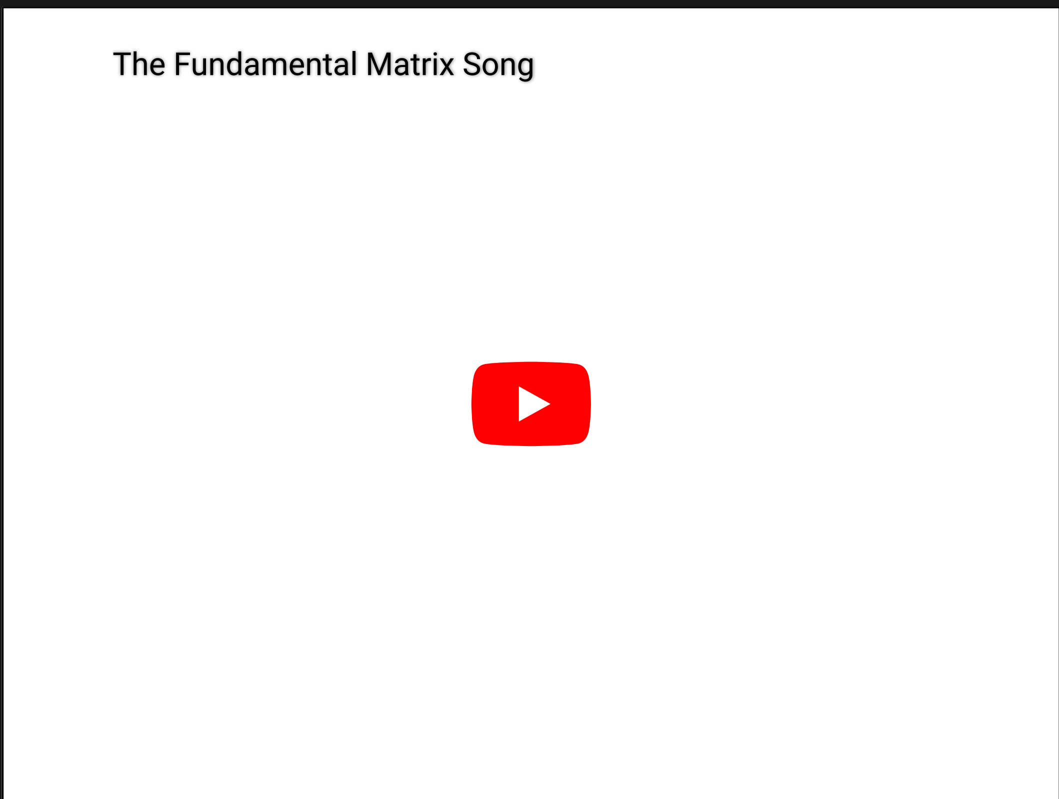
Epipolar geometry (2D-2D)

Dates Taught	October 20, 2020 → October 27, 2020
Links of Videos	L18 (Oct 20), L20-Theory, L20-Discussion (Oct 27)
Comments	Empty
Lecture No.	L18 L20
Module	SLAM: Vision
Related to All Qu…	Empty

> Page Author: Shubodh Sai

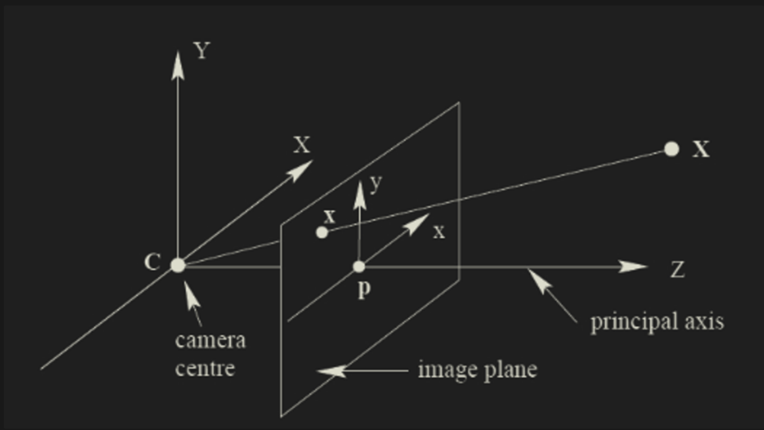


Jokes aside, this song does explain a lot of things!

Source: Cyrill Stachniss videos & Multi-view Geometry (book by Zisserman & Hartley) unless explicitly mentioned.

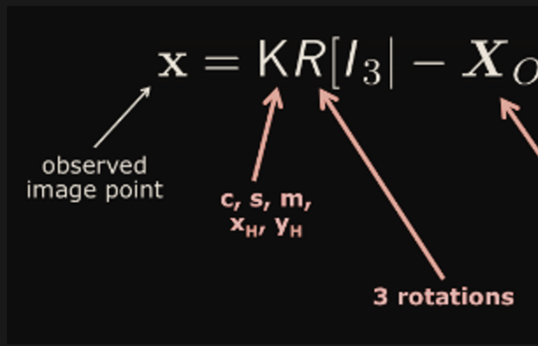
0. Revisiting single view geometry

0.1 Difference between the ray and the image coordinates



Pin hole camera

Few clarifications:



$$\begin{bmatrix} c & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Coordinates of image point in camera coordinate system? Let's call it **normalized image coordinates**.

  Ans: $[x, y, f]^T$

- What is the relation between **normalized image coordinates** and **3D object point** $\mathbf{X}$?

  Related by a scalar

- How to arrive at the above vector from homogenous **image coordinates** $[x, y, 1]$? Forget about the scaling factor.
  Clue ⇒

  $\lambda x = KX$

  $$\text{Ans: } \mathbf{K}^{-1}\text{. Consider } \mathbf{K}^{-1} \times [x, y, 1]^T = ?$$

  So $\beta\ \mathbf{K}^{-1}\mathbf{x}$ is a ray connecting camera center and 3D point.

  $x = KTX$

  $K^{-1}x = T_W^C X_W$ = X_C

1. Epipolar Geometry

1.1 Introduction

1 view, 2 view and n view geometry.

- Intrinsic projective geometry between two views.
- Independent of scene structure.

▼ Dependent only on camera parameters and relative pose.

(Think about F and its relation to x's in both images, scene structure isn't involved there)

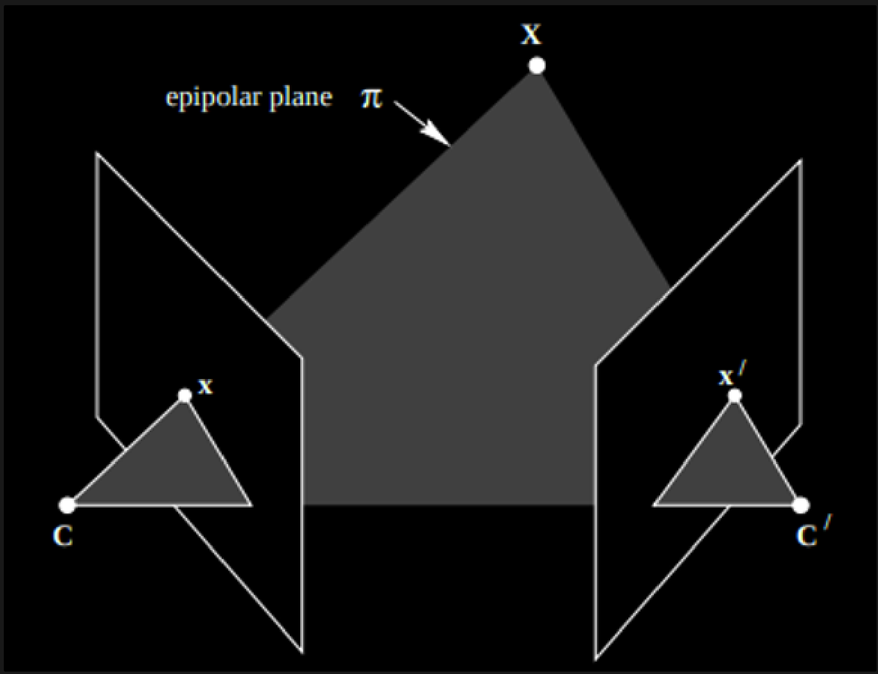
1.2 Motivation

All parameters: World points, image points, camera matrix, relative pose.

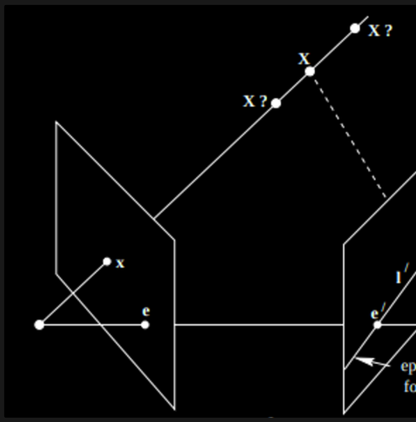
Think of a camera navigating. Typical robotics application.

1. Restricts search for the corresponding points given F. (What is F?)
    - Knowns: Given K, R, t. Unknown: Corresponding point
2. F and its decomposition given enough corresponding points.
    - Knowns: Corresponding Points. Unknown: K, R, t

1.3 Intuition



▼ 1st way: figure



The question: how is **x'** constrained? Two ways of looking at it:

1. Knowns: Given K, R, t. Unknown: Corresponding point

▼ 1st way: Backproject **x**

Joining the 3D ray points to 2nd camera center, you end up with a "epipolar" line.

- 2nd way: 2 lines/rays make a plane
    - Firstly, what epipolar plane is using 1st way (Joining several points on Cx to C' gives a plane)

    Intersection of the plane formed by baseline and first ray with second image plane gives an "epipolar" line.

1.4 Some definitions

- Epipole: ***What is the image of C' in first view?***

    Also, point of intersection of the "baseline" (**CC'**) with the image planes.
- Epipolar plane: a plane containing baseline
- Epipolar line: ***Try using above "epi plane" definition.***
    - Intersection of epipolar plane with the image plane.
    - All epipolar lines intersect at?
- Epipolar lines: Figure

2. Fundamental matrix F

2.1 Derivation of F

- **Notation clarification: R_{21} is same as R_1^2 which is "1st to 2nd frame".**

$$O_1 = \begin{bmatrix} I_{3\times3} & | & 0_{3\times1} \end{bmatrix} \qquad O_2 = \begin{bmatrix} R_{12} & | & t_{12} \end{bmatrix}_{3\times4}$$

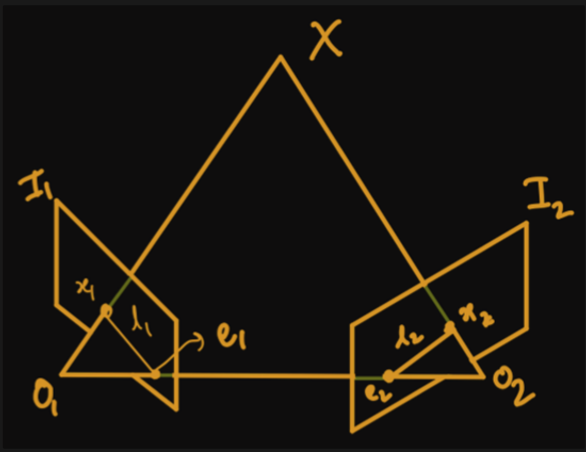
$$\lambda_1 \overrightarrow{p_1} = K[I \quad 0]\overrightarrow{X}_{4\times1} \qquad \lambda_2 \overrightarrow{p_2} = K[R_{21} \quad t_{21}]\overrightarrow{X}_{4\times1}$$

$$\lambda_1 K^{-1}\overrightarrow{p_1} = \overrightarrow{X} \qquad \lambda_2 K^{-1}\overrightarrow{p_2} = [R_{21} \quad t_{21}]\overrightarrow{X}$$

$$where \quad K^{-1}\overrightarrow{p_1} = \overrightarrow{x_1} \qquad where \quad K^{-1}\overrightarrow{p_2} = \overrightarrow{x_2}$$

💡 What really is $K^{-1}\overrightarrow{p}$?

O_1O_2P forms a triangle, the epipolar plane. So?



$$\begin{gathered}
\implies \overrightarrow{O_1P} \cdot (\overrightarrow{O_1O_2} \times \overrightarrow{O_2P}) = 0 \\
\implies \lambda_1 \overrightarrow{x_1} \cdot \left(\overrightarrow{t_{12}} \times R_{12} \lambda_2 \overrightarrow{x_2} \right) = 0 \\
\implies \lambda_1 \lambda_2 \overrightarrow{x_1} \left(\overrightarrow{\mathrm{t}_{12}} \times R_{12} \vec{x}_2 \right) = 0 \\
\implies \overrightarrow{x_1^T} \, [\mathrm{t}_{12}]_\times \, R_{12} \, \vec{x}_2 = 0 \\
\implies \overrightarrow{x_1}^T E_{12} \overrightarrow{x_2} = 0 \\
\implies \overrightarrow{P_1}^T F_{12} \overrightarrow{P_2} = 0 \\
where \; E_{12} = [\mathrm{t}_{12}]_\times \, R_{12} \\
\& \quad F_{12} = K^{-T} [\mathrm{t}_{12}]_\times \, R_{12} K^{-1}
\end{gathered}$$

> F_{12} is the **fundamental matrix** that related 2nd image to the 1st in terms of **pixel coordinates**. E_{12} is the **essential matrix** that related in terms of **normalized pixel coordinates.**

2.2 Computation of F

We have seen:

$$\mathrm{x}'^{\mathrm{T}} \mathrm{F} \mathrm{x} = 0$$

- Now, the natural question is: Can **I** find **F** given **enough** correspondences? **Enough?**

  $$\mathbf{x} = (x, y, 1)^{\mathrm{T}} \text{ and } \mathbf{x}' = (x', y', 1)^{\mathbf{T}}$$

  Expanding:

  $$x'xf_{11} + x'yf_{12} + x'f_{13} + y'xf_{21} + y'yf_{22} + y'f_{23} + xf_{31} + yf_{32} + f_{33} = 0$$

  $$(x'x, x'y, x', y'x, y'y, y', x, y, 1) \, \mathbf{f} = 0$$

From "n" set of correspondences:

$$\mathrm{Af} = \begin{bmatrix} x_1'x_1 & x_1'y_1 & x_1' & y_1'x_1 & y_1'y_1 & y_1' & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ x_n'x_n & x_n'y_n & x_n' & y_n'x_n & y_n'y_n & y_n' & x_n & y_n & 1 \end{bmatrix} \mathbf{f} = \mathbf{0}$$

- **What rank must A have in our case for the solution to be (non-zero and) unique (upto a scaling factor)?**

  8. (Upto a scaling factor: F is homogenous.) So we need 8 corresponding points (8 unknowns in F). (Other algorithms like 7/5/4 point algorithm do exist, we are keeping things simple here)

  However, we want to use more data and with noisy data, the rank will not be 8 but will be 9 (with n=9 or more points). Take the SVD approach (described below).

Just like in DLT: The solution vector $\mathbf{f}$ should minimize $||\mathbf{Af}||$ subject to the condition $||\mathbf{f}|| = 1$.

If $A = UDV^\top$, the Least Squares solution is the last column of V corresponding to the smallest singular value of A.

- See this for understanding why the LS solution is the last column of V corresponding to the smallest singular value of A.

  - *Find the* $\mathbf{x}$ *that minimizes* $\|\mathtt{A}\mathbf{x}\|$ *subject to* $\|\mathbf{x}\| = 1$.

  This problem is solvable as follows. Let $\mathtt{A} = \mathtt{UDV}^\mathsf{T}$. The problem then requires us to minimize $\|\mathtt{UDV}^\mathsf{T}\mathbf{x}\|$. However, $\|\mathtt{UDV}^\mathsf{T}\mathbf{x}\| = \|\mathtt{DV}^\mathsf{T}\mathbf{x}\|$ and $\|\mathbf{x}\| = \|\mathtt{V}^\mathsf{T}\mathbf{x}\|$. Thus, we need to minimize $\|\mathtt{DV}^\mathsf{T}\mathbf{x}\|$ subject to the condition $\|\mathtt{V}^\mathsf{T}\mathbf{x}\| = 1$. We write $\mathbf{y} = \mathtt{V}^\mathsf{T}\mathbf{x}$, and the problem is: minimize $\|\mathtt{D}\mathbf{y}\|$ subject to $\|\mathbf{y}\| = 1$. Now, $\mathtt{D}$ is a diagonal matrix with its diagonal entries in descending order. It follows that the solution to this problem is $\mathbf{y} = (0, 0, \ldots, 0, 1)^\mathsf{T}$ having one non-zero entry, 1 in the last position. Finally $\mathbf{x} = \mathtt{V}\mathbf{y}$ is simply the last column of $\mathtt{V}$. The method is summarized in algorithm A5.4.

  <u>Objective</u>

  Given a matrix $\mathtt{A}$ with at least as many rows as columns, find $\mathbf{x}$ that minimizes $\|\mathtt{A}\mathbf{x}\|$ subject to $\|\mathbf{x}\| = 1$.

  <u>Solution</u>

  $\mathbf{x}$ is the last column of $\mathtt{V}$, where $\mathtt{A} = \mathtt{UDV}^\mathsf{T}$ is the SVD of $\mathtt{A}$.

  Algorithm A5.4. *Least-squares solution of a homogeneous system of linear equations.*

- What are the no of degrees of freedom of F?

  Answer: It is 7.

  It is because F is both homogeneous and singular matrix (as the skew-symmetric matrix of t above is of rank 2).

2.2.1 Normalized 8 point algorithm

The above would've been fine F was a full rank matrix (as in DLT case because K and T are always full rank), but it is not in F's case, so we need to enforce it in our solution.

Objective

Given $n \geq 8$ image point correspondences $\{\mathbf{x_i} \leftrightarrow \mathbf{x_i}'\}$, determine the fundamental matrix F such that $\mathbf{x_i}'^\top \mathrm{F}\mathbf{x_i} = 0$

Algorithm

System: $\mathrm{Af} = 0$

Algorithm EIGHT_POINT

The input is formed by n point correspondences, with $n \geq 8$.

1. Construct system (7.18) from n correspondences. Let A be the $n \times 9$ matrix of the coefficients of the system and $A = UDV^\top$ the SVD of A.
2. The entries of F (up to an unknown, signed scale factor) are the components of the column of V corresponding to the least singular value of A.
3. To enforce the singularity constraint, compute the singular value decomposition of F:

$$F = UDV^\top.$$

4. Set the smallest singular value in the diagonal of D equal to 0; let D' be the corrected matrix.
5. The corrected estimate of F, F', is finally given by

$$F' = UD'V^\top.$$

The output is the estimate of the fundamental matrix, F'.

Source: Trucco, Alessandro Verri. n.d. Introductory Techniques for 3-D Computer Vision-Prentice Hall (1998). Chapter 7, Stereopsis.

> **Normalization has to be done to avoid numerical instabilities!**

For example,

- Apply $\mathrm{Tx} = \hat{\mathrm{x}}$ so that
  - Transform the points: the center of mass of all points is at (0,0)
  - Scale the image: the x and y coordinated are within [-1,1]
- Find fundamental matrix $\hat{\mathrm{F}}$ on $\hat{\mathrm{x}}$.
- Obtain original matrix as $\mathrm{F} = \mathrm{T}^{\mathrm{T}}\hat{\mathrm{F}}\mathrm{T}$.

$$\begin{aligned}
\mathbf{x}'^\top \mathbf{F} \mathbf{x}'' &= \left(\mathbf{T}^{-1}\hat{\mathbf{x}}'\right)^\top \mathbf{F} \left(\mathbf{T}^{-1}\hat{\mathbf{x}}''\right) \\
&= \hat{\mathbf{x}}'^\top \mathbf{T}^{-\top} \mathrm{F} \mathrm{T}^{-1} \hat{\mathbf{x}}'' \\
&= \hat{\mathbf{x}}'^\top \hat{\mathrm{F}} \hat{\mathbf{x}}''
\end{aligned}$$

2.2.2 Decomposition of E

> Called "Chirality"

Refer section 9.5, 9.6 of Zisserman for an elaborate discussion.

Suppose that the SVD of E is

$$\mathrm{U}\,\mathrm{diag}(1,1,0)\mathrm{V}^{\mathrm{T}}$$

$$Recollect$$
$$E_{12} = [\mathrm{t}_{12}]_x\, R_{12}$$

$$Z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad ZW = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\begin{aligned} \mathrm{E} &= U \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \mathrm{V}^\top \\ &= U \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathrm{V}^\top \\ &= UZWV^\top \\ &= UZU^\top UWV^\top \\ &= UZU^\top \qquad UWV^\top \\ &= S_B R^T \end{aligned}$$

Four possibilities to define Z, W:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ = ZW = Z^T W^T \\ = -Z^T W = -ZW^T$$

2 solutions for each S_B and R:

$$\mathrm{S}^1_{\hat{B}} = UZU^\top \quad \mathrm{S}^2_{\hat{B}} = UZ^\top U^\top \\ R_1^\top = UWV^\top \quad R_2^\top = UW^\top V^\top$$

So 4 solutions in total for E.

Summary:

First compute SVD of E, then compute 4 possible solutions (below figure), then test for which solutions all points are in front of cameras and then return only physically plausible unique configuration.

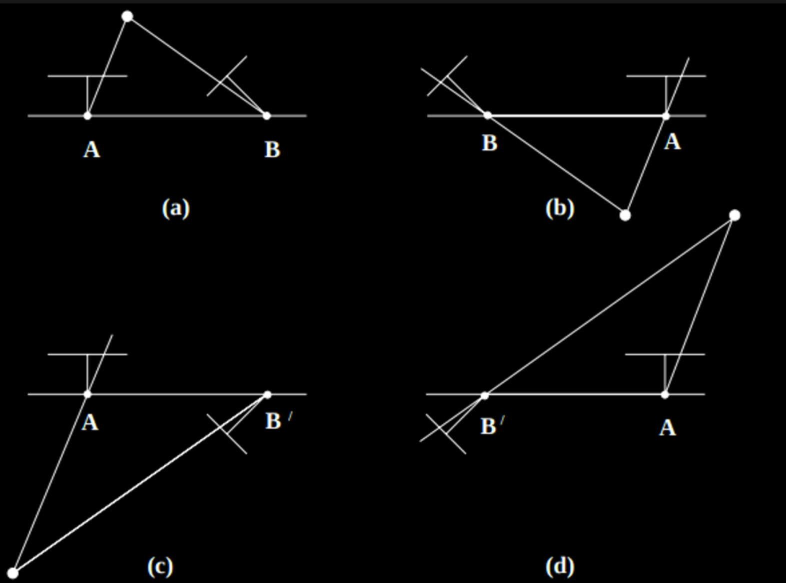


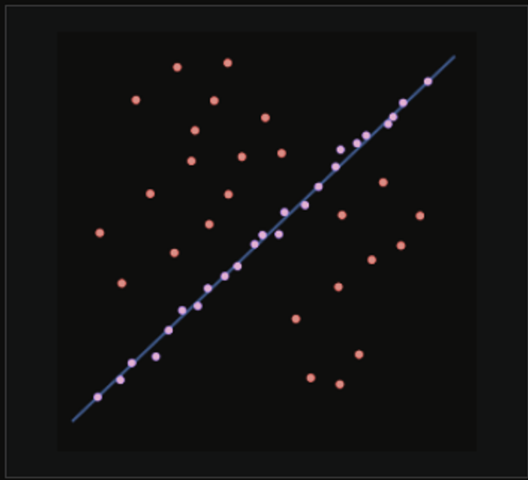
Fig. 9.12. **The four possible solutions for calibrated reconstruction from E.** *Between the left and right sides there is a baseline reversal. Between the top and bottom rows camera B rotates 180° about the baseline. Note, only in (a) is the reconstructed point in front of both cameras.*

Source

2.2.3 RANSAC: A quick overview

Since our data is always noisy, RANSAC eliminates outliers from our data.

A data set with many outliers for which a line has to be fitted.



Fitted line with RANSAC; outliers have no influence on the result.

Algorithm 1 RANSAC

1: Select randomly the minimum number of points required to determine the model parameters.
2: Solve for the parameters of the model.
3: Determine how many points from the set of all points fit with a predefined tolerance ϵ.
4: If the fraction of the number of inliers over the total number points in the set exceeds a predefined threshold τ, re-estimate the model parameters using all the identified inliers and terminate.
5: Otherwise, repeat steps 1 through 4 (maximum of N times).

Source: Konstantinos G. Derpanis

Just given a high level overview in the interest of time, watch the short Cyrill Stachniss video for more clarity.

Classroom discussion

> Questions to solve/discuss

1. Difference in $Ax = 0$ between DLT approach and normalized 8 point approach for F computation?
2. In which frame is $\mathbf{K}^{-1}\mathbf{x}$ in? Global? Camera? Image? Why so?
3. In the second step of our fundamental matrix derivation, shouldn't $R_{12}\lambda_2\overrightarrow{x_2}$ be $T_{12}\lambda_2\overrightarrow{x_2}$ i.e. why aren't we considering the translation term here?
4. Suggest two singularity cases when the above 8 point algorithm fails.
5. At the end of P3P, why is ICP necessary?