Trinity College Dublin

The University of Dublin

V-SENSE

3D computer vision

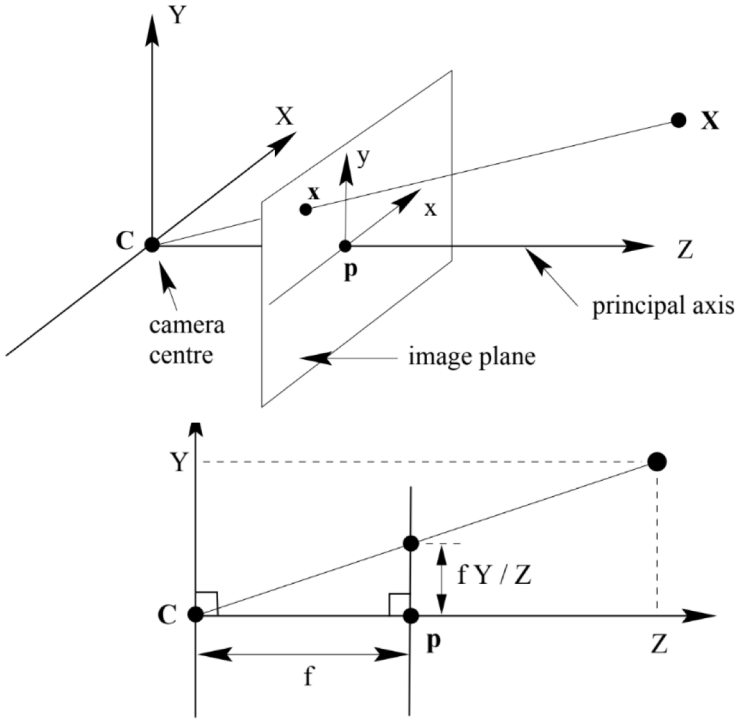
CS7GV1 Computer Vision – guest lecture

Dr Susana Ruano

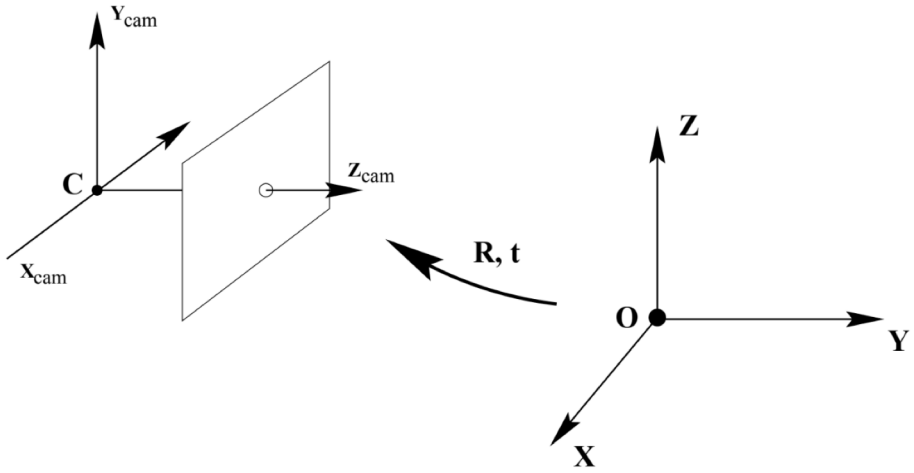
Outline

- **Image-based 3D reconstruction**
  - SfM + MVS
- **Large scale reconstruction: Dublin city**
- **Human reconstruction: PIFu**
- **What else?: NeRF**

Perspective camera model



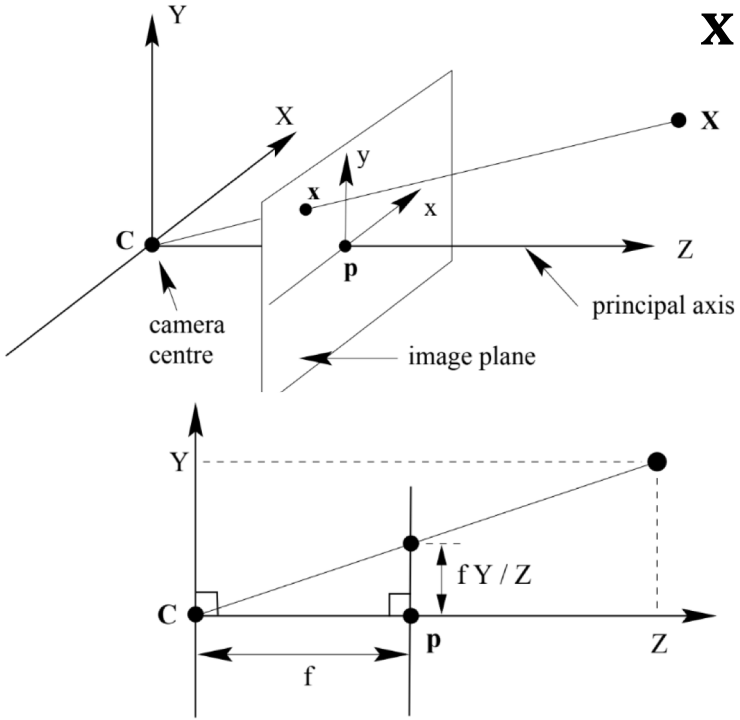
$$\mathbf{x} = \mathbf{P}\boldsymbol{X}$$

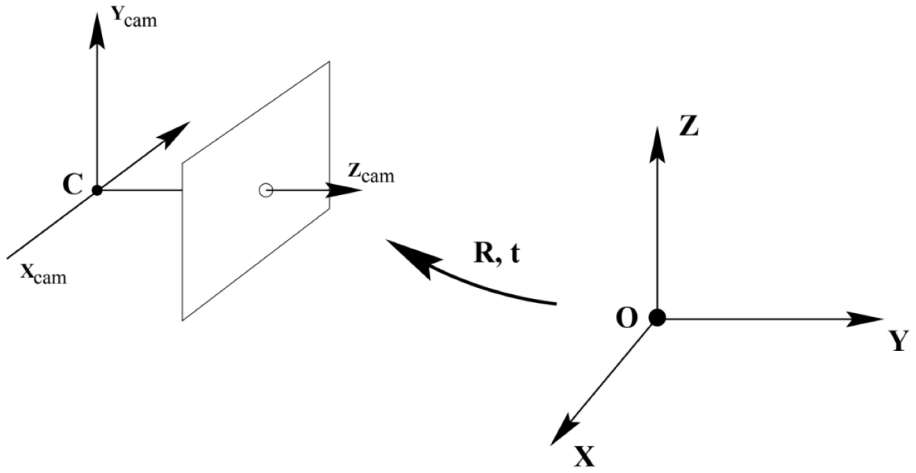


Multiple view geometry. Hartley-Zisserman

Perspective camera model

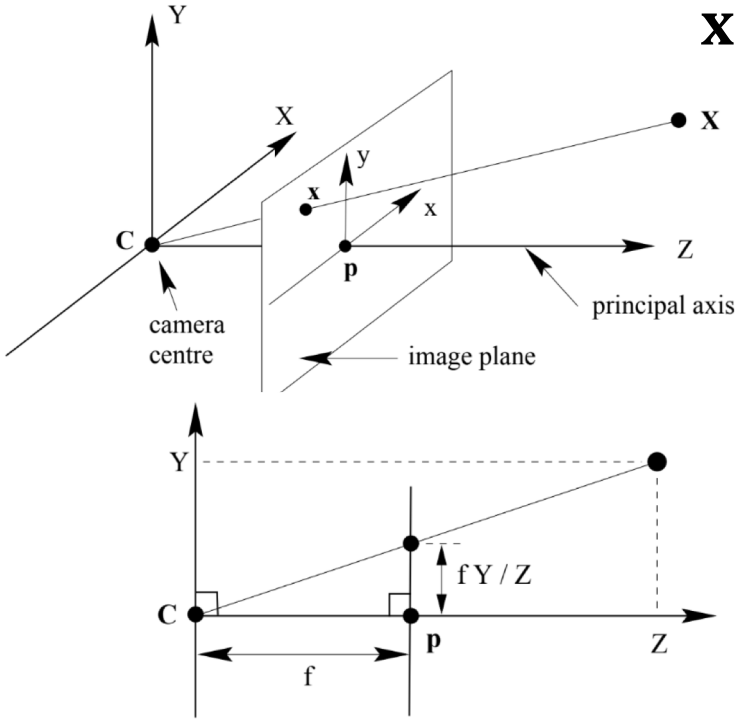
$$\mathbf{x} = \boldsymbol{K}(\boldsymbol{R}|\boldsymbol{t})\boldsymbol{X}$$





Multiple view geometry. Hartley-Zisserman

Perspective camera model



$$\mathbf{x} = \boldsymbol{K}(\boldsymbol{R}|\boldsymbol{t})\boldsymbol{X}$$

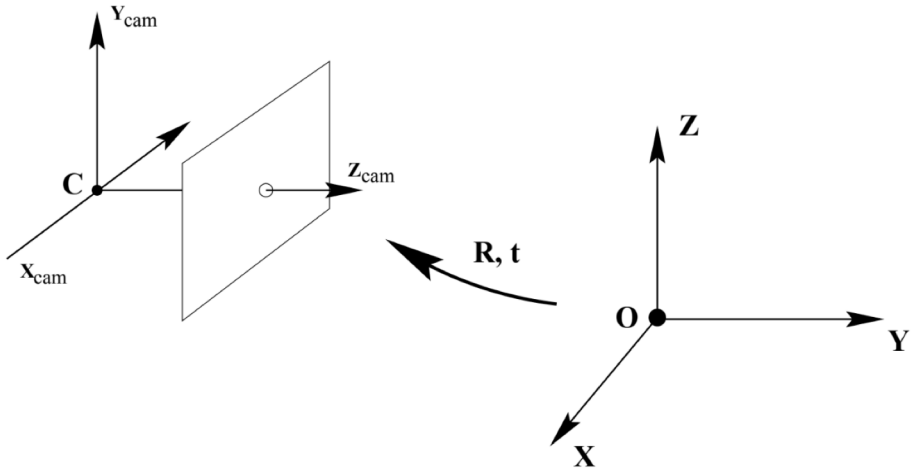
$$\mathrm{K}=\begin{pmatrix} f_x & & p_x \\ & f_y & p_y \\ & & 1 \end{pmatrix}$$

Multiple view geometry. Hartley-Zisserman

Perspective camera model

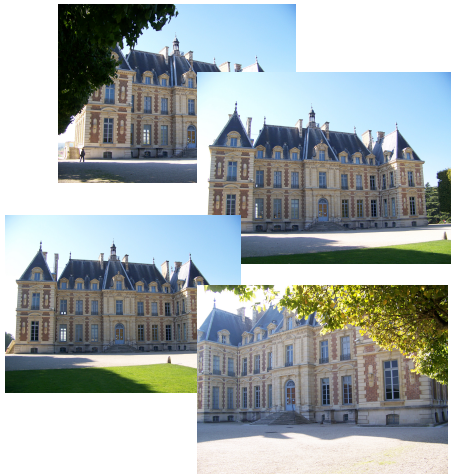
$$\mathbf{x} = \boldsymbol{K}(\boldsymbol{R}|\boldsymbol{t})\boldsymbol{X}$$

$$(\boldsymbol{R}|\boldsymbol{t}) = \begin{pmatrix} \boldsymbol{r_{11}} & \boldsymbol{r_{12}} & \boldsymbol{r_{13}} & \boldsymbol{t_x} \\ \boldsymbol{r_{21}} & \boldsymbol{r_{22}} & \boldsymbol{r_{23}} & \boldsymbol{t_y} \\ \boldsymbol{r_{31}} & \boldsymbol{r_{32}} & \boldsymbol{r_{33}} & \boldsymbol{t_z} \end{pmatrix}$$

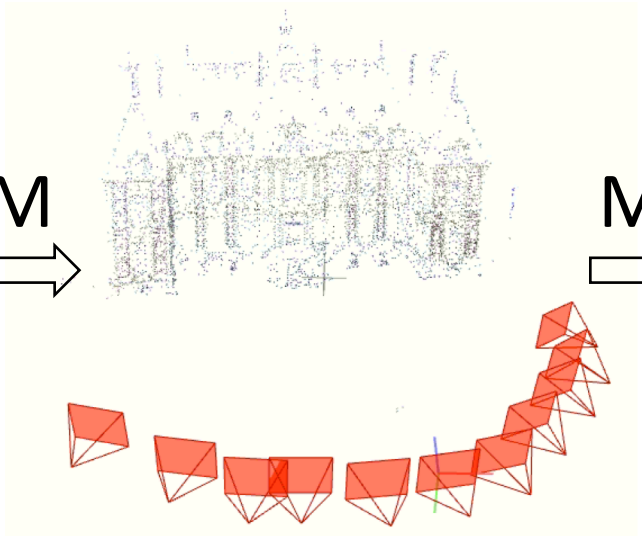


Multiple view geometry. Hartley-Zisserman

Image-based 3D reconstruction



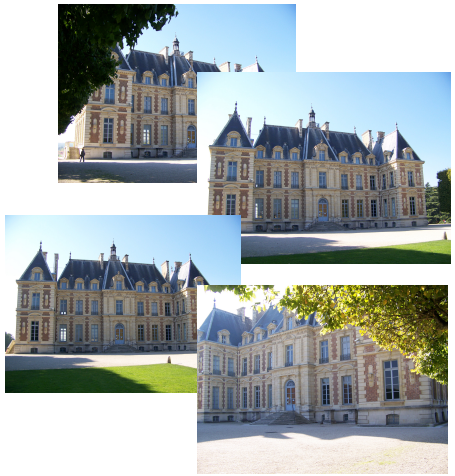
SfM



MVS

Image of "Chateau de Sceaux", Sceaux castle. France.
Photographer: Copyright 2012 Pierre MOULON http://imagine.enpc.fr/~moulonp/

Structure from Motion



⇨ Point correspondences ⇨ SfM ⇨

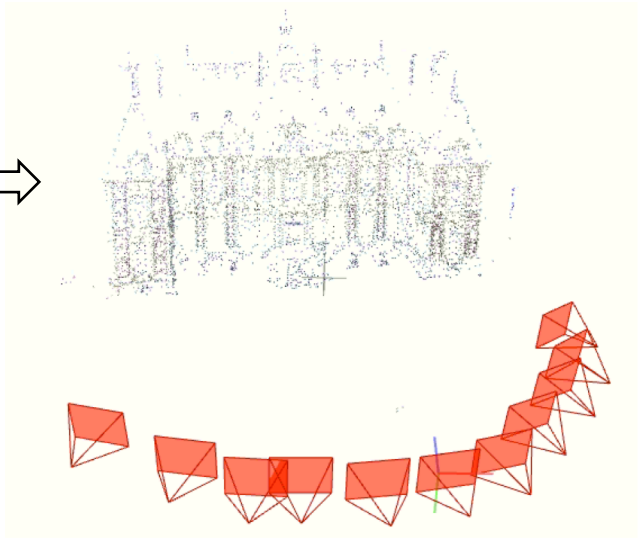
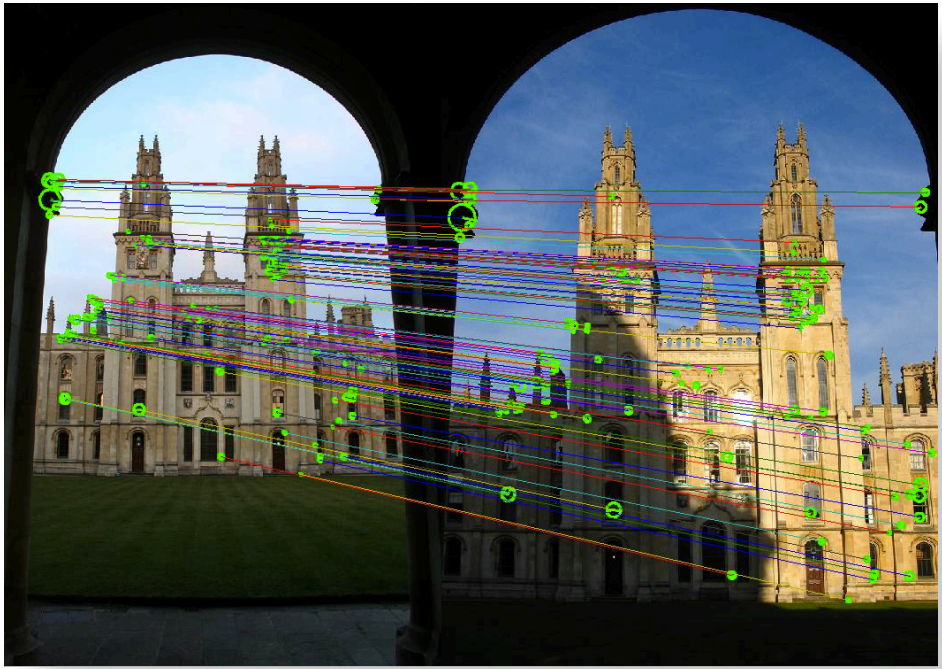


Image of "Chateau de Sceaux", Sceaux castle. France.
Photographer: Copyright 2012 Pierre MOULON http://imagine.enpc.fr/~moulonp/

Feature detection and feature matching

- **Features points are detected and matched across images.**
- **Matches are filtered using geometric and statistics constraints.**
- **SIFT features are probably the most popular, but there are alternatives (ORB, AKAZE, SURF…).**

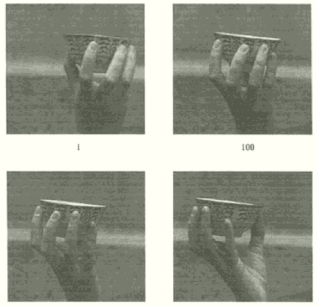


Oxford Visual Geometry Group computer vision practical

SfM in the last decades

1990 — 2006 — 2009 — 2016

SfM by factorization



$W = MS$

Shape and motion without depth.
C. Tomasi and T. Kanade. ICCV 1990

Photo Tourism

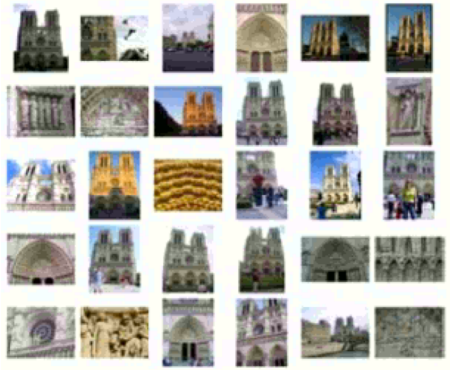


Photo Tourism: Exploring Photo Collections in 3D N. Snavely, S.M Seitz, R. Szeliski. SIGGRAPH 2006

Building Rome in a day

Bundler

Building Rome in a Day
S. Agarwal, N. Snavely, I. Simon, S. M. Seitz and R. Szeliski. ICCV 2009

SfM revisited

COLMAP

Structure-from-motion revisited.
J. Schonberger and JM Frahm. CVPR 2016

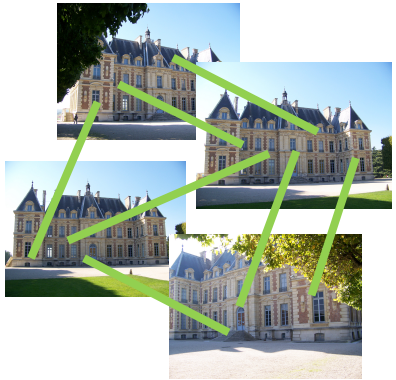
Structure from Motion

Given: 2D correspondences between the images $\{I_1, I_2, \cdots, I_n\}$

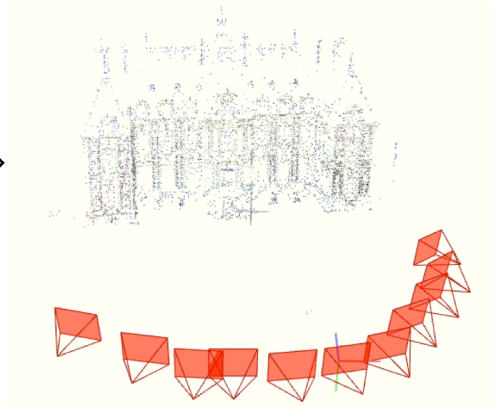
We want to estimate: $\{P_1, P_2, \cdots, P_n\}$ and the 3D points

Different strategies:

- **Incremental SfM**
- **Global SfM**



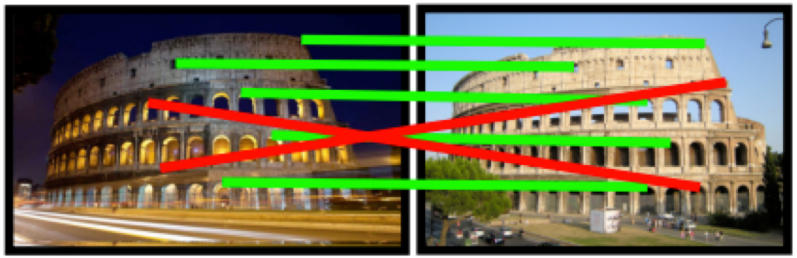
SfM



Incremental SfM

Initialization:

- **Select two views**

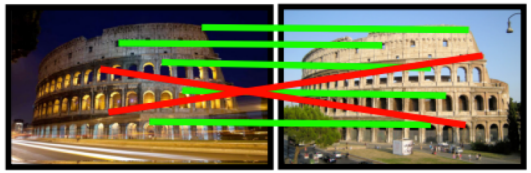


https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

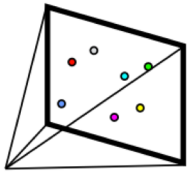
Incremental SfM

Initialization:

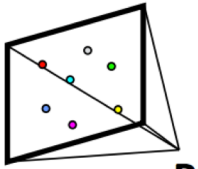
- Select two views
- **Estimate P, P’**



Fundamental matrix -> Essential matrix -> R,t



P = K [I | 0]



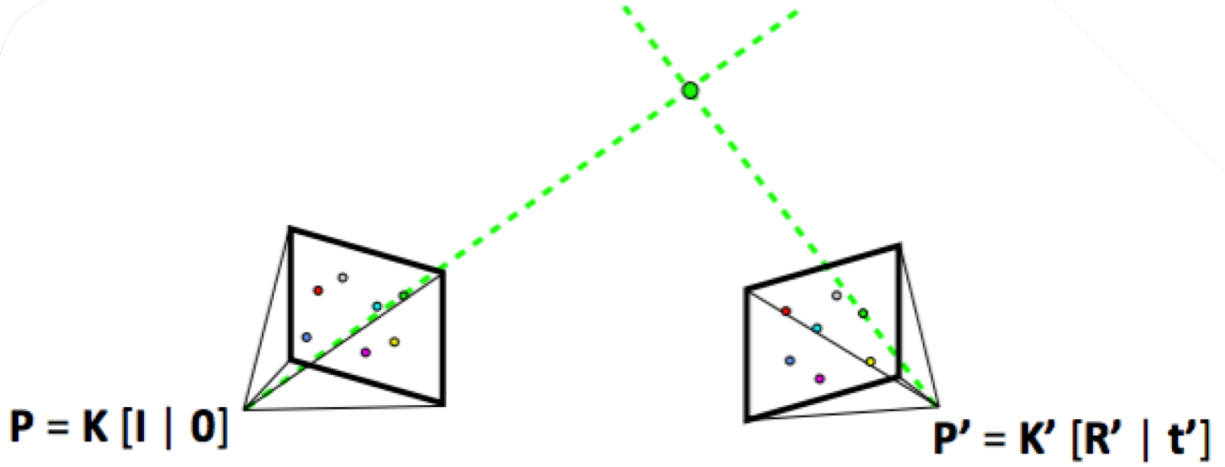
P’ = K’ [R’ | t’]

https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

Incremental SfM

Initialization:

- Select two views
- Estimate P, P’
- **Triangulate points**



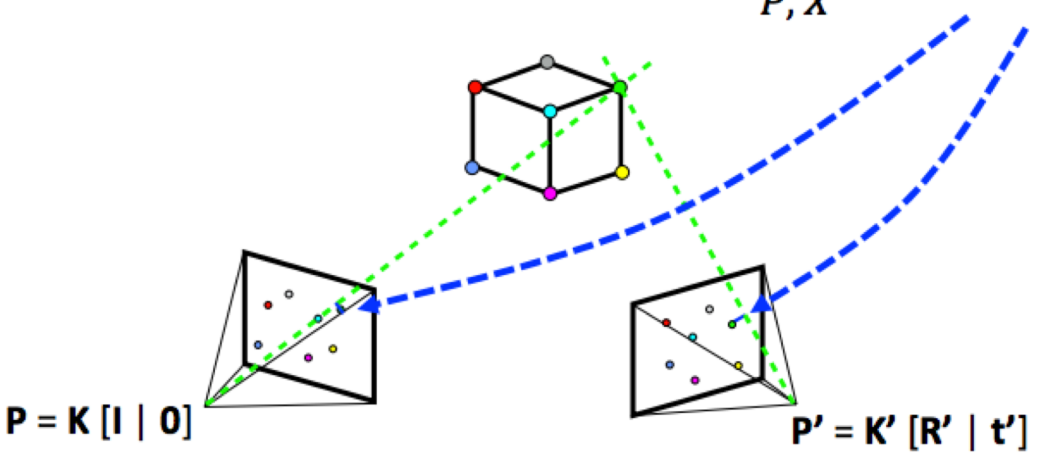
https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

Incremental SfM

Initialization:

- Select two views
- Estimate P, P’
- Triangulate points
- **Bundle adjustment**

- Minimize reprojection error: $\min_{P,X} \|x - \pi(P,X)\|$

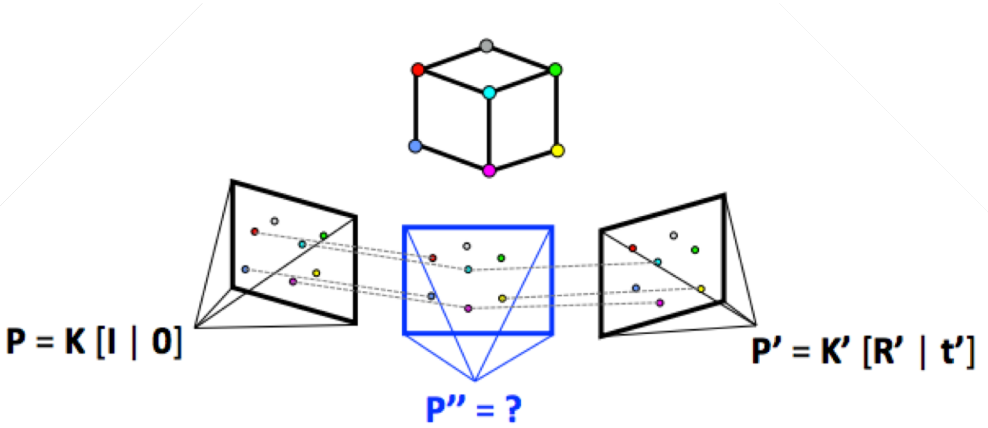


https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

Incremental SfM

Incorporating views:

- **Find 2D-3D correspondences**

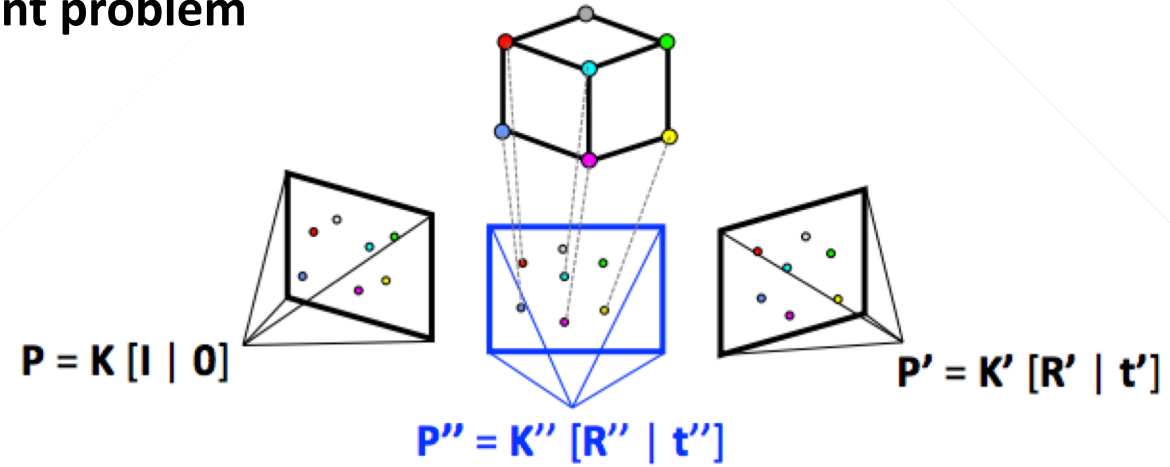


https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

Incremental SfM

Incorporating views:

- Find 2D-3D correspondences
- **Solve Perspective-n-Point problem**

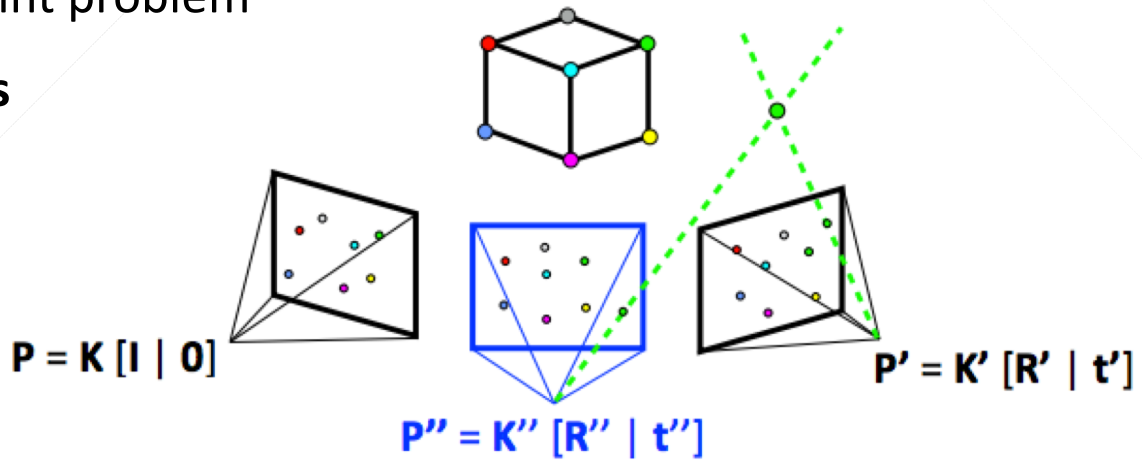


https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

Incremental SfM

Incorporating views:

- Find 2D-3D correspondences
- Solve Perspective-n-Point problem
- **Triangulate new points**



https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

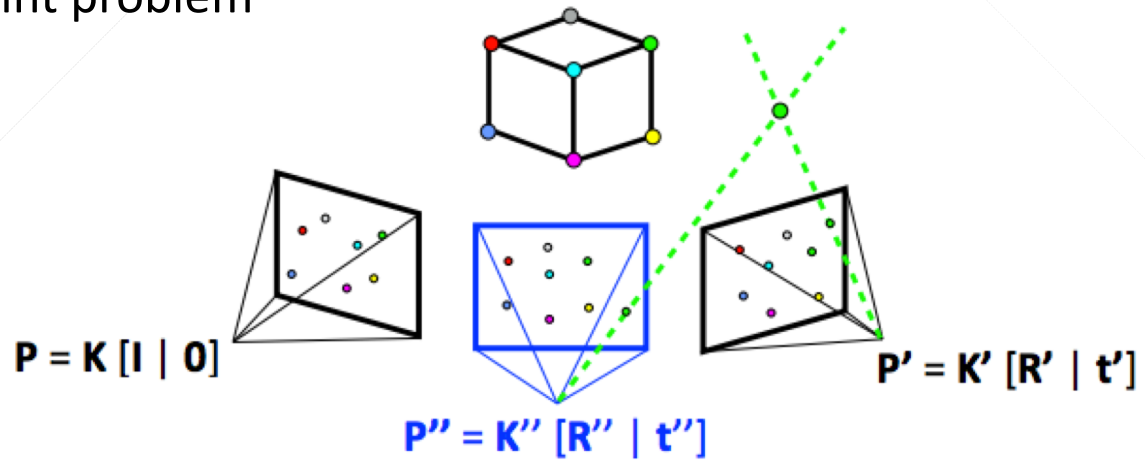
Incremental SfM

Incorporating views:

- Find 2D-3D correspondences
- Solve Perspective-n-Point problem
- Triangulate new points
- **Bundle adjustment**

Minimize reprojection error:

$$\min_{P,X} \|x - \pi(X, P)\|$$

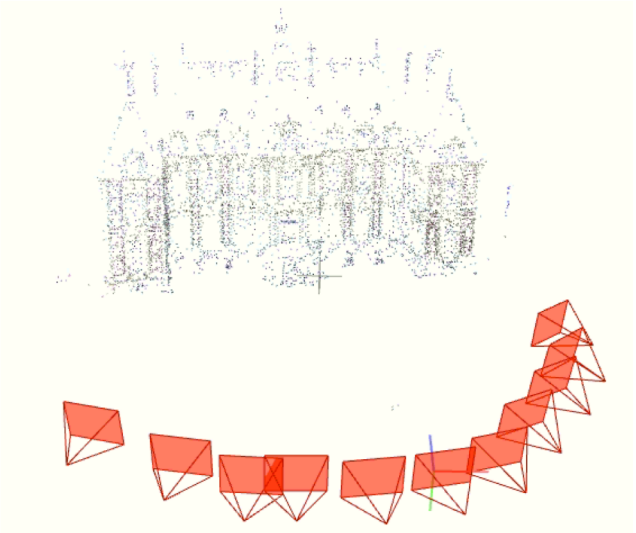


https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

Incremental SfM

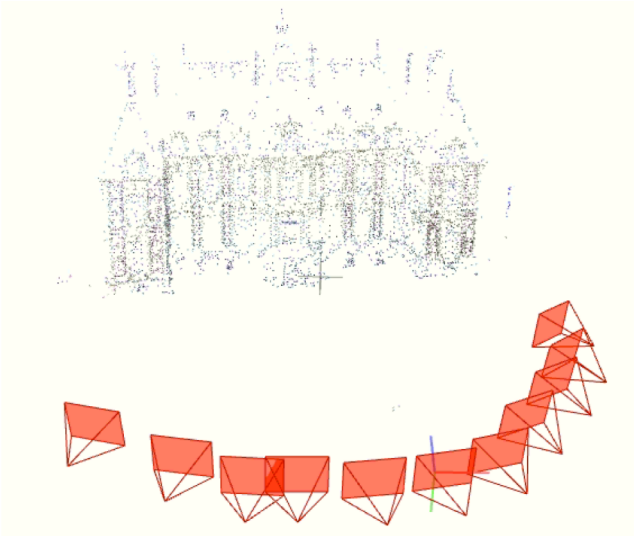
Incorporating views:

- Find 2D-3D correspondences
- Solve Perspective-n-Point problem
- Triangulate new points
- Bundle adjustment
- **Remove outliers**



https://demuc.de/tutorials/cvpr2017/sparse-modeling.pdf

Multi-view stereo



MVS

Image of "Chateau de Sceaux", Sceaux castle. France.
Photographer: Copyright 2012 Pierre MOULON http://imagine.enpc.fr/~moulonp/

Multi-view Stereo

From the original matches:

- **Back project to the original images.**
- **Search for colour correspondences in patches around the feature point.**

Furukawa and Ponce. Accurate, Dense, and Robust Multi-View Stereopsis. *IEEE PAMI* 2010.

Evaluation of image-based reconstruction

Evaluation of the image-based 3D reconstructions:

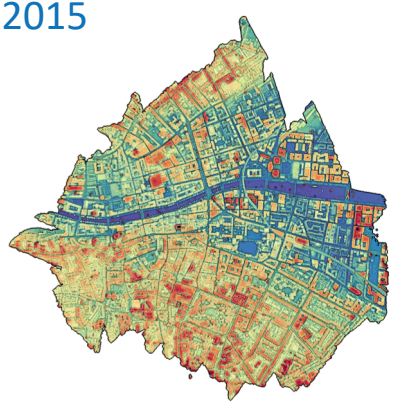
- **How precise it is**
- **How complete it is**

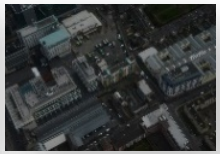
Ground-truth:

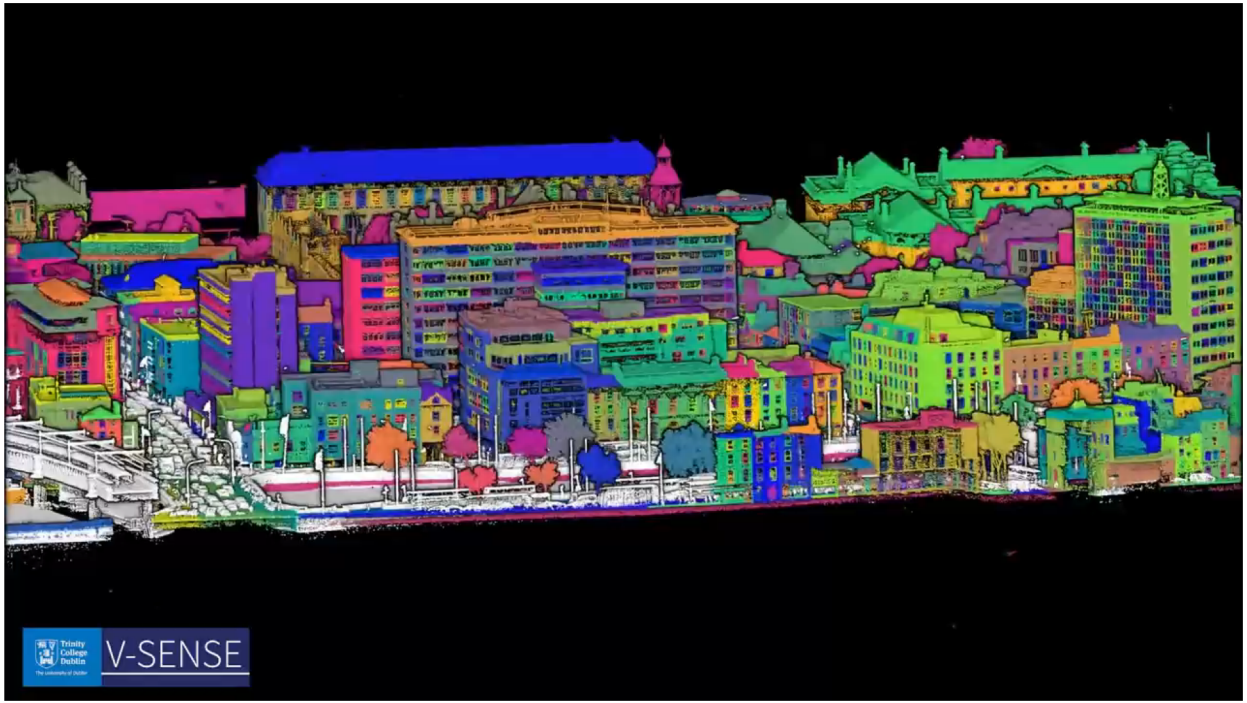
- **LiDAR**
- **Difficult to acquire**

Large scale image-based 3D reconstruction

Initial dataset [1]
2015







[1] Debra F Laefer, Saleh Abuwarda, Anh-Vu Vo, Linh Truong-Hong and Hamid Gharibi. 2015 aerial laser and photogrammetry survey of Dublin city collection record. https://geo.nyu. edu/catalog/nyu-2451-38684.

Image-based 3D reconstruction- Dublin city

The whole extension of the LiDAR point cloud is used as ground-truth for evaluation of image-based 3D reconstructions.

Top view reconstruction

Oblique view reconstruction

Evaluation per tile with the following measurements:
# of points, mean accuracy, precision, recall and F score.

The tile with the highest F score corresponds to an area of the city without river or green areas.

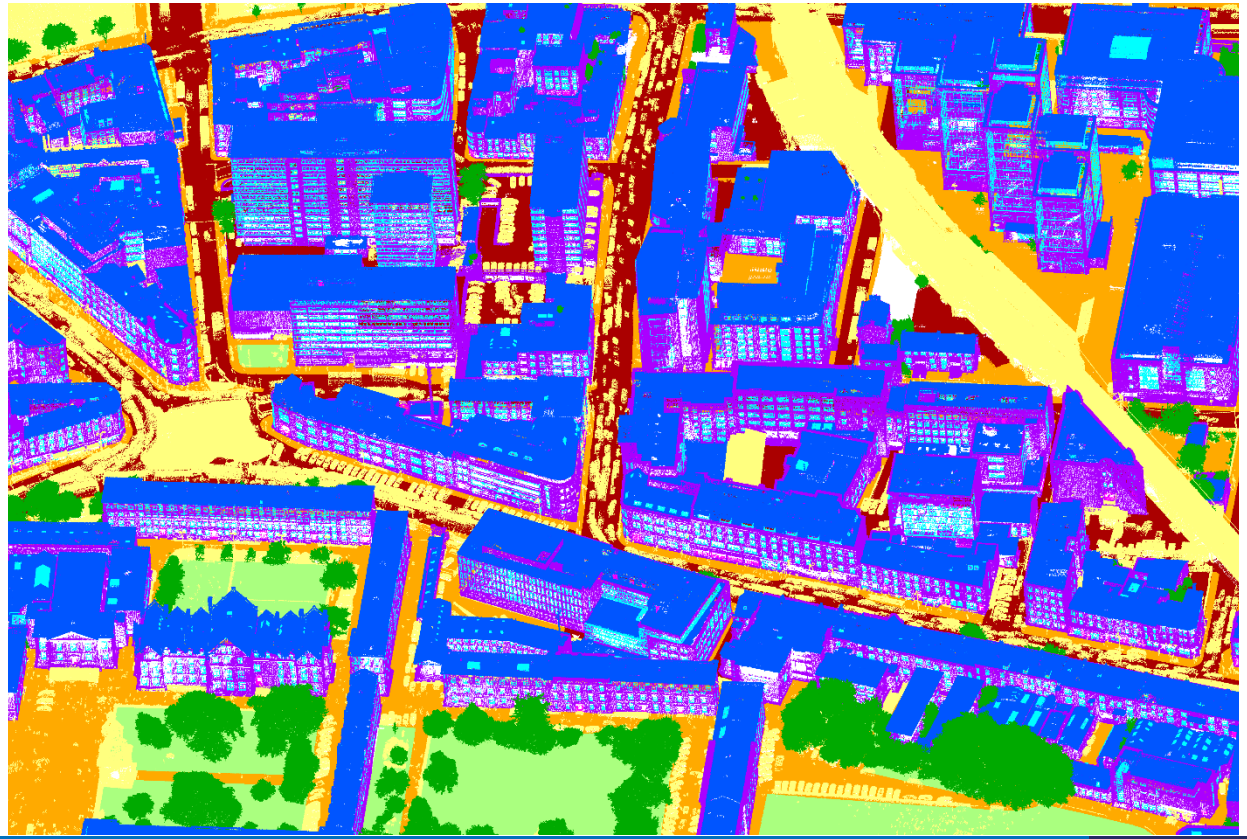
DublinCity: Annotated LiDAR Point Cloud and its Applications I. Zolanvari, S. Ruano, A. Rana, A. Cummins, R.E. da Silva, M. Rahbar, A. Smolic. BMVC 2019

Annotated LiDAR Point Cloud – Dublin city

Scene understanding of 3D models of an urban area remains a challenging task.

It is crucial to have a well-annotated ground-truth data available at city scale.

We present a novel manually annotated point cloud from Aerial Laser Scan data of the city of Dublin.

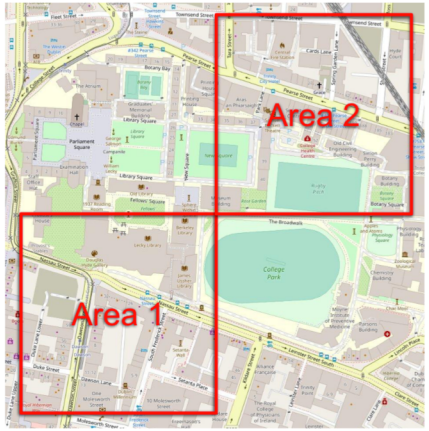


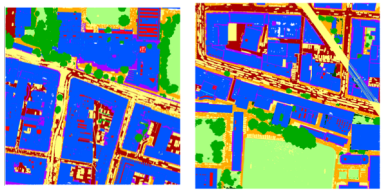
Ai3Dr benchmark

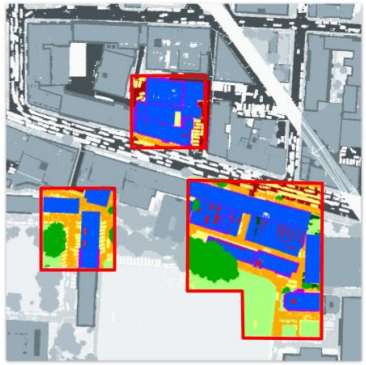
We presented a novel benchmark for evaluating image-based 3D reconstruction pipelines with aerial images in urban environments

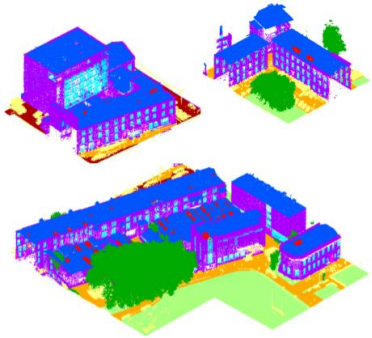
We evaluated state-of-the-art pipelines at scene level and per urban category, supporting previous hypothesis with quantitative measurements

We stimulated the evaluation of new approaches and provided an online evaluation to support the progress of research









A Benchmark for 3D Reconstruction from Aerial Imagery in an Urban Environment. S. Ruano and A. Smolic, VISIGRAPP 2021

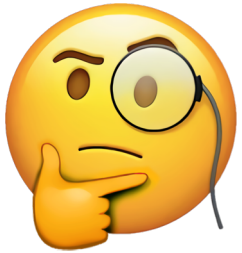
Human Reconstruction

>100 cameras → 12 cameras → just one?

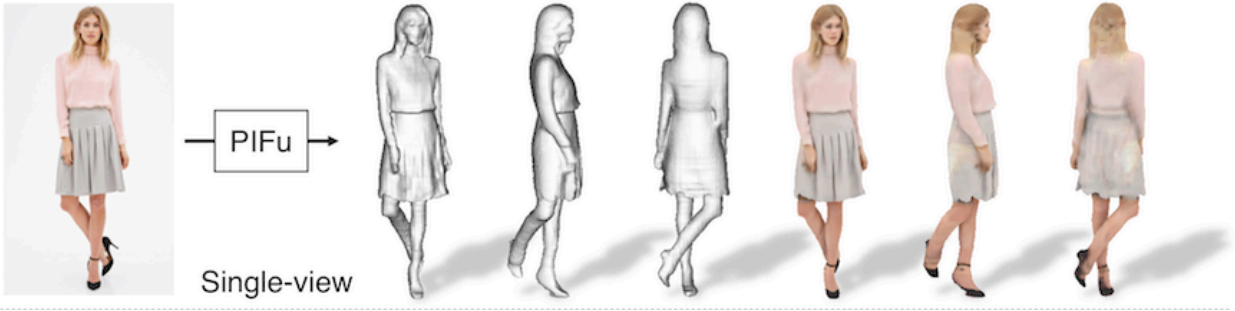
Microsoft

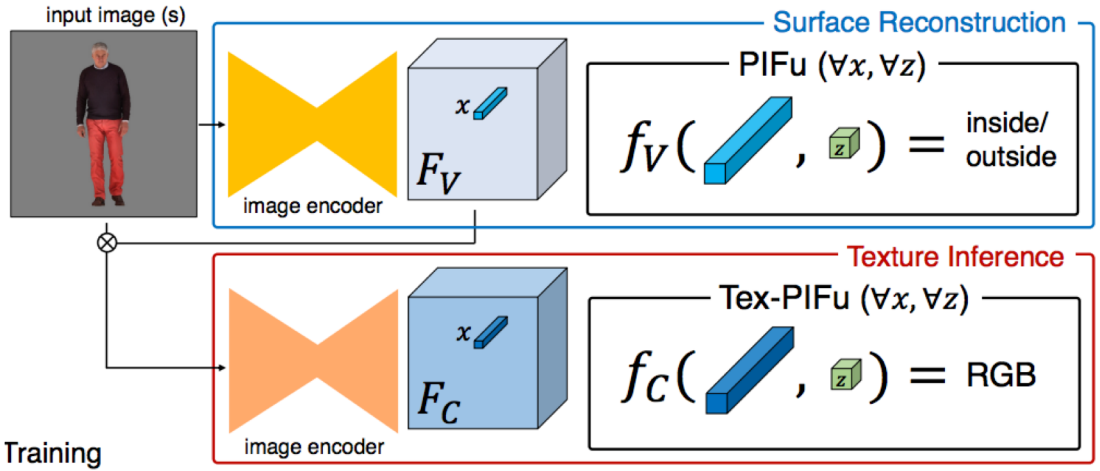
V-SENSE - Volograms

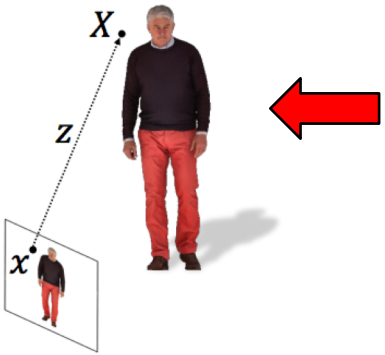
Affordable content creation for free-viewpoint video and VR/AR applications
R Pagés, K Amplianitis, D Monaghan, J Ondřej, A Smolić
Journal of Visual Communication and Image Representation, 2018



Human shape estimation from a single view







3D human scans are needed for training

PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization. Saito*, Huang*, Natsume*, Morishima, Kanazawa, Li. ICCV 2019.

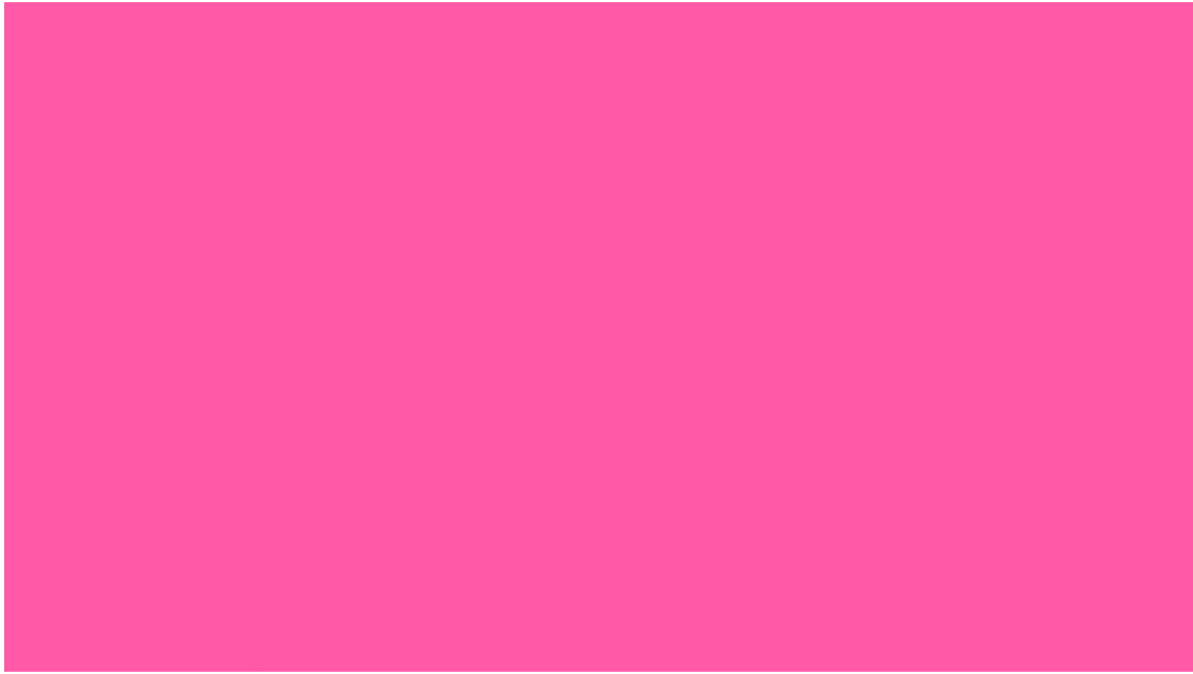
Further uses in the industry -Volograms

VOLOGRAMS

Volumetric video

Objective Making content creation for AR/VR accessible for everyone

Try it!
www.getvolu.com



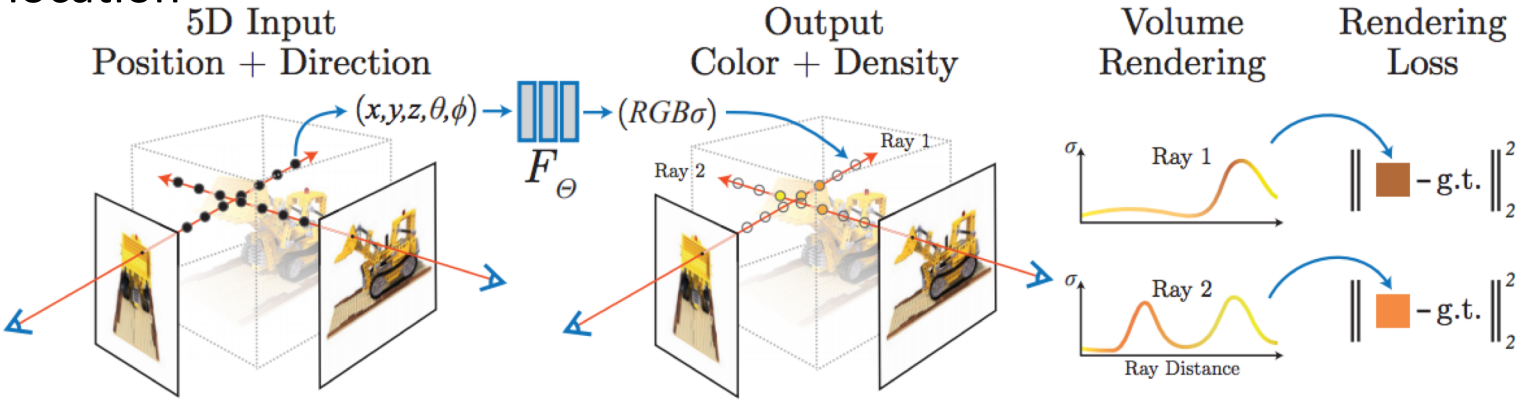
What else? -> View synthesis

NeRF

NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, Ravi Ramamoorthi, Ren Ng. ECCV 2020

What else? -> View synthesis

- **The scene is represented using a fully-connected deep network**
  - Input: 5D coordinate, spatial location (x,y,z) and viewing direction (θ, ϕ)
  - Output: volume density and view-dependent emitted radiance at a spatial location



NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, Ravi Ramamoorthi, Ren Ng. ECCV 2020

What else? -> View synthesis

Input: set of images + **camera poses**

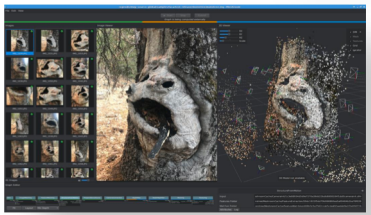
Output: new views

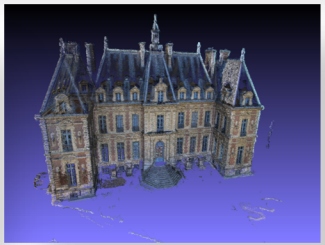
NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, Ravi Ramamoorthi, Ren Ng. ECCV 2020

Try it yourself!

Colmap
https://colmap.github.io/

OpenMVG
https://openmvg.readthedocs.io

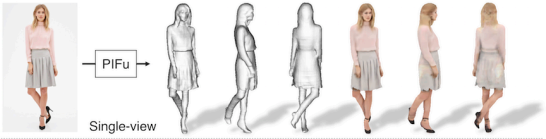




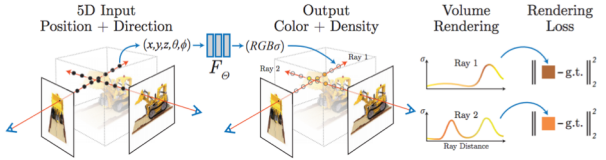
AliceVision - Meshroom
https://alicevision.github.io/

VOLU, AR CONTENT LIKE NEVER BEFORE

www.getvolu.com



https://github.com/shunsukesaito/**PIFu**



https://github.com/bmild/**nerf**

Trinity College Dublin

The University of Dublin

V-SENSE

Many Thanks!