# Object Recognition from Local Scale-Invariant Features

References:

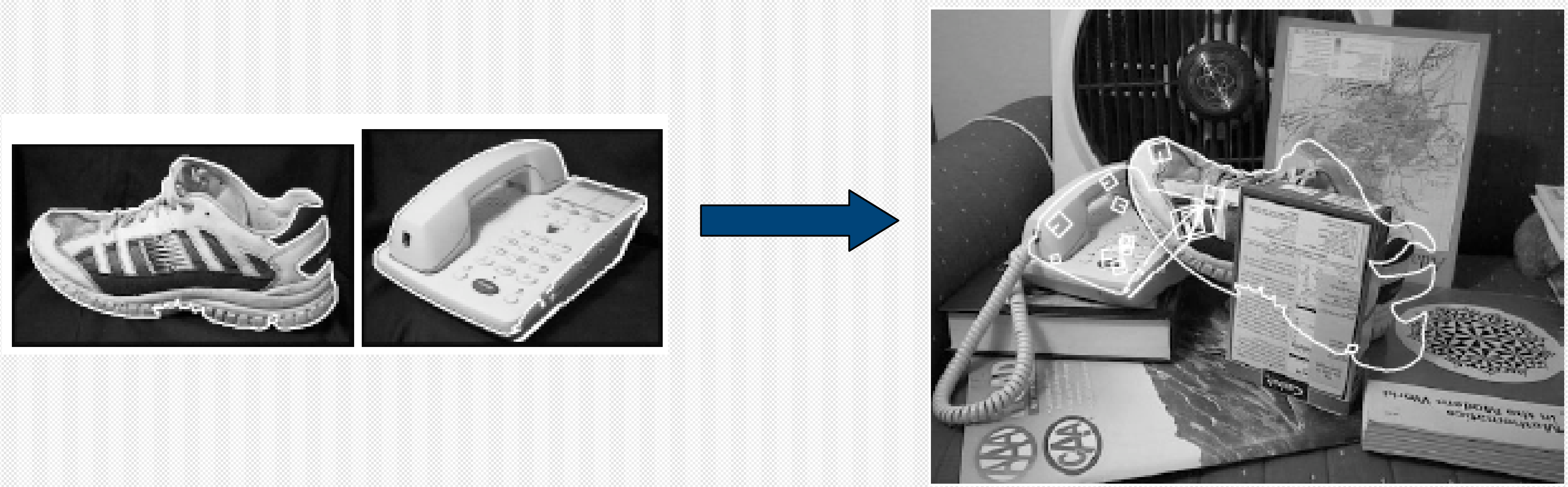
1. D. Lowe, Object recognition from local scale-invariant features. ICCV 1999.
2. M. Brown and D. Lowe, Recognising panoramas. ICCV 2003.

# Motivation

- Recognizing planar patterns is an important and challenging task.
- Requires local image features that are unaffected by
  - Nearby clutter
  - Partial occlusion
  - Invariant to illumination
  - Scaling, translation, and rotation

# Main Contributions

- A new feature generation method: SIFT.
- An unknown image is transformed into a collection of local feature vectors with invariant properties.
- Object models are represented as 2-D locations of the SIFT features that can undergo affine projection.



# SIFT Feature Detection

- Goal: Identify locations in image scale space that are invariant to translation, scaling, rotation, and small distortions.
- Algorithm outline:

Key Localization → Local Description → Matching

# SIFT: Selecting “key” locations

- Key locations are defined as maxima and minima of a difference-of-Gaussian function applied in scale space.
- Gaussian scale space.

Fig. 1. Shoreline of Africa.

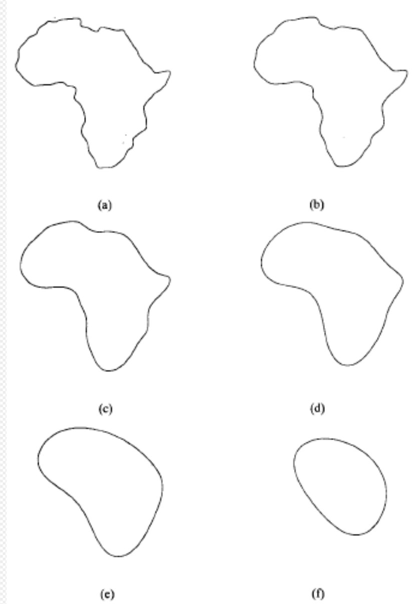



Fig. 2. Africa during evolution: (a) $\sigma = 2$; (b) $\sigma = 4$; (c) $\sigma = 8$; (d) $\sigma = 16$; (e) $\sigma = 32$; (f) $\sigma = 64$.

- Reference:F. Mokhtarian, A theory of multiscale, curvature-based shape representation for planar curves. PAMI, 92.

  D. Lowe. Organization of smooth image curves at multiple scales. IJCV 89

# Difference-of-Gaussian

- Difference-of-Gaussian (DoG) on image I
  - STEP 1. Convolve I with Gaussian kernel with $\sigma=\sqrt{2}$ --> A.
  - STEP 2. Convolve A with Gaussian kernel with $\sigma=\sqrt{2}$ --> B.
  - STEP 3. Define $f(\sqrt{2}) = DoG(I, \sigma) = A - B$.
- To generate the next pyramid level
  - STEP 1. Interpolate B with a pixel spacing of 1.5 --> I'
  - STEP 2. Compute $f(2) = DoG(I', \sigma)$.
- The key location is defined to be an extremum location at all pyramid levels.

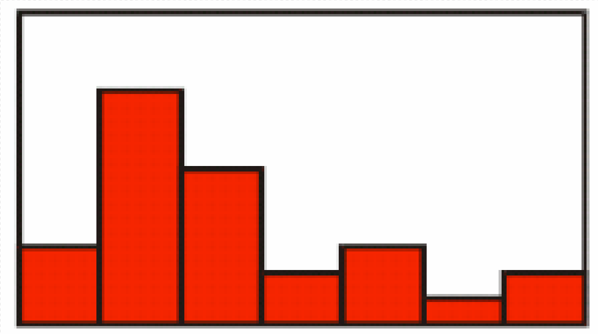
# SIFT Key Stability

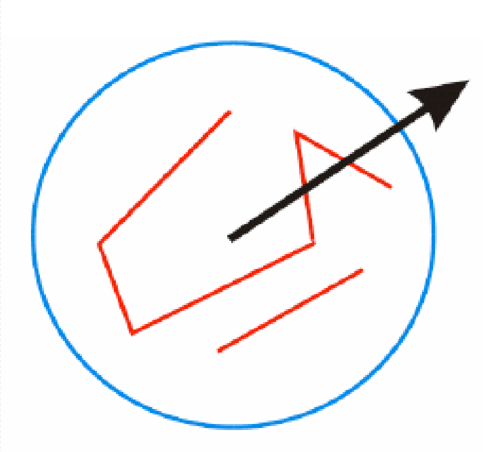
- Compute the gradient magnitude M and orientation R at each pixel.

$$M_{ij} = \sqrt{(A_{ij} - A_{i+1,j})^2 + (A_{ij} - A_{i,j+1})^2}$$

$$R_{ij} = \mathrm{atan2}\,(A_{ij} - A_{i+1,j}, A_{i,j+1} - A_{ij})$$

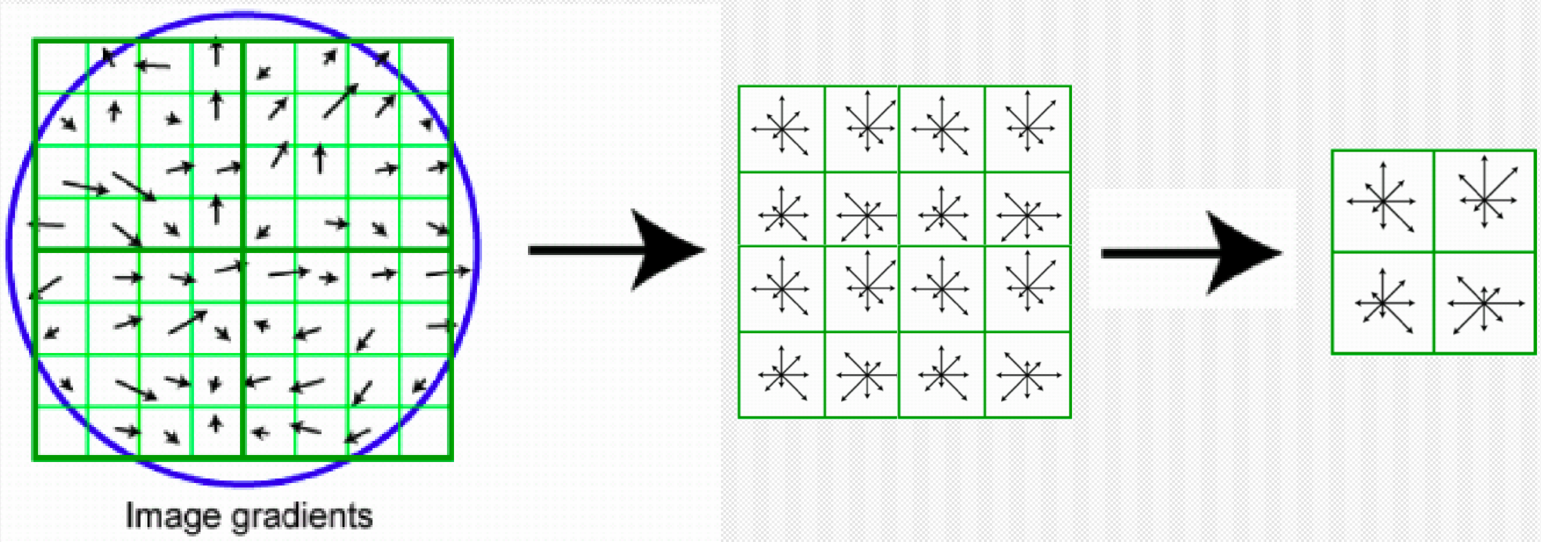
- Assign a canonical orientation at each key location.





# Local Image Description

- Obtain robust descriptor to local affine distortion.
  - For each detected keypoint, consider all pixels in the circle of radius 8 pixels.
  - 8 orientation planes, each sampled over a 2 by 2 grid.



# Indexing and Matching

- Store SIFT keys in sample images and identify matching keys from new images.
  - Difficulty: Identifying most similar keys for high-dimensional vectors.
  - Solution: Best-Bin-First search method.

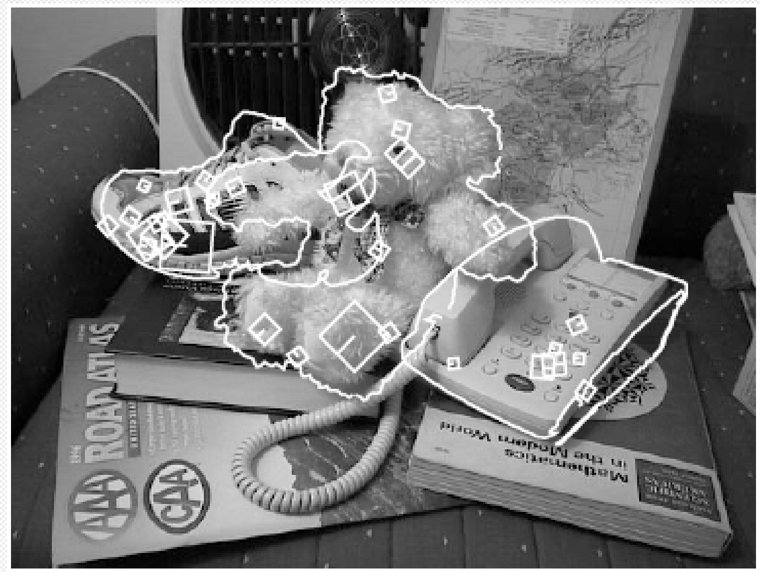
  [3] Beis, el at., *Shape indexing using approximate nearest-neighbour search in high-dimensional space*. CVPR 1997.


# Indexing and Matching

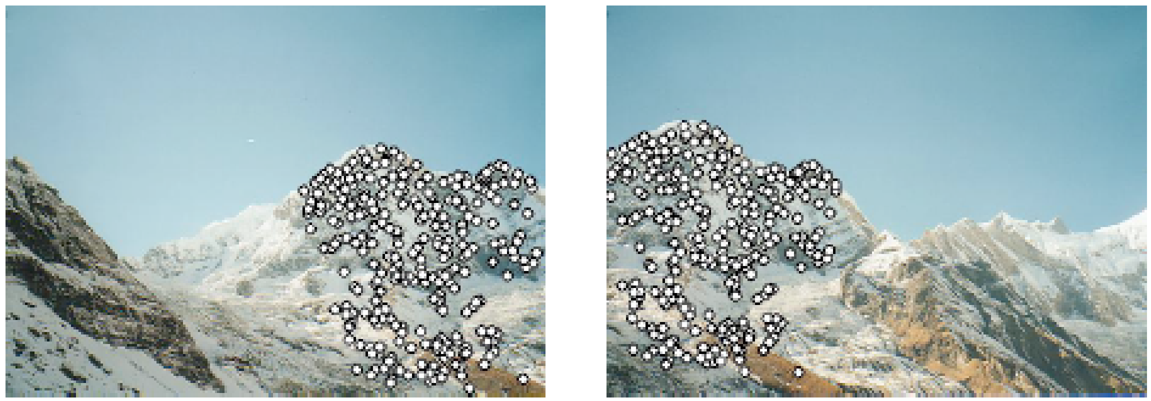
- After each key in the new image is mapped to a key in the example image, use Hough transform to vote for model hypotheses.
- After highest clusters identified in the hash table of the Hough transform, a verification procedure is applied to estimate the affine parameters of the model.
  - One key location provides 2 constraints. Affine model contains 6 unknowns. Therefore, need at least 3 keys per object.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

# Experiments

# Application: Recognizing Panoramas



(c) SIFT matches 1 (d) SIFT matches 2

# Panoramic Recognition Algorithm:

- STEP 1. SIFT feature extraction.
- STEP 2. Match SIFT features on k nearest images.
- STEP 3. Apply RANSAC to solve for the affine model.
- STEP 4. Bundle Adjustment optimization.
- STEP 5. Render panorama.

Advantages:

- Insensitive to the ordering, orientation, scale, and illuminations of the images.
- Insensitive to outlier images.

# RANSAC and Bundle Adjustment

- RANSAC (Random Sample Consensus)
  - Goal: eliminate outlier effects by measuring the consensus of a minimal subsampling set.
  - Repeatedly subsample a subset of samples of minimal size.
  - Estimate an affine model per subset.
  - Measure the consensus of the estimated model.
  - Select the optimal model with the highest consensus.
- Bundle Adjustment on Affine Projection
  - Goal: refining a visual reconstruction to produce a jointly optimal structure across multiple images.

References:

Fischler and Bolles. Random sample consensus. Graphics and Image Processing, 1981.

Triggs et al. Bundle adjustment. Vision Algorithms, 2000.

# Results

# THE END