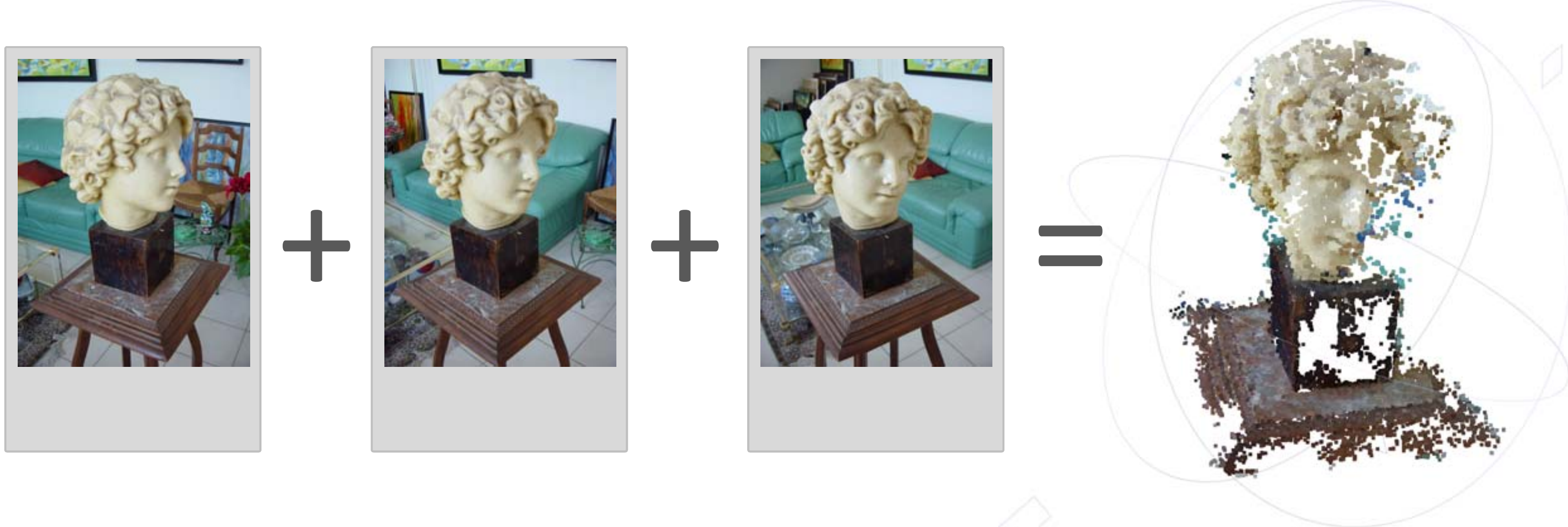


# Multi-view 3D Reconstruction for Dummies

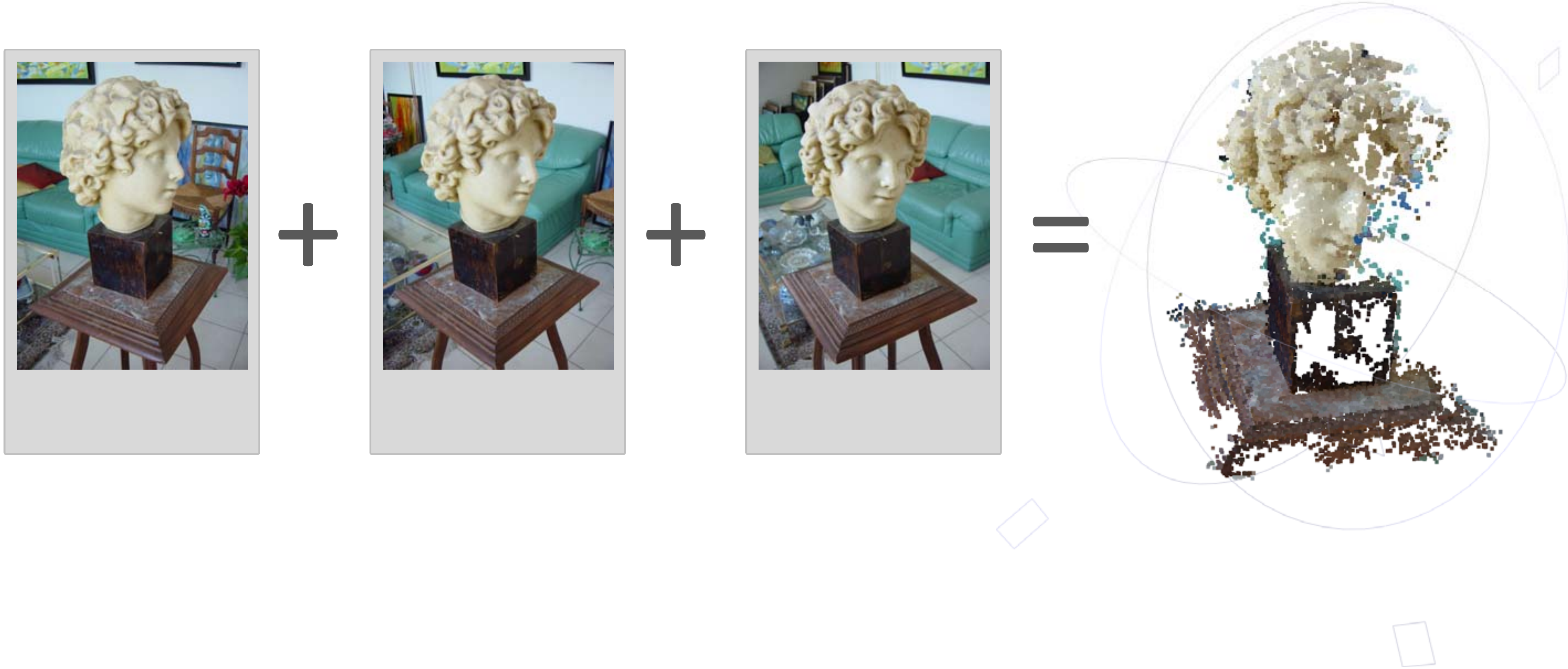

**Jianxiong Xiao**

PRINCETON VISION GROUP

# SFMedu Program with Code

Download from:

http://mit.edu/jxiao/Public/software/SFMedu/



# Camera projection

- When people take a picture of a point:

$$\mathbf{x} = \mathbf{K}\left[\mathbf{R}\middle|\mathbf{t}\right]\mathbf{X}$$

# Camera projection

- When people take two pictures with same camera setting:

$$\mathbf{x}_1 = \mathbf{K}\left[\mathbf{R}_1 | \mathbf{t}_1\right]\mathbf{X}$$

$$\mathbf{x}_2 = \mathbf{K}\left[\mathbf{R}_2 | \mathbf{t}_2\right]\mathbf{X}$$

# Camera projection

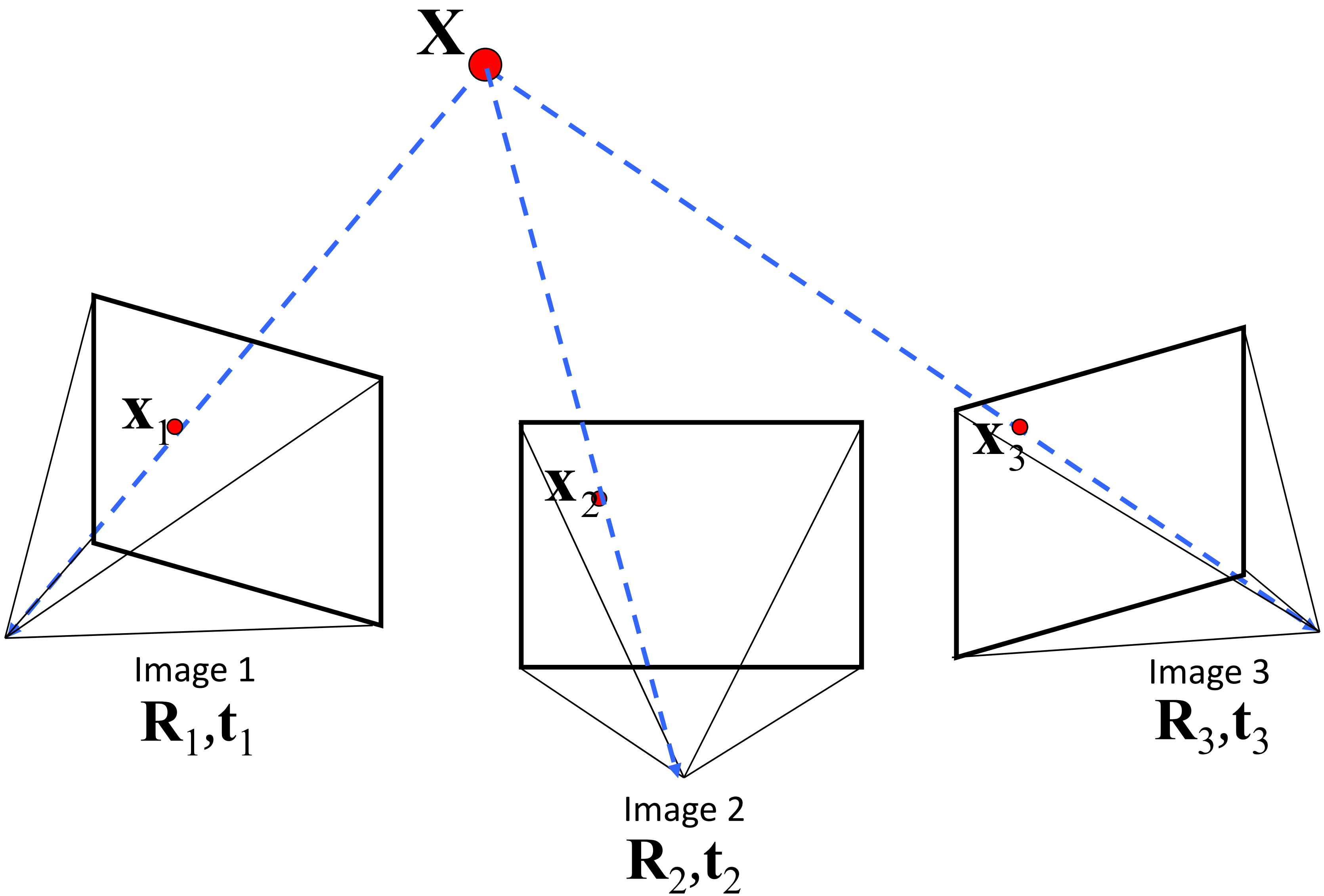
- When people take three pictures with same camera setting:

$$\mathbf{x}_1 = \mathbf{K}\left[\mathbf{R}_1 \middle| \mathbf{t}_1\right]\mathbf{X}$$

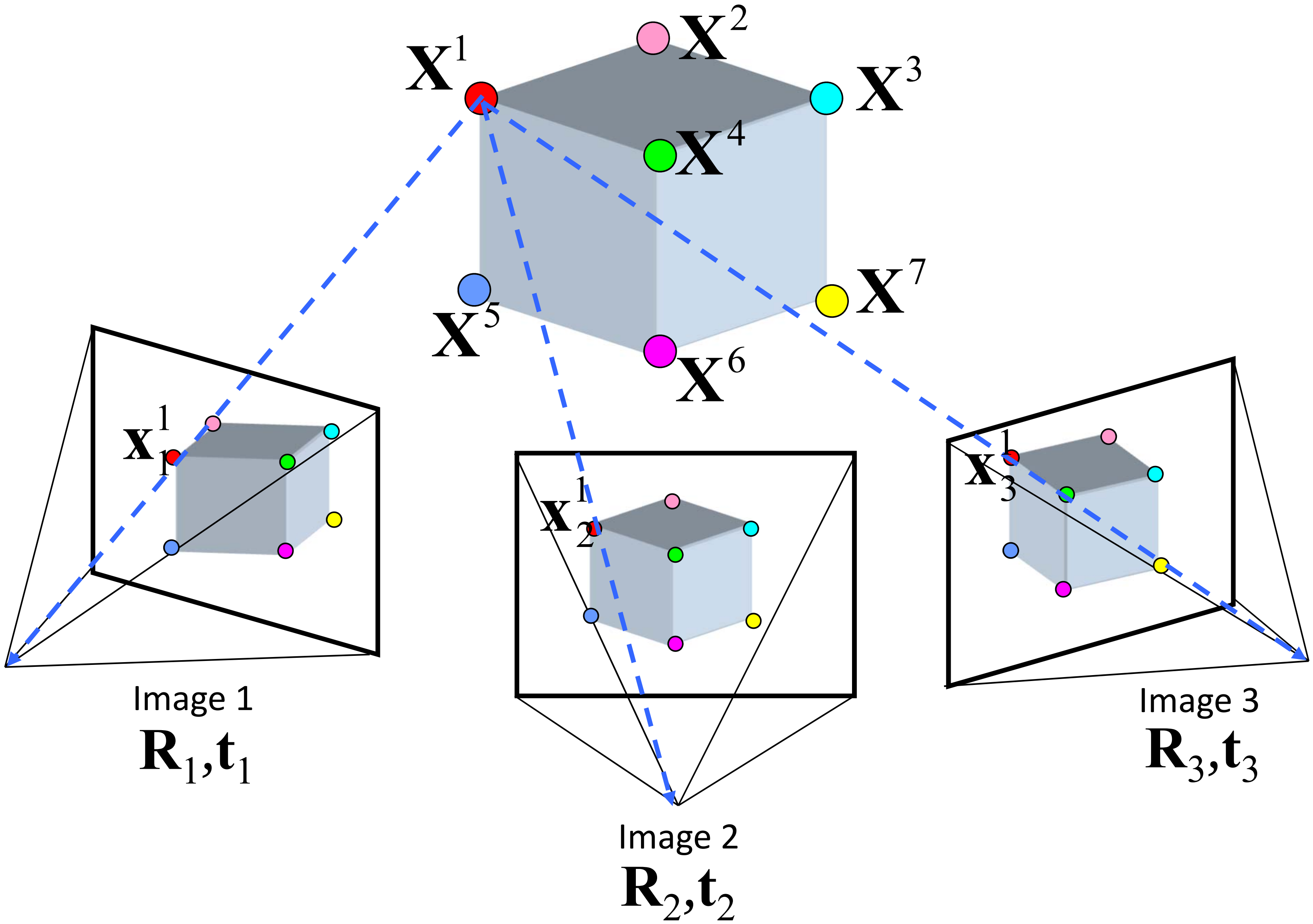
$$\mathbf{x}_2 = \mathbf{K}\left[\mathbf{R}_2 \middle| \mathbf{t}_2\right]\mathbf{X}$$

$$\mathbf{x}_3 = \mathbf{K}\left[\mathbf{R}_3 \middle| \mathbf{t}_3\right]\mathbf{X}$$

# Camera projection



# Camera projection

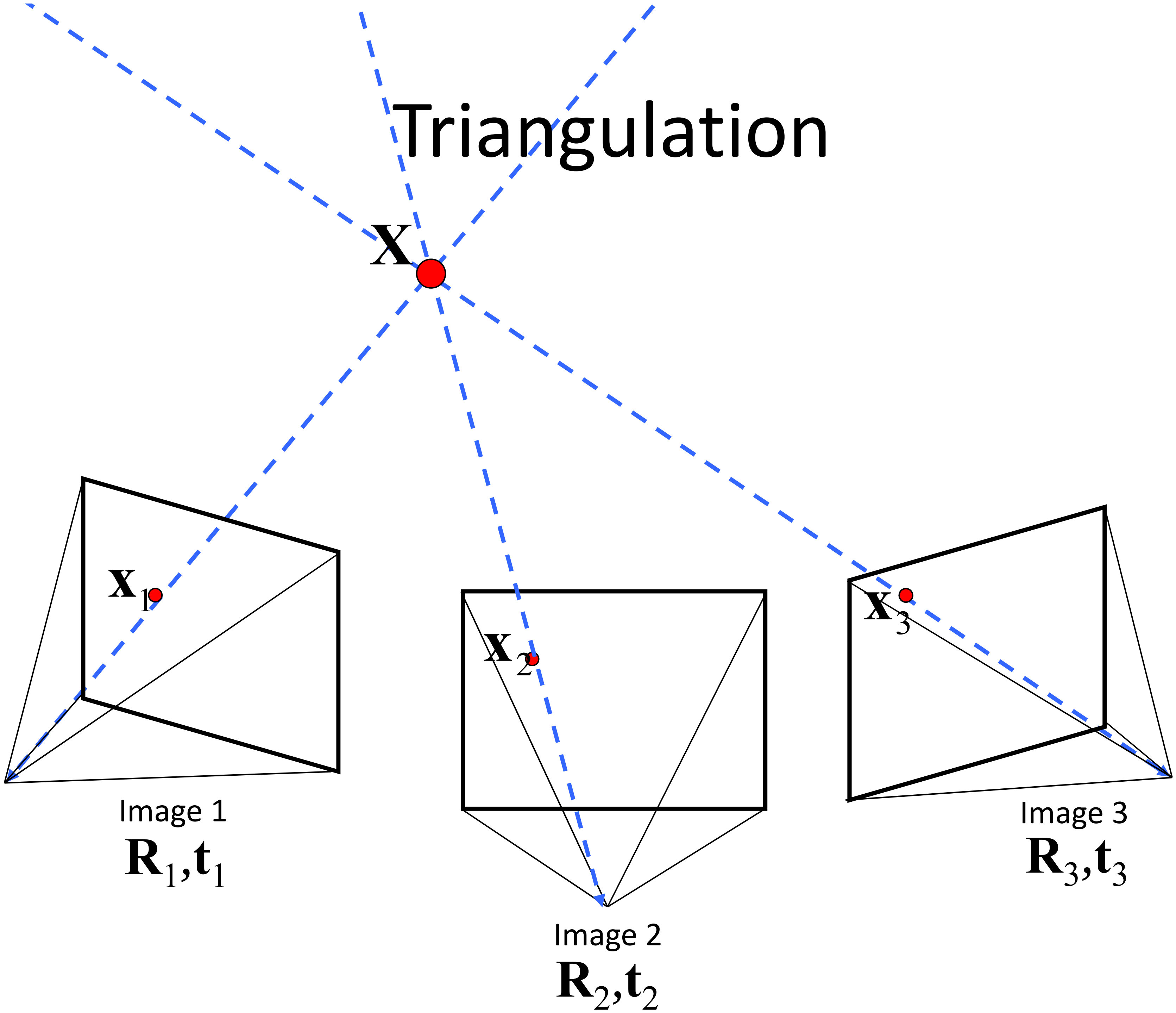


# Camera projection

| | Point 1 | Point 2 | Point 3 |
|---|---|---|---|
| Image 1 | $\mathbf{x}_1^1 = \mathbf{K}\left[\mathbf{R}_1 \middle| \mathbf{t}_1\right]\mathbf{X}^1$ | $\mathbf{x}_1^2 = \mathbf{K}\left[\mathbf{R}_1 \middle| \mathbf{t}_1\right]\mathbf{X}^2$ | |
| Image 2 | $\mathbf{x}_2^1 = \mathbf{K}\left[\mathbf{R}_2 \middle| \mathbf{t}_2\right]\mathbf{X}^1$ | $\mathbf{x}_2^2 = \mathbf{K}\left[\mathbf{R}_2 \middle| \mathbf{t}_2\right]\mathbf{X}^2$ | $\mathbf{x}_2^3 = \mathbf{K}\left[\mathbf{R}_2 \middle| \mathbf{t}_2\right]\mathbf{X}^3$ |
| Image 3 | $\mathbf{x}_3^1 = \mathbf{K}\left[\mathbf{R}_3 \middle| \mathbf{t}_3\right]\mathbf{X}^1$ | | $\mathbf{x}_3^3 = \mathbf{K}\left[\mathbf{R}_3 \middle| \mathbf{t}_3\right]\mathbf{X}^3$ |

Same Camera Same Setting = Same $\mathbf{K}$

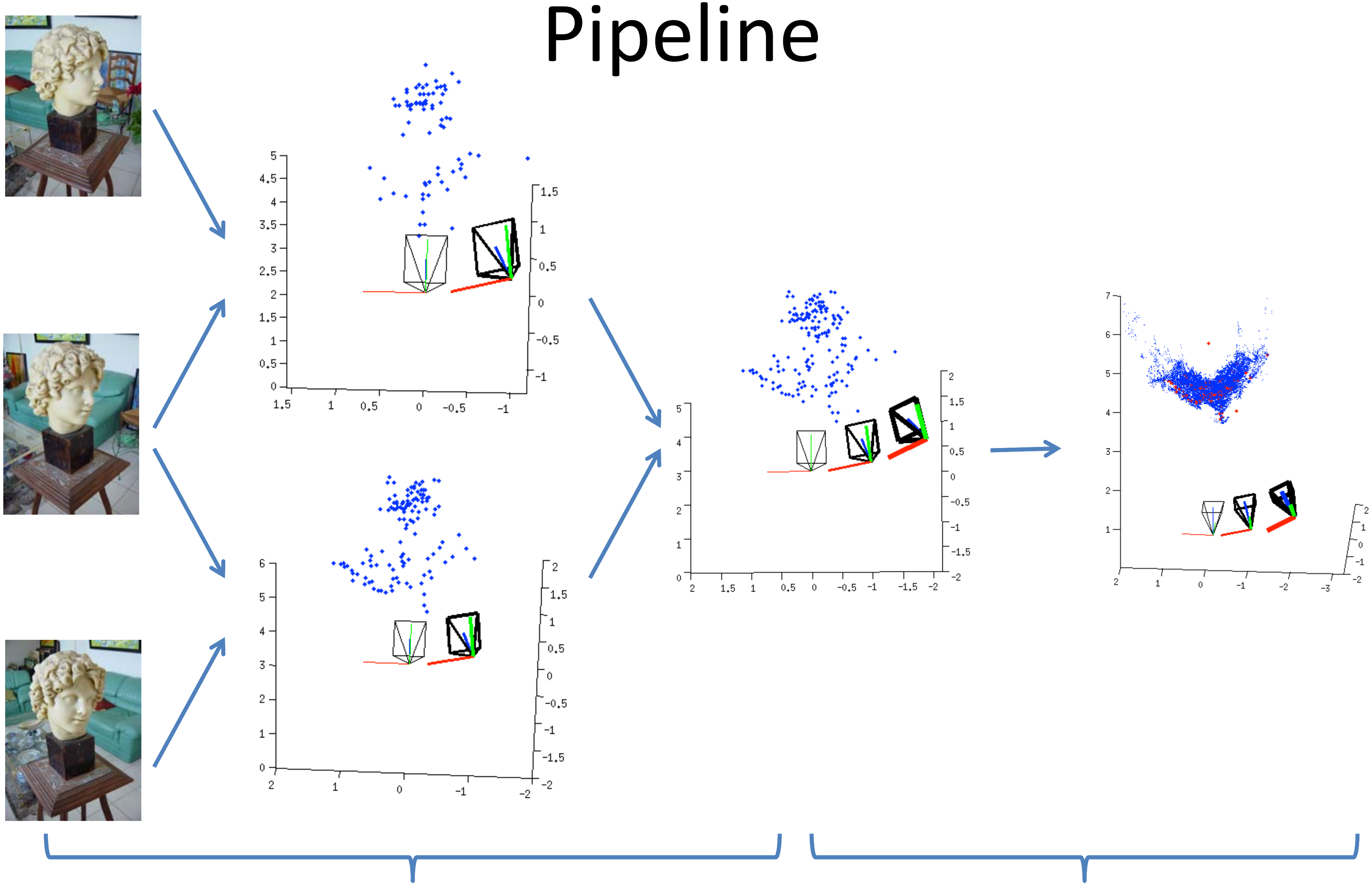
# Triangulation



# Structure From Motion

- Structure = 3D Point Cloud of the Scene
- Motion = Camera Location and Orientation
- SFM = Get the Point Cloud from Moving Cameras
- Structure and Motion: Joint Problems to Solve

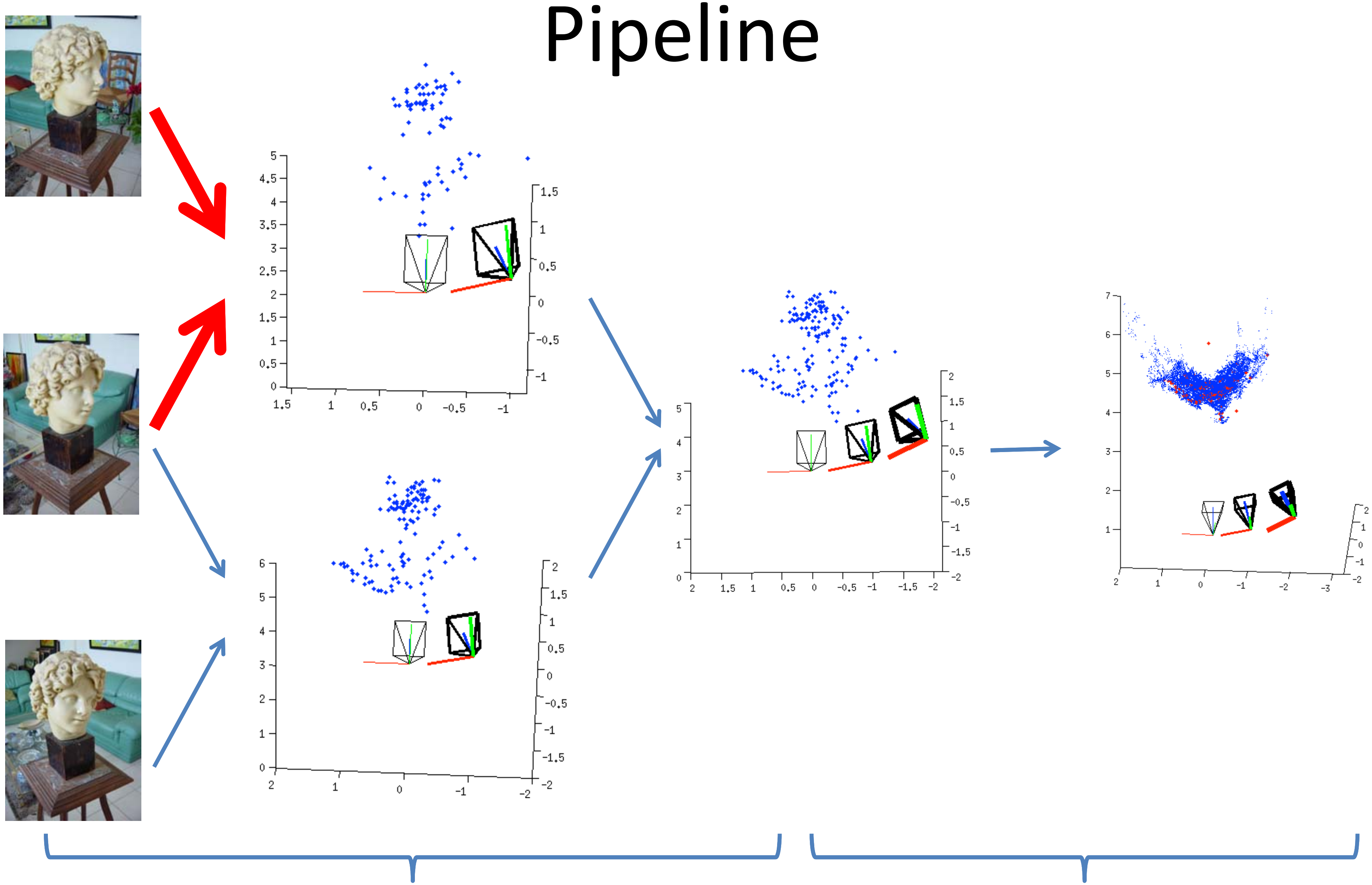
# Pipeline



Structure from Motion (SFM)

Multi-view Stereo (MVS)

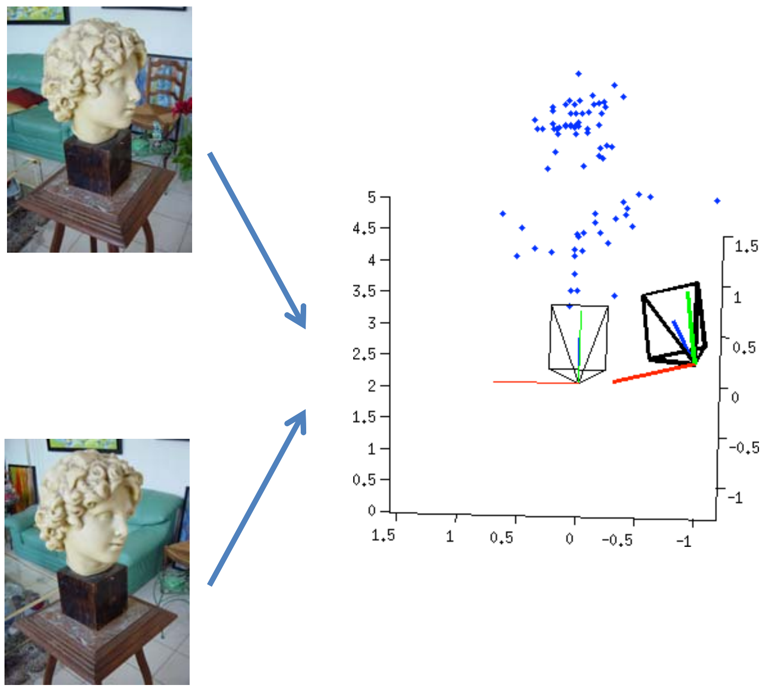
# Pipeline



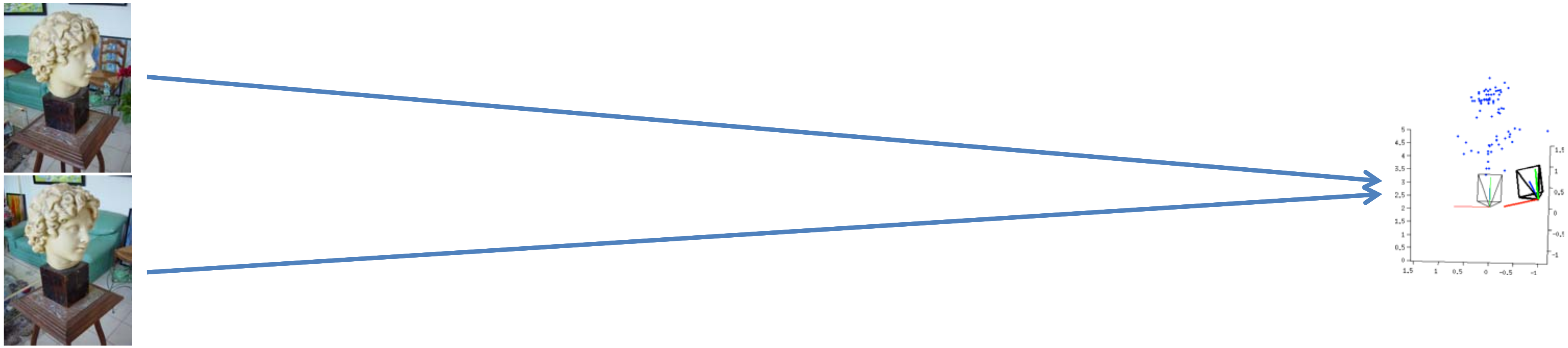
Structure from Motion (SFM)

Multi-view Stereo (MVS)

# Two-view Reconstruction



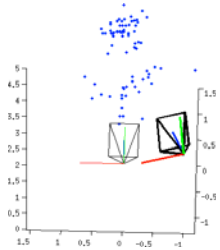
# Two-view Reconstruction



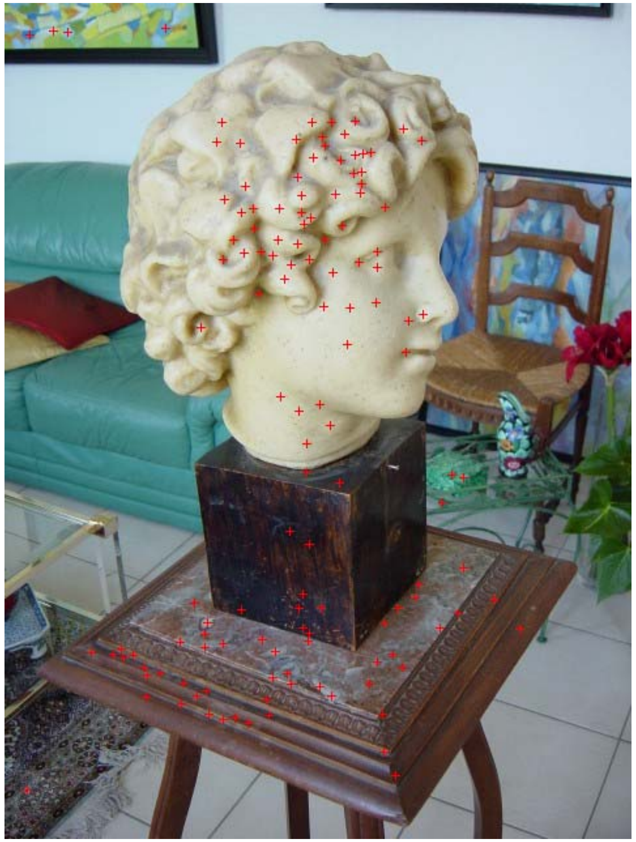
# Two-view Reconstruction

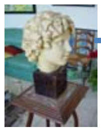
keypoints
keypoints

→ match → fundamental matrix → essential matrix → [R|t] → triangulation →

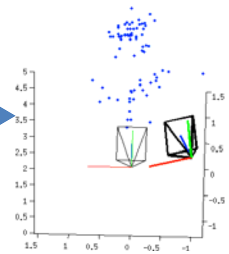


# Keypoints Detection

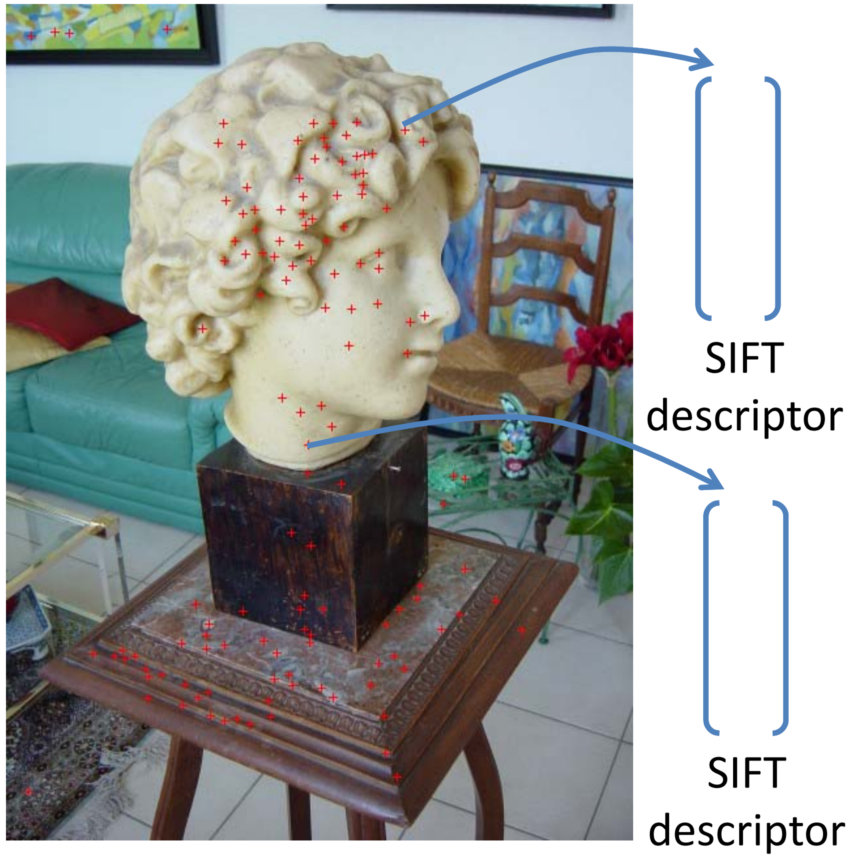


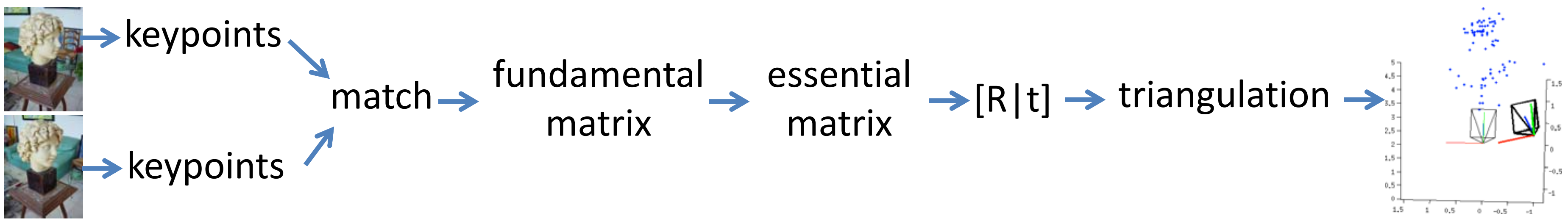
 → keypoints
 → keypoints

match → fundamental matrix → essential matrix → [R|t] → triangulation →

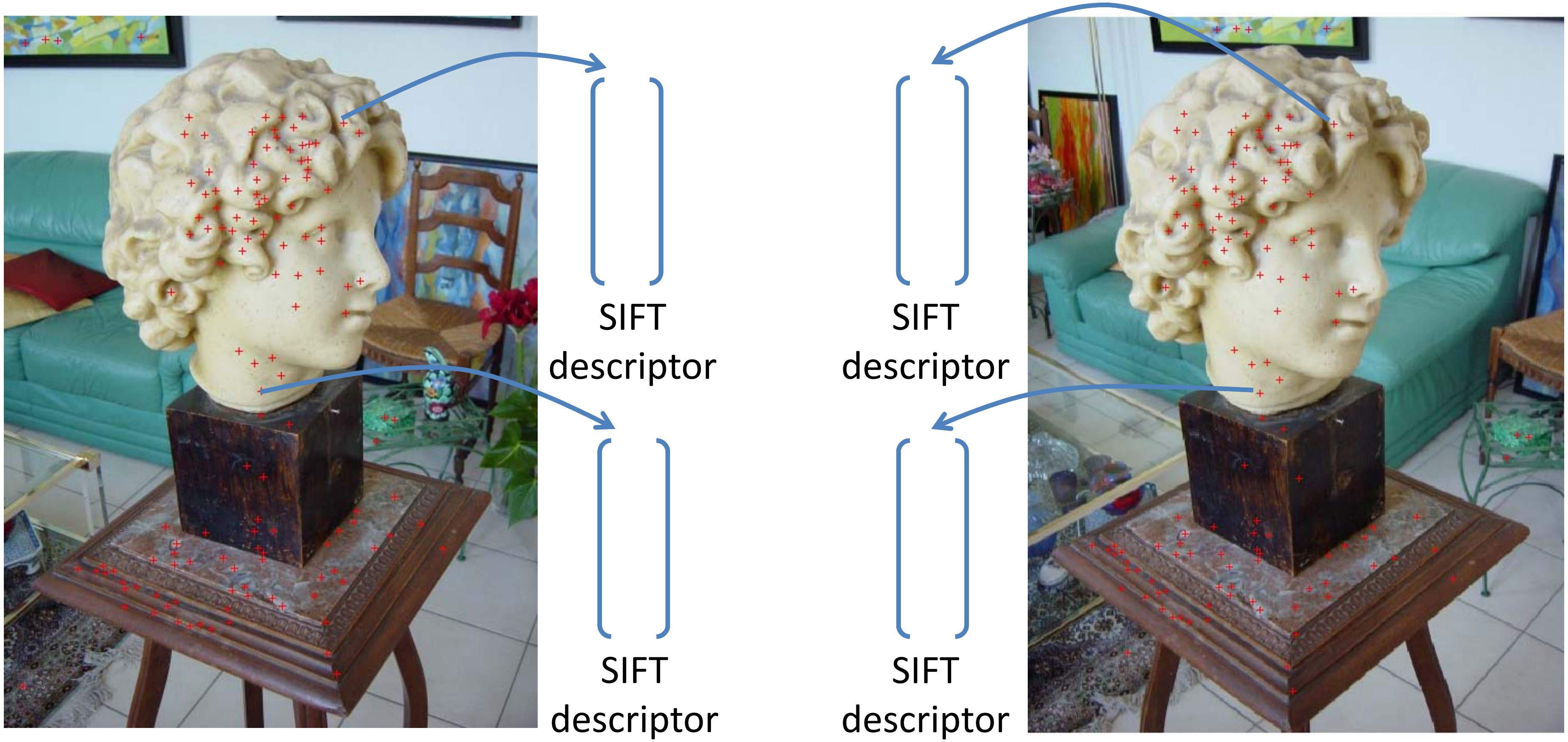


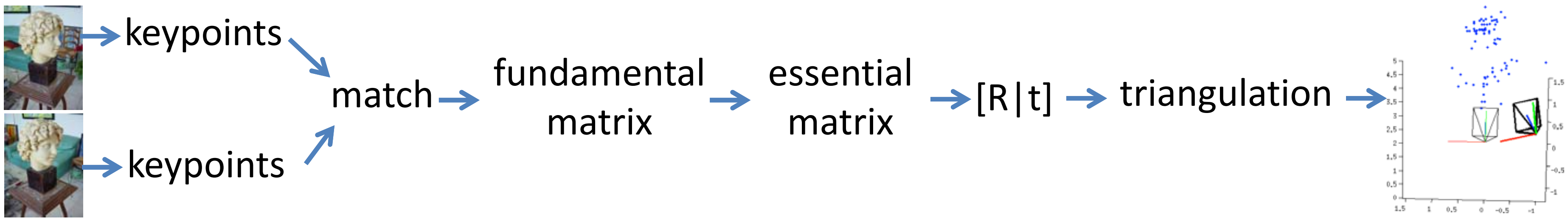
# Descriptor for each point



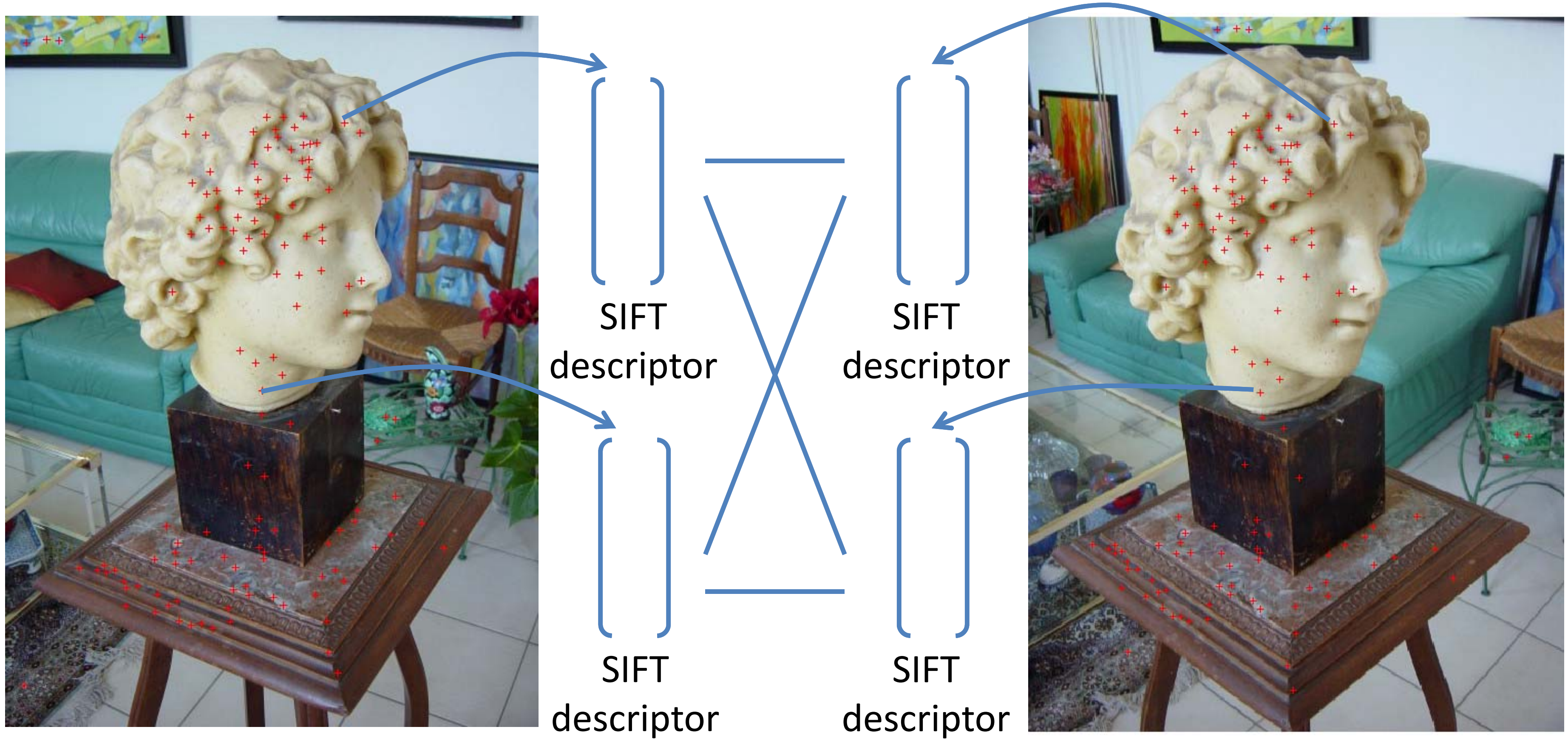



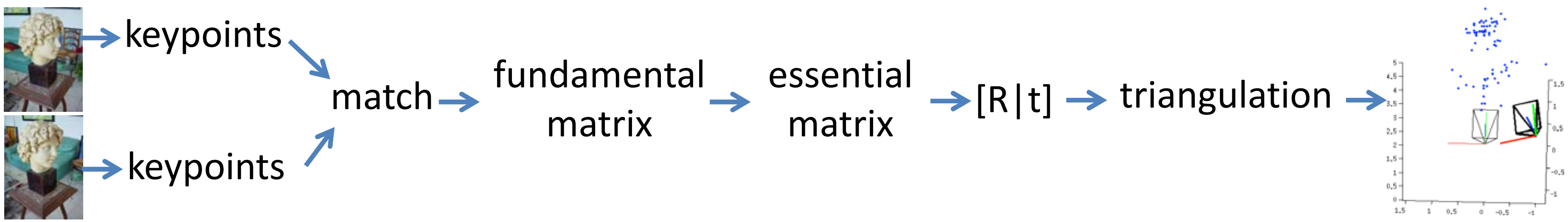
# Same for the other images



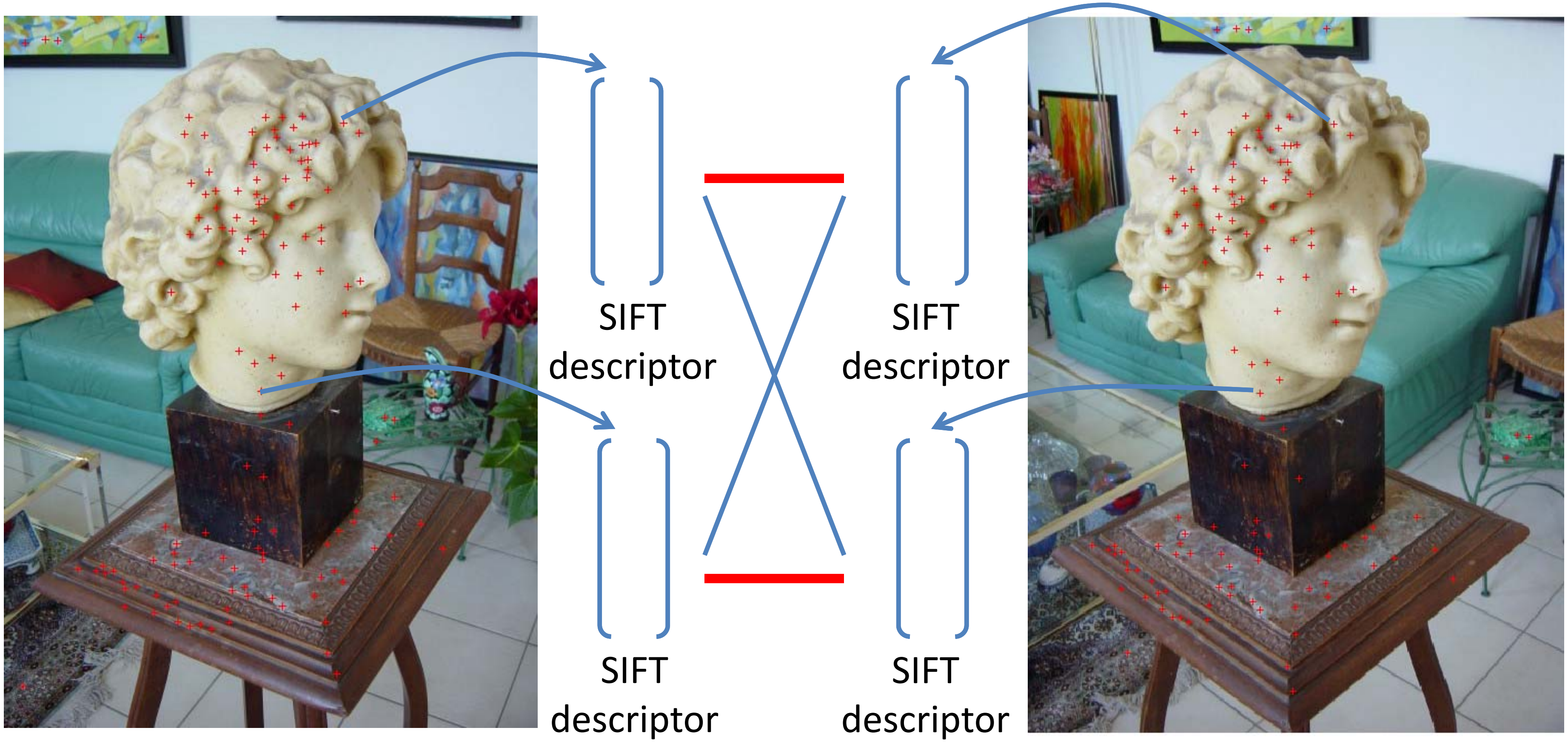



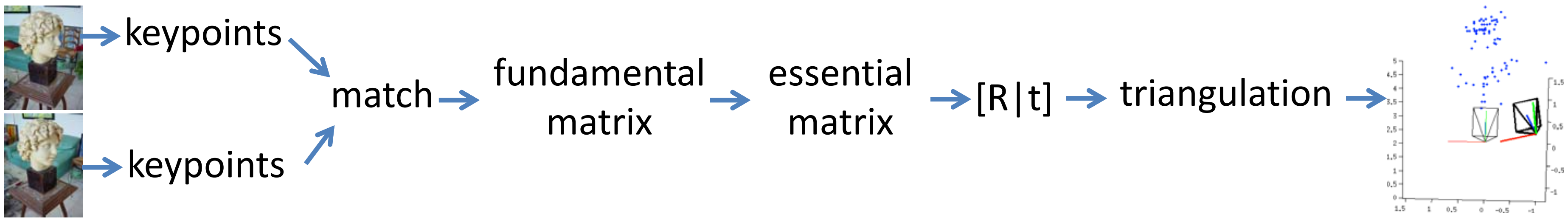
# Point Match for correspondences



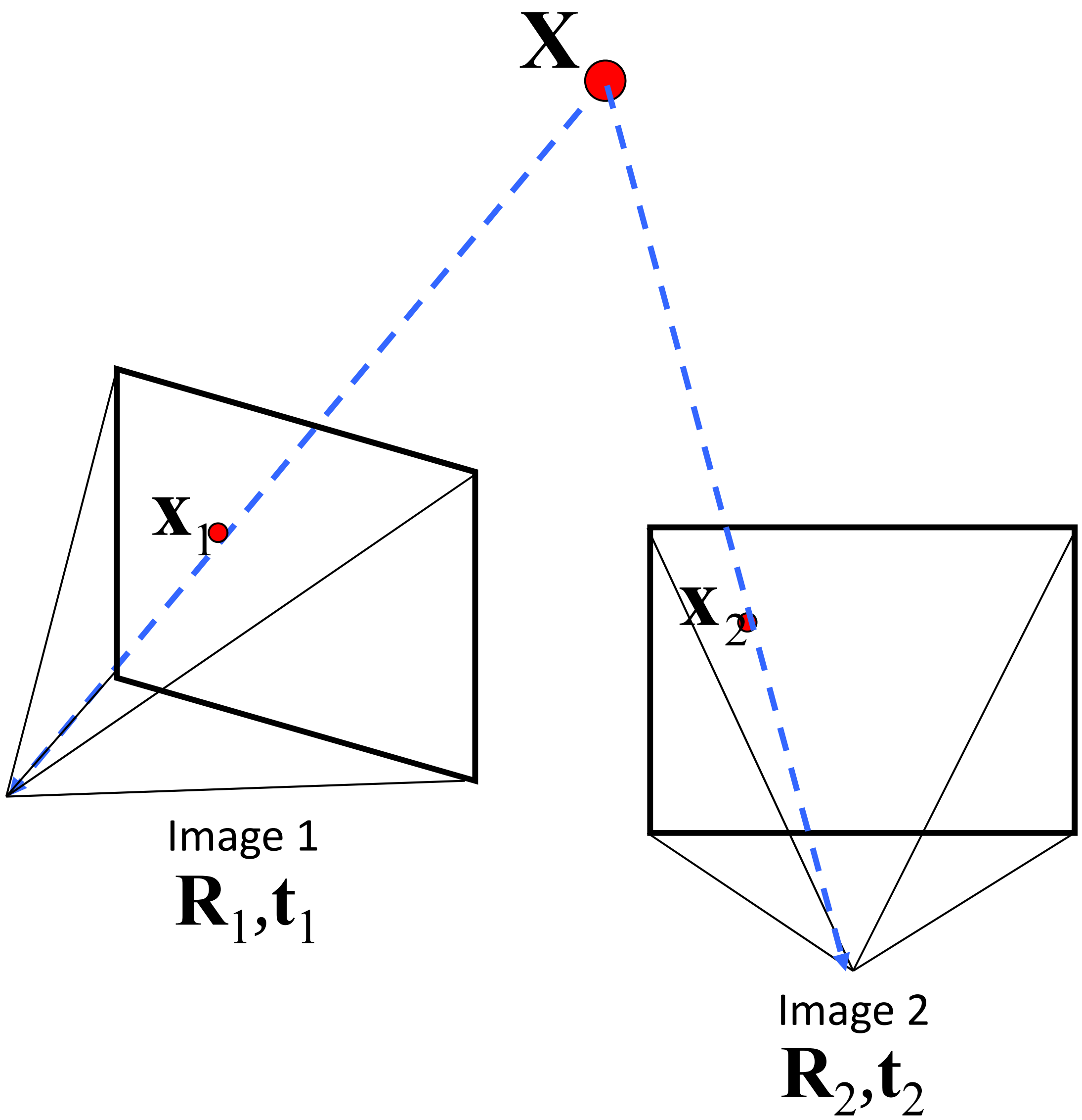



# Point Match for correspondences



# Fundamental Matrix




$$\mathbf{x}_1 \leftrightarrow \mathbf{x}_2$$

$$\mathbf{x}_1^T \mathbf{F} \mathbf{x}_2 = 0$$

# Estimating Fundamental Matrix

- Given a correspondence

$$\mathbf{x}_1 \leftrightarrow \mathbf{x}_2$$

- The basic incidence relation is

$$\mathbf{x}_1^T \mathbf{F} \mathbf{x}_2 = 0 \quad \Longleftrightarrow \quad [x_1x_2, x_1y_2, x_1, y_1x_2, y_1y_2, y_1, x_2, y_2, 1]\begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = 0$$

Need 8 points

# Estimating Fundamental Matrix

$\mathbf{x}_1^T \mathbf{F} \mathbf{x}_2 = 0$ for 8 point correspondences:

$$\mathbf{x}_1^1 \leftrightarrow \mathbf{x}_2^1, \mathbf{x}_1^2 \leftrightarrow \mathbf{x}_2^2, \mathbf{x}_1^3 \leftrightarrow \mathbf{x}_2^3, \mathbf{x}_1^4 \leftrightarrow \mathbf{x}_2^4, \mathbf{x}_1^5 \leftrightarrow \mathbf{x}_2^5, \mathbf{x}_1^6 \leftrightarrow \mathbf{x}_2^6, \mathbf{x}_1^7 \leftrightarrow \mathbf{x}_2^7, \mathbf{x}_1^8 \leftrightarrow \mathbf{x}_2^8$$

$$\begin{bmatrix}
x_1^1 x_2^1 & x_1^1 y_2^1 & x_1^1 & y_1^1 x_2^1 & y_1^1 y_2^1 & y_1^1 & x_2^1 & y_2^1 & 1 \\
x_1^2 x_2^2 & x_1^2 y_2^2 & x_1^2 & y_1^2 x_2^2 & y_1^2 y_2^2 & y_1^2 & x_2^2 & y_2^2 & 1 \\
x_1^3 x_2^3 & x_1^3 y_2^3 & x_1^3 & y_1^3 x_2^3 & y_1^3 y_2^3 & y_1^3 & x_2^3 & y_2^3 & 1 \\
x_1^4 x_2^4 & x_1^4 y_2^4 & x_1^4 & y_1^4 x_2^4 & y_1^4 y_2^4 & y_1^4 & x_2^4 & y_2^4 & 1 \\
x_1^5 x_2^5 & x_1^5 y_2^5 & x_1^5 & y_1^5 x_2^5 & y_1^5 y_2^5 & y_1^5 & x_2^5 & y_2^5 & 1 \\
x_1^6 x_2^6 & x_1^6 y_2^6 & x_1^6 & y_1^6 x_2^6 & y_1^6 y_2^6 & y_1^6 & x_2^6 & y_2^6 & 1 \\
x_1^7 x_2^7 & x_1^7 y_2^7 & x_1^7 & y_1^7 x_2^7 & y_1^7 y_2^7 & y_1^7 & x_2^7 & y_2^7 & 1 \\
x_1^8 x_2^8 & x_1^8 y_2^8 & x_1^8 & y_1^8 x_2^8 & y_1^8 y_2^8 & y_1^8 & x_2^8 & y_2^8 & 1
\end{bmatrix}
\begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = 0$$

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

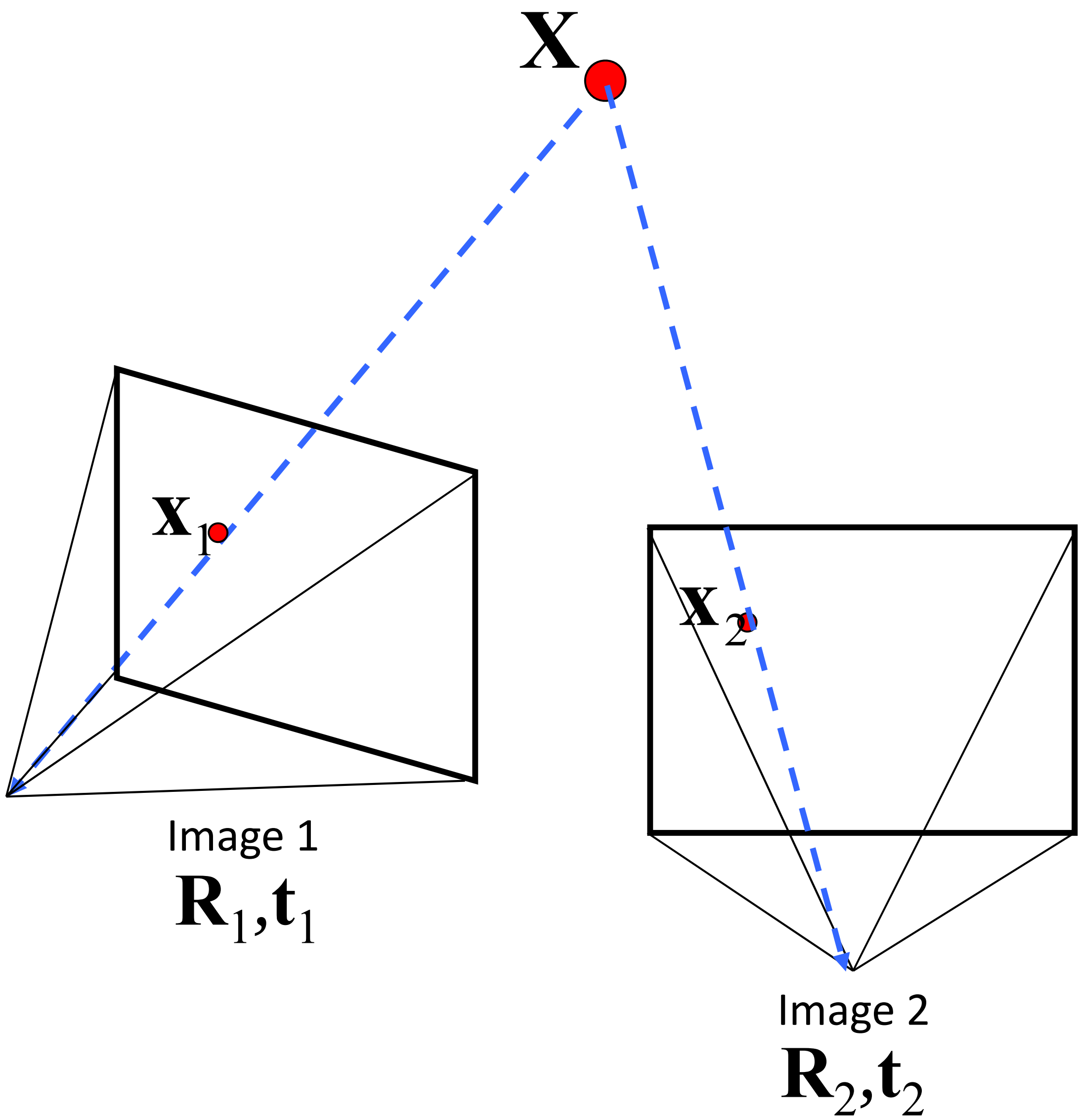
$$\mathbf{A}\mathbf{f} = \mathbf{0}$$

Direct Linear Transformation (DLT)

# RANSAC to Estimate Fundamental Matrix

- For many times
  - Pick 8 points
  - Compute a solution for $\mathbf{F}$ using these 8 points
  - Count number of inliers
- Pick the one with the largest number of inliers

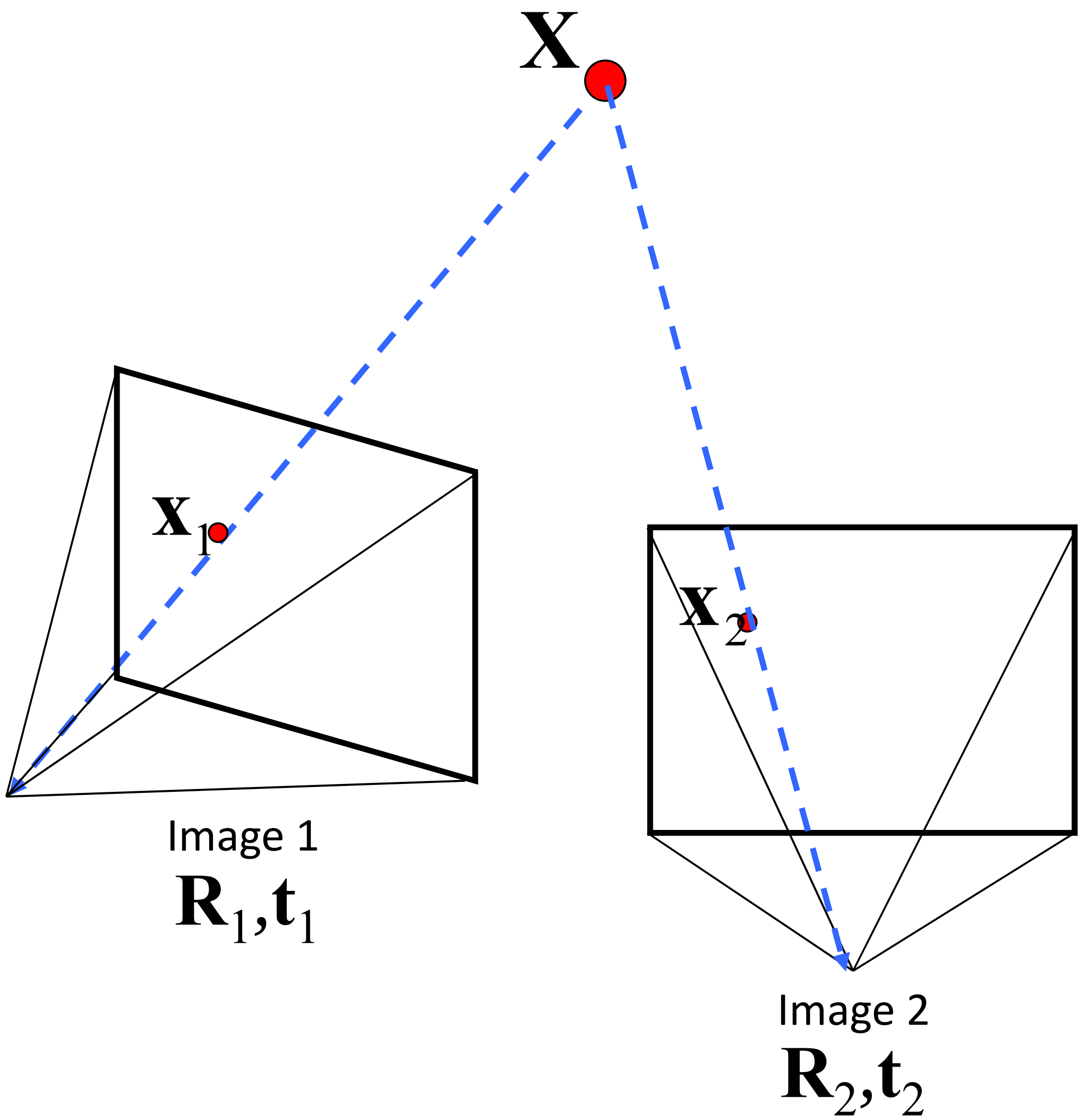
# Fundamental Matrix → Essential Matrix




$$\mathbf{x}_1^T \mathbf{F} \mathbf{x}_2 = 0$$

$$\mathbf{E} = \mathbf{K}_1^T \mathbf{F} \mathbf{K}_2$$

# Essential Matrix → $\left[\mathbf{R}|\mathbf{t}\right]$




$$\mathbf{x}_1^T\mathbf{F}\mathbf{x}_2 = 0$$

$$\mathbf{E} = \mathbf{K}_1^T\mathbf{F}\mathbf{K}_2$$

# Essential Matrix → $\left[\mathbf{R}\middle|\mathbf{t}\right]$

**Result 9.19.** For a given essential matrix

$$\mathbf{E} = \mathbf{U}\mathrm{diag}\left(1,1,0\right)\mathbf{V}^T,$$

and the first camera matrix $\mathbf{P}_1 = \left[\mathbf{I}\middle|\mathbf{0}\right]$, there are four possible choices for the second camera matrix $\mathbf{P}_2$:

$$\mathbf{P}_2 = \left[\mathbf{U}\mathbf{W}\mathbf{V}^T\middle|+\mathbf{u}_3\right]$$

$$\mathbf{P}_2 = \left[\mathbf{U}\mathbf{W}\mathbf{V}^T\middle|-\mathbf{u}_3\right]$$

$$\mathbf{P}_2 = \left[\mathbf{U}\mathbf{W}^T\mathbf{V}^T\middle|+\mathbf{u}_3\right]$$

$$\mathbf{P}_2 = \left[\mathbf{U}\mathbf{W}^T\mathbf{V}^T\middle|-\mathbf{u}_3\right]$$

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

Page 259 of the bible (Multiple View Geometry, 2nd Ed)

# Four Possible Solutions

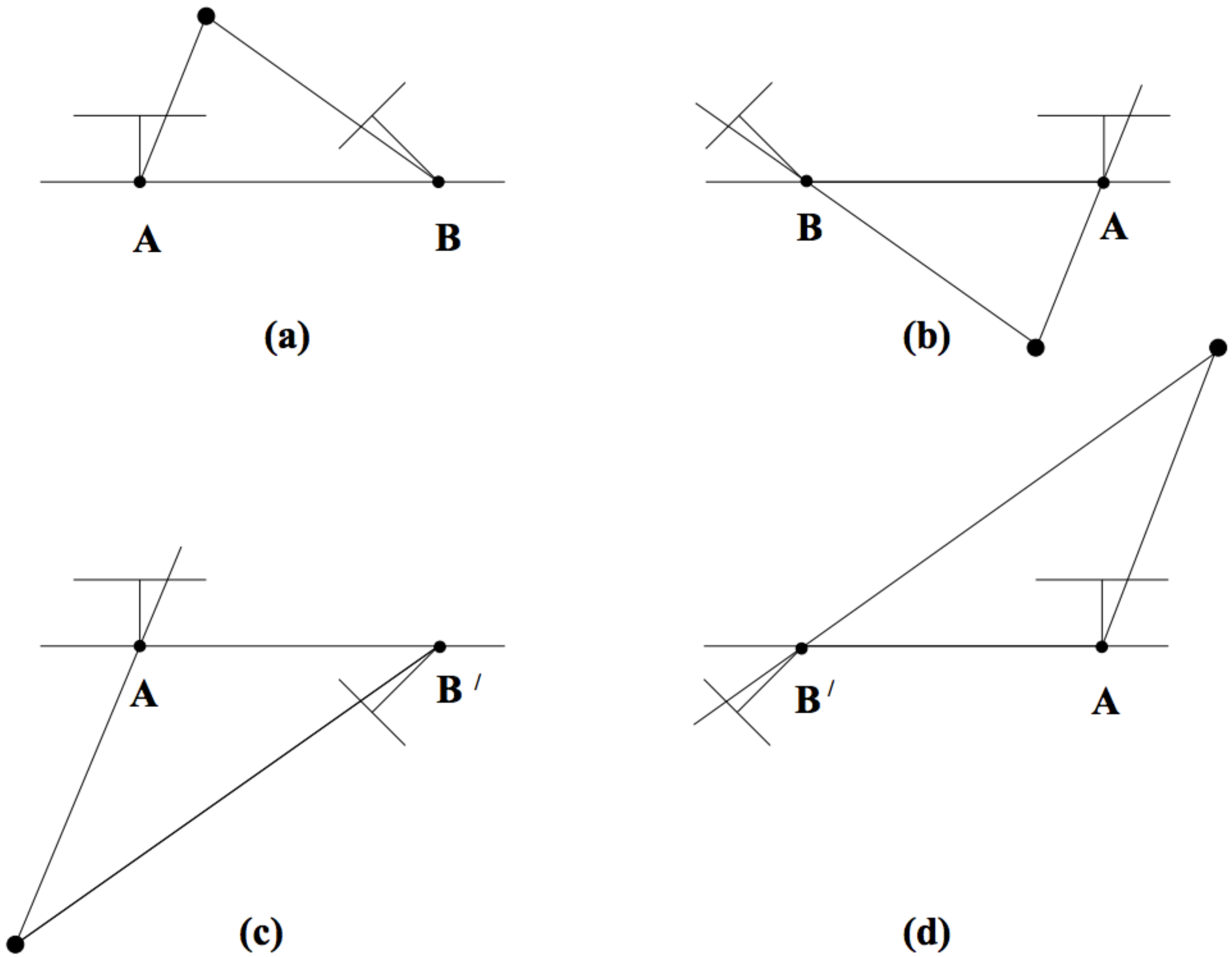



Fig. 9.12. **The four possible solutions for calibrated reconstruction from E.** *Between the left and right sides there is a baseline reversal. Between the top and bottom rows camera B rotates* 180° *about the baseline. Note, only in (a) is the reconstructed point in front of both cameras.*

Page 260 of the bible (Multiple View Geometry, 2nd Ed)

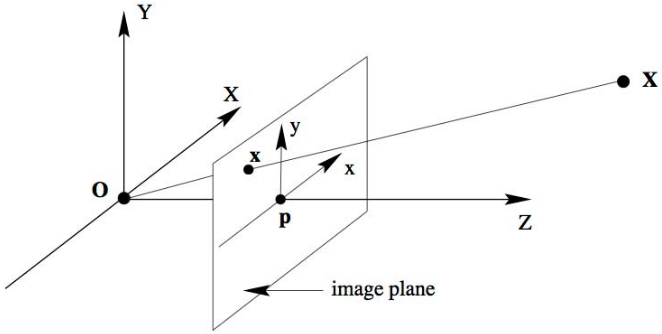
# In front of the camera?

- Camera Extrinsic $\left[\mathbf{R} \middle| \mathbf{t}\right]$

$$\begin{bmatrix} X_{cam} \\ Y_{cam} \\ Z_{cam} \end{bmatrix} = \mathbf{R} \begin{bmatrix} X_{world} \\ Y_{world} \\ Z_{world} \end{bmatrix} + \mathbf{t} \quad \Leftrightarrow \quad \begin{bmatrix} X_{world} \\ Y_{world} \\ Z_{world} \end{bmatrix} = \mathbf{R}^{-1}\left( \begin{bmatrix} X_{cam} \\ Y_{cam} \\ Z_{cam} \end{bmatrix} - \mathbf{t} \right) = \mathbf{R}^{T} \begin{bmatrix} X_{cam} \\ Y_{cam} \\ Z_{cam} \end{bmatrix} - \mathbf{R}^{T}\mathbf{t}$$

- Camera Center

$$\begin{bmatrix} X_{cam} \\ Y_{cam} \\ Z_{cam} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad \mathbf{C} = \begin{bmatrix} X_{world} \\ Y_{world} \\ Z_{world} \end{bmatrix} = \mathbf{R}^{T} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} - \mathbf{R}^{T}\mathbf{t} = -\mathbf{R}^{T}\mathbf{t}$$




- View Direction

$$\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \Rightarrow \quad \left( \mathbf{R}^{T} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \mathbf{R}^{T}\mathbf{t} \right) - (\mathbf{C}) = \left(\mathbf{R}(3,:)^{T} - \mathbf{R}^{T}\mathbf{t}\right) - \left(-\mathbf{R}^{T}\mathbf{t}\right) = \mathbf{R}(3,:)^{T}$$

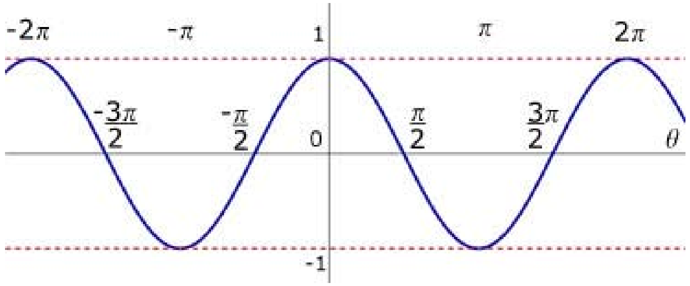
Camera Coordinate System

World Coordinate System

# In front of the camera?

- A point $\mathbf{X}$
- Direction from camera center to point $\mathbf{X}-\mathbf{C}$
- Angle Between Two Vectors

$$\mathbf{A}\cdot\mathbf{B}=\|\mathbf{A}\|\|\mathbf{B}\|\cos\theta$$




- Angle Between $\mathbf{X}-\mathbf{C}$ and View Direction
- Just need to test

$$\left(\mathbf{X}-\mathbf{C}\right)\cdot\mathbf{R}(3,:)^T>0\,?$$

# Pick the Solution

With maximal number of points in front of both cameras.

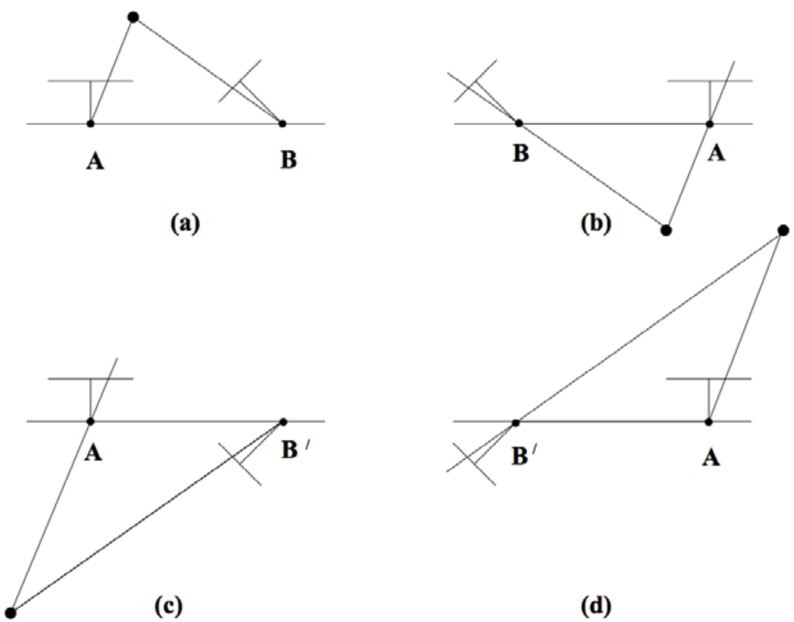
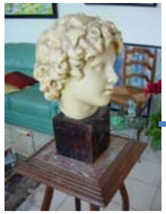



Fig. 9.12. **The four possible solutions for calibrated reconstruction from E.** *Between the left and right sides there is a baseline reversal. Between the top and bottom rows camera B rotates* 180° *about the baseline. Note, only in (a) is the reconstructed point in front of both cameras.*

Page 260 of the bible (Multiple View Geometry, 2nd Ed)

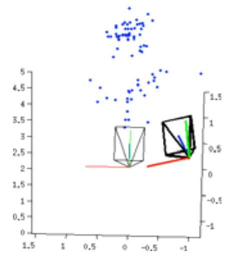
# Two-view Reconstruction



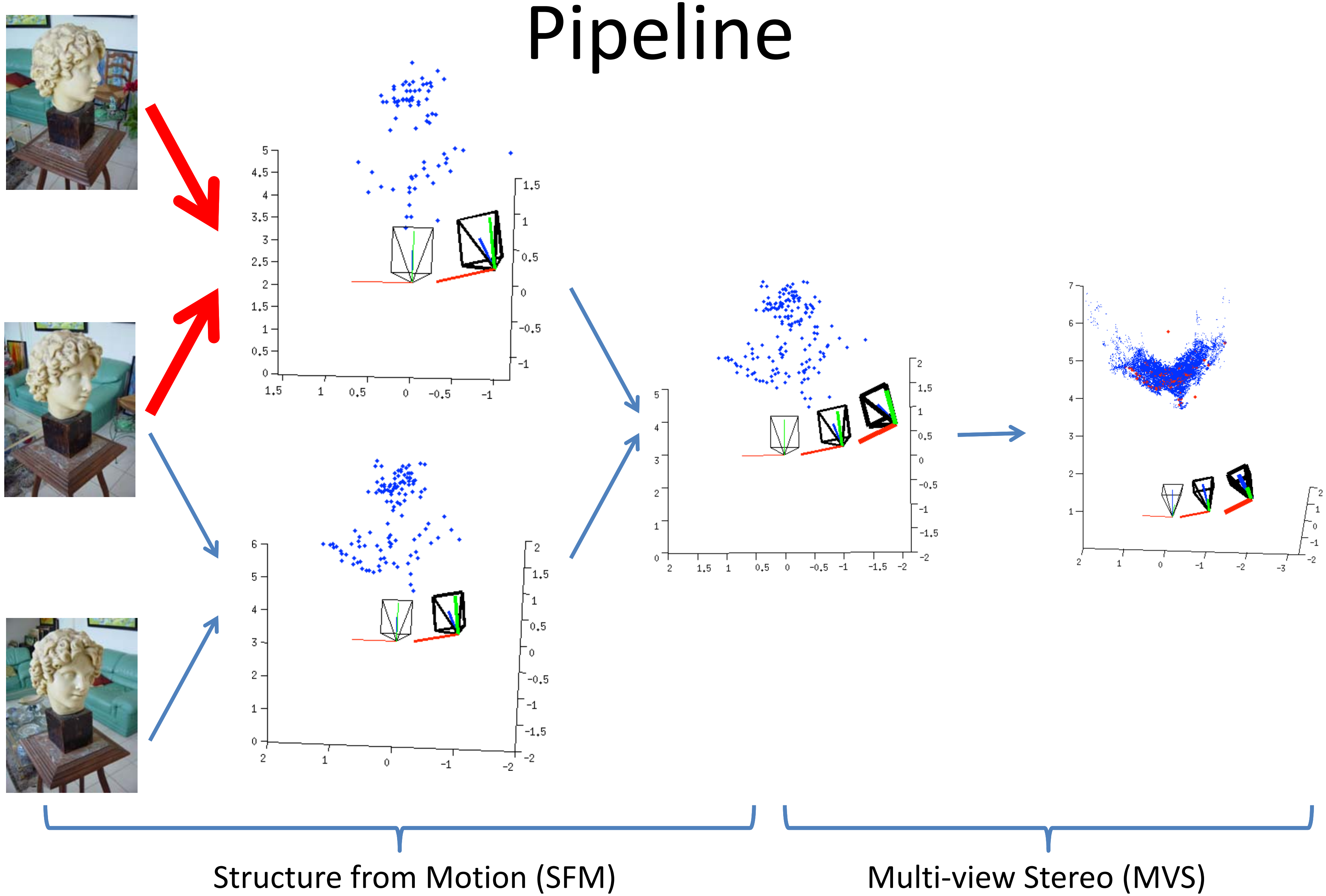
→ keypoints

→ keypoints

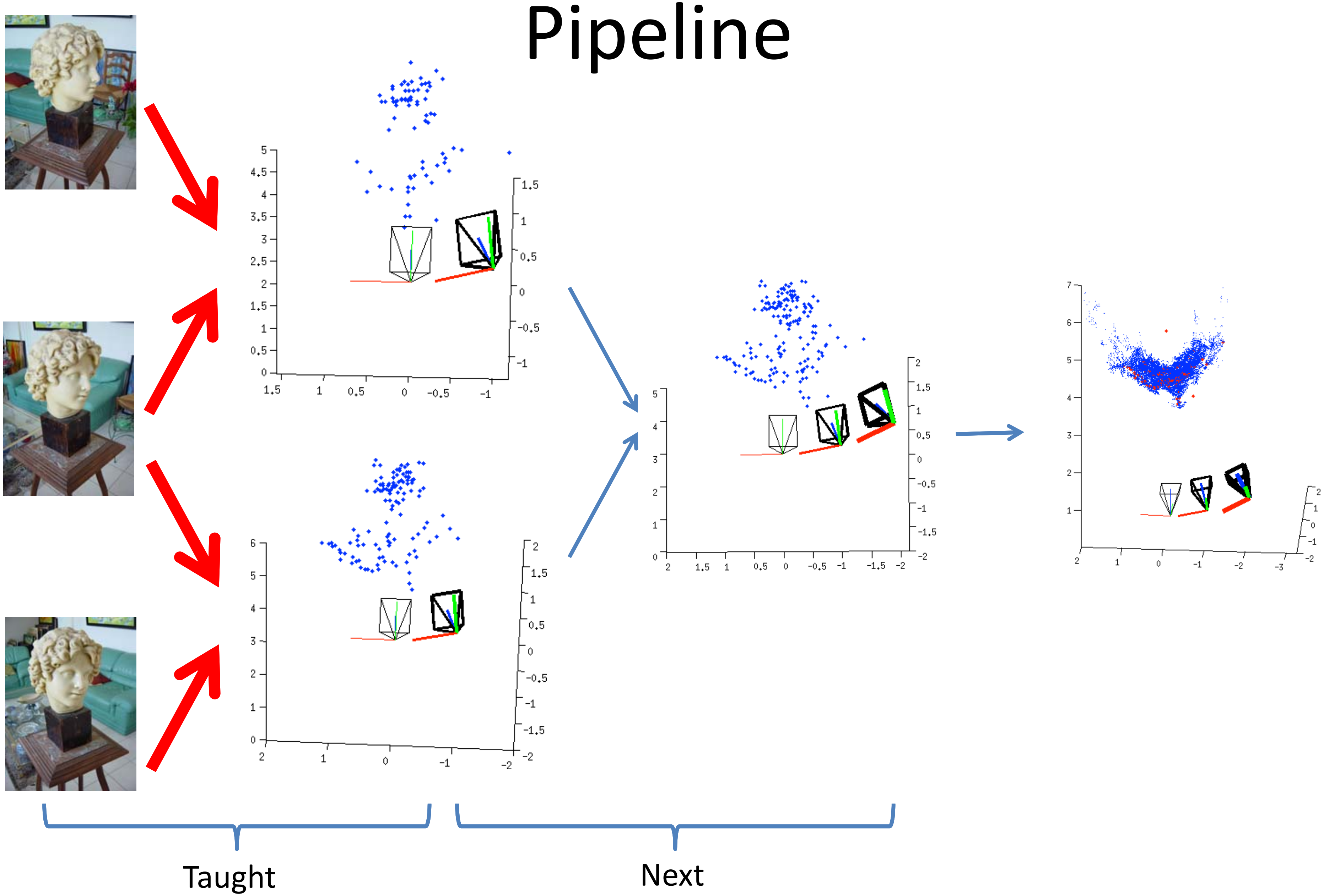
match → fundamental matrix → essential matrix → [R|t] → triangulation →



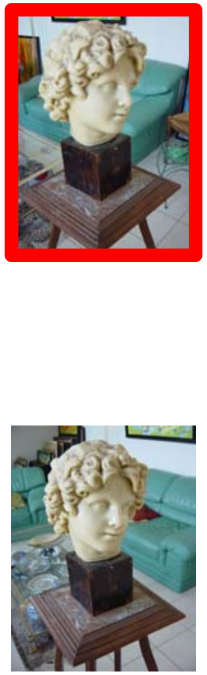
# Pipeline

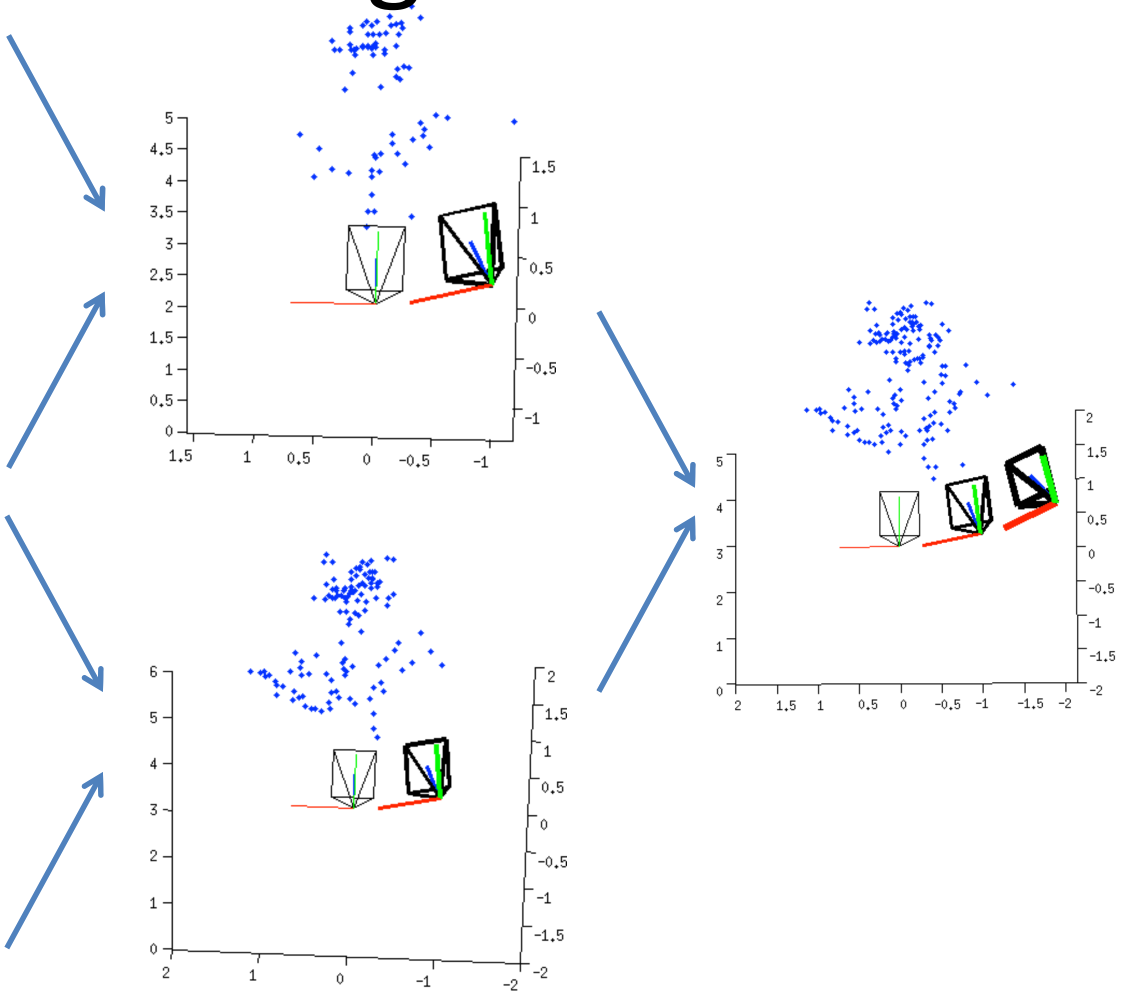


# Pipeline

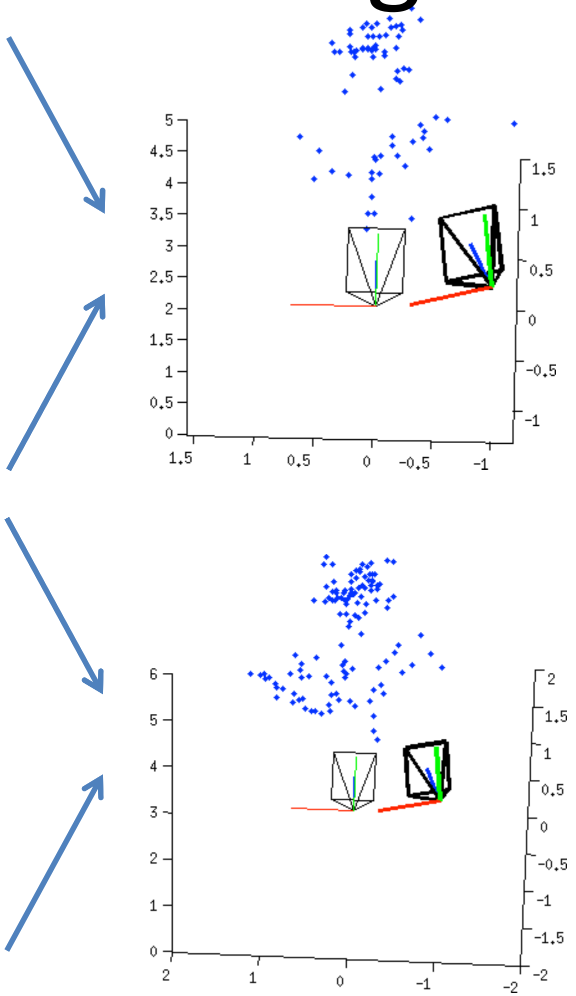


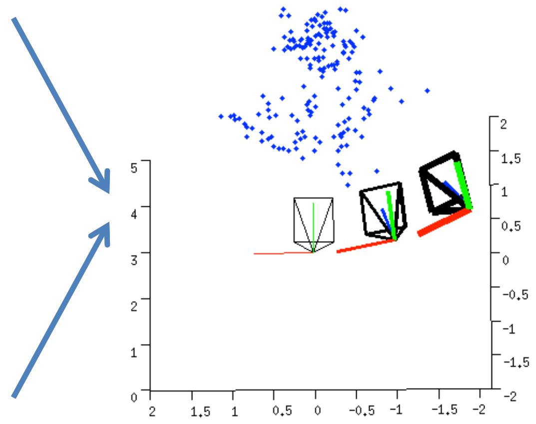
# Merge Two Point Cloud





# Merge Two Point Cloud



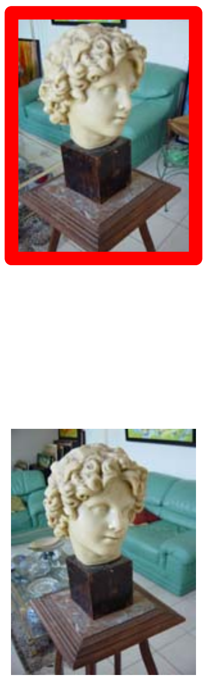


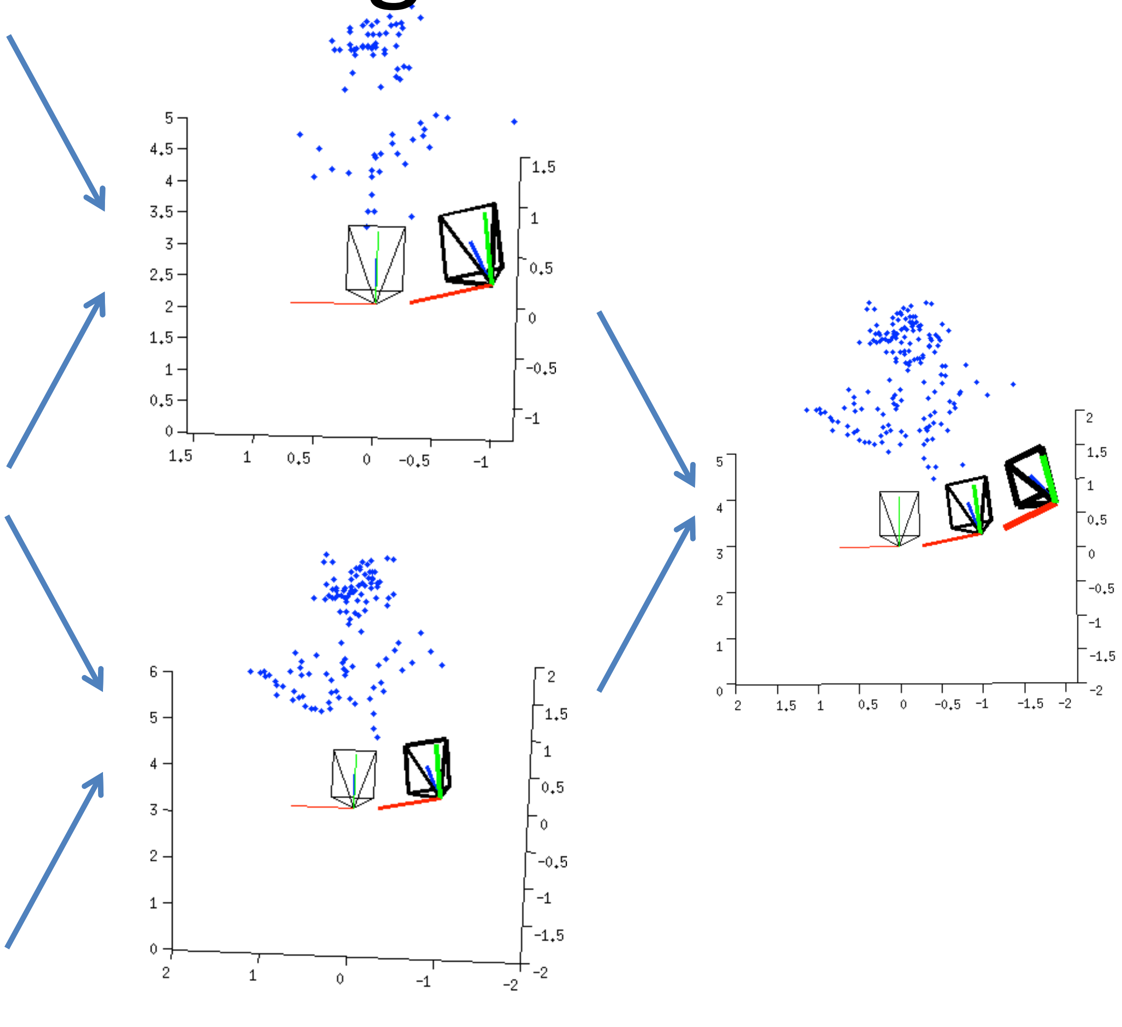
There can be only one $\left[\mathbf{R}_2 | \mathbf{t}_2\right]$

# Merge Two Point Cloud

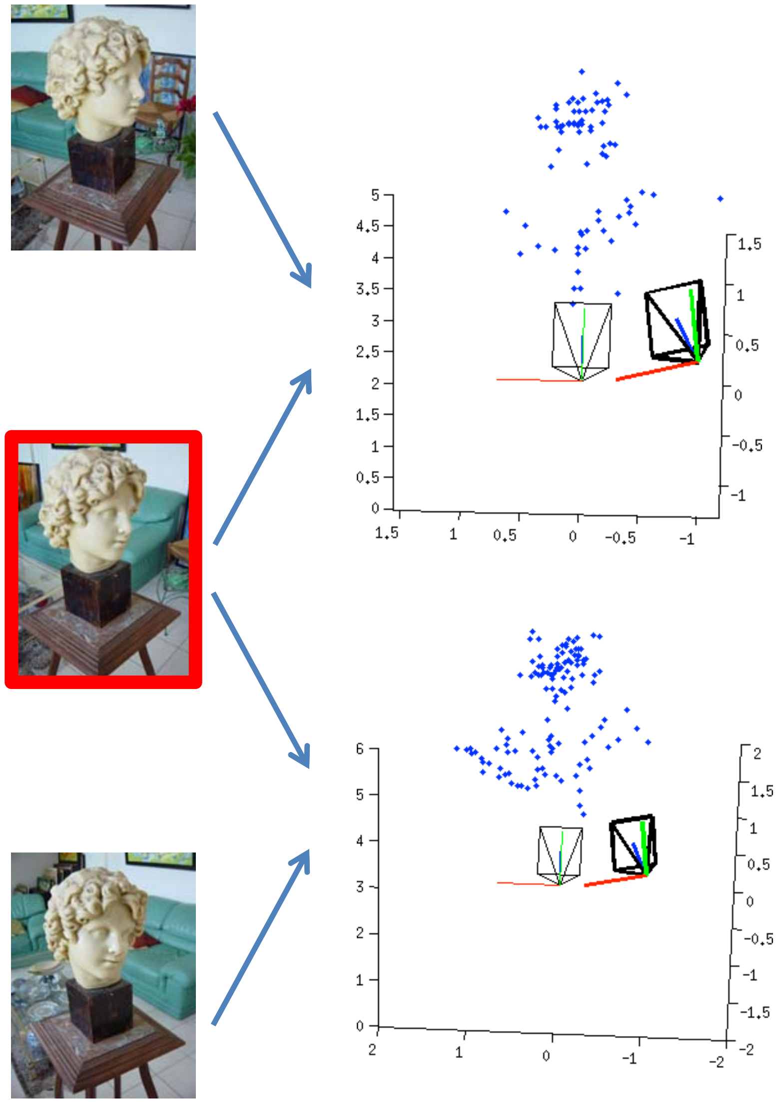
- From the 1st and 2nd images, we have $\left[\mathbf{R}_1 | \mathbf{t}_1\right]$ and $\left[\mathbf{R}_2 | \mathbf{t}_2\right]$
- From the 2nd and 3rd images, we have $\left[\mathbf{R}_2 | \mathbf{t}_2\right]$ and $\left[\mathbf{R}_3 | \mathbf{t}_3\right]$
- **Exercise**: How to transform the coordinate system of the second point cloud to align with the first point cloud so that there is only one $\left[\mathbf{R}_2 | \mathbf{t}_2\right]$?

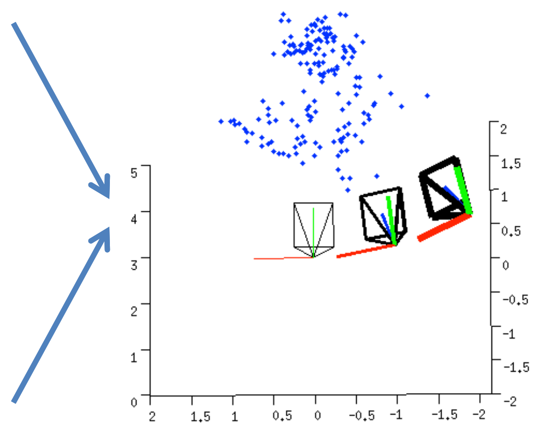
# Merge Two Point Cloud





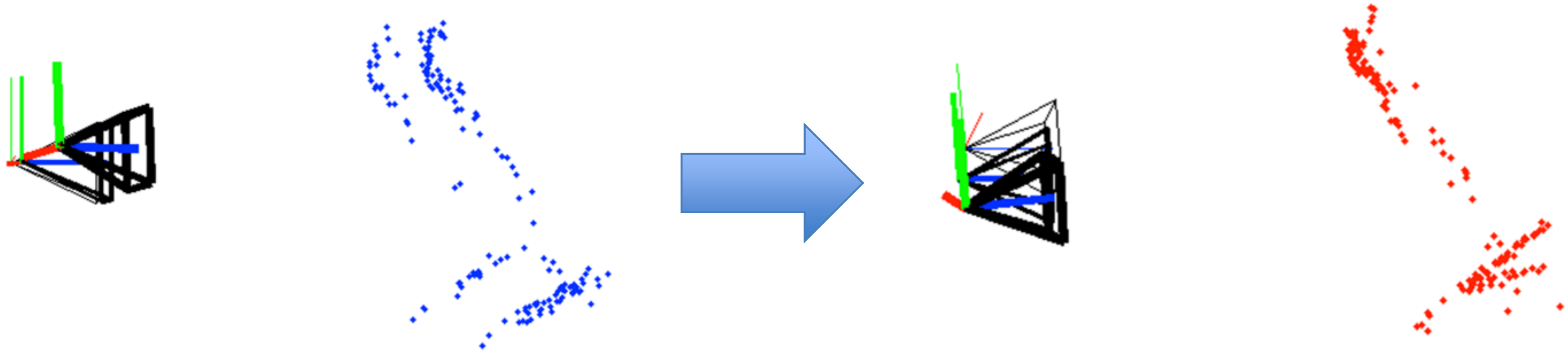
# Oops



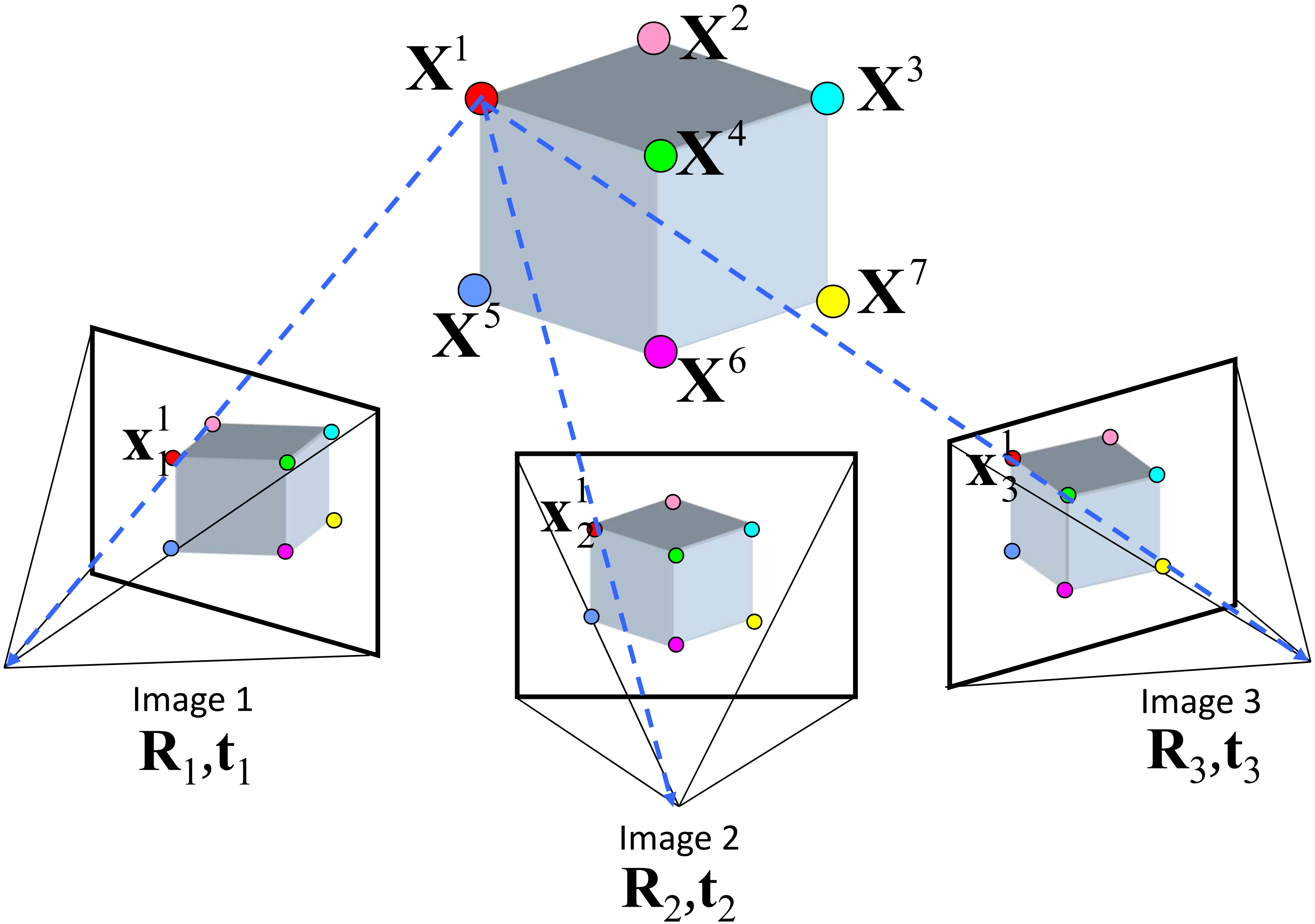




See From a Different Angle

# Bundle Adjustment



# Camera projection



# Camera projection

| | Point 1 | Point 2 | Point 3 |
|---|---|---|---|
| Image 1 | $\mathbf{x}_1^1 = \mathbf{K}\left[\mathbf{R}_1 \middle| \mathbf{t}_1\right]\mathbf{X}^1$ | $\mathbf{x}_1^2 = \mathbf{K}\left[\mathbf{R}_1 \middle| \mathbf{t}_1\right]\mathbf{X}^2$ | |
| Image 2 | $\mathbf{x}_2^1 = \mathbf{K}\left[\mathbf{R}_2 \middle| \mathbf{t}_2\right]\mathbf{X}^1$ | $\mathbf{x}_2^2 = \mathbf{K}\left[\mathbf{R}_2 \middle| \mathbf{t}_2\right]\mathbf{X}^2$ | $\mathbf{x}_2^3 = \mathbf{K}\left[\mathbf{R}_2 \middle| \mathbf{t}_2\right]\mathbf{X}^3$ |
| Image 3 | $\mathbf{x}_3^1 = \mathbf{K}\left[\mathbf{R}_3 \middle| \mathbf{t}_3\right]\mathbf{X}^1$ | | $\mathbf{x}_3^3 = \mathbf{K}\left[\mathbf{R}_3 \middle| \mathbf{t}_3\right]\mathbf{X}^3$ |

Same Camera Same Setting = Same $\mathbf{K}$

# Rethinking the SFM problem

- Input: Observed 2D image position

$$
\begin{array}{ccc}
\tilde{\mathbf{x}}_1^1 & \tilde{\mathbf{x}}_1^2 & \\
\tilde{\mathbf{x}}_2^1 & \tilde{\mathbf{x}}_2^2 & \tilde{\mathbf{x}}_2^3 \\
\tilde{\mathbf{x}}_3^1 & & \tilde{\mathbf{x}}_3^3
\end{array}
$$

- Output:

Unknown Camera Parameters (with some guess)

$$\left[\mathbf{R}_1 | \mathbf{t}_1\right], \left[\mathbf{R}_2 | \mathbf{t}_2\right], \left[\mathbf{R}_3 | \mathbf{t}_3\right]$$

Unknown Point 3D coordinate (with some guess)

$$\mathbf{X}^1, \mathbf{X}^2, \mathbf{X}^3, \cdots$$

# Bundle Adjustment

A valid solution $\left[\mathbf{R}_1|\mathbf{t}_1\right],\left[\mathbf{R}_2|\mathbf{t}_2\right],\left[\mathbf{R}_3|\mathbf{t}_3\right]$ and $\mathbf{X}^1,\mathbf{X}^2,\mathbf{X}^3,\cdots$ must let

$$\text{Re-projection}\left\{\begin{array}{lll}
\mathbf{x}_1^1=\mathbf{K}\left[\mathbf{R}_1|\mathbf{t}_1\right]\mathbf{X}^1 & \mathbf{x}_1^2=\mathbf{K}\left[\mathbf{R}_1|\mathbf{t}_1\right]\mathbf{X}^2 & \\
\mathbf{x}_2^1=\mathbf{K}\left[\mathbf{R}_2|\mathbf{t}_2\right]\mathbf{X}^1 & \mathbf{x}_2^2=\mathbf{K}\left[\mathbf{R}_2|\mathbf{t}_2\right]\mathbf{X}^2 & \mathbf{x}_2^3=\mathbf{K}\left[\mathbf{R}_2|\mathbf{t}_2\right]\mathbf{X}^3 \\
\mathbf{x}_3^1=\mathbf{K}\left[\mathbf{R}_3|\mathbf{t}_3\right]\mathbf{X}^1 & & \mathbf{x}_3^3=\mathbf{K}\left[\mathbf{R}_3|\mathbf{t}_3\right]\mathbf{X}^3
\end{array}\right.$$

$$=$$

$$\text{Observation}\left\{\begin{array}{lll}
\tilde{\mathbf{x}}_1^1 & \tilde{\mathbf{x}}_1^2 & \\
\tilde{\mathbf{x}}_2^1 & \tilde{\mathbf{x}}_2^2 & \tilde{\mathbf{x}}_2^3 \\
\tilde{\mathbf{x}}_3^1 & & \tilde{\mathbf{x}}_3^3
\end{array}\right.$$

# Bundle Adjustment

A valid solution $\left[\mathbf{R}_1 \middle| \mathbf{t}_1\right], \left[\mathbf{R}_2 \middle| \mathbf{t}_2\right], \left[\mathbf{R}_3 \middle| \mathbf{t}_3\right]$ and $\mathbf{X}^1, \mathbf{X}^2, \mathbf{X}^3, \cdots$ must let the Re-projection close to the Observation, i.e. to minimize the reprojection error

$$\min \sum_i \sum_j \left( \tilde{\mathbf{x}}_i^j - \mathbf{K}\left[\mathbf{R}_i \middle| \mathbf{t}_i\right]\mathbf{X}^j \right)^2$$

# Solving This Optimization Problem

- Theory:

  The Levenberg–Marquardt algorithm

  http://en.wikipedia.org/wiki/Levenberg-Marquardt_algorithm

- Practice:

  The Ceres-Solver from Google

  http://code.google.com/p/ceres-solver/