Multi-modal Registration of Visual Data

Massimiliano Corsini
Visual Computing Lab, ISTI - CNR - Italy

Overview

- Introduction and Background
- Features Detection and Description (2D case)
- Features Detection and Description (3D case)
- Image-geometry registration
- Recent Advances and Applications

Overview

- Introduction and Background
- Features Detection and Description (2D case)
- Features Detection and Description (3D case)
- **Image-geometry registration**
  - **Problem formulation**
  - **Fixed-relative methods**
  - **Features-based methods**
  - **Silhouette-based methods**
  - **Statistical methods**
  - **Multi-view methods**
- Recent Advances and Applications

Image-geometry Registration

Geometric Transformation

- Geometric transformation are often expressed as matrix multiplication
- Example (rotation, 2D):

$$P' = \mathbf{R} \cdot P$$

$$P = \begin{bmatrix} x \\ y \end{bmatrix}; \quad P' = \begin{bmatrix} x' \\ y' \end{bmatrix}; \quad \mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Translation (2D)

- Translation cannot be expressed directly as a matrix multiplication

$$P' = P + \mathbf{T}$$

$$P = \begin{bmatrix} x \\ y \end{bmatrix}; \quad P' = \begin{bmatrix} x' \\ y' \end{bmatrix}; \quad \mathbf{T} = \begin{bmatrix} d_x \\ d_y \end{bmatrix};$$

Homogeneous Coordinates

- Point $P = (x,y)$ is represented in *homogeneous coordinates* as (x_h, y_h, w), where:

$$x = x_h/w; \quad y = y_h/w; \quad \text{with} \quad w \neq 0.$$

- Two points *Pa* = (x, y, w) and *Pb* = (x', y', w') in homogeneous coordinates represent the same point if and only if the coordinates of *Pb* are a multiple of *Pb*;
- When $w = 1$ (canonic form) homogeneous coordinates coincide with cartesian coordinates.
- *Points* are represented as $(x, y, w \neq 0)$ while points at infinity (i.e. *directions*) as $(x, y, 0)$.

Transformation in homogeneous coordinates

- Translation (2D):

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & d_x \\ 0 & 1 & d_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

- Rotation (2D):

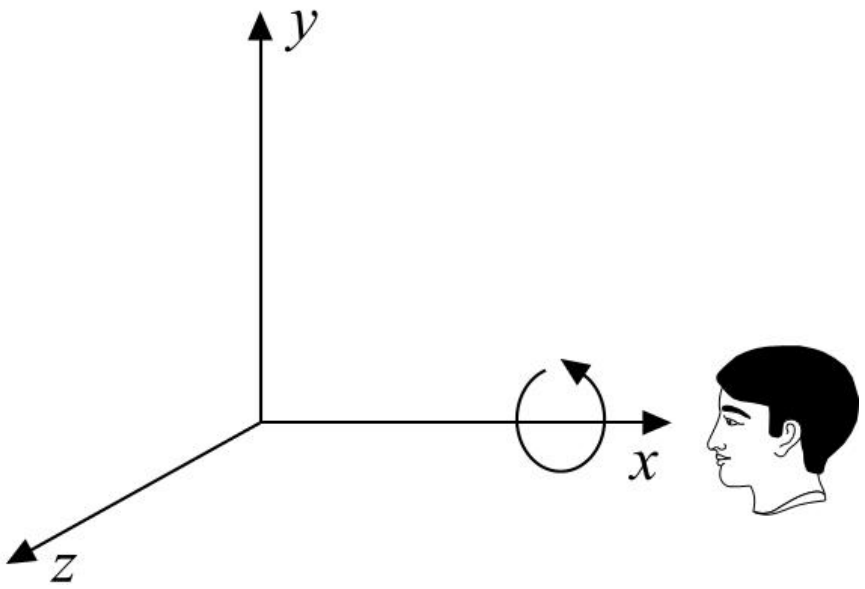
$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

Rotation (3D)

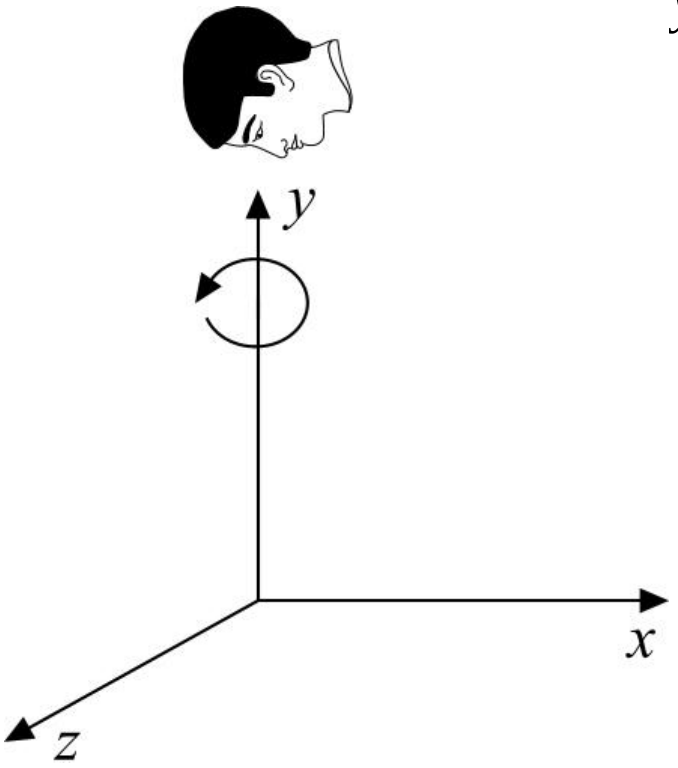
- 2D translation can be naturally extended to 3D
- 2D rotation cannot be extended to the 3D case (an axis must be specified)

Rotation around the X axis

$$\mathbf{R}_x(\theta) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & -sen\theta & 0 \\ 0 & sen\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



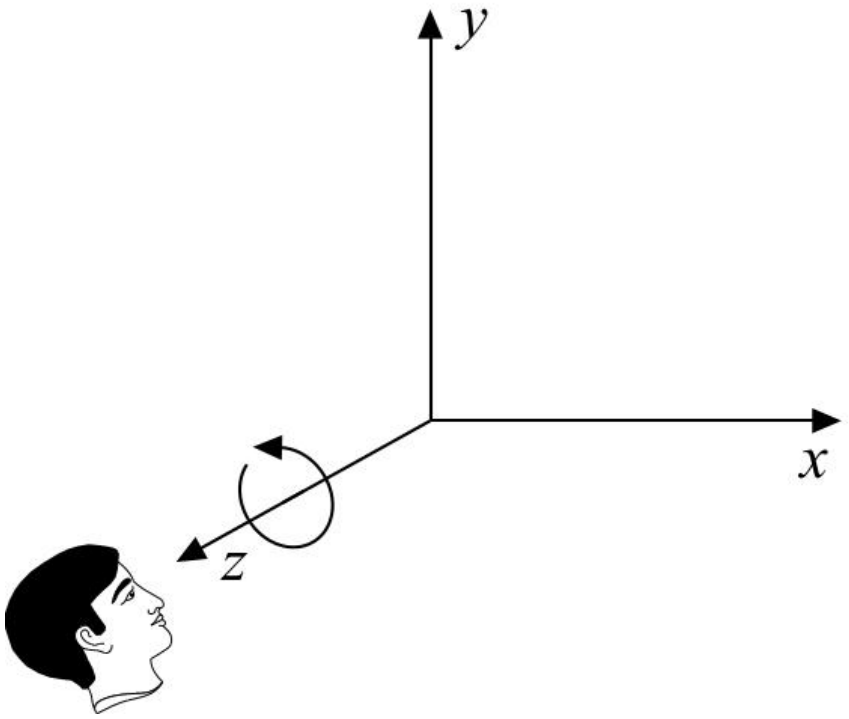
Rotation around the Y axis



$$\mathbf{R}_y(\theta) = \begin{bmatrix} \cos\theta & 0 & \text{sen}\,\theta & 0 \\ 0 & 1 & 0 & 0 \\ -\text{sen}\,\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Rotation around the Z axis

$$\mathbf{R}_z(\theta) = \begin{bmatrix} \cos\theta & -\text{sen}\,\theta & 0 & 0 \\ \text{sen}\,\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



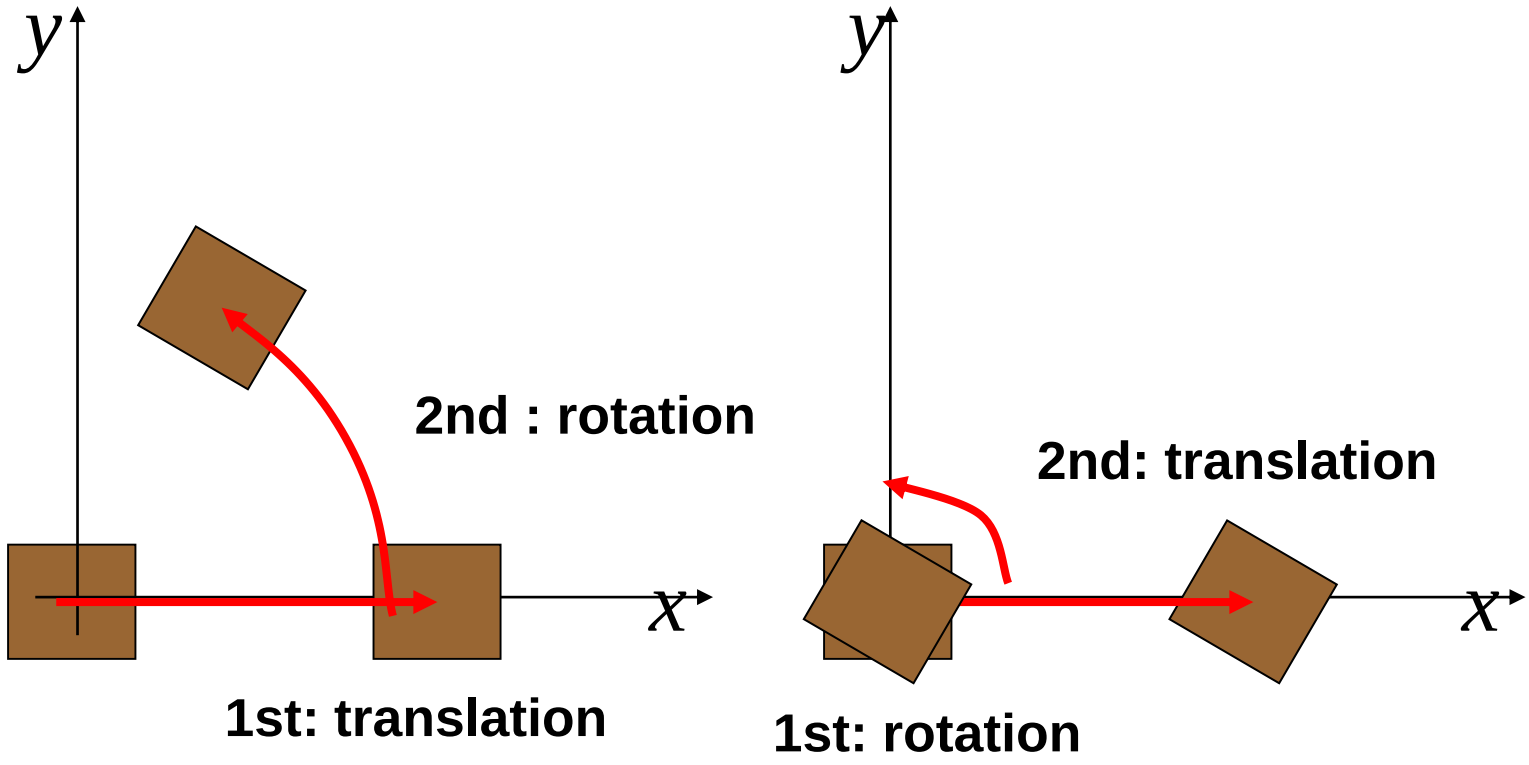
Concatenation of Transformations

- The representation of geometric transformations in homogeneous coordinates permits to concatenate easily two or more transformations:

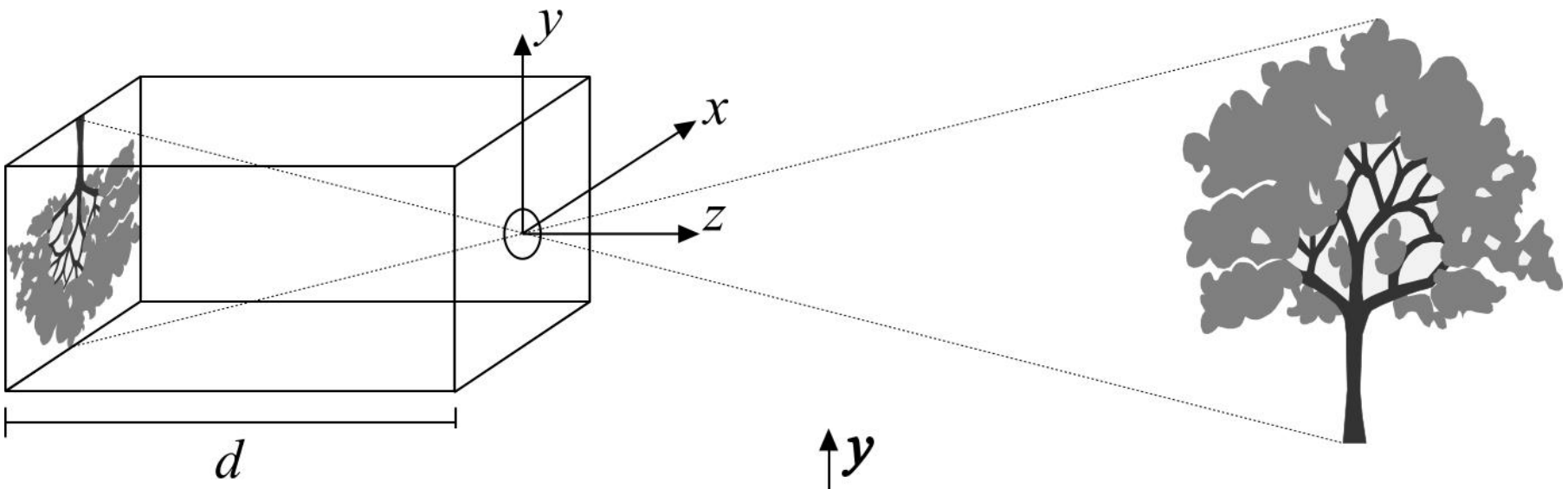
$$T = T_N \cdot T_{N-1} \cdot \ldots \cdot T_2 \cdot T_1$$

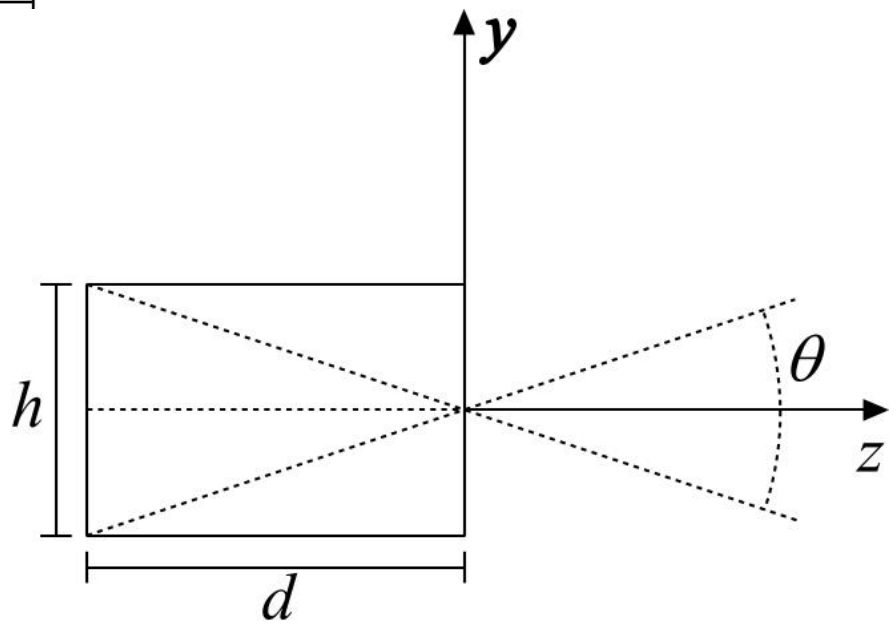
- *Note: the concatenation of geometric transformations is not commutative* (!)

Commutativity

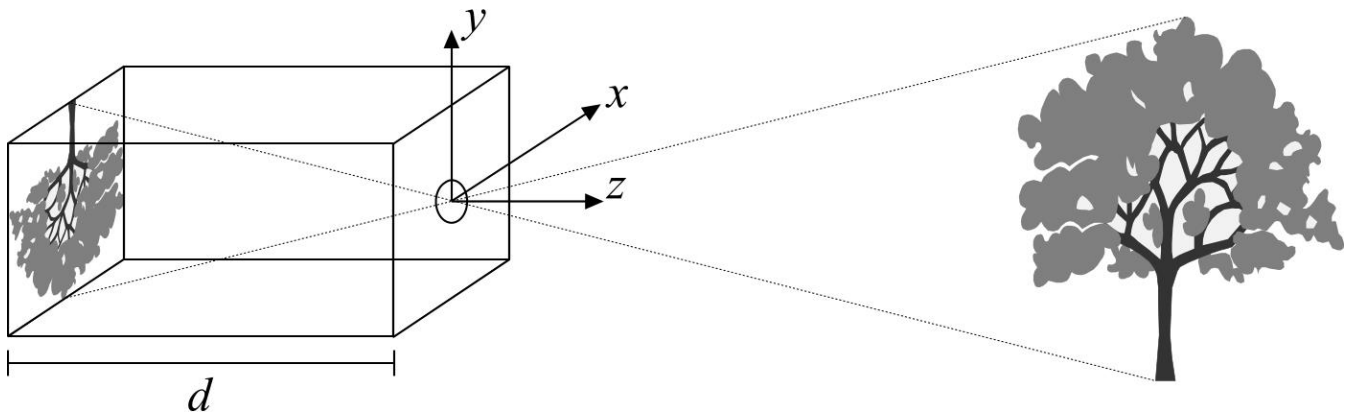


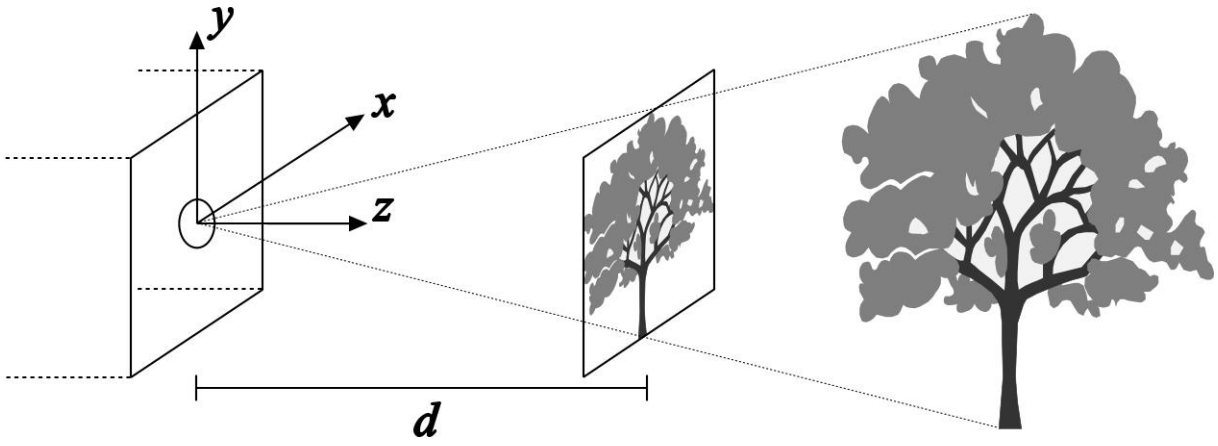
Pinhole Camera



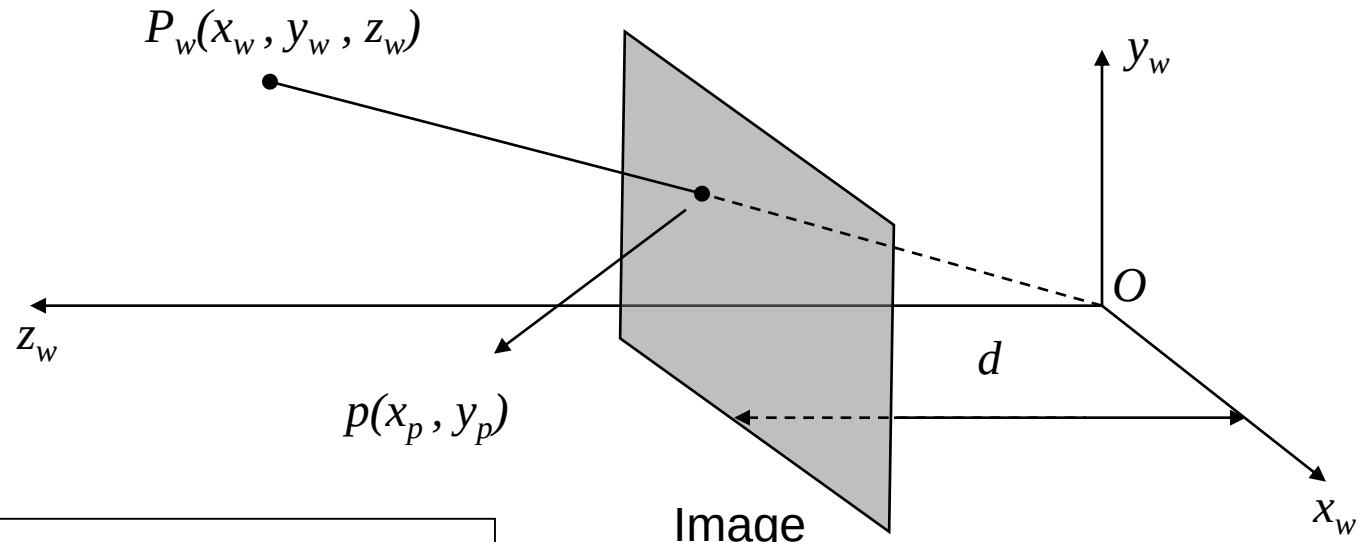


Pinhole Camera





Perspective Projection



$$\frac{x_p}{d} = \frac{x_w}{z_w} \quad , \quad \frac{y_p}{d} = \frac{y_w}{z_w}$$

$$\begin{bmatrix} x_p \\ y_p \\ d \end{bmatrix} = \begin{bmatrix} (x_w d)/z_w \\ (y_w d)/z_w \\ d \end{bmatrix}$$

Coordinate omogenee

$$\begin{bmatrix} x_p \\ y_p \\ d \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1/d & 0 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = T_{\text{persp}} P_w$$

Perspective Projection

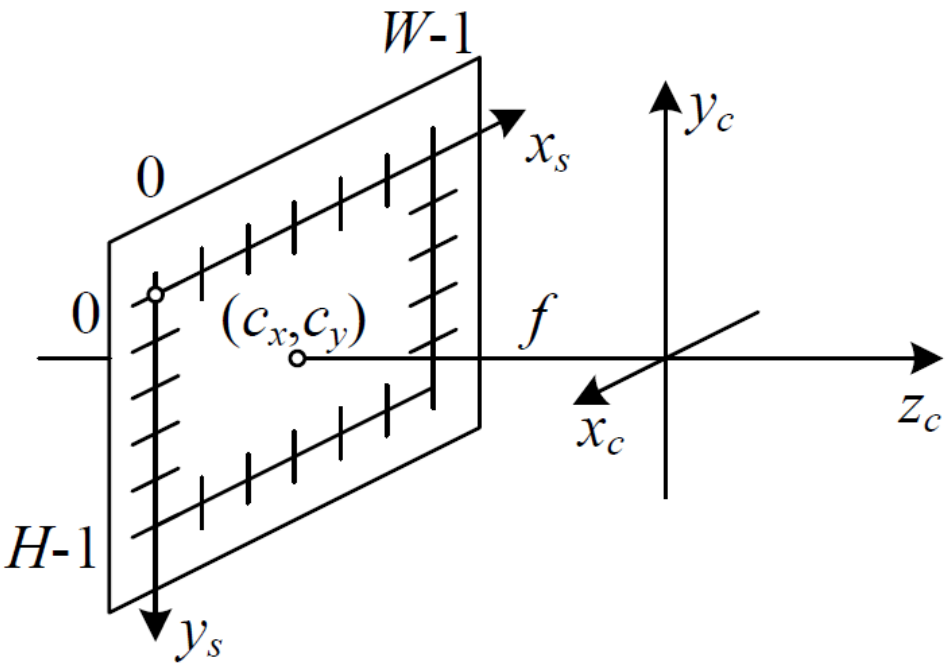
- So, we can write:

$$\begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/d & 0 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix}$$

- Or, equivalently:

$$\begin{bmatrix} x_p \\ y_p \\ 1 \end{bmatrix} = \mathbf{K}\,[\mathbf{I}|0]\,P_w \qquad \mathbf{K} = \begin{bmatrix} d & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & 1 \end{bmatrix} P_w$$

From image plane to pixel coordinates



$$x_s = x_p s_x + c_x$$
$$y_s = y_p s_y + c_y$$

Optical center

S_x and S_y depends on the pixels size

Camera Model

Calibration Matrix

Focal length

$$\begin{bmatrix} x_s \\ y_s \\ 1 \end{bmatrix} = \mathbf{K}\,[\mathbf{I}|0]\,P_w \qquad \mathbf{K} = \begin{bmatrix} fs_x & s & c_x \\ 0 & fs_y & c_y \\ 0 & 0 & 1 \end{bmatrix} P_w$$

$$\mathbf{K} = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix} P_w$$

Usual version
(pixels have right angles – skew coefficient is zero)

Camera Model

- Taking into account the position and the orientation of the camera in the scene:

$$\begin{bmatrix} x_s \\ y_s \\ 1 \end{bmatrix} = \mathbf{K}\,[\mathbf{I}|0]\,\mathbf{M}P_w \qquad \mathbf{M} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 1 & 0 \end{bmatrix}$$

$$\boxed{p_s \simeq P P_w}$$

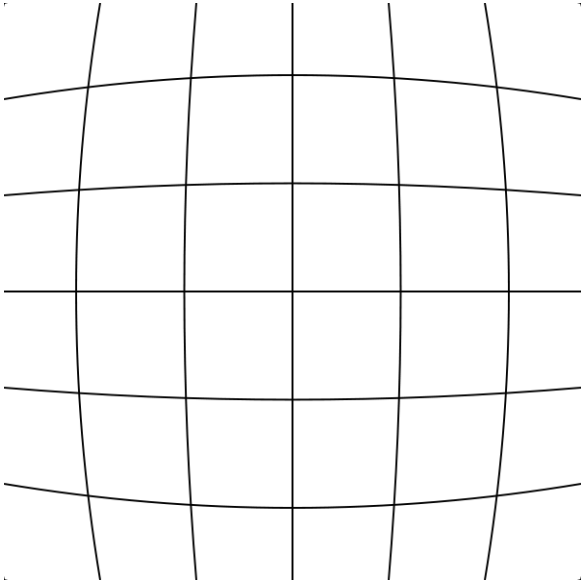
Camera Parameters

- So, we have two sets of parameters that charactize the camera.
- *Intrinsic parameters*: focal length, pixel size, resolution, optical center.
- *Extrinsic parameters*: position and orientation of the camera, i.e. a rotation and a translation.

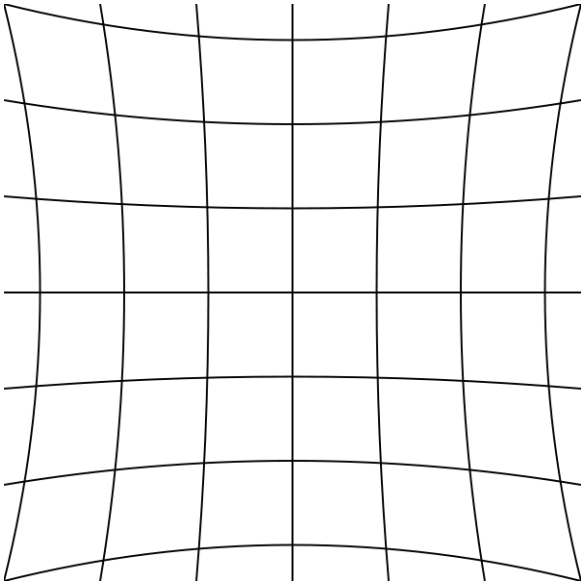
A Note about Lens Distortion

- Camera lenses introduce a certain amount of *radial distortions*.
- We have not considered it in our camera model.

Barrell



Pincushion



From Wikipedia “Distortion (Optics)”.

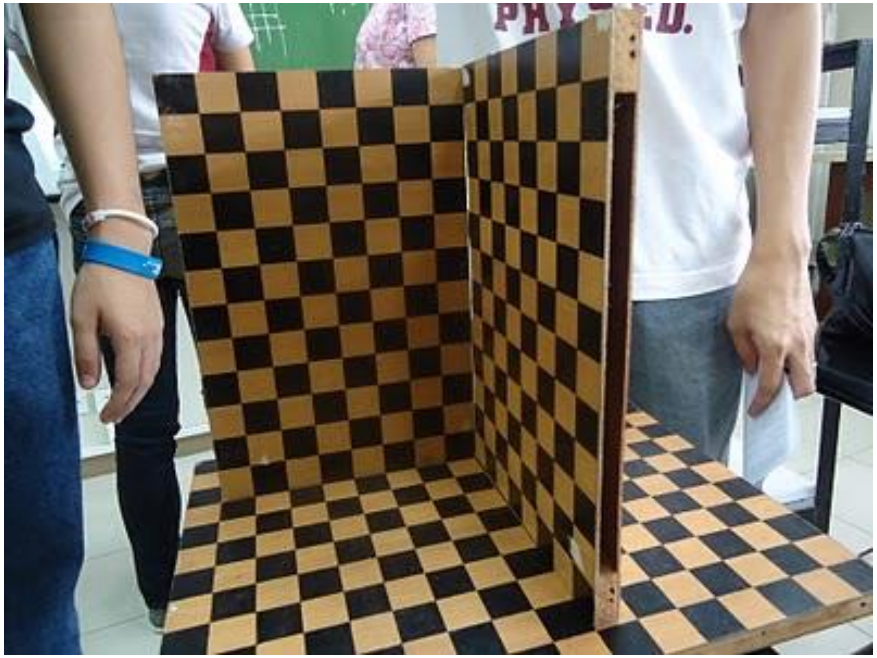
Camera calibration

- *Camera calibration* means to recover both the intrinsic and the extrinsic parameters of the camera.
- Several methods have been developed during the years.
- Tsai's algorithm [Tsai1989] is one of the most famous one.
- Direct Linear Transform (DLT) [Faugeras1993] is another widely used method.

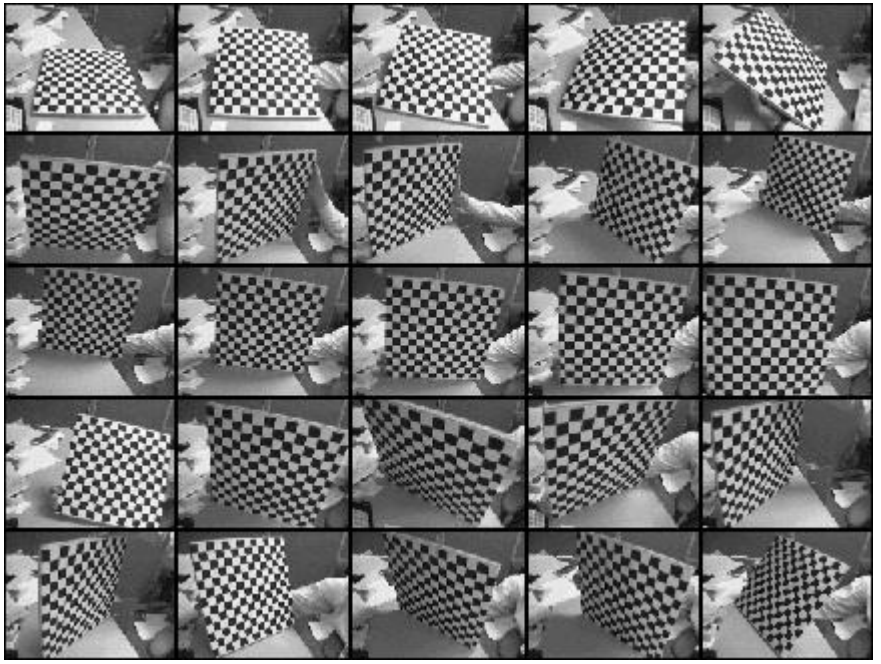
[Tsai1989] R. Tsai, *"A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses"*, IEEE Journal of Robotics and Automation, Vol. RA–3, No. 4, August 1987, pp. 323–344.

[Faugeras1993] O. Faugeras, *"Three dimensional computer vision: A geometric viewpoint"*, MIT Press, Boston, 1993.

Calibration Target



3D Calibration Target



Planar chessboard
(it requires multiple acquisition)

Calibration methods

- After establishing *N* 2D-to-3D point correspondances:

$$p_{s,i} \simeq P P_{w,i}$$

- The idea is to solve a linear system of homogeneous equation (DLT):

$$p_{s,i} \times P P_{w,i} = 0$$

- *DLT* is an *algebraic* method.
- *Tsai* is a *geometric* method.

Problem formulation

- Typically, the skew coefficient is assumed to be zero (*s=0*) and the pixels to be square (*sx=sy*).
- Resolution is known (!)
- *The center of projection (cx, cy) is usually associated to the center of the CCD but this is not always true (!)*
- The **2D/3D registration problem** is often formulated as: *given an image and a 3D model we have to estimate the 7 parameters: f,* t_x, t_y, t_z, r_x, r_y, r_z *(extrinsics + focal length)*

Registration methods

- *During* the geometry acquisition (*co-located camera approach*):
  - Images are perfectly aligned with the shape
  - Some devices have usually low resolution
- *After* the geometry acquisition:
  - Images have to be registered to the 3D model
  - In some cases this is the only viable solution (e.g. nightly acquisition, different set to map on the same 3D model, etc.)

Fixed-relative methods

- *The pose of each camera is related to a known position (co-located methods) or to a previously estimated position.*

[1] Pulli, K., Abi-Rached, H., Duchamp, T., Shapiro, L. G., & Stuetzle,W. Acquisition and visualization of colored 3D objects. In *Proc. of the 14th int. Conf. on Pattern Recognition (ICPR'98)* (Vol. 1, p. 11).

[2] Sequeira, V., & Goncalves, J. G. (2002). 3D reality modelling: Photorealistic 3d models of real world scenes. In *Int. Symposium on 3D data processing visualization and transmission* (p. 776).

[3] Früh, C., & Zakhor, A. (2003). Constructing 3D city models by merging aerial and ground views. *IEEE Computer Graphics and Applications*, *23*, 52–61.

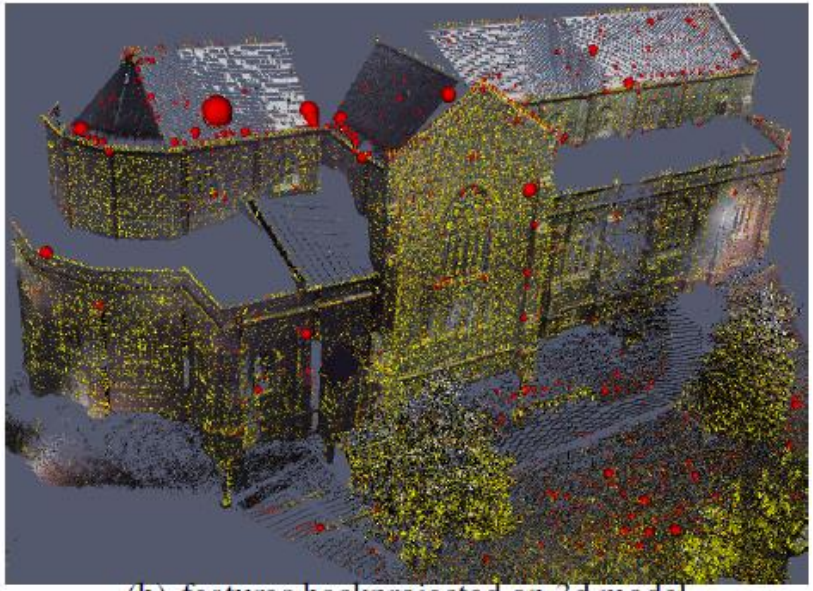
Fixed-relative methods

- Some methods of this type do the registration by align the image to register with the reflectance maps acquired together with the geometry during the acquisition phase.
- Other methods, employ pre-registered image (with a co-located approach). As a representative of these methods, we take a look at the one by Yang et al. [4]

[4] G. Yang, J. Becker, C. V. Stewart, "Estimating the Location of a Camera with Respect to a 3D Model", 3DIM'07, 2007.

Yang et al.[4]

(a) 3d model



(b) features backprojected on 3d model

(c) zoom-in on features

(d) image at night

(e) image after snow

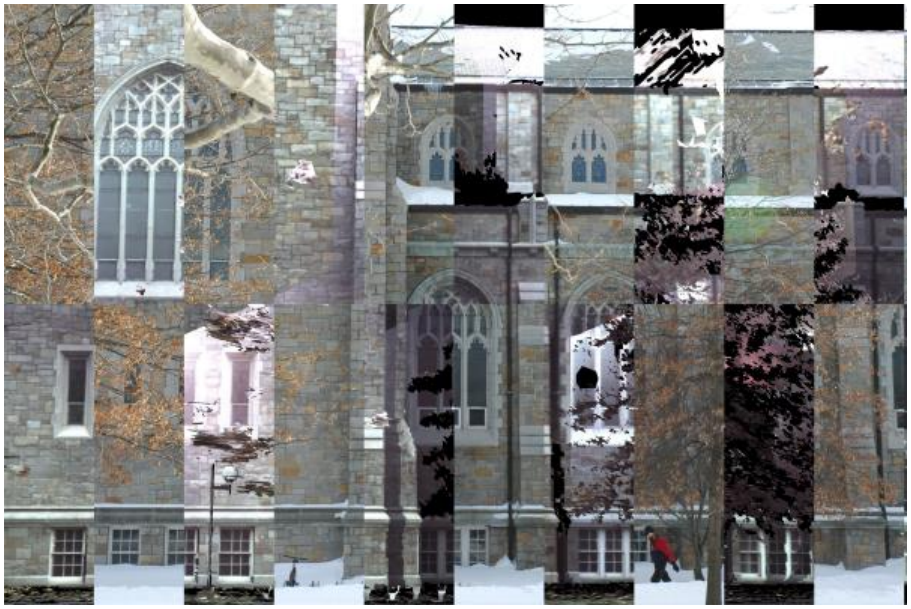
Yang et al.[4]

- The image set is acquired with a calibrated camera having same optical pathway.
- SIFT are extracted for each image and back-projected on the 3D model identify keypoints on its surface (*model keypoints*).
- For each model keypoint a plane is fit using a relatively large portion of the LIDAR points (80 x 80 points).
- Edge-like and corner-like features are also extracted at multiple scale and back-projected (*model features*).

Yang et al.[4] – Algorithm

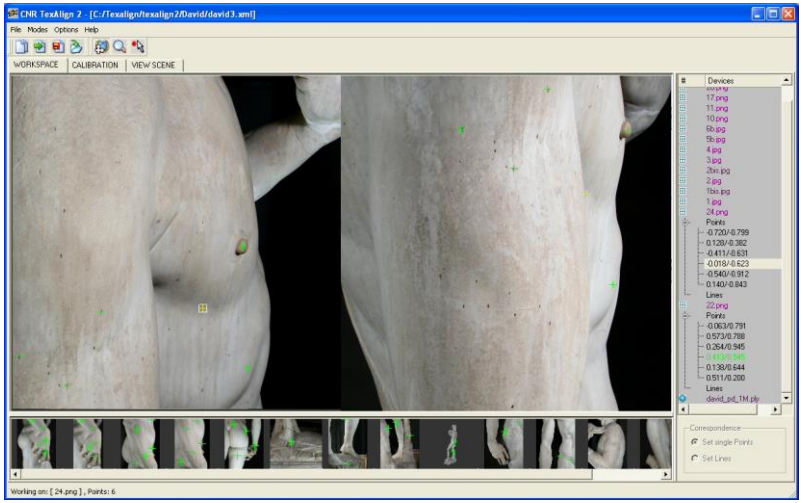
- Step 1: Rank-order keypoints matching
- Step 2: 2D-to-2D registration
- Step 3: 2D-to-3D registration + region growing

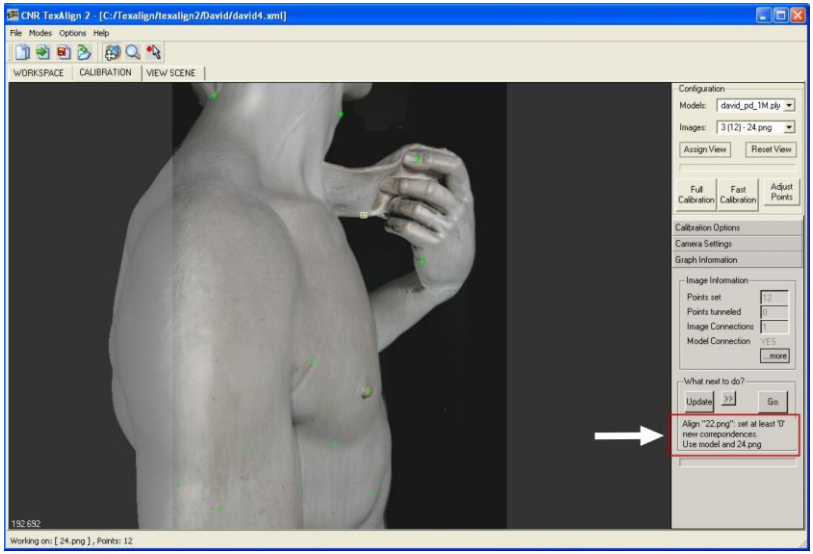
Yang et al.[4] – Results



Semi-automatic Methods

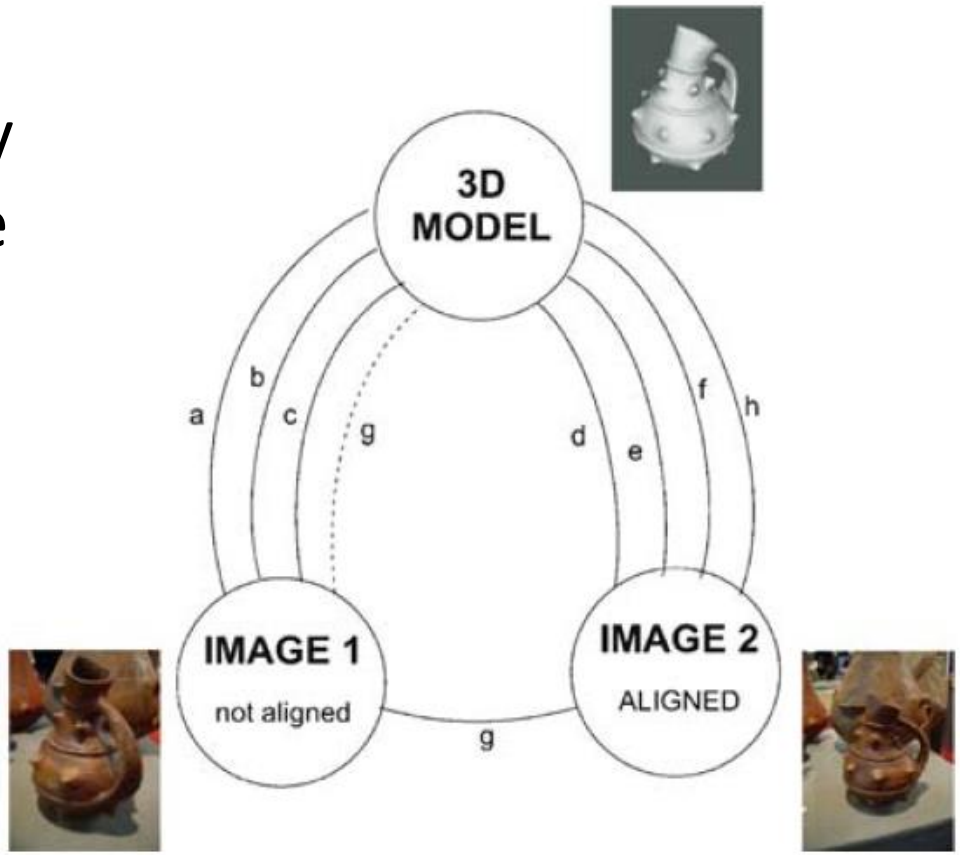
- *They require user intervention but are very robust and general.*
- *An interface permits to manually select pairs of corresponding points between the images and the geometric data.*
- The system computes camera parameters from those corresponding pairs.





Franken et al. 2005[5]

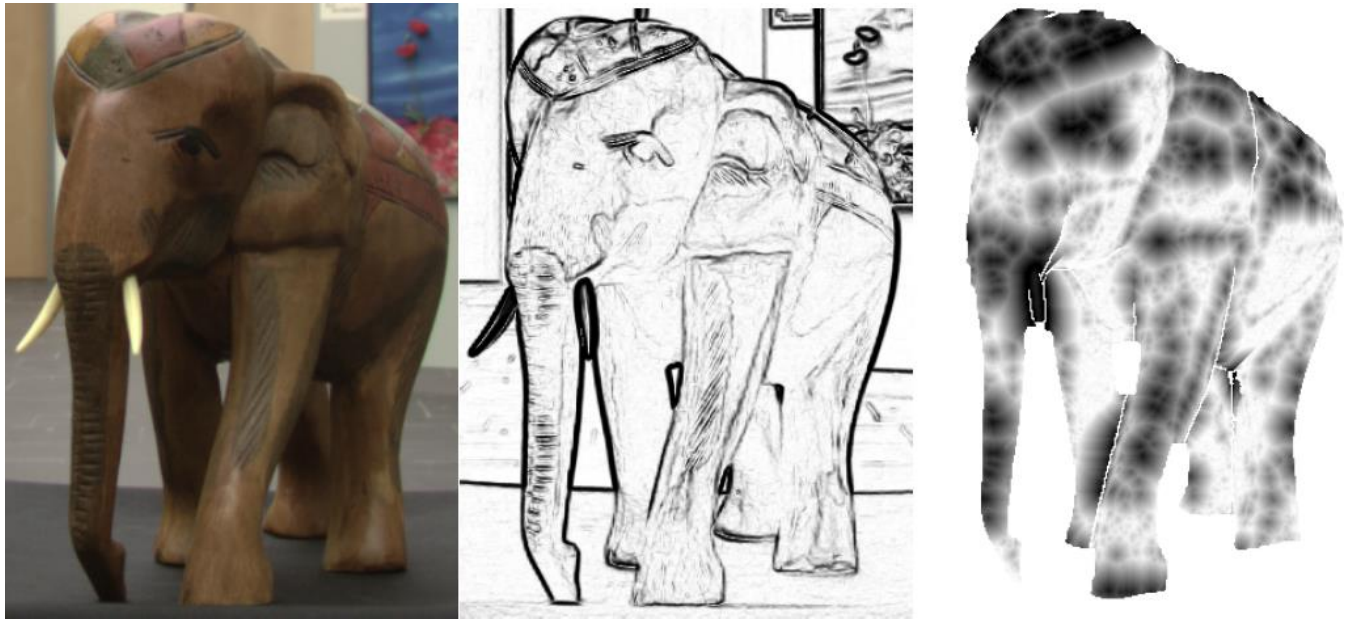
- Exploit the given image-image and image-geometry correspondences to reduce the user effort.
- Use Tsai camera calibration (minimum 12 correspondences).



[5] T. Franken, M. Dellepiane, F. Ganovelli, P. Cignoni, C. Montani, R. Scopigno, "Minimizing user intervention in registering 2D images to 3D models", *The Visual Computer*, Vol. 21(8-10), 2005.

Features-based methods

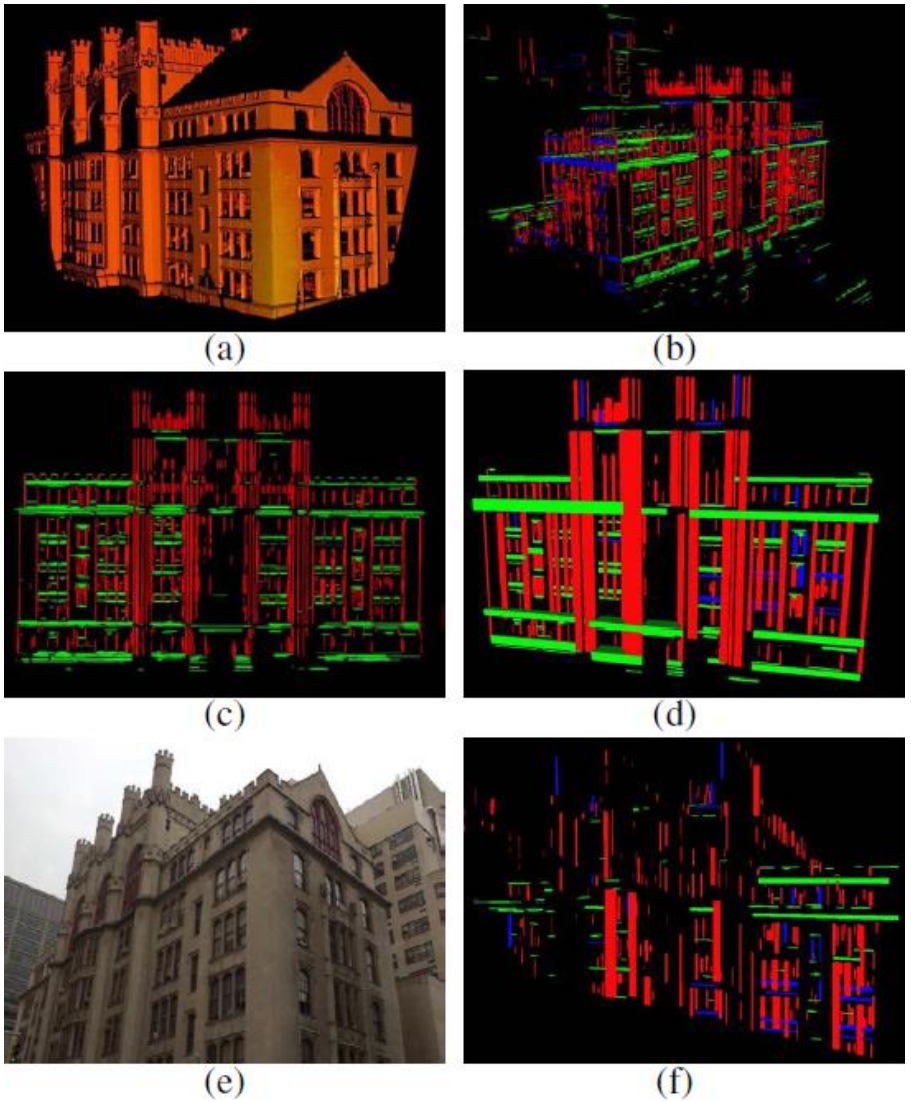
- *Images and geometry is analyzed in order to extract keypoints, lines, rectangles, circles.*
- Neugebauer at al.[6] employed a distance field computed on the extracted edges (Euclidean distance is used on the 3D model).



[6] P. J. Neugebauer and K. Klein, "Texturing 3D models of real world objects from multiple unregistered photographic views", *Computer Graphics Forum*, *18*(3), pp. 245–256, 1999.

Features-based methods

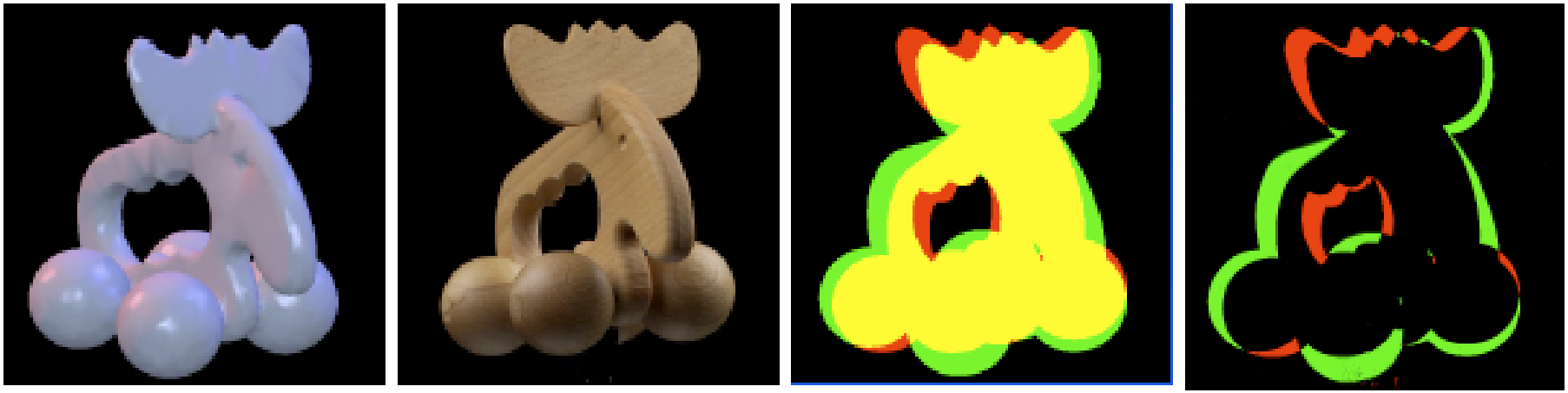
- Method by Liu et al.[7] : parallelepipedes are extracted from the geometric data, rectangles are extracted from the image lines, and then matched.



[7] L. Liu and I. Stamos, "Automatic 3D to 2D registration for the photorealistic rendering of urban scenes", *CVPR'05,* Vol. 2, pp. 137–143, 2005.

Silhouette-based methods

- *The error between the contour/silhoeutte of the object in the image and the projected contour/silhouette of the 3D object is minimized in same way.*
- Lensch et al.[8] proposed to render the silhouette of the object and compare it with the one in the image. The number of pixels covered by just one silhouette (XOR) is minimized. The simplex method is used to minimize the matching error.



3D model | **Image** | **Projected silhouette + image silhouette** | **XOR**

[8] Hendrik P. A. Lensch and Wolfgang Heidrich, "A Silhouette-Based Algorithm for Texture Registration and Stitching", Graphical Models, Vol. 63, pp. 245-262, 2001.

Statistical methods

- One of the most used mathematical tool in multi-modal image registration (in medical ambit in particular) is Mutual Information (MI) which catches the non-linear correlation between the different sources → the idea is to use MI to exploit the non-linear correlation between the image and the geometric properties of the surface.

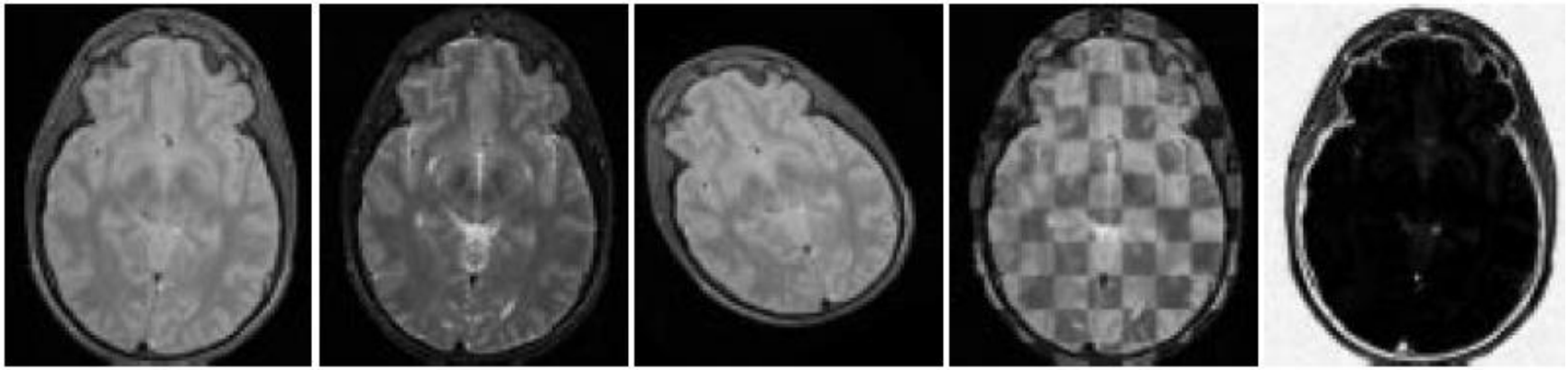
Viola and Wells[9]

- Viola and Wells[9], for first, used Mutual Information (MI) to correlate surface normals with image intensity.
- A (stochastic) gradient descent minimization framework is used for the parameters estimation.

[9] P. Viola and W. M. Wells "Alignment by Maximization of Mutual Information", *Int. J. Computer Vision*, Vol. 24(2), pp. 137-154, 1997.

Viola and Wells[9]

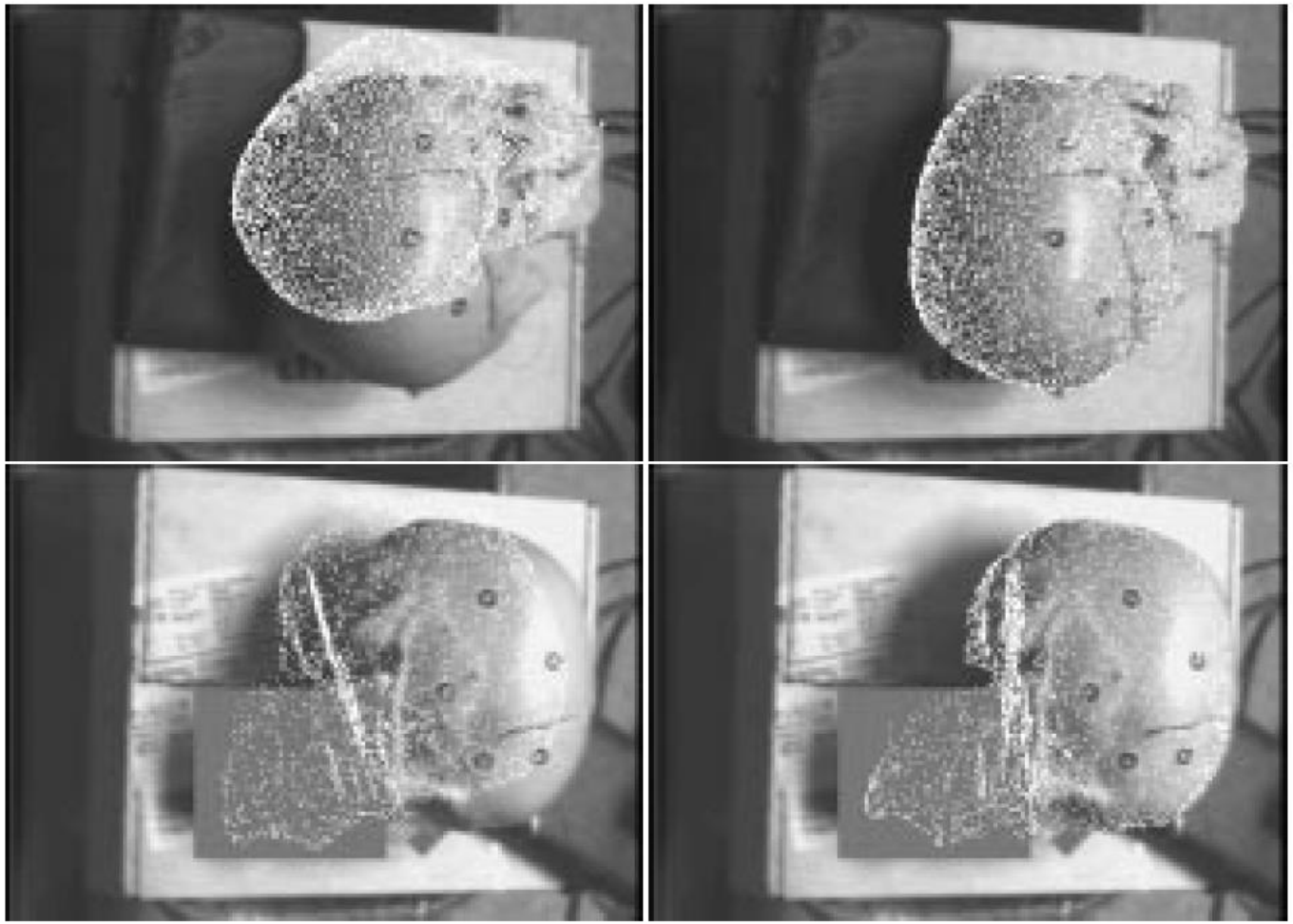
Application on MRI alignment



Proton-density image | **T2-Weighted image** | **Initial Alignment** | **Final Alignment**

Viola and Wells[9]

3D Objects alignment



Corsini et al.[10]

- Corsini et al.[10] extended this idea to other geometric properties related to illumination, such as *ambient occlusion* and *reflection directions*.

[10] M. Corsini, M. Dellepiane, F. Ponchio and R. Scopigno "Image-to-Geometry Registration: a Mutual Information Method exploiting Illumination-related Geometric Properties ", *Computer Graphics Forum*, Vol. 28(7), pp. 1755-1764, 2009.

Measure similarity with MI

- MI is the amount of information about *B* that *A* contains. Using the joint probability it can be expressed as:

$$\mathcal{I}(I_A, I_B) = -\sum_{(a,b)} \log p(a,b) \frac{p(a,b)}{p(a)p(b)}$$

Measure similarity with MI

Joint event (a,b) for the pixel (x0,y0)

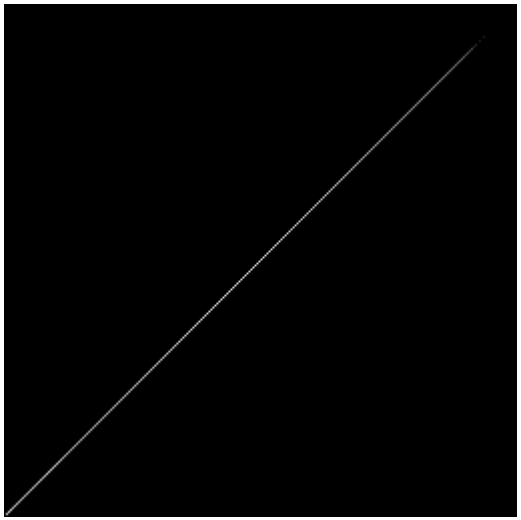
The joint events are stored in a *joint histogram*
The probability of the joint event p(a,b) is the number of occurrences divided the total number of pixels

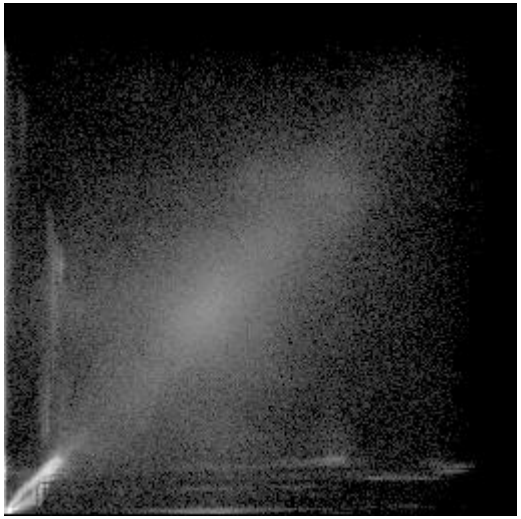
Joint Histogram

Image A

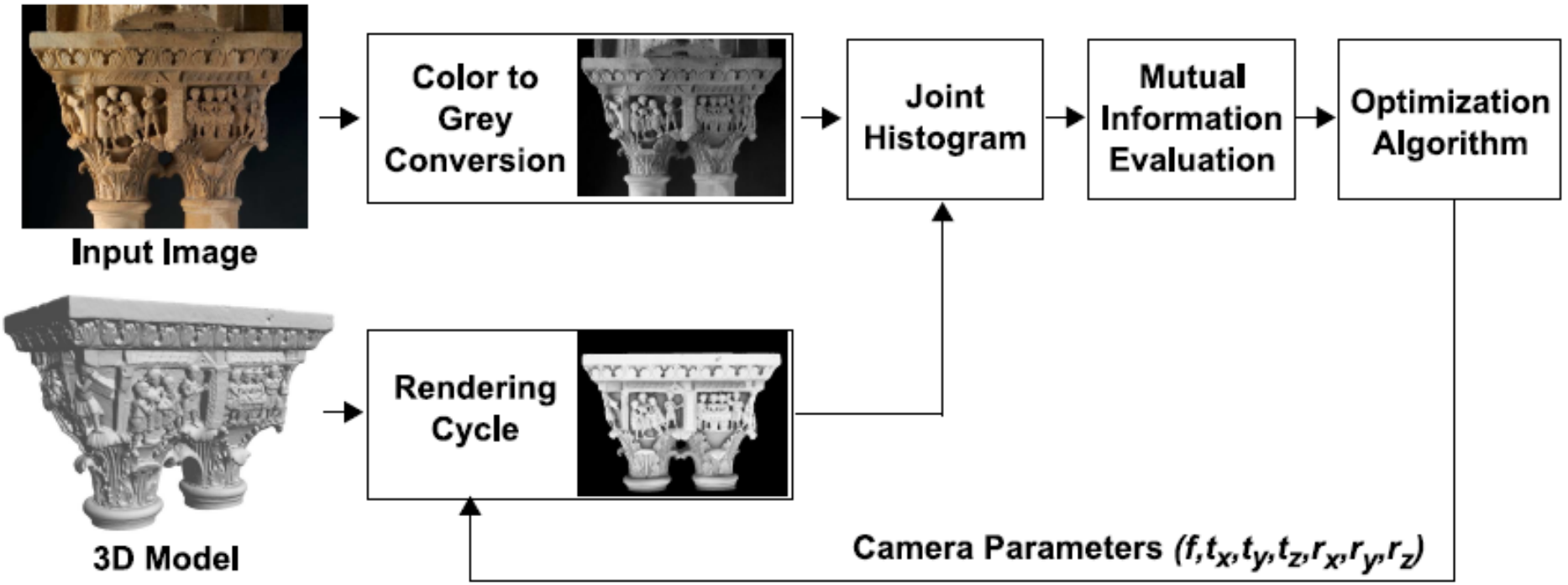
Image B

Joint Histogram





Overview of the algorithm



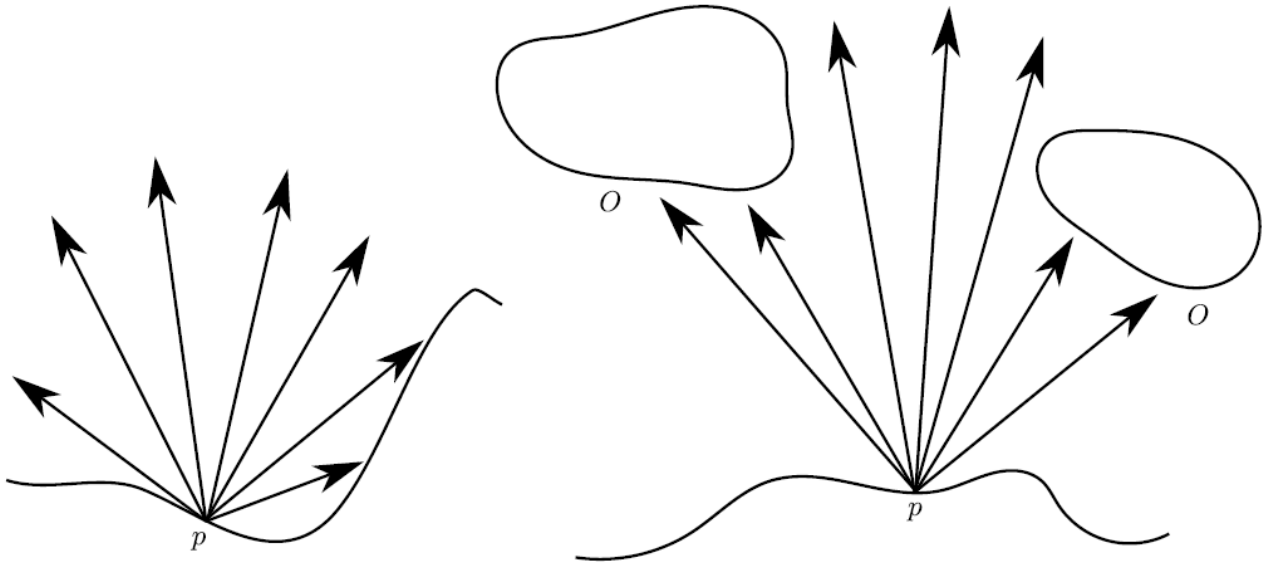
Normal Map

- Independently of the material behavior and the lighting conditions the surface normals correlate (in a nonlinear way) with the shading variations of the image.

Ambient Occlusion

- The idea is to compute a modulation coefficient to darken parts of the geometry which receive less light.

$$\mathcal{A}(p) = 1 - \frac{1}{2\pi} \int_{\Omega} V(p, \omega) \, (\boldsymbol{n}_p \cdot \omega) \, d\omega$$



Ambient Occlusion

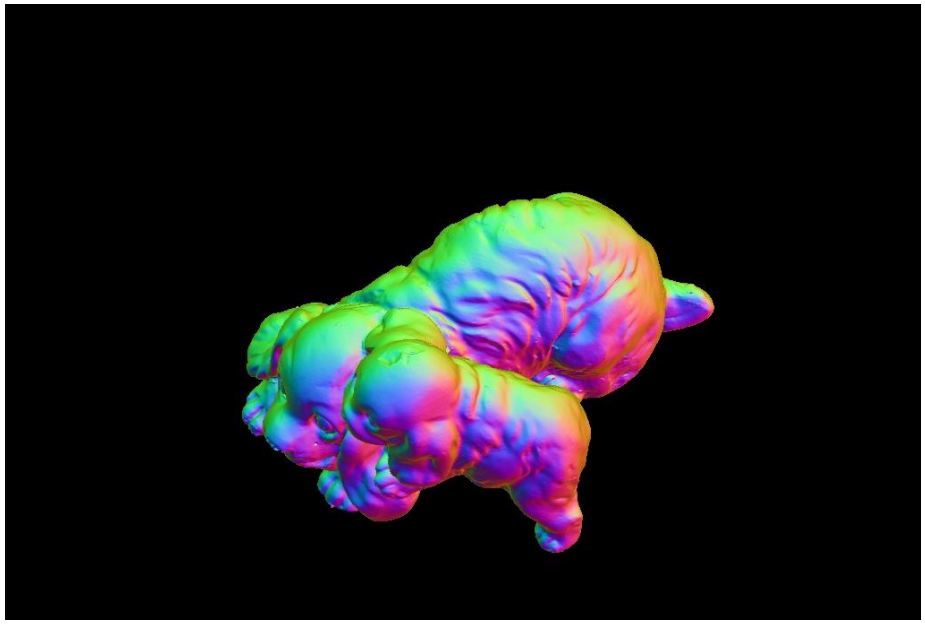
The 3D model is a simplified version of a scanning model of a capital courtesy of the *Kunsthistorisches Institut in Florenz* (*http://www.khi..it/*).

Ambient Occlusion Map

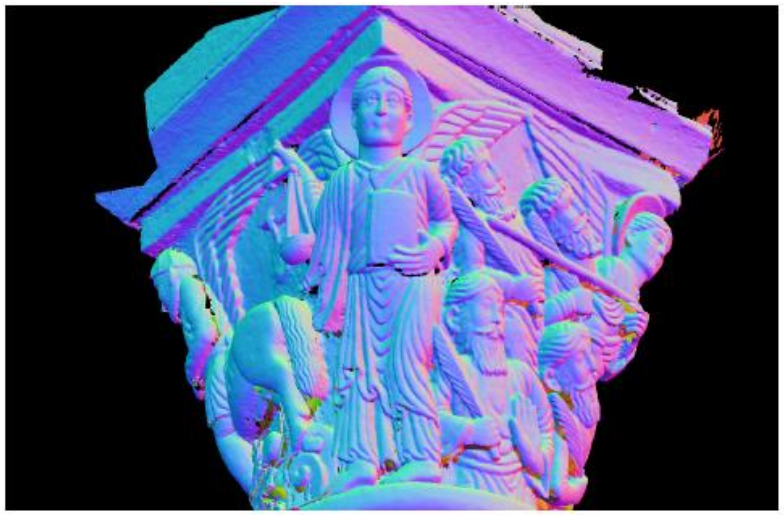
- Ambient occlusion → occluded parts of the geometry receive little light with respect to non-occluded parts. Hence, darker areas of the image may correlate well with not accessible parts of the surface.

Reflection Direction Map

- The direction of the mirror reflection can be computed for each pixel, knowing the surface normal and the viewer position → these directions may correlate well with the highlights of the image.

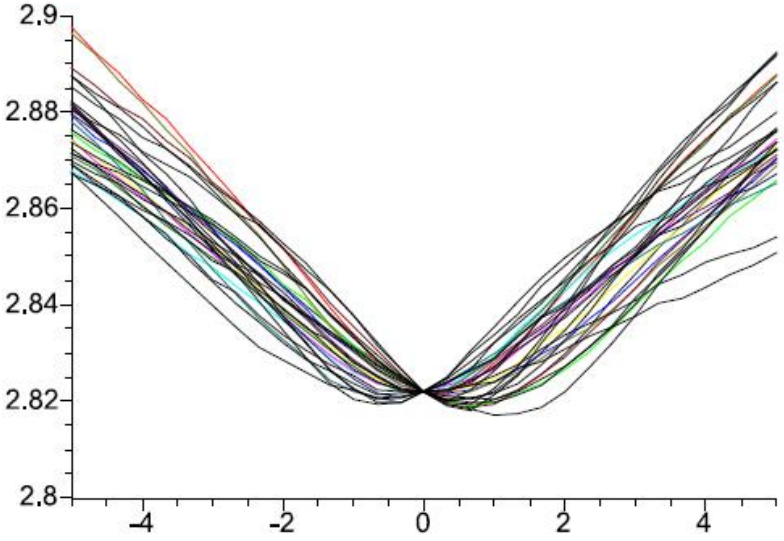


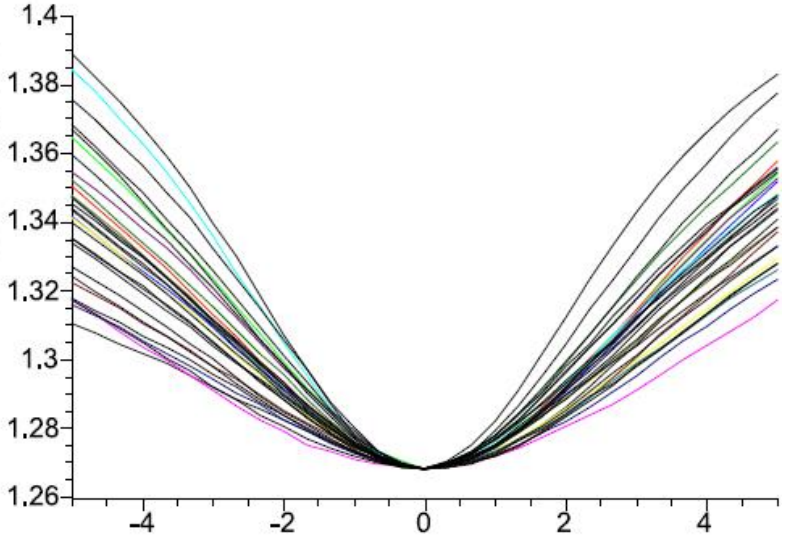
Convergence Analysis



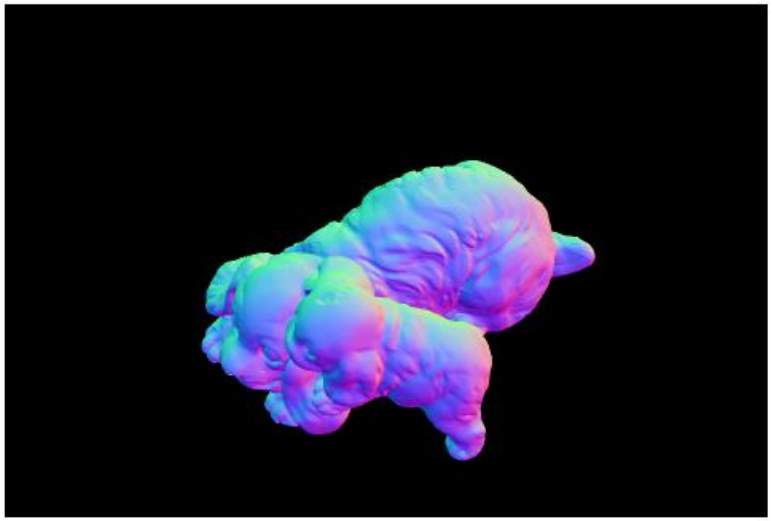
Normal

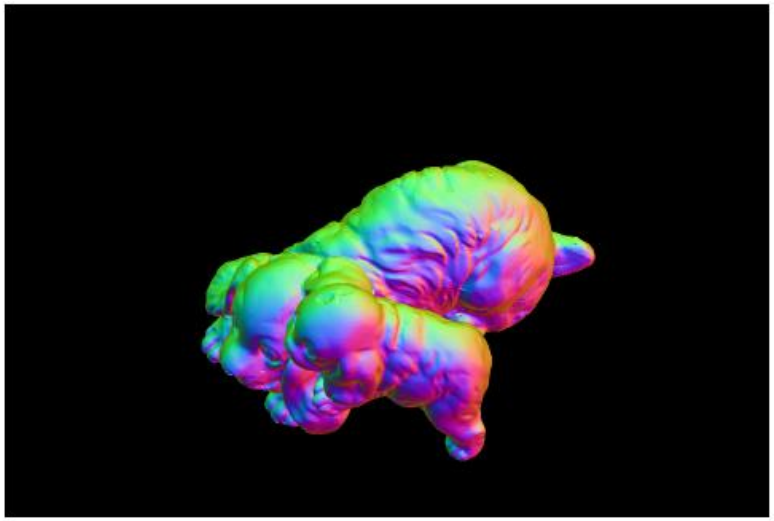
Ambient





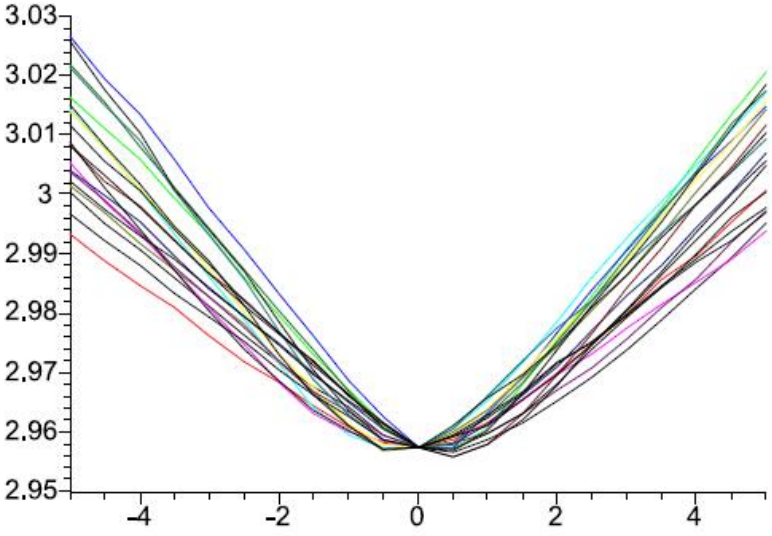
Convergence Analysis

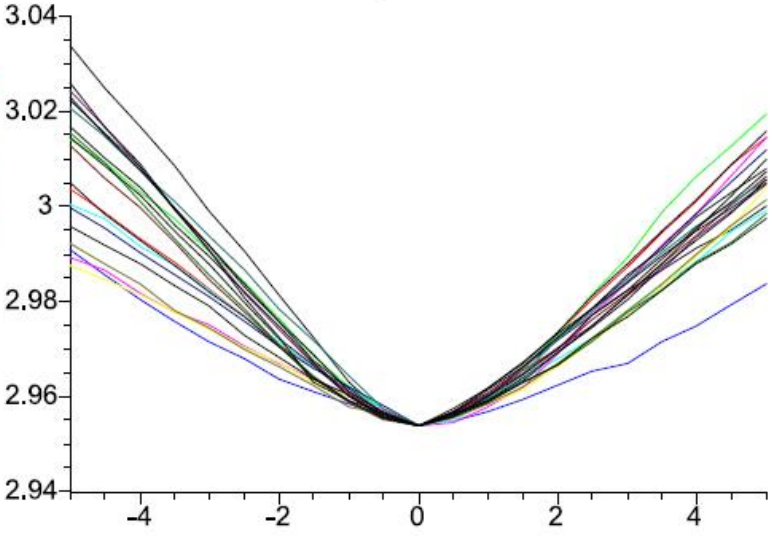




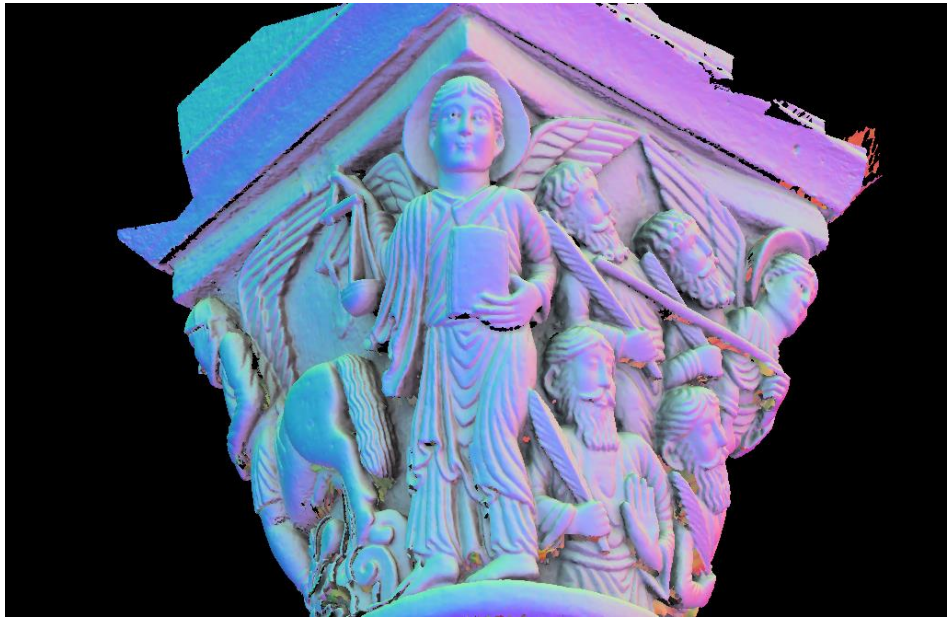
Normal

Specular





Combined Map



Normal map

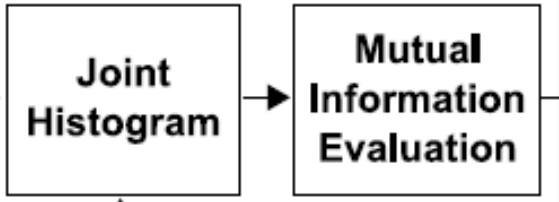
Ambient+Normal map

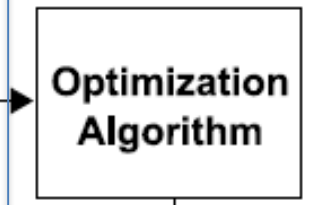
Better generalization than normal map/ambient occlusion only

Optimization algorithm

Input Image

3D Model





Camera Parameters $(f, t_x, t_y, t_z, r_x, r_y, r_z)$

NEWUOA (nonlinear optimization algorithm)

M. J. D. Powell, "The NEWUOA software for unconstrained optimization without derivatives", Large-Scale Nonlinear Optimization - Nonconvex Optimization and Its Applications, Vol. 83, pp. 255-297, 2006.

Results

- Let’s me show you a video..

Multi-view Methods

- *Multi-view methods exploit the information of all the image (not a single image at the time) to obtain a simultaneous alignment of all the image.*
- In general, these methods are more robust than the ones that align one image at a time.

Multi-view registration methods

- Liu et al.[11] improved the performance of its previous system [7] to include models-SFM registration to allow the alignment of the non-registered images.
- Stamos et al.[12] improved again this system by relaxing the orthogonality constraints (circular architectural features are identified and matched).

[11] L. Liu, I. Stamos, G. Yu, G. Wolberg, and S. Zokai, "Multiview geometry for texture mapping 2D images onto 3D range data", *Proc. of CVPR'06.* Vol. 2, pp. 2293–2300, 2006.
[12] I. Stamos,, L. Liu, C. Chen, G. Wolberg, G. Yu and S. Zokai, "Integrating automated range registration with multiview geometry for the photorealistic modeling of large-scale scenes", *Int. J. of Computer Vision*, Vol. *78*, pp. 237–260, 2008.

Multi-view registration methods

- Pintus et al.[13] proposed a semi-automatic similar system (without the fine alignment) but with manual intervention.
- Corsini et al.[14] extended 4PCS and their previous MI-based method [10] to robustly align a set of images on a 3D model.

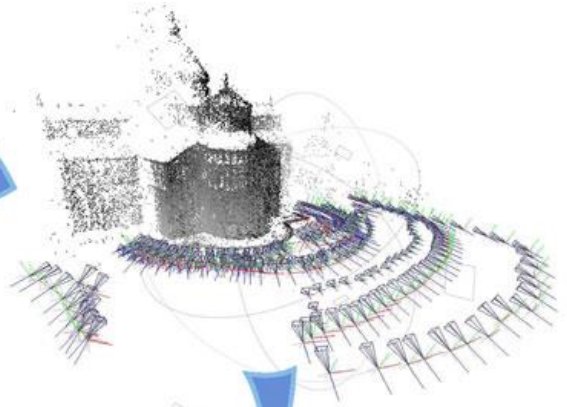
[13] R. Pintus, E. Gobbetti, and R. Combet "Fast and robust semi-automatic registration of photographs to 3D geometry", *Proc. of the 12th international symposium on virtual reality, archaeology and cultural heritage,* 2011.

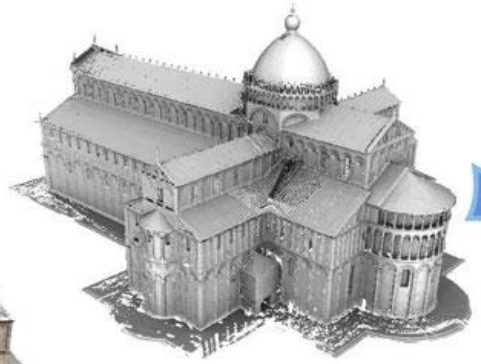
[14] M. Corsini, M. Dellepiane, F. Ganovelli, R. Gherardi, A. Fusiello, and R. Scopigno, "Fully Automatic Registration of Image Sets on Approximate Geometry", *Int. J. Comput. Vision* Vol. 102(1-3), pp. 91-111, March, 2013, 91-111.

The multi-image method

Input Images

Structure from Motion (SfM)

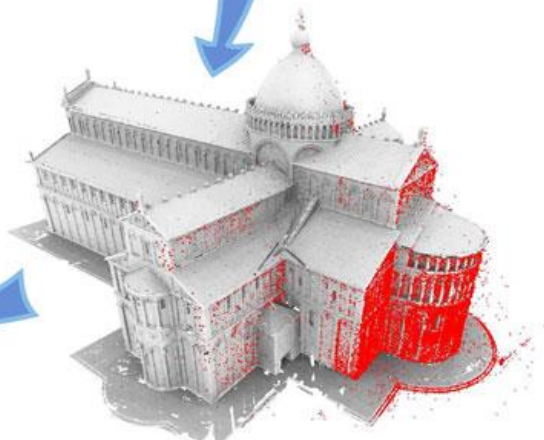




Input 3D Model

Point cloud to 3D model Alignment

4PCS extension



MI-based method [10] extension

Global Refinement

Final Result (model with color)

Point Cloud-to-3D Model Alignment

- An extension of the 4PCS algorithms has been developed to deal with the *problem of scale* → reconstructed 3D points (output of the SFM module) and the 3D model have different, unknown, scale factors.

Mutual Information Extension

- Projected colors are used in the MI-based registration framework described before ([10]).
- Global alignment error is reduced following an approach originally developed for range maps [Pulli1999]

[Pulli1999] Pulli, K, "Multiview registration for large data sets", *Proc. of the 2nd international conference on 3-D digital imaging and modeling (3DIM'99),* pp. 160–168, 1999.

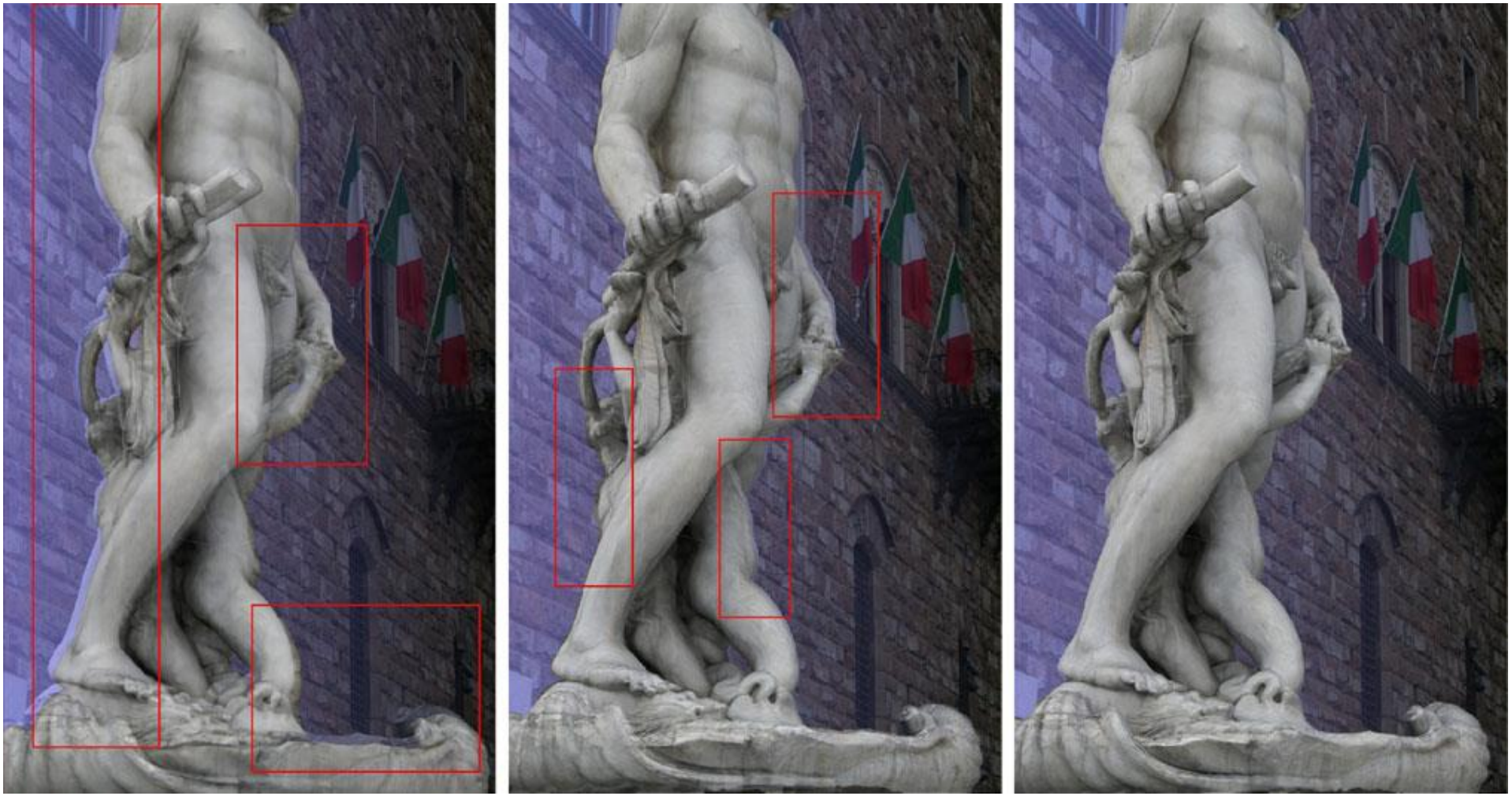
Graph-based optimization approach (Pulli1999)

- A graph is build such that:
  - A node for each image
  - An arc for overlapping image (weighted by MI for that pairs and other factors related to the amount of overlapping)
- The algorithm by Pulli et al. (Pulli1999) distributes the alignment error by re-registering the neighbors starting from the *most important"* node.

Some results

Some results

Global Refinement



Recap

- Many 2D/3D registration methods have been developed.
- Many of them rely on additional information (e.g. pre-registered images, reflectance information) or are specific for certain 3D objects (e.g. urban scenes).
- Typically, multi-view methods are the most general and robust.

Questions ?