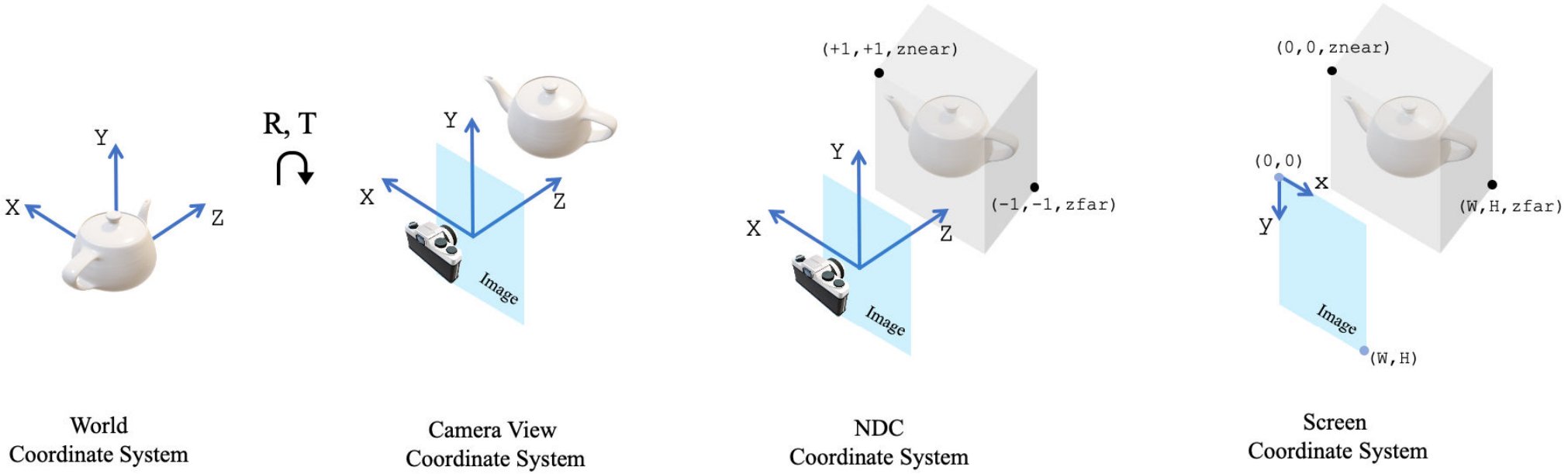
多视角几何重建
3D from Multiview

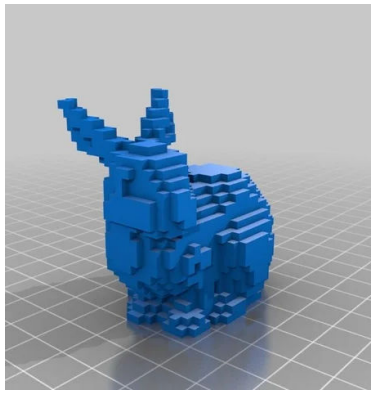
Basic Knowledge — Camera Model

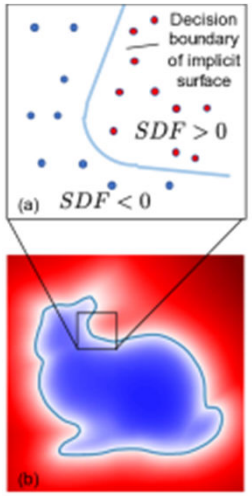


All Homogeneous Matrix

Basic Knowledge — 3D Representation&Rendering

2.5D / Image Based Rendering

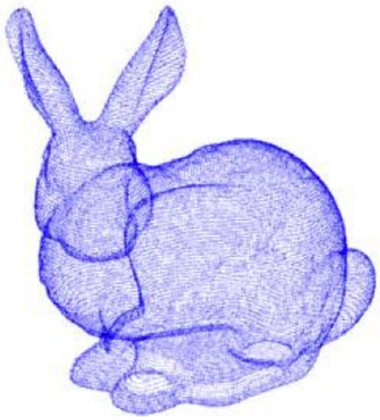




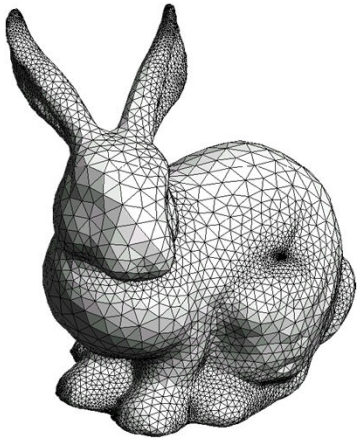
Volumetrics

Explicit

Implicit



Point clouds



Meshes

What we can do now in 3DV

Rendering

3D Surface

Light & Reflectance

Camera Process

3D Representation

What we cover today

Recon

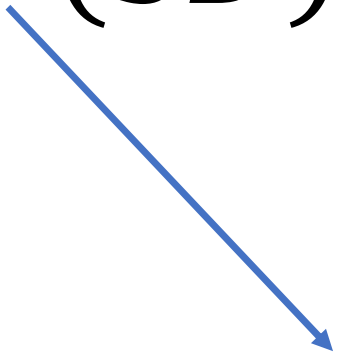
3D Surface

Light & Reflectance

How to obtain good 3D (in certain representation) from Multi-view Images?

Mathematic Formulation

$$F(3D) = 2D$$

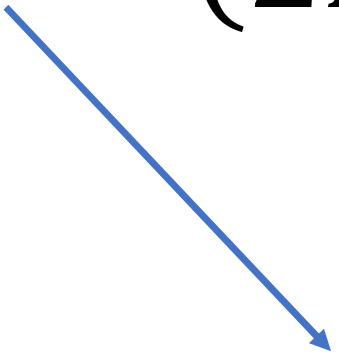


World space to camera space

Camera space projecting to (normalized space to) image space

Mathematic Formulation

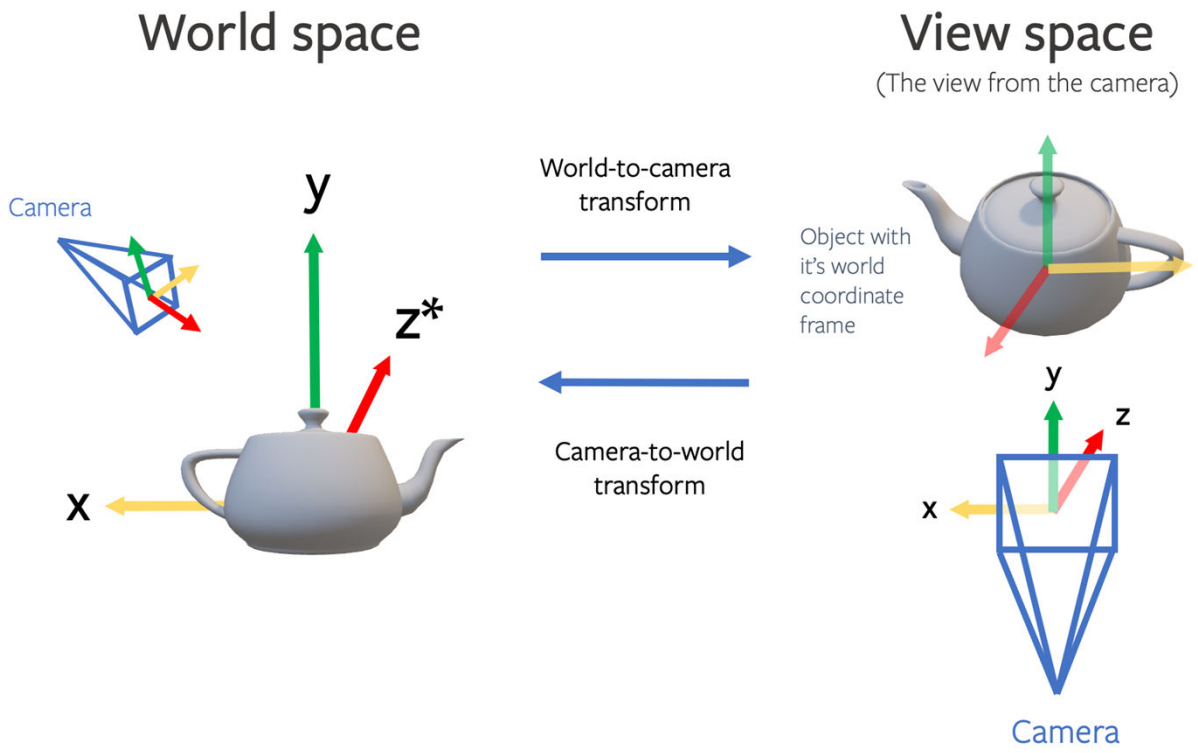
$$F^{-1}(2D) = 3D$$



Reconstruction: the inverse function

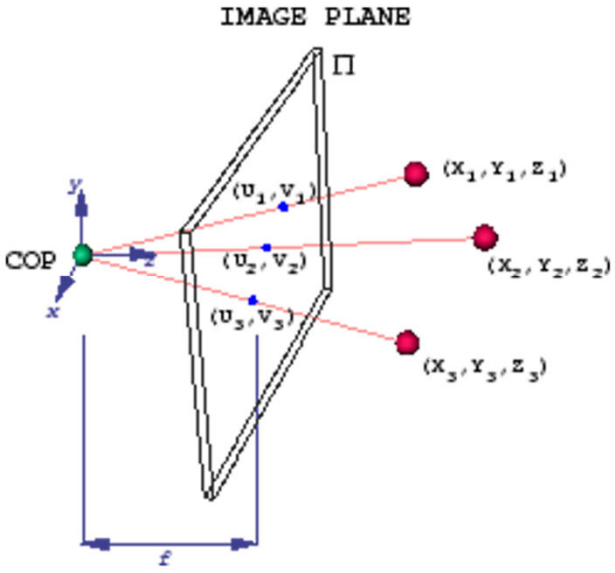
Solve for 3D from 2D observation

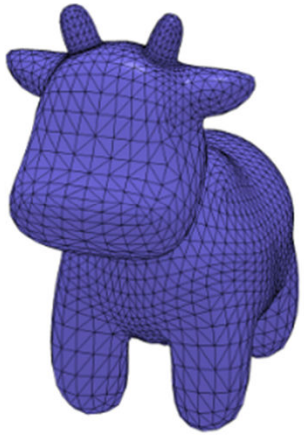
Detailed Formulation of F(3D) = 2D



$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + T$$

Detailed Formulation of F(3D) = 2D





$$u = f \cdot \frac{X_c}{Z_c} + c_x$$

$$v = f \cdot \frac{Y_c}{Z_c} + c_y$$

Detailed Formulation of F(3D) = 2D

$$u = f \cdot \frac{R_{11}X + R_{12}Y + R_{13}Z + T_1}{R_{31}X + R_{32}Y + R_{33}Z + T_3} + c_x$$

$$v = f \cdot \frac{R_{21}X + R_{22}Y + R_{23}Z + T_2}{R_{31}X + R_{32}Y + R_{33}Z + T_3} + c_y$$

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} R & T \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \qquad K = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

$$F(3D) = 2D$$

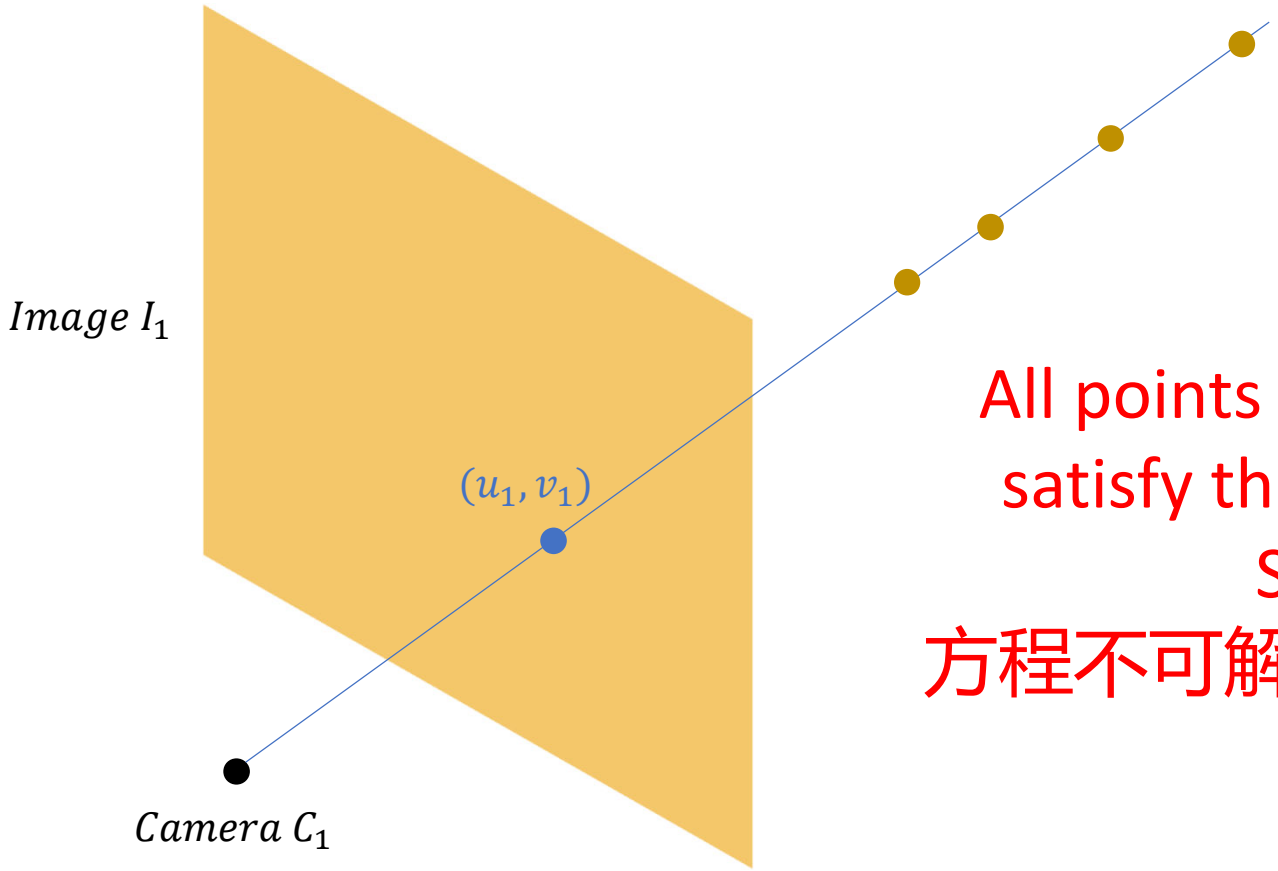
$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} R & T \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \qquad K = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

给定什么样的输入 / 列什么样的等式约束才能让方程的未知量：3D (XYZ)是可解的?

Assume for a Point (X,Y,Z)
We have 1 Camera / Image
and we know the Camera Parameters

$$u = f \cdot \frac{R_{11}X + R_{12}Y + R_{13}Z + T_1}{R_{31}X + R_{32}Y + R_{33}Z + T_3} + c_x$$
$$v = f \cdot \frac{R_{21}X + R_{22}Y + R_{23}Z + T_2}{R_{31}X + R_{32}Y + R_{33}Z + T_3} + c_y$$

two observations / equations (u,v)
three unknown variables (X, Y, Z)



All points on the line satisfy the function

So

方程不可解 / 无唯一解

Assume for a Point (X,Y,Z)
We have 2 Camera / Image
and we know Cameras' Parameters

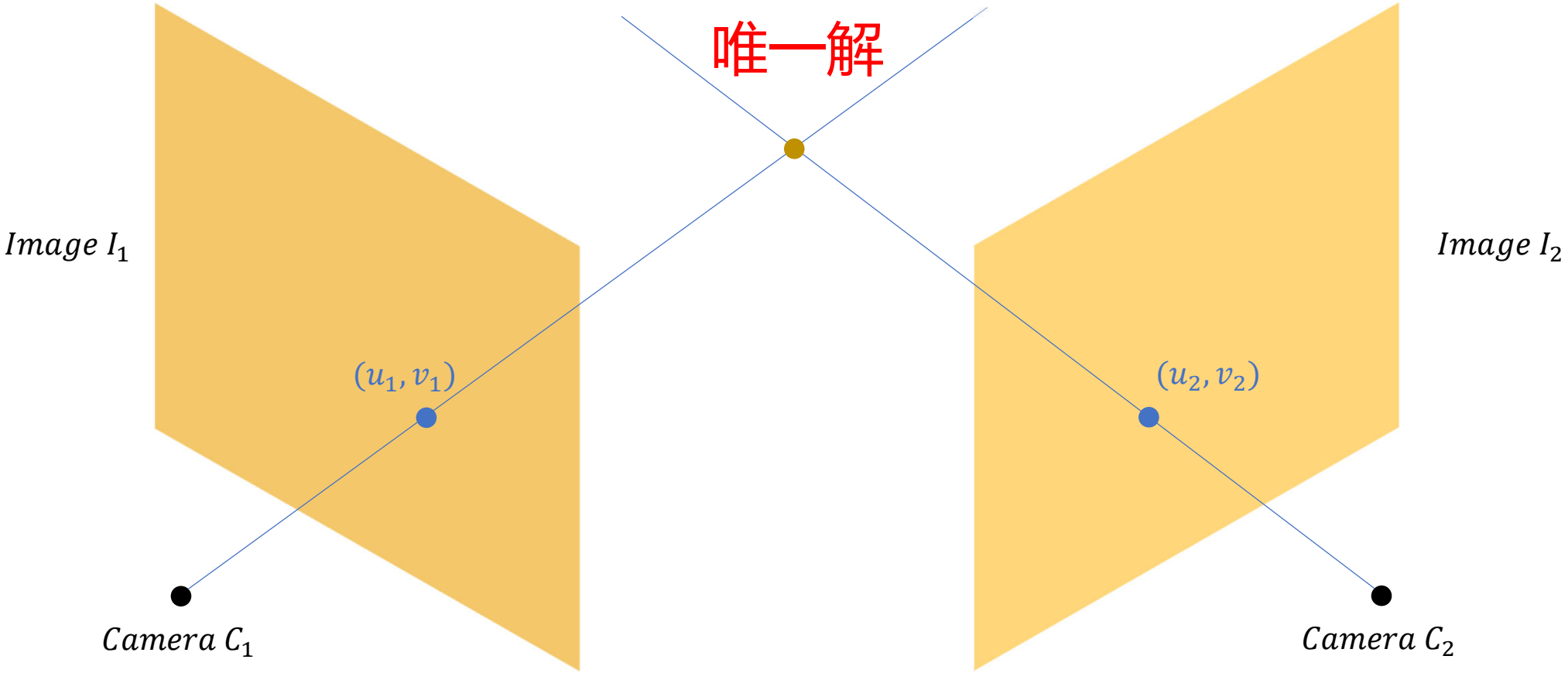
$$u = f \cdot \frac{R_{11}X + R_{12}Y + R_{13}Z + T_1}{R_{31}X + R_{32}Y + R_{33}Z + T_3} + c_x$$

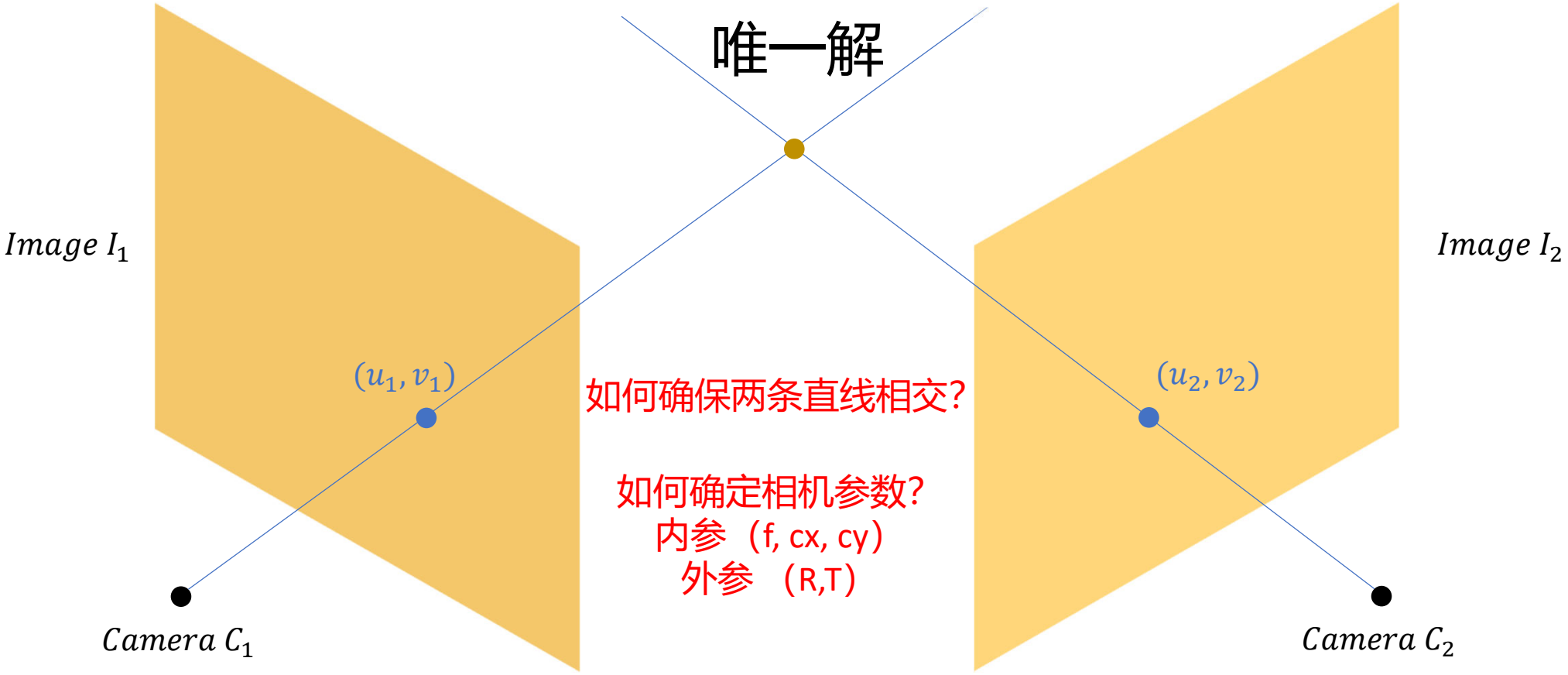
$$v = f \cdot \frac{R_{21}X + R_{22}Y + R_{23}Z + T_2}{R_{31}X + R_{32}Y + R_{33}Z + T_3} + c_y$$

4 observations / equations (u_i, v_i)

3 unknown variables (X, Y, Z), so

数学上是可解的


唯一解
Image I_1
Image I_2
(u_1, v_1)
(u_2, v_2)
Camera C_1
Camera C_2


唯一解
$Image\ I_1$
$Image\ I_2$
(u_1, v_1)
(u_2, v_2)
如何确保两条直线相交?
如何确定相机参数?
内参（f, cx, cy）
外参 （R,T）
$Camera\ C_1$
$Camera\ C_2$

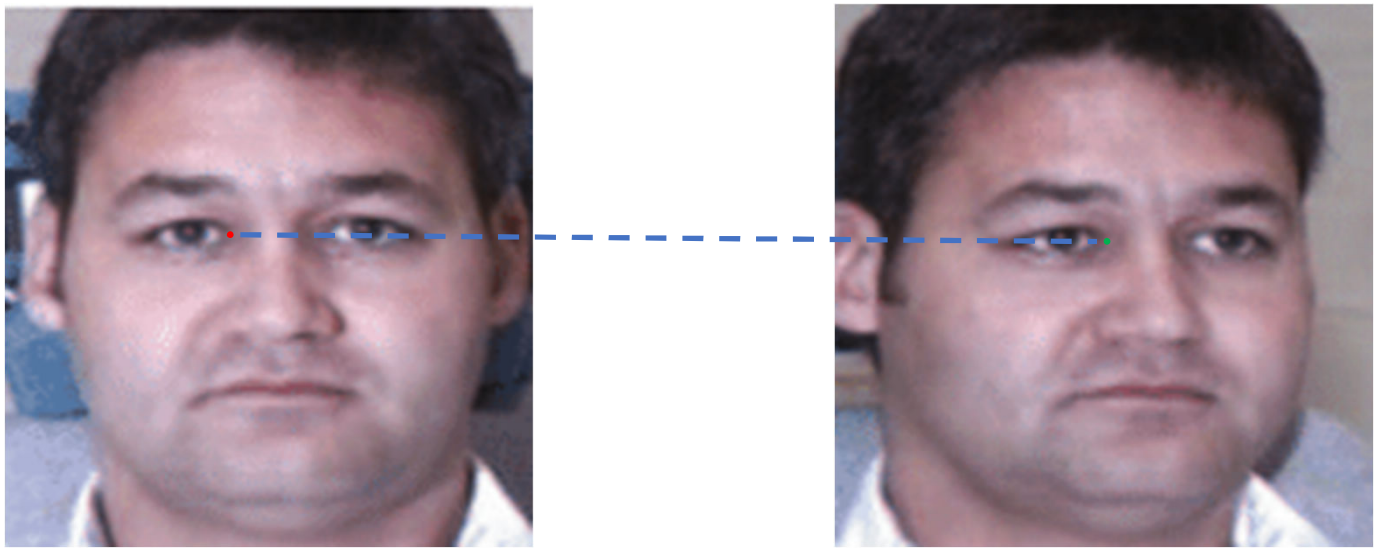
确保相交

如果(ui, vi)精确地是

同一个3D点的投影

那必定会相交在该点

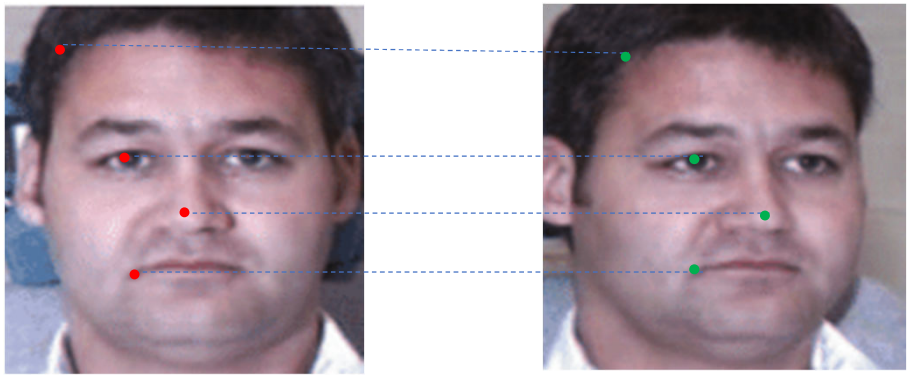
如何确保对应同一个3D点



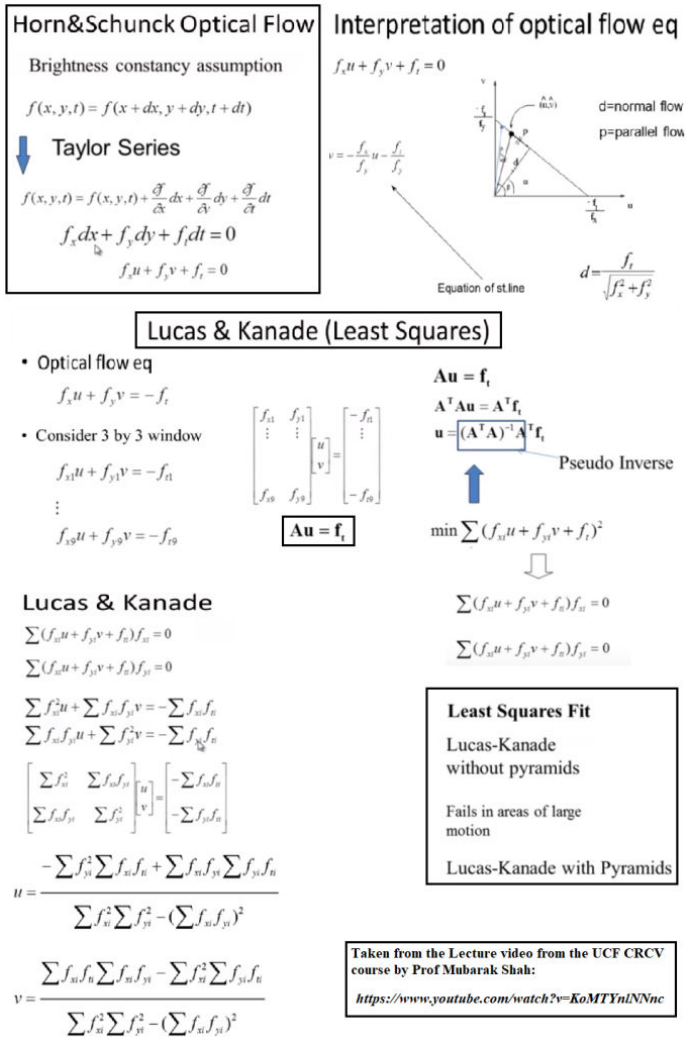
通过视觉（颜色）上的相似度

右图中是否存在和左图中红点同一个（3D）点的像素？

光流（Optical Flow）：局部区域颜色匹配找对应



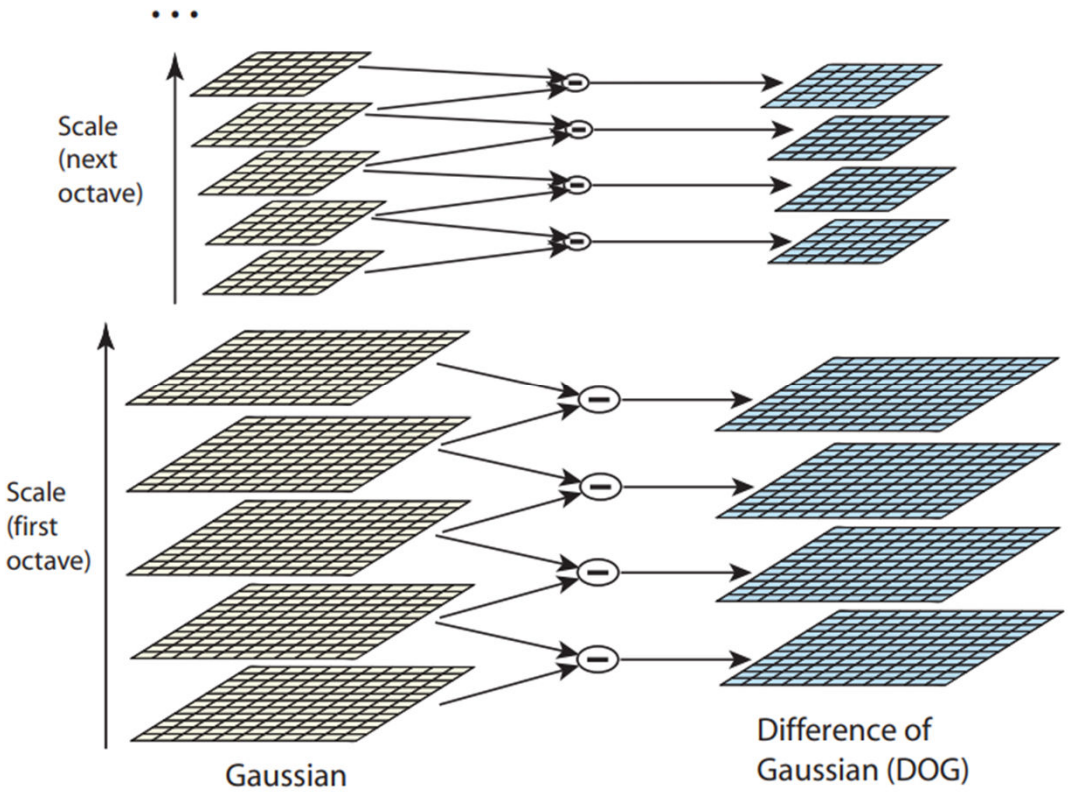
$$I'[y+\Delta y, x+\Delta x] = I[y,x]$$



But 局部颜色对光照、 Scale、 Rotation等不鲁棒

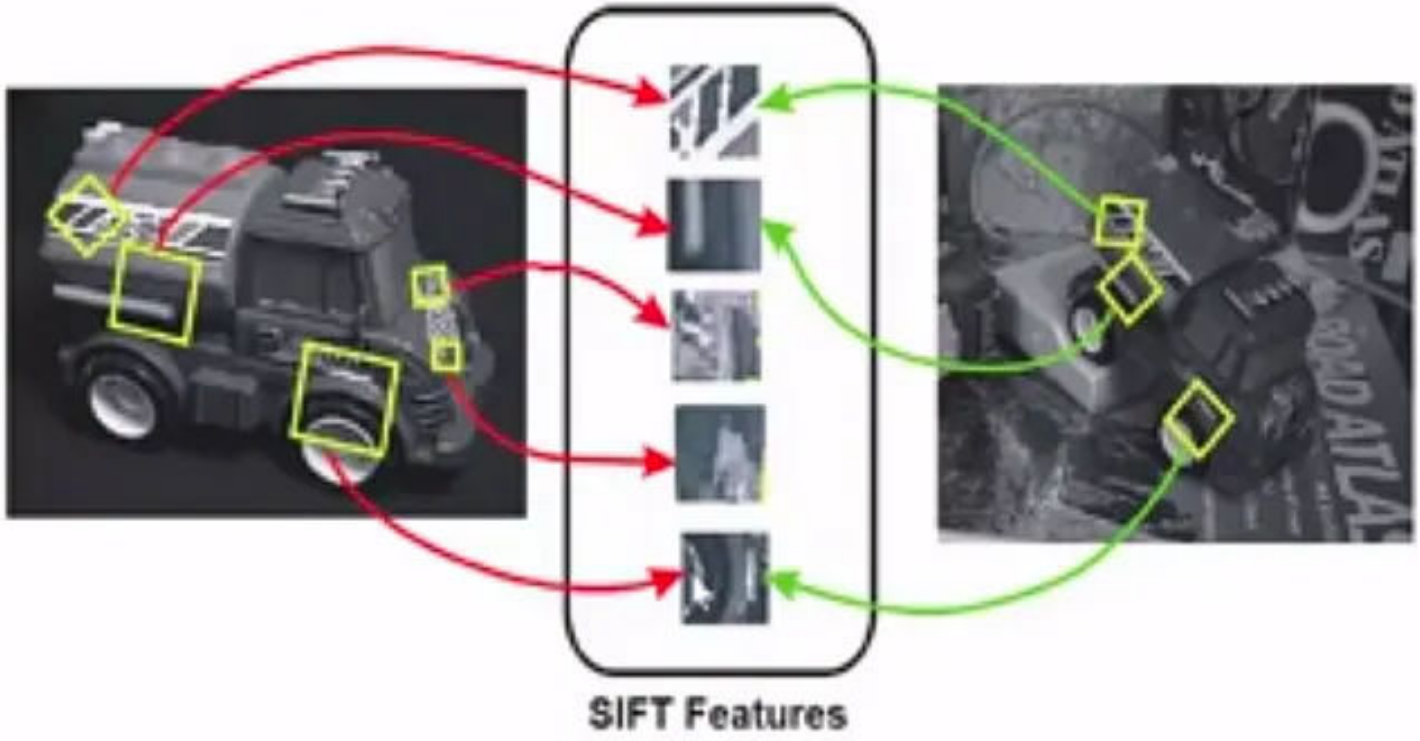
同一个对应点在几种case下的局部颜色度量相差很大

Scale-invariant feature transform (SIFT)



Distinctive Image Features from Scale-Invariant Keypoint. IJCV 2004.

Scale-invariant feature transform (SIFT)



Distinctive Image Features from Scale-Invariant Keypoint. IJCV 2004.

Cases SIFT Fails

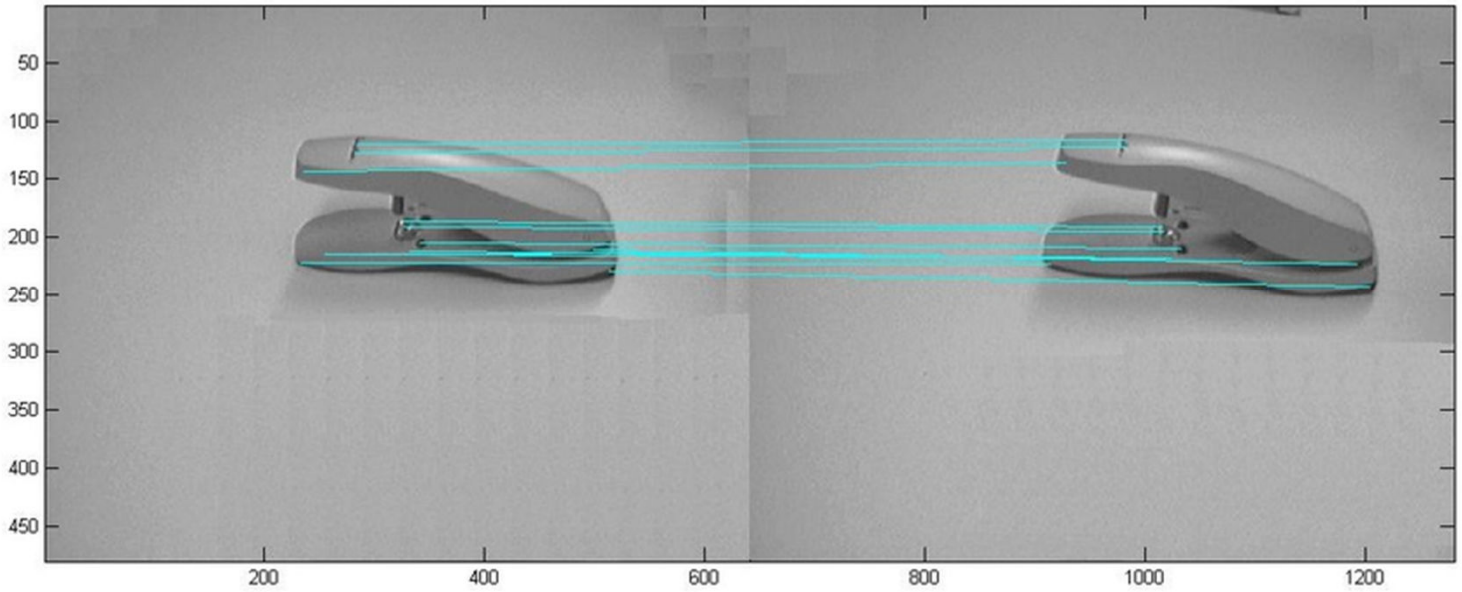
Hand-crafted

Small overlap

Large motion

Textureless

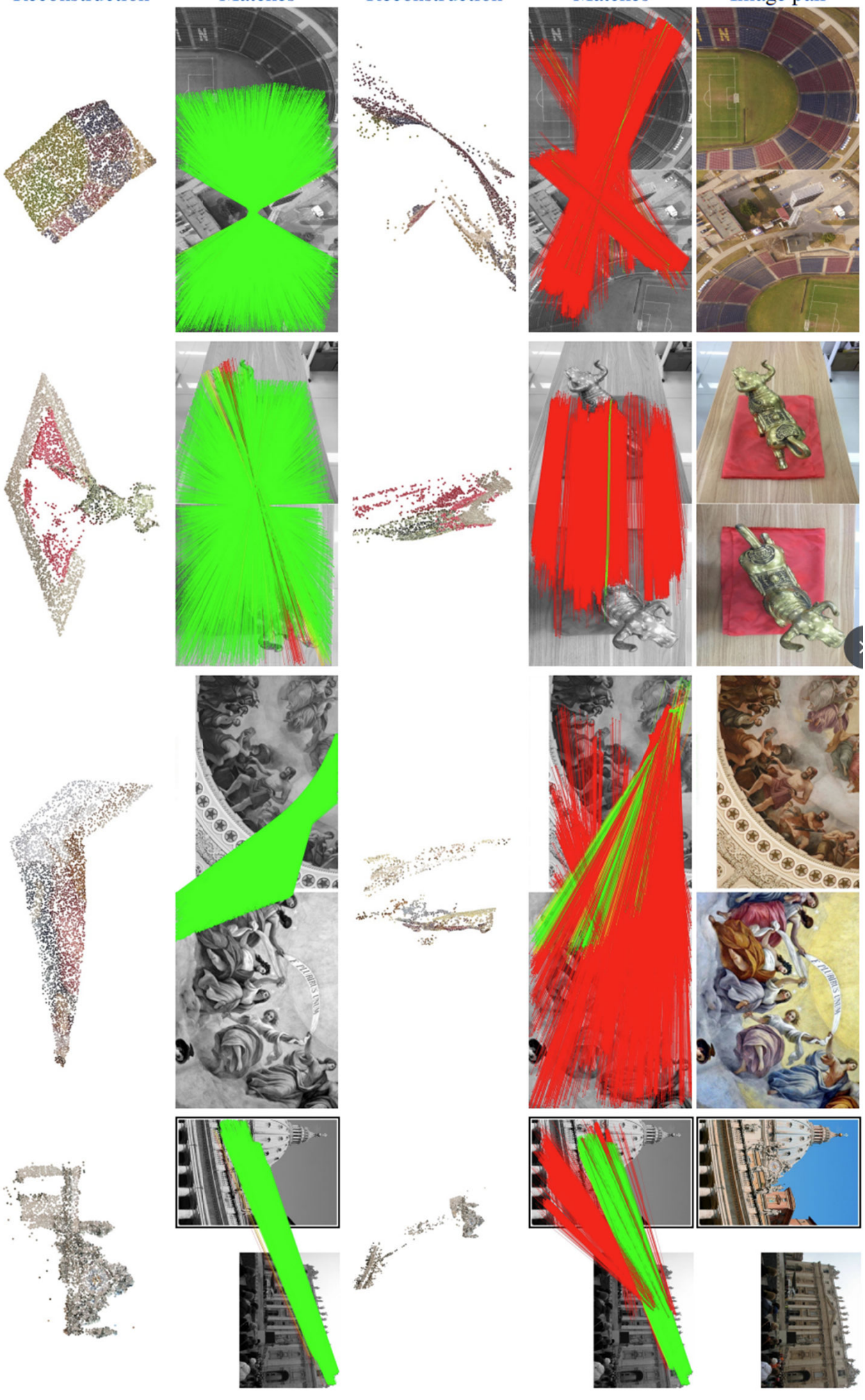
Sparse matching



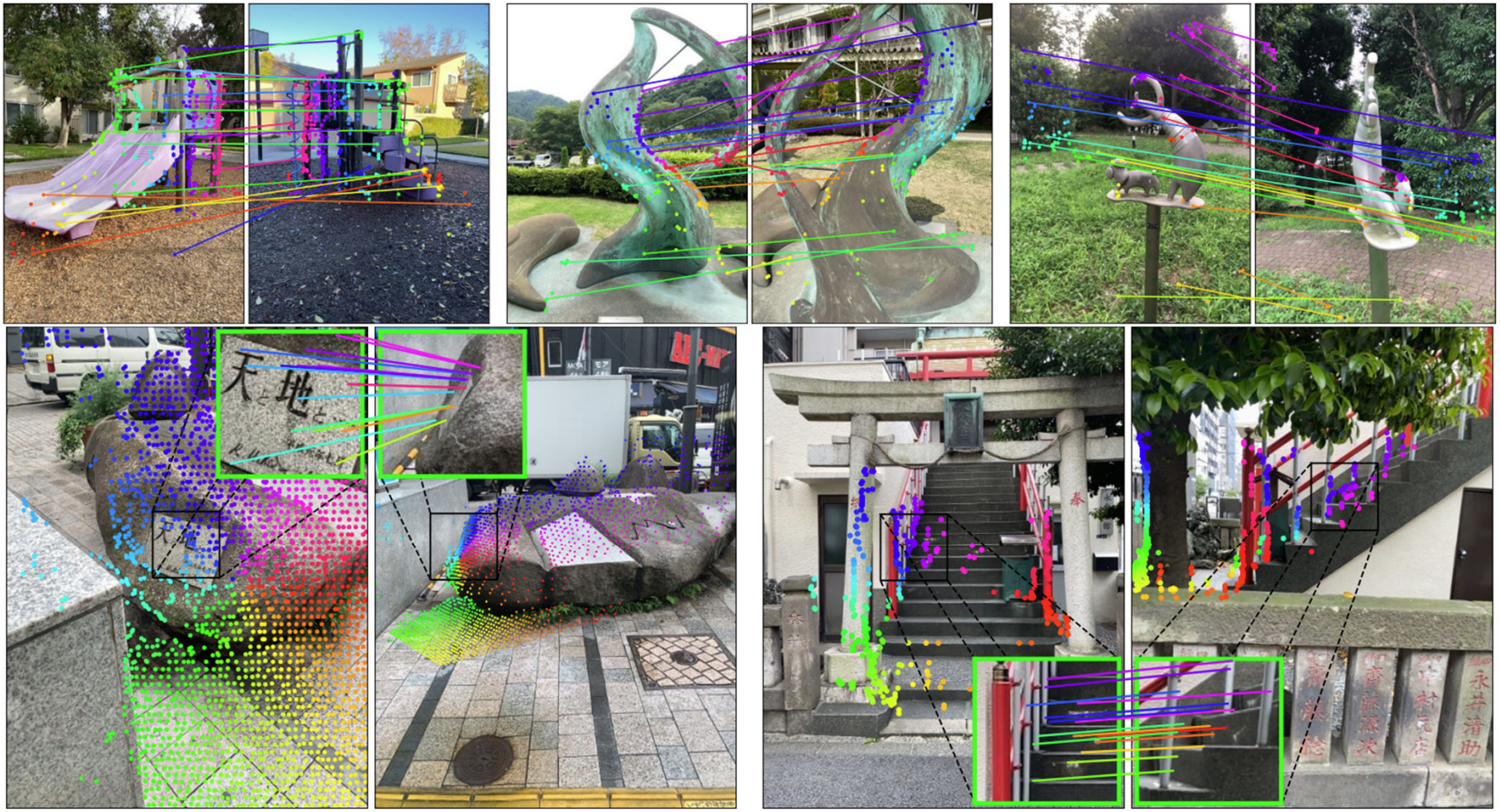
Deep Learning for Correspondence

Figure 2: **GIM framework**. We start by downloading a large amount of internet videos. Then, given a selected architecture, we first train it on standard datasets, and generate correspondences between nearby frames by using the trained model with multiple complementary image matching methods. The self-training signal is then enhanced by 1) filtering outlier correspondences with robust fitting, 2) propagating correspondences to distant frames and 3) injecting strong data augmentations.

GIM: LEARNING GENERALIZABLE IMAGE MATCHER FROM INTERNET VIDEOS. ICLR 2024.

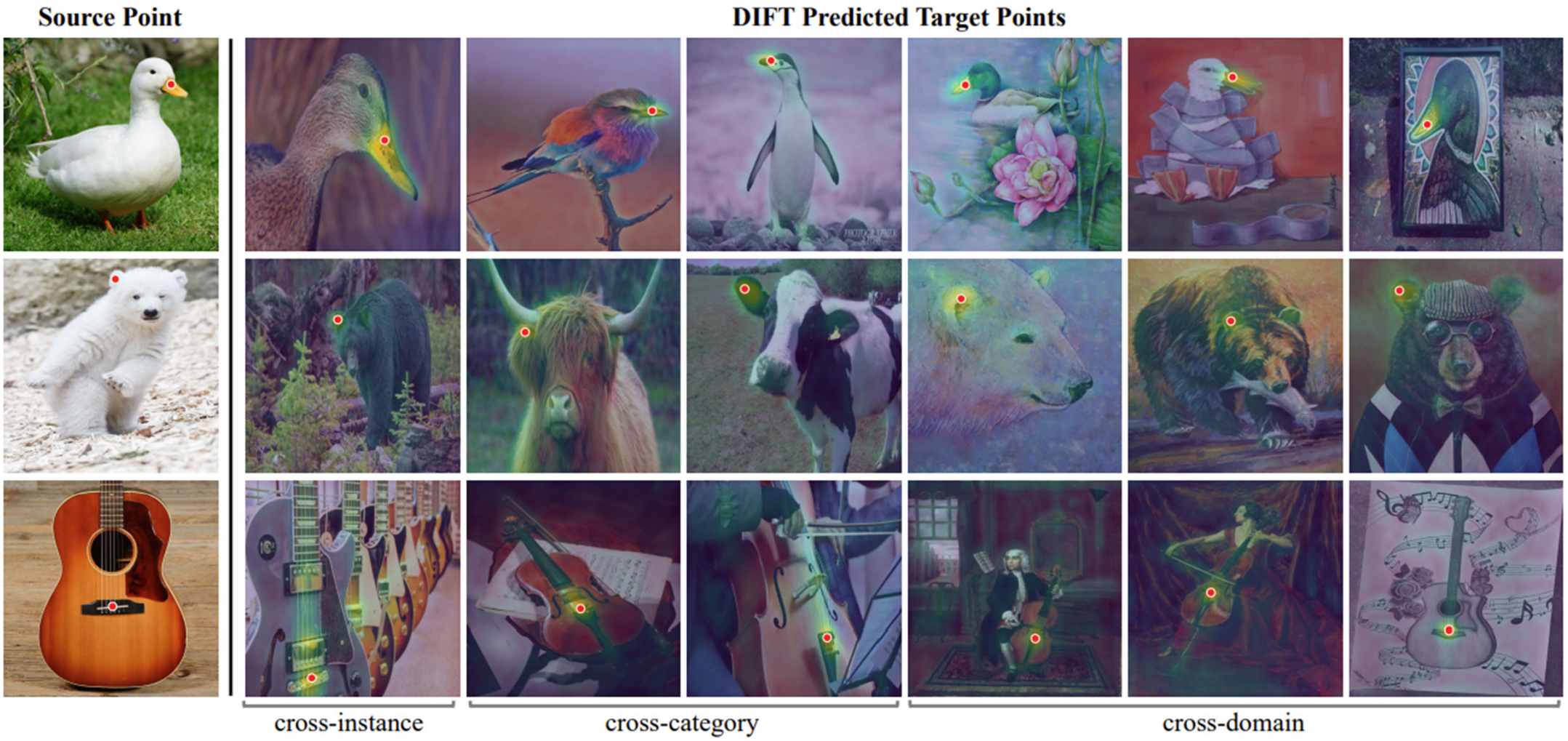

GIM$_{DKM}$ Reconstruction
GIM$_{DKM}$ Matches
DKM (IN) Reconstruction
DKM (IN) Matches
Image pair

Correspondence for Large Pose Difference



Grounding Image Matching in 3D with MASt3R. ArXiv 2024.

Correspondence between Different Objects



Emergent Correspondence from Image Diffusion. NeurIPS 2023.

More Applications of Dense Correspondence

CoDeF: Content Deformation Fields for Temporally Consistent Video Processing. CVPR 2024.


唯一解
$Image\ I_1$
$Image\ I_2$
(u_1, v_1)
(u_2, v_2)
通过精确像素对应保证相交
如何确定相机参数?
内参（f, cx, cy）
外参 （R,T）
$Camera\ C_1$
$Camera\ C_2$

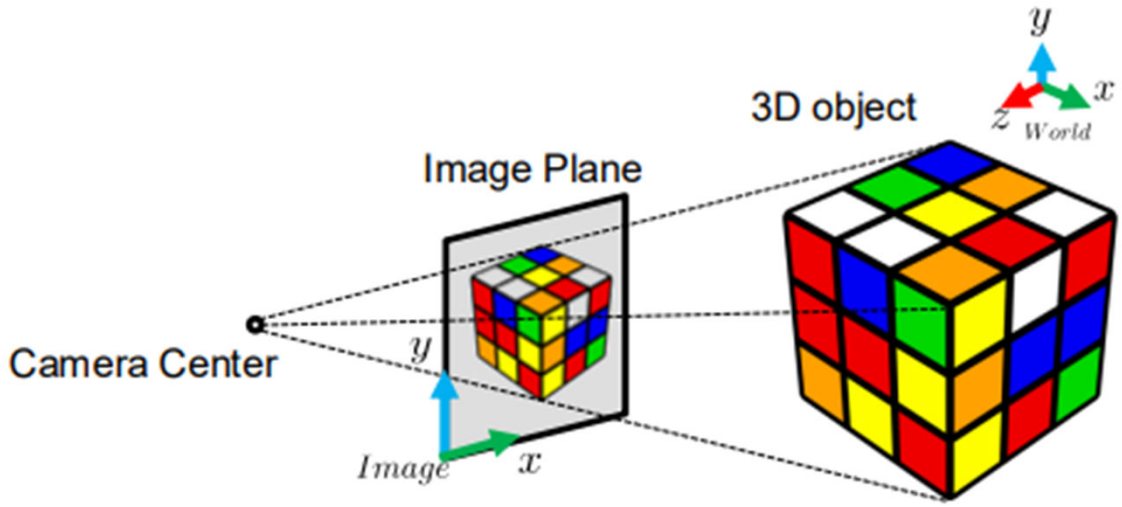
相机标定（Camera Calibration）

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K \begin{bmatrix} R & T \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

$$\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$$

$$\mathbf{P} = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_1 & r_2 & r_3 & t_1 \\ r_4 & r_5 & r_6 & t_2 \\ r_7 & r_8 & r_9 & t_3 \end{bmatrix}$$

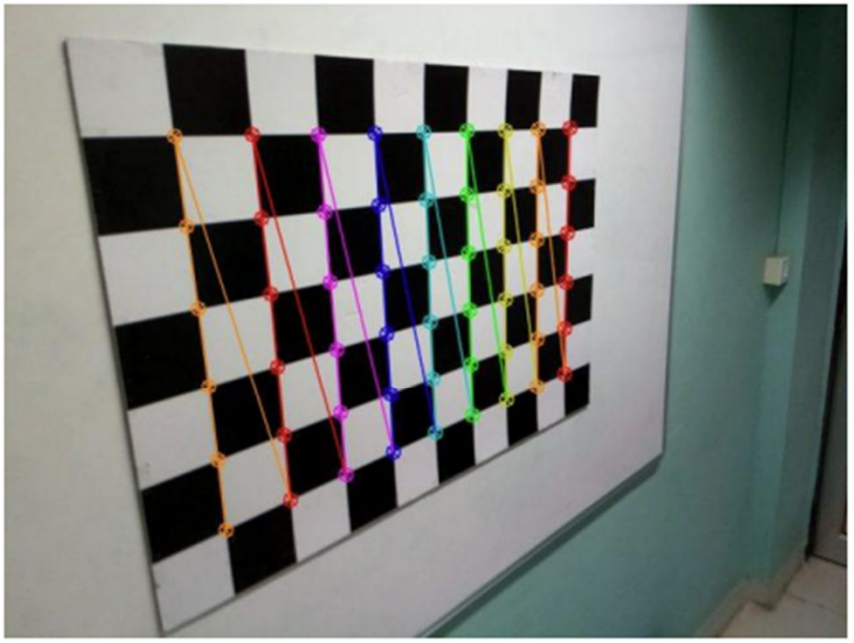
intrinsic parameters　　extrinsic parameters



Assume I have a Checkerboard (标定板)

Capture multiple images of the checkerboard from different viewpoints

Find checkerboard corners



Know the 2D & 3D

If I know N 2D & 3D pairs across K Cameras / Images

$$u = f \cdot \frac{R_{11}X + R_{12}Y + R_{13}Z + T_1}{R_{31}X + R_{32}Y + R_{33}Z + T_3} + c_x$$
$$v = f \cdot \frac{R_{21}X + R_{22}Y + R_{23}Z + T_2}{R_{31}X + R_{32}Y + R_{33}Z + T_3} + c_y$$

2*N*K observations / equations (u_i, v_i)

K*9 unknown variables (3 for (f, cx, cy), 6 for (R, T)), so

If N >=5, it could be solved

How to Solve?

Optimize with Non-Linear Least Squares

$$\min_{\{\mathbf{R}_k,\mathbf{t}_k\},\mathbf{K}} \sum_i \sum_k \|\mathbf{x}_{ik} - \pi_{\mathbf{K}}([\mathbf{R}_k|\mathbf{t}_k]\mathbf{X}_i)\|_2^2$$

Extrinsic

Intrinsic

Detection Corner Points

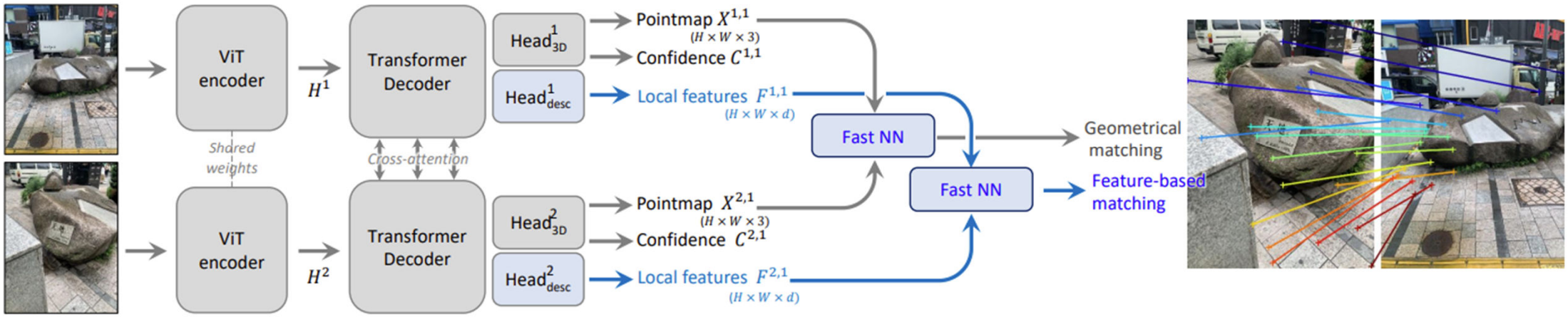
Perspective Projection

Known 3D Location

$$\boldsymbol{\beta}^{(s+1)} = \boldsymbol{\beta}^{(s)} - \left(\mathbf{J_r}^{\mathrm{T}}\mathbf{J_r}\right)^{-1}\mathbf{J_r}^{\mathrm{T}}\mathbf{r}\left(\boldsymbol{\beta}^{(s)}\right)$$

What if we do not have Checkerboard?

But we could have corresponding points



We know 2D

Assume N pairs in 2 Cameras / Images

$$\min_{\{\mathbf{R}_k,\mathbf{t}_k\},\mathbf{K}} \sum_i \sum_k \|\mathbf{x}_{ik} - \pi_{\mathbf{K}}([\mathbf{R}_k|\mathbf{t}_k]\mathbf{X}_i)\|_2^2$$

Extrinsic — Intrinsic — Detection Corner Points — Perspective Projection — Known 3D Location

2*N*2 observations / equations (u_i, v_i)

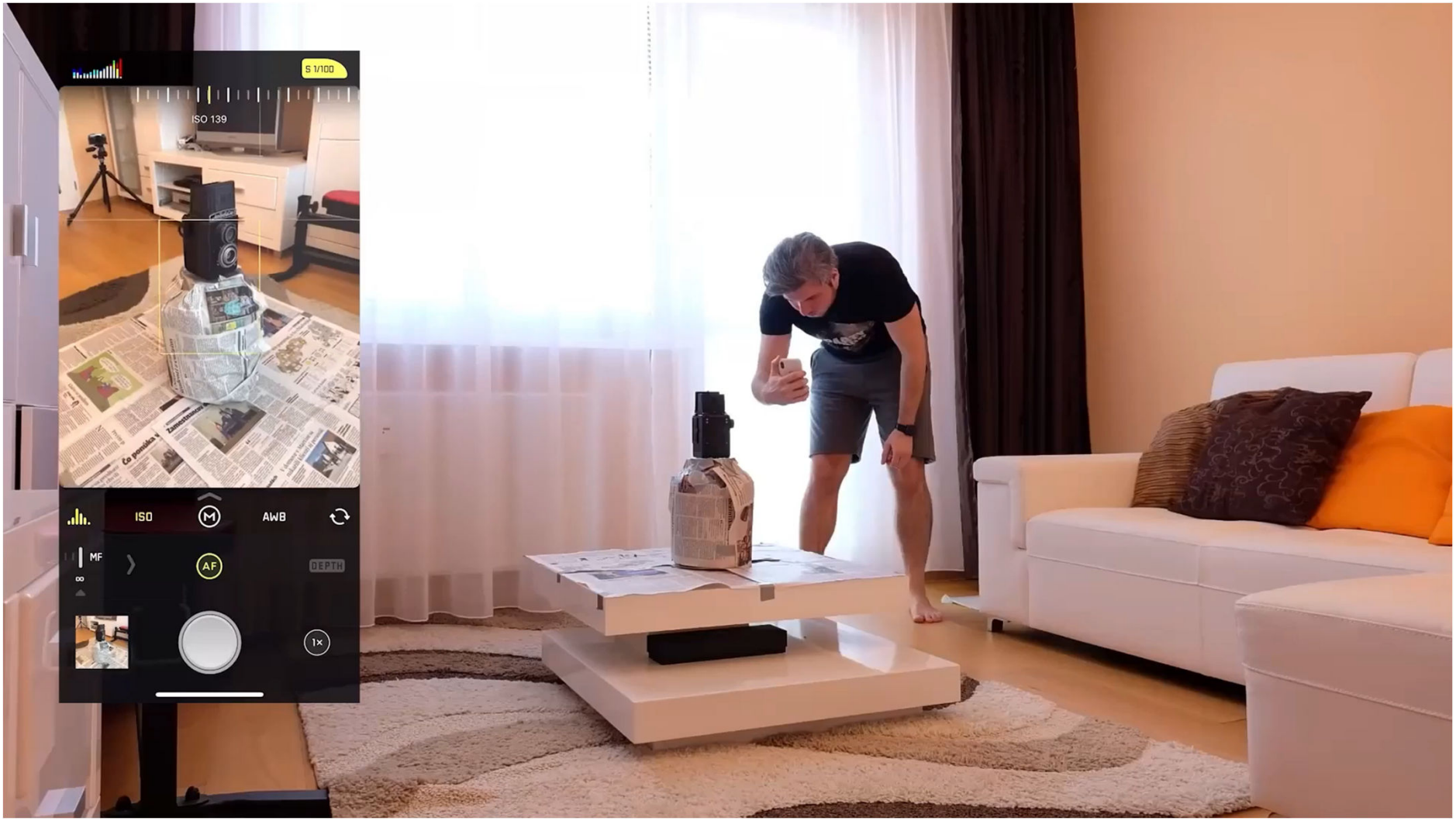
Camera: 2*9 unknown variables

3D: 3*N unknown, so

If N>=18, it colud be solved

Also could be solved with Gaussi-Newton

Extend to Multiple Views — Bundle Adjustment



Step by Step Implementation

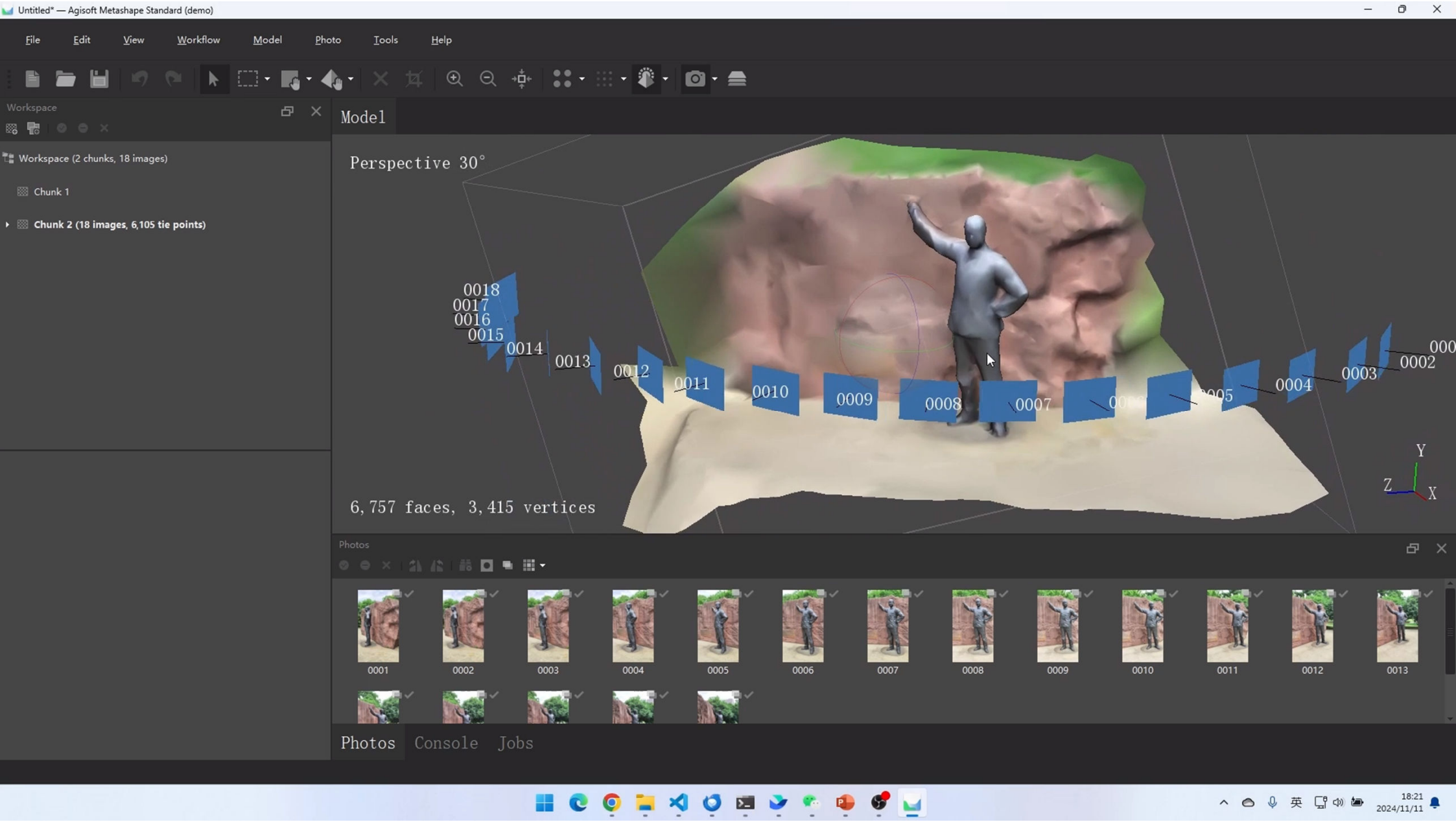
Record Images / Videos

Bundle Adjustment Library / Software

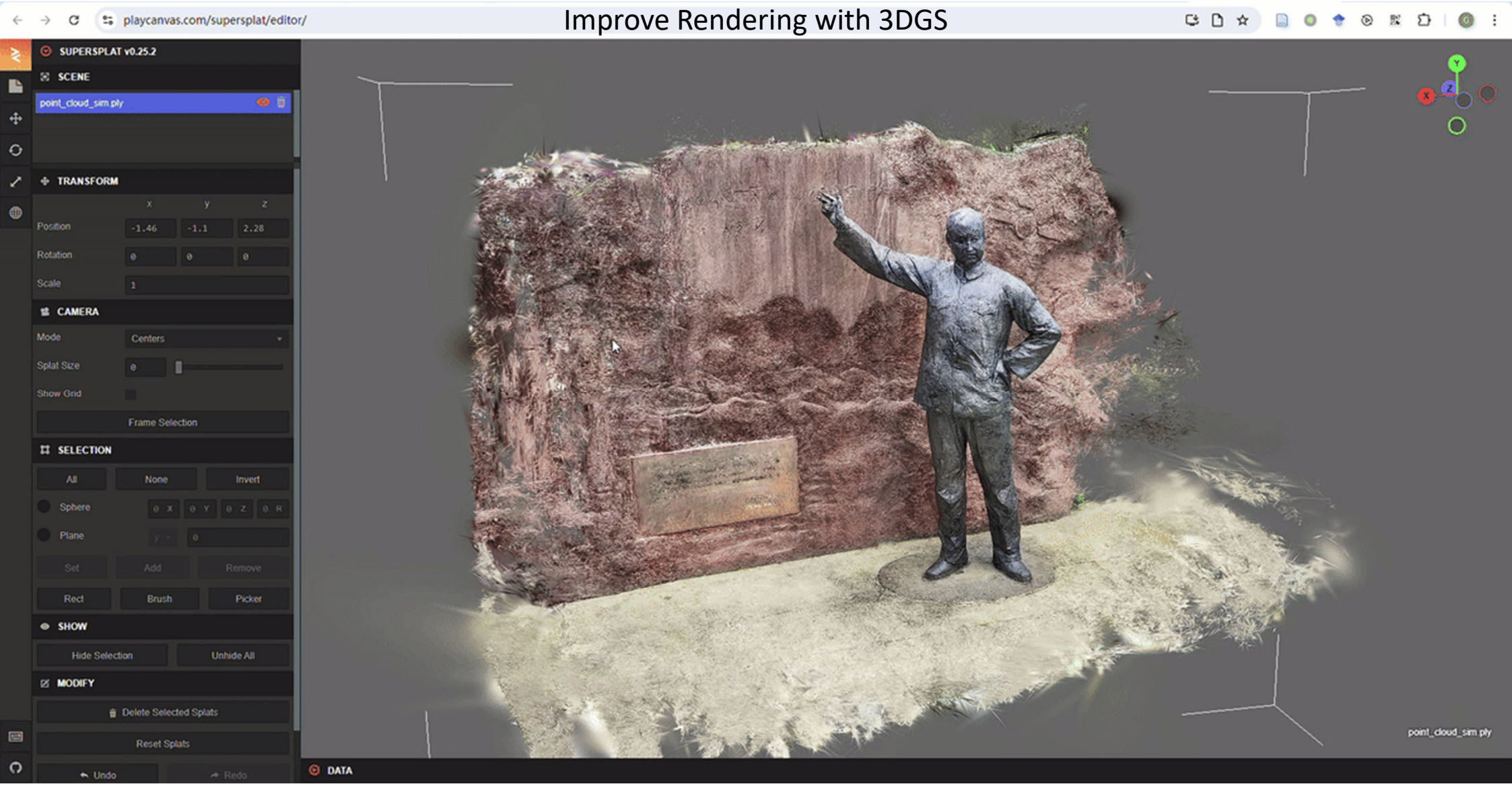
colmap

Meshroom

Metashape


Untitled* — Agisoft Metashape Standard (demo)
File
Edit
View
Workflow
Model
Photo
Tools
Help
Workspace
Workspace (2 chunks, 18 images)
Chunk 1
Chunk 2 (18 images, 6,105 tie points)
Model
Perspective 30°
0018
0017
0016
0015
0014
0013
0012
0011
0010
0009
0008
0007
0004
0003
0002
6,757 faces, 3,415 vertices
Y
Z
X
Photos
0001
0002
0003
0004
0005
0006
0007
0008
0009
0010
0011
0012
0013
Photos
Console
Jobs
18:21
2024/11/11

Improve Rendering with 3DGS



中国科学技术大学

University of Science and Technology of China

谢谢观看!