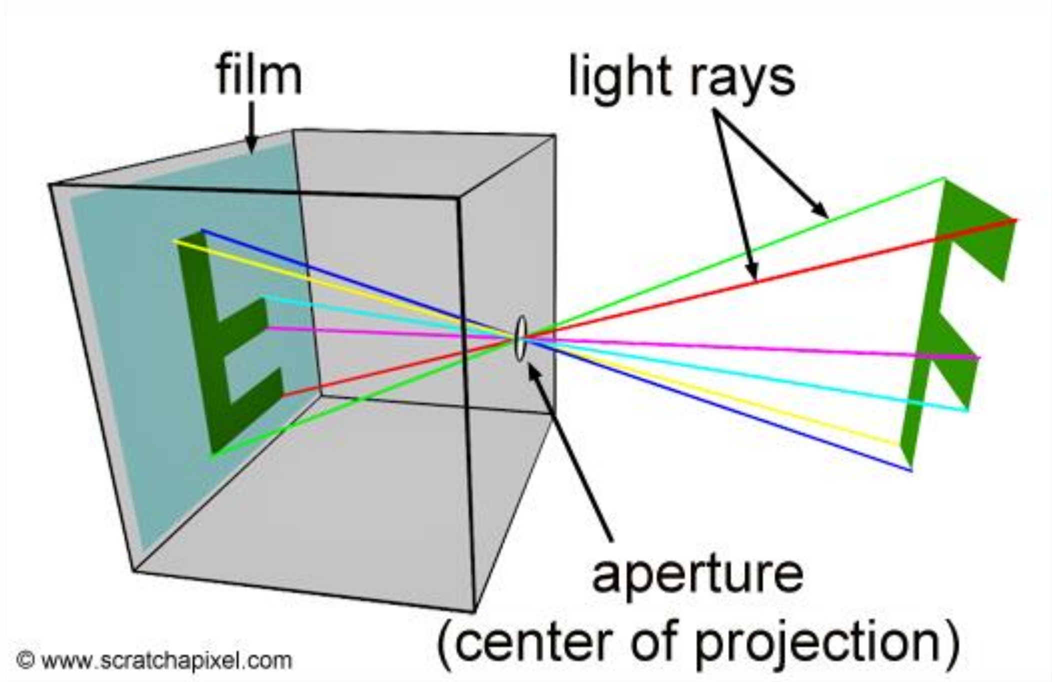
3D reconstruction using structured light

Pinhole camera model

- Pinhole camera — camera with infinitesimal aperture
- Due to very small aperture very limited amount of rays from object point hits “sensor”
- As a result, image is formed since blur quite low
- But more importantly, such camera transforms points in 3D space into 2D points in image plane (projection)
- Transformation partially reversible, because information about distance is lost
- Hence, 2D points in image plane allow to restore only direction to object point (no distance information)



Projective transformation by a pinhole camera model

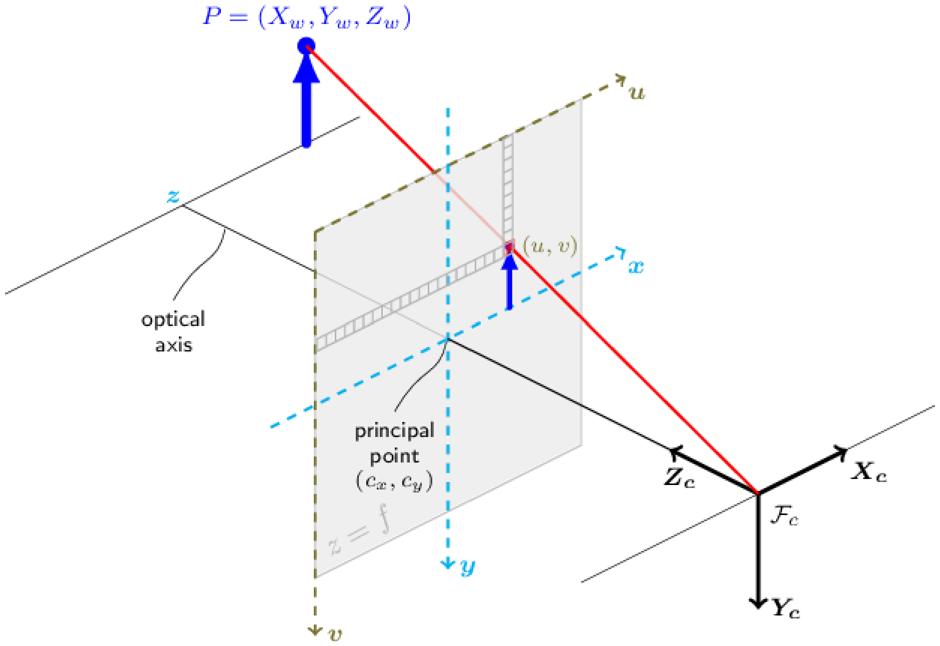
$$s\,p = A\,[R|t]\,P_w$$

$$A = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

$$Z_c \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = [R|t] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix},$$

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$



Projective transformation by a pinhole camera model

Intrinsic matrix

Rotation matrix

Translation vector

Scale factor (distance)

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

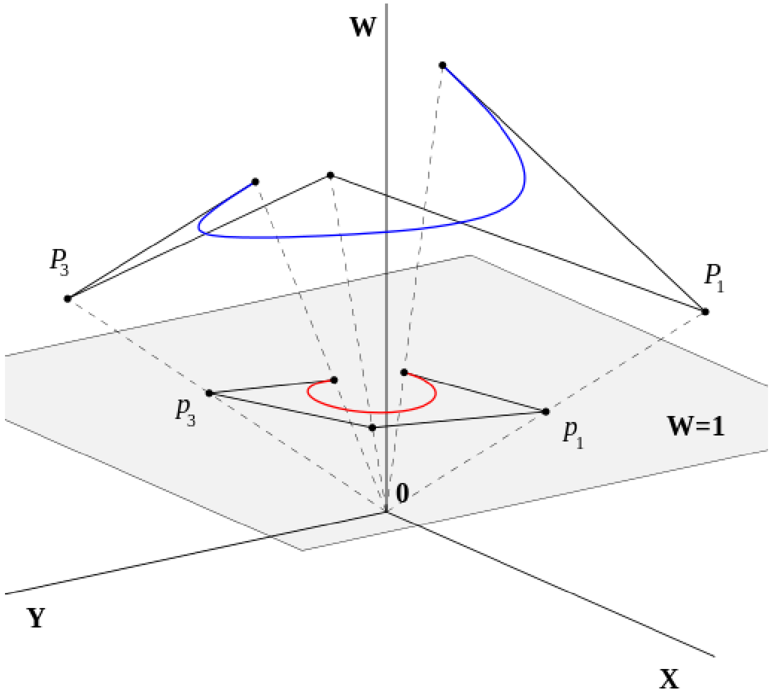
Pixel coordinates

World 3D point coordinates

Homogeneous coordinate system

- Also called projective coordinate system
- Can be thought as the Euclidean space with additional points, which are called points at infinity
- Allow common vector operations such as translation, rotation, scaling and **perspective projection** to be represented as a matrix by which the vector is multiplied

$$\begin{bmatrix} X \\ Y \\ W \end{bmatrix} = s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$$



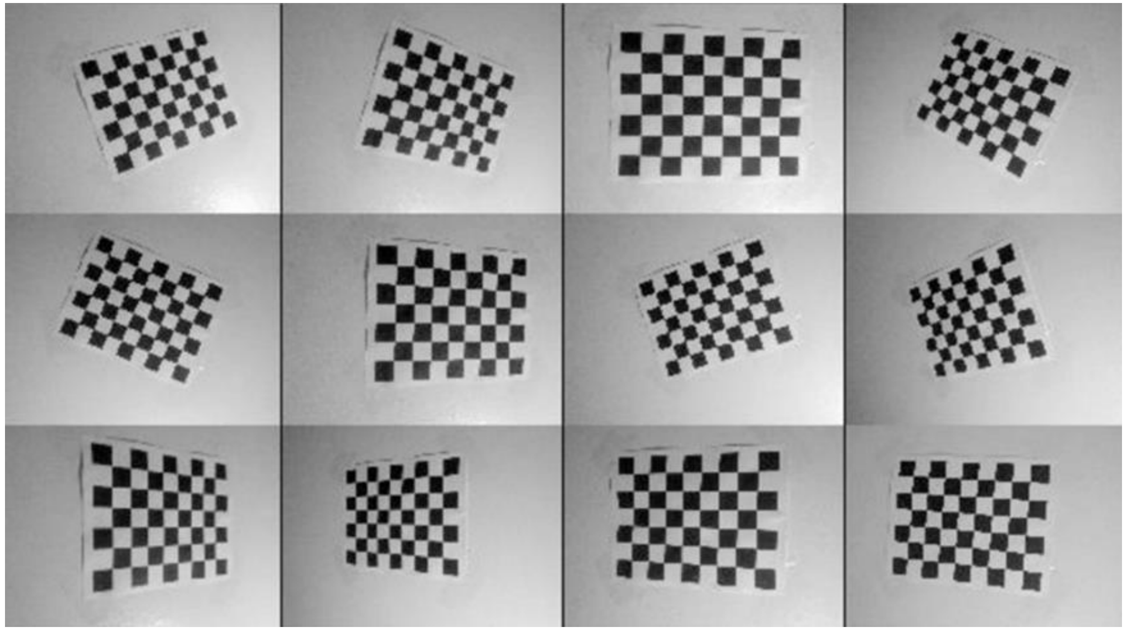
Camera calibration

Method of calibration using 2D checkerboard (Zhang's camera calibration method):

- Known size of cells and structure of checkerboard
- Since it is flat, all detected points lie in same plane
- Setting the world coordinate system to the checkerboard (X and Y in plane, Z perpendicular to plane), detected points will have Z=0

Ultimate sources of information:

- Zhengyou, Zhang. "A flexible new technique for camera calibration." Microsoft Research Technical Report (1998)
- Video by by Cyrill Stachniss with exceptional explanation on method and more https://youtu.be/-9He7Nu3u8s
- OpenCV documentation: https://docs.opencv.org/4.x/d9/d0c/group__calib3d.html



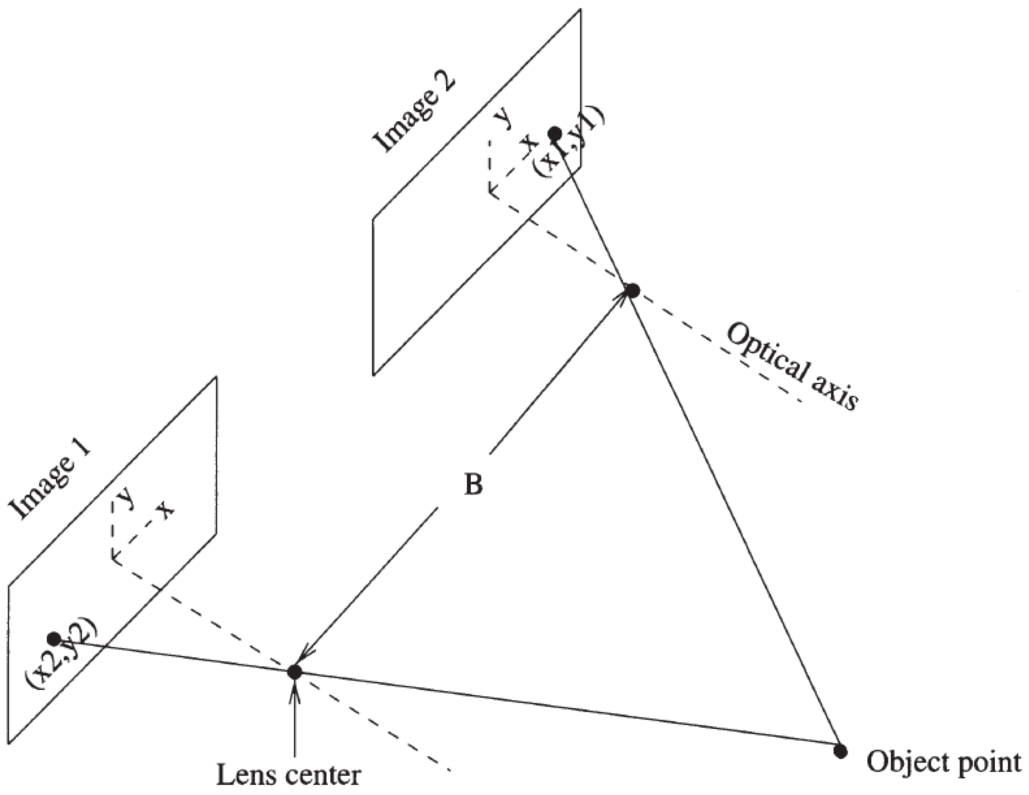
Stereo imaging system

Core principle:

- Single camera can not recover depth, so add another one
- Images of same object point can be used to recover two lines that intersect at that object point in world coordinate frame
- Hence, depth can be recovered

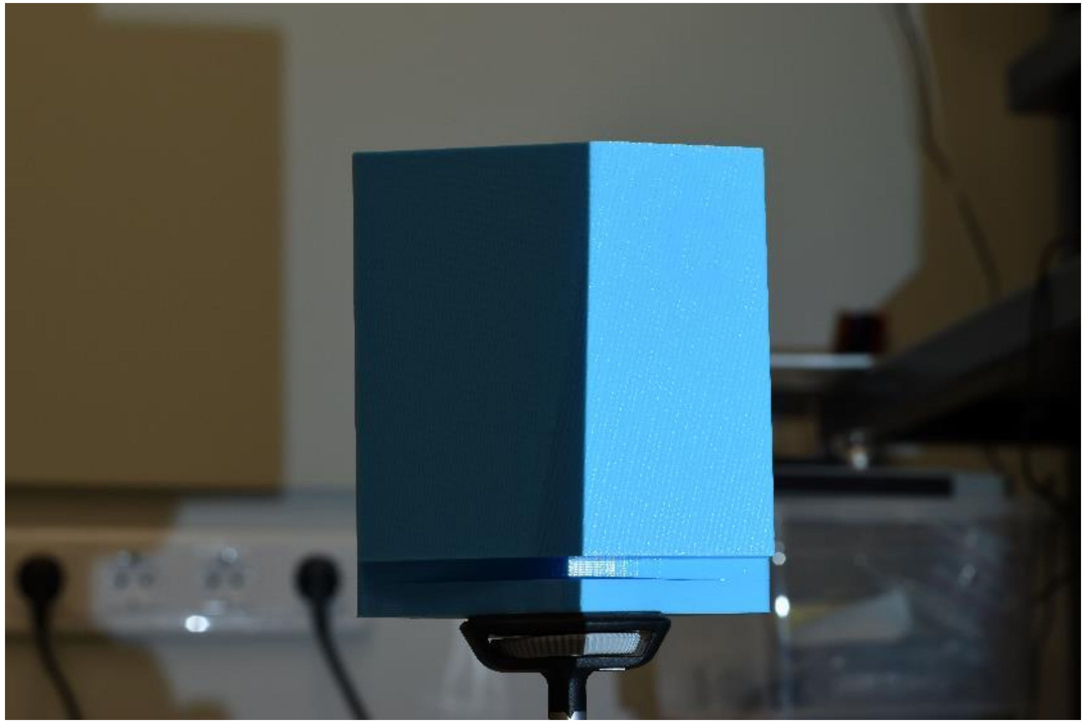
Disadvantages of stereo camera setup

- Need to find same point on both images
- Problems with objects without features to detect
- Very slow pixel by pixel processing



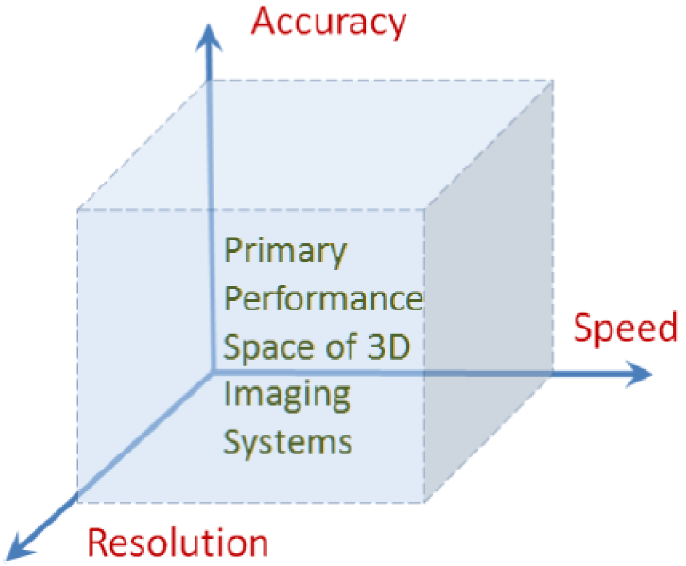
Camera + projector — a more efficient setup for 3D reconstruction

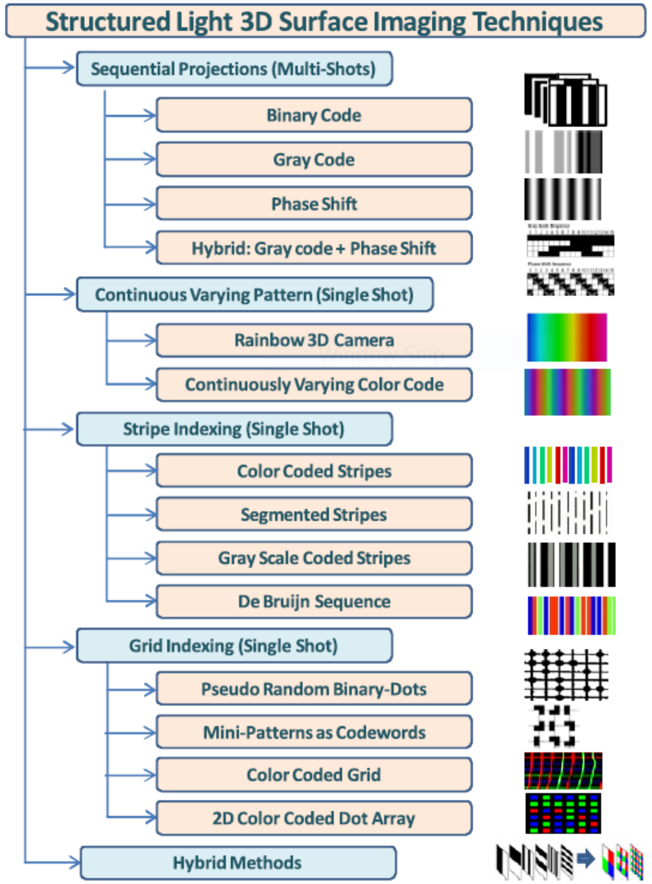
- Second camera can be swapped with projector
- It is a more efficient setup, since projector can project some patterns that we can eventually detect
- Some limitations are applied (discussed later), but overall it is far more efficient



Camera + projector — can we make it even more efficient?

- Yes! Instead of single “line” pattern project multiple “lines” or more complex patterns
- But a solution to resolve ambiguities required
- Check great article of Jason Gang for overview of various options

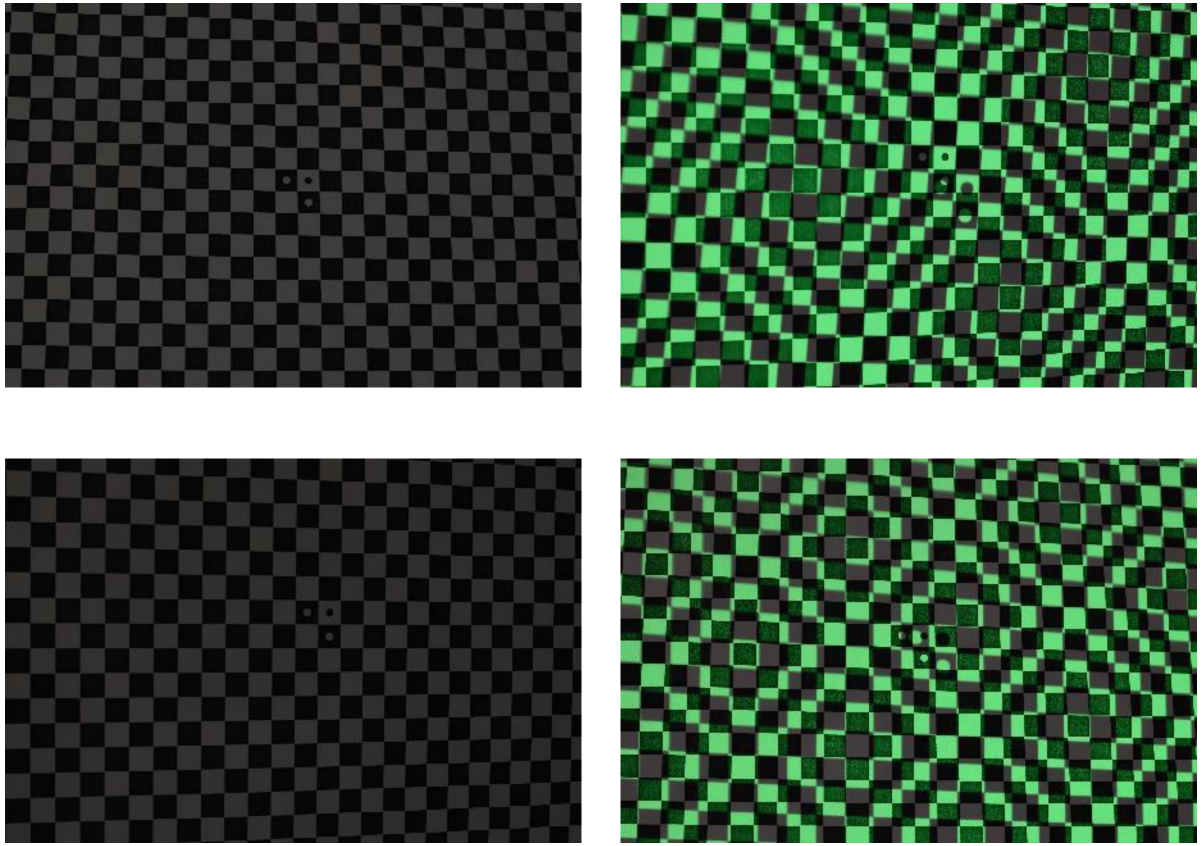




- Geng, Jason. "Structured-light 3D surface imaging: a tutorial." Advances in optics and photonics 3.2 (2011): 128-160.

How to calibration such setup?

- Since projector is inversed camera and can’t capture any data by itself, a camera is necessary
- This camera should be calibrated to get coordinates of pattern features in world coordinate system
- Position and orientation (extrinsic parameters) are part of calibration, therefore, it is important to preserve them
- As a result, calibration of camera and projector should be done simultaneously, unless we don’t have another option to get coordinates of pattern features in world coordinate system
- As on of possible approaches, two different patterns (one embedded into the checkerboard and another projected on checkerboard) can be used, but it is important to separate them on right images



Distortion in pinhole camera model

- Unlike pinhole camera objectives with real lenses have non-linear distortion
- To account distortion in model, additional matrixes that represent these non-linear transformations are added
- Depending of distortion severity in real system different models can be used

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_x x'' + c_x \\ f_y y'' + c_y \end{bmatrix}$$

$$\begin{bmatrix} x'' \\ y'' \end{bmatrix} = \begin{bmatrix} x' \frac{1+k_1 r^2 + k_2 r^4 + k_3 r^6}{1+k_4 r^2 + k_5 r^4 + k_6 r^6} + 2p_1 x' y' + p_2(r^2 + 2x'^2) + s_1 r^2 + s_2 r^4 \\ y' \frac{1+k_1 r^2 + k_2 r^4 + k_3 r^6}{1+k_4 r^2 + k_5 r^4 + k_6 r^6} + p_1(r^2 + 2y'^2) + 2p_2 x' y' + s_3 r^2 + s_4 r^4 \end{bmatrix}$$

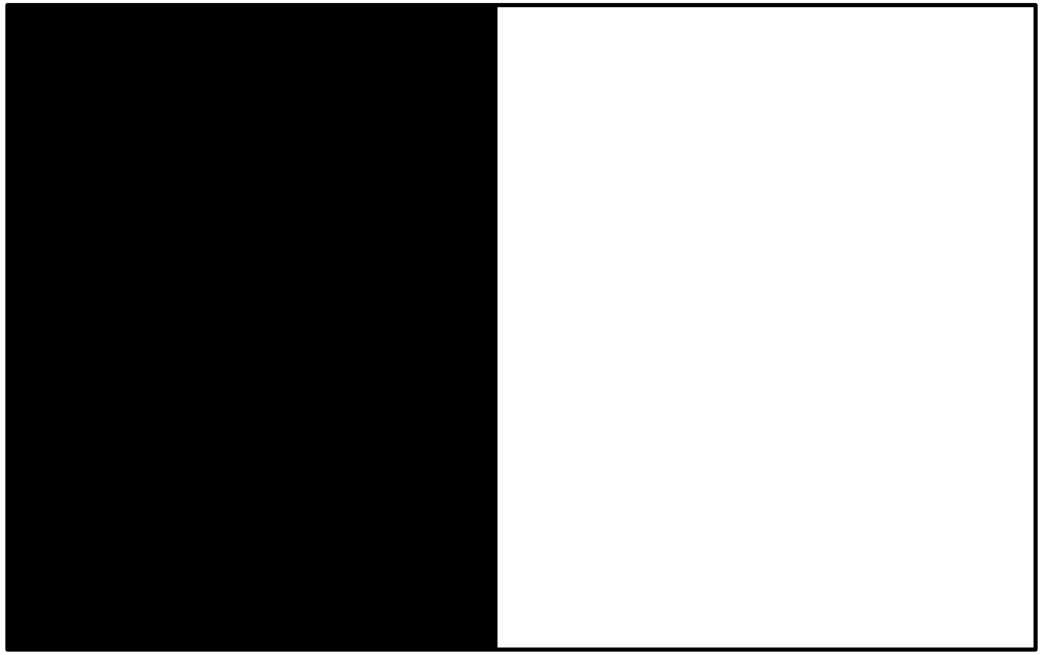
$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} f_x x''' + c_x \\ f_y y''' + c_y \end{bmatrix}$$

$$\begin{bmatrix} x''' \\ y''' \\ 1 \end{bmatrix} = \begin{bmatrix} R_{33}(\tau_x, \tau_y) & 0 & -R_{13}(\tau_x, \tau_y) \\ 0 & R_{33}(\tau_x, \tau_y) & -R_{23}(\tau_x, \tau_y) \\ 0 & 0 & 1 \end{bmatrix} R(\tau_x, \tau_y) \begin{bmatrix} x'' \\ y'' \\ 1 \end{bmatrix}$$

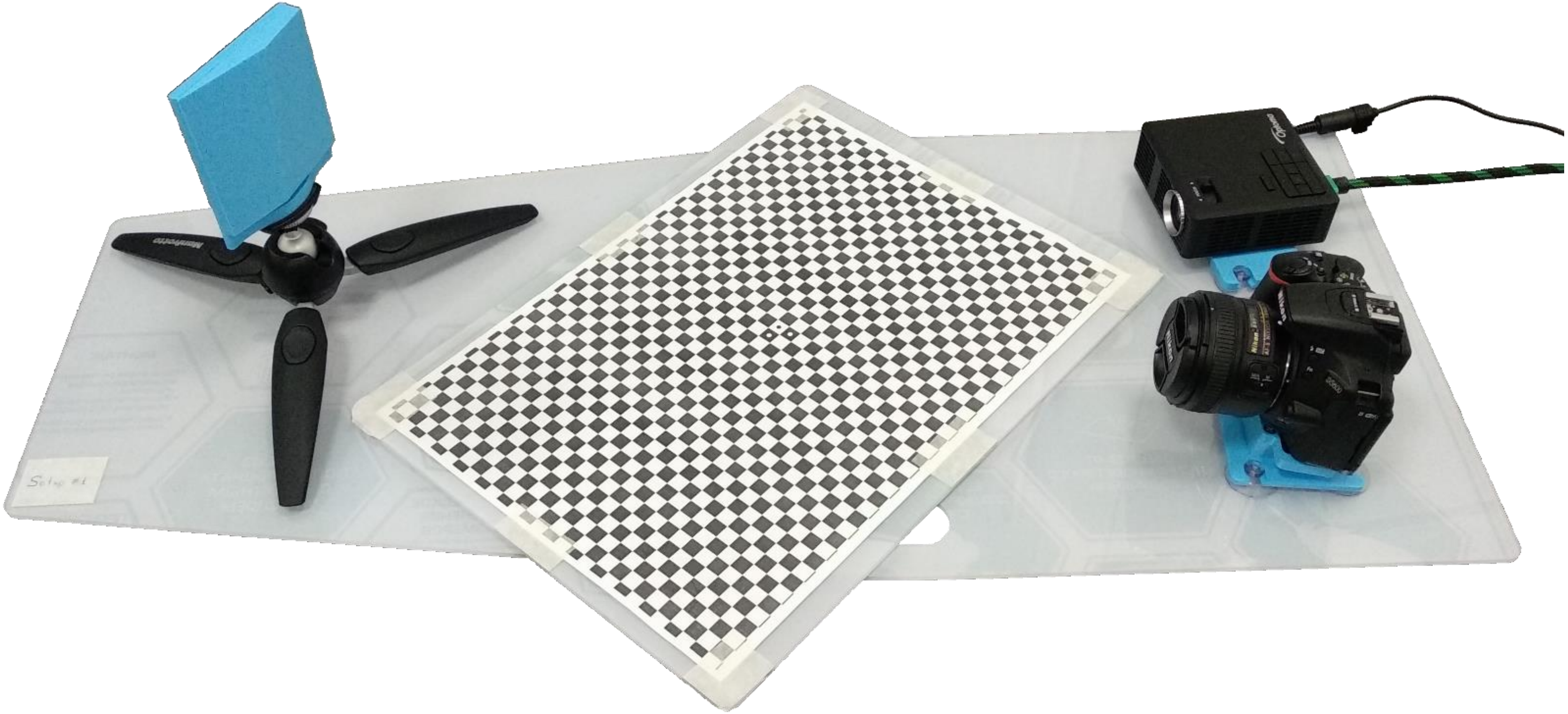
$$R(\tau_x, \tau_y) = \begin{bmatrix} \cos(\tau_y) & 0 & -\sin(\tau_y) \\ 0 & 1 & 0 \\ \sin(\tau_y) & 0 & \cos(\tau_y) \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\tau_x) & \sin(\tau_x) \\ 0 & -\sin(\tau_x) & \cos(\tau_x) \end{bmatrix}$$

Distortion in pinhole camera model

- When distortion is very small (objective of camera has good correction of aberrations including distortion) it can be neglected
- In case of noticeable distortion, dark-bright edge used for scan should be divided into smaller segments where direction recovery will be accurate enough



Example of hardware setup for 3D reconstruction



© Credits

- [Slide 2] www.scratchapixel.com
- [Slide 3] https://docs.opencv.org/4.x/d9/d0c/group__calib3d.html
- [Slide 5] https commons.wikimedia.org/wiki/File:RationalBezier2D.svg
- [Slide 6] Al Isawi , Malik MA, and Jurek Z. Sasiadek . "Pose estimation for mobile and flying robots via vision system." Aerospace Robotics III (2019): 83 96.
- [Slide 7] Chaudhuri, Subhasis , and Ambasamudram N. Rajagopalan . Depth from defocus: a real aperture imaging approach. Springer Science & Business Media,
- [Slide 9] Geng , Jason. "Structured light 3D surface imaging: a tutorial." Advances in optics and photonics 3.2 (2011): 128 160.