RANSAC and SIFT

- Powerful combination to find objects in images
- Exemplar image and image being studied typically have different camera angle or position.
- Recall that:
  - SIFT descriptors are relatively invariant to camera changes
  - SIFT matching leads to lots of “false” matches
- The main idea is that true matches should “agree”
- For planar objects, the definition of “agree” is quite simple
  - They link via a homography (covered soon)
Matching Slides to Presentation Videos

Slides

Keyframes
SIFT (Scale Invariant Feature Transformation) keypoints review

local feature descriptors
location, scale, orientation and a feature vector with 128 elements

![Image showing keypoints and their descriptors]

Only a quarter of keypoints shown!
Nearest neighbor ratio has many outliers
Planar Homography

Mappings of points on a plane in 3D satisfy a simple relation

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

frame keypoints

slide keypoints

\[ X' = H X \]
Derivation of Planar Homography

Consider a point on a plane given by

\[ X = X_o + sX_1 + tX_2 \]

under the two projective transforms

\[ P = [A \ b] \quad \text{and} \quad P' = [A' \ b'] \]

This leads to two image points, \( \lambda p \) and \( \lambda' p' \).
Derivation of Planar Homography

\[
\lambda \mathbf{p} = [A \ b] \begin{bmatrix} X_o + sX_1 + tX_2 \ \ 1 \end{bmatrix}
\]

\[
= [A \ b] \begin{bmatrix} X_1 & X_2 & X_o \ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s \\ t \\ 1 \end{bmatrix}
\]

\[
= [AX_1 \ AX_2 \ AX_o + b] \begin{bmatrix} s \\ t \\ 1 \end{bmatrix}
\]

\[
= V \begin{bmatrix} s \\ t \\ 1 \end{bmatrix}
\]
Derivation of Planar Homography

Similarly, \[ \lambda' p' = V' \begin{bmatrix} s \\ t \\ 1 \end{bmatrix} \]

and so \[ \lambda' p' = V' V^{-1} \lambda p = H \lambda p \]
Constraining matches by homography

Fit homography with RANSAC

Only a quarter of matches shown!
RANSAC approach

Repeat many times

Randomly select enough matches to fit homography

Compute homography

Using that homography, measure error on best (say) 50%

Output best one found
Computing Homography

Seek $H$ where

$$
\begin{bmatrix}
u' \\ v' \\ w'
\end{bmatrix} = H
\begin{bmatrix}
x \\ y \\ 1
\end{bmatrix}
X = \begin{bmatrix}
x \\ y \\ 1
\end{bmatrix}
$$

$H$ is only determined up to a scale factor (eight unknowns).

Let the rows of $H$ be $h_1^T, h_2^T, h_3^T$.

$$
x' = \frac{u'}{w'} \text{ so } x'w' = u'. \text{ Similarly, } y'w' = v'
$$

Also, $u' = h_1^T X$ and $v' = h_2^T X$ and $w' = h_3^T X$
Computing Homography

Each match then gives two linear equations

\[ x' h_3^T X = h_1^T X \quad \text{and} \quad y' h_3^T X = h_2^T X \]

Hence four matches are OK.

This can be solved with homogenous least squares*, but this is a bit unstable. A better way is the DLT (direct linear transform) method.

* Doing this is part of the homework. You may want to review how we set up the equations for camera calibration. Notice that in the expressions above, \( X \) can be a row vector and the \( h_x \) column vectors. Referring to camera calibration, \( X \) is playing the role that \( P \) did, and \( h_x \) are playing the role of \( m_x \) (rows of \( M \)).