ECG782: MULTIDIMENSIONAL DIGITAL SIGNAL PROCESSING

INTEREST POINTS

OUTLINE

- Interest Point Detection
  - Feature Detection
  - Feature Description
  - Feature Matching
  - Feature Tracking
- Maximally Stable Regions

DETERMINING FEATURES TO MATCH

- What can help establish correspondences between images?

DETECTION OF CORNERS (INTEREST POINTS)

- Useful for fundamental vision techniques
  - Image matching or registration
- Correspondence problem needs to find all pairs of matching pixels
  - Typically a complex problem
  - Can be made easier only considering a subset of points
- Interest points are these important image regions that satisfy some local property
  - Corners are a way to get to interest points/keypoints

FEATURE DETECTION AND MATCHING

- Essential component of computer vision
  - E.g. alignment for image stitching, correspondences for 3D model construction, object detection, stereo, etc.

- Need to establish some features that can be detected and matched

DIFFERENT TYPES OF FEATURES

- (a) keypoints/interest points/corners
- (b) regions
- (c) edges
- (d) straight lines

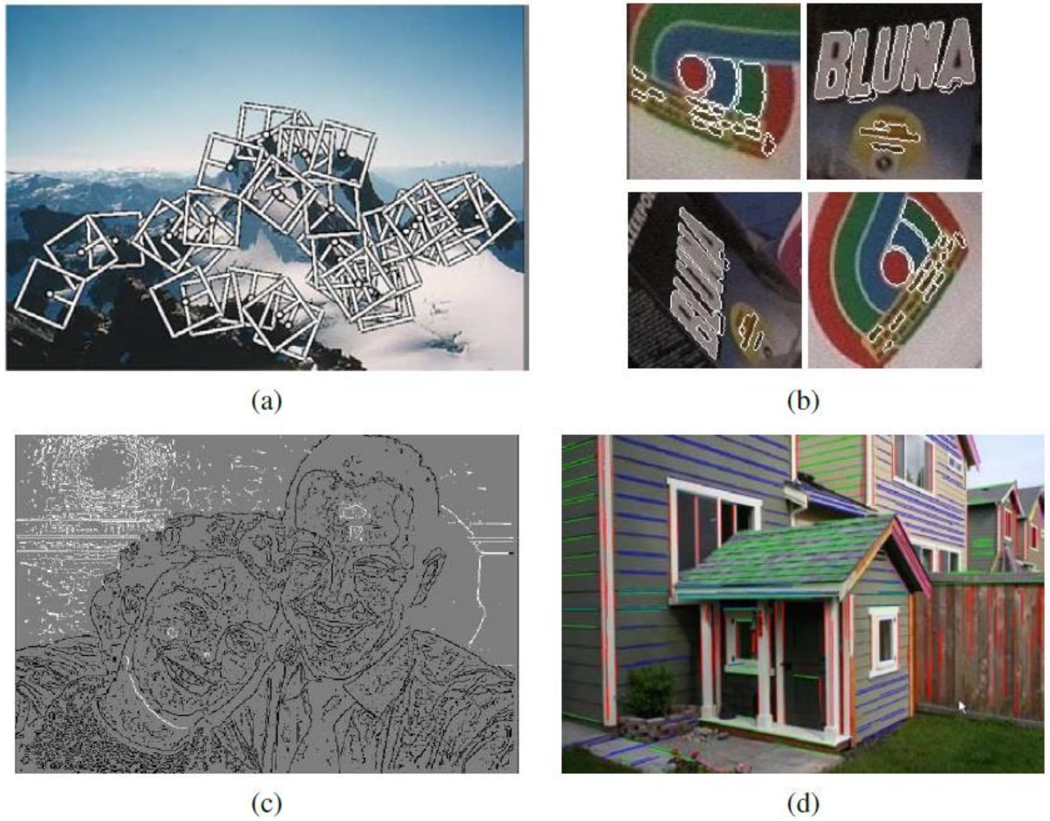


Figure 4.1 A variety of feature detectors and descriptors can be used to analyze, describe and match images: (a) point-like interest operators (Brown, Szeliski, and Winder 2005) © 2005 IEEE; (b) region-like interest operators (Matas, Chum, Urban *et al.* 2004) © 2004 Elsevier; (c) edges (Elder and Goldberg 2001) © 2001 IEEE; (d) straight lines (Sinha, Steedly, Szeliski *et al.* 2008) © 2008 ACM.

DIFFERENT TYPES OF FEATURES

- Points and patches
- Edges
- Lines

- Which features are best?
  - Depends on the application
- Want features that are robust
  - Descriptive and consistent (can readily detect)

POINTS AND PATCHES

- Maybe most generally useful feature for matching
  - E.g. Camera pose estimation, dense stereo, image stitching, video stabilization, tracking
  - Object detection/recognition

- Key advantages:
  - Matching is possible even in the presence of clutter (occlusion)
  - Can handle large scale and orientation changes

POINT CORRESPONDENCE TECHNIQUES

- Detection and tracking
  - Initialize by detecting features in a single image
  - Track features through localized search
  - Best for images from similar viewpoint or video
- Detection and matching
  - Detect features in all images
  - Match features across images based on local appearance
  - Best for large motion or appearance change

KEYPOINT PIPELINE

- Feature detection (extraction)
  - Search for image locations that are likely to be matched in other images
- Feature description
  - Regions around a keypoint are represented as a compact and stable descriptor
- Feature matching
  - Descriptors are compared between images efficiently
- Feature tracking
  - Search for descriptors in small neighborhood
  - Alternative to matching stage best suited for video

FEATURE DETECTORS

- Must determine image locations that can be reliably located in another image

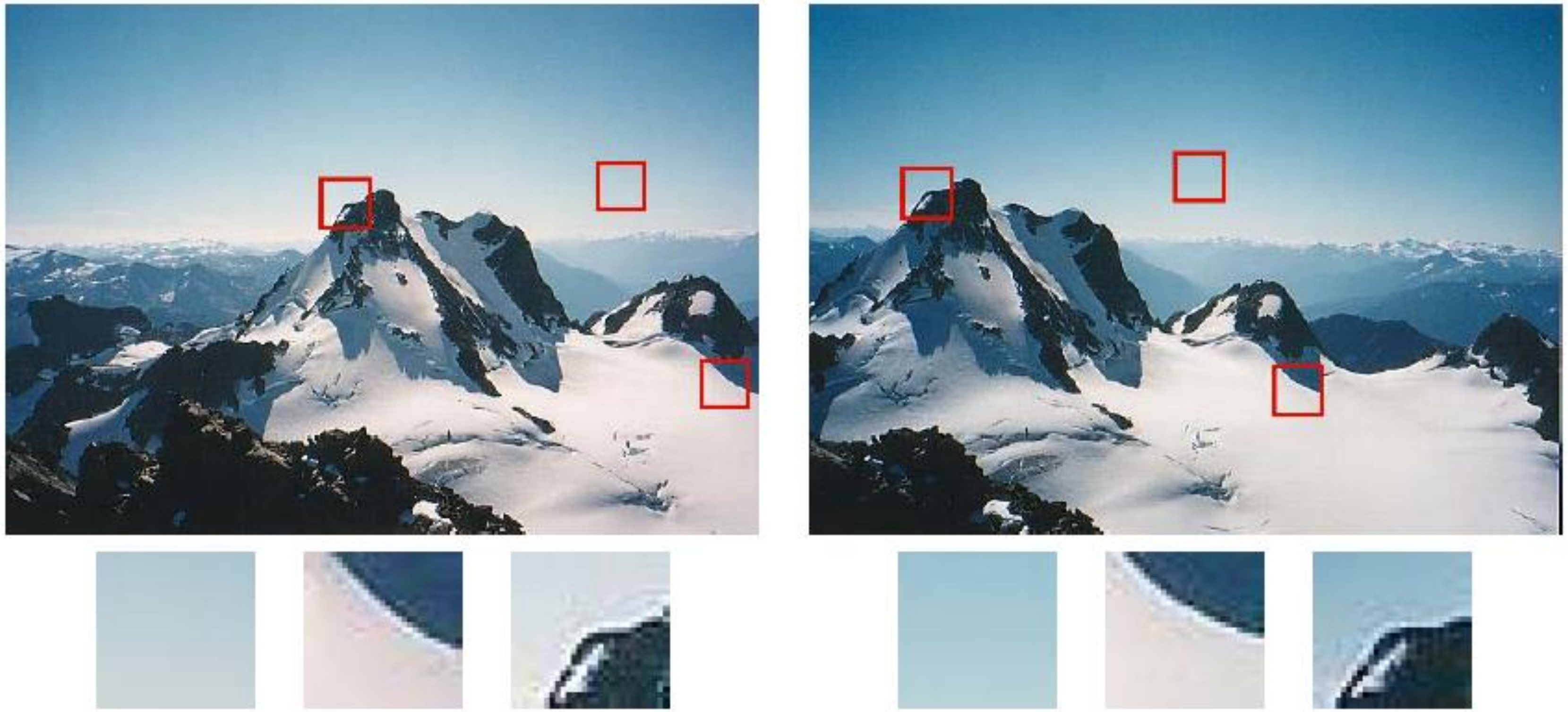


Figure 4.3 Image pairs with extracted patches below. Notice how some patches can be localized or matched with higher accuracy than others.

COMPARISON OF IMAGE PATCHES

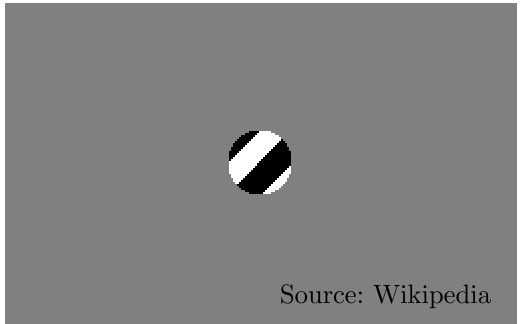
- Textureless patches
  - Nearly impossible to localize and match
    - Sky region "matches" to all other sky areas

- Edge patches
  - Large contrast change (gradient)
  - Suffer from aperture problem
    - Only possible to align patches along the direction normal the edge direction

- Corner patches
  - Contrast change in at least two different orientations
  - Easiest to localize

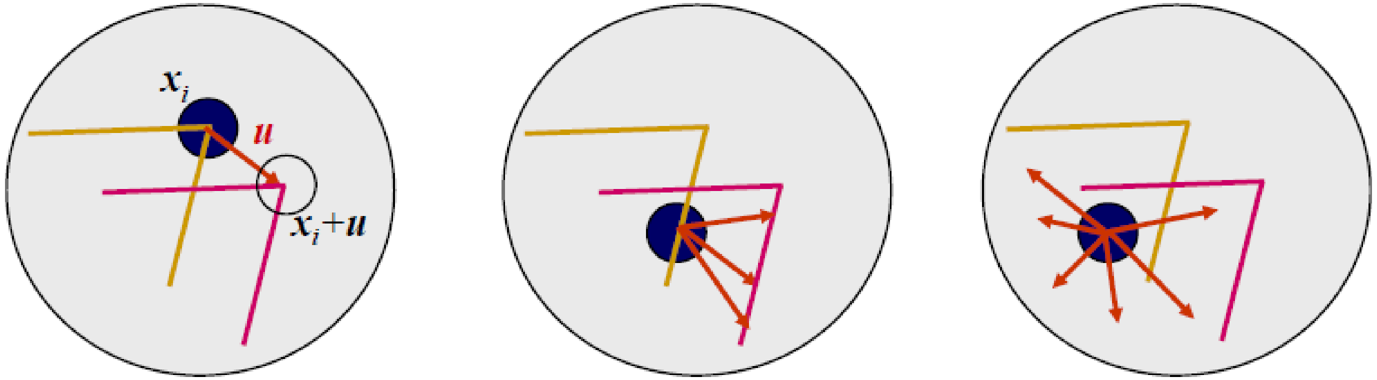
APERTURE PROBLEM I

- Only consider a small window of an image
  - Local view does not give global structure
  - Causes ambiguity
- Best visualized with motion (optical flow)
  - Imagine seeing the world through a straw hole
  - Aperture Problem – Demo
  - Also known as the barber pole effect



APERTURE PROBLEM II

- Corners have strong matches
- Edges can have many potential matches – Constrained upon a line
- Textureless regions provide no useful information



(a) (b) (c)

Figure 4.4 Aperture problems for different image patches: (a) stable ("corner-like") flow; (b) classic aperture problem (barber-pole illusion); (c) textureless region. The two images I_0 (yellow) and I_1 (red) are overlaid. The red vector $\boldsymbol{u}$ indicates the displacement between the patch centers and the $w(\boldsymbol{x}_i)$ weighting function (patch window) is shown as a dark circle.

WSSD MATCHING CRITERION

- Weighted summed squared difference
  - $E_{WSSD}(\boldsymbol{u}) = \sum_i w(\boldsymbol{x}_i)\,[I_1(\boldsymbol{x}_i - \boldsymbol{u}) - I_0(\boldsymbol{x}_i)]^2$
    - I_1, I_0 - two image patches to compare
    - $\boldsymbol{u} = (u, v)$ – displacement vector
    - $w(\boldsymbol{x})$ - spatial weighting function
- Normally we do not know the image locations to perform the match
- Calculate the autocorrelation in small displacements of a single image
  - Gives a measure of stability of patch
  - $E_{AC}(\Delta\boldsymbol{u}) = \sum_i w(\boldsymbol{x}_i)\,[I_0(\boldsymbol{x}_i - \Delta\boldsymbol{u}) - I_0(\boldsymbol{x}_i)]^2$

IMAGE PATCH AUTOCORRELATION

- $E_{AC}(\Delta\boldsymbol{u}) = \sum_i w(\boldsymbol{x}_i)\,[I_0(\boldsymbol{x}_i - \Delta\boldsymbol{u}) -$

- Example autocorrelation

IMAGE AUTOCORRELATION II

- (b) High texture
  - Alignment peak (min)
- (c) Line edge
  - Align along edge
- (d) Low texture
  - No clear alignment

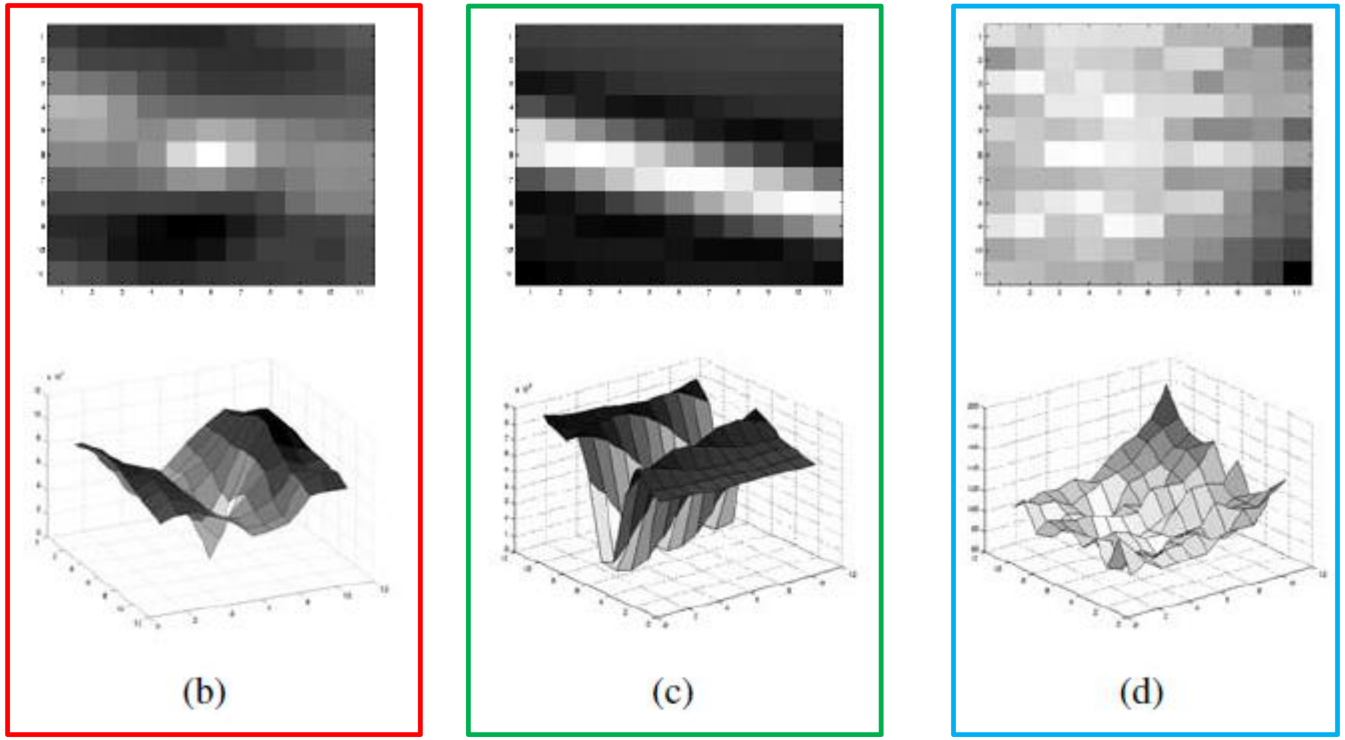
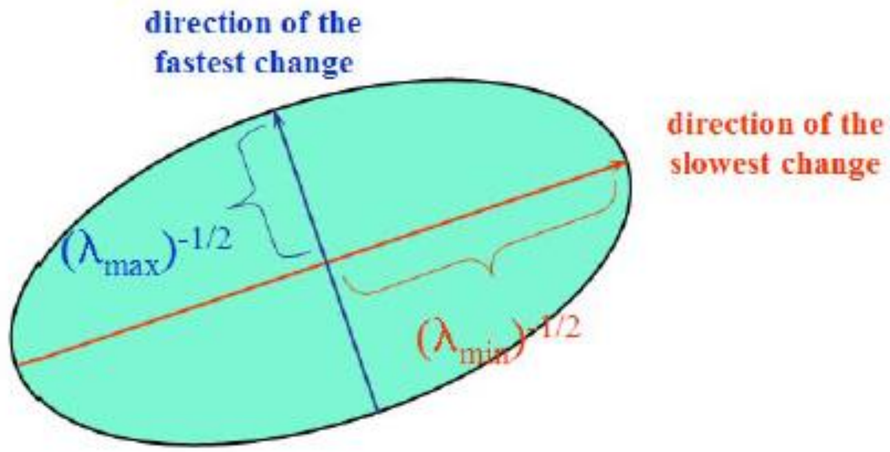


Figure 4.5 Three auto-correlation surfaces $E_{AC}(\Delta u)$ shown as both grayscale images and surface plots: (a) The original image is marked with three red crosses to denote where the auto-correlation surfaces were computed; (b) this patch is from the flower bed (good unique minimum); (c) this patch is from the roof edge (one-dimensional aperture problem); and (d) this patch is from the cloud (no good peak). Each grid point in figures b–d is one value of Δu.

IMAGE AUTOCORRELATION III

- The matrix A provides a measure of uncertainty in location of the patch
- Do eigenvalue decomposition
  - Get eigenvalues and eigenvector directions
- Good features have both eigenvalues large
  - Indicates gradients in orthogonal directions (e.g. a corner)
- Uncertainty ellipse



- Many different methods to quantify uncertainty
  - Easiest: look for maxima in the smaller eigenvalue

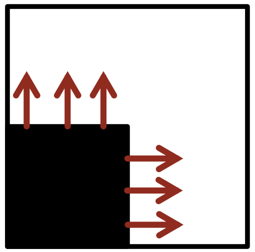
BASIC FEATURE DETECTION ALGORITHM

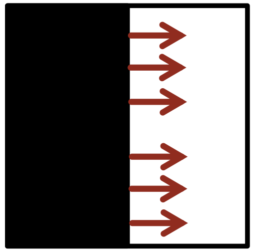
1. Compute the horizontal and vertical derivatives of the image I_x and I_y by convolving the original image with derivatives of Gaussians (Section 3.2.3).
2. Compute the three images corresponding to the outer products of these gradients. (The matrix $\boldsymbol{A}$ is symmetric, so only three entries are needed.)
3. Convolve each of these images with a larger Gaussian.
4. Compute a scalar interest measure using one of the formulas discussed above.
5. Find local maxima above a certain threshold and report them as detected feature point locations.

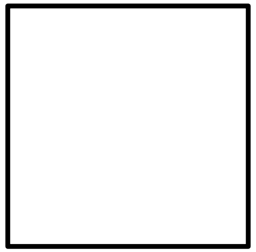
Algorithm 4.1 Outline of a basic feature detection algorithm.

INTEREST POINT DETECTION

- The correlation matrix gives a measure of edges in a patch
- Corner
  - Gradient directions
    - $\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}$
  - Correlation matrix
    - $A \propto \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$
- Edge
  - Gradient directions
    - $\begin{bmatrix}1\\0\end{bmatrix}$
  - Correlation matrix
    - $A \propto \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}$
- Constant
  - Gradient directions
    - $\begin{bmatrix}0\\0\end{bmatrix}$
  - Correlation matrix
    - $A \propto \begin{bmatrix}0 & 0\\0 & 0\end{bmatrix}$







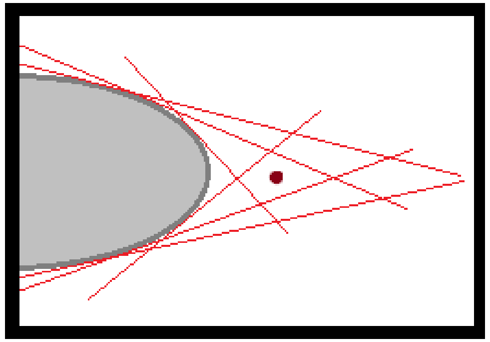
HARRIS CORNERS

IMPROVING FEATURE DETECTION

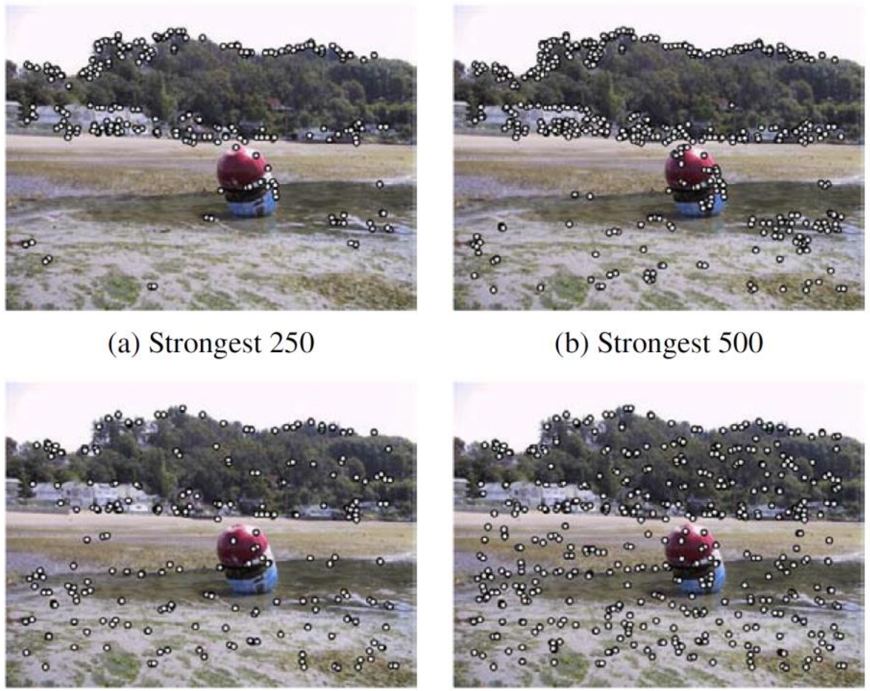
- Multiple responses to a corner
- Scale invariance
- Rotational invariance and orientation estimation
- Affine invariance

MULTIPLE CORNER REPONSE

- Corners may produce more than one strong response (due to neighborhood)
- Estimate corner with subpixel accuracy – use edge tangents (Förstner)



- Non-maximal suppression – only select features that are far enough away
  - Create more uniform distribution – can be done through blocking as well



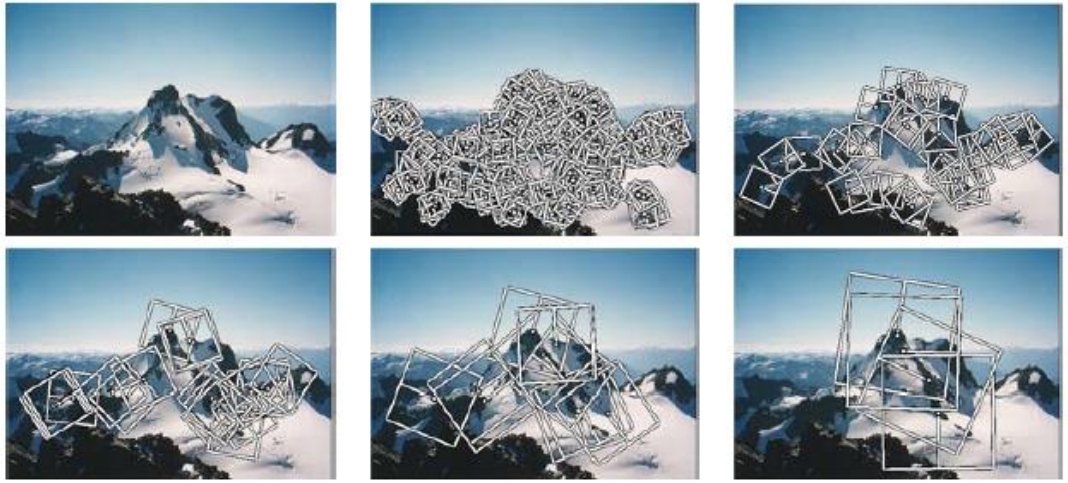
(a) Strongest 250 (b) Strongest 500

(c) ANMS 250, $r = 24$ (d) ANMS 500, $r = 16$

Figure 4.9 Adaptive non-maximal suppression (ANMS) (Brown, Szeliski, and Winder 2005) © 2005 IEEE: The upper two images show the strongest 250 and 500 interest points, while the lower two images show the interest points selected with adaptive non-maximal suppression, along with the corresponding suppression radius r. Note how the latter features have a much more uniform spatial distribution across the image.

SCALE INVARIANCE

- Use an image pyramid
  - Represent at different scale through subsampling
- Compute Hessian of difference of Gaussian (DoG) image
  - Laplacian of Gaussian (LoG) as another variant
- Analyze scale space [SIFT – Lowe 2004]
  - Sub-octave (quarter-octave) pyramid
  - Will see more later



(a) (b) (c)

Figure 4.8 Interest operator responses: (a) Sample image, (b) Harris response, and (c) DoG response. The circle sizes and colors indicate the scale at which each interest point was detected. Notice how the two detectors tend to respond at complementary locations.

ROTATIONAL INVARIANCE

- Need to estimate the orientation of the feature by examining gradient information
  - Find dominant orientation
- Average gradient in neighborhood of keypoint
  - Simplest possible solution
- Histogram of orientations
  - More later with SIFT and HOG

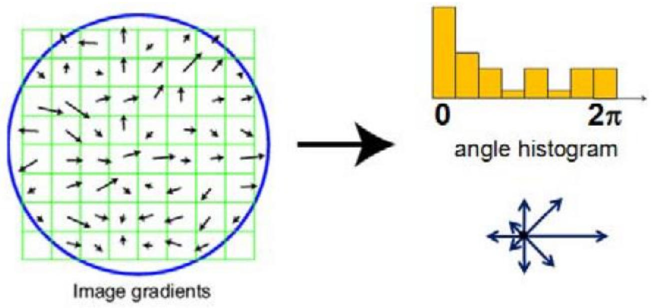


Figure 4.12 A dominant orientation estimate can be computed by creating a histogram of all the gradient orientations (weighted by their magnitudes or after thresholding out small gradients) and then finding the significant peaks in this distribution (Lowe 2004) © 2004 Springer.

AFFINE INVARIANCE

- Handle scale and orientation but also affine deformations
  - Local perspective distortion for small patches
  - Closer to appearance change due to perspective distortion
- Fit ellipse to autocorrelation matrix and use it as an affine coordinate frame
- Maximally stable region (MSER) [Matas 2004] – regions that do not change much through thresholding

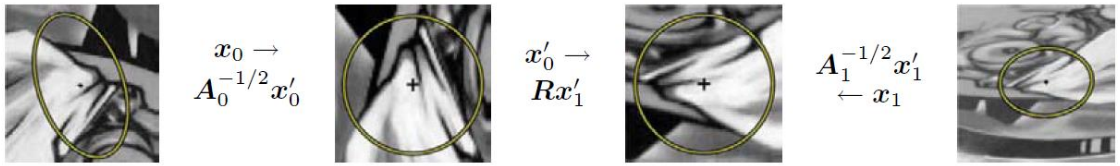


Figure 4.14 Affine normalization using the second moment matrices, as described by Mikolajczyk, Tuytelaars, Schmid *et al.* (2005) © 2005 Springer. After image coordinates are transformed using the matrices $A_0^{-1/2}$ and $A_1^{-1/2}$, they are related by a pure rotation R, which can be estimated using a dominant orientation technique.

Figure 4.15 Maximally stable extremal regions (MSERs) extracted and matched from a number of images (Matas, Chum, Urban *et al.* 2004) © 2004 Elsevier.

FEATURE DESCRIPTORS

- After keypoint detection, must match
- Small motion and limited appearance change
  - Use simple error metrics (SSD, NCC)
- Generally, rotation, orientation, and affine deformations
  - Need to compensate for distortion
- Invariant descriptors
  - Lots of variants

- MOPS

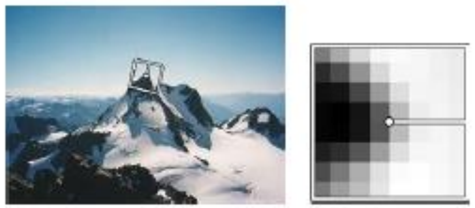
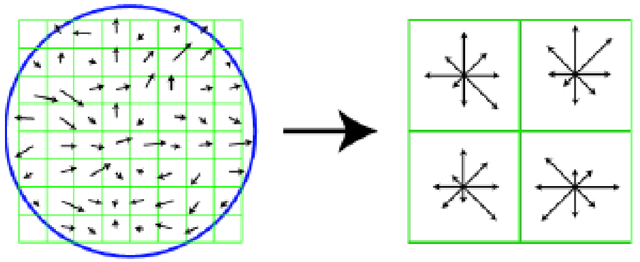


Figure 4.17 MOPS descriptors are formed using an 8×8 sampling of bias and gain nor-

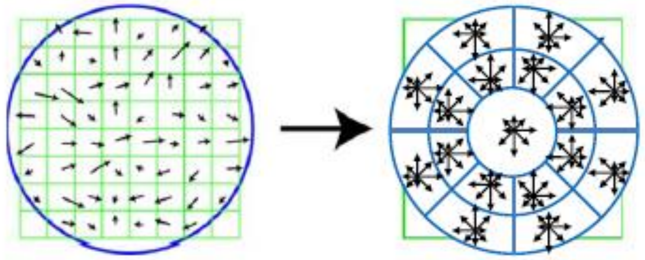
- SIFT



(a) image gradients (b) keypoint descriptor

Figure 4.18 A schematic representation of Lowe's (2004) scale invariant feature transform

- GLOH



(a) image gradients (b) keypoint descriptor

Figure 4.19 The gradient location-orientation histogram (GLOH) descriptor uses log-polar

FEATURE MATCHING

- Given extracted features, match
- Euclidean distance (vector magnitude) for ranking
- Simple threshold to get candidates
  - Hard to select global threshold
- Nearest neighbor matching
  - Handles variation within scene (distance from camera)
  - Use threshold to prune

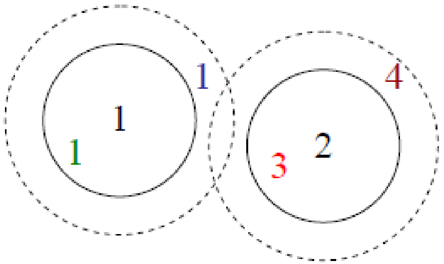


Figure 4.22 False positives and negatives: The black digits 1 and 2 are features being matched against a database of features in other images. At the current threshold setting (the solid circles), the green 1 is a *true positive* (good match), the blue 1 is a *false negative* (failure to match), and the red 3 is a *false positive* (incorrect match). If we set the threshold higher (the dashed circles), the blue 1 becomes a true positive but the brown 4 becomes an additional false positive.

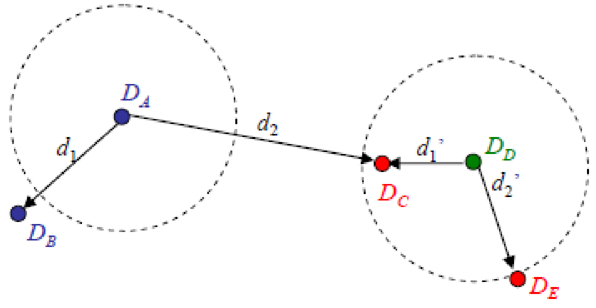


Figure 4.24 Fixed threshold, nearest neighbor, and nearest neighbor distance ratio matching. At a fixed distance threshold (dashed circles), descriptor D_A fails to match D_B and D_D incorrectly matches D_C and D_E. If we pick the nearest neighbor, D_A correctly matches D_B but D_D incorrectly matches D_C. Using nearest neighbor distance ratio (NNDR) matching, the small NNDR d_1/d_2 correctly matches D_A with D_B, and the large NNDR d_1'/d_2' correctly rejects matches for D_D.

QUANTIFYING PERFORMANCE

- Confusion matrix-based metrics
  - Binary {1,0} classification tasks

		actual value		
predicted outcome		p	n	total
	p'	TP	FP	P'
	n'	FN	TN	N'
	total	P	N	

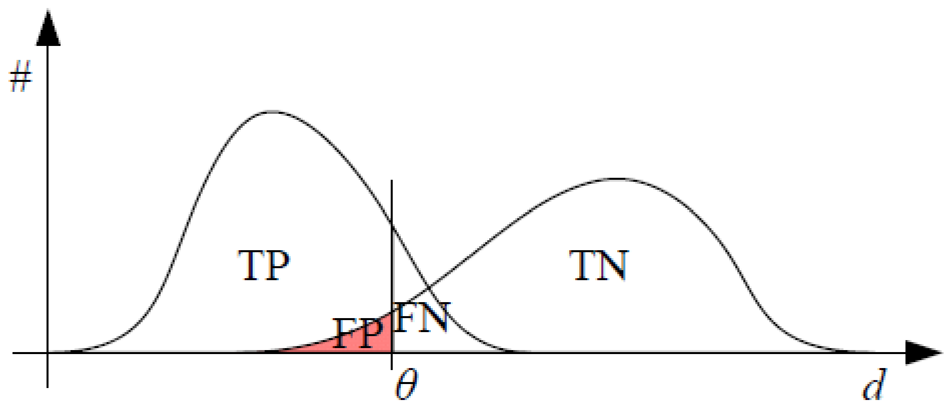
- True positives (TP) - # correct matches
- False negatives (FN) - # of missed matches
- False positives (FP) - # of incorrect matches
- True negatives (TN) - # of non-matches that are correctly rejected

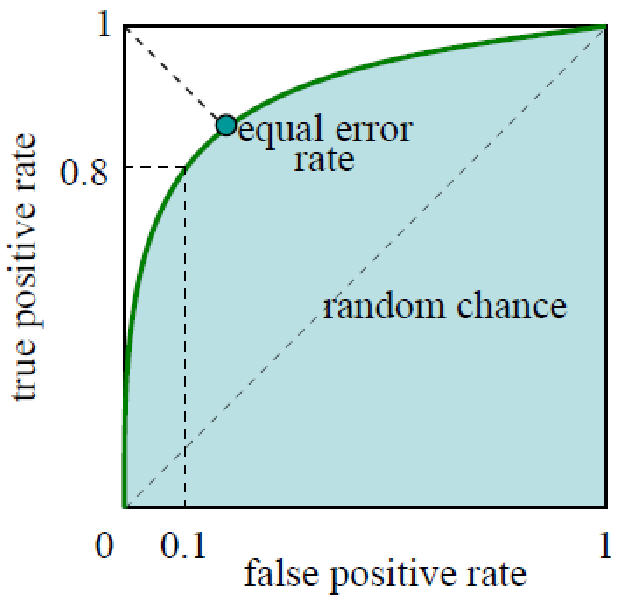
- A wide range of metrics can be defined

- True positive rate (TPR) (sensitivity)
  - $TPR = \frac{TP}{TP+FN} = \frac{TP}{P}$
  - Document retrieval → recall – fraction of relevant documents found
- False positive rate (FPR)
  - $FPR = \frac{FP}{FP+TN} = \frac{FP}{N}$
- Positive predicted value (PPV)
  - $PPV = \frac{TP}{TP+FP} = \frac{TP}{P'}$
  - Document retrieval → precision – number of relevant documents are returned
- Accuracy (ACC)
  - $ACC = \frac{TP+TN}{P+N}$

RECEIVER OPERATING CHARACTERISTIC (ROC)

- Evaluate matching performance based on threshold
  - Examine all thresholds θ to map out performance curve
- Best performance in upper left corner
  - Area under the curve (AUC) is a ROC performance metric





FEATURE TRACKING

- Similar to feature detection-matching
  - Find good points and then follow them in subsequent images
- Useful for video
  - Small motion and appearance change → NCC
- With longer tracking → more appearance change
  - Deal with affine motion model

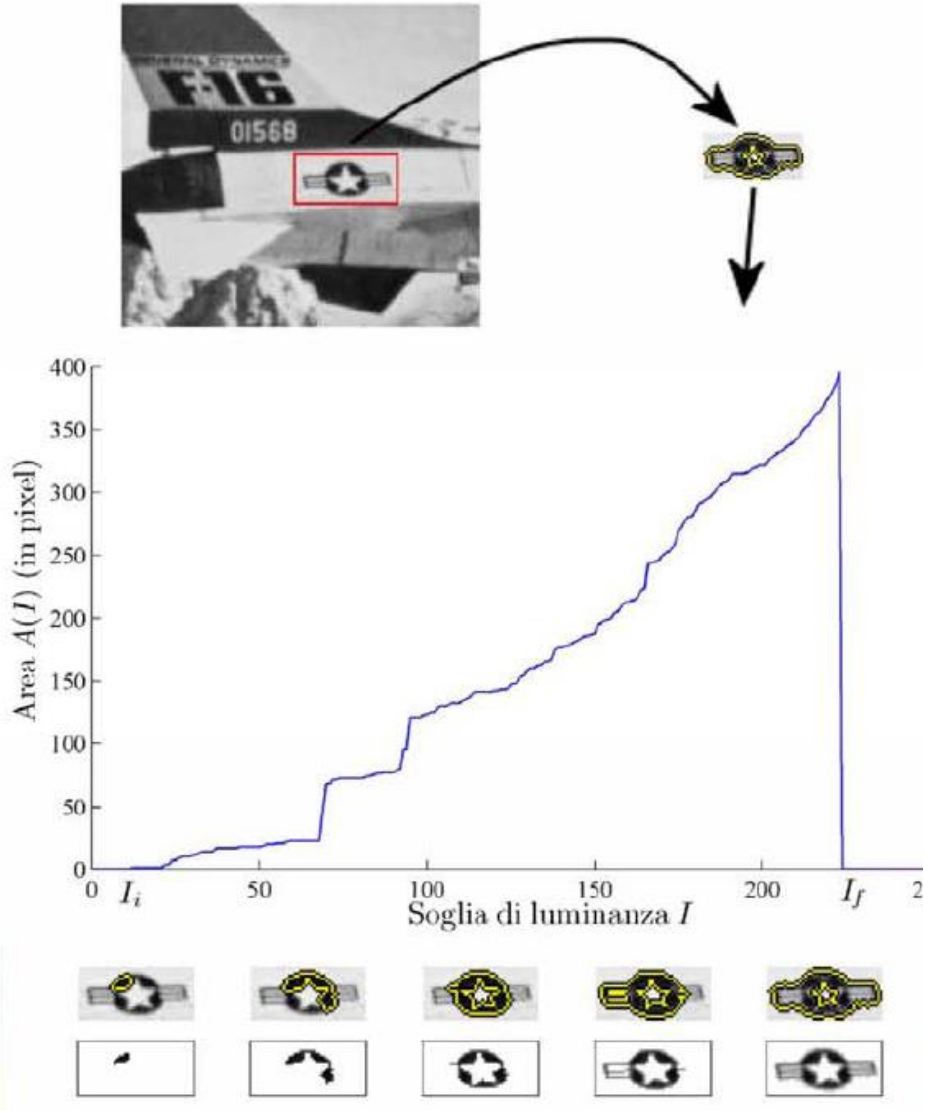
- Kanade-Lucas-Tomasi (KLT) tracker first estimates motion and then does affine warp
  - Search performed in area around predicted location

Figure 4.28 Feature tracking using an affine motion model (Shi and Tomasi 1994) © 1994 IEEE, Top row: image patch around the tracked feature location. Bottom row: image patch after warping back toward the first frame using an affine deformation. Even though the speed sign gets larger from frame to frame, the affine transformation maintains a good resemblance between the original and subsequent tracked frames.

MAXIMALLY STABLE EXTREMAL REGIONS

- MSERs are image structures that can be recovered after translations, rotations, similarity (scale), and affine (shear) transforms
- Connected areas characterized by almost uniform intensity, surrounded by contrasting background
- Constructed based on a watershed-type segmentation
  - Threshold image a multiple different values
  - MSERs are regions with shape that does not change much over thresholds
- Each region is a connected component but no global or optimal threshold is selected

MSER



- Red borders from increasing intensity
- Green boarders from decreasing intensity

MSER INVARIANCE

- Fit ellipse to area and normalize into circle