Recognising Panoramas

M. Brown and D. Lowe,
University of British Columbia
Introduction

• Are you getting the whole picture?
  – Compact Camera FOV = 50 x 35°
Introduction

- Are you getting the whole picture?
  - Compact Camera FOV = 50 x 35°
  - Human FOV = 200 x 135°
Introduction

- Are you getting the whole picture?
  - Compact Camera FOV = 50 x 35°
  - Human FOV = 200 x 135°
  - Panoramic Mosaic = 360 x 180°
Why “Recognising Panoramas”? 
Why “Recognising Panoramas”? 

- 1D Rotations ($\theta$)
  - Ordering $\Rightarrow$ matching images
Why “Recognising Panoramas”?

• 1D Rotations ($\theta$)
  – Ordering $\Rightarrow$ matching images
Why “Recognising Panoramas”?

- 1D Rotations ($\theta$)
  - Ordering $\Rightarrow$ matching images
Why “Recognising Panoramas”?

- **1D Rotations ($\theta$)**
  - Ordering $\Rightarrow$ matching images

- **2D Rotations ($\theta, \phi$)**
  - Ordering $\not\Rightarrow$ matching images
Why “Recognising Panoramas”?

• 1D Rotations ($\theta$)
  - Ordering $\Rightarrow$ matching images

• 2D Rotations ($\theta$, $\phi$)
  - Ordering $\not\Rightarrow$ matching images
Why “Recognising Panoramas”?

- 1D Rotations ($\theta$)
  - Ordering $\Rightarrow$ matching images

- 2D Rotations ($\theta$, $\phi$)
  - Ordering $\nRightarrow$ matching images
Why “Recognising Panoramas”? 
Overview

- Feature Matching
- Image Matching
- Bundle Adjustment
- Multi-band Blending
- Results
- Conclusions
Outline

- Input Images
- Feature Extraction
- Image Matching
- Bundle Adjustment
  and Multi-band Blending
Overview

- Feature Matching
- Image Matching
- Bundle Adjustment
- Multi-band Blending
- Results
- Conclusions
Overview

• Feature Matching
  – SIFT Features
  – Nearest Neighbour Matching

• Image Matching
• Bundle Adjustment
• Multi-band Blending
• Results
• Conclusions
Overview

• Feature Matching
  – SIFT Features
  – Nearest Neighbour Matching

• Image Matching

• Bundle Adjustment

• Multi-band Blending

• Results

• Conclusions
Invariant Features

SIFT Features – Location

- Extrema of DoG
- Discard Low Contrast
- Discard Edge Points

Output: x,y,σ

Picture Credit: http://en.wikipedia.org/wiki/Scale-invariant_feature_transform
SIFT Features – Orientation

• **Difference of Gaussian Function:**
  
  – \( G(x, y, k\sigma) = \text{Gaussian Kernel with SD = } k\sigma \)
  
  – \( L(x, y, k\sigma) = l(x, y) \ast G(x, y, k\sigma) \)
  
  – \( D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, k_j\sigma) \)

• **Gradient Magnitude:**
  
  – \( m(x,y) = \sqrt{ (L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2 } \)
  
  – \( \Theta(x, y) = \tan^{-1}{ (L(x, y + 1) - L(x, y - 1)) / (L(x + 1, y) - L(x - 1, y)) } \)
SIFT Features – Orientation

Local Gradients

Adjusted Gradients

Gaussian Weighting

Orientation Histogram – 36 bins with 10 degrees each
SIFT Features – Orientation

Local Gradients

Adjusted Gradients

Gaussian Weighting

Orientation Histogram – 36 bins with 10 degrees each
SIFT Features – Descriptor Vector

- 4 x 4 region around point oriented according to $\theta$
- Take 8 versions of the gradient image corresponding to 8 different directions.
  - Each version has only gradients that fall in corresponding direction range
  - Concatenate the 8 versions (each 4 x 4) together to get 128 dimensional descriptor vector
SIFT Features

• Invariant Features
  – Establish invariant frame
    • Maxima/minima of scale-space DOG ⇒ x, y, s
    • Maximum of distribution of local gradients ⇒ θ
  – Form descriptor vector
    • Histogram of smoothed local gradients
    • 128 dimensions

• SIFT features are...
  – Geometrically invariant to similarity transforms,
    • some robustness to affine change
  – Photometrically invariant to affine changes in intensity
Overview

• Feature Matching
  – SIFT Features
  – Nearest Neighbour Matching

• Image Matching
• Bundle Adjustment
• Multi-band Blending
• Results
• Conclusions
Nearest Neighbour Matching

- Find k-NN for each feature
  - \( k \approx \) number of overlapping images (we use \( k = 4 \))
- Use k-d tree
  - k-d tree recursively bi-partitions data at mean in the dimension of maximum variance
  - Approximate nearest neighbours found in \( O(n\log n) \)
Overview

- Feature Matching
  - SIFT Features
  - Nearest Neighbour Matching
- Image Matching
- Bundle Adjustment
- Multi-band Blending
- Results
- Conclusions
Overview

- Feature Matching
- Image Matching
- Bundle Adjustment
- Multi-band Blending
- Results
- Conclusions
Overview

- Feature Matching
- Image Matching
- Bundle Adjustment
- Multi-band Blending
- Results
- Conclusions
Overview

• Feature Matching
• Image Matching
  – RANSAC for Homography
  – Probabilistic model for verification
• Bundle Adjustment
• Multi-band Blending
• Results
• Conclusions
Overview

• Feature Matching
• Image Matching
  – RANSAC for Homography
  – Probabilistic model for verification
• Bundle Adjustment
• Multi-band Blending
• Results
• Conclusions
**Defn – Homography:**
A one-to-one mapping between two images. In computer vision, it is typically used to describe the correspondence between two images taken of the same scene from different camera angles.

**For Images:**
Homography has four parameters, $\theta_1 \theta_2 \theta_3 f$, corresponding to the three angles of camera rotation and the focal length.

Picture Credit: http://weboflife.nasa.gov/learningResources/vestibularbrief.htm
bestH = 0% success
While i < max iterations:
   - Select a random sample of the data
   - currentH = solve for the homography using sample data
   - for each pair of points, test to see if currentH is a suitable solution
   - if more points work with currentH than with bestH
     set bestH = currentH
RANSAC for Homography
RANSAC for Homography
Overview

- Feature Matching
- Image Matching
  - RANSAC for Homography
  - Probabilistic model for verification
- Bundle Adjustment
- Multi-band Blending
- Results
- Conclusions
Probabilistic model for verification
Probabilistic model for verification

• Want to solve for the probability that the image match is correct given the set of RANSAC inliers and outliers.
  
  \[
  p( \text{match} | \text{inliers} )
  \]

• Using Bayes’ Rule:
  
  \[
  p( \text{match} | \text{inliers} ) = p( \text{inliers} | \text{match} ) \cdot p( \text{match} ) / p( \text{inliers} )
  \]

• This can be simplified if we make some assumptions
  
  \[
  p( \text{inliers} | \text{match} ) = B( n_i; n_f, p_1)
  \]
  
  \[
  p( \text{inliers} | \sim \text{match} ) = B( n_i; n_f, p_0)
  \]

  (B is a binomial distribution function)
Probabilistic model for verification

- Compare probability that this set of RANSAC inliers/outliers was generated by a correct/false image match

\[
\frac{B(n_i; n_f, p_1)}{B(n_i; n_f, p_0)} \begin{cases} \text{accept} & \frac{1}{p_{\text{min}}} - 1 \\ \text{reject} & \frac{1}{p_{\text{min}}} - 1 \end{cases}
\]

- \( n_i = \#\text{inliers}, \; n_f = \#\text{features} \)
- \( p_1 = p(\text{inlier} \mid \text{match}), \; p_0 = p(\text{inlier} \mid \sim\text{match}) \)
- \( p_{\text{min}} = \text{acceptance probability} \)

- Choosing values for \( p_1, p_0 \) and \( p_{\text{min}} \)

\[
n_i > 5.9 + 0.22n_f
\]
Finding the panoramas
Finding the panoramas
Finding the panoramas
Finding the panoramas
Overview

• Feature Matching
• Image Matching
  – RANSAC for Homography
  – Probabilistic model for verification
• Bundle Adjustment
• Multi-band Blending
• Results
• Conclusions
Overview

- Feature Matching
- **Image Matching**
- Bundle Adjustment
- Multi-band Blending
- Results
- Conclusions
Overview

- Feature Matching
- Image Matching
- Bundle Adjustment
- Multi-band Blending
- Results
- Conclusions
Bundle Adjustment

- Overlay the images together one at a time.

- For each image, solve for the camera parameters $\theta_1 \theta_2 \theta_3 f$ to align it with the current compilation using Levenberg-Marquardt to minimize an error function.
Overview

• Feature Matching
• Image Matching
• Bundle Adjustment
  – Error function
• Multi-band Blending
• Results
• Conclusions
Overview

• Feature Matching
• Image Matching
• Bundle Adjustment
  – Error function
• Multi-band Blending
• Results
• Conclusions
Error function

- **Sum of squared projection errors**

\[ e = \sum_{i=1}^{n} \sum_{j \in I(i)} \sum_{k \in F(i,j)} f(r_{ij}^k)^2 \]

- \( n = \#\text{images} \)
- \( I(i) = \text{set of image matches to image } i \)
- \( F(i, j) = \text{set of feature matches between images } i, j \)
- \( r_{ij}^k = \text{residual of } k^{\text{th}} \text{ feature match between images } i, j \)

- **Robust error function**

\[
 f(x) = \begin{cases} 
 |x|, & \text{if } |x| < x_{max} \\
 x_{max}, & \text{if } |x| \geq x_{max} 
\end{cases}
\]
Homography for Rotation

- Parameterise each camera by rotation and focal length

\[
R_i = e^{[\theta_i]_\times}, \quad [\theta_i]_\times = \begin{bmatrix}
0 & -\theta_{i3} & \theta_{i2} \\
\theta_{i3} & 0 & -\theta_{i1} \\
-\theta_{i2} & \theta_{i1} & 0
\end{bmatrix}
\]

\[
K_i = \begin{bmatrix}
f_i & 0 & 0 \\
0 & f_i & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

- This gives pairwise homographies

\[
\tilde{u}_i = H_{ij} \tilde{u}_j, \quad H_{ij} = K_i R_i R_j^T K_j^{-1}
\]
Bundle Adjustment

- New images initialised with rotation, focal length of best matching image
Bundle Adjustment

• New images initialised with rotation, focal length of best matching image
Overview

• Feature Matching
• Image Matching
• Bundle Adjustment
  – Error function
• Multi-band Blending
• Results
• Conclusions
Overview

- Feature Matching
- Image Matching
- Bundle Adjustment
- Multi-band Blending
- Results
- Conclusions
Overview

- Feature Matching
- Image Matching
- Bundle Adjustment
- Multi-band Blending
- Results
- Conclusions
Multi-band Blending

- Burt & Adelson 1983
  - Blend frequency bands over range $\propto \lambda$
2-band Blending

Low frequency ($\lambda > 2$ pixels)

High frequency ($\lambda < 2$ pixels)
2-band Blending
Overview

- Feature Matching
- Image Matching
- Bundle Adjustment
- Multi-band Blending
- Results
- Conclusions
Overview

• Feature Matching
• Image Matching
• Bundle Adjustment
• Multi-band Blending
• Results
• Conclusions
Results
Overview

• Feature Matching
• Image Matching
• Bundle Adjustment
• Multi-band Blending
• Results
• Conclusions
Overview

- Feature Matching
- Image Matching
- Bundle Adjustment
- Multi-band Blending
- Results
- Conclusions
Conclusions

• Fully automatic panoramas
  – A recognition problem...

• Invariant feature based method
  – SIFT features, RANSAC, Bundle Adjustment, Multi-band Blending
  – O(nlogn)

• Future Work
  – Advanced camera modelling
    • radial distortion, camera motion, scene motion, vignetting, exposure, high dynamic range, flash ...
  – Full 3D case – recognising 3D objects/scenes in unordered datasets

http://www.cs.ubc.ca/~mbrown/panorama/panorama.html
Analytical computation of derivatives

\[
\frac{\partial p_{ij}^k}{\partial \theta_{i1}} = \frac{\partial p_{ij}^k}{\partial \tilde{p}_{ij}^k} \frac{\partial \tilde{p}_{ij}^k}{\partial \theta_{i1}}
\]

\[
\frac{\partial p_{ij}^k}{\partial \tilde{p}_{ij}^k} = \frac{\partial}{\partial \begin{bmatrix} x/z & y/z \\ x & y & z \end{bmatrix}} = \begin{bmatrix} 1/z & 0 & -x/z^2 \\ 0 & 1/z & -y/z^2 \end{bmatrix}
\]

\[
\frac{\partial \tilde{p}_{ij}^k}{\partial \theta_{i1}} = K_i \frac{\partial R_i}{\partial \theta_{i1}} R_j K_j^{-1} \tilde{u}_l^j
\]

\[
\frac{\partial R_i}{\partial \theta_{i1}} = \frac{\partial}{\partial \theta_{i1}} e[\theta_i] \times = e[\theta_i] \times \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}
\]
Levenberg-Marquardt

• Iteration step of form

\[ \Theta = (J^T J + \sigma^2 C_p^{-1})^{-1} J^T r \]