Motion

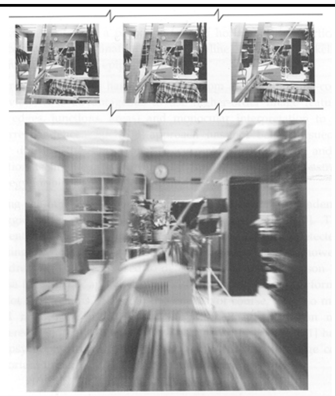
Computer Vision I

CSE 252A

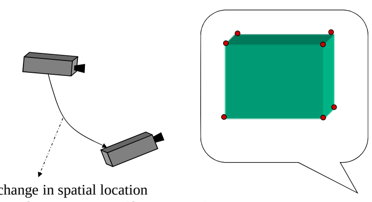
Lecture 15

Announcements

- Read Trucco and Verri
  - Course reserves
- Homework 3 is due Dec 4, 11:59 PM



Structure from Motion



MOVING CAMERAS ARE LIKE STEREO

Structure-from-Motion (SFM)

Goal: Take as input two or more images or video w/o any information on camera position/motion, and estimate camera position and 3-D structure of scene.

Two Approaches

1. Discrete motion (wide baseline)
   1. Orthographic (affine) vs. Perspective
   2. Two view vs. Multi-view
   3. Calibrated vs. Uncalibrated
2. Continuous (Infinitesimal) motion

Discrete Motion: Some Counting

Consider *M* images of *N* points, how many unknowns?

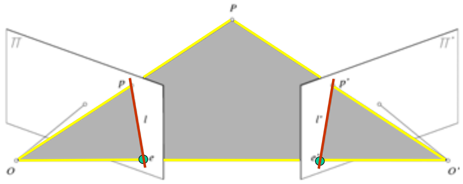
1. Affix coordinate system to location of first camera location: (M-1)*6 Unknowns
2. 3-D Structure: 3*N Unknowns
3. Can only recover structure and motion up to scale. Why?

Total number of unknowns: (M-1)*6+3*N-1

Total number of measurements: 2*M*N

Solution is possible when (M-1)*6+3*N-1 ≤ 2*M*N

Epipolar Constraint: Calibrated Case



$$\overrightarrow{Op}\cdot[\overrightarrow{OO'}\times\overrightarrow{O'p'}]=0 \;\Rightarrow\; \boldsymbol{p}\cdot[\boldsymbol{t}\times(\mathcal{R}\boldsymbol{p}')]=0 \quad \text{with} \begin{cases} \boldsymbol{p}=(u,v,1)^T \\ \boldsymbol{p}'=(u',v',1)^T \\ \mathcal{M}=(\mathrm{Id}\;\; \boldsymbol{0}) \\ \mathcal{M}'=(\mathcal{R}^T, -\mathcal{R}^T\boldsymbol{t}) \end{cases}$$

Essential Matrix (Longuet-Higgins, 1981) ⇐ $\boldsymbol{p}^T\mathcal{E}\boldsymbol{p}'=0$ with $\mathcal{E}=[\boldsymbol{t}_\times]\mathcal{R}$

where $[t_\times]=\begin{bmatrix}0 & -t_z & t_y\\ t_z & 0 & -t_x\\ -t_y & t_x & 0\end{bmatrix}$

The Eight-Point Algorithm (Longuet-Higgins, 1981)

$\boldsymbol{p}^T\mathcal{E}\boldsymbol{p}'=0$ with $\mathcal{E}=[\boldsymbol{t}_\times]\mathcal{R}$

$$[u,v,1]\begin{bmatrix}E_{11} & E_{12} & E_{13}\\ E_{21} & E_{22} & E_{23}\\ E_{31} & E_{32} & E_{33}\end{bmatrix}\begin{bmatrix}u'\\ v'\\ 1\end{bmatrix}=0 \;\Rightarrow\; [uu' \;\; uv' \;\; u \;\; vu' \;\; vv' \;\; v \;\; u' \;\; v' \;\; 1]\begin{bmatrix}E_{11}\\E_{12}\\E_{13}\\E_{21}\\E_{22}\\E_{23}\\E_{31}\\E_{32}\\E_{33}\end{bmatrix}=0$$

- Set E_{33} to 1
- Use 8 points (u_i,v_i), i=1..8

$$\begin{bmatrix}
u_1u'_1 & u_1v'_1 & u_1 & v_1u'_1 & v_1v'_1 & v_1 & u'_1 & v'_1\\
u_2u'_2 & u_2v'_2 & u_2 & v_2u'_2 & v_2v'_2 & v_2 & u'_2 & v'_2\\
u_3u'_3 & u_3v'_3 & u_3 & v_3u'_3 & v_3v'_3 & v_3 & u'_3 & v'_3\\
u_4u'_4 & u_4v'_4 & u_4 & v_4u'_4 & v_4v'_4 & v_4 & u'_4 & v'_4\\
u_5u'_5 & u_5v'_5 & u_5 & v_5u'_5 & v_5v'_5 & v_5 & u'_5 & v'_5\\
u_6u'_6 & u_6v'_6 & u_6 & v_6u'_6 & v_6v'_6 & v_6 & u'_6 & v'_6\\
u_7u'_7 & u_7v'_7 & u_7 & v_7u'_7 & v_7v'_7 & v_7 & u'_7 & v'_7\\
u_8u'_8 & u_8v'_8 & u_8 & v_8u'_8 & v_8v'_8 & v_8 & u'_8 & v'_8
\end{bmatrix}\begin{bmatrix}E_{11}\\E_{12}\\E_{13}\\E_{21}\\E_{22}\\E_{23}\\E_{31}\\E_{32}\end{bmatrix}=-\begin{bmatrix}1\\1\\1\\1\\1\\1\\1\\1\end{bmatrix}$$

⇒ Solve For E ⇒ Solve For **R** and **t**

Sketch of Two View SFM Algorithm

Input: Two images

1. Detect feature points
2. Find 8 matching feature points (easier said than done)
3. Compute the Essential Matrix E using Normalized 8-point Algorithm
4. Compute R and T (recall that E=RS where S is skew symmetric matrix)
5. Perform stereo matching using recovered epipolar geometry expressed via E.
6. Reconstruct 3-D geometry of corresponding points.

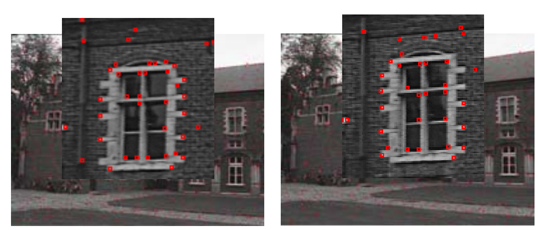
Essential Matrix

- Number of point correspondences and solutions
  - 5 point correspondences, up to 10 (real) solutions
  - 6 point correspondences, 1 solution
  - 7 point correspondences, 1 or 3 real solutions (and 2 or 0 complex ones)
  - 8 or more point correspondences, 1 solution

Fundamental Matrix

- Number of point correspondences and solutions
  - 7 point correspondences, 1 or 3 real solutions (and 2 or 0 complex ones)
  - 8 or more point correspondences, 1 solution

Feature points

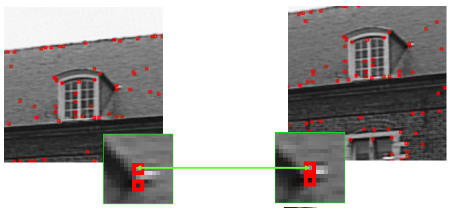


Select strongest features **(e.g. 1000/image)**

Feature matching

Evaluate normalized cross correlation (or sum of squared differences) for all features with similar coordinates

e.g. $(x',y') \in \left[x-\frac{w}{10}, x+\frac{w}{10}\right] \times \left[y-\frac{h}{10}, y+\frac{h}{10}\right]$



Keep mutual best matches

Still many wrong matches!

Comments

- Greedy Algorithm:
  - Given feature in one image, find best match in second image irrespective of other matches.
  - OK for small motions, little rotation, small search window
- Otherwise
  - Must compare descriptor over rotation
  - Can't consider O(n^8) potential pairings (way too many), so
    - Manual correspondence (e.g., façade, photogrammetry).
    - use random sampling (RANSAC)
    - More descriptive features (line segments,SIFT, larger regions,color).
    - Use video sequence to track, but perform SFM w/ first and last image.

RANSAC

Slides shamelessly taken from

Frank Dellaert and Marc Pollefeys and modified

Motivation

- Estimating motion models
- Typically: points in two images
- Candidates:
  - Translation
  - Affine
  - Homography

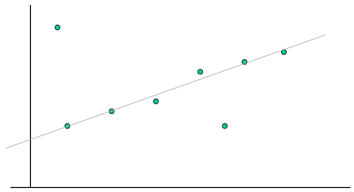
Mosaicing: Homography Estimate with RANSAC

www.cs.cmu.edu/~dellaert/mosaicking

Fundamental Matrix Estimate with RANSAC

Simpler Example

- Fitting a straight line



- Inliers
- Outliers

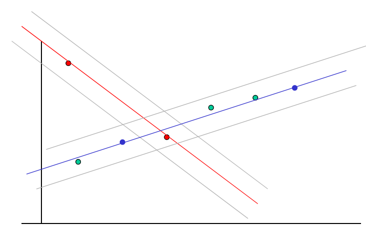
Discard Outliers

- No point with d>t
- RANSAC:
  - RANdom SAmple Consensus
  - Fischler & Bolles 1981
  - Copes with a large proportion of outliers

Main Idea

- Select 2 points at random
- Fit a line
- “Support” = number of inliers
- Line with most inliers wins

Why will this work ?



Best Line has most support

- More support -> better fit

RANSAC

Objective

Robust fit of model to data set S which contains outliers

Algorithm

(i) Randomly select a sample of *s* data points from S and instantiate the model from this subset.

(ii) Determine the set of data points S_i which are within a distance threshold *t* of the model. The set S_i is the **consensus set** of samples and defines the inliers of S.

(iii) If the size of S_i is greater than some threshold *T*, re-estimate the model using all the points in S_i and terminate

(iv) If the size of S_i is less than *T*, select a new subset and repeat the above.

(v) After *N* trials the largest consensus set S_i is selected, and the model is re-estimated using all the points in the subset S_i

How many samples?

Choose N (number of samples) so that, with probability p, at least one random sample is free from outliers. e.g. p=0.99

$$\left(1-(1-e)^s\right)^N = 1-p$$
$$N = \log(1-p)/\log\left(1-(1-e)^s\right)$$

e: proportion of outliers

s: Number of points needed for the model

	proportion of outliers e						
s	5%	10%	20%	25%	30%	40%	50%
2	2	3	5	6	7	11	17
3	3	4	7	9	11	19	35
4	3	5	9	13	17	34	72
5	4	6	12	17	26	57	146
6	4	7	16	24	37	97	293
7	4	8	20	33	54	163	588
8	5	9	26	44	78	272	1177

Acceptable consensus set?

- Typically, terminate when inlier ratio reaches expected ratio of inliers

$$T = (1-e)N$$

Using RANSAC to estimate the Fundamental Matrix

- What is the model?
- How many "points" are needed, and where do they come from?
- What distance do we use to compute the consensus set?
- How often do outliers occur

Is motion estimation inherent in humans? Demo

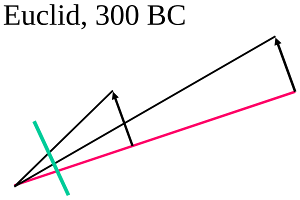
http://michaelbach.de/ot/cog-hiddenBird/index.html

Small Motion

Motion

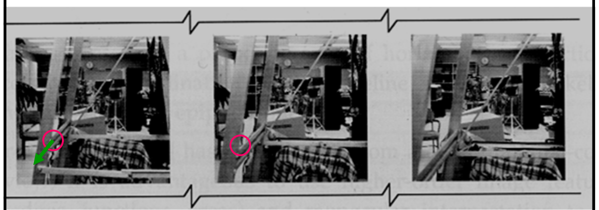
"When objects move at equal speed, those more remote seem to move more slowly."

- Euclid, 300 BC



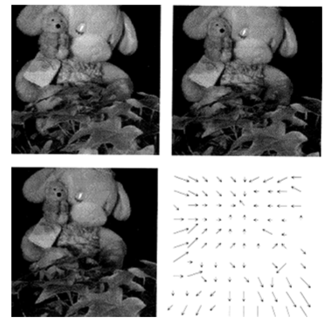
The Motion Field

Where in the image did a point move?



Down and left

The Motion Field



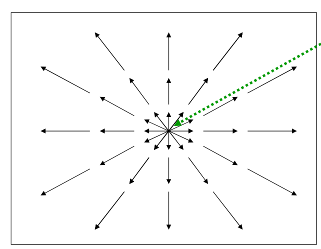
What causes a motion field?

1. Camera moves (translates, rotates)
2. Object's in scene move rigidly
3. Objects articulate (pliers, humans, animals)
4. Objects bend and deform (fish)
5. Blowing smoke, clouds

THE MOTION FIELD

The "instantaneous" velocity of all points in an image

LOOMING



The *Focus of Expansion (FOE)*

Intersection of velocity vector with image plane

With just this information it is possible to calculate:

1. Direction of motion
2. Time to collision