3D reconstruction using Stereo vision

Yeraly Kalel

Eötvös Loránd University

yeraly.kalel@nu.edu.kz

December 17, 2020

Applications of 3D reconstruction

- Computer graphics
- Medical imaging
- Computer animation
- Computational science
- Virtual reality

Point representations

- 2D point $(x, y)^T$ in homogeneous coordinates
  - 3D vector $(X, Y, W)^T$, where $x = X/W$ and $y = Y/W$
- 3D point $(x, y, z)^T$ in homogeneous coordinates
  - 4D vector $(X, Y, Z, W)^T$, where $x = X/W$, $y = Y/W$ and $z = Z/W$

Relationship of 3D world point with 2D image point

In pinhole camera model, a mapping from world coordinates into pixel coordinates:

$$\mathbf{p} = \mathbf{PX} \tag{1}$$

where $\mathbf{p}$ is 2D point in image ($\begin{bmatrix} u & v & 1 \end{bmatrix}^T$); $\mathbf{X}$ is 3D world point ($\begin{bmatrix} X & Y & Z & 1 \end{bmatrix}^T$); $\mathbf{P} \in \mathbb{R}^{3x4}$ is projection matrix.

$$\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}] \tag{2}$$

where $\mathbf{K} \in \mathbb{R}^{3x3}$ is camera matrix; $\mathbf{R} \in \mathbb{R}^{3x3}$ and $\mathbf{t} \in \mathbb{R}^{3x1}$ are rotation matrix and translation vector between world and camera coordinates, respectively.

Estimating 3D point from 2D point using stereo vision

Eq. 1 can be rewritten as follows:

$$\mathbf{p} \times \mathbf{PX} = \mathbf{0} \tag{3}$$

By writing all three resultant equations:

$$\begin{aligned} u\mathbf{p_3}^T\mathbf{X} - \mathbf{p_1}^T\mathbf{X} &= 0 \\ v\mathbf{p_3}^T\mathbf{X} - \mathbf{p_2}^T\mathbf{X} &= 0 \\ u\mathbf{p_2}^T\mathbf{X} - v\mathbf{p_1}^T\mathbf{X} &= 0, \end{aligned} \tag{4}$$

where $\mathbf{p}_i$ is i-th row of projection matrix $\mathbf{P}$. For stereo vision:

$$\begin{bmatrix} u^{(1)}(\mathbf{p_3}^T)^{(1)} - (\mathbf{p_1}^T)^{(1)} \\ v^{(1)}(\mathbf{p_3}^T)^{(1)} - (\mathbf{p_2}^T)^{(1)} \\ u^{(2)}(\mathbf{p_3}^T)^{(2)} - (\mathbf{p_1}^T)^{(2)} \\ v^{(2)}(\mathbf{p_3}^T)^{(2)} - (\mathbf{p_2}^T)^{(2)} \end{bmatrix} \mathbf{X} = \mathbf{0} \tag{5}$$

Epipolar geometry

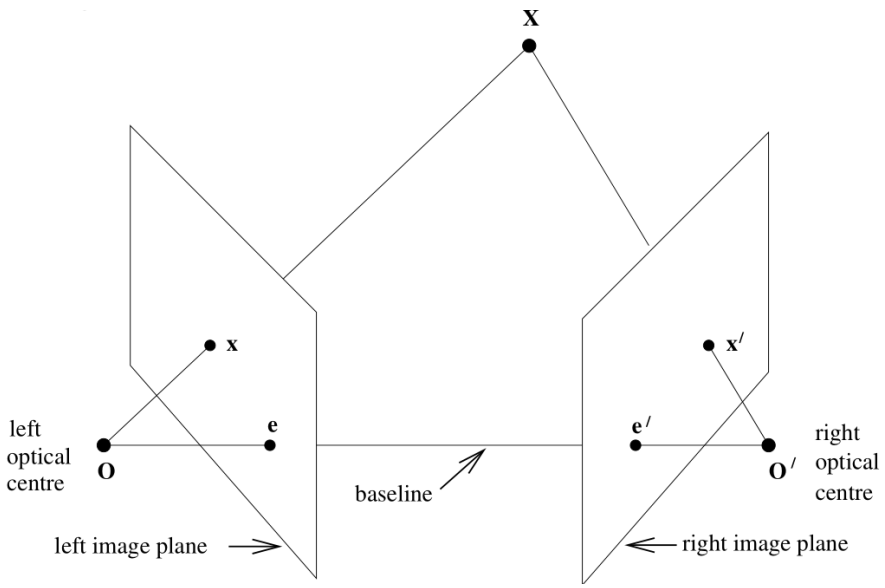


Figure 1: Representation of epipolar geometry [1]

Epipolar geometry - Fundamental matrix

Fundamental matrix encapsulates intrinsic projective geometry in stereo vision. Besides, each point correspondence must satisfy the following relation:

$$\begin{bmatrix} u^{(2)} & v^{(2)} & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u^{(1)} \\ v^{(1)} \\ 1 \end{bmatrix} = 0 \tag{6}$$

where **f** is a singular 3x3 fundamental matrix. For arbitrary n correspondences (6) can be rearranged to the following form:

$$\begin{bmatrix} u_1^{(2)}u_1^{(1)} & u_1^{(2)}v_1^{(1)} & u_1^{(2)} & v_1^{(2)}u_1^{(1)} & v_1^{(2)}v_1^{(1)} & v_1^{(2)} & u_1^{(1)} & v_1^{(1)} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_n^{(2)}u_n^{(1)} & u_n^{(2)}v_n^{(1)} & u_n^{(2)} & v_n^{(2)}u_n^{(1)} & v_n^{(2)}v_n^{(1)} & v_n^{(2)} & u_n^{(1)} & v_n^{(1)} & 1 \end{bmatrix} \tag{7}$$

$$\begin{bmatrix} f_{11} & f_{12} & f_{13} & f_{21} & f_{22} & f_{23} & f_{31} & f_{32} & f_{33} \end{bmatrix}^T = \mathbf{0}$$

Epipolar geometry - Epipolar lines

The point that has corresponding point on the second image, must locate on the line specified by the Fundamental matrix and that corresponding point. The lines are called epipolar lines and they can be formulated as:

$$\begin{aligned} \mathbf{l}^{(2)} &= \begin{bmatrix} a' \\ y' \\ c' \end{bmatrix} = \mathbf{F}\mathbf{p}^{(1)} \\ \mathbf{l}^{(1)} &= \begin{bmatrix} a \\ y \\ c \end{bmatrix} = \mathbf{F}^T\mathbf{p}^{(2)} \end{aligned} \tag{8}$$

Epipolar geometry - Essential matrix

Essential matrix is the special case of the fundamental matrix when image coordinates normalized by camera:

$$(\hat{\mathbf{p}}^{(2)})^T \mathbf{E} \hat{\mathbf{p}}^{(1)} = 0, \tag{9}$$

where $\hat{\mathbf{p}}^{(i)}$ is $(\mathbf{K}^{-1})^{(i)}\hat{\mathbf{p}}^{(i)}$ for $i = 1, 2$ and $\mathbf{E}$ is an 3x3 essential matrix.

Essential matrix decomposition

The essential matrix defines rotation and translation variables between two cameras:

$$\mathbf{E} = [\mathbf{t}]_{\times}\mathbf{R} \tag{10}$$

Suppose that

- $\mathbf{P}^{(1)} = \mathbf{K}^{(1)}[\mathbf{I}|\mathbf{0}]$
- SVD of $\mathbf{E}$ is $\mathbf{U}diag(1,1,0)\mathbf{V}^T$

In this case, four possible solutions are

1. $\mathbf{P}^{(2)} = \mathbf{K}^{(2)}[\mathbf{U}\mathbf{W}\mathbf{V}^{\mathbf{T}}|+\mathbf{u_3}]$
2. $\mathbf{P}^{(2)} = \mathbf{K}^{(2)}[\mathbf{U}\mathbf{W}\mathbf{V}^{\mathbf{T}}|-\mathbf{u_3}]$
3. $\mathbf{P}^{(2)} = \mathbf{K}^{(2)}[\mathbf{U}\mathbf{W}^{\mathbf{T}}\mathbf{V}^{\mathbf{T}}|+\mathbf{u_3}]$
4. $\mathbf{P}^{(2)} = \mathbf{K}^{(2)}[\mathbf{U}\mathbf{W}^{\mathbf{T}}\mathbf{V}^{\mathbf{T}}|-\mathbf{u_3}]$

$+\mathbf{u_3}$ is third column of matrix $\mathbf{U}$ and $\mathbf{W}$ is orthogonal matrix:

$$\mathbf{W} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{11}$$

Essential matrix decomposition (cont.)

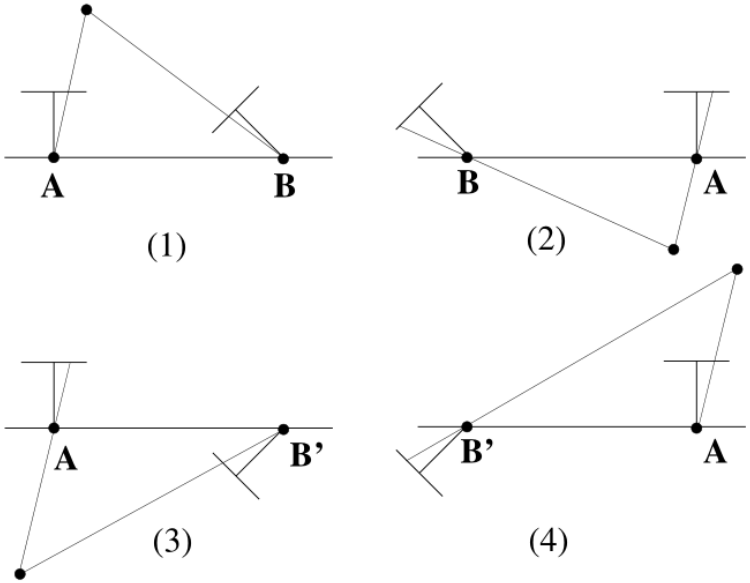


Figure 2: The four possible solutions of the projections [1]

Clustering algorithm improvement

- The maximum number of iterations can be predicted under some confidence level by:

$$k = \frac{\log(1-p)}{\log(1-w^n)}, \tag{12}$$

  where k is the predicted maximum number of iterations, p is the confidence level, w is the inlier ratio, and n is the minimum number of elements needed to construct the model.

- Local optimization can be applied to fasten RANSAC [2]. Specifically, If the better model is found, least-square method is applied with new best inliers to estimate hypothesized model. Better model is chosen among those two.

Testing if projection matrix is unknown

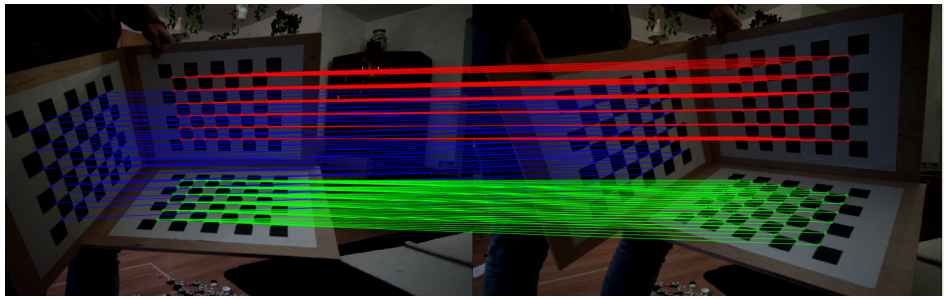
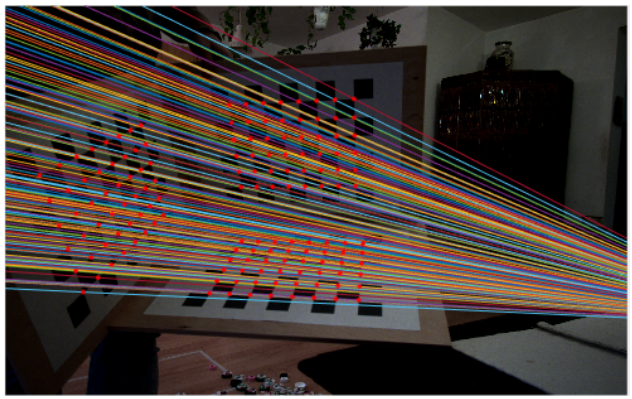
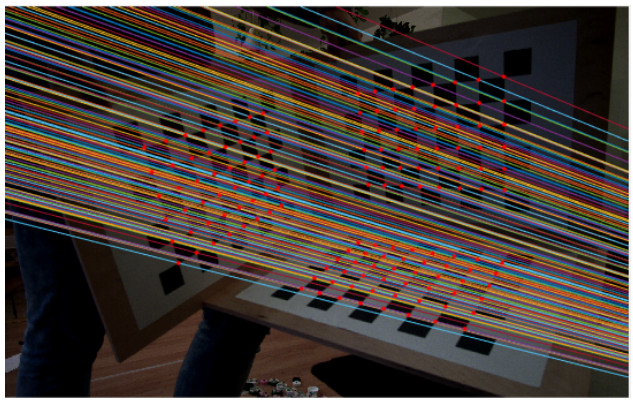


Figure 3: Synthetic data and correspondences in them



(a) First image



(b) Second image

Figure 4: Epipolar lines for the first and second image for chessboard case

Testing if projection matrix is unknown

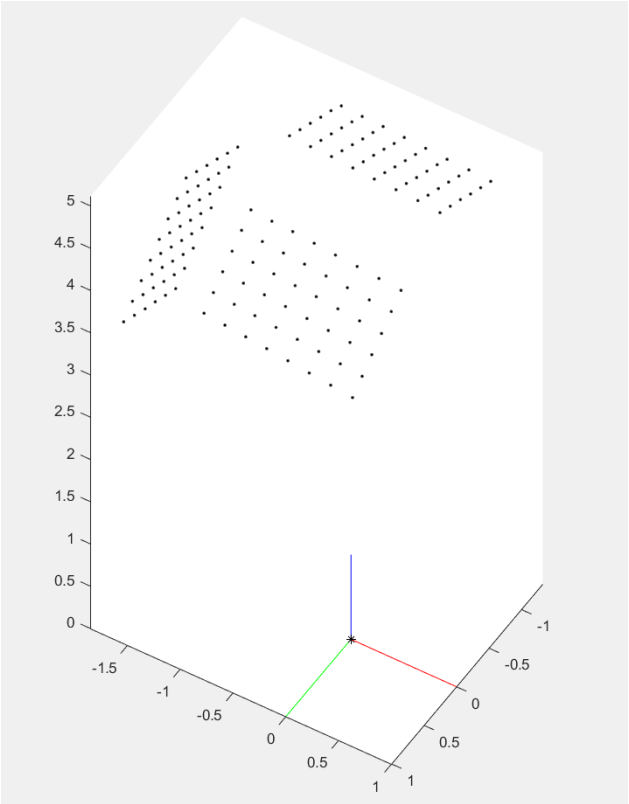
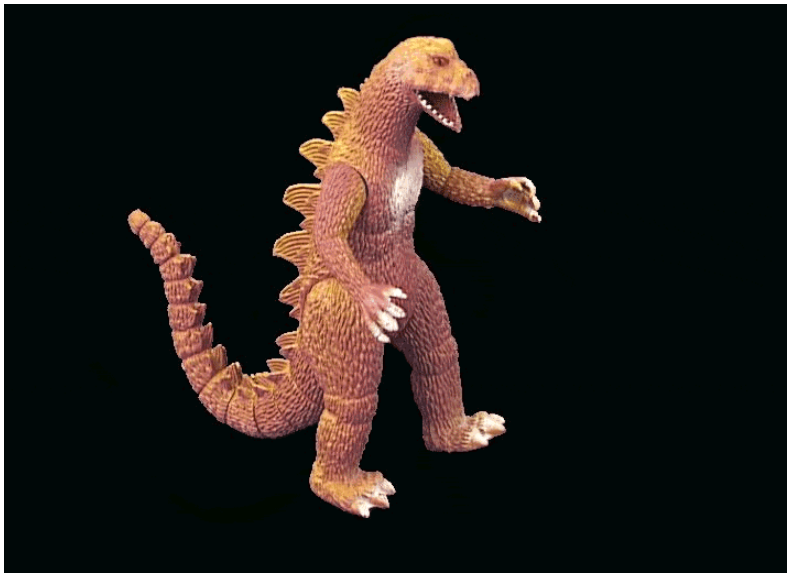
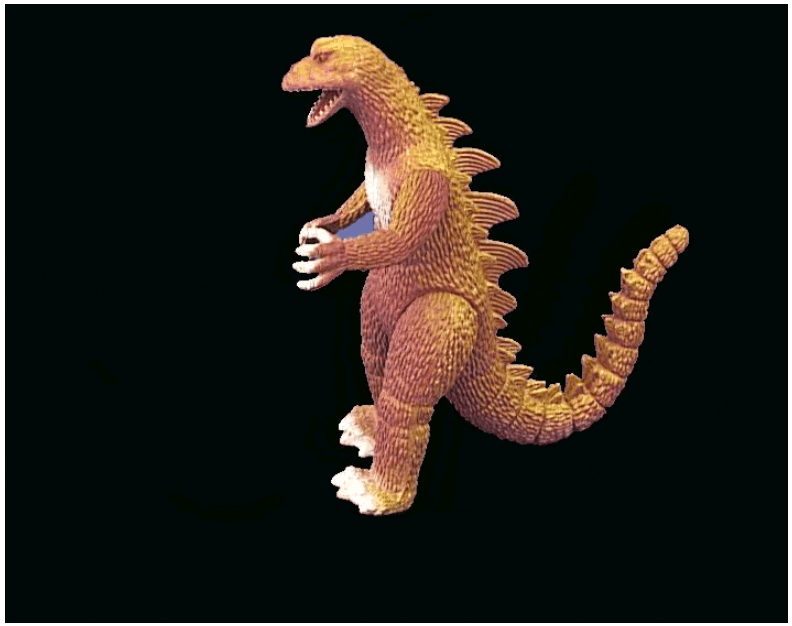


Figure 5: 3D point cloud of the 3 planes of the chessboard. Red, greed and blue lines represent x, y and z axes, respectively

Testing if projection matrix is known



(a) One view



(b) Second view

Figure 6: Representation of dinosaur [3]

Testing if projection matrix is known

(a) One view

(b) Second view

Figure 7: 3D point clouds of dinosaur

References

Richard Hartley and Andrew Zisserman (2003)

Multiple View Geometry in Computer Vision (Second Edition)

Cambridge University Press.

Ondrej Chum, Jiri Matas and Josef Kittler (2003)

Locally optimized RANSAC

Multi-view and Oxford Colleges building reconstruction (2009)

`https://www.robots.ox.ac.uk/~vgg/data/mview/`. Accessed: 2020-12-07

Thank you for attention!

Homogeneous system of linear equations

- Homogeneous system of linear equations has the following form:

$$\mathbf{A}\mathbf{x} = \mathbf{0} \tag{13}$$

  where $\mathbf{A} \in \mathbb{R}^{k \times n}$ is matrix of known variables, $\mathbf{x} \in \mathbb{R}^{n\mathrm{x}1}$ is a vector of independent unknown variables.
- This can be solved by using lagrange-multipliers if $k \geqslant n - 1$.
- The solution of $\mathbf{x}$ is the (one dimensional) kernel of $\mathbf{A}$ and it is an eigenvector with at least eigenvalue of $\mathbf{A}^T\mathbf{A}$ subjected to $\left\|\mathbf{x}\right\| = 1$.

Clustering algorithm

Outiler detector method random sample consensus (RANSAC) is used. Each iteration:

1. Randomly selects a subset from the dataset, with size n, which is a minimum number of elements to describe the model, i.e., the DoF of the model.
2. Define the model using those hypothetical inliers.
3. Test dataset using the defined model according to some loss functions. If an element has a loss value under the specified threshold, then it is considered as an inlier; otherwise, it is an outlier.
4. The model is identified as a better one among previously defined models if it fits more data than them.

Point normalization

Point normalization have a positive effect on increase of condition number of the the coefficient matrix.
Procedure:

1. Translate the points such that centroid is at the origin
2. Scale points so that the average distance from origin is $\sqrt{dimension}$