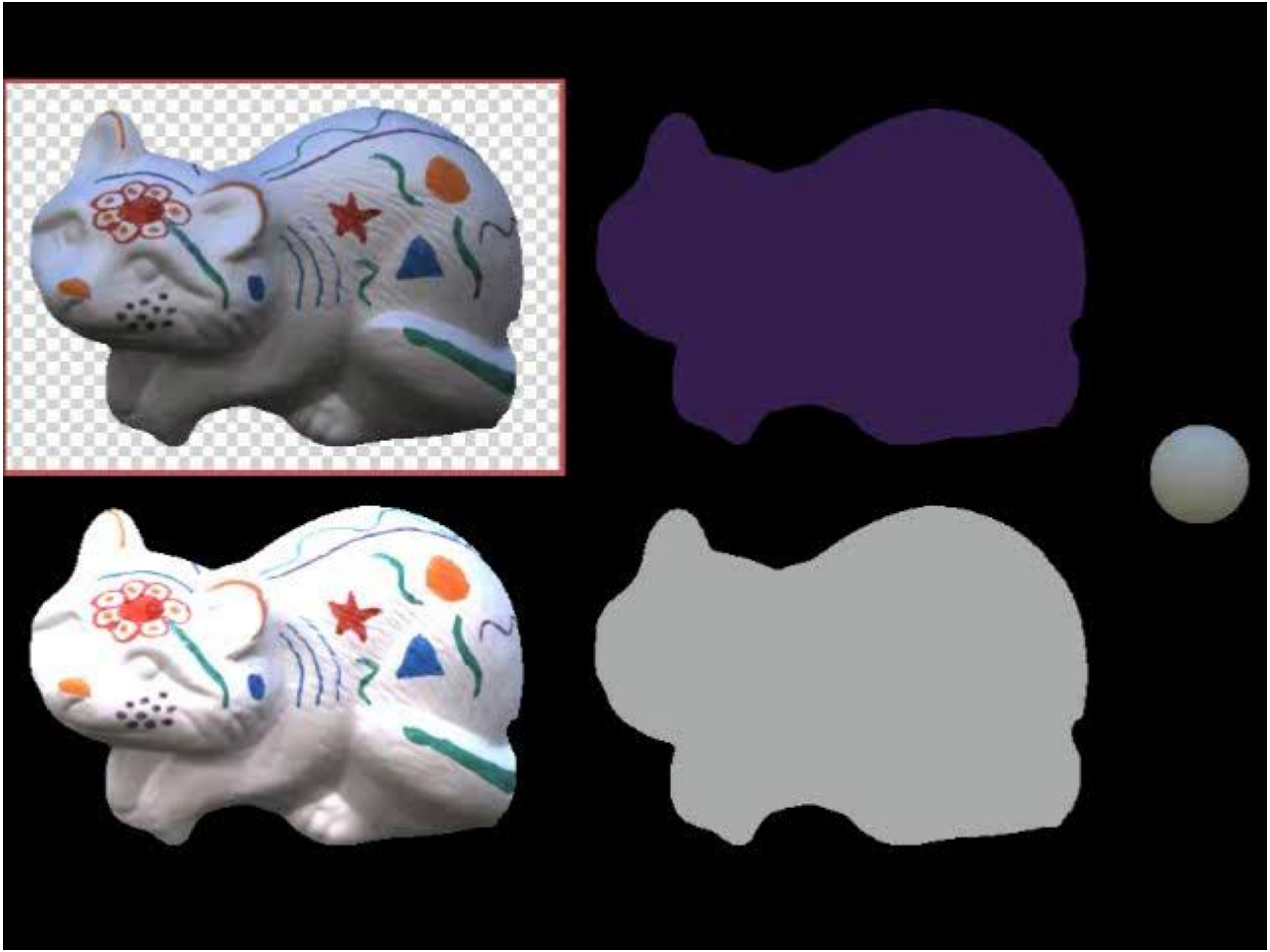


4. Natural lighting prior



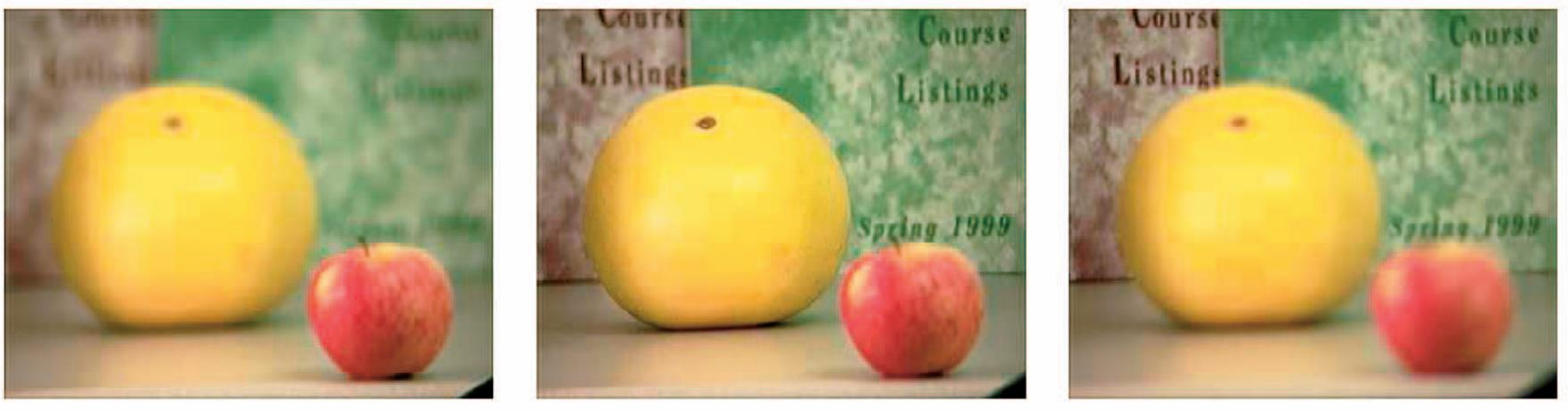
Author: Jon Barron

Optimization with coarse-to-fine Quasi-Newton method



Author: Jon Barron

Shape from defocus



- Given a series of images, where the focus was set to different depths, one can reconstruct the local depth in the image (Favaro-Soatto 2005).

- Estimate the relative blur $d\sigma$ needed to match the other image (Subbarao-Surya 1992)

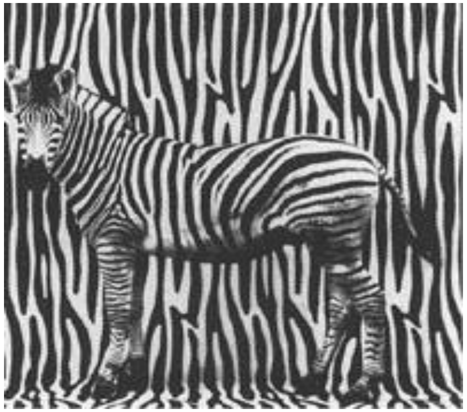
- The amount of blur needed determines the depth

Shape from texture

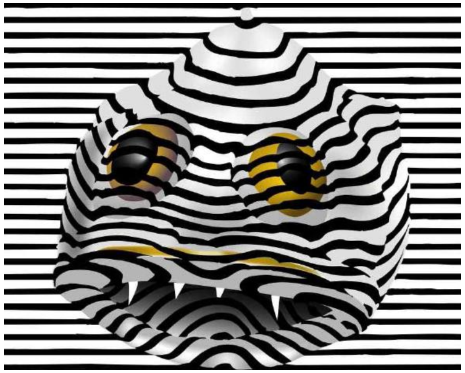
- If the appearance of a texture is known, e.g. a regular repetitive pattern, the shape of the surface can be reconstructed from the texture’s deformation.

- Usually the texture is not known. Then practical assumptions are that the texture is homogenous, isotropic, and stationary.

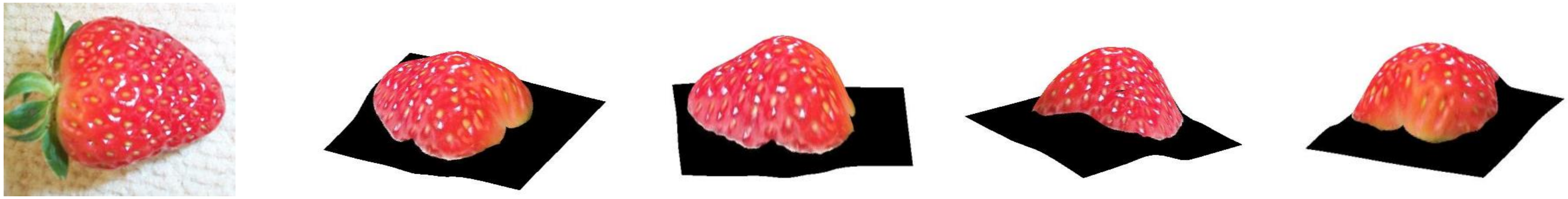
- From the repetitive nature of texture elements under these conditions we can estimate the non-deformed shape of such an element and the surface that causes its deformation.



Source: CV Lab, UC Berkeley



Source: MPI Saarbrücken



Author: Angelina Loh

Summary

- Reconstructing the depth of points that has been lost by the projection to the image plane is one of the central computer vision tasks.

- There are multiple cues for reconstructing 3D shapes:
  - stereo images
  - camera motion
  - multiple silhouette images
  - shading
  - defocus
  - texture

- Most prominent are stereo reconstruction and structure from motion.

References

- O. Faugeras, Q.-T. Luong: *The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications*, MIT Press, 2004.
- R. Hartley, A. Zisserman: *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2nd edition, 2004.
- A. Yezzi, S. Soatto: Stereoscopic segmentation, *International Journal of Computer Vision*, 53(1):31-43, 2003.
- K. Ikeuchi, B. Horn: Numerical shape from shading and occluding boundaries, *Artificial Intelligence*, 17:141-184, 1981.
- J. Barron, J. Malik: Color constancy, intrinsic images, and shape estimation, *European Conference on Computer Vision*, 2012.
- P. Favaro, S. Soatto: A geometric approach to shape from defocus, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):406-417, 2005.
- A. M. Loh: *The recovery of 3-D structure using visual texture patterns*, PhD thesis, 2006.