2-view Alignment and RANSAC

CSE P576

Dr. Matthew Brown
AutoStitch iPhone

“Create gorgeous panoramic photos on your iPhone”
- Cult of Mac

“Raises the bar on iPhone panoramas”
- TUAW

“Magically combines the resulting shots”
- New York Times

Available on the iPhone App Store
Case study – Image mosaicing

Any two images of a general scene with the same camera centre are related by a planar projective transformation given by:

$$\tilde{w}' = KRK^{-1}\tilde{w}$$

where $K$ represents the camera calibration matrix and $R$ is the rotation between the views.

This projective transformation is also known as the homography induced by the plane at infinity. A minimum of four image correspondences can be used to estimate the homography and to warp the images onto a common image plane. This is known as mosaicing.
Scale Invariant Feature Transform

• Extract SIFT features from an image

• Each image might generate 100’s or 1000’s of SIFT descriptors
Feature Matching

- Each SIFT feature is represented by 128 numbers
- Feature matching becomes task of finding a nearby 128-d vector
- All nearest neighbours:

  \[ \forall j \quad \text{NN}(j) = \arg \min_i \| \mathbf{x}_i - \mathbf{x}_j \|, \quad i \neq j \]

- Solving this exactly is \( O(n^2) \), but good approximate algorithms exist
- e.g., [Beis, Lowe '97] Best-bin first k-d tree
- Construct a binary tree in 128-d, splitting on the coordinate dimensions
- Find approximate nearest neighbours by successively exploring nearby branches of the tree
2-view Rotational Geometry

- Feature matching returns a set of noisy correspondences
- To get further, we will have to understand something about the geometry of the setup
2-view Rotational Geometry

- Recall the projection equation for a pinhole camera

\[
\tilde{u} = \begin{bmatrix}
K & R & t
\end{bmatrix}
\begin{bmatrix}
\tilde{X}
\end{bmatrix}
\]

\[
\tilde{u} \sim [u, v, 1]^T \quad : \text{Homogeneous image position}
\]

\[
\tilde{X} \sim [X, Y, Z, 1]^T \quad : \text{Homogeneous world coordinates}
\]

\[
K (3 \times 3) \quad : \text{Intrinsic (calibration) matrix}
\]

\[
R (3 \times 3) \quad : \text{Rotation matrix}
\]

\[
t (3 \times 1) \quad : \text{Translation vector}\]
2-view Rotational Geometry

- Consider two cameras at the same position (translation)
- WLOG we can put the origin of coordinates there

\[ \tilde{u}_1 = K_1 [ R_1 \mid t_1 ] \tilde{X} \]

- Set translation to 0

\[ \tilde{u}_1 = K_1 [ R_1 \mid 0 ] \tilde{X} \]

- Remember \( \tilde{X} \sim [X, Y, Z, 1]^T \) so

\[ \tilde{u}_1 = K_1 R_1 X \]

(where \( X = [X, Y, Z]^T \))
2-view Rotational Geometry

- Add a second camera (same translation but different rotation and intrinsic matrix)

\[
\tilde{u}_1 = K_1 R_1 X \\
\tilde{u}_2 = K_2 R_2 X
\]

- Now eliminate \( X \)

\[
X = R_1^T K_1^{-1} \tilde{u}_1
\]

- Substitute in equation 1

\[
\tilde{u}_2 = K_2 R_2 R_1^T K_1^{-1} \tilde{u}_1
\]

This is a 3x3 matrix -- a (special form) of homography.
Computing H: Quiz

\[
\begin{bmatrix}
  u \\ v \\ 1
\end{bmatrix}
= 
\begin{bmatrix}
  h_{11} & h_{12} & h_{13} \\
  h_{21} & h_{22} & h_{23} \\
  h_{31} & h_{32} & h_{33}
\end{bmatrix}
\begin{bmatrix}
  x \\ y \\ 1
\end{bmatrix}
\]

• Each correspondence between 2 images generates _____ equations
• A homography has _____ degrees of freedom
• _____ point correspondences are needed to compute the homography
• Rearranging to make H the subject leads to an equation of the form

\[\mathbf{Mh} = 0\]

• This can be solved by _____
Computing $H$: Quiz

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

- Each correspondence between 2 images generates ____ equations
- A homography has ____ degrees of freedom
- ____ point correspondences are needed to compute the homography
- Rearranging to make $H$ the subject leads to an equation of the form
  $$Mh = 0$$
- This can be solved by ____
Finding Consistent Matches

- Raw SIFT correspondences (contains outliers)
Finding Consistent Matches

• SIFT matches consistent with a rotational homography
Finding Consistent Matches

- Warp images to common coordinate frame
2-view Alignment + RANSAC

- 2-view alignment: linear equations
- Least squares and outliers
- Robust estimation via sampling

[Szeliski 6.1]
Image Alignment

- Find corresponding (matching) points between the images

\[ u = Hx \]

2 points for Similarity
3 for Affine
4 for Homography
Image Alignment

- In practice we have many noisy correspondences + outliers
RANSAC algorithm

1. Match feature points between 2 views
2. Select minimal subset of matches*
3. Compute transformation $T$ using minimal subset
4. Check consistency of all points with $T$ — compute projected position and count #inliers with distance $<$ threshold
5. Repeat steps 2-4 to maximise #inliers

* Similarity transform = 2 points, Affine = 3, Homography = 4
2-view Rotation Estimation

- Find features + raw matches, use RANSAC to find Similarity
2-view Rotation Estimation

- Remove outliers, can now solve for $R$ using least squares
2-view Rotation Estimation

- Final rotation estimation