Image Stitching

Hong Shang & Jesus Avila

EE225B, 1 April 2014

Outline

- Motivation
- Stitching Steps
  - Coordinate System and Motion Modeling
  - Alignment: Direct and Featured-based
  - Compositing
- Summary

Why Mosaic?

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

Slide from Brown & Lowe

Why Mosaic?

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

Slide from Brown & Lowe

Why Mosaic?

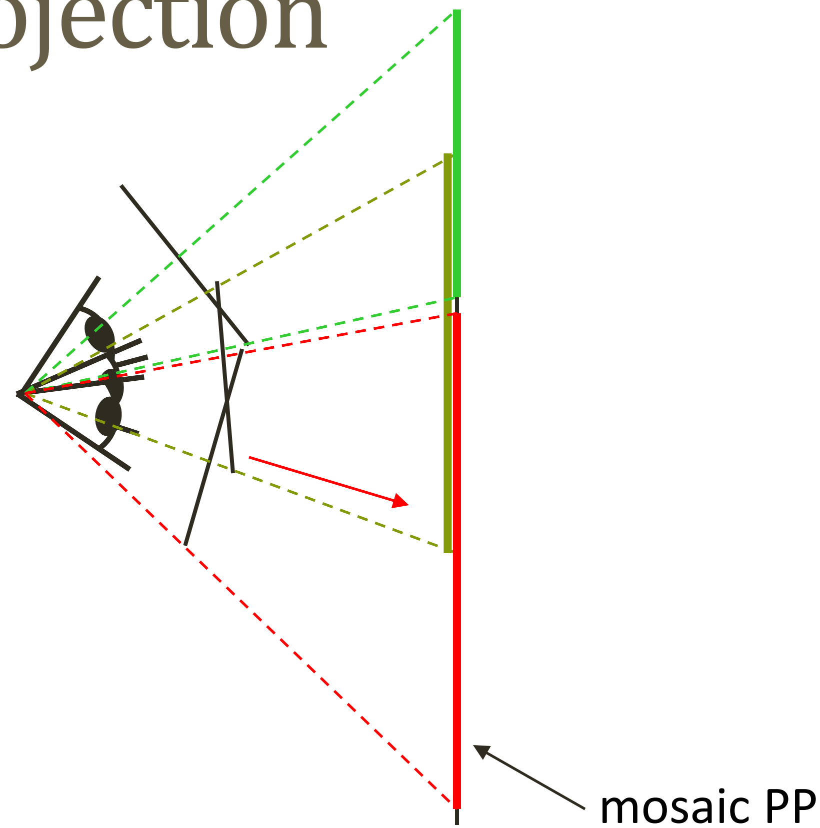
- Are you getting the whole picture?
  - Compact Camera FOV = 50 x 35°
  - Human FOV = 200 x 135°
  - Panoramic Mosaic = 360 x 180°

Slide from Brown & Lowe

Outline

- Motivation
- **Stitching Steps**
  - **Coordinate System/Motion Modeling**
  - Alignment: Direct and Featured-based
  - Compositing
- Conclusions

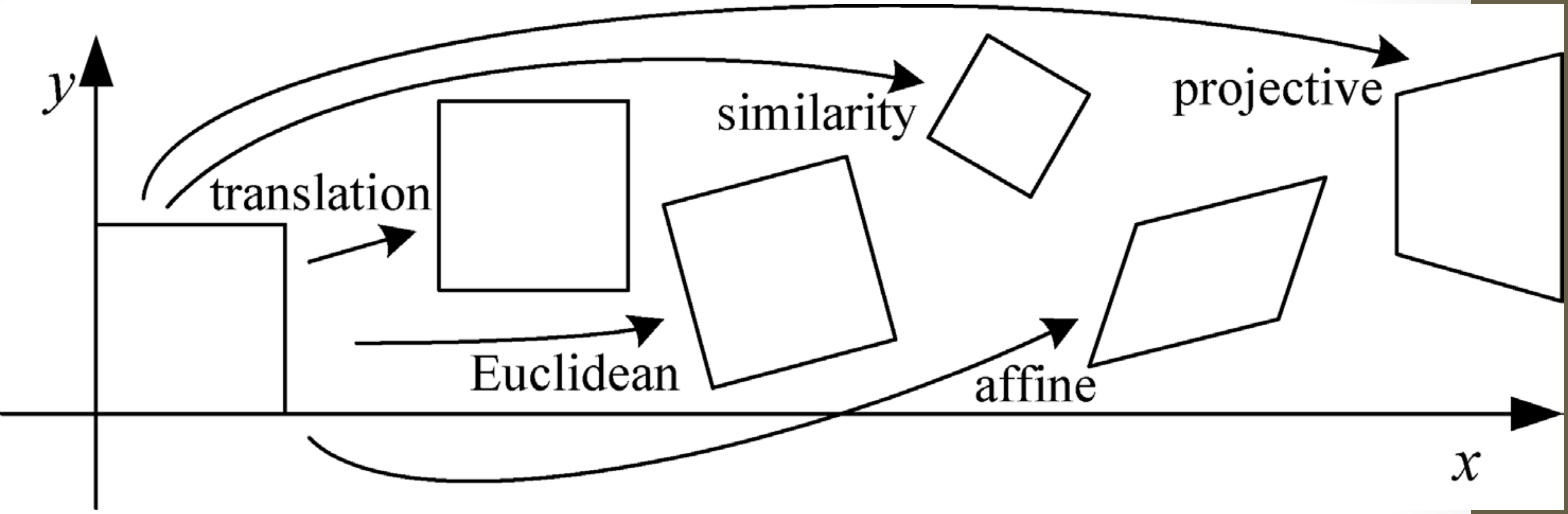
Motion Modeling
Image reprojection



- The mosaic has a natural interpretation in 3D
  - The images are reprojected onto a common plane
  - The mosaic is formed on this plane
  - Mosaic is a *synthetic wide-angle camera*

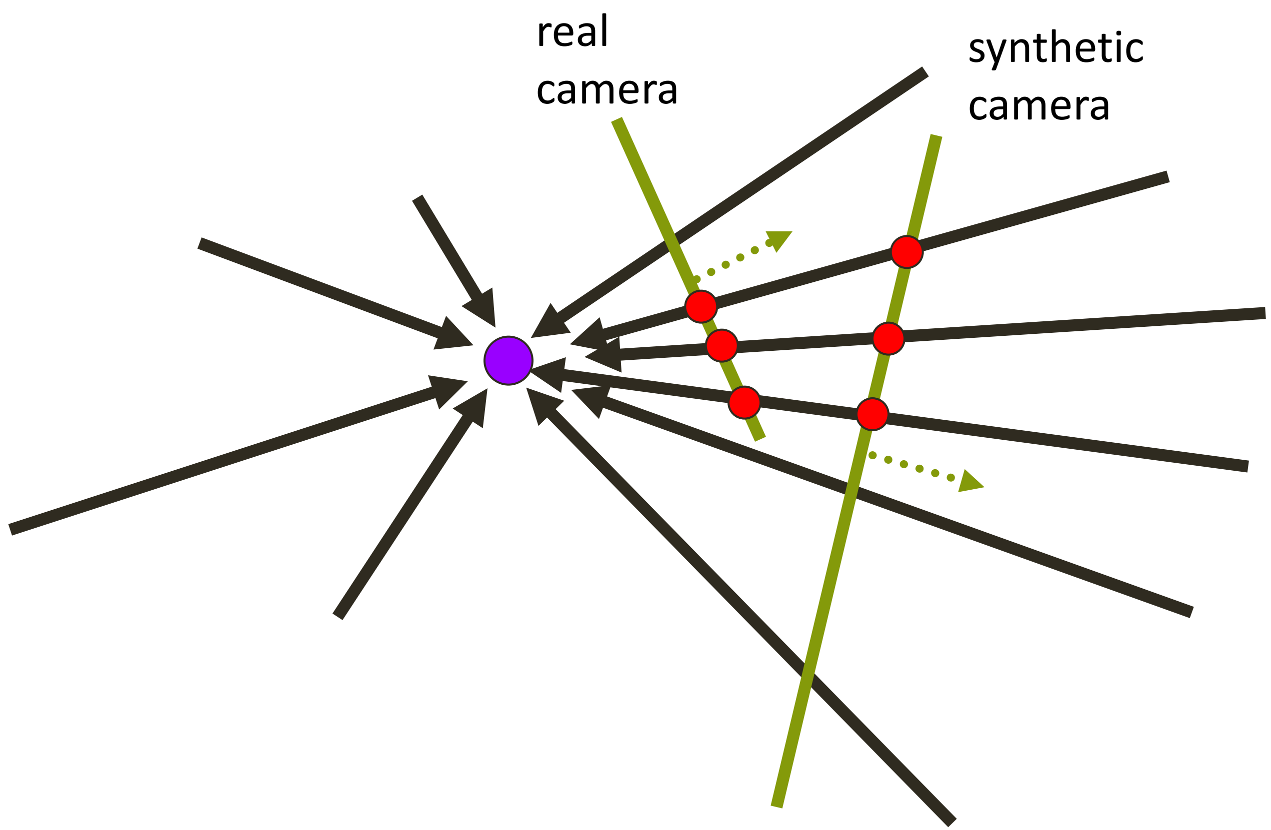
Alexei Efros, CMU

Motion Modeling: 2-D



Motion Modeling: 3-D

A pencil of rays contains all views



Can generate any synthetic camera view
as long as it has **the same center of projection**!

Alexei Efros, CMU

Outline

- Motivation
- Stitching Steps
  - Coordinate System and Motion Modeling
  - **Alignment: Direct and Feature-based**
  - Compositing
- Conclusions

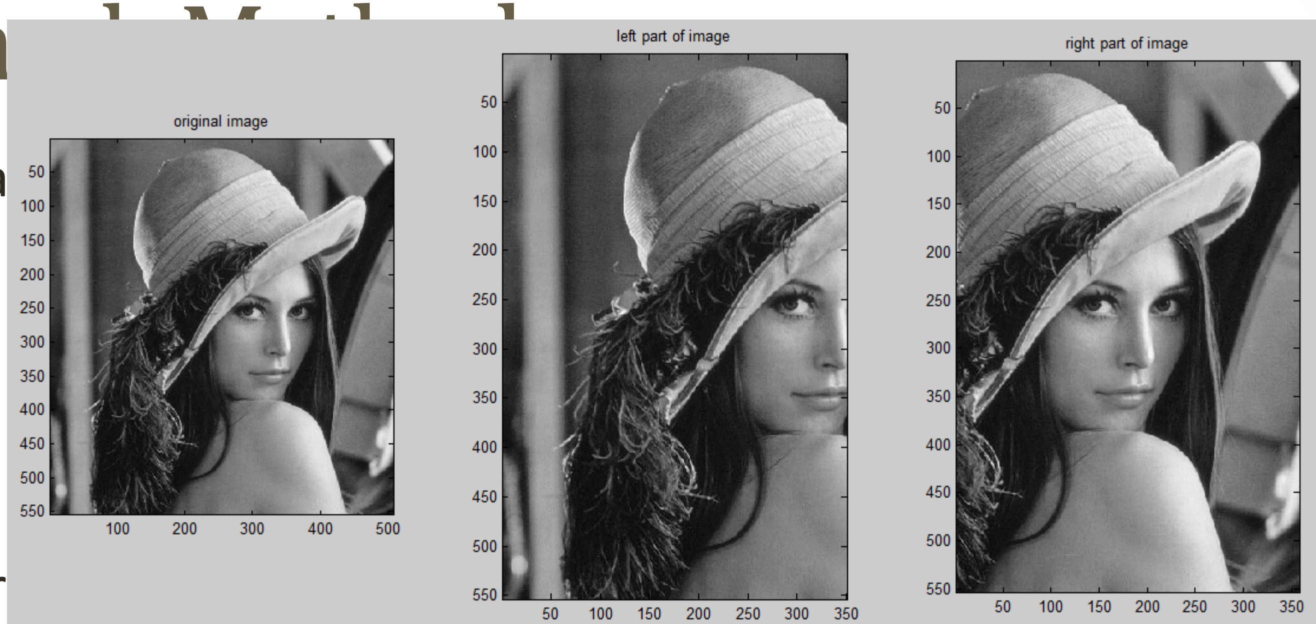
Direct (Pixel-based) Alignment

Determine Error Metric → Use search technique to find possible alignments

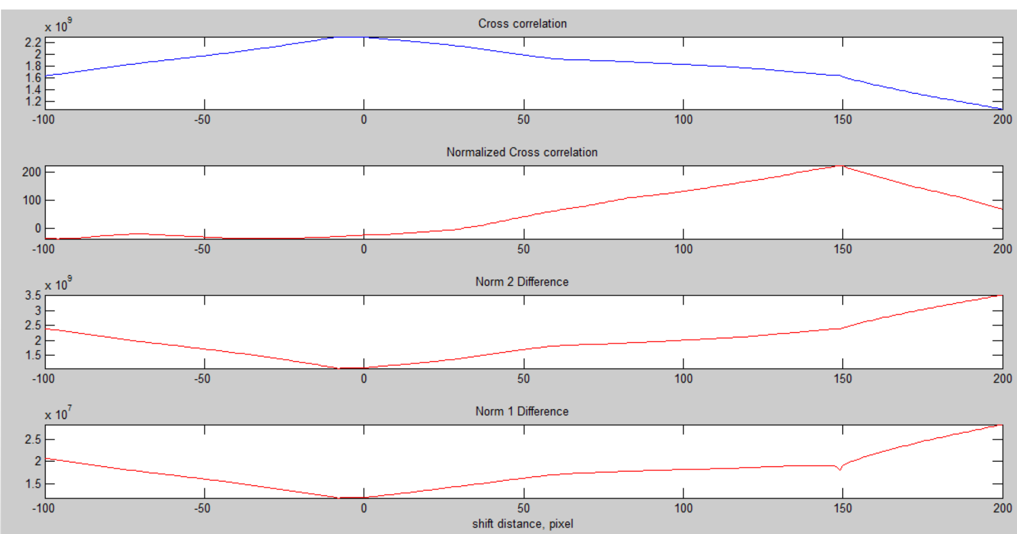
Complexity of error metric depends on complexity of motion model

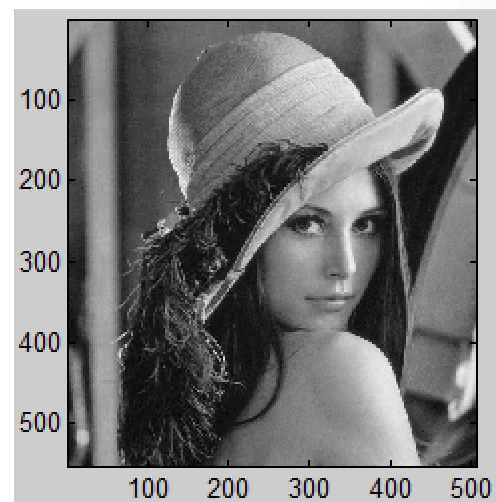
Example of Simple Error Metric & Sea

- Ma



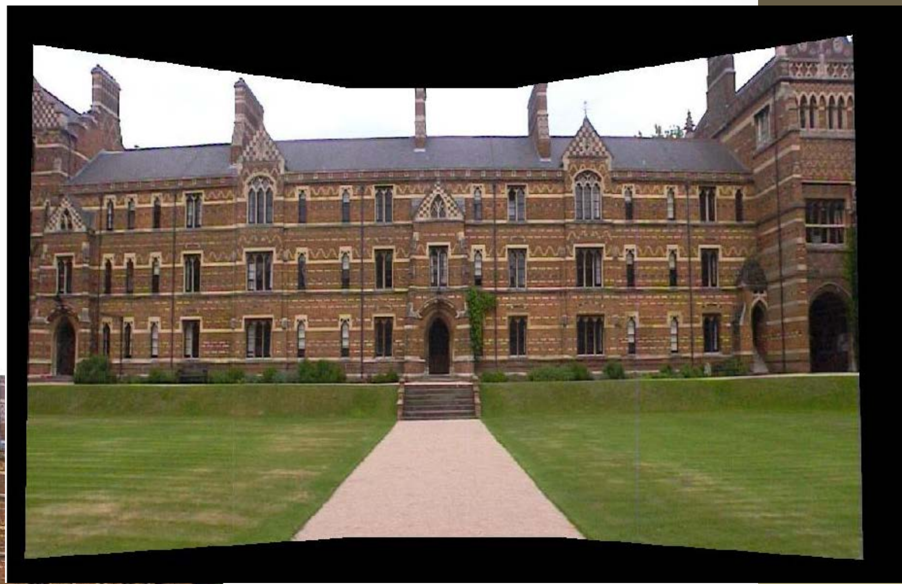
- Do
  cor

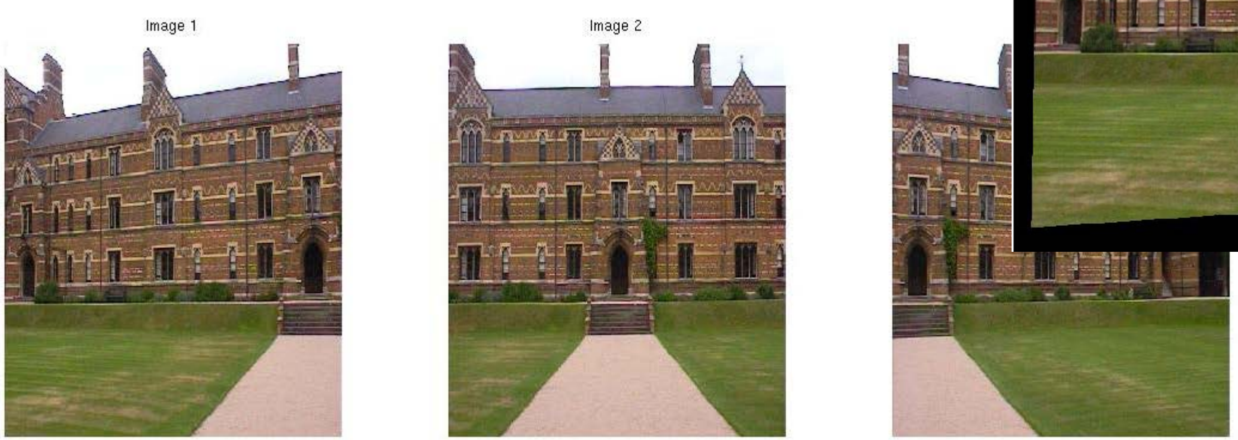




Feature-based approach

1. Detect local features
2. Extract feature descriptor
3. Match feature between two images
4. Estimate homography using RANSAC
5. Warping
6. compositing





http://www.di.ens.fr/willow/teaching/recvis10/assignment2/

Feature-based approach

1. Detect local features

a simple single scale Harris corner detector

The Hessian and eigenvalue images can be efficiently evaluated using a sequence of filters and algebraic operations

$$G_x(\boldsymbol{x}) = \frac{\partial}{\partial x} G_{\sigma_d}(\boldsymbol{x}) * I(\boldsymbol{x}),$$

$$G_y(\boldsymbol{x}) = \frac{\partial}{\partial y} G_{\sigma_d}(\boldsymbol{x}) * I(\boldsymbol{x}),$$

$$\boldsymbol{B}(\boldsymbol{x}) = \begin{bmatrix} G_x^2(\boldsymbol{x}) & G_x(\boldsymbol{x})G_y(\boldsymbol{x}) \\ G_x(\boldsymbol{x})G_y(\boldsymbol{x}) & G_y^2(\boldsymbol{x}) \end{bmatrix},$$

$$\boldsymbol{A}(\boldsymbol{x}) = G_{\sigma_i}(\boldsymbol{x}) * \boldsymbol{B}(\boldsymbol{x}),$$

$$\lambda_{0,1}(\boldsymbol{x}) = \frac{a_{00} + a_{11} \mp \sqrt{(a_{00} - a_{11})^2 + a_{01}a_{10}}}{2},$$

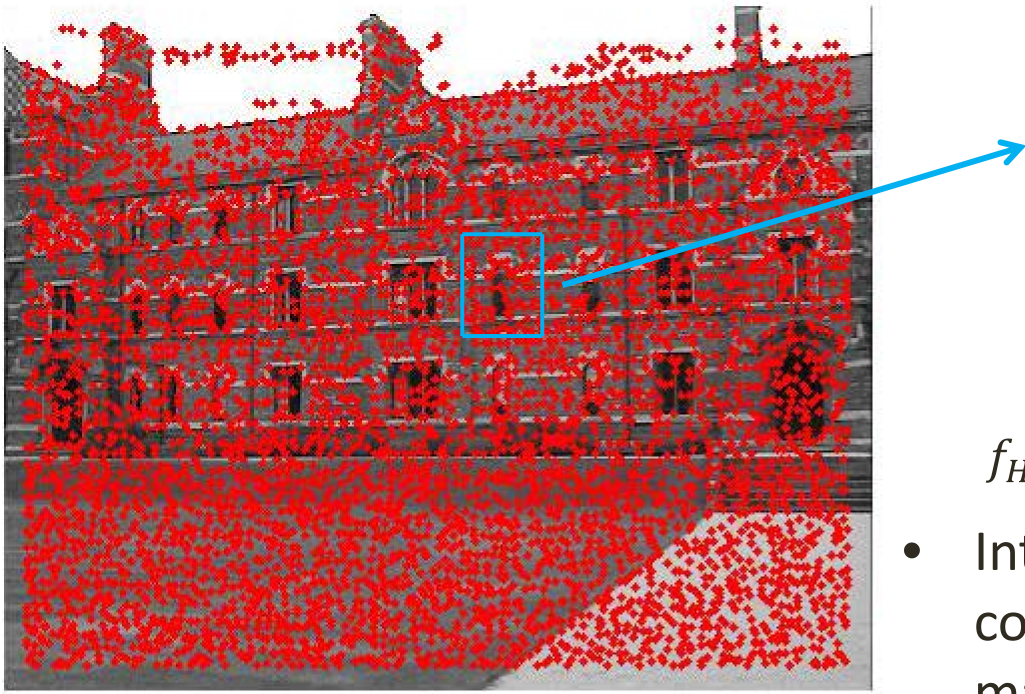
G_{σ_d} is a noise-reducing pre-smoothing "derivative" Gaussian filter of width $\sigma_d (= 1)$

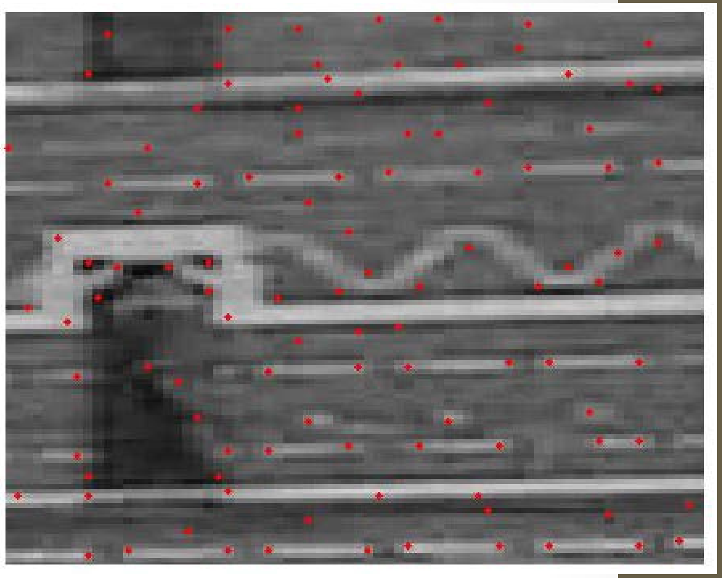
G_{σ_i} is the integration Gaussian filter whose scale $\sigma_i (= 1.5)$ controls the effective patch size.

Szeliski, Richard. "Image alignment and stitching: A tutorial." *Foundations and Trends® in Computer Graphics and Vision* 2.1 (2006): 1-104.

Feature-based approach

1. Detect local features





$$f_{HM}(x,y) = \frac{\det A(x,y)}{tr\ A(x,y)} = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2}$$

- Interest point are located where corner strength is a local maximum in 3*3 neighbourhood
- Non-maximal supresion for spreading out interest points

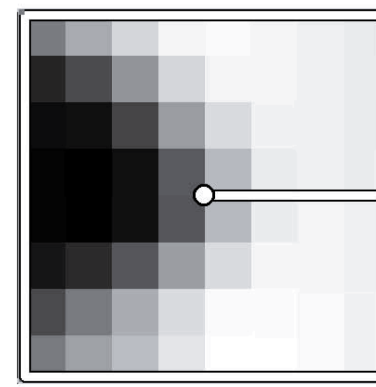
Brown et.al, "Multi-image matching using multi-scale oriented patches." *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 1. IEEE, 2005.

Feature-based approach

2. Extract feature descriptor

- 8x8 pixel patches patch around each detected feature to form a 64-dimensional descriptor (image intensity itself).
- sample from the larger 40x40 window to have a nice large blurred descriptor.

- bias/gain-normalize intensity.
- Invariant to intensity change, but sensitive to scale change, rotation, not a problem for the basic case.



Mikolajczyk et al, "A performance evaluation of local descriptors." *Pattern Analysis and Machine Intelligence, IEEE Transactions on*27.10 (2005): 1615-1630.

Feature-based approach

3. Match feature descriptors between two images

$$distance(i,j) = ||d_1(i) - d_2(j)||_2$$

Fixed i in image 1 and search for all the j in image 2

$$\text{match: } \frac{\min(\text{distance}, 1)}{\min(distance, 2)} < threshold$$

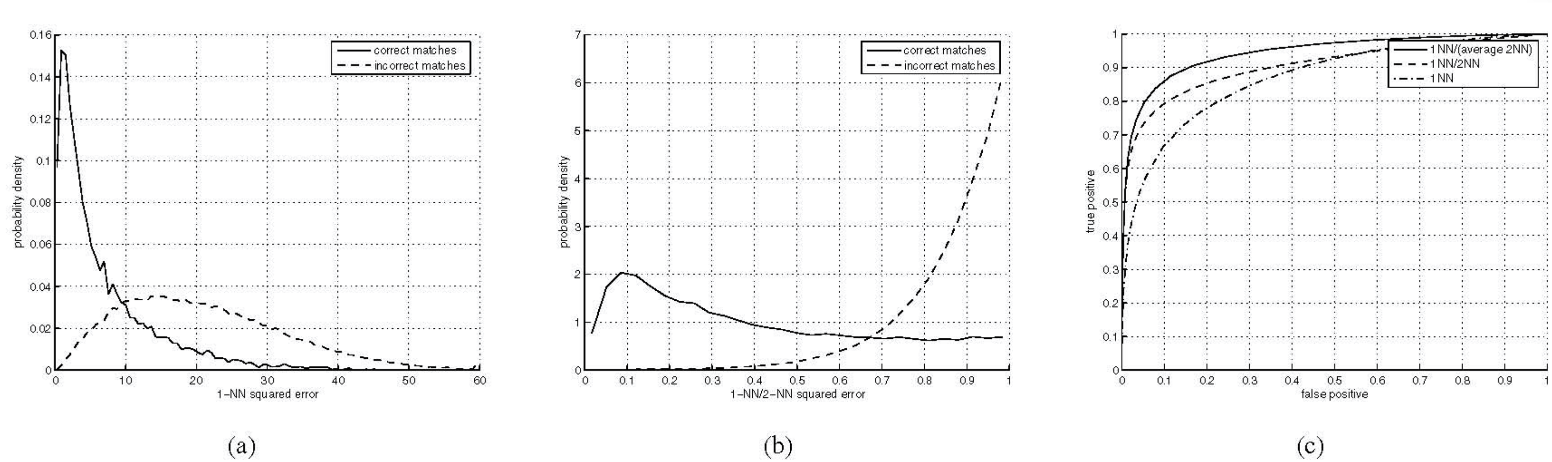
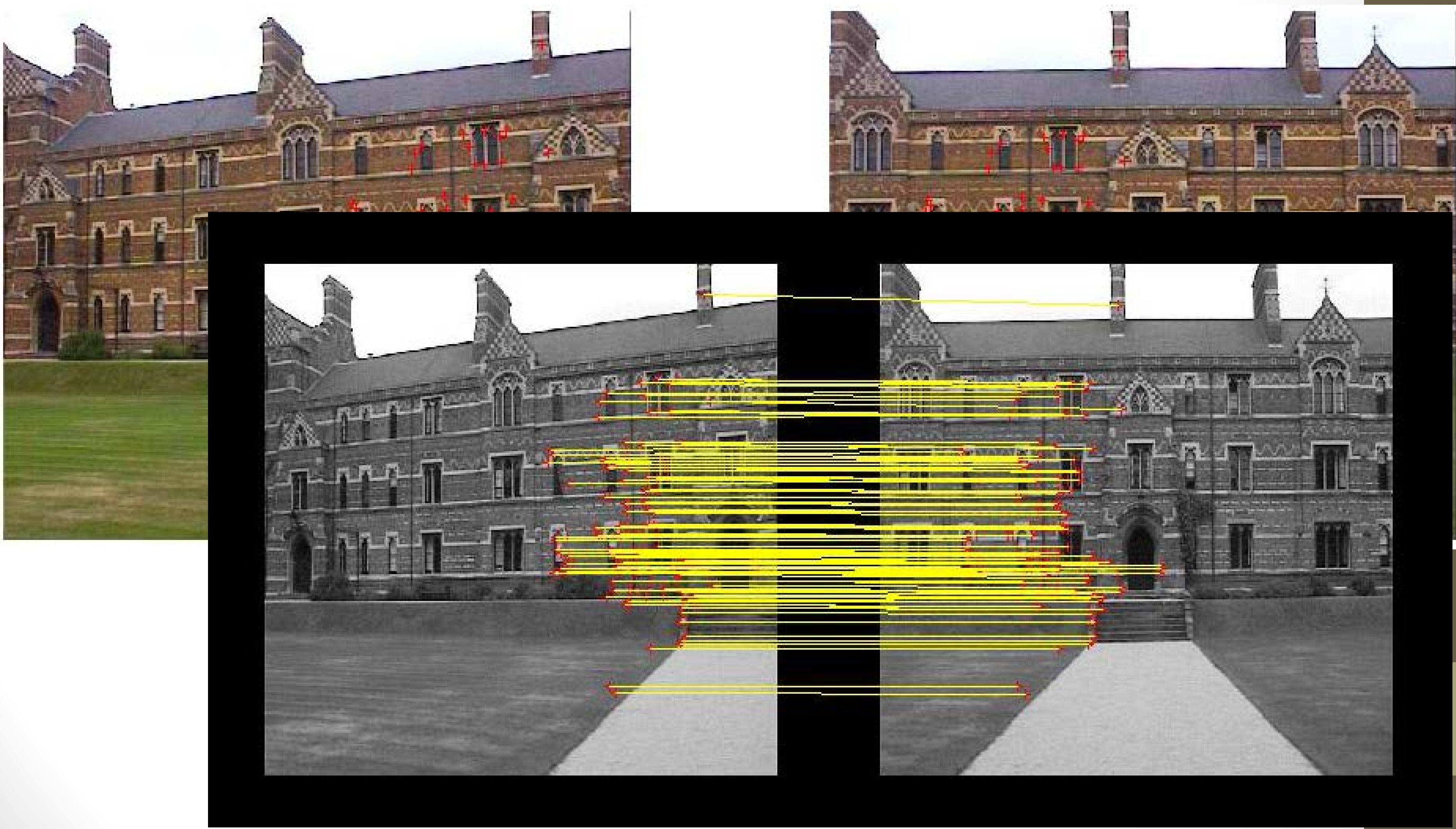


Figure 6. Distributions of matching error for correct and incorrect matches. Note that the distance of the closest match (the 1-NN) is a poor metric for distinguishing whether a match is correct or not (figure (a)), but the ratio of the closest to the second closest (1-NN/2-NN) is a good metric (figure (b)). We have found that using an average of 2-NN distances from multiple images (1NN/(average 2-NN)) is an even better metric (figure (c)). These results were computed from 18567 features in 20 images of the Abbey dataset, and have been verified for several other datasets.

Lowe, David G. "Distinctive image features from scale-invariant keypoints."*International journal of computer vision* 60.2 (2004): 91-110.

Feature-based approach

3. Match feature descriptors between two images



Feature-based approach

4. Estimate homography using RANSAC

$$\begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \Leftrightarrow \mathbf{x}_2 = H\mathbf{x}_1$$

$$\begin{aligned} \mathbf{h} &= (H_{11}, H_{12}, H_{13}, H_{21}, H_{22}, H_{23}, H_{31}, H_{32}, H_{33})^T \\ \mathbf{a}_x &= (-x_1, -y_1, -1, 0, 0, 0, x_2' x_1, x_2' y_1, x_2')^T \\ \mathbf{a}_y &= (0, 0, 0, -x_1, -y_1, -1, y_2' x_1, y_2' y_1, y_2')^T . \end{aligned}$$

$$\mathbf{a}_x^T \mathbf{h} = \mathbf{0}$$
$$\mathbf{a}_y^T \mathbf{h} = \mathbf{0}$$

$$Ah = 0 \quad A = U\Sigma V^T \quad h = v_n$$

note written by David Kriegman for Computer Vision I CSE 252A, UCSD, Winter 2007

Feature-based approach

4. Estimate homography using RANSAC

RANdom Sample Consensus, or RANSAC

- Basic idea: first find a good starting set of *inlier* correspondences, i.e., points that are all consistent with some particular motion estimate

1. random selecting a subset of *k* correspondences
2. use it to compute a motion estimate H
3. counts the number of *inliers* that are within of their predicted location, i.e., $||r_i|| \leq \epsilon$
4. repeated *S* times, and the sample set with largest number of inliers is kept as the final solution.

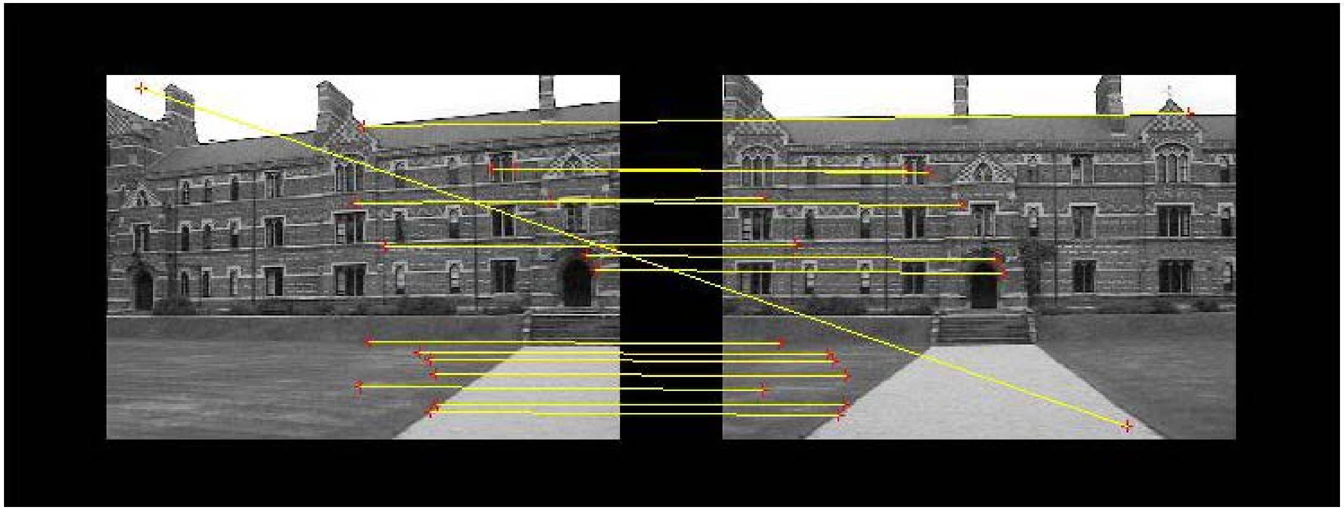
- In our implementation k=4, S = 50, $\epsilon = 3$

Fischler et al "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography." *Communications of the ACM* 24.6 (1981): 381-395.

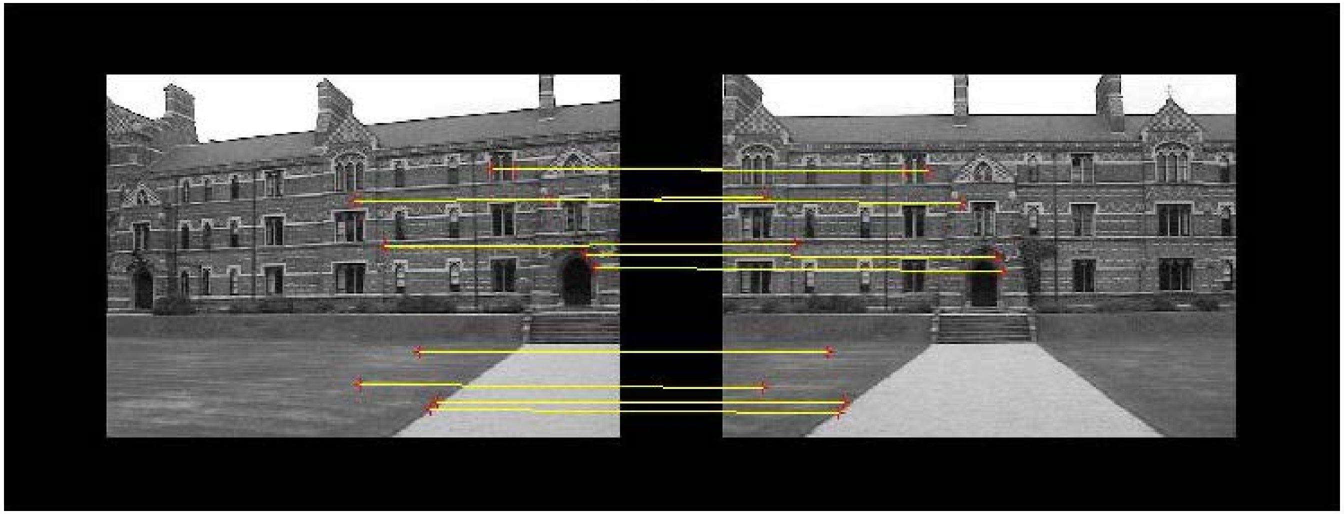
Feature-based approach

4. Estimate homography using RANSAC

all the matched features



only inliers

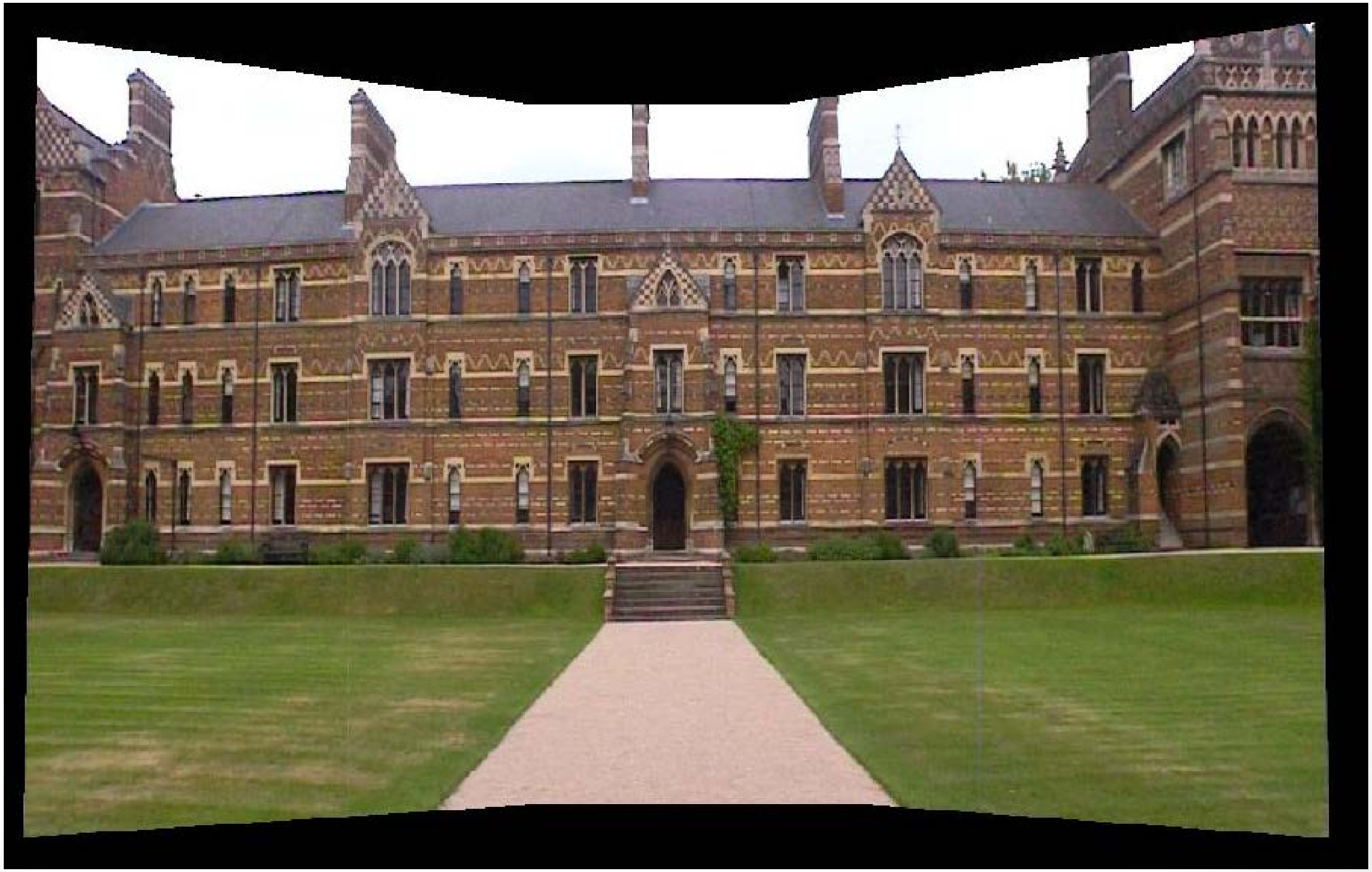


Note the robustness to pick out the real match features

Feature-based approach

5. Warping and compositing

using the estimated homography matrix H to warp
Choose the maximum value where pixels overlap



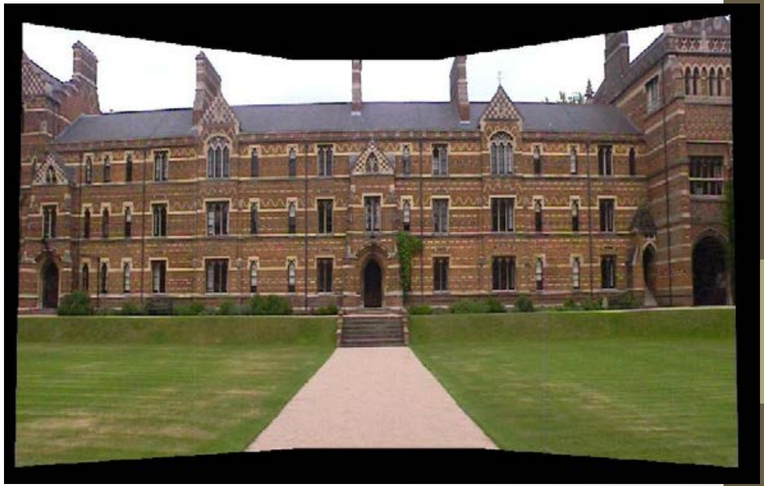
Feature-based approach

6. Speed up

Use the Adaptive Non-Maximal Suppression algorithm
# of Features ~ 100
time: 8.97 s

Control (use all the features)

# of Features ~ 5000
time: 128.85 s



Compositing: putting everything together

- Select final compositing surface and reference image
- Select which pixels contribute to final composite and optimally blend them to minimize visible seams (exposure differences), blur (mis-registration), and ghosting (moving objects)

http://www.blamethemonkey.com/hdr-photography-panorama-tutorial

Summary

- Image stitching
- Choice of algorithms depends on complexity of image differences
- Trade off between robustness, accuracy, computation time