Relative pose estimation and fusion of 2D spectral data and 3D point clouds

Zoltan Kato
University of Szeged, Hungary

Content

1. Introduction
2. Camera geometry & calibration
3. Perspective camera calibration without special target or point correspondences
4. Omnidirectional camera pose estimation
5. CH use case

+ Demo & Lab. in the afternoon by Robert Frohlich

Problem statement

- Fusion of depth sensor images (e.g. Lidar) with perspective camera pictures are widely used to construct rich 3D models or for environment mapping.
- Independent sensory data must be transformed into a common coordinate frame
  - Lidar: 3D depth (geometry) of the scene
  - Perspective: 2D projection (radiometry) of the scene
- Either pose of full camera matrix needs to be estimated w.r.t. the 3D data.

Cultural heritage applications

COLOR & SPACE IN CULTURAL HERITAGE

- 3D photorealistic models of CH objects
  - Documentation and visualization
- Reconstruct historical state of objects/buildings using archive (2D) photographs and 3D scans
  - 3D visualization of lost surface paintings/ornaments
- Use 3D data as a common reference for multimodal 2D measurements
  - analyse multiple aspects of the same object (or parts)
  - joint evaluation of various measurements

3D photorealistic models

- Given the 3D geometry of an object obtained from a 3D depth sensor
- and 2D color calibrated images of the object
- Produce a high quality 3D textured model of the object
  - CH documentation
  - Visualization for the public

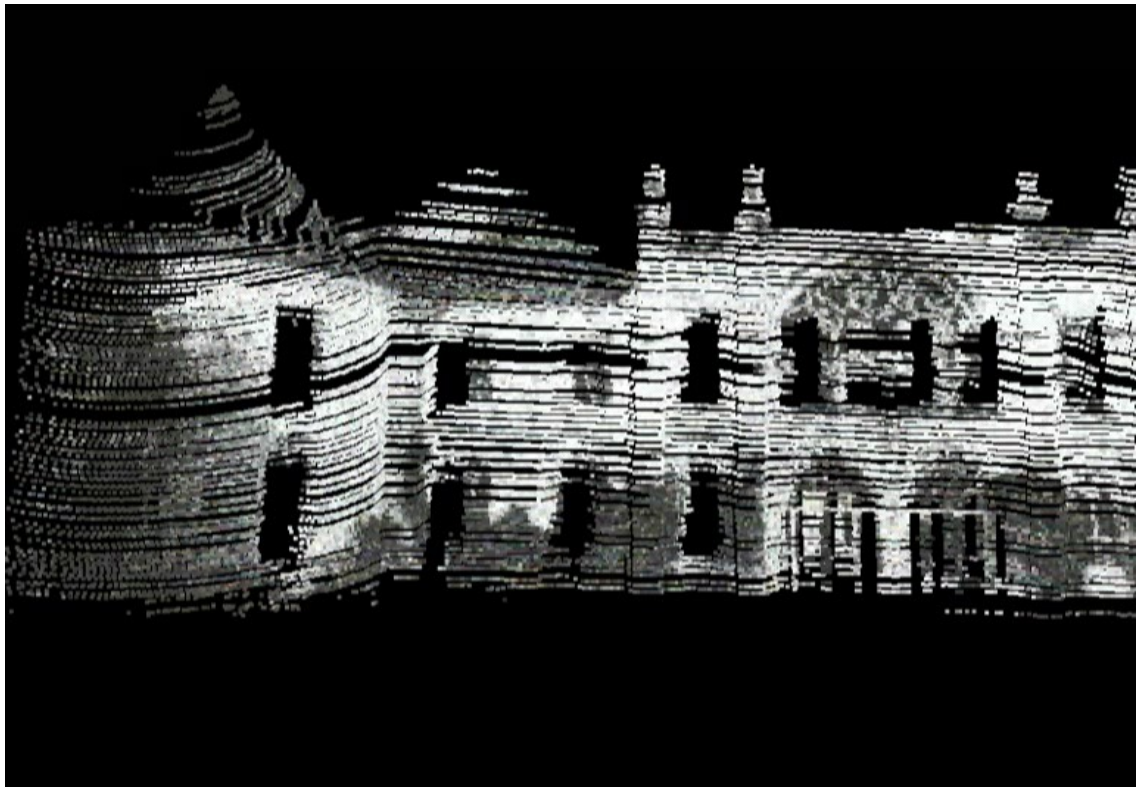
Cultural heritage applications

COLOR & SPACE IN CULTURAL HERITAGE

- 3D photorealistic models of CH objects
  - Documentation and visualization
- Reconstruct historical state of objects/buildings using archive (2D) photographs and 3D scans
  - 3D visualization of lost surface paintings/ornaments
- Use 3D data as a common reference for multimodal 2D measurements
  - analyse multiple aspects of the same object (or parts)
  - joint evaluation of various measurements

3D depth data & 2D archive images

COLOR & SPACE IN CULTURAL HERITAGE



Archive photo backprojected onto today's 3D structure

Archive image taken in 1935 (unknown camera)

How it looks today (2010)

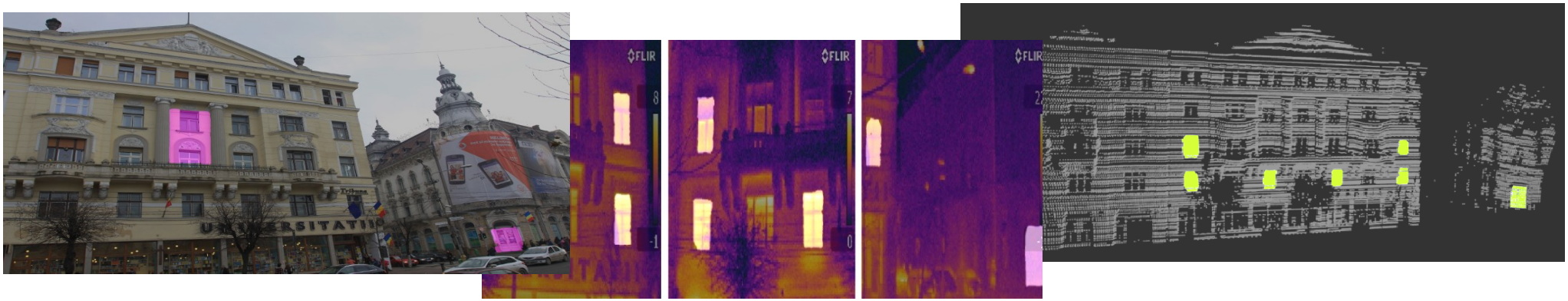
Cultural heritage applications

COLOR & SPACE IN CULTURAL HERITAGE

- 3D photorealistic models of CH objects
  - Documentation and visualization
- Reconstruct historical state of objects/buildings using archive (2D) photographs and 3D scans
  - 3D visualization of lost surface paintings/ornaments
- Use 3D data as a common reference for multimodal 2D measurements
  - analyse multiple aspects of the same object (or parts)
  - joint evaluation of various measurements

Fusion of RGB, IR & Lidar data

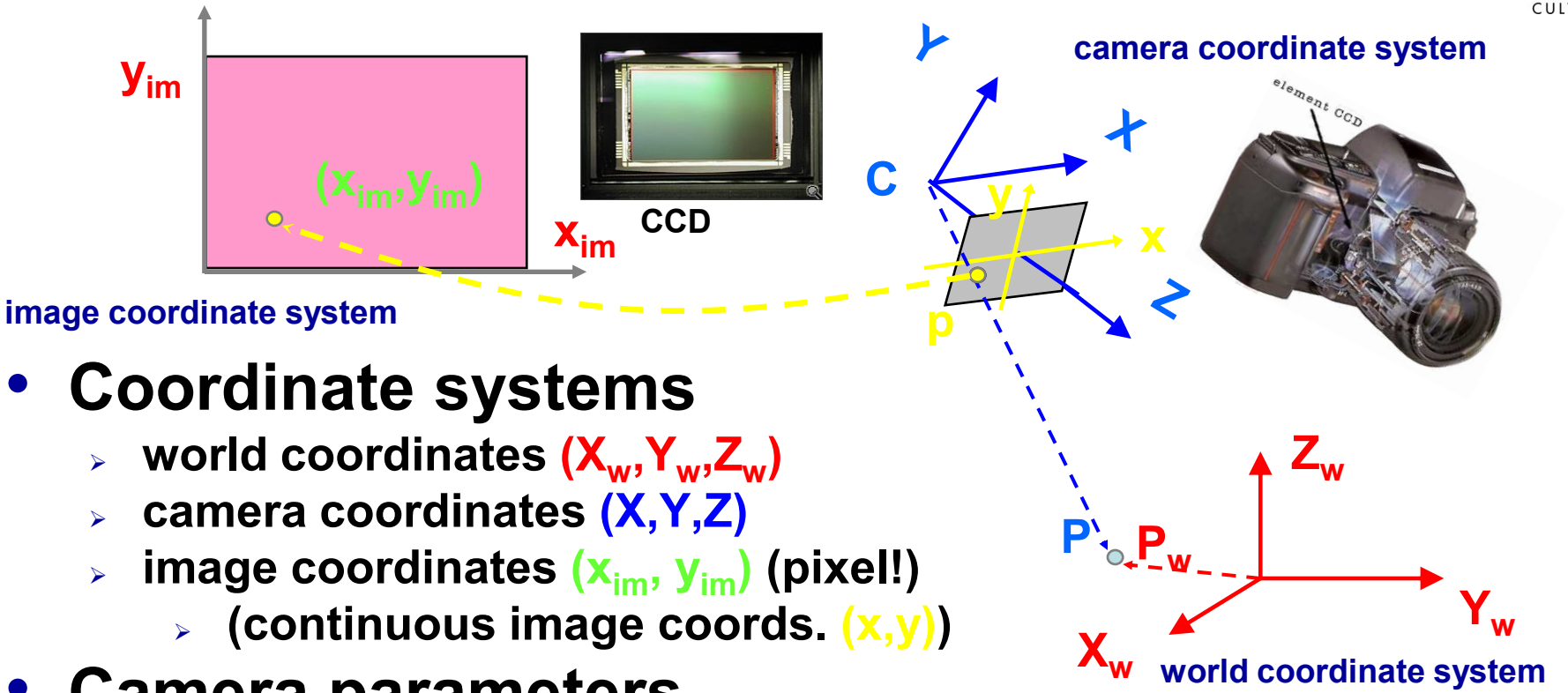
COLOR & SPACE IN CULTURAL HERITAGE



Standard textbook material

CAMERA INTERNAL CALIBRATION AND POSE

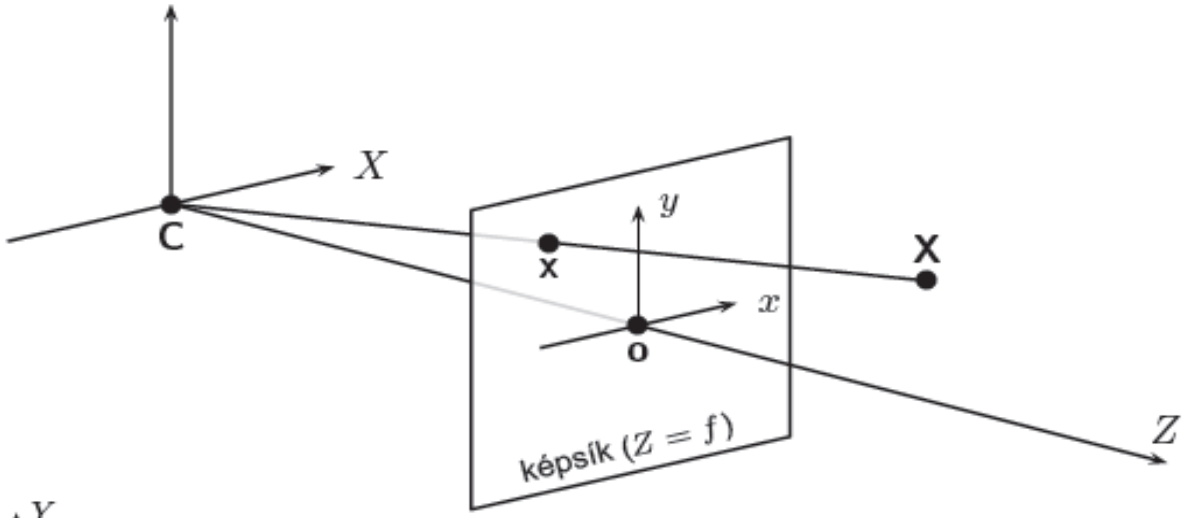
Camera parameters

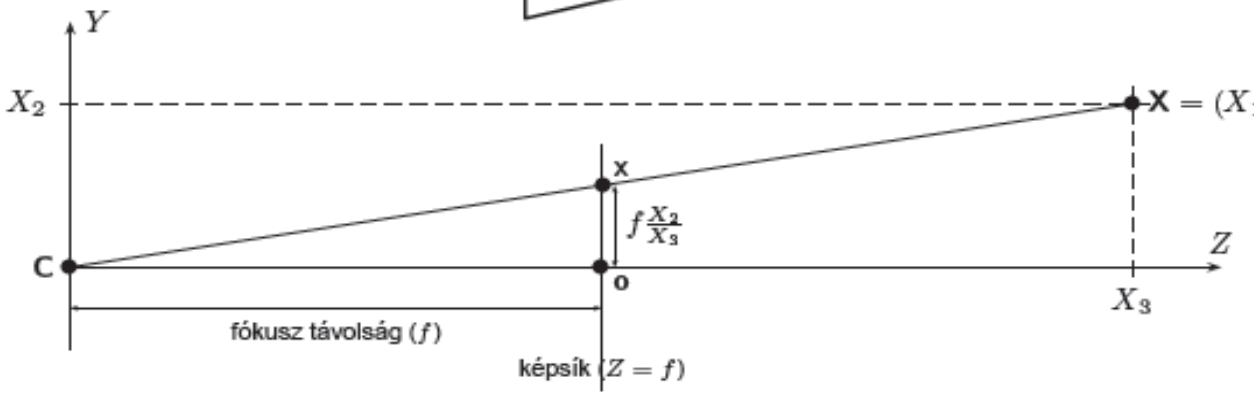


- **Coordinate systems**
  - world coordinates (X_w,Y_w,Z_w)
  - camera coordinates (X,Y,Z)
  - image coordinates (x_{im}, y_{im}) (pixel!)
    - (continuous image coords. (x,y))
- **Camera parameters**
  - **Intrinsic parameters** : 3D→2D map from the camera coordinate frame into the image coordinate frame.
  - **Extrinsic parameters (pose)**: the position and orientation of the camera coordinate frame w.r.t. world coordinate frame (3D translation and rotation)

3D point (X) and its image (x)

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fX_1 \\ fX_2 \\ X_3 \end{pmatrix} = \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{pmatrix}$$

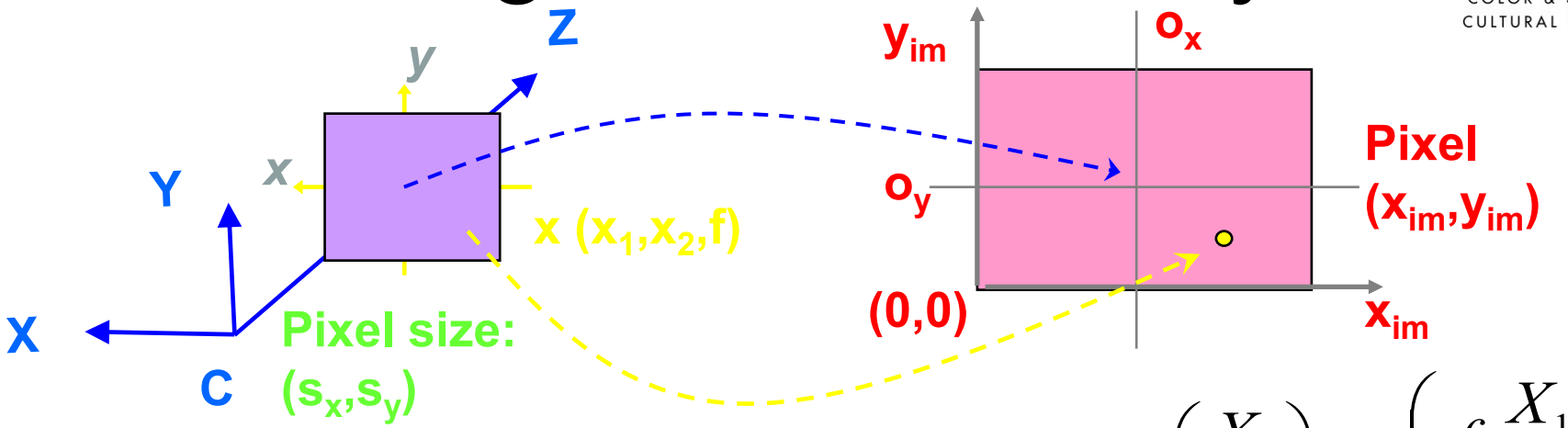




- Nonlinear map in the Euclidean space!
- Homogeneous coords. → projective space, where it is a linear map.
- We get the image points in a 2D euclidean coordinate frame whose
  - Origin is *o* camera center,
  - *x* & *y* axes are || with the camera coordinate frame's *X* & *Y axes.*

$$\frac{X_2}{X_3} = \frac{x_2}{f} \Rightarrow x_2 = f\frac{X_2}{X_3}$$

x in the image coordinate system



- Intrinsic parameters
  - (o_x, o_y): camera center (~image center)
  - (s_x, s_y): pixel size
  - f: focal length

$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \mapsto \begin{pmatrix} f\dfrac{X_1}{X_3} \\ f\dfrac{X_2}{X_3} \end{pmatrix}$$

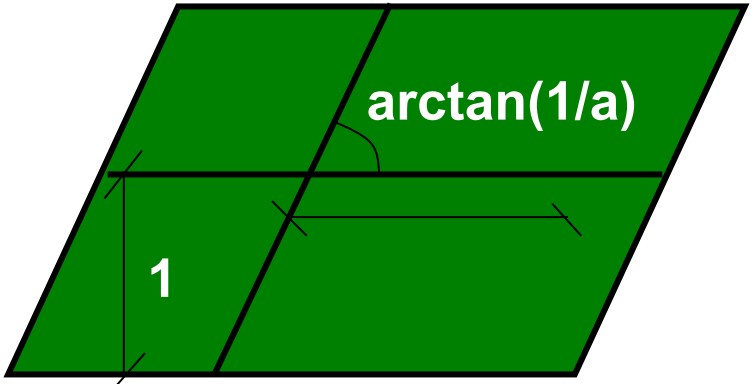
$$\begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} f_x X_1 + X_3 o_x \\ f_y X_2 + X_3 o_y \\ X_3 \end{pmatrix} = \begin{bmatrix} fs_x & & o_x & 0 \\ & fs_y & o_y & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{pmatrix}$$

Slide adopted from Zhigang Zhu Computer Vision - CSC I6716

Calibration matrix K

$$\mathbf{K} = \begin{bmatrix} f_x & a & o_x \\ & f_y & o_y \\ & & 1 \end{bmatrix}$$

$$\mathbf{P} = [\mathbf{K}|\mathbf{0}] = \mathbf{K}[\mathbf{I}|\mathbf{0}]$$



a ➔ skew
For CCD/CMOS: **a=0**

- P is a 3×4 homogeneous camera matrix
- 3×3 upper triangular K is the *camera calibration mxtrix*
  - 5 intrinsic parameters: f_x, f_y, a, o_x, o_y
  - Generic pixels: parallelogram
  - Rectangular pixels: a=0
  - Square pixels: $f_x = f_y$, a=0

Camera calibration

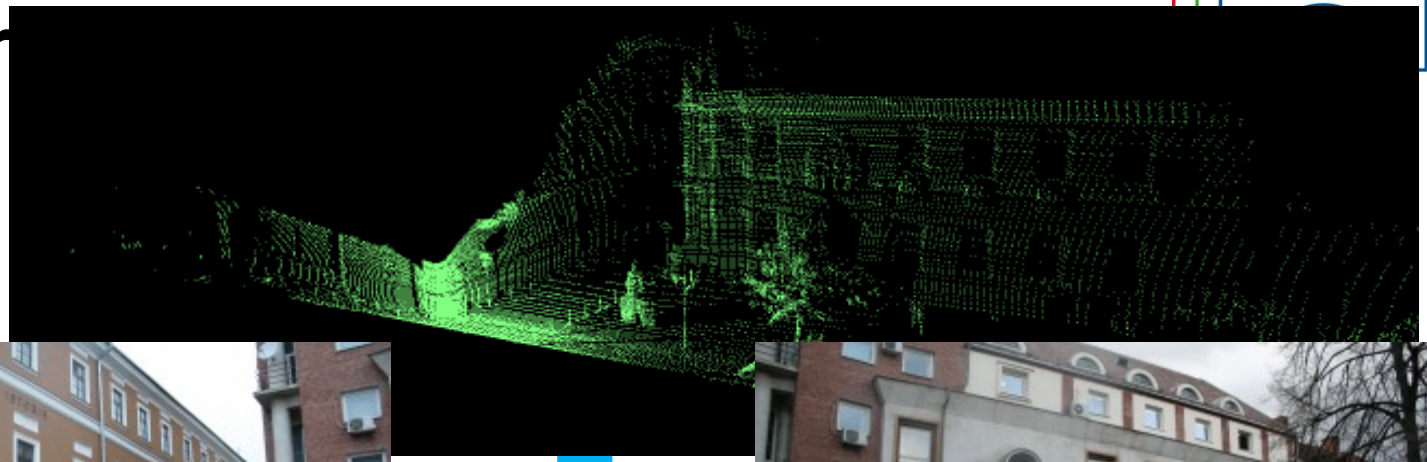
- Estimating the internal parameters (i.e. K) can be done with a calibration pattern and appropriate programs
  - Matlab Calibration Toolbox - http://www.vision.caltech.edu/bouguetj/calib_doc/
  - If optics do not change (no zooming either) then it has to be done only once – K remains constant.
- Extrinsic calibration (***pose estimation***) has to be done using the actual image only!

with Levente Tamas

TARGETLESS CALIBRATION OF A LIDAR-PERSPECTIVE CAMERA PAIR

Pr

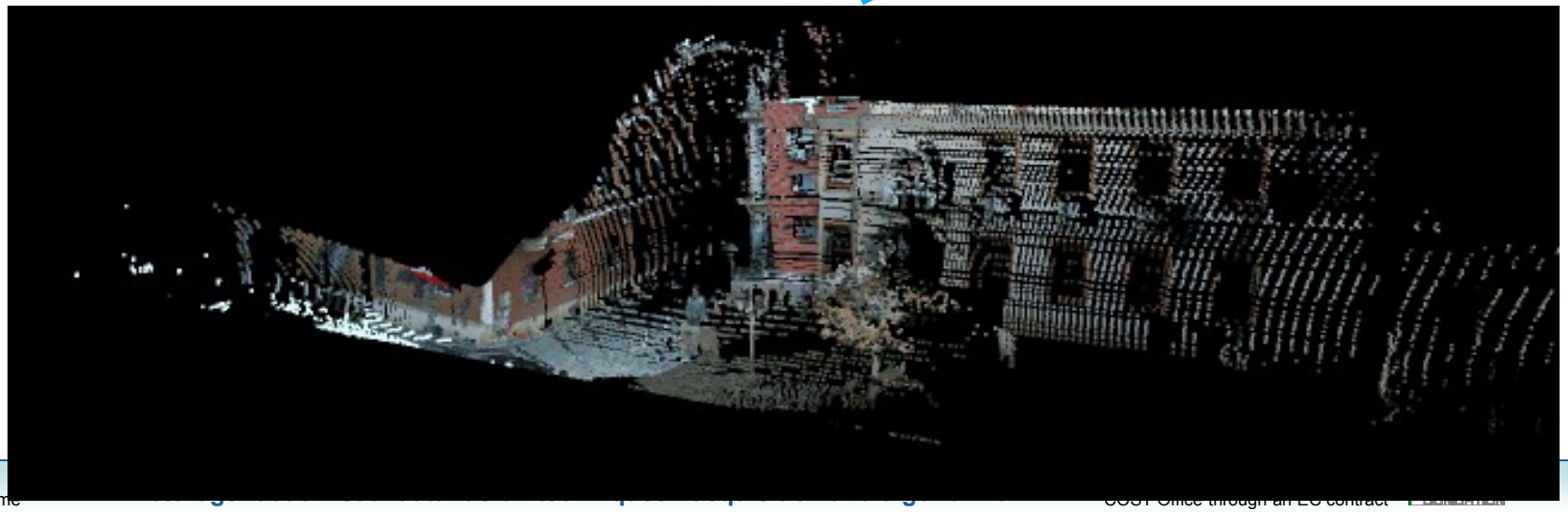
Depth image (Lidar)



2D perspective images

Pose estimation/ calibration

Fused 3D/2D image



COST is supported by the EU RTD Framework Programme

Challenges

- Different functionality of sensors yields both radiometrically and geometrically unrelated data:
  - Lidar results in a 3D point cloud with depth only data or depth+intensity data in each point
  - Perspective camera results in a regular 2D image with a radiometric measurement (intensity, RGB, Infrared, etc.) at each pixel
- Matching and correspondence hard to solve

Calibration problem formulation

- Lidar provides 3D measurements which can be expressed as 3D point coordinates **X** in Lidar's 3D coordinate system ➔ world coordinate system
- Perspective camera projection from world to image plane: $\mathbf{x} = \mathbf{PX}$
- **P** 3X4 camera matrix can be factorized as

$$\mathbf{P} = \mathbf{KR}[\mathbf{I} \mid \mathbf{t}]$$

  - when calibration matrix **K** is known ➔ determine extrinsic parameters **R**,**t** (*pose estimation*)
    - 6 unknowns
    - Need at least 1 planar region
  - when **K** is unknown ➔ determine intrinsic & extrinsic parameters
    - 11 unknowns
    - Need at least 2 non-coplanar planes

Typical solutions

- Calibration target-based solutions
  - Markers visible in both sensors
  - Requires special target and setup at location
- Point correspondences
  - Lidar intensity ←→ RGB image feature matching
  - point correspondences for standard PnP problem
  - Requires a Lidar which records intensity too
- GPS, IMU or other external data source
  - not precise enough for accurate calibration
  - Good for an initial alignment

Proposed solution

- Formulate calibration as a 3D → 2D registration problem which works **without**
  - special target
  - point correspondences
  - additional intensity data from Lidar
- The registration is solved as a nonlinear system of equations
  - Equations are constructed by integrating nonlinear functions over corresponding planar regions

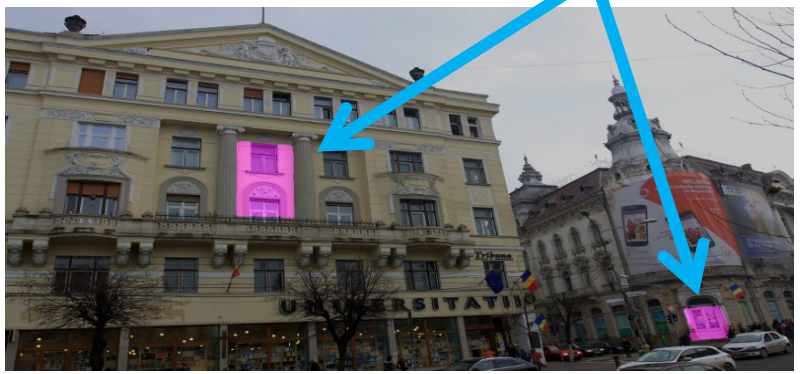
Solution without correspondences

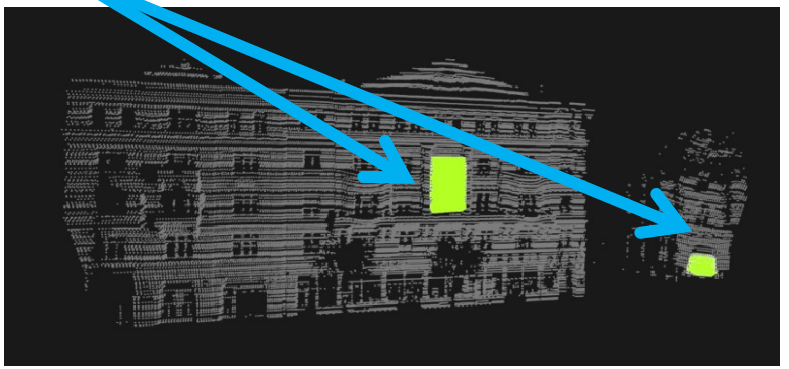
Identity relation for each Lidar-camera point pair

$$\mathbf{x} = \mathbf{PX}$$

integrate out individual point matches over segmented regions

$$\int_D x dx = \int_{P\mathcal{F}} z dz$$

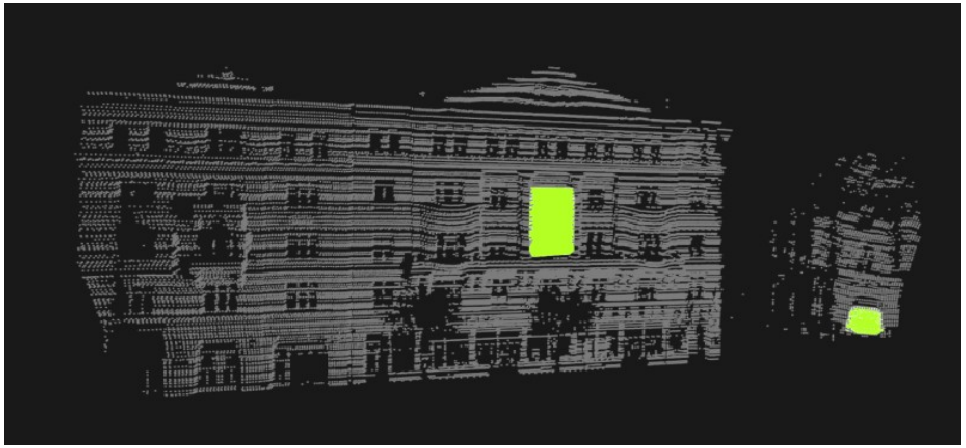




Provides only 2 equations but more is needed!

Regions instead of points

- Instead of extracting keypoints use planar regions
  - planar ➔ no occlusion!!
    - But “smooth enough” is OK if no occluded points!
  - segmentation may be available in a processing pipeline
  - Otherwise planar regions are easy to find in Lidar data
    - Corresponding regions in 2D image?
      - A 1 clique user input is enough to match + initialize a region growing algorithm



Generate equations

- Identity relation remains valid, when an ω function is applied to both sides
  - yields the following form of the integrals:

$$\int_D \omega(x)dx = \int_{P\mathcal{F}} \omega(z)dz$$

- generate sufficiently many equations by making use of a set of non-linear functions
  - Generate overdetemined system and find an LSE solution

C. Domokos, J. Nemeth, and Z. Kato. **Nonlinear Shape Registration without Correspondences**. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(5):943--958, May 2012

Efficient implementation

- Computing integrals at each iteration might be expensive
  - 2D image: pixel representation
  - 3D data: triangulated (planar) surface
- Make use of polynomial $\varpi_i = x_1^{n_i} x_2^{m_i}$ functions
  - yields an efficient computational scheme
    - Left hand side: constant (compute only once)
    - Right hand side: recursive formula exists for Δs.

$$\int_{\mathcal{D}} x_1^{n_i} x_2^{m_i} dx =$$

$$\int_{P\mathcal{F}} z_1^{n_i} z_2^{m_i} dz \approx \sum_{\forall \Delta \in \mathcal{F}^{\Delta}} \int_{\Delta} z_1^{n_i} z_2^{m_i} dz$$

Algorithm overview

Input: 3D point cloud and 2D binary image representing the same region (and the calibration matrix **K** if available)
Output: Parameters of the camera matrix **P**

1. Normalize the input 3D points into the unit cube and the 2D points into the unit square centered in the origin.
2. Triangulate the region represented by the 3D point cloud
3. Construct the nonlinear system of equations
4. Initialize the camera matrix as $\mathbf{P} = \mathbf{K}[\mathbf{I} \mid \mathbf{0}]$
5. Solve the nonlinear system of equations using the Levenberg-Marquardt algorithm
6. Unnormalize the solution

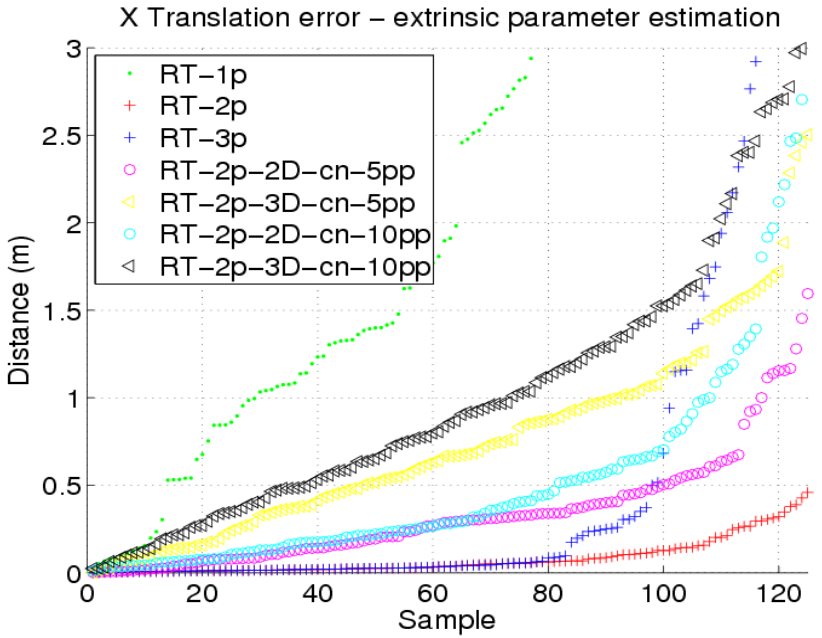
Quantitatvie evaluation

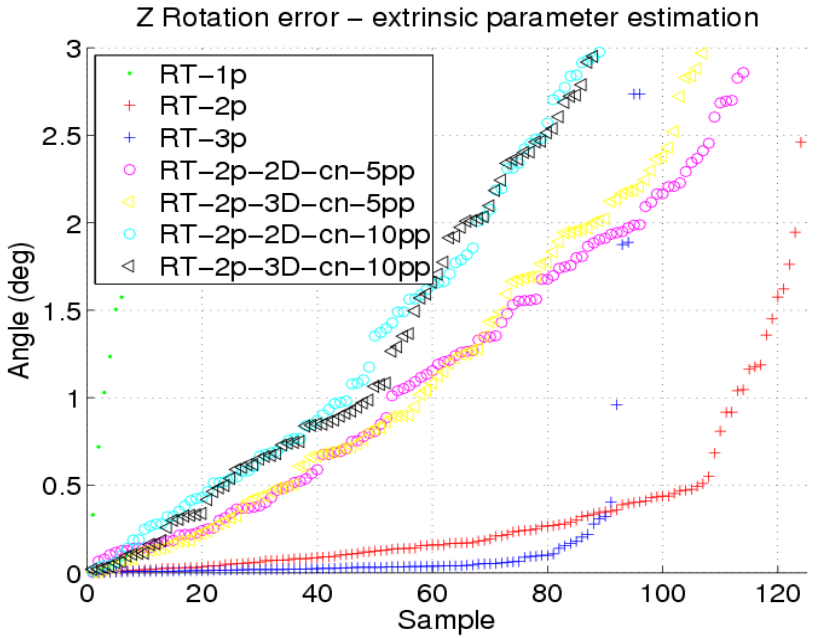
- Synthetic 2D-3D dataset with typical street scene settings (target region is about 1m)
- More than 2500 data pair registration test:
  - Extrinsic parameter estimation
  - Intrinsic-extrinsic parameter estimation
  - Segmentation error robustness test

5% 10%

Extrinsic parameter estimation

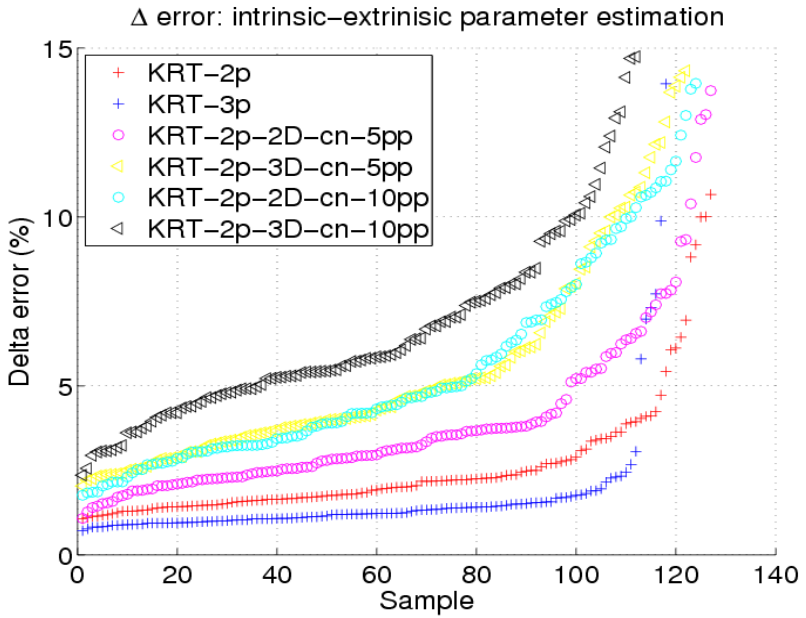
- More than +/- 45º rotation,
- Random 2-10m translation
- Stable and robust results with 2 planes:
  - in 70% of the cases less than 0.5º/3cm error





Intrinsic-extrinsic estimation

- With at least 2 different planes
- Focal length in range of 400-1600
- in 70% cases < 5% backprojection error



Comparative results on synthetic data

Method	*Backprojection error (%)*	*Runtime (s)*
[HSA2012] - 6 regions	7.4	8.1
[RU2005] - 3 regions	8.3	15.9
[RU2005] - 6 regions	1.6	32.7
Proposed - 3 regions	0.9	54.2
Proposed - 6 regions	0.8	49.6

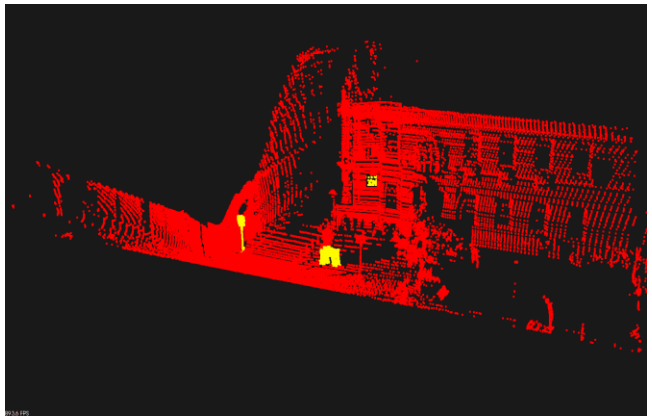
- [HSA2012]: min. 6 planes, sensitive to errors in point cloud
- [RU2005]: performance drops for <6 planes, only pose estimation

[**HSA2012**] H. S. Alismail, L. D. Baker, and B. Browning. Automatic calibration of a range sensor and camera system. In *Second Joint 3DIM/3DPVT Conference: 3D Imaging, Modeling, Processing, Visualization and Transmission*, pages 286–292, Zurich, Switzerland, October 2012. IEEE

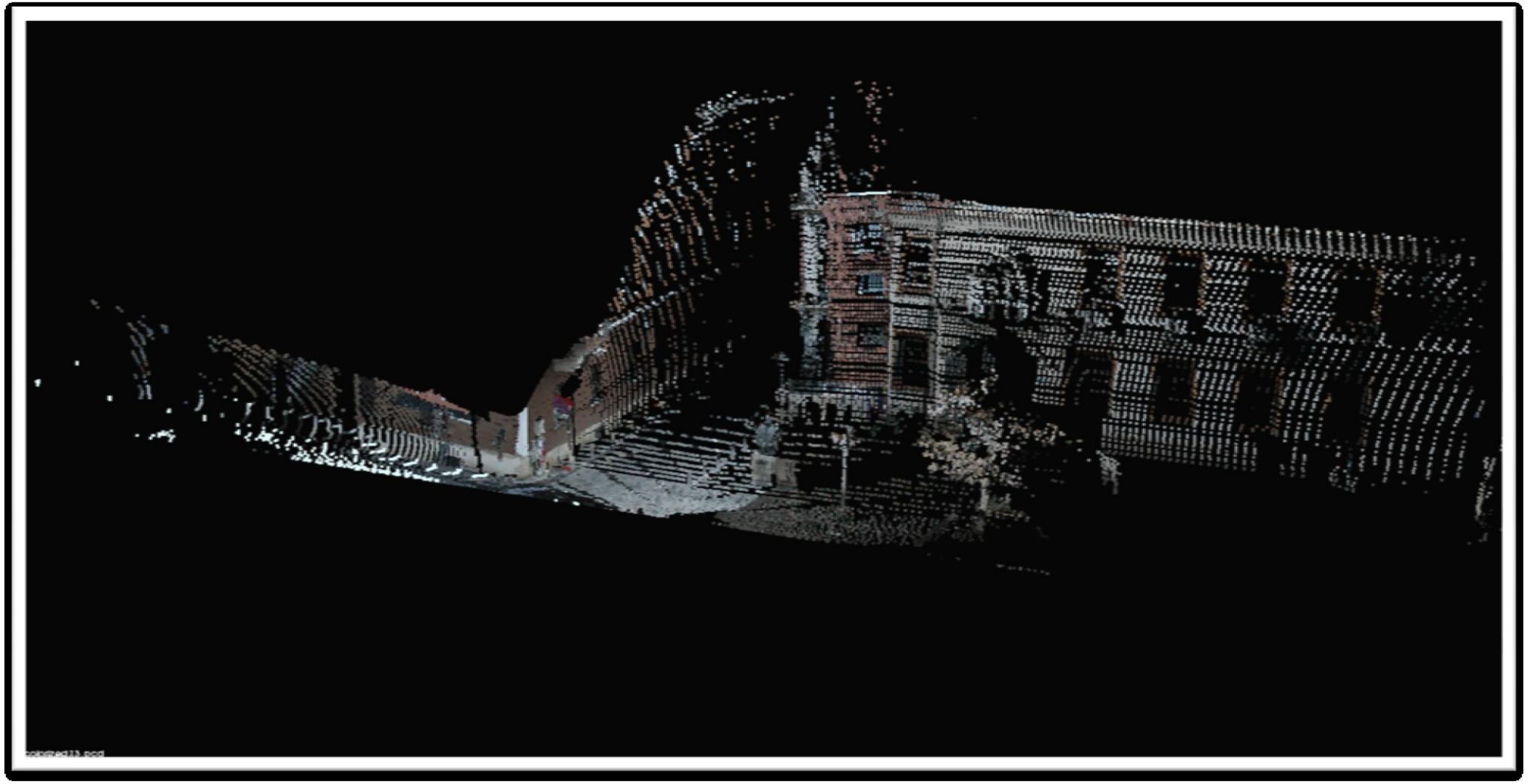
[**RU2005**] R. Unnikrishnan and M. Hebert. Fast extrinsic calibration of a laser rangefinder to a camera. Technical report, Carnegie Mellon University, 2005

Real data experiments

- With our custom 3D and RGB camera
  - 0.5º rotation and 1cm depth error in 3D data
  - Standard 5Mpixel RGB camera
  - Automatic segmentation
  - 1 Lidar + 2 RGB images

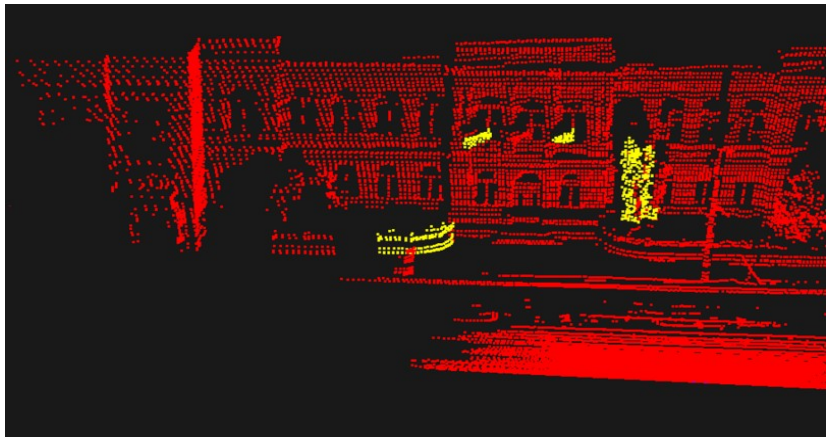


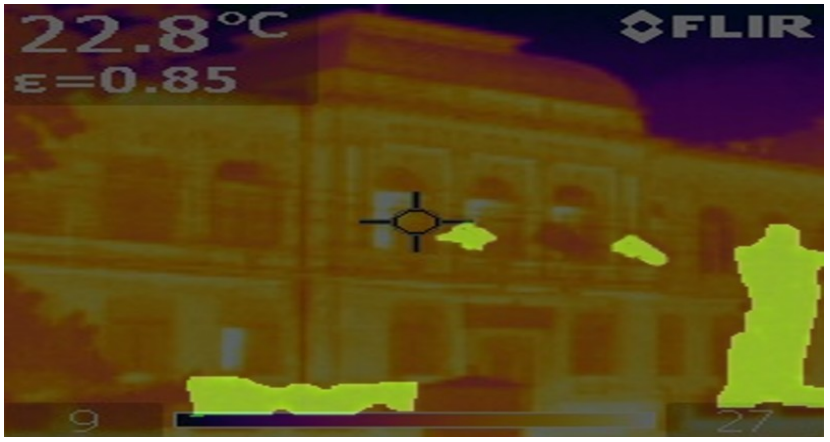
3D and RGB camera fused output



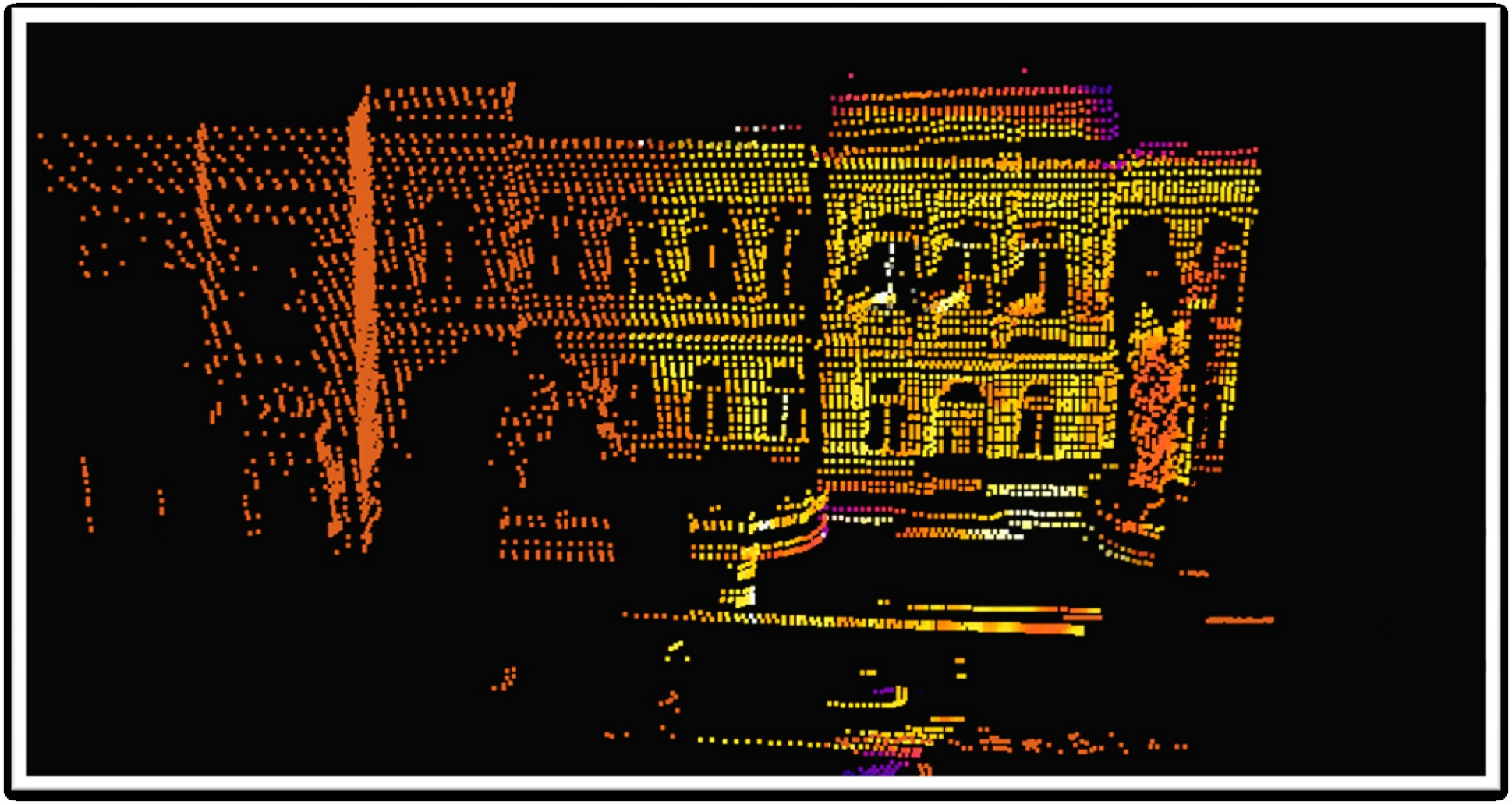
Real data experiments

- With our custom 3D and IR camera
  - 0.5º rotation, 1cm depth error in 3D data
  - Flir IR camera with 240x240 resolution
  - Image taken in night time
  - Large displacement & rotation



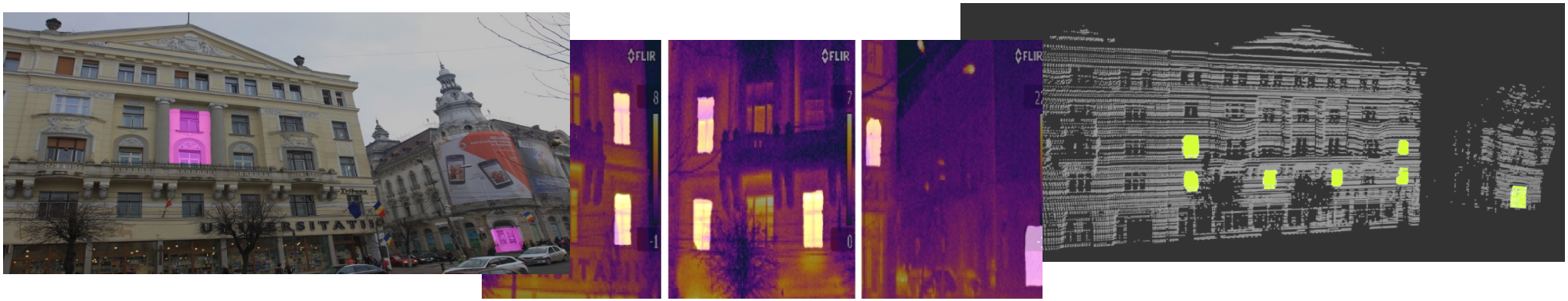


3D and IR camera fused output



Fusion of RGB, IR & Lidar data

COLOR & SPACE IN CULTURAL HERITAGE



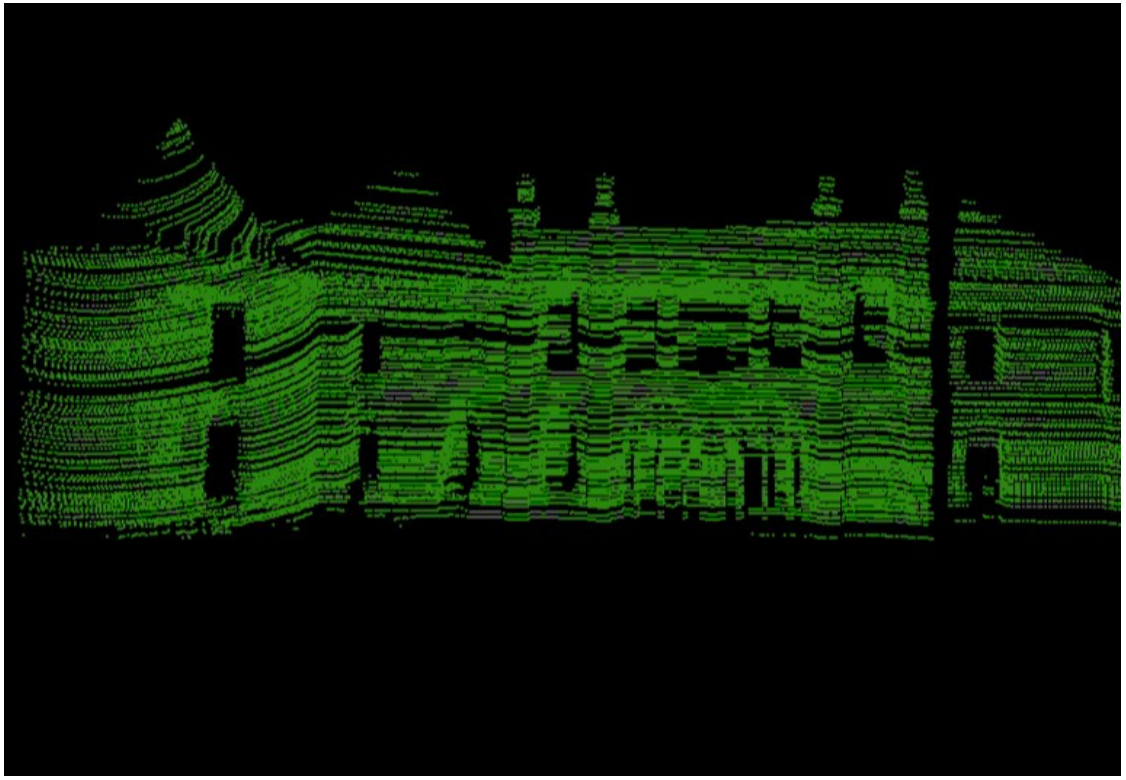
Real data experiments

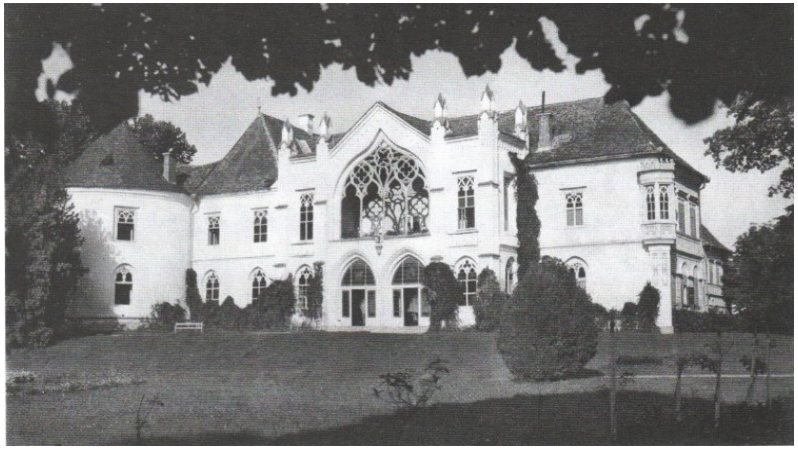
- 3D depth data & 2D perspective images

  Archive image taken in 1935 (unknown camera)

  Modern image taken after 2010 (known camera)

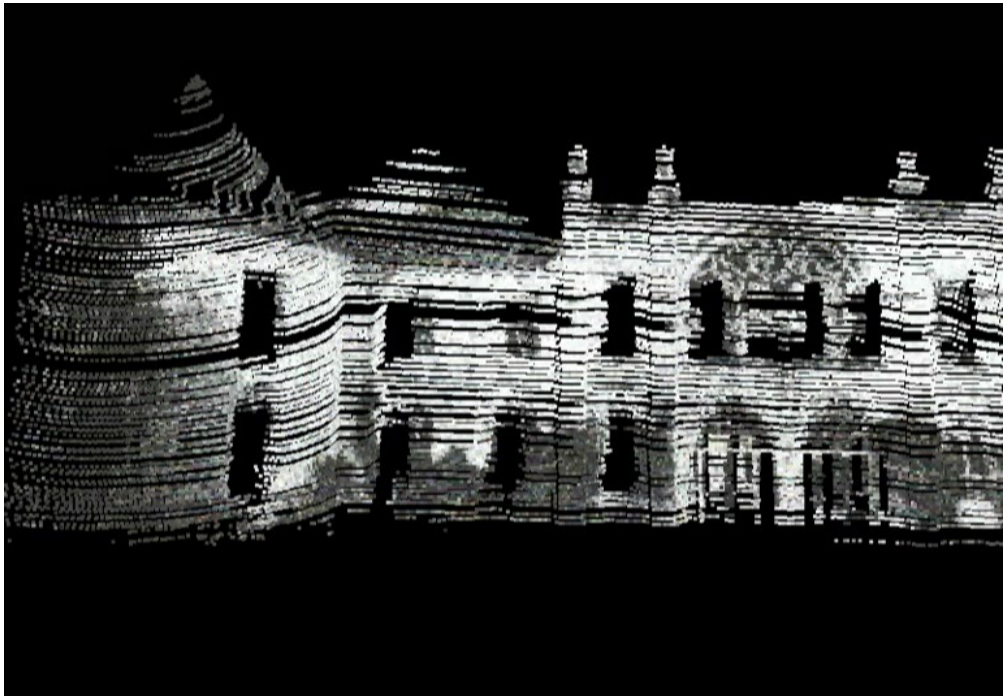
  - ❖ Considerable structural differences

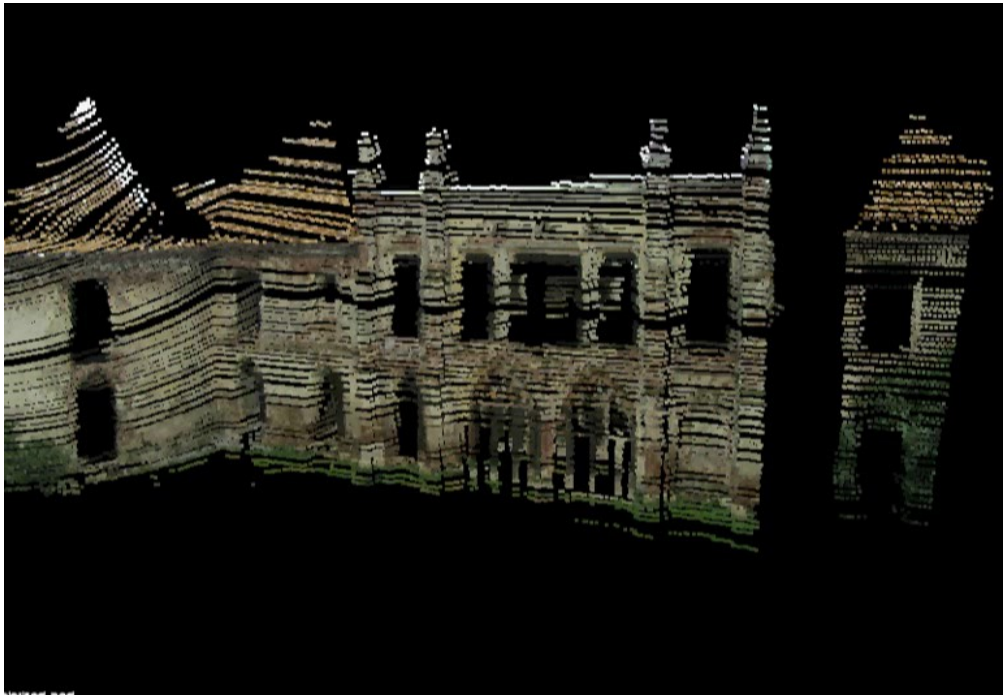




Fusion result after calibration

COLOR & SPACE IN CULTURAL HERITAGE





Publications:

1. F. Boochs, A. Bentkowska-Kafel, C. Degrigny, M. Karaszewski, A. Karmacharya, Z. Kato, M. Picollo, R. Sitnik, A. Trémeau, D. Tsiafaki, L. Tamas **Colour and Space in Cultural Heritage: Key Questions in 3D Optical Documentation of Material Culture for Conservation, Study and Preservation**. In*Proceedings of International Conference on Progress in Cultural Heritage. Documentation, Preservation, and Protection (**EuroMed**), Lecture Notes in Computer Science, Vol. 8740*, Limassol, Cyprus, pages 11-24, November 2014. Springer. (Winner of ***The Werner Weber Award***).

2. Levente Tamas and Zoltan Kato. **Targetless Calibration of a Lidar - Perspective Camera Pair**. In *Proceedings of ICCV Workshop on Big Data in 3D Computer Vision (**ICCV-BigData3DCV**)*, Sydney, Australia, pages 668-675, December 2013. IEEE.

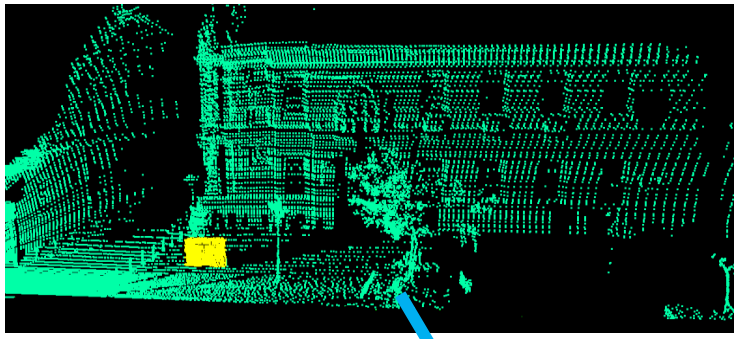
with Robert Frohlich and Levente Tamas

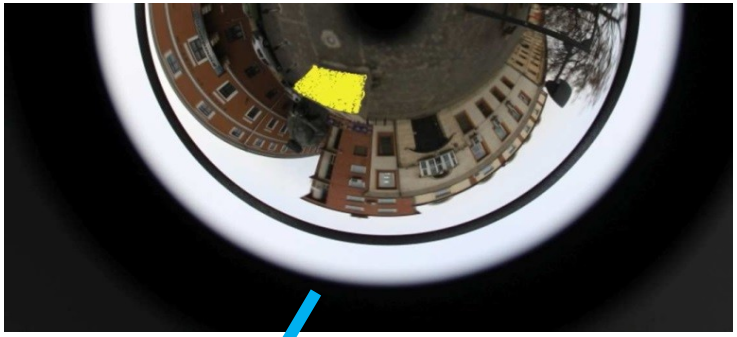
RELATIVE POSE ESTIMATION OF AN OMNIDIRECTIONAL – LIDAR CAMERA PAIR

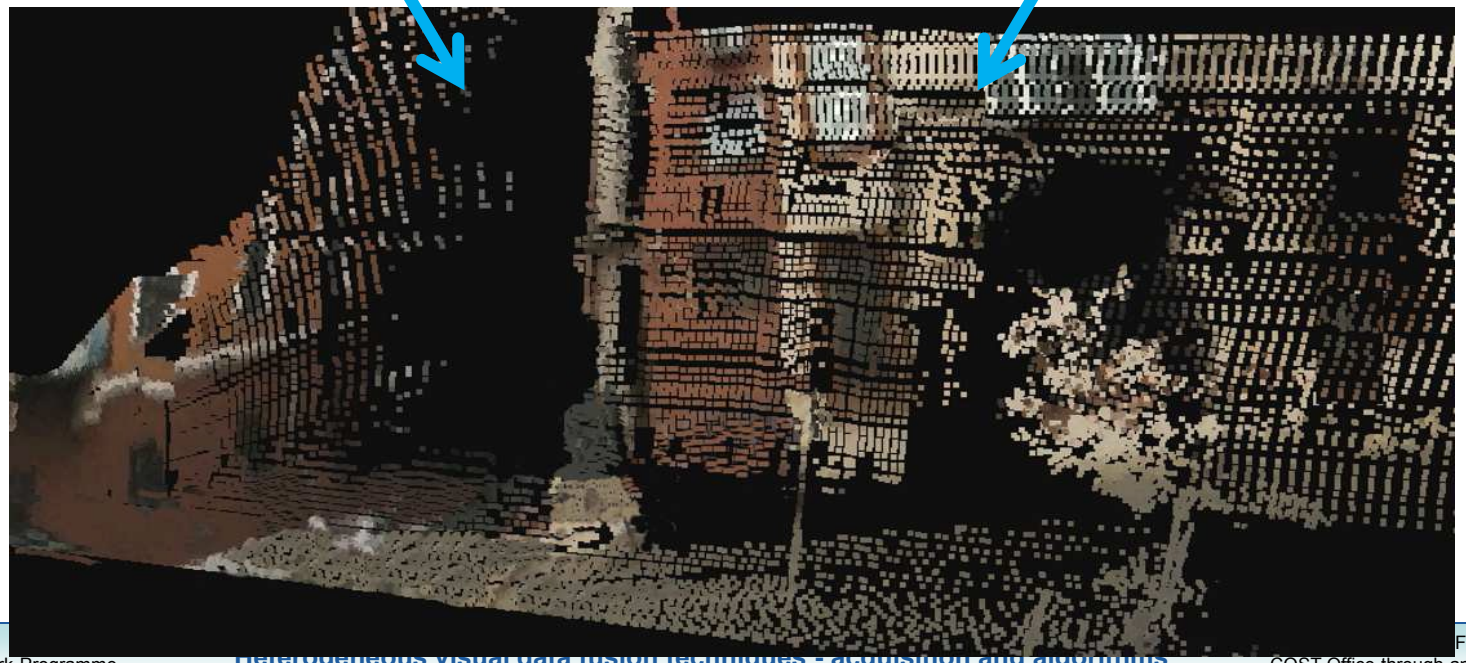
Problem statement

COLOR & SPACE IN CULTURAL HERITAGE

- Estimate the relative pose of an omnidirectional camera with respect to a 3D Lidar coordinate frame and fuse the different sensor data.







Problem statement

Classical solution:

- estimate point correspondences in the 3D-2D domain (eg. in [1], the calibration is performed in natural scenes, selecting point correspondences in a semisupervised manner
- [2] tackles calibration as an observability problem using a planar fiducial marker
- In [3] and [4] methods are proposed based on mutual information (MI)

1. Scaramuzza, D., Harati, A., Siegwart, R.: Extrinsic Self Calibration of a Camera and a 3D Laser Range Finder from Natural Scenes. In: International Conference on Intelligent Robots and Systems. pp. 4164–4169. San Diego, USA (October 2007)
2. Mirzaei, F.M., Kottas, D.G., Roumeliotis, S.I.: 3D LIDAR-camera intrinsic and extrinsic calibration: Identifiability and analytical least-squares-based initialization. International Journal of Robotics Research 31(4), 452–467 (2012)
3. Pandey, G., McBride, J.R., Savarese, S., Eustice, R.M.: Automatic Targetless Extrinsic Calibration of a 3D Lidar and Camera by Maximizing Mutual Information. In: AAAI National Conference on Artificial Intelligence. pp. 2053–2059. Toronto, Canada (July 2012)
4. Taylor, Z., Nieto, J.: A Mutual Information Approach to Automatic Calibration of Camera and Lidar in Natural Environments. In: Australian Conference on Robotics and Automation. pp. 3–8. Wellington, Australia (December 2012)

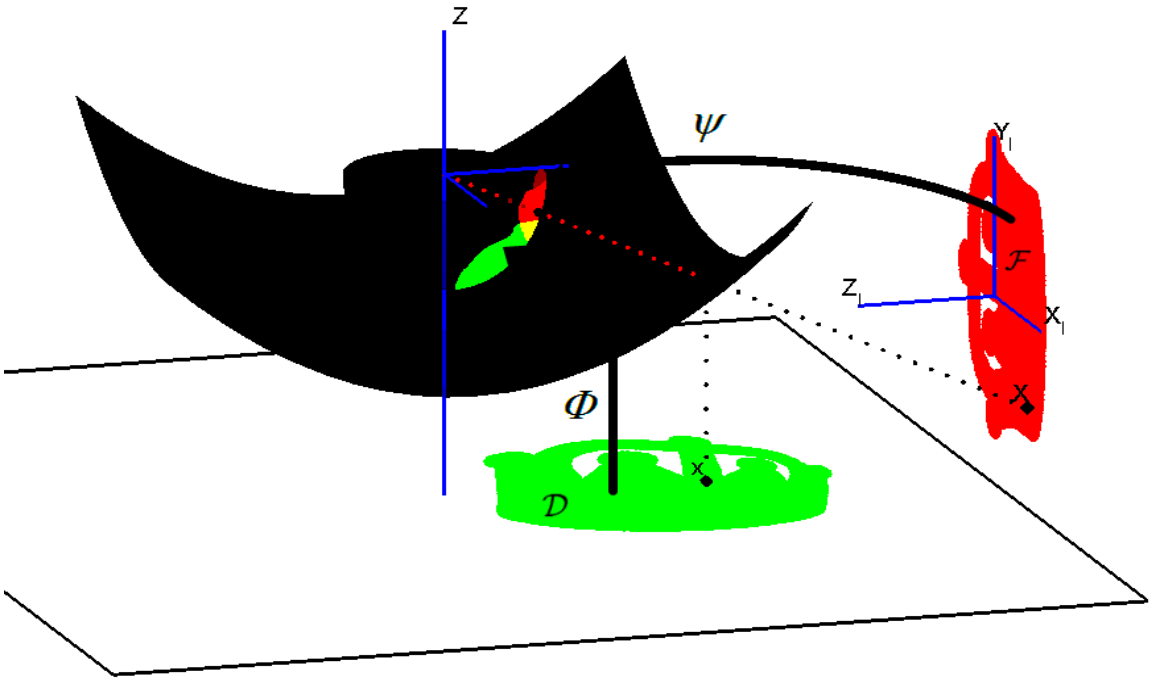
- no specific setup, only a pair of 2D – 3D data needed
- Internal parameters are known ➔ calibrated omni camera
- Pose parameters obtained by solving a system of nonlinear equations
- Based on shape alignment of planar segmented patches

Camera model

The omnidirectional camera is represented as a projection onto the surface of a unit sphere[5].

The image plane I maps to the surface of sphere S by Φ through:

- lifting the image point x onto the g surface
- centrally projecting x_g onto the sphere S



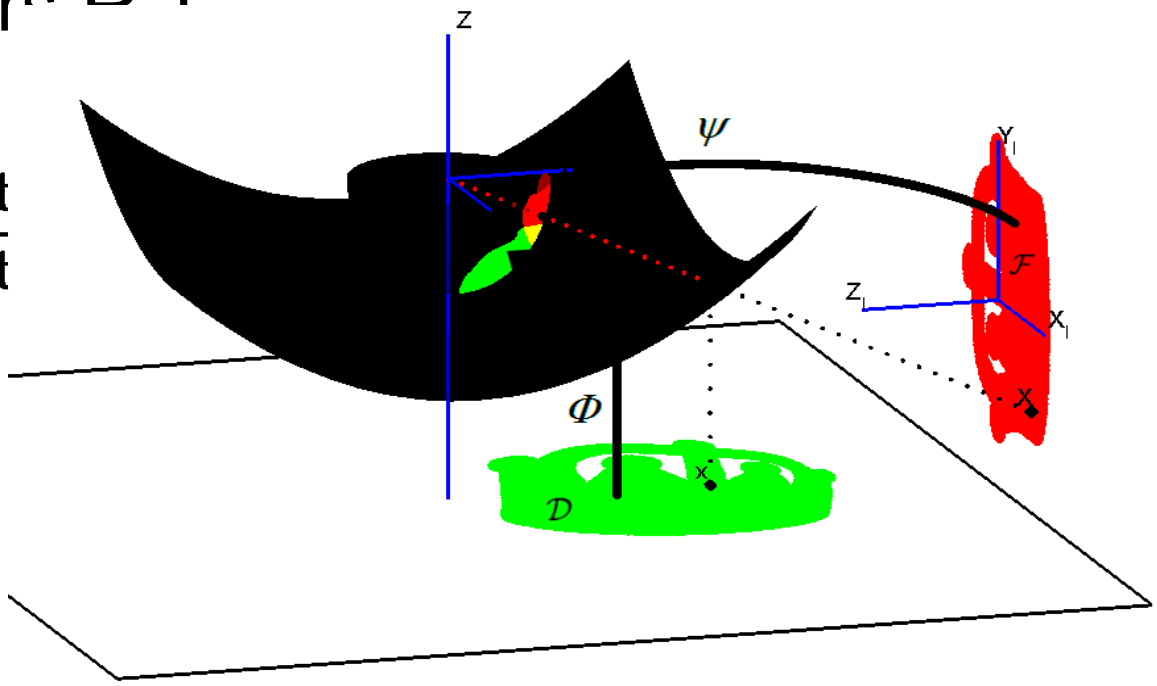
5. Scaramuzza, D., Martinelli, A., Siegwart, R.: A Flexible Technique for Accurate Omnidirectional Camera Calibration and Structure from Motion. In: International Conference on Computer Vision Systems, Washington, USA (January 2006) 45–51

Camera model

The omnidirectional camera is represented as a projection onto the surface of a unit sphere[5].

A 3D world point X projects onto S considering the extrinsic pose parameters R, t:

$$\Phi(\mathbf{x}) = \mathbf{x}_{\mathcal{S}} = \Psi(\mathbf{X}) = \frac{\mathbf{RX} + \mathbf{t}}{\|\mathbf{RX} + \mathbf{t}\|}$$



5. Scaramuzza, D., Martinelli, A., Siegwart, R.: A Flexible Technique for Accurate Omnidirectional Camera Calibration and Structure from Motion. In: International Conference on Computer Vision Systems, Washington, USA (January 2006) 45–51

Proposed solution

Point matches not available ⇒ integrate out individual point pairs over spherical surface patches D_S and F_S

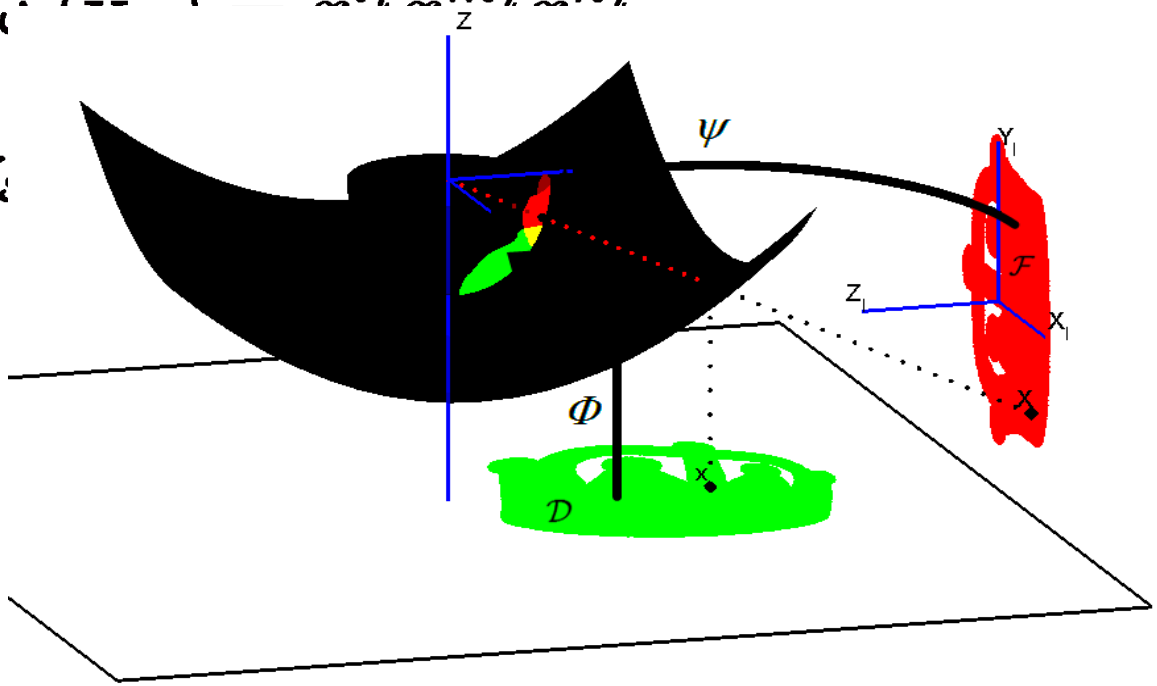
$$\iint_{\mathcal{D}_S} \mathbf{x}_S \, \mathrm{d}\mathcal{D}_S = \iint_{\mathcal{F}_S} \mathbf{z}_S \, \mathrm{d}\mathcal{F}_S$$

⇒ This gives us 2 equations only, but pose has 6 parameters

We generate independent equations by applying a set of nonlinear functions on both sides[6].

$$\iint_{\mathcal{D}_S} \omega(\mathbf{x}_S) \, \mathrm{d}\mathcal{D}_S = \iint_{\mathcal{F}_S} \omega(\mathbf{z}_S) \, \mathrm{d}\mathcal{F}_S$$

We obtain an overdetermined system of 15 equations.



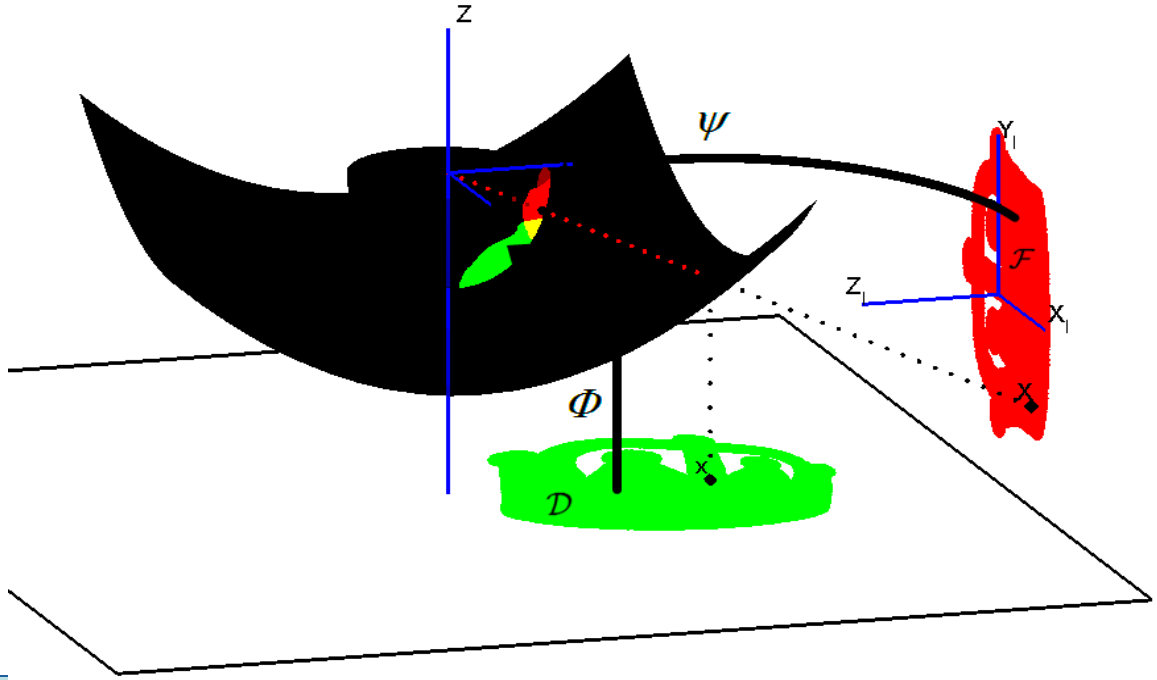
6. Domokos, C., Nemeth, J., Kato, Z.: Nonlinear Shape Registration without Correspondences. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(5) (2012) 943–958

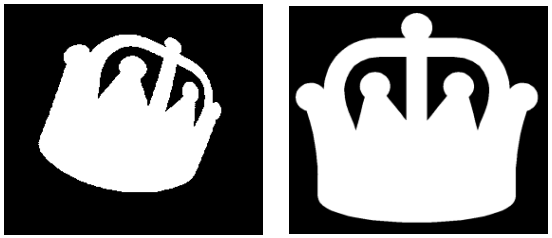
Proposed solution

The explicit form of the equation is obtained by parameterizing the surface patches D_s and F_s via Φ and ψ over the planar regions D and F:

$$\iint_{\mathcal{D}} \omega_i(\Phi(\mathbf{x})) \left\| \frac{\partial \Phi}{\partial x_1} \times \frac{\partial \Phi}{\partial x_2} \right\| \mathrm{d}x_1 \, \mathrm{d}x_2 = \iint \omega_i(\Psi(\mathbf{X})) \left\| \frac{\partial \Psi}{\partial X_1} \times \frac{\partial \Psi}{\partial X_2} \right\| \mathrm{d}X_1 \, \mathrm{d}X_2$$

The above equation can be solved by LM algorithm.

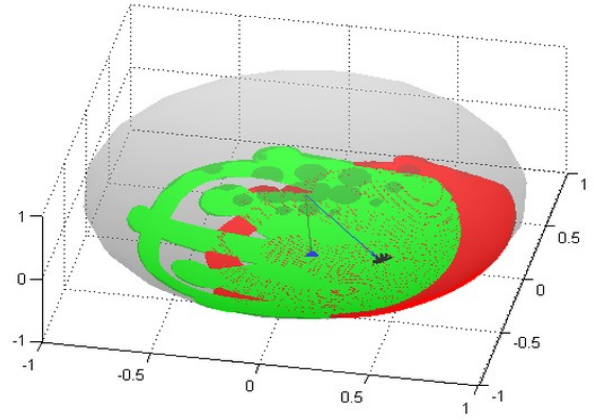


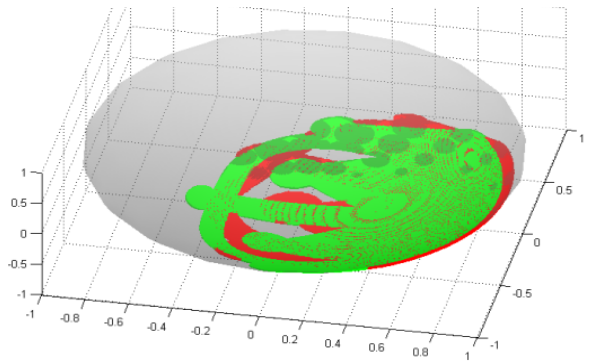


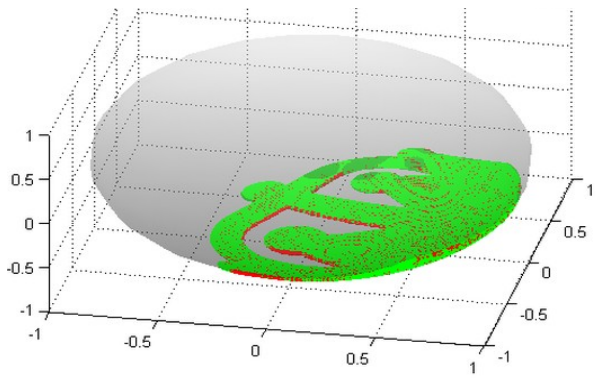
Algorithm

Good initialization of R,t parameters is crucial for optimal results

1. project both data on the unit sphere, calculate centroid of shapes
2. Initialize R as the rotation between the centroids, t as the distance from where the area of the projected patches is equal
3. Solve the system using LM algorithm

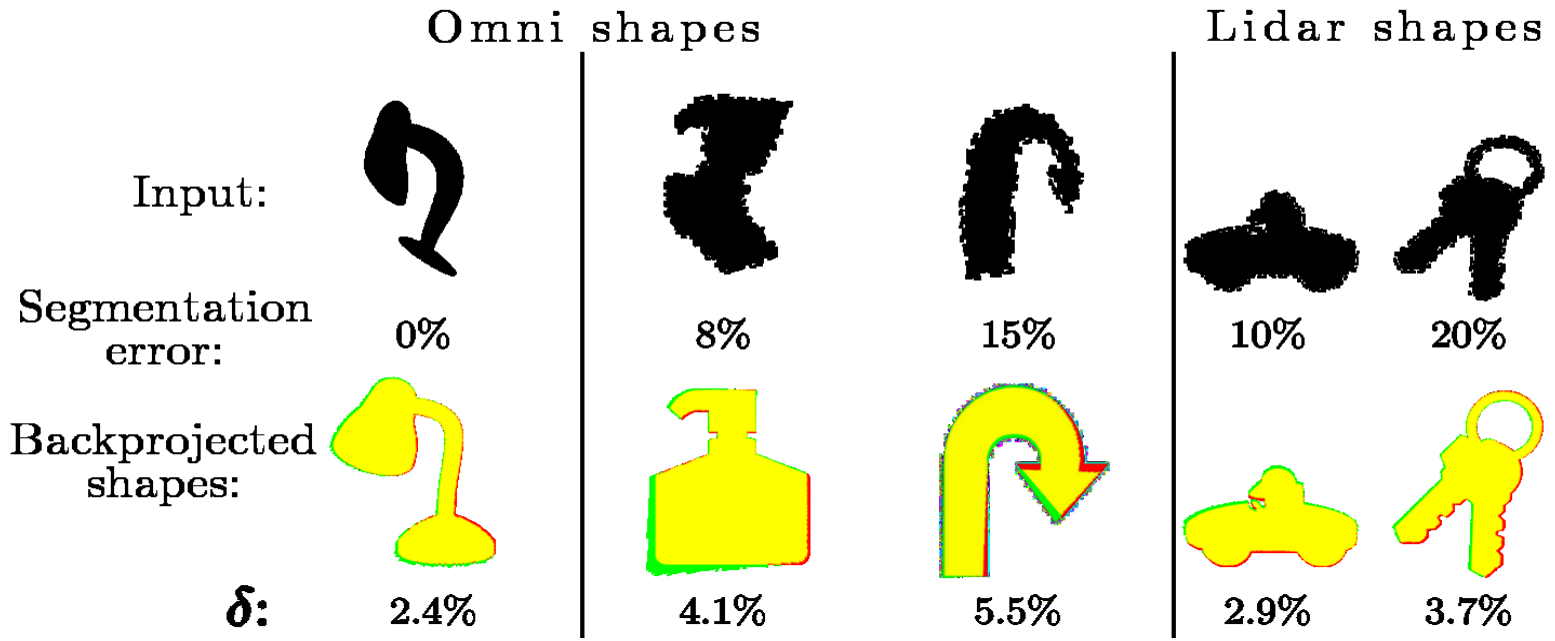






Evaluation on synthetic data

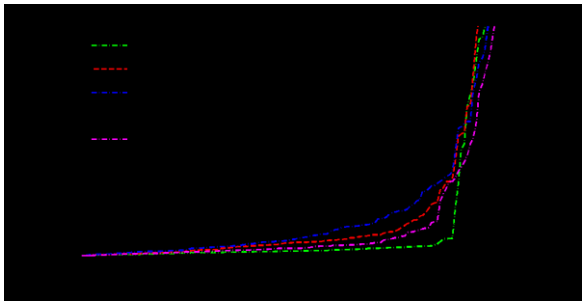
- benchmark dataset of 2500 2D-3D synthetic image pairs
- simulating segmentation errors around the contour
- alignment error(δ) measured as the percentage of

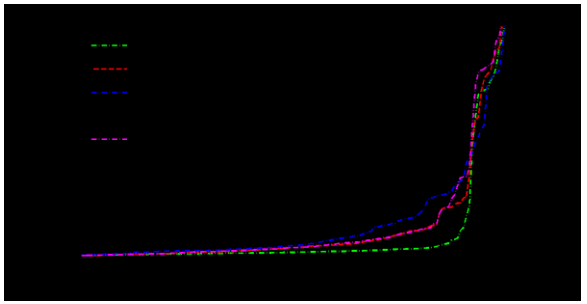


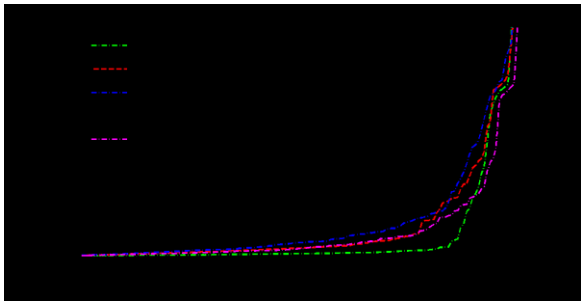
COSCH
COLOR & SPACE IN CULTURAL HERITAGE

Evaluation on synthetic data

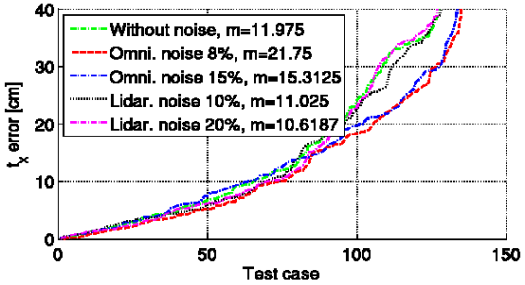
Rotation errors typically < 2.5 degrees along the 3 axis.

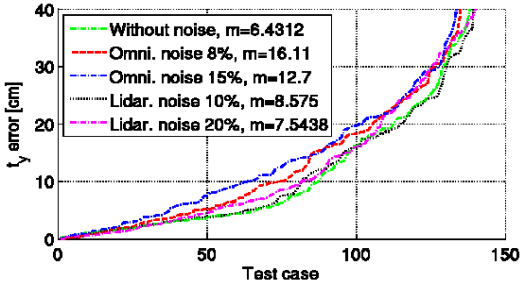


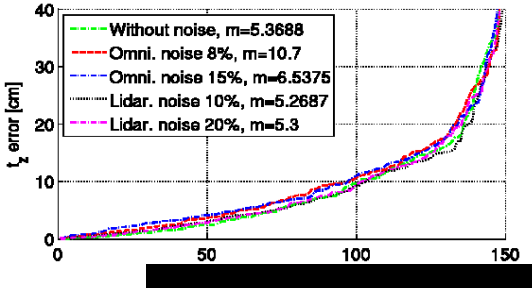


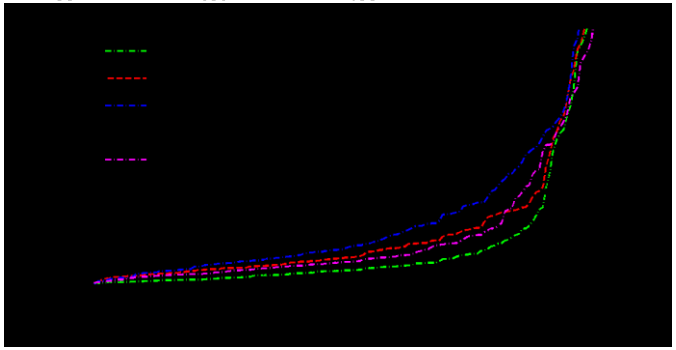


Translation error typically < 15 cm along the 3 axis.





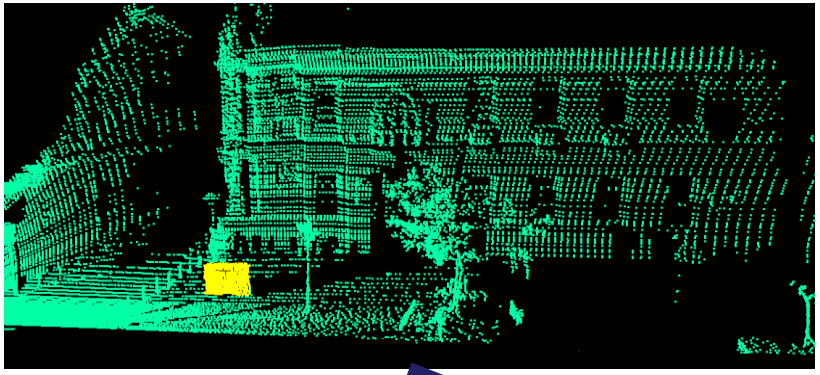


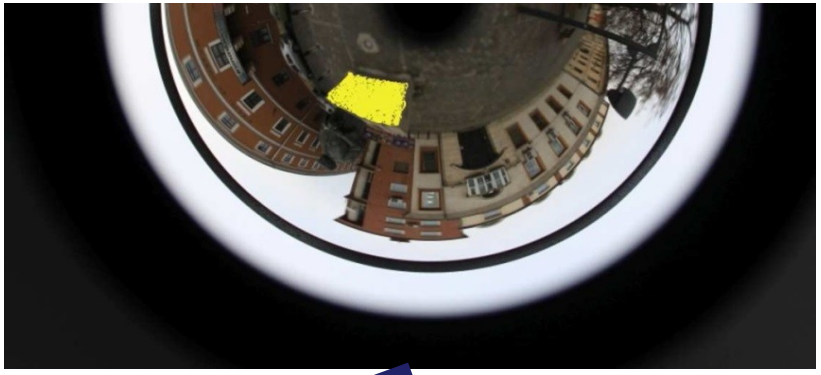


Alignment error plot for different segmentation errors.
The method proves to be robust for errors up to 15%.

Evaluation on real data

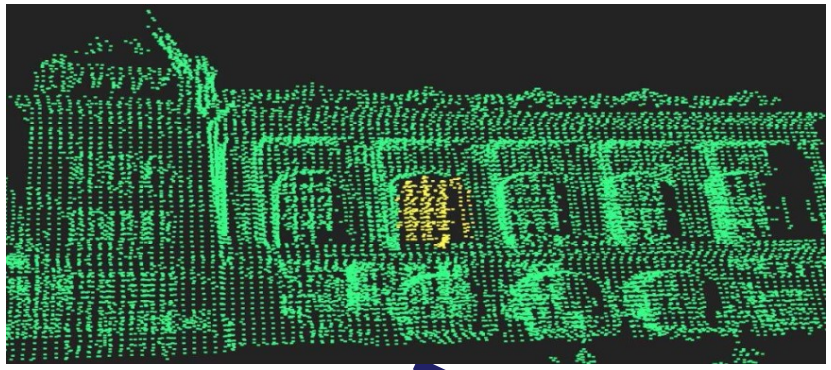
3D Lidar + catadioptric camera





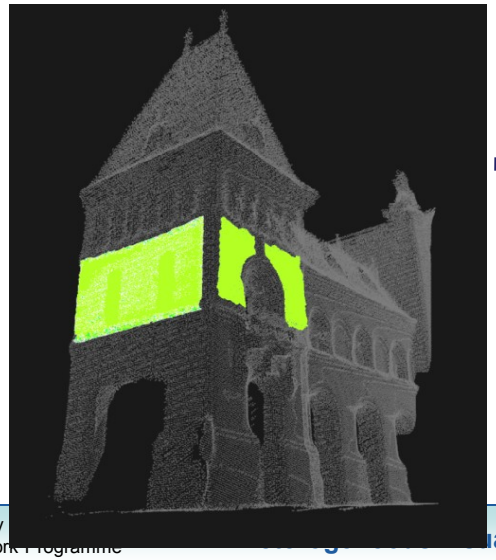
Evaluation on real data

3D Lidar + fisheye camera



Evaluation on real data

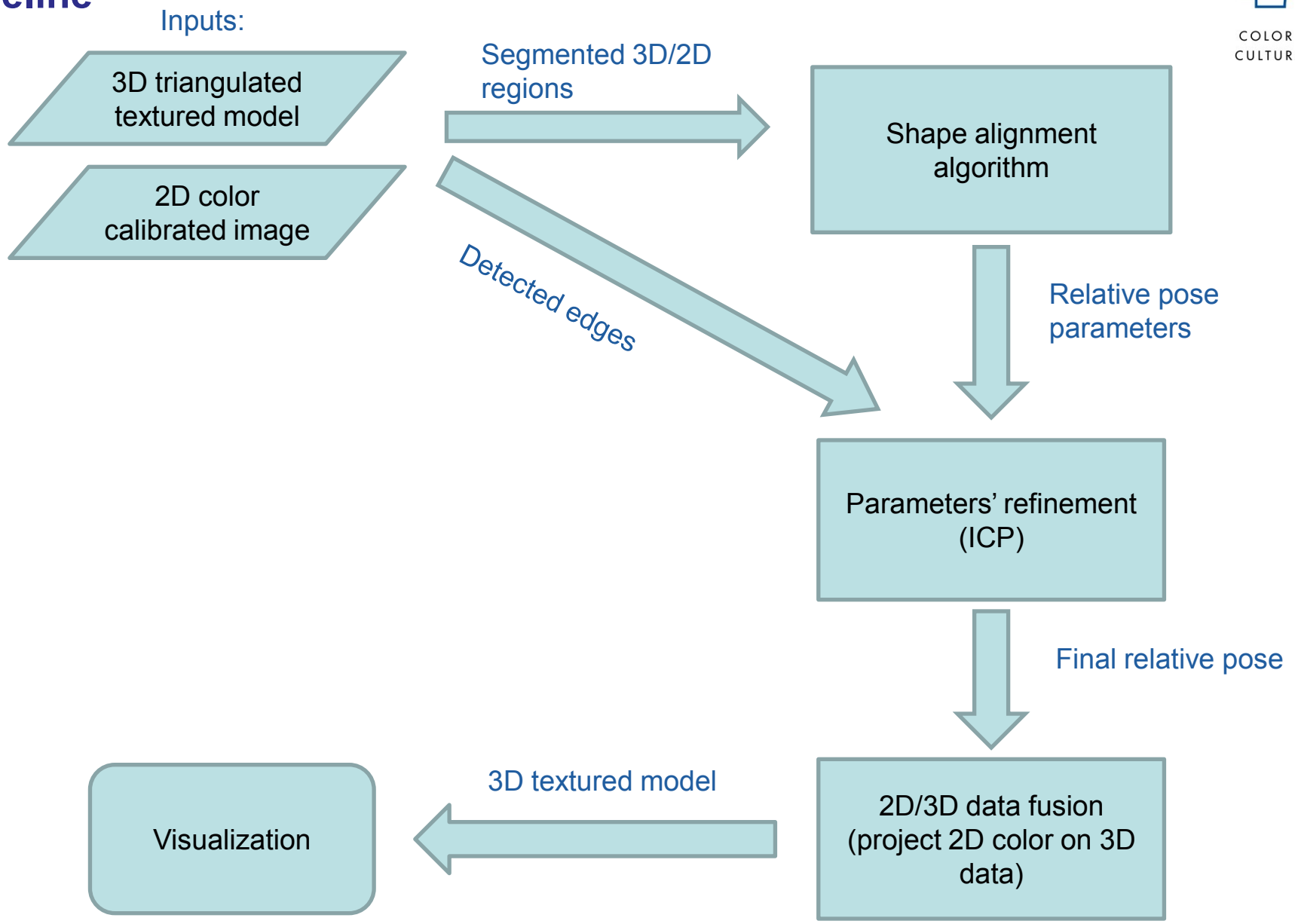
3D Lidar + fisheye camera



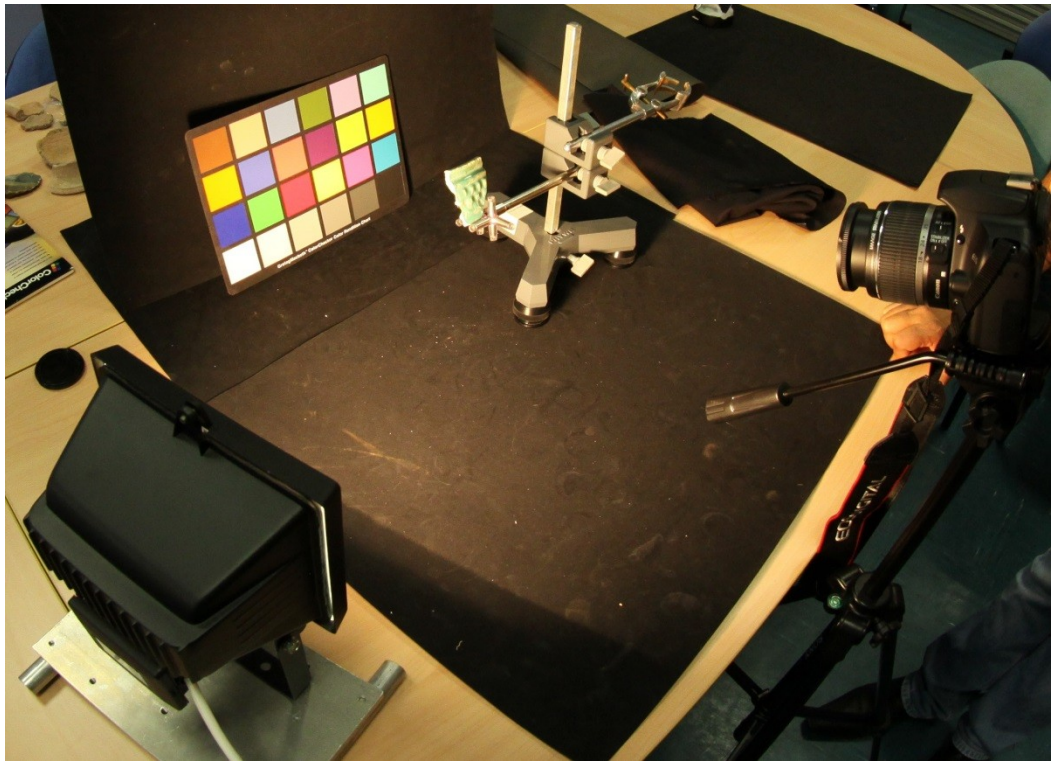
with Robert Frohlich & Alain Tremeau

USE CASE: COLOR-CALIBRATED 3D MODELS OF SMALL CH OBJECTS

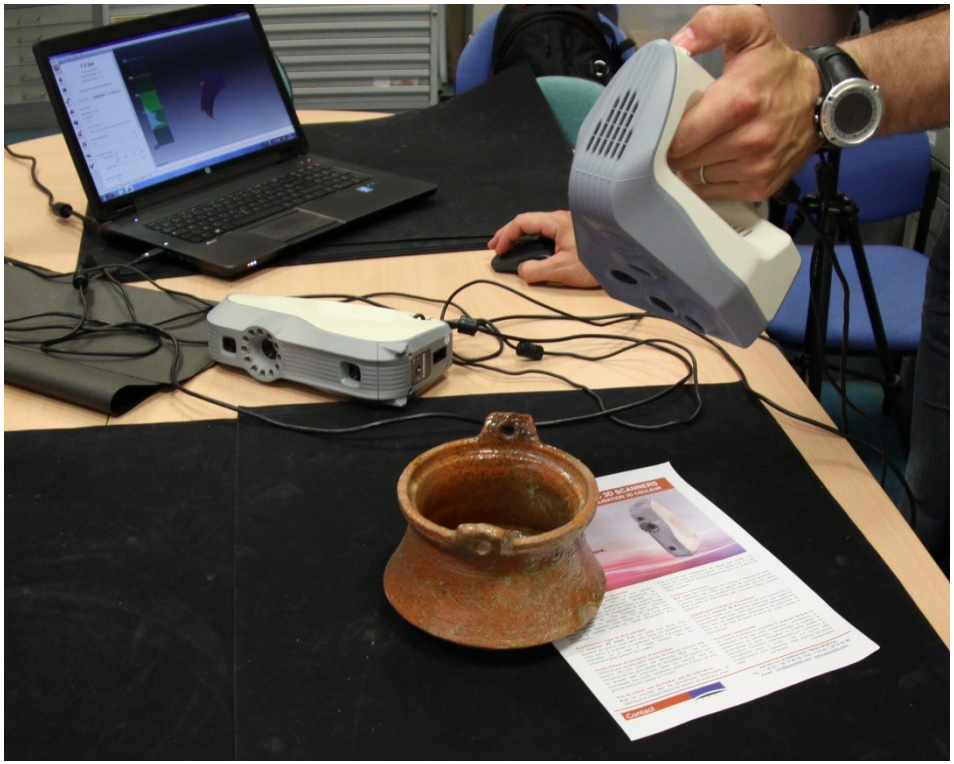
Pipeline


Inputs:
3D triangulated textured model
2D color calibrated image
Segmented 3D/2D regions
Shape alignment algorithm
Detected edges
Relative pose parameters
Parameters' refinement (ICP)
Final relative pose
2D/3D data fusion (project 2D color on 3D data)
3D textured model
Visualization

Data acquisition: setup

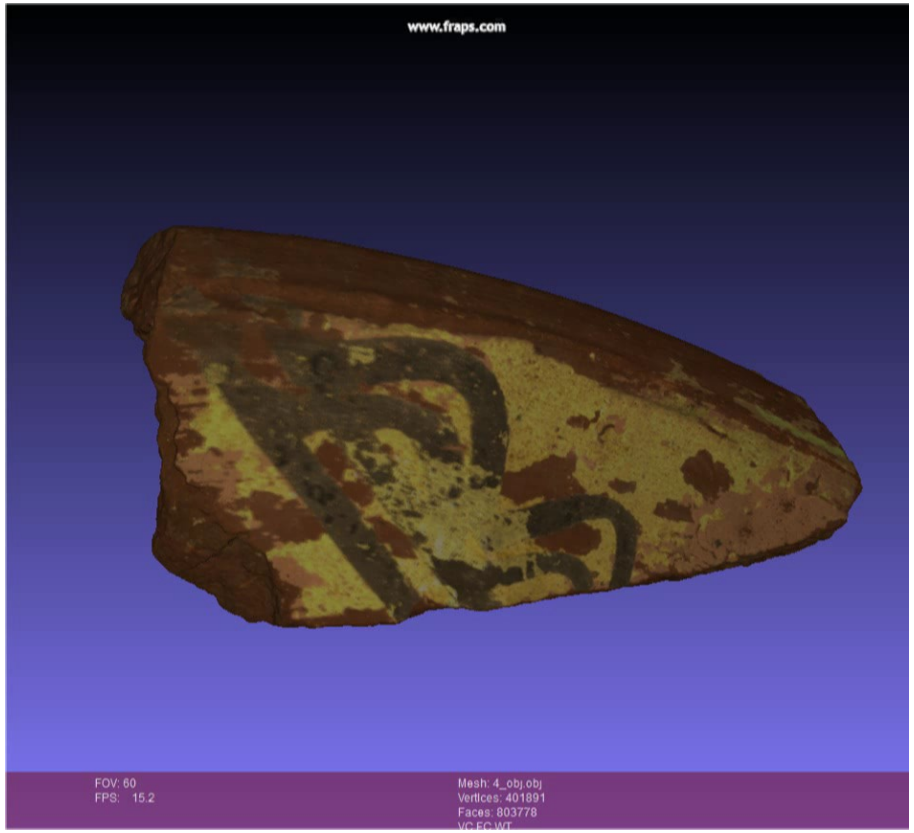


Two 2D cameras, light source,
color calibration chart, stand



3D scanner: Artec Spider, Artec Eva

Data acquisition: 2D and 3D data



Segmentation

- Use regions of big flat surfaces

  Edges of flat regions may not be well defined

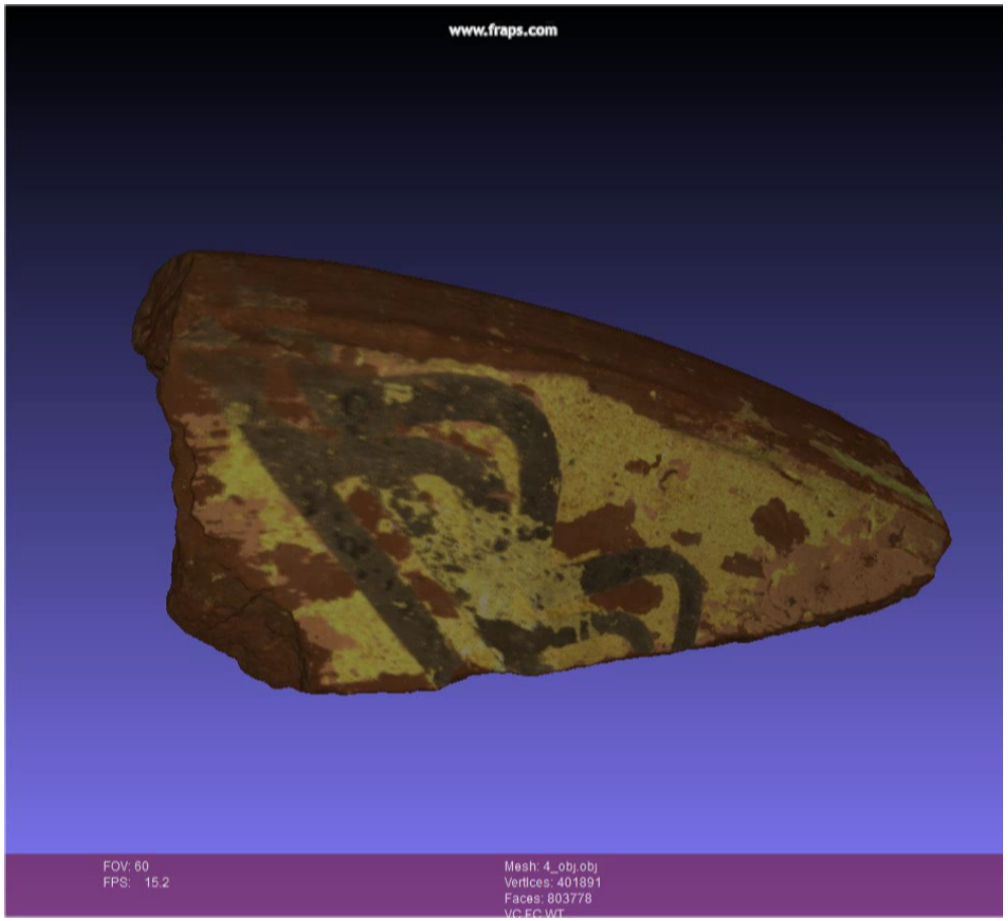
Segmentation

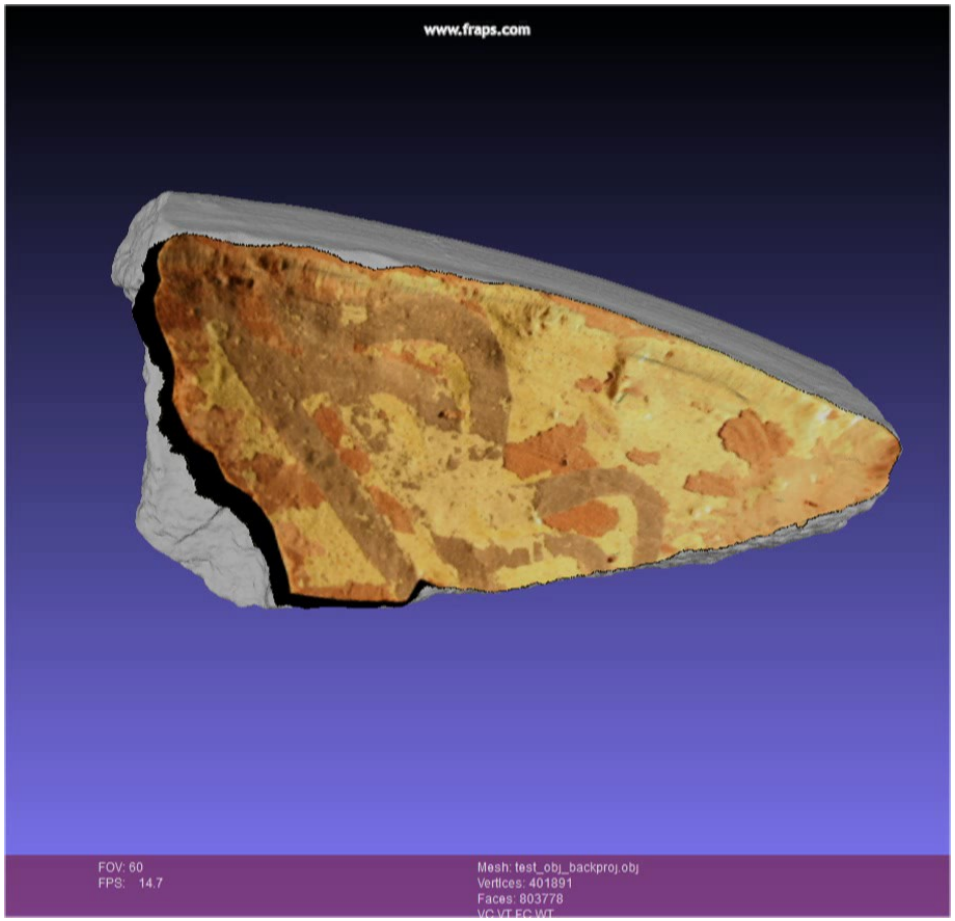
- Try to use color information as well

  Color is imprecise, low resolution, washed out on 3D data

Final result

- Color projected back from a single 2D image onto the 3D pointcloud





Conclusions

- Targetless 2D-3D calibration framework
- Region-based registration without point correspondences
- No need for special targets or intensity Lidar data
  - works on any existing dataset
- Tested on large synthetic data set
- Experimental proof on different devices

Publications

- Levente Tamas and Zoltan Kato. **Targetless Calibration of a Lidar - Perspective Camera Pair**. In *Proceedings of ICCV Workshop on Big Data in 3D Computer Vision (**ICCV-BigData3DCV**)*, Sydney, Australia, pages 668-675, December 2013. IEEE.

- Levente Tamas, Robert Frohlich, and Zoltan Kato. **Relative Pose Estimation and Fusion of Omnidirectional and Lidar Cameras**. In *Proceedings of the ECCV Workshop on Computer Vision for Road Scene Understanding and Autonomous Driving (**ECCV-CVRSUAD**), Lecture Notes in Computer Science*, Zurich, Switzerland, September 2014. Springer.

Acknowledgement

***Collaborators**:*

- *Levente Tamas (UTC, Cluj)*
- *Robert Frohlich (SZTE)*

***This research was supported by**:*

- *the EU and the State of Hungary, co-financed by the European Social Fund*
  - *in the framework of "National Excellence Program" TAMOP-4.2.4. A/2-11-1-2012-0001 as well as*
  - *through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013)*
- *SCIEX-NMS-CH project no. 12.239*
- *MTA Domus.*