Projective geometry for 3D Computer Vision

Subhashis Banerjee

Computer Science and Engineering
IIT Delhi

NCVPRIPG 2015, IIT Patna
Dec 16, 2015

Overview

- Pin-hole camera
- Why projective geometry?
- Reconstruction

Computer vision geometry: main problems

Correspondence problem: Match image projections of a 3D configuration.

Reconstruction problem: Recover the structure of the 3D configuration from image projections.

Re-projection problem: Is a novel view of a 3D configuration consistent with other views? (Novel view generation)

An infinitely strange perspective

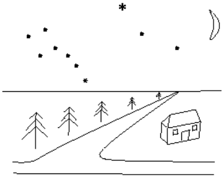


Figure 1.1: Landscape with horizon

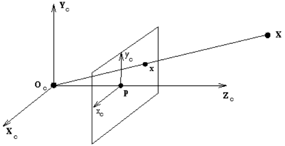
- Parallel lines in 3D space converge in images.
- The line of the horizon is formed by 'infinitely' distant points (vanishing points).
- Any pair of parallel lines meet at a point on the horizon corresponding to their common direction.
- All 'intersections at infinity' stay constant as the observer moves.

3D reconstruction from pin-hole projections

La Flagellazione di Cristo (1460) Galleria Nazionale delle Marche by Piero della Francesca (1416-1492) (Robotics Research Group, Oxford University, 2000)

Pin-hole camera

- The effects can be modelled mathematically using the 'linear perspective' or a 'pin-hole camera' (realized first by Leonardo?)



- If the world coordinates of a point are (X, Y, Z) and the image coordinates are (x, y), then

$$x = fX/Z \text{ and } y = fY/Z$$

- **The model is non-linear**.

In terms of projective coordinates

$$\lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

where,

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \in \mathcal{P}^2 \text{ and } \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \in \mathcal{P}^3$$

are **homogeneous coordinates**.

Euclidean and Affine geometries

- Given a coordinate system, n-dimensional real **affine space** is the set of all points parameterized by $\mathbf{x} = (x_1, \ldots, x_n)^t \in \mathbb{R}^n$.
- An affine transformation is expressed as

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

where $\mathbf{A}$ is a $n \times n$ (usually) non-singular matrix and $\mathbf{b}$ is a $n \times 1$ vector representing a translation.

- By *SVD*

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T = (\mathbf{U}\mathbf{V}^T)(\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^T) = R(\theta)R(-\phi)\boldsymbol{\Sigma}R(\phi)$$

where where

$$\boldsymbol{\Sigma} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

Euclidean and Affine geometries

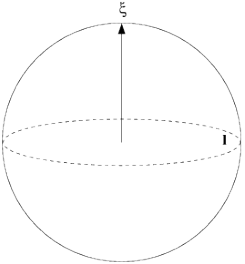
- In the special case of when $\mathbf{A}$ is a rotation (i.e., $\mathbf{A}\mathbf{A}^t = \mathbf{A}^t\mathbf{A} = \mathbf{I}$, then the transformation is *Euclidean*.
- An affine transformation preserves parallelism and ratios of lengths along parallel directions.
- An Euclidean transformation, in addition to the above, also preserves lengths and angles.
- *Since an affine (or Euclidean) transformation preserves parallelism it cannot be used to describe a pinhole projection*.

Spherical geometry

- **The space $\mathcal{S}^2$:**

$$\mathcal{S}^2 = \left\{\mathbf{x} \in \mathbb{R}^3 : ||x|| = 1\right\}$$

- **lines in $\mathcal{S}^2$:** Viewed as a set in $\mathbb{R}^3$ this is the intersection of $\mathcal{S}^2$ with a plane through the origin. We will call this great circle a line in $\mathcal{S}^2$. Let ξ be a unit vector. Then, $\mathbf{l} = \left\{\mathbf{x} \in \mathcal{S}^2 : \xi^t\mathbf{x} = 0\right\}$ is the line with pole ξ.



Spherical geometry

- Lines in $\mathcal{S}^2$ cannot be parallel. Any two lines intersect at a pair of antipodal points.
- A point on a line:

$$\mathbf{l} \cdot \mathbf{x} = 0 \text{ or } \mathbf{l}^T \mathbf{x} = 0 \text{ or } \mathbf{x}^T \mathbf{l} = 0$$

- Two points define a line:

$$\mathbf{l} = \mathbf{p} \times \mathbf{q}$$

- Two lines define a point:

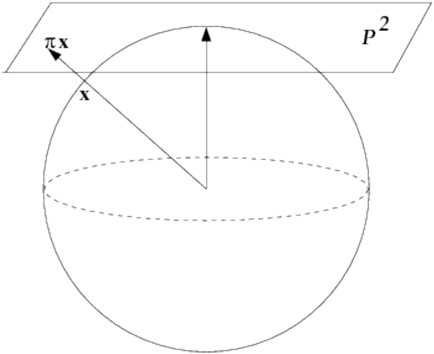
$$\mathbf{x} = \mathbf{l} \times \mathbf{m}$$

Projective geometry

- *The projective plane $\mathcal{P}^2$ is the set of all pairs $\{\mathbf{x}, -\mathbf{x}\}$ of antipodal points in $\mathcal{S}^2$.*
- Two alternative definitions of $\mathcal{P}^2$, equivalent to the preceding one are
  1. The set of all lines through the origin in $\mathbb{R}^3$.
  2. The set of all equivalence classes of ordered triples (x_1, x_2, x_3) of numbers (i.e., vectors in $\mathbb{R}^3$) not all zero, where two vectors are equivalent if they are proportional.

Projective geometry

The space $\mathcal{P}^2$ can be thought of as the infinite plane tangent to the space $\mathcal{S}^2$ and passing through the point $(0, 0, 1)^t$.



Projective geometry

- Let $\pi : \mathcal{S}^2 \to \mathcal{P}^2$ be the mapping that sends $\mathbf{x}$ to $\{\mathbf{x}, -\mathbf{x}\}$. The π is a two-to-one map of $\mathcal{S}^2$ onto $\mathcal{P}^2$.
- A line of $\mathcal{P}^2$ is a set of the form $\pi\mathbf{l}$, where $\mathbf{l}$ is a line of $\mathcal{S}^2$. Clearly, $\pi\mathbf{x}$ lies on $\pi\mathbf{l}$ if and only if $\xi^t\mathbf{x} = 0$.
- **Homogeneous coordinates:** In general, points of real n-dimensional **projective space**, $\mathcal{P}^n$, are represented by $n+1$ component column vectors $(x_1, \ldots, x_n, x_{n+1}) \in \mathbb{R}^{n+1}$ such that at least one x_i is non-zero and $(x_1, \ldots, x_n, x_{n+1})$ and $(\lambda x_1, \ldots, \lambda x_n, \lambda x_{n+1})$ represent the same point of $\mathcal{P}^n$ for all $\lambda \neq 0$.
- $(x_1, \ldots, x_n, x_{n+1})$ is the homogeneous representation of a projective point.

Canonical injection of $\mathbb{R}^n$ into $\mathcal{P}^n$

- Affine space $\mathbb{R}^n$ can be embedded in $\mathcal{P}^n$ by

$$(x_1, \ldots, x_n) \to (x_1, \ldots, x_n, 1)$$

- Affine points can be recovered from projective points with $x_{n+1} \neq 0$ by

$$(x_1, \ldots, x_n) \sim (\frac{x_1}{x_{n+1}}, \ldots, \frac{x_n}{x_{n+1}}, 1) \to (\frac{x_1}{x_{n+1}}, \ldots, \frac{x_n}{x_{n+1}})$$

- A projective point with $x_{n+1} = 0$ corresponds to a **point at infinity**.
- The ray $(x_1, \ldots, x_n, 0)$ can be viewed as an additional **ideal point** as $(x_1, \ldots, x_n)$ recedes to infinity in a certain direction. For example, in $\mathcal{P}^2$,

$$\lim_{T \to 0} (X/T, Y/T, 1) = \lim_{T \to 0} (X, Y, T) = (X, Y, 0)$$

Lines in $\mathcal{P}^2$

- A line equation in $\mathbb{R}^2$ is

$$a_1 x_1 + a_2 x_2 + a_3 = 0$$

- Substituting by homogeneous coordinates $x_i = X_i / X_3$ we get a homogeneous linear equation

$$(a_1, a_2, a_3) \cdot (X_1, X_2, X_3) = \sum_{i=1}^{3} a_i X_i = 0, \ \mathbf{X} \in \mathcal{P}^2$$

- A line in $\mathcal{P}^2$ is represented by a homogeneous 3-vector (a_1, a_2, a_3).
- A point on a line: $\mathbf{a} \cdot \mathbf{X} = 0$ or $\mathbf{a}^T \mathbf{X} = 0$ or $\mathbf{X}^T \mathbf{a} = 0$
- Two points define a line: $\mathbf{l} = \mathbf{p} \times \mathbf{q}$
- Two lines define a point: $\mathbf{x} = \mathbf{l} \times \mathbf{m}$

The line at infinity

- The **line at infinity** ($\mathbf{l}_\infty$): is the line of equation $X_3 = 0$. Thus, the homogeneous representation of $\mathbf{l}_\infty$ is $(0, 0, 1)$.
- The line (u_1, u_2, u_3) intersects $\mathbf{l}_\infty$ at the point $(-u_2, u_1, 0)$.
- Points on $\mathbf{l}_\infty$ are directions of affine lines in the embedded affine space (can be extended to higher dimensions).

Conics in $\mathcal{P}^2$

A **conic** in affine space (inhomogeneous coordinates) is

$$ax^2 + bxy + cy^2 + dx + ey + f = 0$$

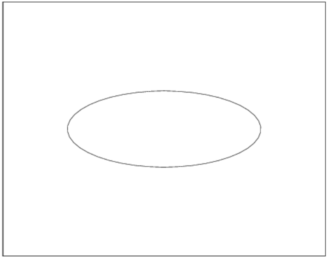
Homogenizing this by replacements $x = X_1/X_3$ and $y = Y_1/Y_3$, we obtain

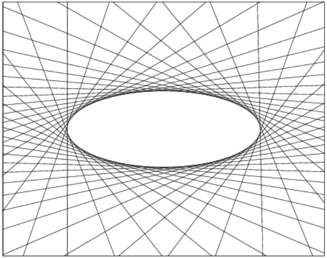
$$aX_1^2 + bX_1X_2 + cX_2^2 + dX_1X_3 + eX_2X_3 + fX_3^2 = 0$$

which can be written in matrix notation as $\mathbf{X}^T\mathbf{C}\mathbf{X} = 0$ where C is symmetric and is the *homogeneous representation* of a **conic**.

Conics in $\mathcal{P}^2$

- The line $\mathbf{l}$ tangent to a conic $\mathbf{C}$ at any point $\mathbf{x}$ is given by $\mathbf{l} = \mathbf{C}\mathbf{x}$.
- $\mathbf{x}^t\mathbf{C}\mathbf{x} = 0 \implies (\mathbf{C}^{-1}\mathbf{l})^t\mathbf{C}((\mathbf{C}^{-1}\mathbf{l}) = \mathbf{l}^t\mathbf{C}^{-1}\mathbf{l} = 0$ (because $\mathbf{C}^{-t} = \mathbf{C}^{-1}$). This is the equation of the *dual conic*.





Conics in $\mathcal{P}^2$

- The *degenerate conic* of rank 2 is defined by two line $\mathbf{l}$ and $\mathbf{m}$ as
$$\mathbf{C} = \mathbf{l}\mathbf{m}^t + \mathbf{m}\mathbf{l}^t$$
Points on line $\mathbf{l}$ satisfy $\mathbf{l}^t\mathbf{x} = 0$ and are hence on the conic because $(\mathbf{x}^t\mathbf{l})(\mathbf{m}^t\mathbf{x}) + (\mathbf{x}^t\mathbf{m})(\mathbf{l}^t\mathbf{x}) = 0$. (Similarly for $\mathbf{m}$). The dual conic $\mathbf{x}\mathbf{y}^t + \mathbf{y}\mathbf{x}^t$ represents lines passing through $\mathbf{x}$ and $\mathbf{y}$.

Projective basis

Projective basis: A **projective basis** for $\mathcal{P}^n$ is any set of $n+2$ points no $n+1$ of which are linearly dependent.

Canonical basis:

$$\underbrace{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots}_{\text{points at infinity along each axis}} \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}}_{\text{origin}}, \underbrace{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{\text{unit point}}$$

Projective basis

Change of basis: Let $\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_{n+1}}, \mathbf{e_{n+2}}$ be the standard basis and $\mathbf{a_1}, \mathbf{a_2}, \ldots, \mathbf{a_{n+1}}, \mathbf{a_{n+2}}$ be any other basis. There exists a non-singular transformation $[\mathbf{T}]_{(n+1)\times(n+1)}$ such that:

$$\mathbf{T}\mathbf{e_i} = \lambda_i \mathbf{a_i}, \forall i = 1, 2. \ldots, n+2$$

$\mathbf{T}$ is unique up to a scale.

Homography

The invertible transformation $\mathbf{T} : \mathcal{P}^n \to \mathcal{P}^n$ is called a **projective transformation** or **collineation** or **homography** or **perspectivity** and is completely determined by $n+2$ point correspondences.

- Preserves straight lines and cross ratios
- Given four collinear points $\mathbf{A}_1$, $\mathbf{A}_2$,$\mathbf{A}_3$ and $\mathbf{A}_4$, their **cross ratio** is defined as

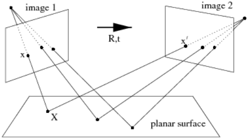
$$\frac{\overline{\mathbf{A}_1\mathbf{A}_3}\;\overline{\mathbf{A}_2\mathbf{A}_4}}{\overline{\mathbf{A}_1\mathbf{A}_4}\;\overline{\mathbf{A}_2\mathbf{A}_3}}$$

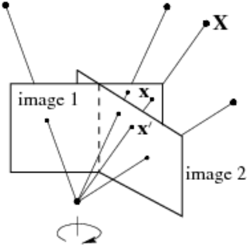
- If $\mathbf{A}_4$ is a point at infinity then the cross ratio is given as

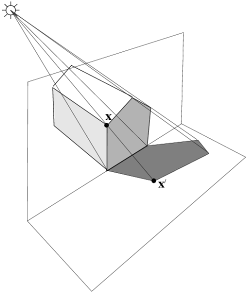
$$\frac{\overline{\mathbf{A}_1\mathbf{A}_3}}{\overline{\mathbf{A}_2\mathbf{A}_3}}$$

- The cross ratio is independent of the choice of the projective coordinate system.

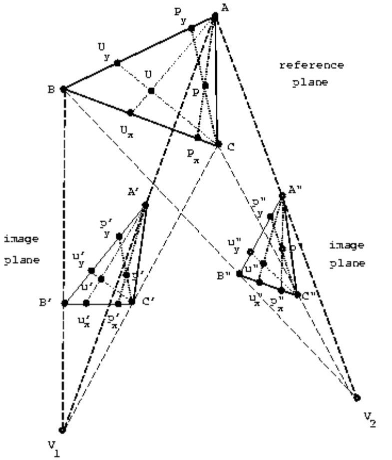
Homography







Homography



Projective mappings of lines

- If the points $\mathbf{x}_i$ lie on the line $\mathbf{l}$, we have $\mathbf{l}^T\mathbf{x}_i = 0$.
- Since, $\mathbf{l}^T\mathbf{H}^{-1}\mathbf{H}\mathbf{x}_i = 0$ the points $\mathbf{H}\mathbf{x}_i$ all lie on the line $\mathbf{H}^{-T}\mathbf{l}$.
- Hence, if points are transformed as $\mathbf{x'}_i = \mathbf{H}\mathbf{x}_i$, lines are transformed as $\mathbf{l}' = \mathbf{H}^{-T}\mathbf{l}$.

Projective mappings of conics

- Note that a conic is represented (homogeneously) as

$$\mathbf{x}^T \mathbf{C} \mathbf{x} = 0$$

- Under a point transformation $\mathbf{x}' = \mathbf{H}\mathbf{x}$ the conic becomes

$$\mathbf{x}^T \mathbf{C} \mathbf{x} = \mathbf{x}'^T [\mathbf{H}^{-1}]^T \mathbf{C} \mathbf{H}^{-1} \mathbf{x}' = \mathbf{x}'^T \mathbf{H}^{-T} \mathbf{C} \mathbf{H}^{-1} \mathbf{x}' = 0$$

- This is the quadratic form of $\mathbf{x}'^T \mathbf{C}' \mathbf{x}'$ with $\mathbf{C}' = \mathbf{H}^{-T} \mathbf{C} \mathbf{H}^{-1}$. This gives the transformation rule for a conic.

The affine subgroup

In an affine space $\mathcal{A}^n$ an **affine transformation** defines a correspondence $\mathbf{X} \leftrightarrow \mathbf{X}'$ given by:

$$\mathbf{X}' = \mathbf{A}\mathbf{X} + \mathbf{b}$$

where $\mathbf{X}$, $\mathbf{X}'$ and $\mathbf{b}$ are n-vectors, and $\mathbf{A}$ is an $n \times n$ matrix. Clearly this is a subgroup of the projective group. Its projective representation is

$$\mathbf{T} = \begin{bmatrix} \mathbf{C} & \mathbf{c} \\ \mathbf{0}_n^T & t_{33} \end{bmatrix}$$

where $\mathbf{A} = \frac{1}{t_{33}}\mathbf{C}$ and $\mathbf{b} = \frac{1}{t_{33}}\mathbf{c}$.

Affine transformations preserve the plane/line at infinity

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{0}^t & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} = \begin{pmatrix} \mathbf{A} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ 0 \end{pmatrix}$$

A general projective transformation moves points at infinity to finite points.

The Euclidean subgroup

- The absolute conic: The conic $\mathbf{\Omega}_\infty$ is intersection of the quadric of equation:

$$\sum_{i=1}^{n+1} x_i^2 = x_{n+1} = 0 \text{ with } \pi_\infty$$

- In a metric frame $\pi_\infty = (0,0,0,1)^T$, and points on $\mathbf{\Omega}_\infty$ satisfy

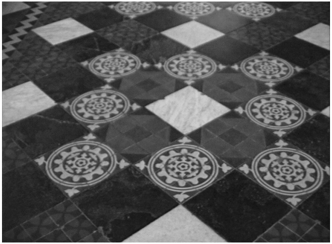
$$\left.\begin{array}{r} X_1^2 + X_2^2 + X_3^3 \\ X_4 \end{array}\right\} = 0$$

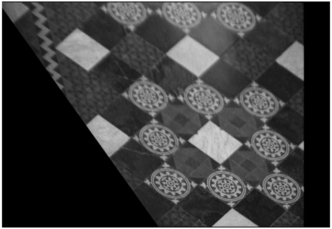
- For directions on π_∞ (with $X_4 = 0$), the absolute conic $\mathbf{\Omega}_\infty$ can be expressed as

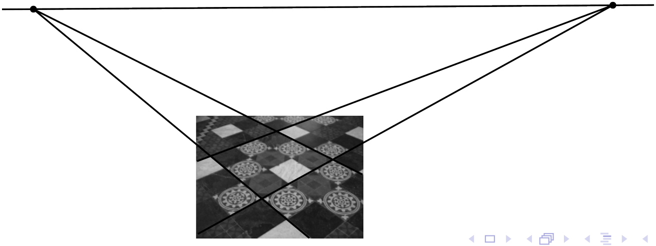
$$(X_1, X_2, X_3)\mathbf{I}(X_1, X_2, X_3)^T = 0$$

- **The absolute conic, $\mathbf{\Omega}_\infty$, is fixed under a projective transformation H if and only if H is an Euclidean transformation.**

Affine calibration of a plane







Affine calibration of a plane

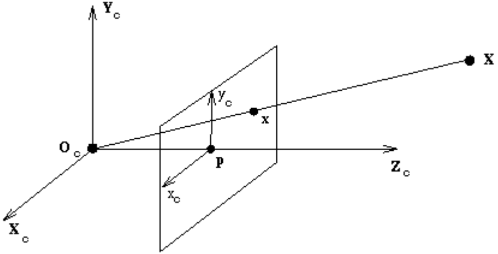
If the imaged line at infinity is $\mathbf{l} = (l_1, l_2, l_3)^t$, then provided $l_3 \neq 0$ a suitable projective transformation that maps $\mathbf{l}$ back to $\mathbf{l}_\infty$ is

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_1 & l_2 & l_3 \end{bmatrix} \mathbf{H_A}$$

Euclidean calibration of a plane

How?

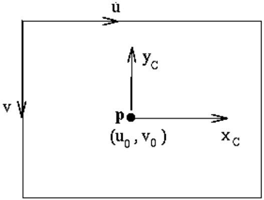
Pin-hole camera revisited



$$\begin{bmatrix} x_c \\ y_c \\ f \end{bmatrix} = k \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} x_c \\ y_c \\ f \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}$$

$$k = f/Z_c$$

Intrinsic/internal parameters

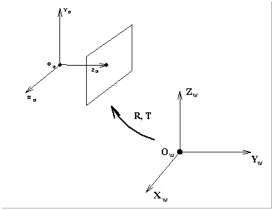


$$\begin{aligned} k_u x_c &= u - u_0 \\ k_v y_c &= v_0 - v \end{aligned}$$

or

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} fk_u & 0 & u_0 \\ 0 & -fk_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ f \end{bmatrix} = \mathbf{K} \begin{bmatrix} x_c \\ y_c \\ f \end{bmatrix}$$

External parameters



$$\mathbf{X_c} = \mathbf{R}\mathbf{X_w} + \mathbf{T}$$

or

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0}_3^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

Calibrated camera

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{K}\left[\mathbf{R} \mid \mathbf{T}\right] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

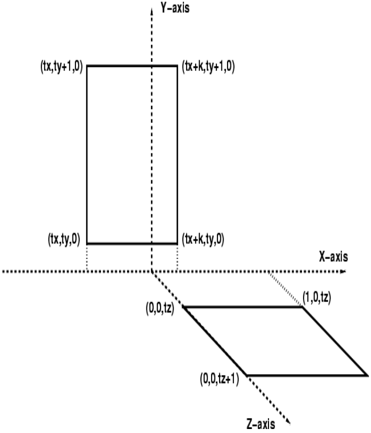
Image to world correspondences of six points in general position gives camera calibration.

What does nternal calibration give?

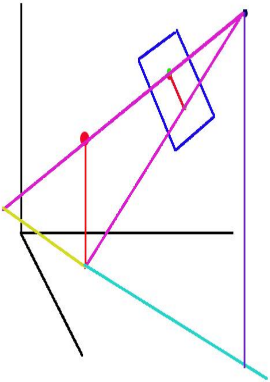
- An image point $\mathbf{x}$ back projects to a ray defined by $\mathbf{x}$ and the camera center. Calibration relates the image point to the ray's direction.
- Suppose points on the ray are written as $\tilde{\mathbf{X}} = \lambda \mathbf{d}$ in the camera Euclidean frame. Then these points map to the point $\mathbf{x} = \mathbf{K}\,[\mathbf{I} \mid \mathbf{0}]\,(\lambda \mathbf{d}^T, ?)^T$
- Thus, $\mathbf{K}$ is the (affine) transformation between $\mathbf{x}$ and the ray's direction $\mathbf{d} = \mathbf{K}^{-1}\mathbf{x}$ measured in the cameras Euclidean frame.
- The angle between two rays $\mathbf{d}_1$ and $\mathbf{d}_2$ corresponding to image points $\mathbf{x}_1$ and $\mathbf{x}_2$ may be obtained as (by the cosine formula)

$$\cos\theta = \frac{\mathbf{d}_1{}^T \mathbf{d}_2}{\sqrt{\mathbf{d}_1{}^T \mathbf{d}_1}\sqrt{\mathbf{d}_2{}^T \mathbf{d}_2}} = \frac{\mathbf{x}_1{}^T(\mathbf{K}^{-T}\mathbf{K}^{-1})\mathbf{x}_1}{\sqrt{\mathbf{x}_1{}^T(\mathbf{K}^{-T}\mathbf{K}^{-1})\mathbf{x}_1}\sqrt{\mathbf{x}_2{}^T(\mathbf{K}^{-T}\mathbf{K}^{-1})\mathbf{x}_2}}$$

Reconstruction

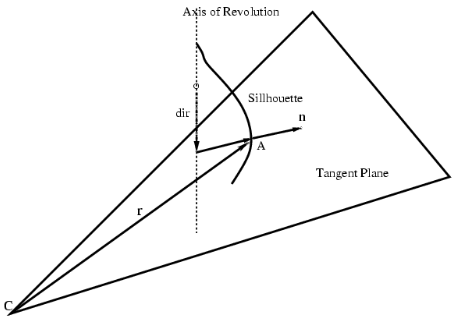


Camera recovery



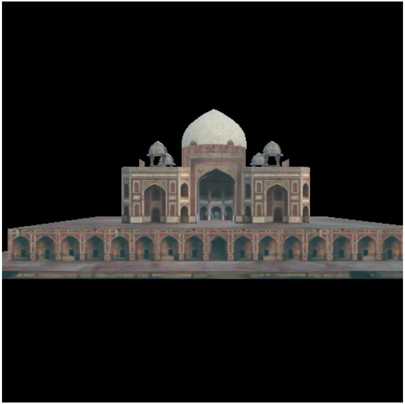
Metrology

Surfaces of revolution

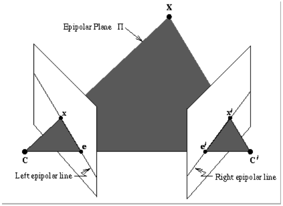


Modeling of structured scenes

A walkthrough



Multiple views: epipolar geometry



epipole: The epipole is the image in one camera of the optical center of the other camera.

epipolar plane: is the plane defined by a 3D point and the optical centers.

epipolar line: is the line of intersection of the epipolar plane with the image plane.

Epipolar constraint

- Given the two cameras 1 and 2, we have the camera equations:
$$\mathbf{x_1} = \tilde{\mathbf{P}}_1\mathbf{X} \text{ and } \mathbf{x_2} = \tilde{\mathbf{P}}_2\mathbf{X}$$
- The **optical center** projects as
$$\tilde{\mathbf{P}}_i\mathbf{X} = \mathbf{0}$$
- Writing
$$\tilde{\mathbf{P}}_i = [\mathbf{P}_i \mid -\mathbf{P}_i\mathbf{t}_i]$$
where $\mathbf{P_i}$ is 3×3 non-singular we have that $\mathbf{t_i}$ is the optical center.
$$[\mathbf{P}_i \mid -\mathbf{P}_i\mathbf{t}_i]\begin{bmatrix} \mathbf{t}_i \\ 1 \end{bmatrix} = \mathbf{0}$$

Epipolar constraint

- The epipole $\mathbf{e_2}$ in the second image is the projection of the optical center of the first image:

$$\mathbf{e_2} = \tilde{\mathbf{P_2}} \begin{bmatrix} \mathbf{t_1} \\ 1 \end{bmatrix}$$

- The projection of point on infinity along the optical ray $< \mathbf{t_1}, \mathbf{x_1} >$ on to the second image is given by:

$$\mathbf{x_2} = \mathbf{P}_2 \mathbf{P_1}^{-1} \mathbf{x_1}$$

- The epipolar line $< \mathbf{e_2}, \mathbf{x_2} >$ is given by the cross product $\mathbf{e_2} \times \mathbf{x_2}$.
- If $[\mathbf{e_2}]_\times$ is the 3×3 antisymmetric matrix representing cross product with $\mathbf{e_2}$, then we have that the epipolar line is given by

$$[e_2]_\times \mathbf{P}_2 \mathbf{P_1}^{-1} \mathbf{x_1} = \mathbf{F} \mathbf{x_1}$$

Epipolar constraint

- Any point $\mathbf{x_2}$ on this epipolar line satisfies

$$\mathbf{x_2}^T \mathbf{F} \mathbf{x_1} = 0$$

- **F** is called the **fundamental matrix**. It is of rank 2 and has 7 *DOF*. Can be computed from 8 (7) point correspondences.
- Clearly $\mathbf{F}\mathbf{e_1} = \mathbf{0}$ (degenerate epipolar line) and $\mathbf{e_2}^T \mathbf{F} = \mathbf{0}$. **The epipoles are obtained as the null spaces of F.**

Epipolar Geometry: calibrated case

Consider the two image scenario where projection matrices are

$$
\begin{aligned}
P &= K\,[I|\mathbf{0}] \quad P' = K'\,[\boldsymbol{R}|\boldsymbol{t}] \\
\Rightarrow P^{+} &= \begin{bmatrix} K^{-1} \\ \mathbf{0}^{T} \end{bmatrix} \quad \boldsymbol{C} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix} \\
\Rightarrow F &= \left[P'\boldsymbol{C}\right]_{\times} P' P^{+} \\
&= \left[K'\boldsymbol{t}\right]_{\times} K'\boldsymbol{R}K^{-1} = \underline{K'^{-T}[\boldsymbol{t}]_{\times}\boldsymbol{R}K^{-1}} \\
&= K'^{-T}\boldsymbol{R}[\boldsymbol{R}\boldsymbol{t}]_{\times}K^{-1} = K'^{-T}\boldsymbol{R}K^{T}\left[K\boldsymbol{R}^{T}\boldsymbol{t}\right]_{\times}
\end{aligned}
$$

Epipolar Geometry: calibrated case

Where are the epipoles ?

$$\begin{aligned} \boldsymbol{e} &= P \begin{pmatrix} -\boldsymbol{R}^T \boldsymbol{t} \\ 1 \end{pmatrix} = K \boldsymbol{R}^T \boldsymbol{t} \\ \boldsymbol{e}' &= P' \begin{pmatrix} \boldsymbol{0} \\ 1 \end{pmatrix} = K' \boldsymbol{t} \end{aligned}$$

Also, the most useful form of $\boldsymbol{F}$ is

$$\boldsymbol{F} = K'^{-T} [\boldsymbol{t}]_\times \boldsymbol{R} K^{-1}$$

Epipolar Geometry: calibrated case

Consider the representation of the fundamental matrix

$$\begin{aligned} \boldsymbol{F} &= K'^{-T}[\boldsymbol{t}]_\times \boldsymbol{R} K^{-1} \\ \Rightarrow \boldsymbol{x}'^T \boldsymbol{F} \boldsymbol{x} &= 0 \\ \Rightarrow \boldsymbol{x}'^T K'^{-T}[\boldsymbol{t}]_\times \boldsymbol{R} K^{-1} \boldsymbol{x} &= 0 \end{aligned}$$

This can be interpreted as

$$\underbrace{\boldsymbol{x}'^T K'^{-T}}[\boldsymbol{t}]_\times \boldsymbol{R} \overbrace{K^{-1}\boldsymbol{x}} = 0$$

- Terms denoted are the calibrated image points
- Central term is the *essential matrix*

Epipolar Geometry: calibrated case

Consider camera pairs $P = [\boldsymbol{I}|\boldsymbol{0}]$ and $P' = [\boldsymbol{R}|\boldsymbol{t}]$

Essential matrix is given by

$$\boldsymbol{E} = [\boldsymbol{t}]_\times \boldsymbol{R} = \boldsymbol{R}\left[\boldsymbol{R}^T\boldsymbol{t}\right]_\times$$

$$\hat{\boldsymbol{x}}'^T \boldsymbol{E} \hat{\boldsymbol{x}} = 0$$

As shown earlier, we have

$$\boldsymbol{E} = \boldsymbol{K}'^T \boldsymbol{F} \boldsymbol{K}$$

Essential Matrix properties

- Has fewer degrees of freedom than the fundamental matrix
- Defined by rotation and translation, i.e. six degrees of freedom
- However there's an overall scale ambiguity
- The *essential matrix* has 5 degrees of freedom
- Implies a translation scale ambiguity
- We can only solve for heading direction and not actual translation

Epipolar Geometry: Essential Matrix

A 3×3 matrix is an essential matrix if and only if two if its singular values are equal, and the third is zero

Essential Matrix properties

- Rank-2 constraint implies third singular value is zero
- Canonical form is singular values of $(1, 1, 0)$

Epipolar Geometry: Essential Matrix

Proof Consider the decomposition $\boldsymbol{E} = [\boldsymbol{t}]_\times \boldsymbol{R} = \boldsymbol{S}\boldsymbol{R}$

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

- W is orthogonal
- Z is skew-symmetric
- Skew-symmetric $\boldsymbol{S}$ can always be written as $\boldsymbol{S} = k\boldsymbol{U}Z\boldsymbol{U}^T$
- $\boldsymbol{U}$ is orthogonal
- Z = diag(1,1,0)W upto scale
- Implies $\boldsymbol{S} = \boldsymbol{U}diag(1,1,0)W\boldsymbol{U}^T$
- $\boldsymbol{E} = \boldsymbol{S}\boldsymbol{R} = \boldsymbol{U}diag(1,1,0)W\boldsymbol{U}^T\boldsymbol{R}$
- Above is a singular value decomposition of $\boldsymbol{E}$ (Q.E.D.)

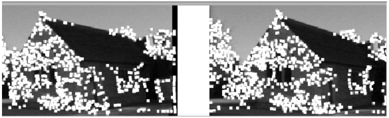
Decomposing Essential Matrix

- SVD of $\boldsymbol{E}$ is $\boldsymbol{U}\, diag(1, 1, 0)\, \boldsymbol{V}^T$
- Ignoring signs, there are two factorisations
  - $\boldsymbol{S} = \boldsymbol{U} Z \boldsymbol{U}^T$ and $\boldsymbol{R} = \boldsymbol{U} W \boldsymbol{V}^T$
  - $\boldsymbol{S} = \boldsymbol{U} Z \boldsymbol{U}^T$ and $\boldsymbol{R} = \boldsymbol{U} W^T \boldsymbol{V}^T$
- $\boldsymbol{t}$ is (upto scale) the last column of $\boldsymbol{U}$
- Sign of $\boldsymbol{t}$ is ambiguous
- Results in 4 possible decomposition pairs
- Need to verify *depth positivity* of one point to disambiguate
- *Depth positivity* will give an unambiguous solution

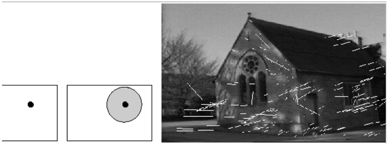
Epipolar geometry: example

Epipolar geometry: computation

- Given corners:



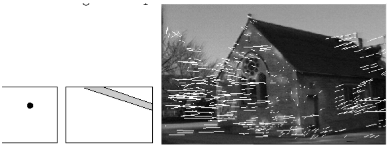
- **Unguided matching:** Obtain a small number of seed matches using cross-correlation and pessimistic thresholds.



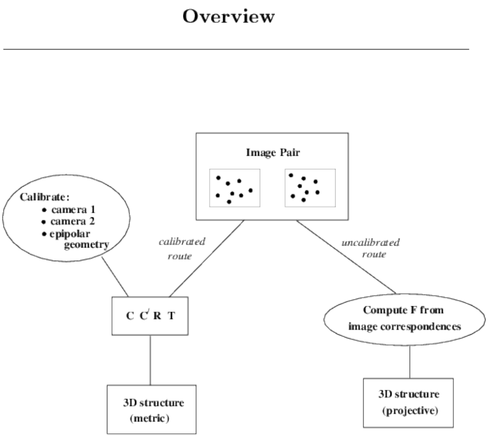
Epipolar geometry: computation

- Compute epipolar geometry and reject outliers.

- **Guided matching:** Search for matches in a band about the epipolar lines.



Structure computation: overview

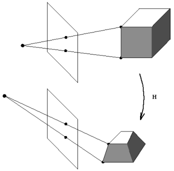


Structure computation

1. Compute the fundamental matrix $\mathbf{F}$ from $\mathbf{x_i} \leftrightarrow \mathbf{x_i'}$.
2. Decompose $\mathbf{F}$ as $\mathbf{F} = \mathbf{e'}_{\times}\mathbf{M'}$.
3. Compute 3D points $\mathbf{X_i}$ by intersecting back-projected rays using

$$\mathbf{P} = [\mathbf{I} \mid \mathbf{0}] \text{ and } \mathbf{P'} = \left[\mathbf{M'} \mid \mathbf{e'}\right]$$

Projective ambiguity



- Suppose $\mathbf{P}$ and $\mathbf{P}'$ are two matrices consistent with $\mathbf{F}$, then

$$\mathbf{x} = \mathbf{PX} \text{ and } \mathbf{x}' = \mathbf{P}'\mathbf{X}'$$

- But, if $\mathbf{H}$ is any arbitrary homography of 3-space, then

$$\mathbf{x} = (\mathbf{PH}^{-1})(\mathbf{HX}) \text{ and } \mathbf{x}' = (\mathbf{P}'\mathbf{H}^{-1})(\mathbf{HX}')$$

- Thus, we can only recover projective structure modulo an unknown homography.
- How to find the homography that will upgrade the projective structure to Euclidean?

Auto-calibration: Kruppa's equations

- *SVD* of the fundamental matrix gives

$$F = \underbrace{U \begin{pmatrix} r & & \\ & s & \\ & & 1 \end{pmatrix}}_{A'^T} \underbrace{\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}}_{E} \underbrace{\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} V^T}_{A}$$

- Setting $\hat{u} = Au$ and $\hat{u}' = A'u'$, we get $\hat{u}'^T E \hat{u} = 0$.
- E is a very special fundamental matrix with the properties
  1. The two epipoles are at origin.
  2. Corresponding epipolar lines are identical in the two images.

Auto-calibration: Kruppa's equations

- Consider a plane passing through the two camera centers, tangent to the absolute conic.
- Such a plane will project to a pair of corresponding epipolar lines in the two images, and these two lines will be tangent to the IAC.
- Let $(\lambda, \mu, 0)^T$ be a tangent to the IAC. Since D is the DIAC in the first image, this tangential relationship may be written as $(\lambda, \mu, 0)D(\lambda, \mu, 0)^T = 0$.
- Similarly $(\lambda, \mu, 0)D'(\lambda, \mu, 0)^T = 0$.
- Enforcing same solutions from the two equations we obtain:

$$\frac{d_{11}}{d'_{11}} = \frac{d_{12}}{d'_{12}} = \frac{d_{22}}{d'_{22}}$$

Auto-calibration: Kruppa's equations

- Writing $D = ACA^T$ and $D' = A'CA'^T$, we obtain

$$\frac{a_1^T C a_1}{a'^T_1 C a'_1} = \frac{a_1^T C a_2}{a'^T_1 C a'_2} = \frac{a_2^T C a_2}{a'^T_2 C a'_2}$$

- Define $\mathbf{x} = (x, y, z)^T$. $(\mathbf{x}, 0)$ is on the absolute conic *if and only if* $x^2 + y^2 + z^2 = 0$ or $\mathbf{x}\mathbf{x}^T = 0$.
- $\mathbf{u} = P(x, y, z, 0)^T = K(R| - R\mathbf{t})(x, y, z, 0)^T = KR\mathbf{x}$. Thus $\mathbf{x} = R^T K^{-1}\mathbf{u}$.
- $\mathbf{u}$ is on IAC if $\mathbf{x}\mathbf{x}^T = \mathbf{u}^T K^{-T} R R^T K^{-1}\mathbf{u} = \mathbf{u}^T K^{-T} K^{-1}\mathbf{u} = 0$.
- $\mathbf{u}$ must lie on conic represented by $K^{-T}K^{-1}$. Equivalently, the DIAC $KK^T = C$.
- K can be compute by Cholesky decomposition.

Sparse bundle adjustment

Simultaneously refine

- 3D coordinates describing the scene geometry
- Parameters of the relative motion.
- Optical characteristics of the camera.

from corresponding image projections of all points. The process minimizes the reprojection error defined by

$$\min_{\mathbf{a_j},\mathbf{b_i}} \sum_{i=1}^{n} \sum_{j=1}^{m} v_{ij} d(\mathbf{P}(\mathbf{a}_j, \mathbf{b}_i), \mathbf{x}_{ij})^2 \qquad (0)$$

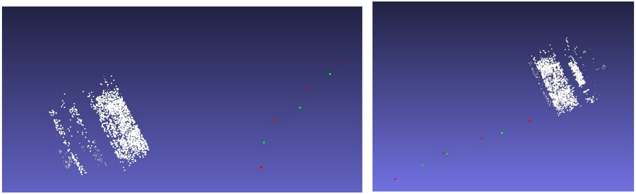
where, $v_{ij} = 1$ if point i is visible in image j.
$\mathbf{P}$ is the projection matrix, d is the euclidean distance, $\mathbf{x_{ij}}$ is projection of $\mathbf{a_j^{th}}$ point on image $\mathbf{b_i}$

Sparse bundle adjustment

1. Initially choose two cameras having most inliers using fundamental matrix
2. Obtain structure and motion by minimization using *LM*.
3. Remove the points and correspondences with large reprojection errors
4. Remove points which violate the cheirality condition
5. Add images that are more consistent with the already recovered cameras.
6. Apply DLT to initialise the new camera internal and external calibration.
7. Go to 2

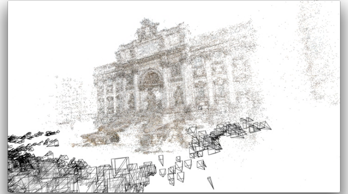
Sparse bundle adjustment

Sparse bundle adjustment



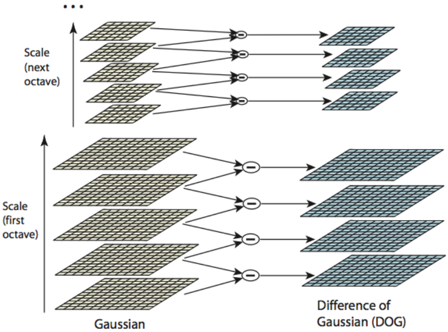
Build Rome in a day?

Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz and Richard Szeliski, 2009

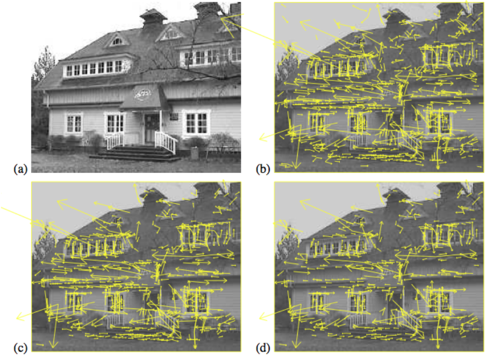


SIFT

Features are scale-space maximas in DOGs.

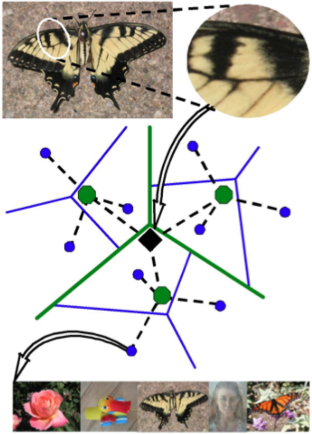


SIFT:example

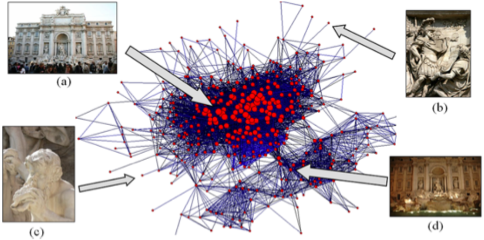


SIFT: matching

Vocabulary tree



Vocabulary tree match graph



Divide and conquer: Efficient large scale SFM using graph paritioning

We use multiway **Normalised Cut** (Shi and Malik, 2000):

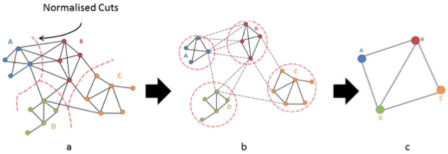
$$Ncut(A, B) = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(A, B)}{assoc(B, V)}$$

where

$$cut(A, B) = \sum_{u \in A, v \in B} w(u, v)$$

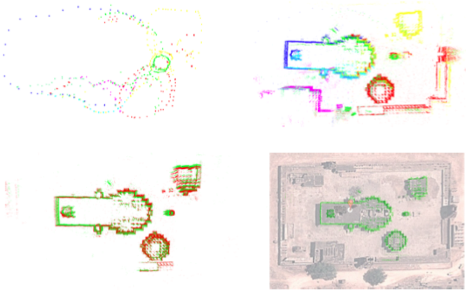
$$assoc(A, V) = \sum_{u \in A, t \in V} w(u, t)$$

Component-wise reconstruction and registration

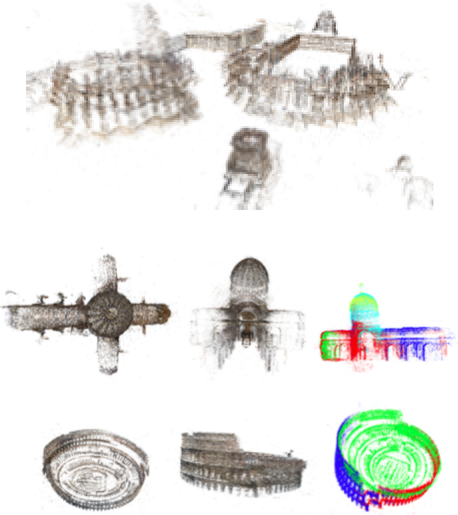


- Incremental bundle adjustment for reconstruction of each component.
- Estimation of epipolar geometry, rotation and translation of cur edges.
- Global motion averaging and registration.

Segmentation and reconstruction example



Reconstruction examples

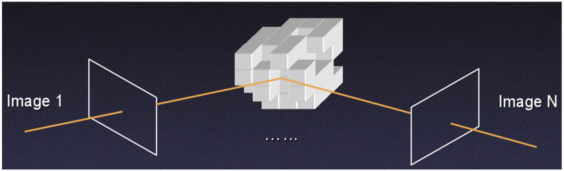


One order of magnitude speed-up

Data set	Match graph creation using vocabulary tree (mins)	Pairwise matching (mins)	Reconstruction and registration (mins)	Total Time by us (mins)	Pairwise matching by VSFM (mins)	Reconstruction by VSFM (mins)	Total time by VSFM (mins)
Rome	768	502	**27**	**1297**	N/A	N/A	N/A
Hampi	481	424	**8**	**913**	9522	59	9581
St Peter's Basilica	98	22	**4**	**124**	1385	10	1395
Colosseum	83	24	**3**	**110**	1394	9	1403

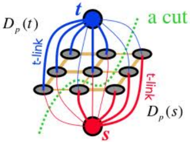
Dense Reconstruction

- Initialize voxel space using the sparse reconstruction
- Voxel labeling:
  - label *s* indicates voxel on surface front,
  - label *a* indicates voxel between camera and surface front,
  - label *b* indicates voxel is behind surface front.



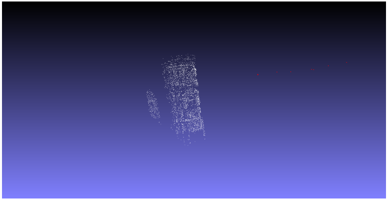
Dense Reconstruction

- Graph-cut formulation using photo consistency measure.
- α expansion based optimization for multi-label graph-cut.

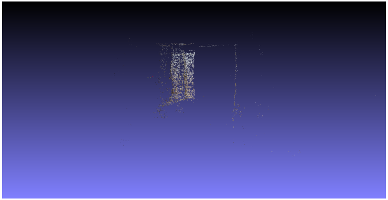


Dense reconstruction: Qutab pillar

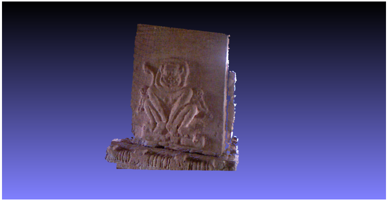


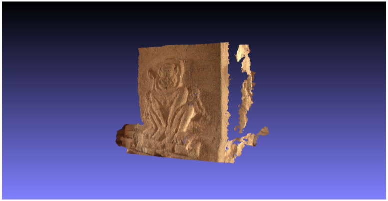




Dense reconstruction: Hampi pillar



Kinect super-resolution





Occlusion is a challenge