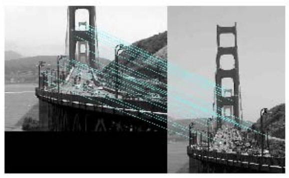
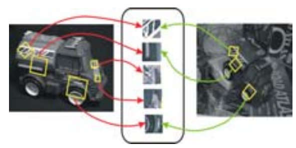
EE 6882
Visual Search Engine

Prof. Shih-Fu Chang, Feb. 6th 2012
Lecture #3

- Evaluation Metrics
- Local Features: Corner Detector, SIFT
- Local Feature Matching

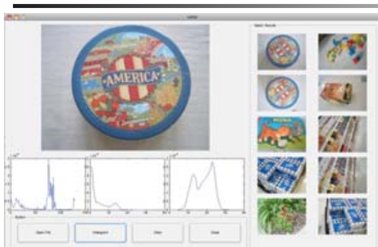


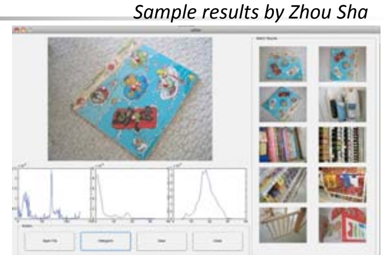


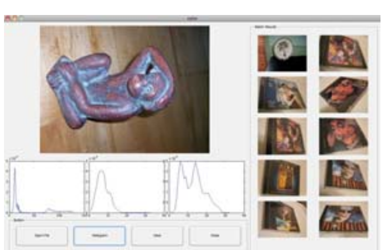
(Many slides from A. Efors, W. Freeman, C. Kambhamettu, L. Xie, and likely others)
(Slides preparation assisted by Rong-Rong Ji)

Performance evaluation

Sample results by Zhou Sha







Given a search result list, how to measure performance?

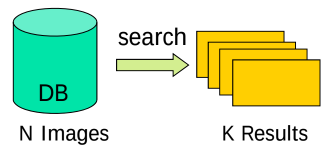
2

Evaluation

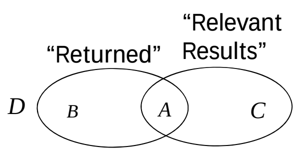
<u>Ground truth</u>

$V_n = 1$ "Relevant"

$\quad\; 0$ "Irrelevant" $\quad n = 0 \cdots N-1$



- Detection $A = \sum_{n=0}^{K-1} V_n$
- False Alarms $B = \sum_{n=0}^{K-1} (1 - V_n)$
- Misses $C = (\sum_{n=0}^{N-1} V_n) - A$
- Correct Dismissals $D = (\sum_{n=0}^{N-1} (1 - V_n)) - B$



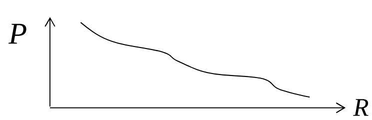
- Recall $R = A/(A + C)$
- Precision $P = A/(A+B)$
- False $F = B/(A+B)$
- Combined $F_1 = \frac{P \cdot R}{(P+R)/2}$

Evaluation Measures

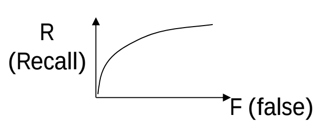
Precision at depth K

$$P_k = (\textstyle\sum_{n=0}^{k-1} V_n) / K$$

Precision Recall Curve



Receiver Operating Characteristic (ROC Curve)



Average Precision

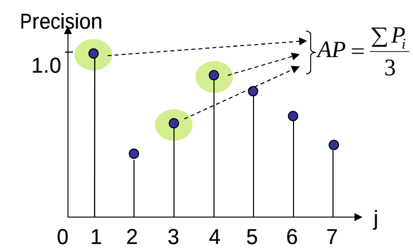
AP approximates areas under PR curve

$$AP = \frac{1}{\min(K, R)} \sum_{j=1}^{K} (P_j \cdot V_j) \qquad R: total\ \#\ \text{of relevant data}$$

Example:

Ranked list of results

	D_{15}	D_8	D_{63}	D_{21}	...	...	D_s
Ground truth	*1*	*0*	*1*	*1*	*0*	*0*	*0*
Precision	*1/1*	*1/2*	*2/3*	*3/4*	*3/5*	*3/6*	*3/7*



Evaluation Metric: Average Precision

- Observations (AP)
  - AP depends on the rankings of relevant data and the size of the relevant data set. E.g., R=10

Case I: + + + + + + + + + + - - - - - - - - - - -

Case II: - + - + - + - + - + - + - + - + - + - +

Case III: - - - - - - - - - - - + + + + + + + + +

Affected by ranks

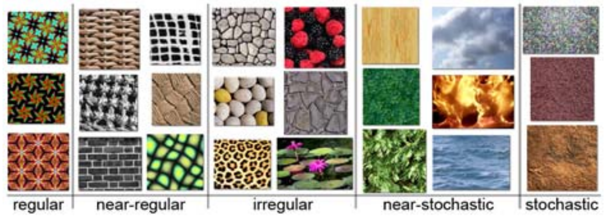
Case IV: + + + - - - - - - - - - - ... + + +

Case V: + + + - - - - - - - - - - ... + + + + + +

Same top ranks, different recall

Texture

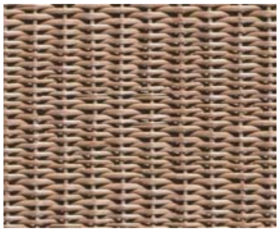
- What is texture?
  - "stylized subelements, repeated in meaningful ways"
  - May have quasi-stochastic macro structure (e.g. bricks), each with stochastic micro structure
- Why texture?
  - Application to satellite images, medical images
  - Useful for describing and reproducing contents of real world images, i.e., clouds, fabrics, surfaces, wood, stone
- Challenging issues
  - Rotation and scale variance (3D)
  - Segmentation/extraction of texture regions from images
  - Texture in noise

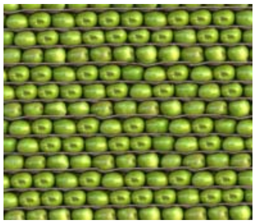


A wide range of filters for textures

Randen, T. and Husøy, J. H. 1999. Filtering for Texture Classification: A Comparative Study. *IEEE Trans. Pattern Anal. Mach. Intell.* 21, 4 (Apr. 1999), 291-310.

- Tamura Texture
- Zernike moments
- Steerable filters
- Ring/wedge filters
- Gabor filter banks
- wavelet transforms
- Quadrature mirror filters
- Discrete cosine transform
- Co-occurrence matrices



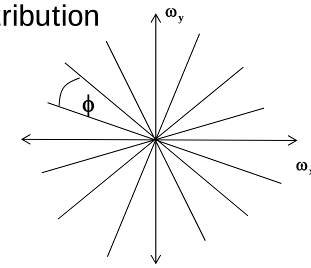
Filter approaches for texture description

- Fourier Transform *F(u,v)* Energy Distribution
  - Angular features (directionality)

$$V_{\theta_1\theta_2} = \iint |F(u,v)|^2 \, dudv$$

where ,

$$\theta_1 \le \tan^{-1}\left[\frac{v}{u}\right] \le \theta_2$$

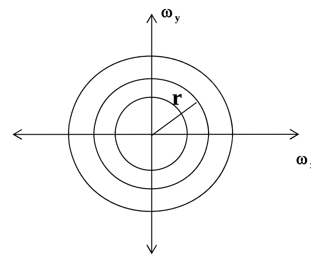


  - Radial features (coarseness)

$$V_{r_1 r_2} = \iint |F(u,v)|^2 \, dudv$$

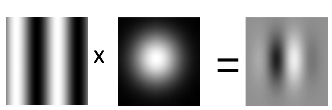
where ,

$$r_1 \le u^2 + v^2 < r_2$$

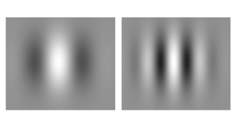


Gabor filters

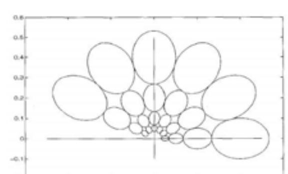
- Gaussian windowed Fourier Transform
  - Dyadic partitions of the spatio-frequency space
  - Basis filters are product/conv of Fourier basis images and Gaussians



Gabor filters



Different Frequency



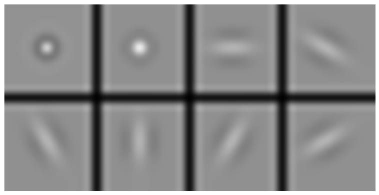
$$g(x,y) = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right)\exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi jWx\right]$$

$$G(u,v) = \exp\left\{-\frac{1}{2}\left[\frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}\right]\right\}$$

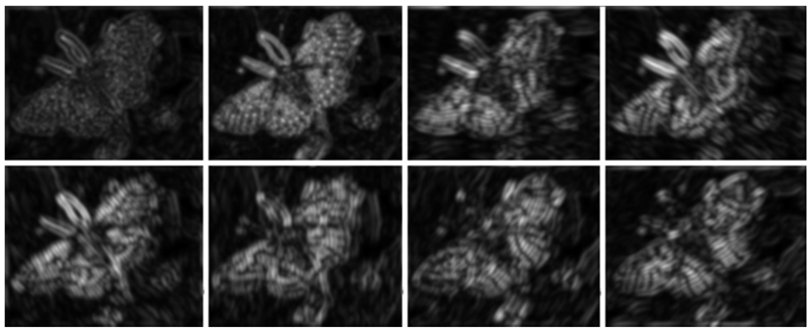
$$\sigma_u = 1/(2\pi\sigma_x); \; \sigma_v = 1/(2\pi\sigma_y)$$

Example: Filter Responses

Filter bank



Input image

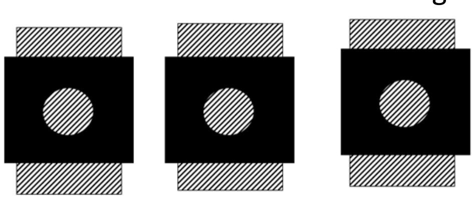


from Forsyth & Ponce

Local Features

- What are good local features?
  - Distinct, interesting content
  - Repeatable (invariant)
  - Precise locations – sensitive to position shift

- Aperture Problem – information within a small window often insufficient for determining true motion or matching

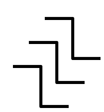


shift up   shift left

Images of Elizabeth Johnson

The barber pole illusion

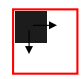
Multiple motion hypotheses exist



Corners help

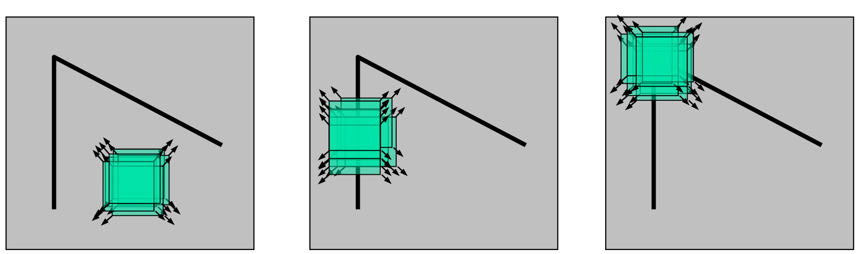
Corner Detection

- Types of local image windows
  - *Flat*: Little or no brightness change
  - *Edge*: Strong brightness change in single direction
  - *Flow*: Parallel stripes
  - *Corner/spot*: Strong brightness changes in orthogonal directions
- Basic idea
  - Find points where two edges meet
  - Look at the gradient behavior over a small window



(Slide of A. Efros)

Harris Detector: Basic Idea



"flat" region:
no change in all directions

"edge":
no change along the edge direction

"corner":
significant change in all directions

(Slide of A. Efros)

Harris Detector: Mathematics

Change of intensity for the shift [u,v]:

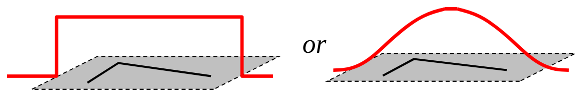
$$E(u,v) = \sum_{x,y} w(x,y)\left[I(x+u, y+v) - I(x,y)\right]^2$$

Window function

Shifted intensity

Intensity

Window function $w(x,y)$ =



Harris Corner Detector

- Taylor's expansion

$$E(u,v) = \sum_{x,y} w(x,y)[I(x+u, y+v) - I(x,y)]^2$$

$$= \sum_{x,y} w(x,y)[I_x u + I_y v + O(u^2, v^2)]^2$$

$$E(u,v) = Au^2 + 2Cuv + Bv^2$$

$$A = \sum_{x,y} w(x,y) I_x^2(x,y)$$

$$B = \sum_{x,y} w(x,y) I_y^2(x,y)$$

$$C = \sum_{x,y} w(x,y) I_x(x,y) I_y(x,y)$$

Harris Detector: Mathematics

Taylor's Expansion: For small shifts [u,v] we have a bilinear approximation:

$$E(u,v) \cong [u,v] \; M \begin{bmatrix} u \\ v \end{bmatrix}$$

where M is a 2×2 matrix computed from image derivatives:

$$M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

Harris Detector: Mathematics

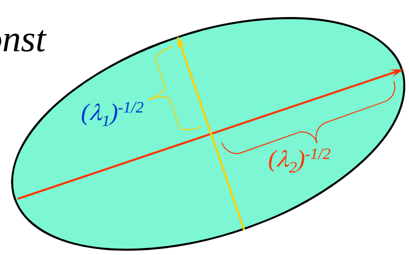
Intensity change in shifting window: eigenvalue analysis

$$E(u,v) \cong [u,v] \; M \begin{bmatrix} u \\ v \end{bmatrix}$$

$\lambda_1 > \lambda_2$ – eigenvalues of M

If we try every possible shift, the direction of fastest change is λ_1

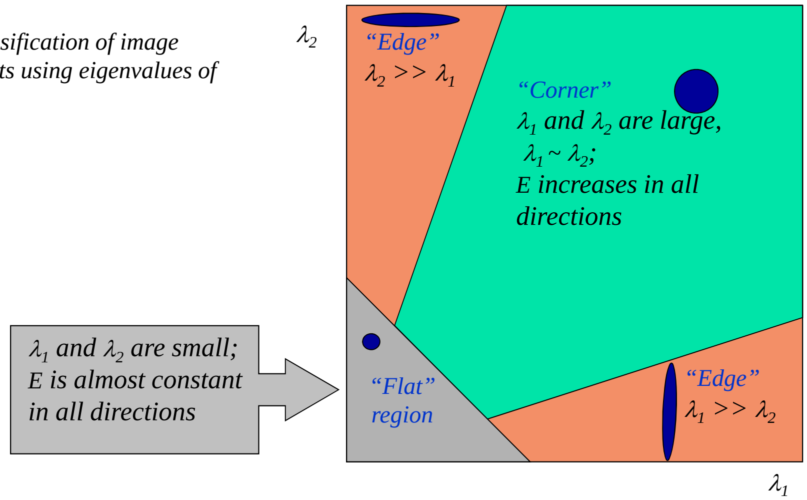
Ellipse E(u,v) = const



(Slide of K. Efros)

Harris Detector: Mathematics

Classification of image points using eigenvalues of M:



(Slide of A. Efros)

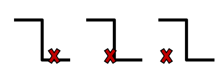
Harris Detector: Another Interpretation- Optical Flow

- Model image sequence as a spatiotemporal function I(x,y,t)

$$\frac{dI}{dt} = \frac{\partial I}{\partial x}\frac{dx}{dt} + \frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial t}$$

- Assume no change at displaced point, similar optical flow in local area

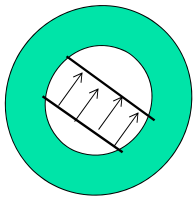
$$\frac{dI}{dt} = 0 \quad \Rightarrow \quad I_x u + I_y v + I_t = 0$$



where $(u,v) = (\frac{dx}{dt}, \frac{dy}{dt})$ Optical Flow

Harris Detector: Interpretation based on Tracking

- Lucas-Kanade Tracking: local consistency of optical flow



– (u,v) remain constant within a local patch Ω

$$\underset{(u,v)}{Min}\, E(u,v) = \underset{(u,v)}{Min} \sum_{x,y\in\Omega} (I_x u + I_y v + I_t)^2$$

$$\begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_x & \sum I_y I_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix}$$

M

- Optimal (u, v) satisfies Lucas-Kanade equation

(Slide of W. Freeman)

Conditions for solvability

- Optimal (u, v) satisfies Lucas-Kanade equation
- When is this Solvable?
  - M should be invertible
  - eigenvalues λ_1, λ_2 of M should not be too small
  - M should be well-conditioned
    - λ_1/λ_2 should not be too large (λ_1 = larger eigenvalue)

$$M = \begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix} = \sum \begin{bmatrix} I_x \\ I_y \end{bmatrix} [I_x \quad I_y]$$

(Slide of Khurram Hassan-Shafique)

Harris Detector: Mathematics

Measure of corner response:

$$R = \det M - k\left(\operatorname{trace} M\right)^2$$

$$\det M = \lambda_1 \lambda_2$$
$$\operatorname{trace} M = \lambda_1 + \lambda_2$$

Or

$$R = \frac{\det M}{\operatorname{Trace} M}$$

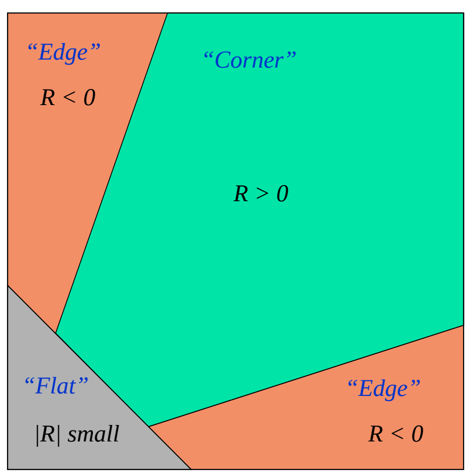
$$\det M = \lambda_1 \lambda_2$$
$$\operatorname{trace} M = \lambda_1 + \lambda_2$$

(k – empirical constant, k = 0.04-0.06)

Harris Detector: Mathematics

- *R depends only on eigenvalues of M*
- *R is large for a corner*
- *R is negative with large magnitude for an edge*
- *|R| is small for a flat region*

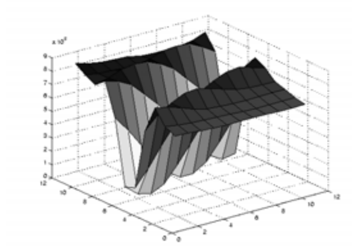
λ_2



λ_1

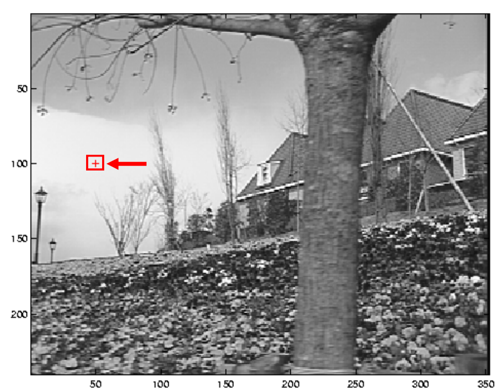
(Slide of K. Efros)

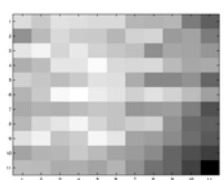
Edge

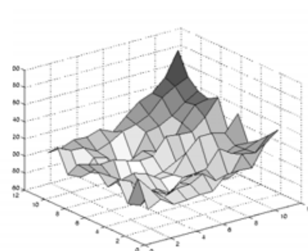


– *large gradients, all the same*
– *large λ_1, small λ_2*

Low texture region

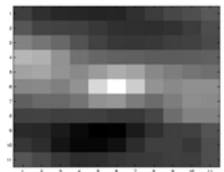


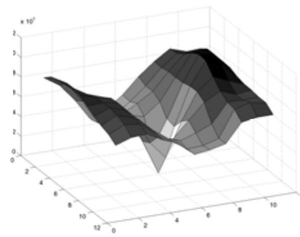




– *gradients have small magnitude*
– *small λ_1, small λ_2*

High textured region





– *gradients are different, large magnitudes*
– *large* λ_1*, large* λ_2

Harris Detector: Summary

- Average intensity change in direction $[u,v]$ can be expressed as a bilinear form:

$$E(u,v) \cong [u,v] \ \ M \begin{bmatrix} u \\ v \end{bmatrix}$$

- Describe a point in terms of eigenvalues of M: *measure of corner response*

$$M = \sum_{x,y} w(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

- A good (corner) point should have a *large intensity change* in *all directions*, i.e. R should be large positive

$$R = \lambda_1 \lambda_2 - k\left(\lambda_1 + \lambda_2\right)^2$$

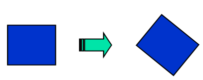
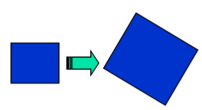
(k – empirical constant, k = 0.04-0.06)

(Slide of K. Efros)

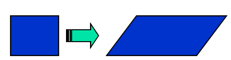
Harris Detector

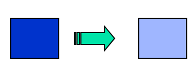
- The Algorithm:
  - Find points with large corner response function R (R > threshold)
  - Take the points of local maxima of R

Models of Image Change

- Geometry
  - Rotation 
  - Similarity (rotation + uniform scale) 
  - Affine (scale dependent on direction) valid for: orthographic camera, locally planar object

$$p' = \mathbf{H}p$$

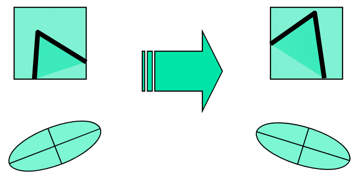


- Photometry
  - Affine intensity change ($I \rightarrow a\,I + b$) 

(Slide of C. Kambhamettu)

Harris Detector: Some Properties

- Rotation invariance



Ellipse rotates but its shape (i.e. eigenvalues) remains the same

Corner response R is invariant to image rotation

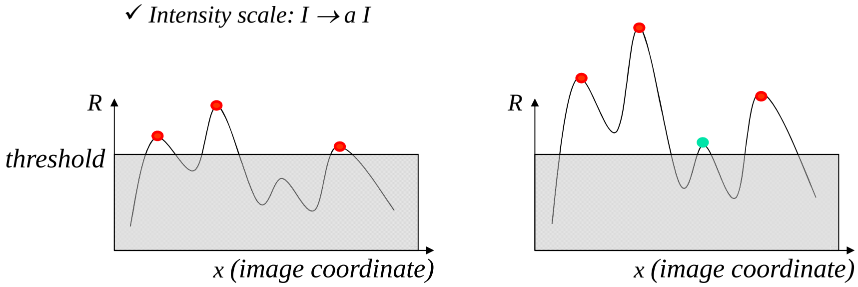
(Slide of C. Kambhamettu)

Harris Detector: Some Properties

- Partial invariance to *affine intensity* change

✓ *Only derivatives are used => invariance to intensity shift $I \rightarrow I + b$*

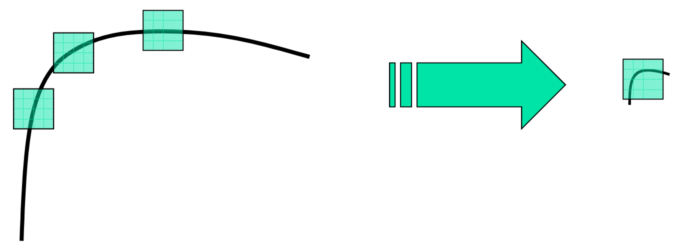
✓ *Intensity scale: $I \rightarrow a\, I$*



(Slide of C. Kambhamettu)

Harris Detector: Some Properties

- But: non-invariant to *image scale*!



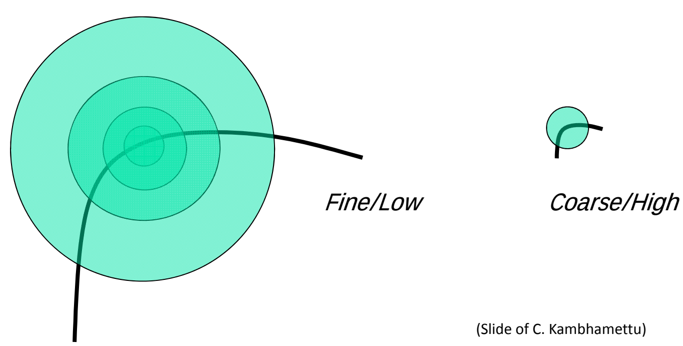
All points will be classified as edges

Corner !

(Slide of C. Kambhamettu)

Scale Invariant Detection

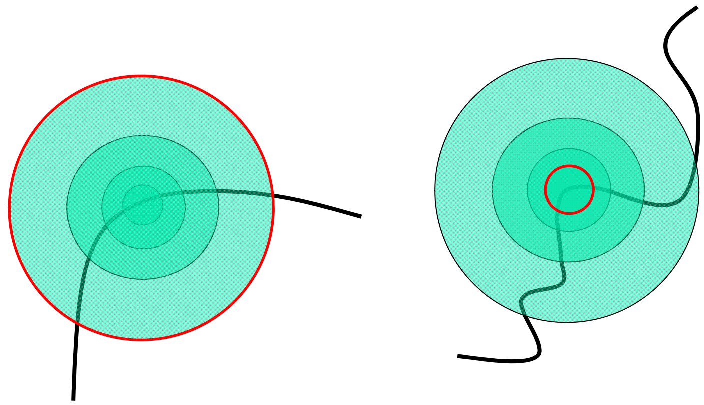
- Consider regions (e.g. circles) of different sizes around a point
- Regions of corresponding sizes (at different scales) will look the same in both images



(Slide of C. Kambhamettu)

Scale Invariant Detection

- The problem: how do we choose corresponding circles ***independently*** in each image?



(Slide of C. Kambhamettu)

Reading List

- Rui, Y., T.S. Huang, and S.-F. Chang, *Image retrieval: current techniques, promising directions and open issues. Journal of Visual Communication and Image Representation, 1999. 10(4): p. 39-62.*
- Smith, J.R. and S.-F. Chang. *VisualSEEk: a Fully Automated Content-Based Image Query System. in ACM International Conference on Multimedia. 1996. Boston, MA.*
- David G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, 60(2), 2004, pp91-110.
- Randen, T. and J. Husoy, *Filtering for texture classification: A comparative study. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2002. 21(4): p. 291-310.*
- Mikolajczyk, K. and C. Schmid, *A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005: p. 1615-1630.*
- Brown, M., R. Szeliski, and S. Winder. *Multi-image matching using multi-scale oriented patches. in IEEE CVPR. 2005.*

- *Interesting demo:*
  Georg Klein and David Murray, "Parallel Tracking and Mapping for Small AR Workspaces," ISMAR 2007.