Computer Vision

18AI742

Dr. rer. nat. D. Antony Louis Piriyakumar,
Dean (R&D),
Cambridge Institute of Technology.

BRIDGE TO THE FUTURE

CAMBRIDGE
INSTITUTE OF TECHNOLOGY

Structure from Motion

Dr. D. Antony Louis Piriyakumar,
Dean (Research & Development)
Registered Indian patent agent (IN/PA 3041)

www.cambridge.edu.in

Structure from Motion

Contents

1) Internally calibrated perspective cameras
2) Uncalibrated weak-perspective cameras
3) Uncalibrated perspective cameras
4) Conclusion
5) Q&A

$$\nabla \cdot \mathbf{E} = \rho / \varepsilon_0$$

$$\nabla \cdot \mathbf{B} = 0$$

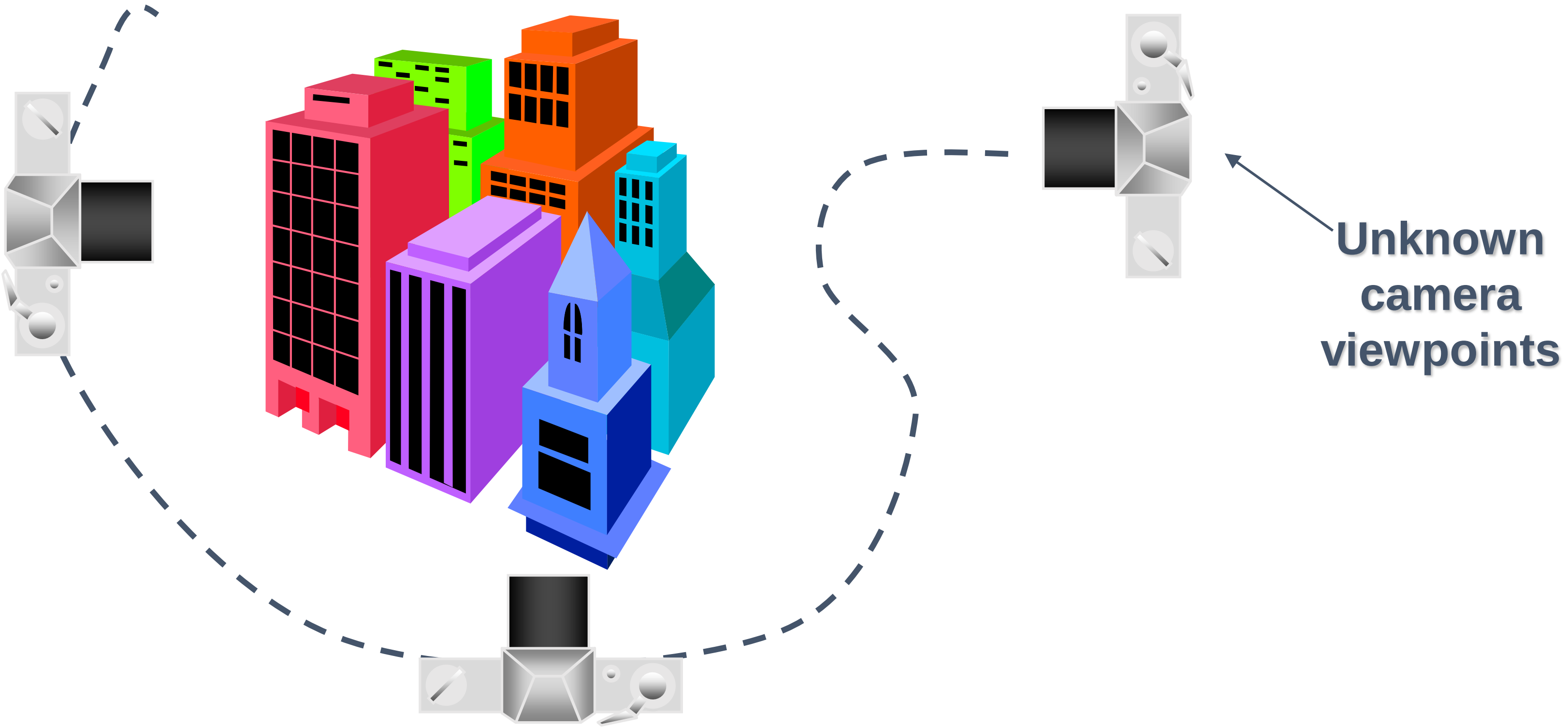
$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

$$\nabla \times \mathbf{B} = \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \mu_0 \boldsymbol{j}_c$$

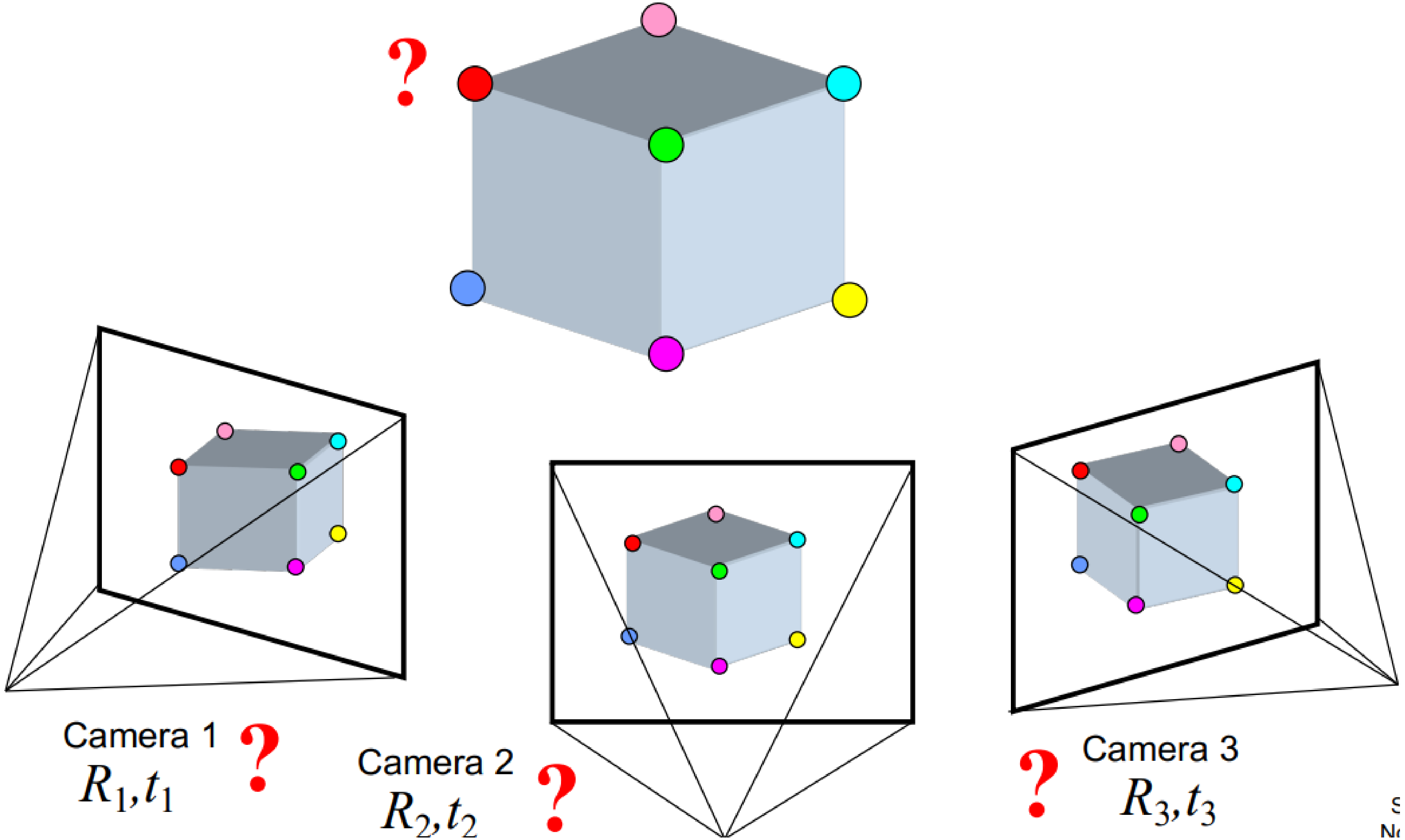
where

$$\nabla = \hat{\mathbf{i}} \frac{\partial}{\partial x} + \hat{\mathbf{j}} \frac{\partial}{\partial y} + \hat{\mathbf{k}} \frac{\partial}{\partial z}$$

Structure from motion



Compute the camera parameters and the 3D point coordinates



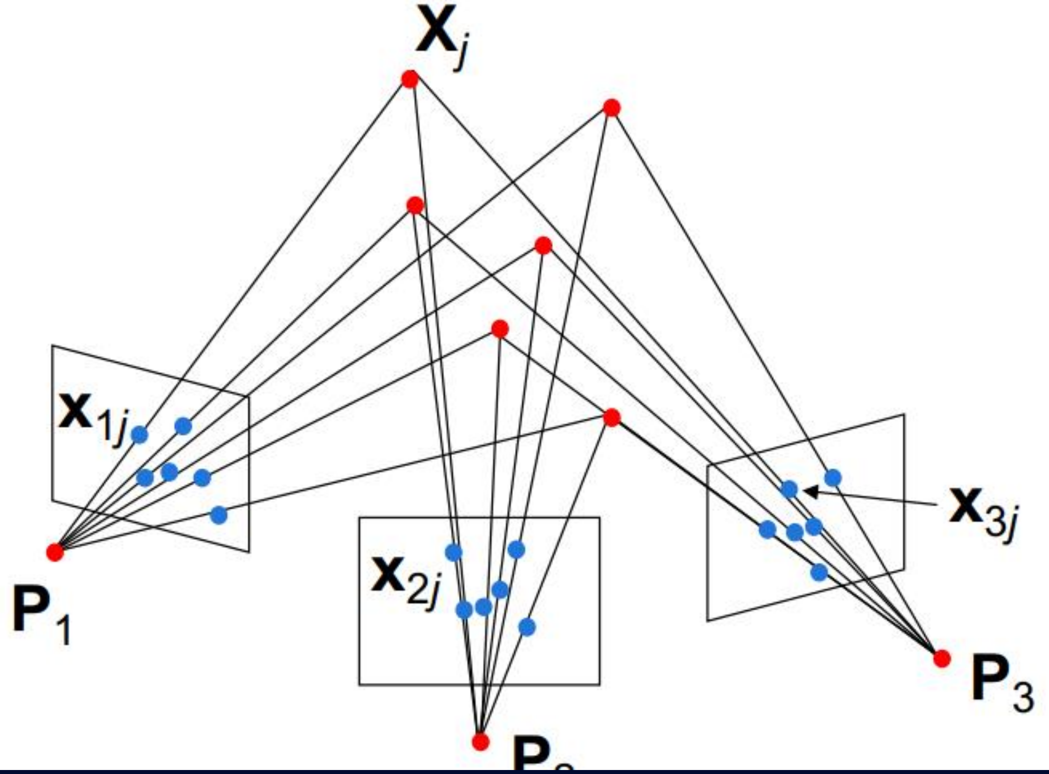
CAMBRIDGE
INSTITUTE OF TECHNOLOGY

Structure from motion

Given: m images of n fixed 3D points

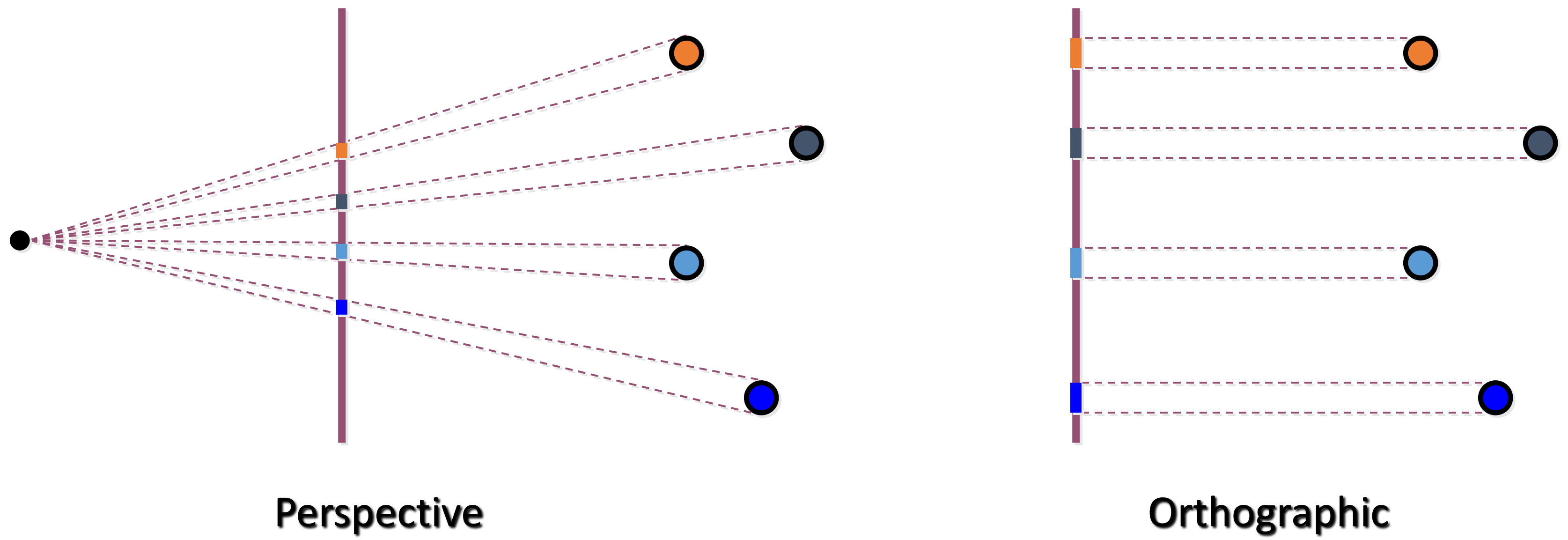
$$\lambda_{ij}\mathbf{x}_{ij} = \mathbf{P}_i \mathbf{X}_j, \quad i = 1, \dots, m, \quad j = 1, \dots, n$$

Problem: estimate m projection matrices $\mathbf{P}_i$ and n 3D points $\mathbf{X}_j$ from the mn correspondences $\mathbf{x}_{ij}$

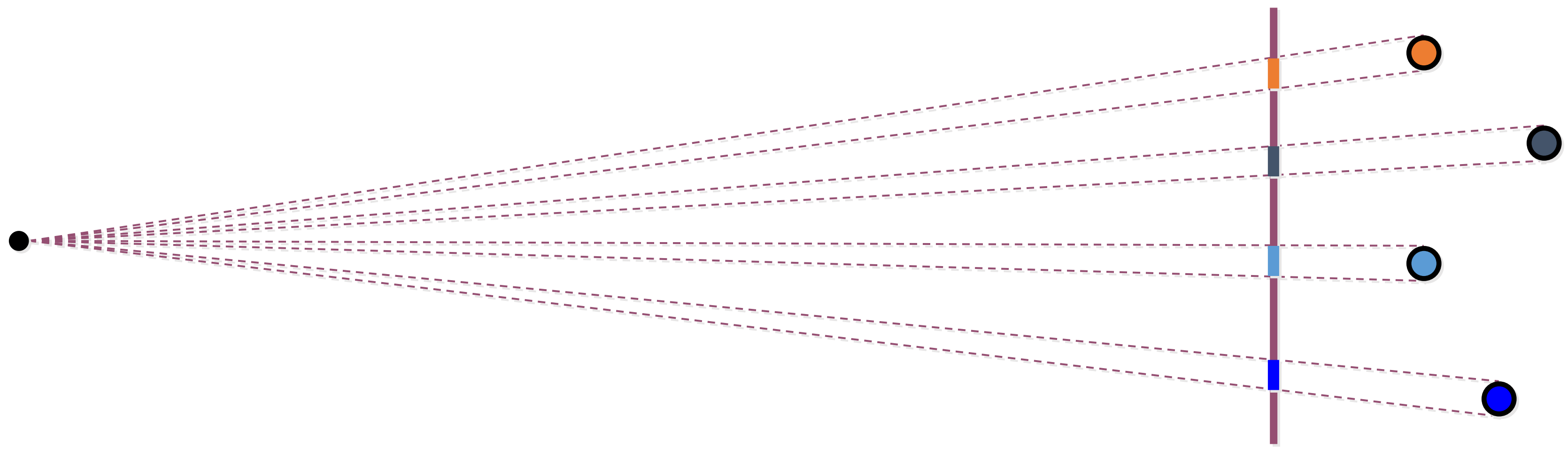


Obama’s head in 3D

Camera approximated by orthographic projection

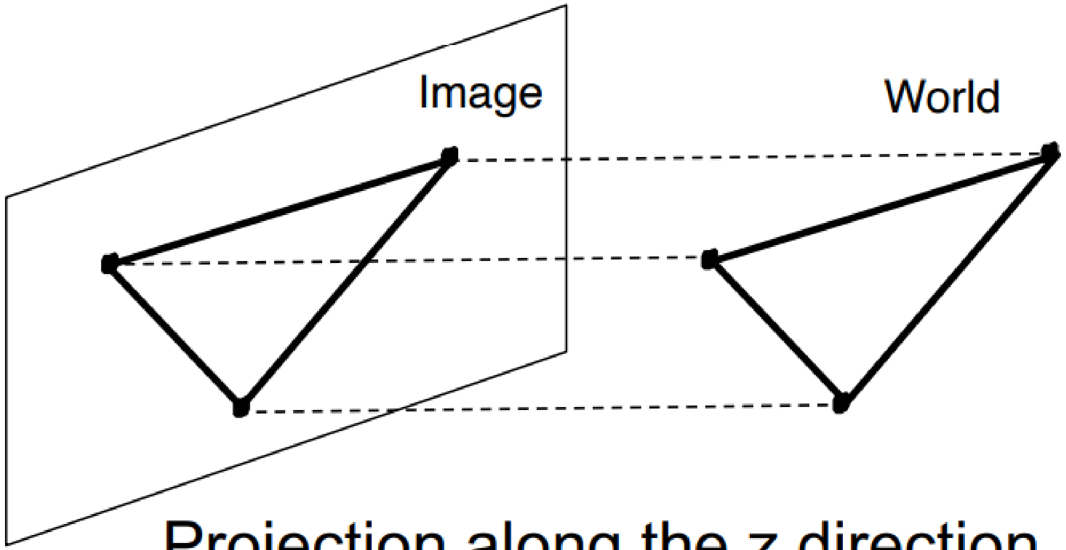


An orthographic assumption is sometimes well approximated by a telephoto lens



Weak Perspective

Orthographic projection



Projection along the z direction

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \Rightarrow (x, y)$$

SFM under orthographic projection

$$\underset{2\times 1}{\mathbf{u}} = \underset{2\times 3}{\mathbf{\Pi}}\ \underset{3\times 1}{\mathbf{X}} + \underset{2\times 1}{\mathbf{t}}$$

image point | **projection matrix** | **scene point** | **image offset**

More generally: weak perspective, para-perspective, affine

Trick

Choose scene origin to be centroid of 3D points

Choose image origins to be centroid of 2D points

Allows us to drop the camera translation:

$$\underset{2\times 1}{\mathbf{u}} = \underset{2\times 3}{\mathbf{\Pi}}\ \underset{3\times 1}{\mathbf{X}}$$

Shape by factorization

projection of *n* features in one image:

$$\underset{2\times n}{\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n\end{bmatrix}} = \underset{2\times 3}{\prod} \underset{3\times n}{\begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_n\end{bmatrix}}$$

projection of *n* features in *f* images

$$\underset{2f\times n}{\begin{bmatrix}\mathbf{u}_1^1 & \mathbf{u}_2^1 & \cdots & \mathbf{u}_n^1\\ \mathbf{u}_1^2 & \mathbf{u}_2^2 & \cdots & \mathbf{u}_n^2\\ \vdots & \vdots & \ddots & \vdots\\ \mathbf{u}_1^f & \mathbf{u}_2^f & \cdots & \mathbf{u}_n^f\end{bmatrix}} = \underset{2f\times 3}{\begin{bmatrix}\mathbf{\Pi}^1\\ \mathbf{\Pi}^2\\ \vdots\\ \mathbf{\Pi}^f\end{bmatrix}} \underset{3\times n}{\begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_n\end{bmatrix}}$$

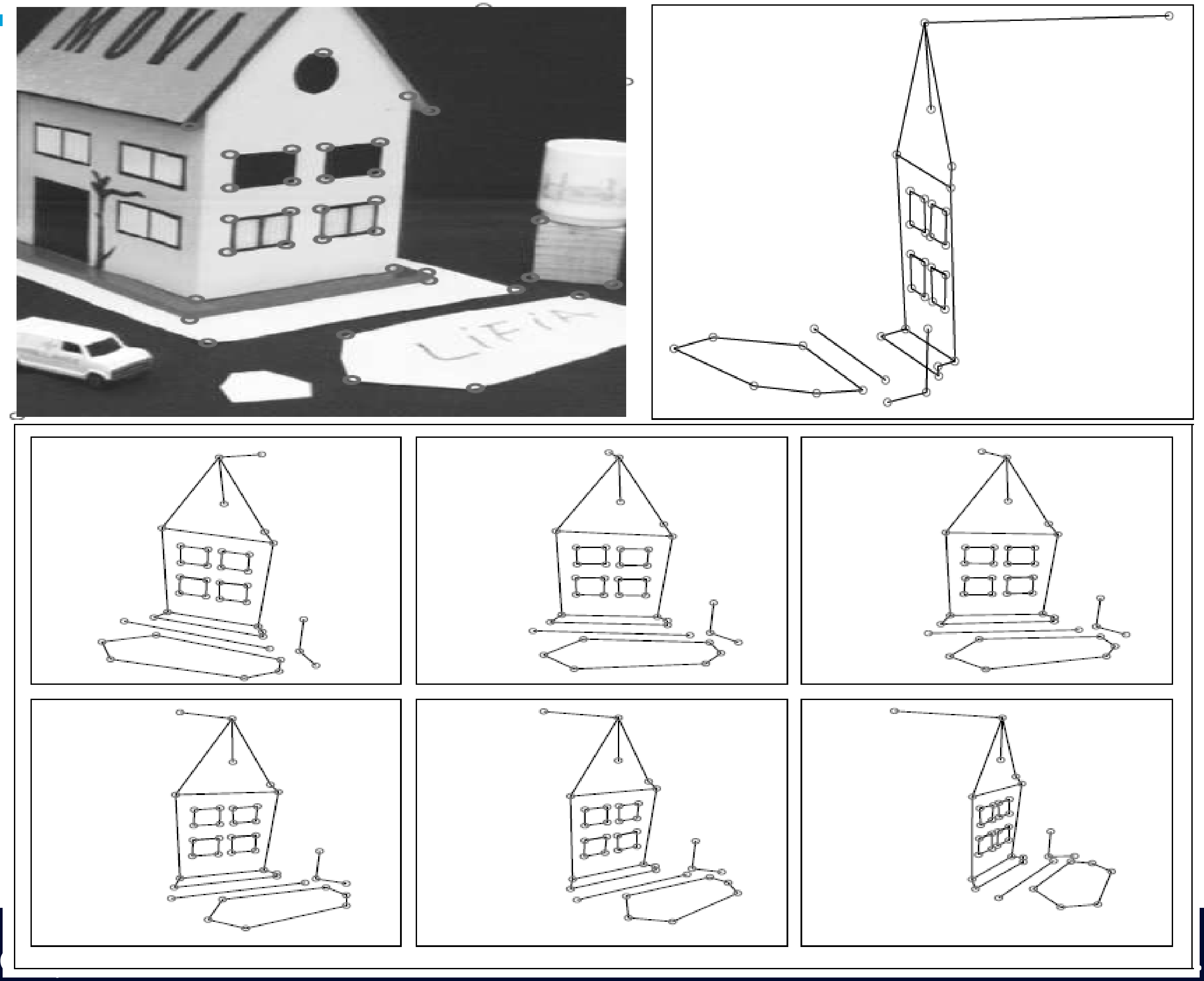
W measurement **M** motion **S** shape

Key Observation: *rank*(**W**) <= 3

Structure from motion

- estimating the three-dimensional shape of a scene from multiple pictures.
- stereopsis, the cameras are normally calibrated (intrinsic parameters are known)
- Extrinsic have been determined relative to some fixed world coordinate system.
- simplifies the reconstruction process and explains the emphasis put on binocular

A “wireframe” display of the corresponding ground-truth 3D points, observed from some arbitrary viewpoint



CAMBRIDGE
INSTITUTE OF TECHNOLOGY

Internally calibrated perspective cameras

- m pinhole perspective cameras with known intrinsic parameters
- observing a scene that consists of n fixed points Pj (j = 1, . . . , n).
- work in normalized image coordinates
- assume that correspondences have been established between the m images

Internally calibrated perspective cameras

- mn homogeneous coordinate vectors pij = ˆpij = (xij, yij , 1)T (i = 1, . . .,m)
- projections of the points Pj are known.
- Ri and ti are respectively the rotation matrix and the translation vector
- representing position and orientation of camera number i in a fixed coordinate system,
- Pj is the nonhomogeneous coordinate vector of the point Pj in thatc oordinate system,
- Zij is the depth of that point relative to camera number i.

CAMBRIDGE
INSTITUTE OF TECHNOLOGY

Euclidean structure from motion

$$p_{ij} = \frac{1}{Z_{ij}} \begin{pmatrix} \mathcal{R}_i & t_i \end{pmatrix} \begin{pmatrix} P_j \\ 1 \end{pmatrix},$$

the problem of estimating the n vectors Pj,
together with the m rotation matrices Ri and
translation vectors ti,
from the mn image correspondences pij

CAMBRIDGE
INSTITUTE OF TECHNOLOGY

Natural Ambiguity of the problem

$$p_{ij} = \frac{1}{Z_{ij}} (\mathcal{R}_i \quad t_i) \begin{pmatrix} P_j \\ 1 \end{pmatrix},$$

$$p_{ij} = \frac{1}{Z_{ij}} \left((\mathcal{R}_i \quad t_i) \begin{pmatrix} \mathcal{R} & t \\ 0^T & 1 \end{pmatrix} \right) \left(\begin{pmatrix} \mathcal{R}^T & -\mathcal{R}^T t \\ 0^T & 1 \end{pmatrix} \begin{pmatrix} P_j \\ 1 \end{pmatrix} \right) = \frac{1}{Z_{ij}} (\mathcal{R}'_i \quad t'_i) \begin{pmatrix} P'_j \\ 1 \end{pmatrix},$$

where $\mathcal{R}'_i = \mathcal{R}_i \mathcal{R}$, $t'_i = \mathcal{R}_i t + t_i$, and $P'_j = \mathcal{R}^T (P_j - t)$.

$$p_{ij} = \frac{1}{\lambda Z_{ij}} (\mathcal{R}_i \quad \lambda t_i) \begin{pmatrix} \lambda P_j \\ 1 \end{pmatrix} = \frac{1}{Z'_{ij}} (\mathcal{R}_i \quad t'_i) \begin{pmatrix} P'_j \\ 1 \end{pmatrix},$$

ti' = λ**t**$_i$,
P'$_j$ = λ**P**$_j$,
Z'ij = λZij.

Recovery of the Euclidean shape of the observed scene, along with the corresponding perspective projection matrices

- 2mn constraints on
- the 6m extrinsic parameters of the matrices Mi
- 3n parameters of the vectors Pj,
- Admits a finite number of solutions as soon as $2mn \geq 6m + 3n - 7$.
- For m = 2, five point correspondences should thus be sufficient to determine

An approximate solution can be found by minimizing the mean-squared error

$$E = \frac{1}{mn} \sum_{i,j} \left|\left| p_{ij} - \frac{1}{Z_{ij}} (\mathcal{R}_i \quad t_i) \begin{pmatrix} P_j \\ 1 \end{pmatrix} \right|\right|^2$$

require a reasonable initial guess to converge
something close to the global minimum of the error function

Euclidean Structure and Motion from Two Images (weak calibration)

Let us start with the uncalibrated case. The epipolar constraint can be written

$$p^T \mathcal{F} p' = [u, v, 1] \begin{pmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{pmatrix} \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = 0. \quad (8.3)$$

$$\mathcal{U} = \begin{pmatrix} x_1 x_1' & x_1 y_1' & x_1 & y_1 x_1' & y_1 y_1' & y_1 & x_1' & y_1' & 1 \\ x_2 x_2' & x_2 y_2' & x_2 & y_2 x_2' & y_2 y_2' & y_2 & x_2' & y_2' & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_n x_n' & x_n y_n' & x_n & y_n x_n' & y_n y_n' & y_n & x_8' & y_n' & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{f} = \begin{pmatrix} F_{11} \\ F_{12} \\ F_{13} \\ F_{21} \\ F_{22} \\ F_{23} \\ F_{31} \\ F_{32} \\ F_{33} \end{pmatrix}.$$

Solving this equation in the least-squares sense amounts to minimizing

$$E = \frac{1}{n} ||\mathcal{U} \boldsymbol{f}||^2 = \frac{1}{n} \sum_{i=1}^{n} (p_i^T \mathcal{F} p_i')^2 \quad (8.4)$$

CAMBRIDGE
INSTITUTE OF TECHNOLOGY

Hartley normalization

- Transform image coordinates using T : pi→ ˜pi and T ′ : p′i→ ˜p′i
- use linear least squares to compute the matrix F˜ minimizing
- $$\frac{1}{n}\sum_{i=1}^{n}(\tilde{p}_i^T \tilde{\mathcal{F}} \tilde{p}_i')^2.$$
- rank-2 matrix F¯ minimizing the Frobenius norm of F˜ − F¯ is simply F¯ = Udiag(r, s, 0)VT
- sets F = T TF¯T ′ as the final estimate of the fundamental matrix.

From Essential Matrix to Camera Motion

- Assume that essential matrix E is known. $E = [t\times]R$.
- Solve inverse problem of recovering R and t from E.
- Because $E^T = V\,\mathrm{diag}(1, 1, 0)U^T$ such that $E^T v = 0$—is Ru3
- because $E^T t = 0$, $t' = u3$ and $t'' = -u3$.

From Essential Matrix to Camera Motion

Let us now show that there are also two solutions for the rotational part of the essential matrix, namely

$$\mathcal{R}' = \mathcal{U}\mathcal{W}\mathcal{V}^T \quad \text{and} \quad \mathcal{R}'' = \mathcal{U}\mathcal{W}^T\mathcal{V}^T, \quad \text{where} \quad \mathcal{W} = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Now let $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ denote the first two columns of $\mathcal{U}$. Because $\boldsymbol{t}' = \boldsymbol{u}_3$ and $\boldsymbol{U}$ is a rotation matrix, we have $\boldsymbol{t}' \times \boldsymbol{u}_1 = \boldsymbol{u}_2$ and $\boldsymbol{t}' \times \boldsymbol{u}_2 = -\boldsymbol{u}_1$. In particular,

$$[t'_\times]\mathcal{R}' = (\boldsymbol{u}_2 \quad -\boldsymbol{u}_1 \quad 0)\mathcal{W}\mathcal{V}^T = -(\boldsymbol{u}_1 \quad \boldsymbol{u}_2 \quad 0)\mathcal{V}^T = -\mathcal{U}\mathrm{diag}(1,1,0)\mathcal{V}^T = -\mathcal{E}.$$

Algorithm 8.1: The Longuet-Higgins Eight-Point Algorithm for Euclidean Structure and Motion from Two Views.

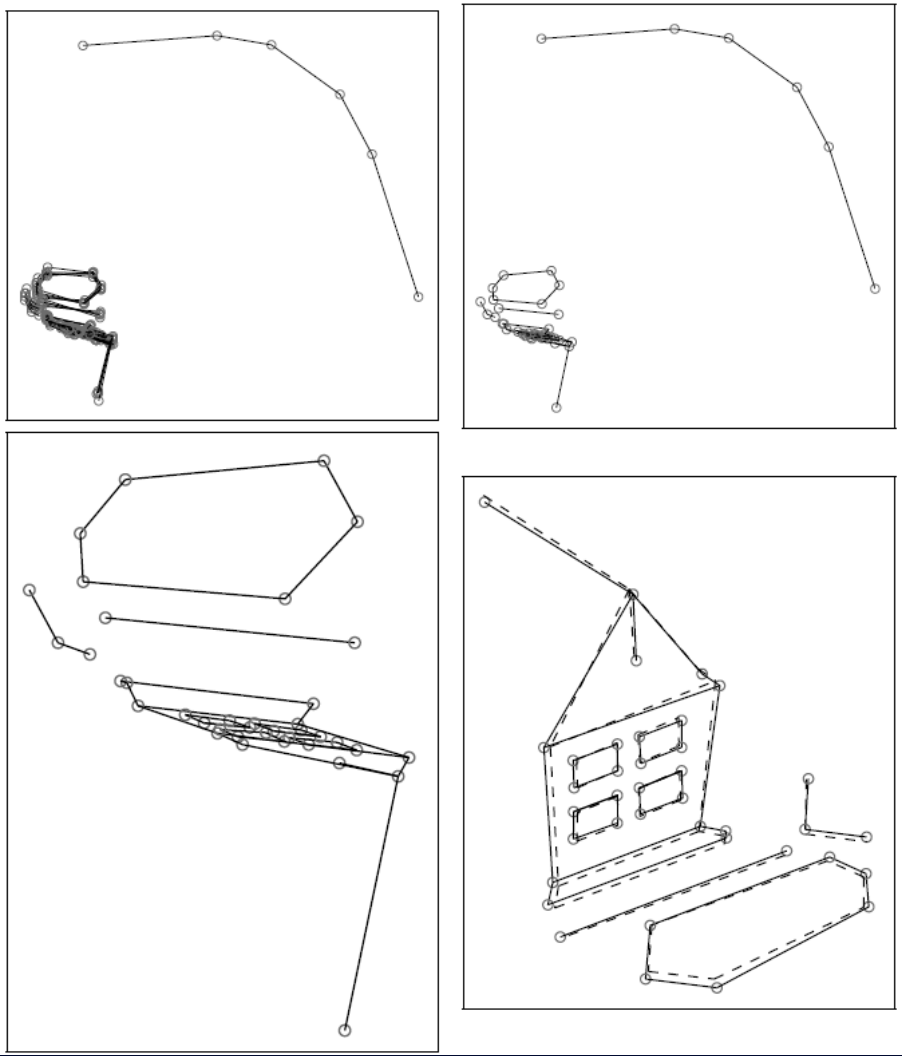
1. **Estimate $\mathcal{F}$.**
   (a) Compute Hartley's normalization transformation $\mathcal{T}$ and $\mathcal{T}'$, and the corresponding points $\tilde{\boldsymbol{p}}_i$ and $\tilde{\boldsymbol{p}}'_i$.
   (b) Use homogeneous linear least squares to estimate the matrix $\tilde{\mathcal{F}}$ minimizing $\frac{1}{n}\sum_{i=1}^{n}(\tilde{\boldsymbol{p}}_i^T \tilde{\mathcal{F}} \tilde{\boldsymbol{p}}'_i)^2$ under the constraint $||\tilde{\mathcal{F}}||_F^2 = 1$.
   (c) Compute the singular value decomposition $\mathcal{U}\text{diag}(r, s, t)\mathcal{V}^T$ of $\tilde{\mathcal{F}}$, and set $\bar{\mathcal{F}} = \mathcal{U}\text{diag}(r, s, 0)\mathcal{V}^T$.
   (d) Output the fundamental matrix $\mathcal{F} = \mathcal{T}^T \bar{\mathcal{F}} \mathcal{T}'$.
2. **Estimate $\mathcal{E}$.**
   (a) Compute the matrix $\tilde{\mathcal{E}} = \mathcal{K}^T \mathcal{F} \mathcal{K}'$.
   (b) Set $\mathcal{E} = \mathcal{U}\,\text{diag}(1, 1, 0)\mathcal{V}^T$, where $\mathcal{U}\mathcal{W}\mathcal{V}^T$ is the singular value decomposition of the matrix $\tilde{\mathcal{E}}$.
3. **Compute $\mathcal{R}$ and $\boldsymbol{t}$.**
   (a) Compute the rotation matrices $\mathcal{R}' = \mathcal{U}\mathcal{W}\mathcal{V}^T$ and $\mathcal{R}'' = \mathcal{U}\mathcal{W}^T\mathcal{V}^T$, and the translation vectors $\boldsymbol{t}' = \boldsymbol{u}_3$ and $\boldsymbol{t}'' = -\boldsymbol{u}_3$, where $\boldsymbol{u}_3$ is the third column of the matrix $\mathcal{U}$.
   (b) Output the combination of the rotation matrices $\mathcal{R}'$, $\mathcal{R}''$, and the translation vectors $\boldsymbol{t}'$, $\boldsymbol{t}''$ such that the reconstructed points lie in front of both cameras.

Euclidean Structure and Motion from Multiple Images

- graph whose nodes correspond to image pairs and
- whose edges link two images that share at least three points.
- Homogeneous coordinate vectors kPj and lPj in the corresponding camera frames
- 3 × 4 similarity transformation Skl separating the coordinate systems

$$\frac{1}{n_{kl}} \sum_{j \in J_{kl}} ||^k P_j - \mathcal{S}_{kl}\, {}^l P_j||^2$$

Euclidean Structure and Motion from Multiple Images



Structure from Motion

Contents

1) Internally calibrated perspective cameras
2) Uncalibrated weak-perspective cameras
3) Uncalibrated perspective cameras
4) Conclusion
5) Q&A

$$\nabla \cdot \mathbf{E} = \rho / \varepsilon_0$$

$$\nabla \cdot \mathbf{B} = 0$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

$$\nabla \times \mathbf{B} = \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \mu_0 \mathbf{j}_c$$

where

$$\nabla = \hat{\mathbf{i}} \frac{\partial}{\partial x} + \hat{\mathbf{j}} \frac{\partial}{\partial y} + \hat{\mathbf{k}} \frac{\partial}{\partial z}$$

Uncalibrated weak-perspective cameras

P_j $(j = 1, \dots, n)$ observed by m affine cameras with unknown intrinsic and extrinsic parameters, and the corresponding *mn nonhomogeneous* coordinate vectors $\boldsymbol{p}_{ij}$ of their images, we can rewrite the corresponding weak-perspective projection equations as

$$\boldsymbol{p}_{ij} = \mathcal{M}_i \begin{pmatrix} \boldsymbol{P}_j \\ 1 \end{pmatrix} = \mathcal{A}_i \boldsymbol{P}_j + \boldsymbol{b}_i \quad \text{for} \quad i = 1, \dots, m \quad \text{and} \quad j = 1, \dots, n, \tag{8.5}$$

where $\mathcal{M}_i = (\mathcal{A}_i \quad \boldsymbol{b}_i)$ is a general rank-2 2×4 matrix, and the vector $\boldsymbol{P}_j$ in $\mathbb{R}^3$ is the position of the point P_j in some fixed coordinate system. We define *affine structure from motion* as the problem of estimating the m matrices $\mathcal{M}_i$ and the n vectors $\boldsymbol{P}_j$ from the mn image correspondences $\boldsymbol{p}_{ij}$.

Uncalibrated weak-perspective cameras - ambiguity

$$\mathcal{M}'_i = \mathcal{M}_i \mathcal{Q}, \quad \begin{pmatrix} \boldsymbol{P}'_j \\ 1 \end{pmatrix} = \mathcal{Q}^{-1} \begin{pmatrix} \boldsymbol{P}_j \\ 1 \end{pmatrix}, \tag{8.6}$$

and $\mathcal{Q}$ is an arbitrary *affine transformation* matrix; that is, it can be written as

$$\mathcal{Q} = \begin{pmatrix} \mathcal{C} & \boldsymbol{d} \\ \boldsymbol{0}^T & 1 \end{pmatrix} \quad \text{with} \quad \mathcal{Q}^{-1} = \begin{pmatrix} \mathcal{C}^{-1} & -\mathcal{C}^{-1}\boldsymbol{d} \\ \boldsymbol{0}^T & 1 \end{pmatrix}, \tag{8.7}$$

where C is a nonsingular 3×3 matrix and d is a vector in R3.

Affine Epipolar Geometry

$$\begin{cases} p = \mathcal{A}P + b \\ p' = \mathcal{A}'P + b' \end{cases} \quad \text{as} \quad \begin{pmatrix} \mathcal{A} & p - b \\ \mathcal{A}' & p' - b' \end{pmatrix} \begin{pmatrix} P \\ -1 \end{pmatrix} = 0.$$

$$\operatorname{Det} \begin{pmatrix} \mathcal{A} & p - b \\ \mathcal{A}' & p' - b' \end{pmatrix} = 0,$$

$$\alpha x + \beta y + \alpha' x' + \beta' y' + \delta = 0,$$

$$\alpha' x' + \beta' y' + \gamma' = 0, \text{ where } \gamma' = \alpha x + \beta y + \delta$$

The affine epipolar constraint can be rewritten in the familiar form

$$(x, y, 1)\mathcal{F} \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = 0, \quad \text{where} \quad \mathcal{F} = \begin{pmatrix} 0 & 0 & \alpha \\ 0 & 0 & \beta \\ \alpha' & \beta' & \delta \end{pmatrix} \qquad (8.9)$$

is the *affine fundamental matrix*. This suggests that the affine epipolar geometry

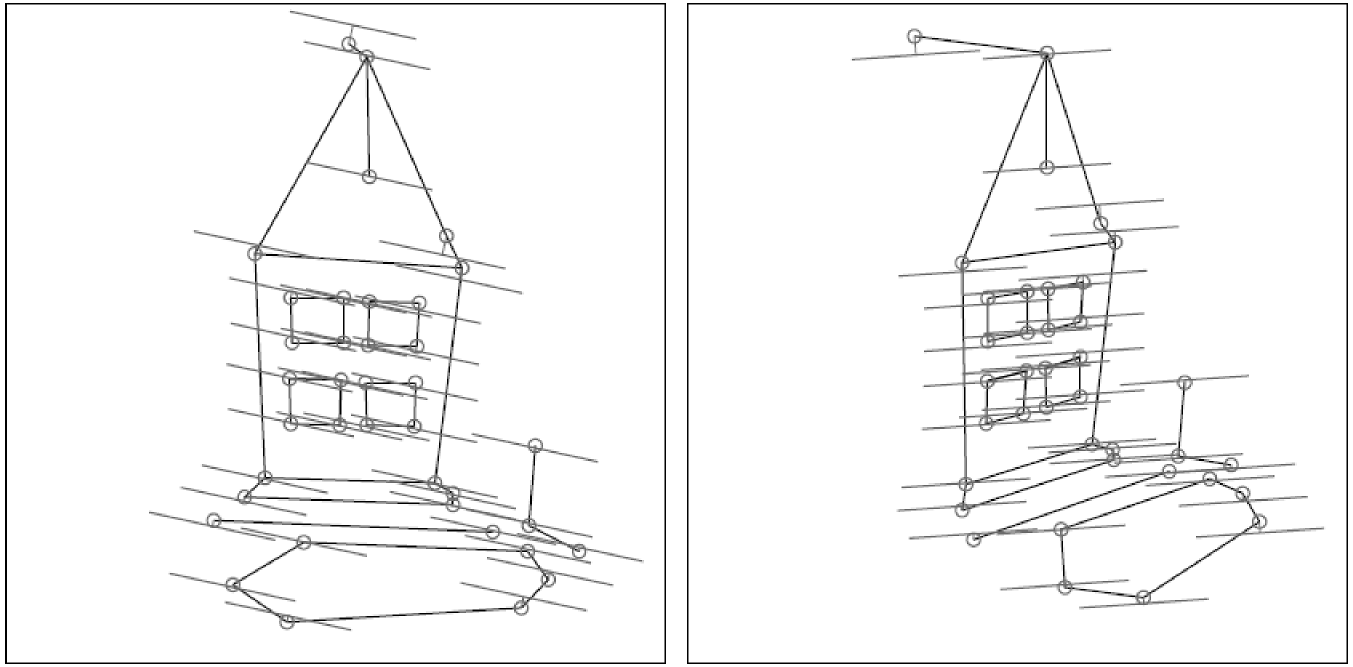
where α, β, α′, β′, and δ are constants depending on A, b, A′, and b′.
This is the affine epipolar constraint

Affine Weak Calibration Uf = 0

$$\mathcal{U} = \begin{pmatrix} x_1 & y_1 & x_1' & y_1' & 1 \\ x_2 & y_2 & x_2' & y_2' & 1 \\ \dots & \dots & \dots & \dots & \dots \\ x_n & y_n & x_n' & y_n' & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{f} = \begin{pmatrix} \alpha \\ \beta \\ \alpha' \\ \beta' \\ \delta \end{pmatrix}.$$

solving this equation in the least-squares sense amounts to computing the eigenvector f associated with the smallest eigenvalue of $U^T U$.

Affine weak-calibration experiment using two images of the house sequence and linear least squares, together with Hartley’s normalization



From the Affine Fundamental Matrix to Camera Motion.

calculations: according to Equations (8.6) and (8.7), if $\mathcal{M} = (\mathcal{A} \quad b)$ and $\mathcal{M}' = (\mathcal{A}' \quad b')$ are solutions of our problem, so are $\tilde{\mathcal{M}} = \mathcal{M}\mathcal{Q}$ and $\tilde{\mathcal{M}}' = \mathcal{M}'\mathcal{Q}$, where

$$\mathcal{Q} = \begin{pmatrix} \mathcal{C} & d \\ 0^T & 1 \end{pmatrix}$$

is an arbitrary affine transformation. The new projection matrices can be written as $\tilde{\mathcal{M}} = (\mathcal{A}\mathcal{C} \quad \mathcal{A}d + b)$ and $\tilde{\mathcal{M}}' = (\mathcal{A}'\mathcal{C} \quad \mathcal{A}'d + b')$. Note that, according to Equation (8.7), applying this transformation to the projection matrices amounts to applying the inverse transformation to every scene point P, whose position $\boldsymbol{P}$ is replaced by $\tilde{\boldsymbol{P}} = \mathcal{C}^{-1}(\boldsymbol{P} - \boldsymbol{d})$.

From the Affine Fundamental Matrix to Camera Motion.

$\mathcal{C}$ and $\boldsymbol{d}$ so that the two projection matrices take the canonical forms:

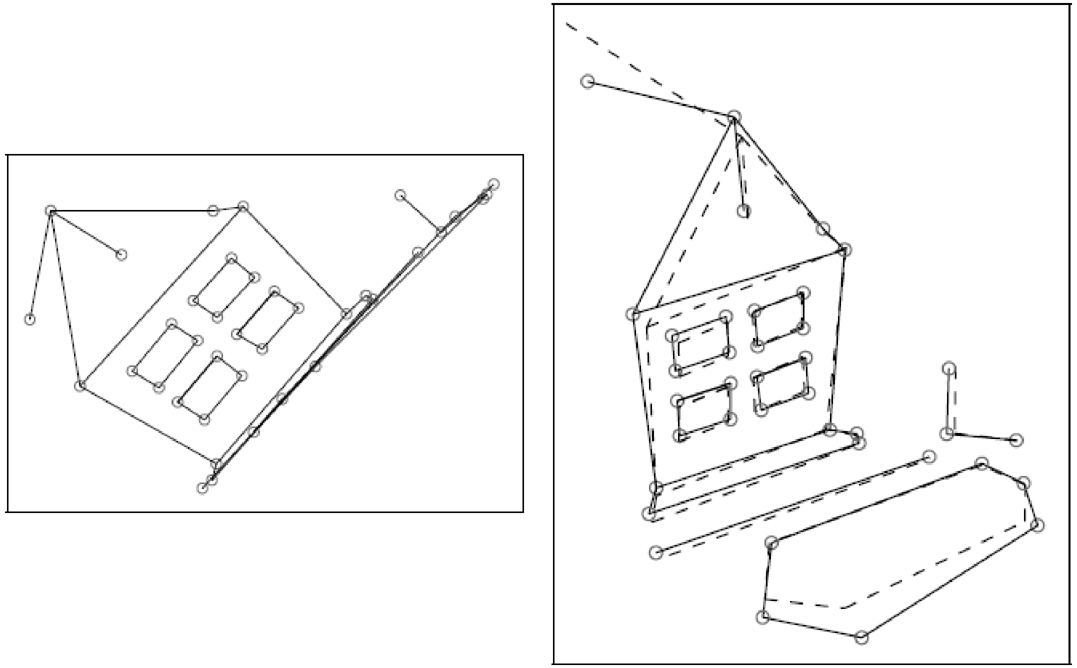
$$\tilde{\mathcal{M}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \tilde{\mathcal{M}}' = \begin{pmatrix} 0 & 0 & 1 & 0 \\ a & b & c & d \end{pmatrix},$$

which allows us to rewrite the epipolar constraint as

$$\text{Det} \begin{pmatrix} 1 & 0 & 0 & x \\ 0 & 1 & 0 & y \\ 0 & 0 & 1 & x' \\ a & b & c & y'-d \end{pmatrix} = -ax - by - cx' + y' - d = 0,$$

$$\begin{pmatrix} 1 & 0 & 0 & x \\ 0 & 1 & 0 & y \\ 0 & 0 & 1 & x' \\ a & b & c & y'-d \end{pmatrix} \begin{pmatrix} \tilde{\boldsymbol{P}} \\ -1 \end{pmatrix} = 0,$$

The affine reconstruction of the house from two views.



Affine Structure and Motion from Multiple Images

let P0 denote the center of mass of the n points P1, . . . , Pn, and let pi0 denote its projection into image number i,

$$\boldsymbol{p}_{i0} = \mathcal{A}_i \boldsymbol{P}_0 + \boldsymbol{b}_i, \quad \text{and thus} \quad \boldsymbol{p}_{ij} - \boldsymbol{p}_{i0} = \mathcal{A}_i(\boldsymbol{P}_j - \boldsymbol{P}_0).$$

$$\boldsymbol{p}_{ij} = \mathcal{A}_i \boldsymbol{P}_j \quad \text{for} \quad i = 1, \ldots, m \quad \text{and} \quad j = 1, \ldots, n,$$

$$\mathcal{D} = \mathcal{A}\mathcal{P}, \text{ where } \mathcal{D} = \begin{pmatrix} \boldsymbol{p}_{11} & \ldots & \boldsymbol{p}_{1n} \\ \ldots & \ldots & \ldots \\ \boldsymbol{p}_{m1} & \ldots & \boldsymbol{p}_{mn} \end{pmatrix}, \ \mathcal{A} = \begin{pmatrix} \mathcal{A}_1 \\ \vdots \\ \mathcal{A}_m \end{pmatrix}, \text{ and } \mathcal{P} = \begin{pmatrix} \boldsymbol{P}_1 & \ldots & \boldsymbol{P}_n \end{pmatrix}.$$

$$E = \sum_{i,j} ||\boldsymbol{p}_{ij} - \mathcal{A}_i \boldsymbol{P}_j||^2 = \sum_j ||\boldsymbol{q}_j - \mathcal{A}\boldsymbol{P}_j||^2 = ||\mathcal{D} - \mathcal{A}\mathcal{P}||_F^2,$$

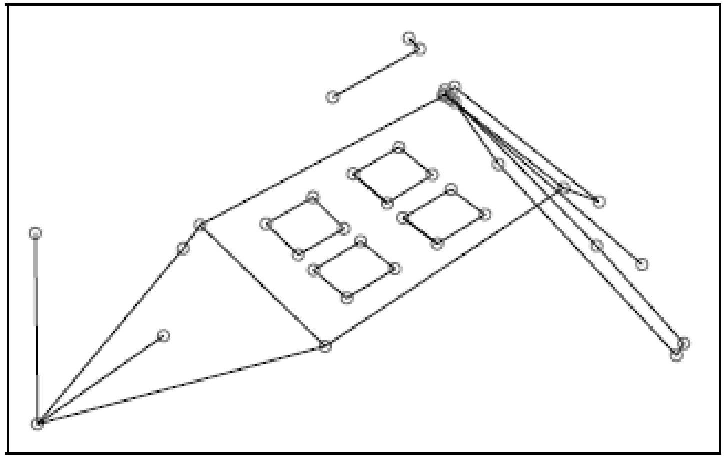
Algorithm 8.2: The Tomasi–Kanade Factorization Algorithm for Affine Shape from Motion.

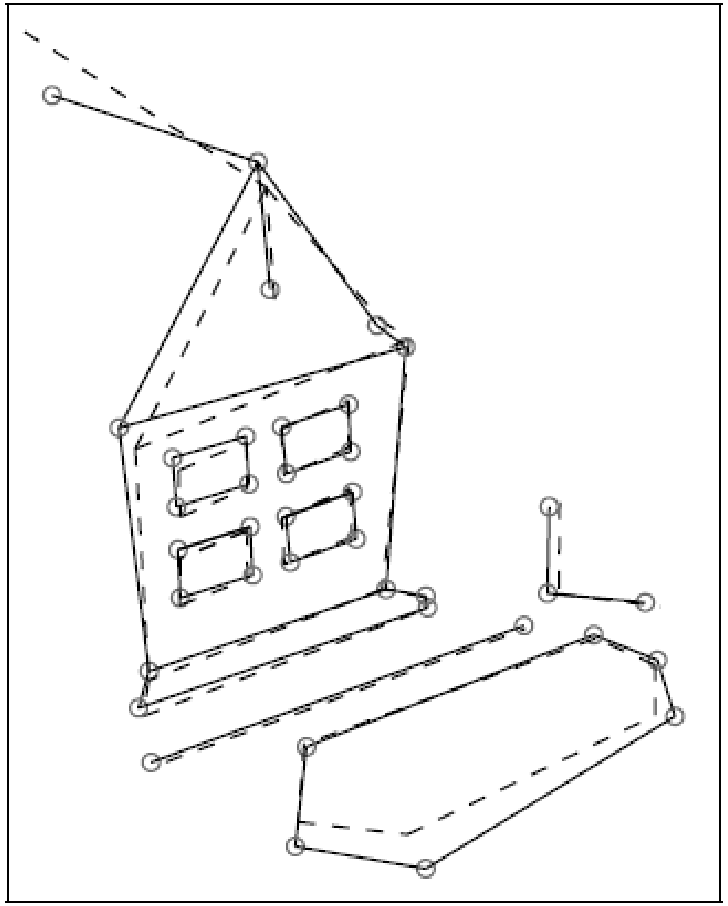
1. Compute the singular value decomposition $\mathcal{D} = \mathcal{U}\mathcal{W}\mathcal{V}^T$.
2. Construct the matrices $\mathcal{U}_3$, $\mathcal{V}_3$, and $\mathcal{W}_3$ formed by the three leftmost columns of the matrices $\mathcal{U}$ and $\mathcal{V}$, and the corresponding 3×3 submatrix of $\mathcal{W}$.
3. Define

$$\mathcal{A}_0 = \mathcal{U}_3\sqrt{\mathcal{W}_3} \quad \text{and} \quad \mathcal{P}_0 = \sqrt{\mathcal{W}_3}\mathcal{V}_3^T;$$

the $2m \times 3$ matrix $\mathcal{A}_0$ is an estimate of the camera motion, and the $3 \times n$ matrix $\mathcal{P}_0$ is an estimate of the scene structure.

The affine reconstruction of the house from multiple views





From Affine to Euclidean Shape

$$\mathcal{M} = \frac{1}{Z_r}\begin{pmatrix} k & s \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \mathcal{R}_2 & t_2 \end{pmatrix},$$

Zr is the depth of the reference point, k and s are aspect-ratio and skew parameters, R2 is the 2×3 matrix formed by the first two rows of a rotation matrix, and t2 is a vector in R2

$$\hat{\mathcal{M}} = \begin{pmatrix} \hat{\mathcal{A}} & \hat{b} \end{pmatrix} = \frac{1}{Z_r}\begin{pmatrix} \mathcal{R}_2 & t_2 \end{pmatrix}. \qquad \hat{a}_1 \cdot \hat{a}_2 = 0 \quad \text{and} \quad ||\hat{a}_1||^2 = ||\hat{a}_2||^2.$$

$$\begin{cases} \hat{a}_{i1} \cdot \hat{a}_{i2} = 0, \\ ||\hat{a}_{i1}||^2 = ||\hat{a}_{i2}||^2, \end{cases} \Longleftrightarrow \begin{cases} a_{i1}^T \mathcal{C}\mathcal{C}^T a_{i2} = 0, \\ a_{i1}^T \mathcal{C}\mathcal{C}^T a_{i1} = a_{i2}^T \mathcal{C}\mathcal{C}^T a_{i2}, \end{cases} \quad \text{for} \quad i = 1, \ldots, m,$$

Structure from Motion

Contents

1) Internally calibrated perspective cameras
2) Uncalibrated weak-perspective cameras
3) Uncalibrated perspective cameras
4) Conclusion
5) Q&A

$$\nabla \cdot \mathbf{E} = \rho / \varepsilon_0$$

$$\nabla \cdot \mathbf{B} = 0$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$$

$$\nabla \times \mathbf{B} = \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \mu_0 \mathbf{j}_c$$

where

$$\nabla = \hat{\mathbf{i}} \frac{\partial}{\partial x} + \hat{\mathbf{j}} \frac{\partial}{\partial y} + \hat{\mathbf{k}} \frac{\partial}{\partial z}$$

Uncalibrated perspective cameras

Given n fixed points Pj (j = 1, . . . , n) observed by m cameras and the corresponding mn homogeneous coordinate vectors pij = (xij, yij , 1)T of their images, the corresponding perspective projection equations as

$$\begin{cases} x_{ij} = \dfrac{\boldsymbol{m}_{i1} \cdot \boldsymbol{P}_j}{\boldsymbol{m}_{i3} \cdot \boldsymbol{P}_j} \\ y_{ij} = \dfrac{\boldsymbol{m}_{i2} \cdot \boldsymbol{P}_j}{\boldsymbol{m}_{i3} \cdot \boldsymbol{P}_j} \end{cases} \quad \text{for} \quad i = 1, \ldots, m \quad \text{and} \quad j = 1, \ldots, n,$$

Projective Structure and Motion from Two Images

$$Zp = \mathcal{A}(\mathrm{Id} \quad 0)\tilde{P} + b = Z'\mathcal{A}p' + b.$$

$$p^T \mathcal{F} p' = 0 \quad \text{where} \quad \mathcal{F} = [b_\times]\mathcal{A}.$$

linear least-squares solution of $\mathcal{F}^T b = 0$ with unit norm, and we pick $\mathcal{A}_0 = -[b_\times]\mathcal{F}$ as the value of $\mathcal{A}$. It is easy to show that, for any vector a, $[a_\times]^2 = aa^T - ||a||^2\mathrm{Id}$, thus:

$$[b_\times]\mathcal{A}_0 = -[b_\times]^2\mathcal{F} = -bb^T\mathcal{F} + ||b||^2\mathcal{F} = \mathcal{F},$$

since $\mathcal{F}^T b = 0$ and $||b||^2 = 1$. This shows that $\tilde{\mathcal{M}} = (\mathcal{A}_0 \quad b)$ is *a* solution of Equation (8.17).[5] As shown in the problems, there is in fact a four-parameter family of solutions whose general form is

$$\tilde{\mathcal{M}} = (\mathcal{A} \quad b) \quad \text{with} \quad \mathcal{A} = \lambda\mathcal{A}_0 + (\ \mu b \mid \nu b \mid \tau b\ ). \tag{8.18}$$

Projective factorization. Given m images of n points,

(8.18) as

$$\mathcal{D} = \mathcal{M}\mathcal{P}, \tag{8.19}$$

where

$$\mathcal{D} = \begin{pmatrix} Z_{11}\boldsymbol{p}_{11} & Z_{12}\boldsymbol{p}_{12} & \dots & Z_{1n}\boldsymbol{p}_{1n} \\ Z_{21}\boldsymbol{p}_{21} & Z_{22}\boldsymbol{p}_{22} & \dots & Z_{2n}\boldsymbol{p}_{2n} \\ \dots & \dots & \dots & \dots \\ Z_{m1}\boldsymbol{p}_{m1} & Z_{m2}\boldsymbol{p}_{m2} & \dots & Z_{mn}\boldsymbol{p}_{mn} \end{pmatrix}, \quad \mathcal{M} = \begin{pmatrix} \mathcal{M}_1 \\ \mathcal{M}_2 \\ \dots \\ \mathcal{M}_m \end{pmatrix} \text{ and } \mathcal{P} = \begin{pmatrix} \boldsymbol{P}_1 & \boldsymbol{P}_2 & \dots & \boldsymbol{P}_n \end{pmatrix},$$

and thus formulate projective structure from motion as the minimization of

$$E = \sum_{i,j} ||Z_{ij}\boldsymbol{p}_j - \mathcal{M}_i\boldsymbol{P}_j||^2 = ||\mathcal{D} - \mathcal{M}\mathcal{P}||_F^2 \tag{8.20}$$

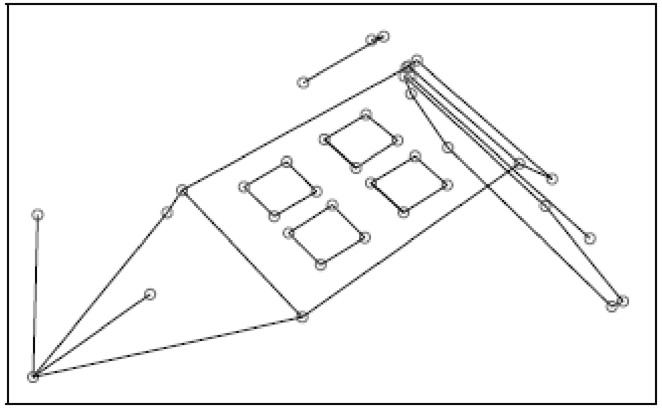
Bundle adjustment

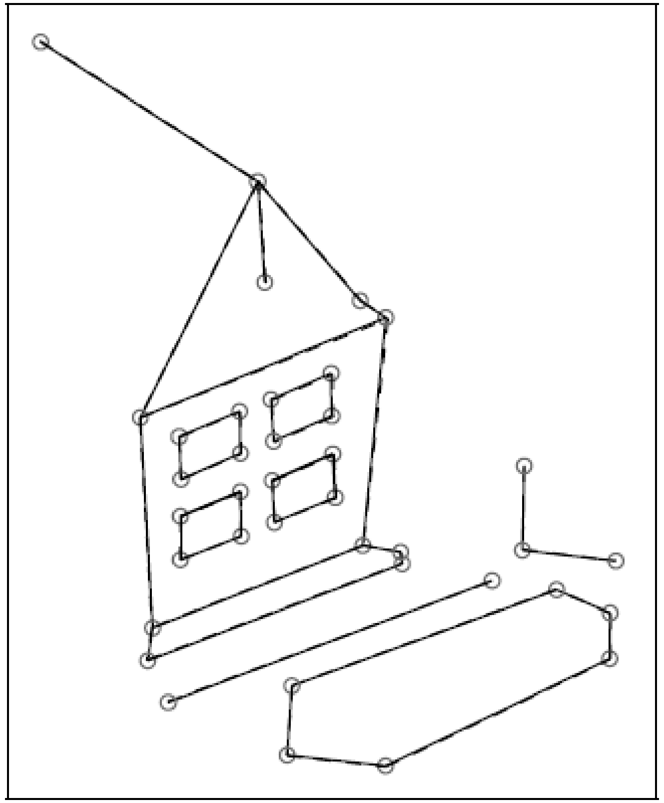
and use nonlinear least squares to minimize directly

$$E = \frac{1}{mn} \sum_{i,j} \left[\left(x_{ij} - \frac{\boldsymbol{m}_{i1} \cdot \boldsymbol{P}_j}{\boldsymbol{m}_{i3} \cdot \boldsymbol{P}_j} \right)^2 + \left(y_{ij} - \frac{\boldsymbol{m}_{i2} \cdot \boldsymbol{P}_j}{\boldsymbol{m}_{i3} \cdot \boldsymbol{P}_j} \right)^2 \right]$$

with respect to the matrices $\mathcal{M}_i$ $(i = 1, \ldots, m)$ and vectors $\boldsymbol{P}_j$ $(j = 1, \ldots, n)$.

Bundle adjustment





From Projective to Euclidean Shape

$$\hat{\mathcal{M}}_i = \rho_i \mathcal{K}_i (\mathcal{R}_i \quad t_i),$$

$$\mathcal{M}_i \mathcal{Q}_3 = \rho_i \mathcal{K}_i \mathcal{R}_i.$$

$\mathcal{Q} = (\mathcal{Q}_3 \quad \boldsymbol{q}_4)$, where $\mathcal{Q}_3$ is a 4×3 matrix and $\boldsymbol{q}_4$ is a vector in $\mathbb{R}^4$,

$$\begin{cases} \boldsymbol{m}_{i1}^T \mathcal{A} \boldsymbol{m}_{i2} = 0, \\ \boldsymbol{m}_{i2}^T \mathcal{A} \boldsymbol{m}_{i3} = 0, \\ \boldsymbol{m}_{i3}^T \mathcal{A} \boldsymbol{m}_{i1} = 0, \\ \boldsymbol{m}_{i1}^T \mathcal{A} \boldsymbol{m}_{i1} - \boldsymbol{m}_{i2}^T \mathcal{A} \boldsymbol{m}_{i2} = 0, \\ \boldsymbol{m}_{i2}^T \mathcal{A} \boldsymbol{m}_{i2} - \boldsymbol{m}_{i3}^T \mathcal{A} \boldsymbol{m}_{i3} = 0, \end{cases}$$

where $\mathcal{A} = \mathcal{Q}_3 \mathcal{Q}_3^T$.

$$\mathcal{M}_i \mathcal{A} \mathcal{M}_i^T = \rho_i^2 \mathcal{K}_i \mathcal{K}_i^T.$$

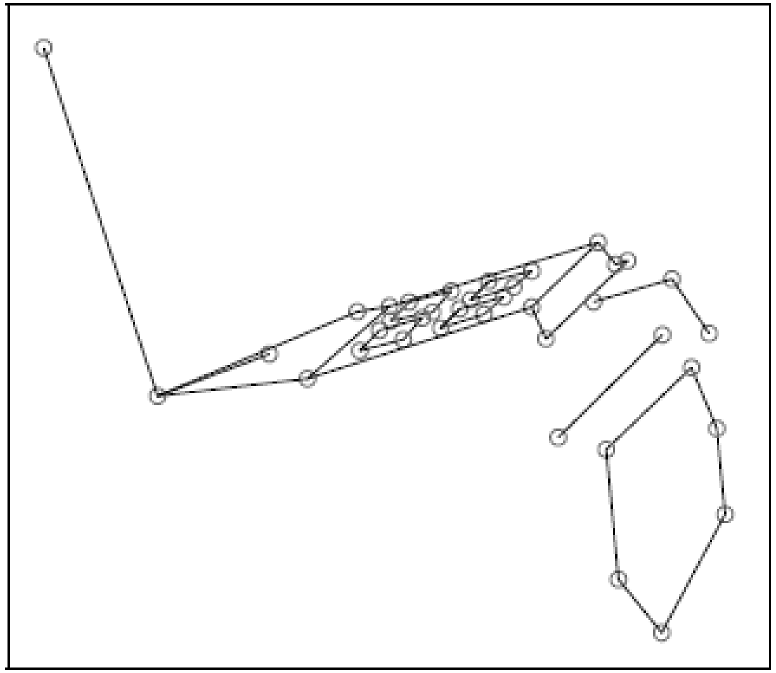
From Projective to Euclidean Shape

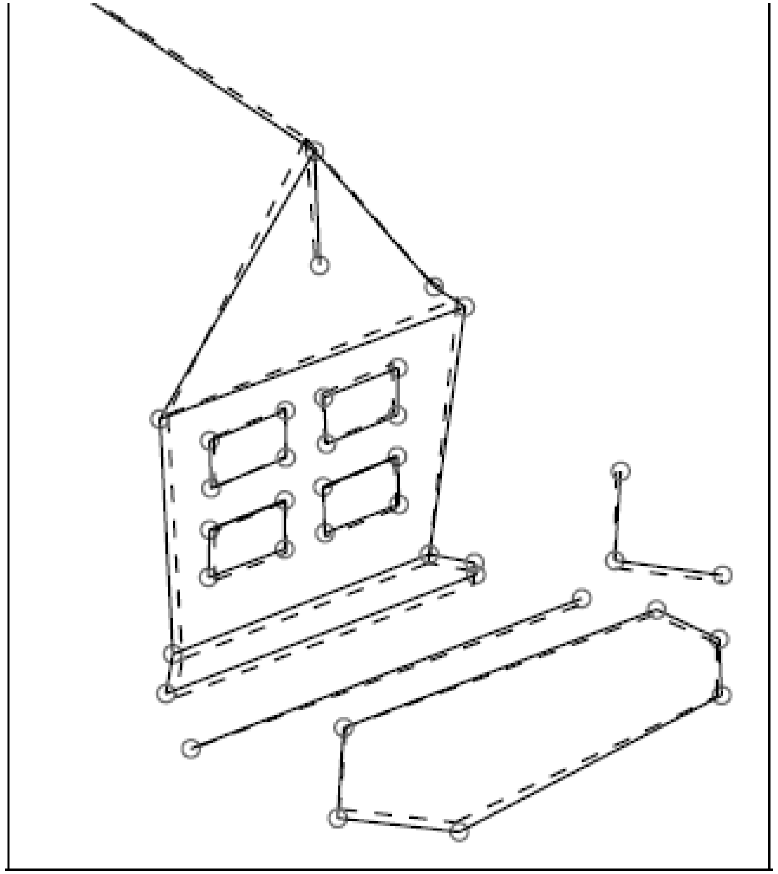
$$\mathcal{K}_i\mathcal{K}_i^T = \begin{pmatrix} \alpha_i^2 \dfrac{1}{\sin^2\theta_i} & -\alpha_i\beta_i \dfrac{\cos\theta_i}{\sin^2\theta_i} & 0 \\ -\alpha_i\beta_i \dfrac{\cos\theta_i}{\sin^2\theta_i} & \beta_i^2 \dfrac{1}{\sin^2\theta_i} & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

$\mathcal{K}_i\mathcal{K}_i^T$ provides two independent linear equations in the 10 coefficients of the 4×4 symmetric matrix $\mathcal{A}$:

$$\begin{cases} m_{i1}^T \mathcal{A} m_{i3} = 0, \\ m_{i2}^T \mathcal{A} m_{i3} = 0. \end{cases}$$

Euclidean reconstruction of the house obtained by a Euclidean upgrade of the projective reconstruction obtained with bundle adjustment





Conclusion

Plethora of applicatcions of computer vision

- code
- test
- learn

Q&A

Contact

- **Prof. D. Antony Louis Piriyakumar**
  Dean (Research and development)
  Cambridge Institute of Technology
- K.R. Puram,
  560036 Bengaluru, India
- Mobile: +91 98459 25132
- E-mail:
  dean_rd@Cambridge.edu.in