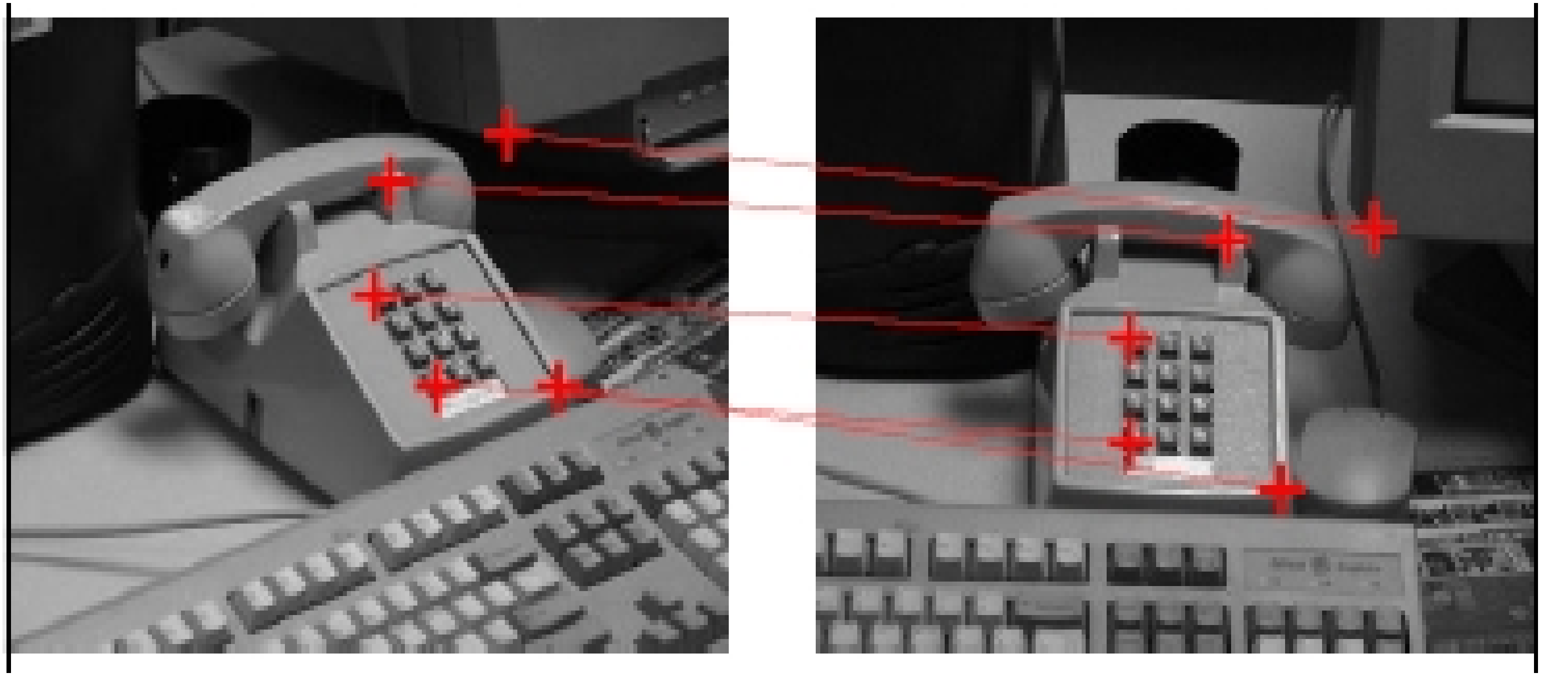
Interest Points

CS 554 – Computer Vision
Pinar Duygulu
Bilkent University

Image matching

- Image matching is a fundamental aspect of many problems in computer vision
  - Object or scene recognition
  - Solving for 3D structure from multiple images
  - Stereo correspondence
  - Motion tracking

Matching



First step toward 3-D reconstruction: find correspondences between feature points in two images of a scene
Object recognition: Find correspondences between feature points in training and test images

Adapted from Martial Hebert, CMU

Applications – Stereo correspondence

Applications – Image Retrieval

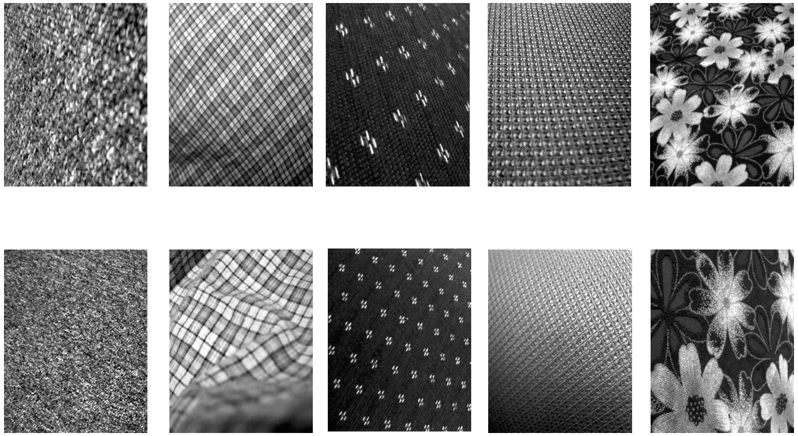
• • •

> 5000 images

change in viewing angle

Adapted from Cordelia Schmid and David Lowe, CVPR

Applications – Recognition

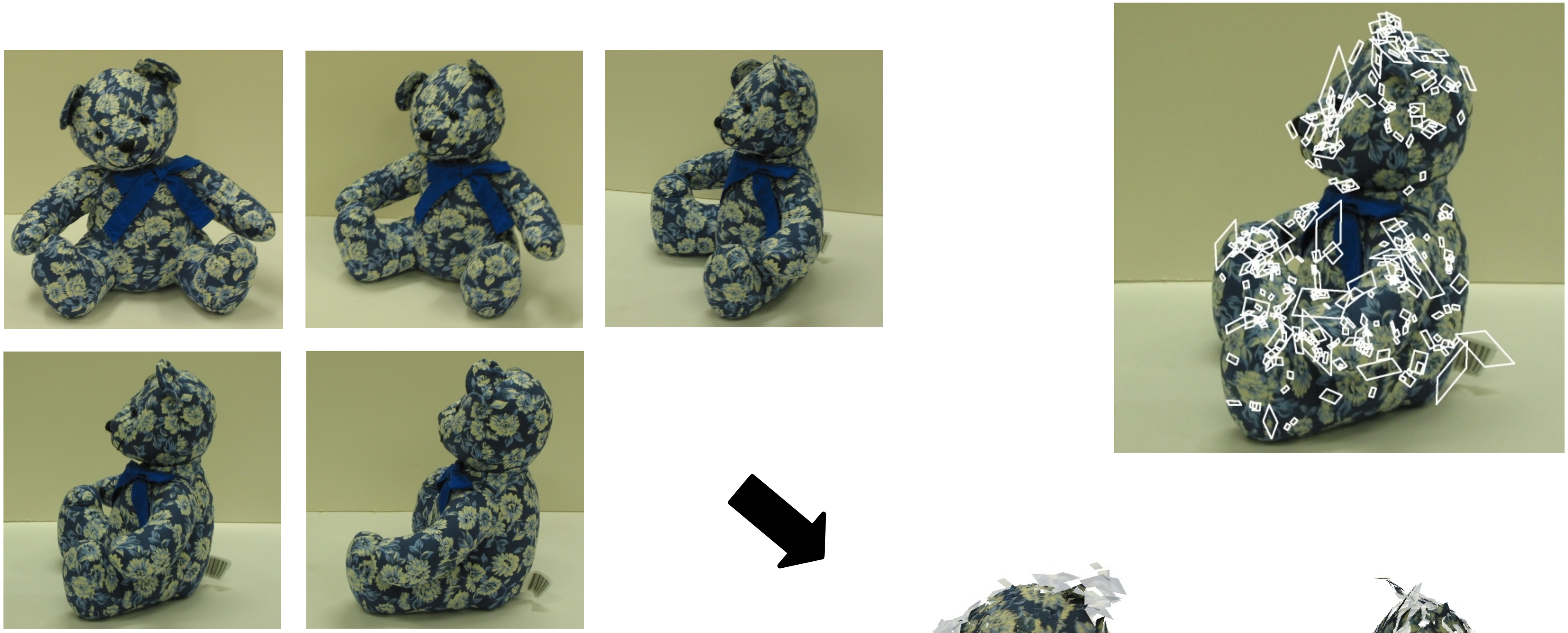


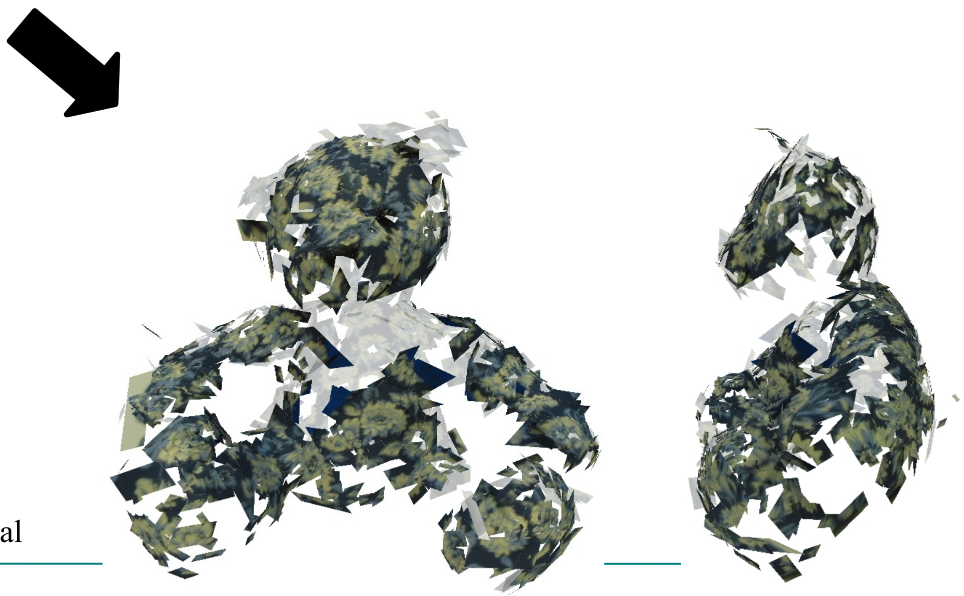
texture recognition

car detection

Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Applications – 3D Recognition





Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Matching

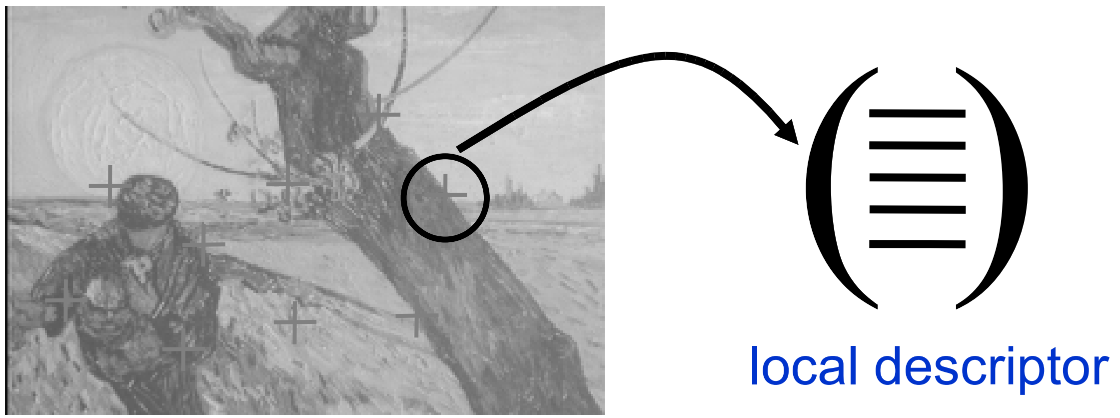
- Matching based on a form of continuum like texture, edge pixels or line segments
  - Not very discriminant
- Solution : matching with interest points & correlation
  - discrete, reliable and meaningful

Matching

- There are two important requirements for feature points to have a better correspondence for matching:
  - points corresponding to the same scene points should be extracted consistently over the different views
  - They should be invariant to image scaling, rotation and to change in illumination and 3D camera viewpoint
  - there should be enough information in the neighborhood of the points so that corresponding points can be automatically matched.

Interest Points

Local invariant photometric descriptors



Local : robust to occlusion/clutter + no segmentation

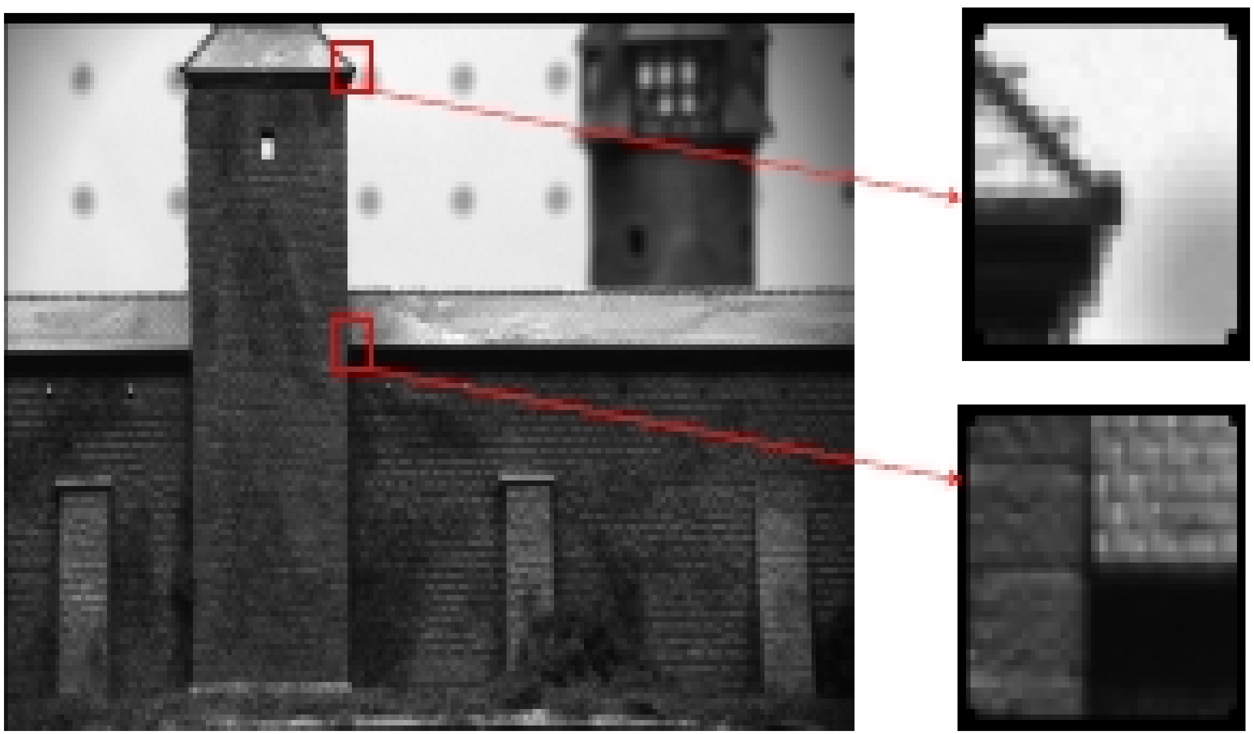
Photometric : distinctive

Invariant : to image transformations + illumination changes

Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Interest points

Intuitively junctions or contours
Generally more stable features over changes of view point
Intuitively large variations in the neighborhood of the point in all directions



Adapted from Martial Hebert, CMU

Edges vs. Corners

At sharp corners partial derivative estimates are poor, because their support will cross the corner

At the corners gradient swings sharply

The statistics of the gradient in an image neighborhood yields quite useful description of the image neighborhood:

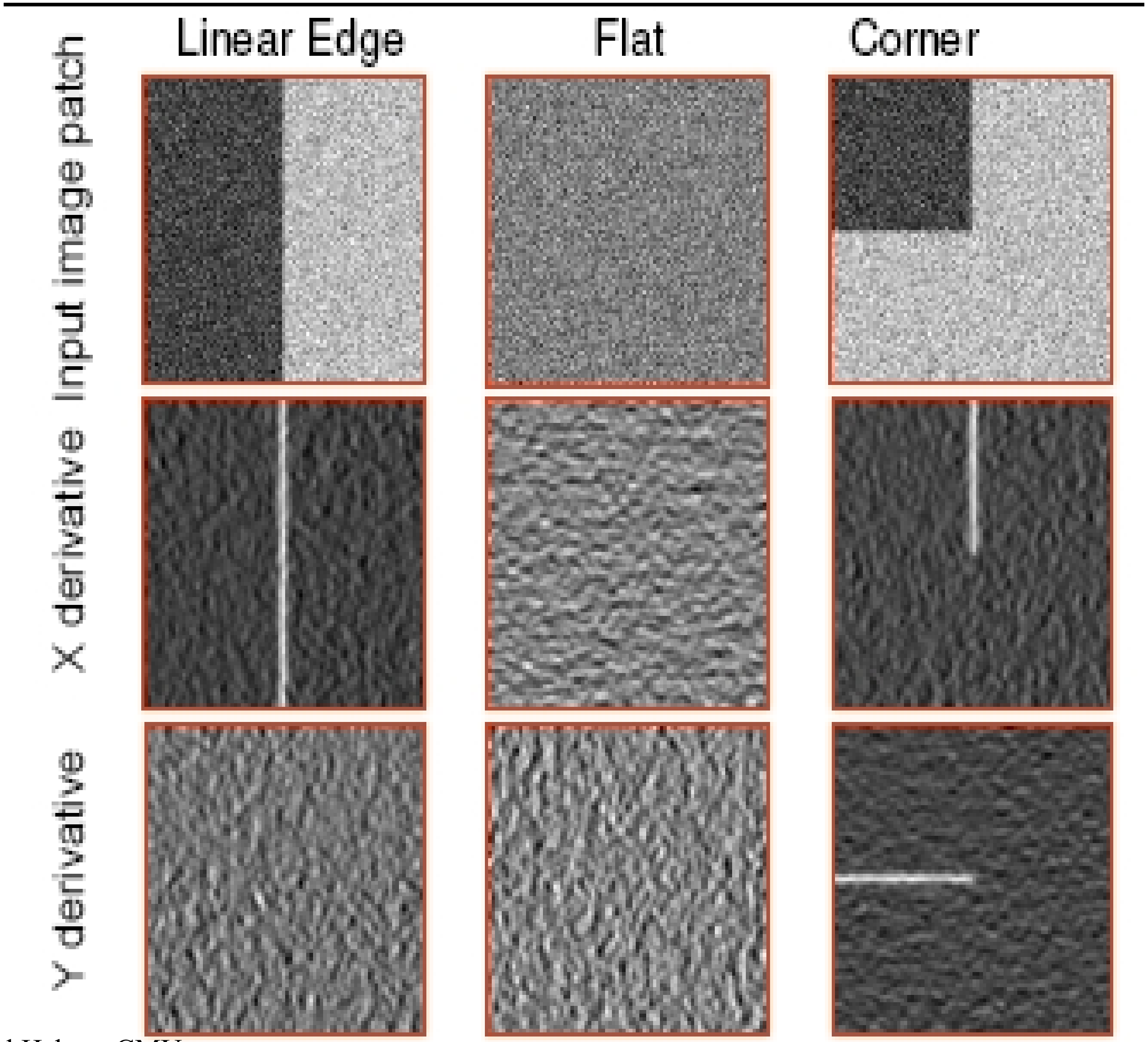
Constant windows

Edge windows

Flow windows : several parallel stripes

2D windows : spots or corners

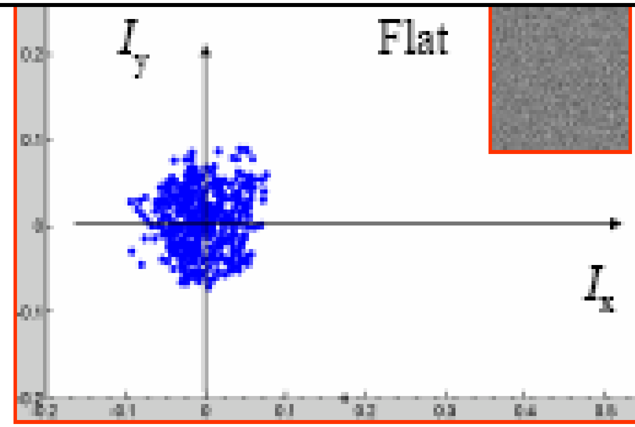
Edges vs. Corners

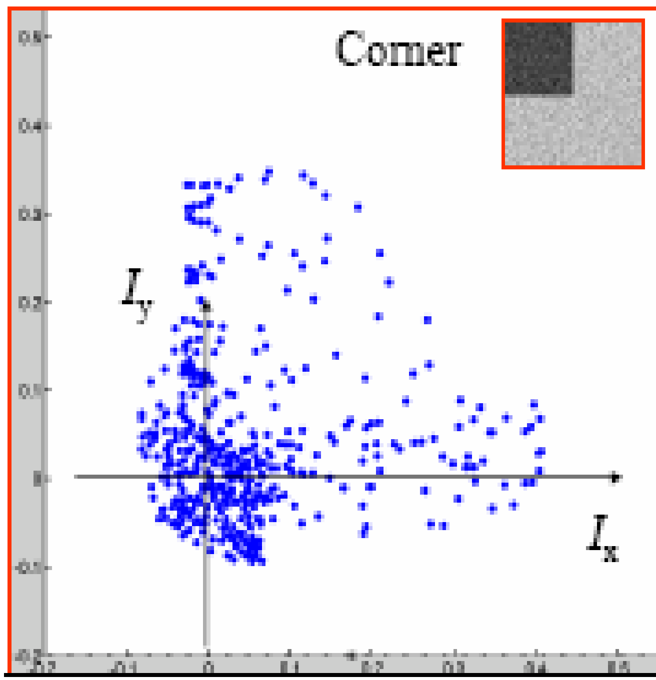


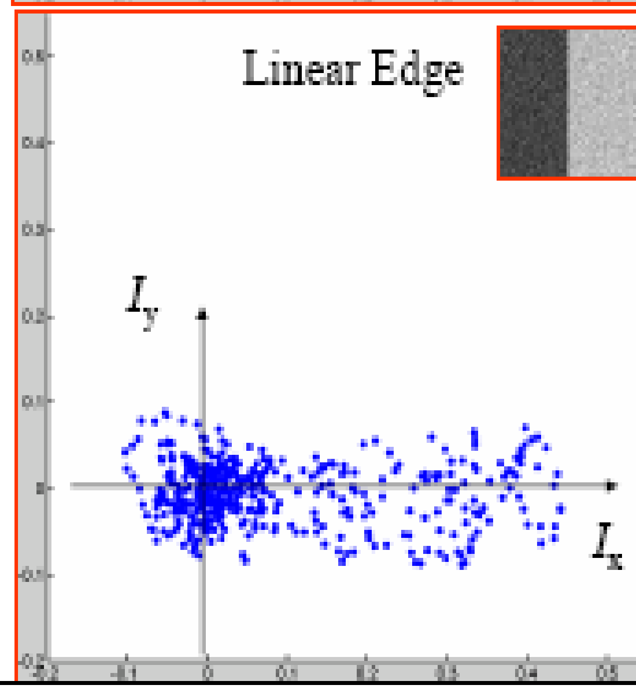
Adapted from Martial Hebert, CMU

Edges vs. Corners

The distribution of the x and y derivatives is very different for all three types of patches

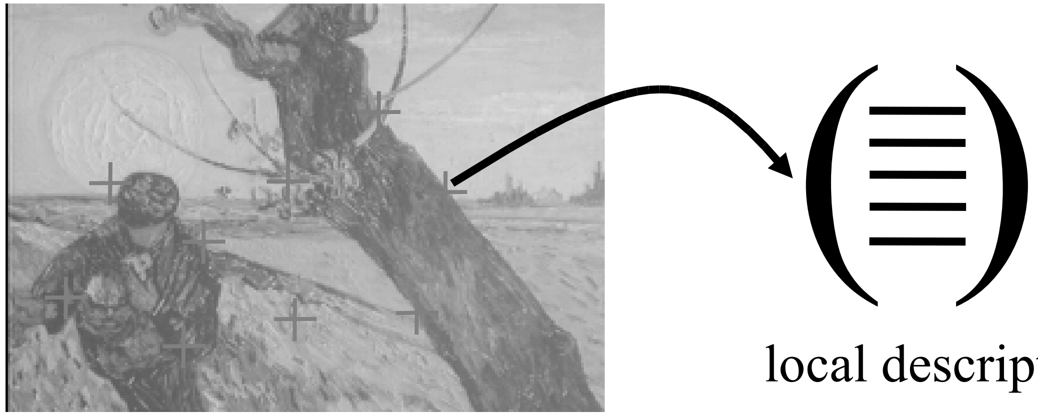






Adapted from Martial Hebert, CMU

Overview of the approach



1) Extraction of interest points (characteristic locations)

2) Computation of local descriptors

3) Determining correspondences

4) Selection of similar images

Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Moravec's Corner Detector

Shift a local window over the image to determine the average intensity changes

-If the windowed image patch is flat (i.e. approximately constant in intensity) then all shifts will result in only a small change

-If the window straddles an edge then a shift along the edge will result in a small change, but a shift perpendicular to the edge will result in a large change

-If the windowed patch is a corner or isolated point then all shifts will result in a large change. A corner can thus be detected by finding when the minimum change produced by any of the shifts is large

Moravec's Corner Detector

$$E(x,y) = \sum W(u,v)\, |I(x+u,y+v) - I(u,v)|^2$$

w: window

I: image intensities

E: the change produced by a shift (x,y)

Moravec's corner detector : look for local maxima

Moravec's Corner Detector

Problems:

- The response is anisotropic because only a discrete set of shifts at every 45 degrees is considered

-The response is noisy because the window is binary and rectangular

-The operator responds to readily to edges because only the minimum of E is taken into account

Harris & Stephens, 1988

Based on the idea of auto-correlation

Important difference in all directions => interest point

Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Harris & Stephens, 1988

Rewrite E for small shifts as

$$E(x,y) = Ax^2 + 2Cxy + By^2$$

$$A = X^2 * w$$
$$B = Y^2 * w$$
$$C = (XY) * w$$

$$X = I * (-1, 0, 1) = \partial I/\partial x$$
$$Y = I * (-1, 0, 1)_T = \partial I/\partial y$$

Harris & Stephens, 1988

$$E(x,y) = (x,y)M(x,y)^T$$

$$M = \begin{bmatrix} A & C \\ C & B \end{bmatrix}$$

E is related to local auto correlation function, with M describing its shape at the origin

Harris & Stephens, 1988

$$M = \begin{pmatrix} \left(\frac{\partial I}{\partial x}\right)^2 & \left(\frac{\partial I}{\partial x}\right)\left(\frac{\partial I}{\partial y}\right) \\ \left(\frac{\partial I}{\partial x}\right)\left(\frac{\partial I}{\partial y}\right) & \left(\frac{\partial I}{\partial y}\right)^2 \end{pmatrix}$$

Let λ1 and λ2 be the eigenvalues of M

$$\mathrm{M} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

λ1 and λ2 will be proportional to the principal curvatures of the local auto-correlation function, and form a rotationally invariant description of M

Harris Corner Detector

If both curvatures are small, so that auto-correlation function is flat, then the windowed image region is of approximately constant intensity --> arbitrary shifts of the image patch cause little change in E

If one curvature is high and the other low, so that the auto-correlation function is ridge shaped, then only shifts along the ridge (along the edge) cause little changes in E --> this indicates an edge

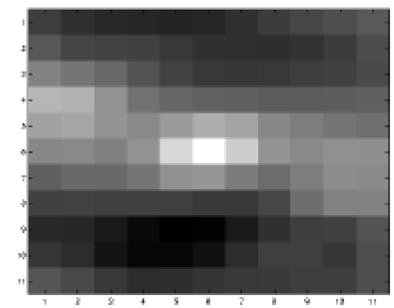
If both curvatures are high, so that the local auto-correlation function is sharply peaked then shifts in any direction will increase E --> this indicates a corner

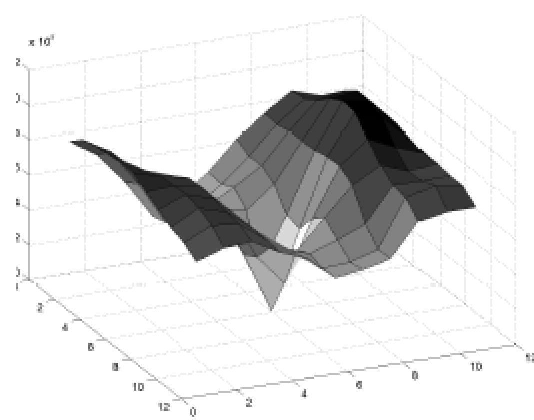
Harris Corner Detector

Three cases may occur:

- In a constant window both eigenvalues are small
- In an edge window there is one large eigenvalue
- In a 2D window both eigenwindows are large

Harris Corner Detector

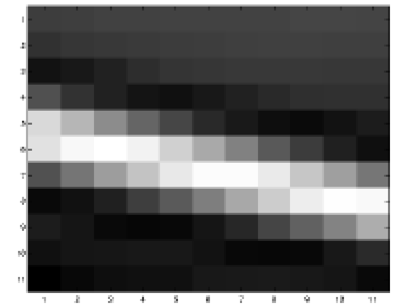


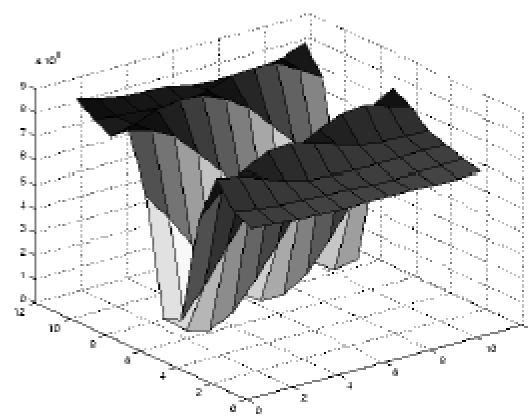


λ_1 and λ_2 are large

Adapted from Trevor Darrell, MIT

Harris Corner Detector

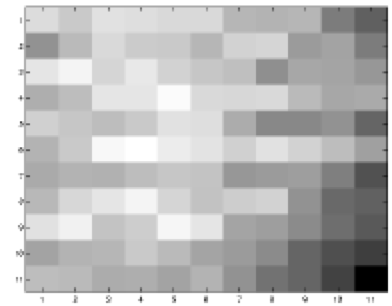


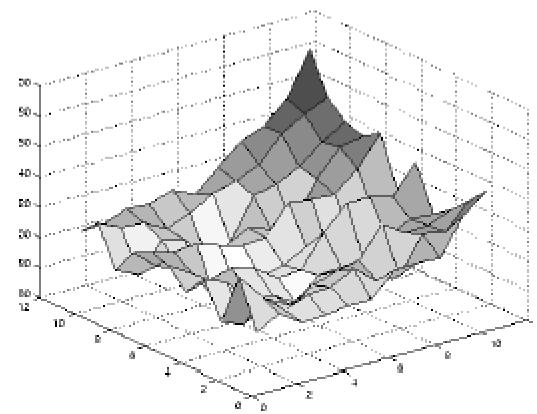


large λ_1, small λ_2

Adapted from Trevor Darrell, MIT

Harris Corner Detector





small λ_1, small λ_2 31

Adapted from Trevor Darrell, MIT

Harris detector

To measure the corner quality:

If at a certain point the two eigenvalues of the matrix M are large, then a small motion in any direction will cause an important change of grey level. This indicates that the point is a corner.

The corner response function (R) is given by:

$$R = \det(M) - k(\mathrm{trace}(M))^2$$

$$\mathrm{trace}(M) = \lambda 1 + \lambda 2$$

$$\mathrm{Det}(M) = \lambda 1 \,.\, \lambda 2$$

R is positive for corners
negative in edge regions
small in flat regions

Interest points

1. Compute x and y derivatives of image

$$I_x = G_\sigma^x * I \quad I_y = G_\sigma^y * I$$

2. Compute products of derivatives at every pixel

$$I_{x2} = I_x . I_x \quad I_{y2} = I_y . I_y \quad I_{xy} = I_x . I_y$$

3. Compute the sums of the products of derivatives at each pixel

$$S_{x2} = G_{\sigma'} * I_{x2} \quad S_{y2} = G_{\sigma'} * I_{y2} \quad S_{xy} = G_{\sigma'} * I_{xy}$$

4. Define at each pixel (x, y) the matrix

$$H(x, y) = \begin{bmatrix} S_{x2}(x, y) & S_{xy}(x, y) \\ S_{xy}(x, y) & S_{y2}(x, y) \end{bmatrix}$$

5. Compute the response of the detector at each pixel

$$R = Det(H) - k(Trace(H))^2$$

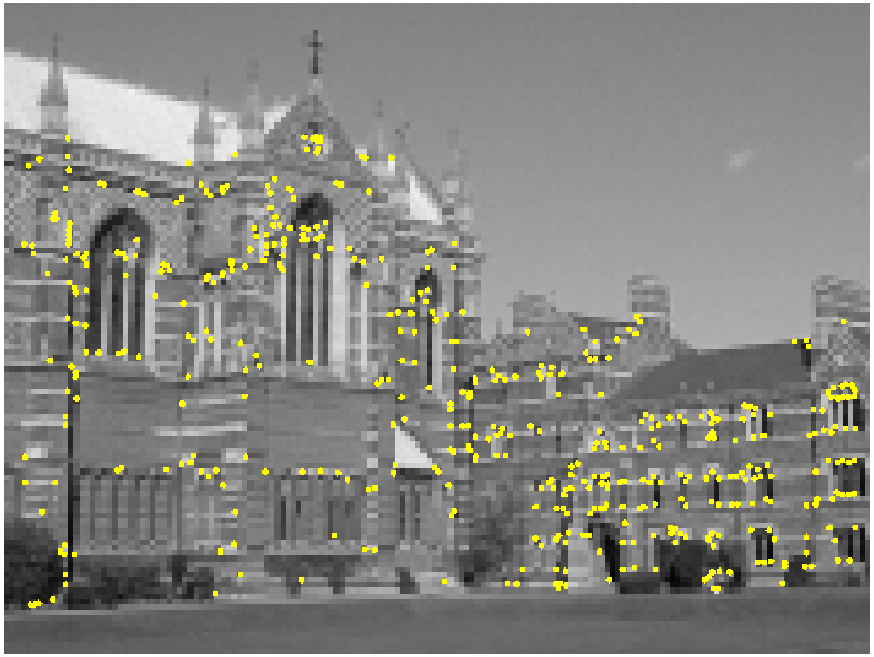
Adapted from Martial Hebert, (

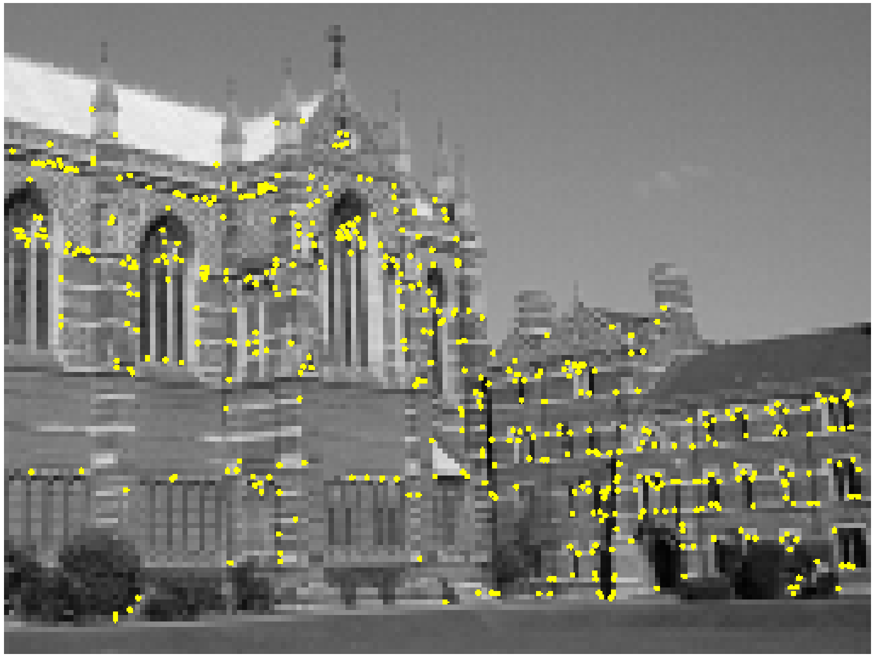
Harris detector

In practice often far too much corners are extracted.

- first restrict the numbers of corners before trying to match them.
  - One possibility consists of only selecting the corners with a value above a certain threshold.
  - This threshold can be tuned to yield the desired number of features.
  - Since for some scenes most of the strongest corners are located in the same area, it can be interesting to refine this scheme further to ensure that in every part of the image a sufficient number of corners are found.

Harris Detector



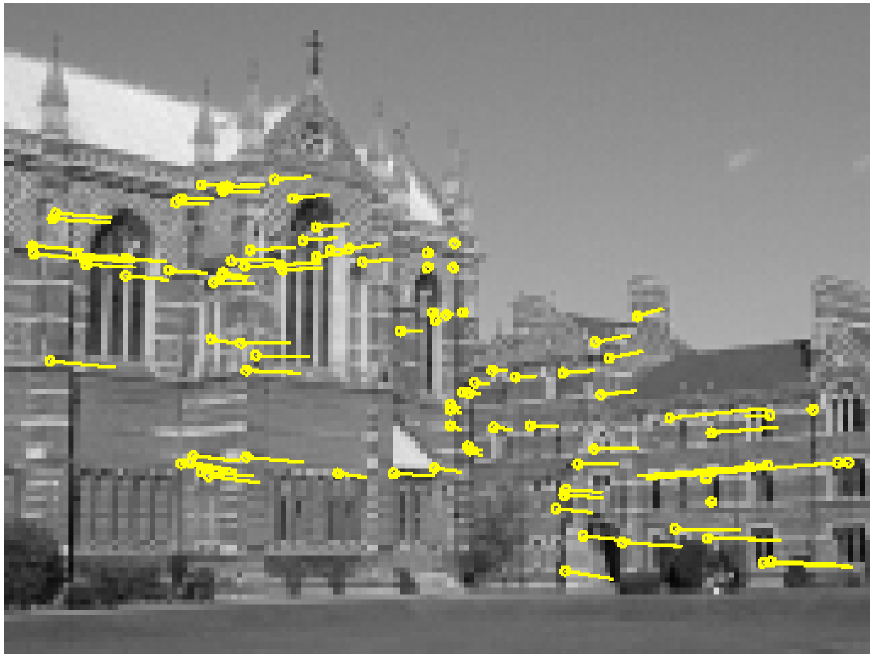


Interest points extracted with Harris (~ 500 points)

Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Harris Detector

Robust estimation of the fundamental matrix



99 inliers

89 outliers

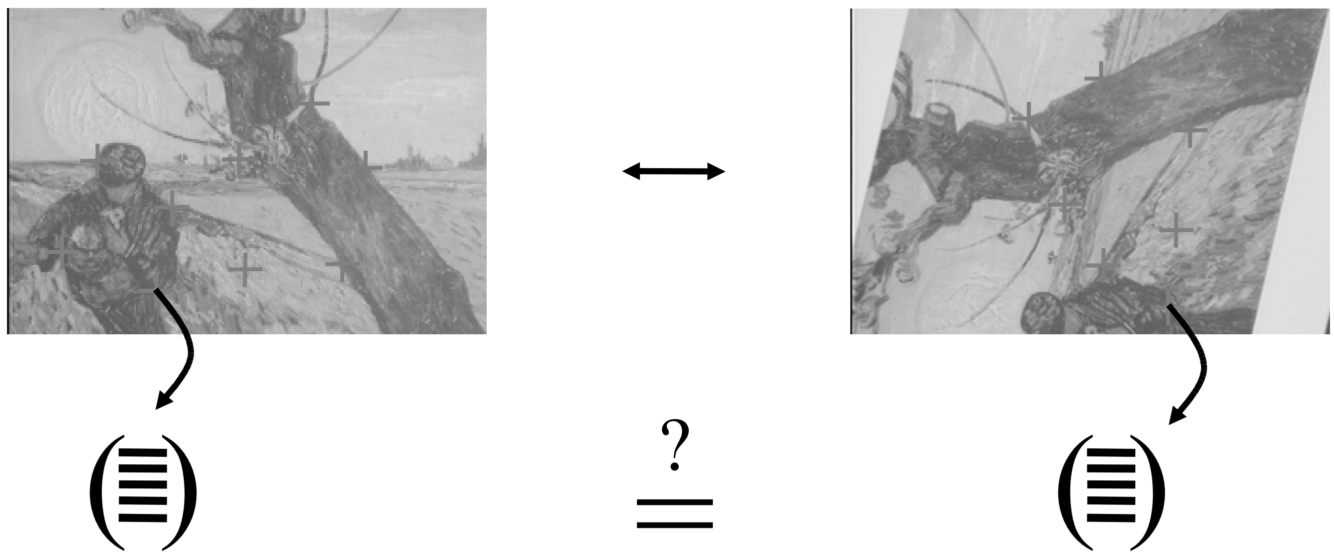
Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Approach for Matching and Recognition

- Detection of interest points/regions
  - Harris detector
- Computation of descriptors for each point
- Similarity of descriptors
  - correlation, Mahalanobis distance, Euclidean distance
- Semi-local constraints
- Global verification

Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Determining Correspondences



Vector comparison using the Mahalanobis distance

$$dist_M(\mathbf{p},\mathbf{q}) = \sqrt{(\mathbf{p}-\mathbf{q})^T \Lambda^{-1} (\mathbf{p}-\mathbf{q})}$$

Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

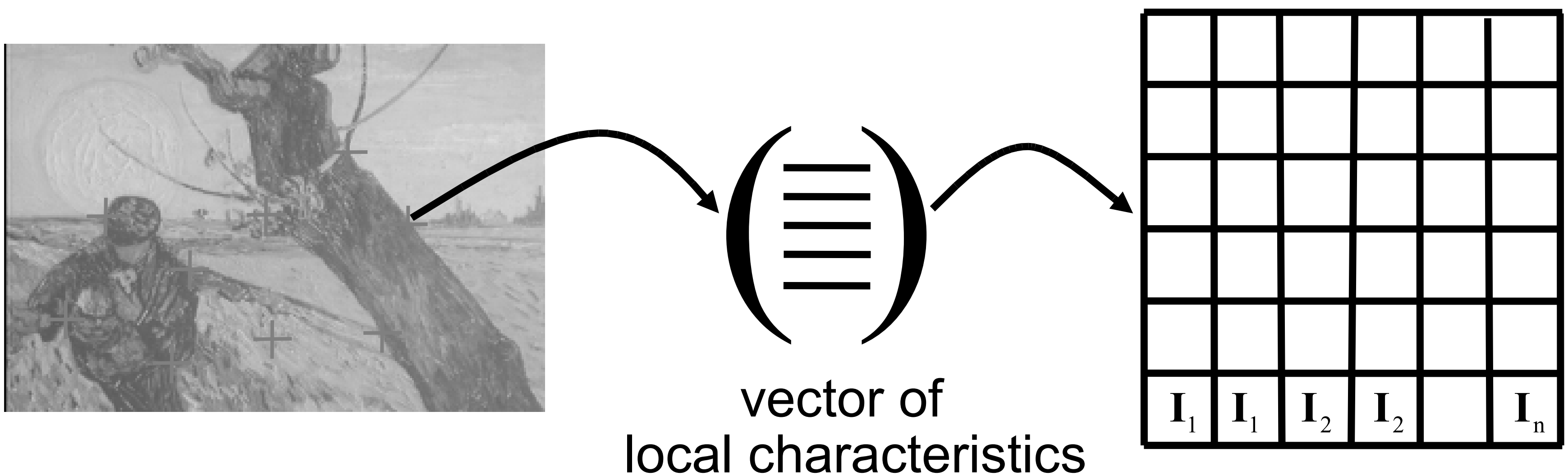
Selection of Similar Images

- In a large database
  - voting algorithm
  - additional constraints

- Rapid access with an indexing mechanism

Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Voting Algorithm



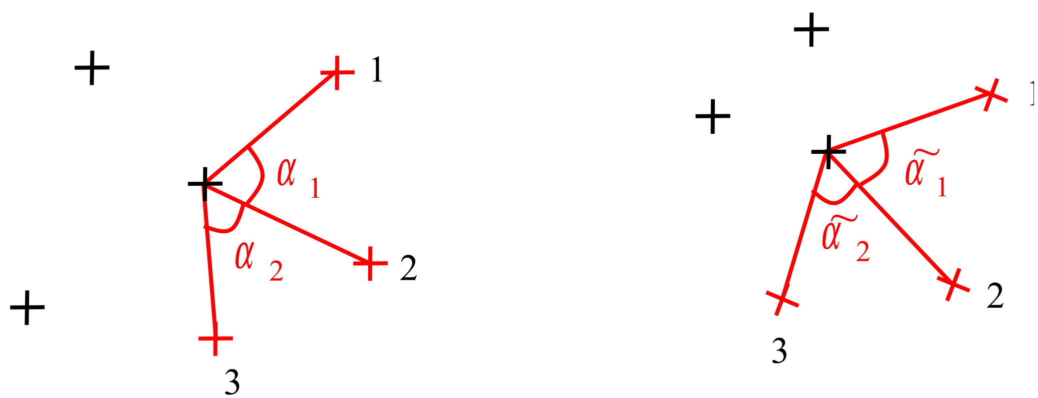
Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Voting Algorithm

- Compute a set of invariant features V around each interest point for each image in the database
- For a query image compute the same model
- Compare the vectors for each of the interest points in the query image with all the models in the database
- If distance is below some threshold then give a vote to the corresponding model

Additional Constraints

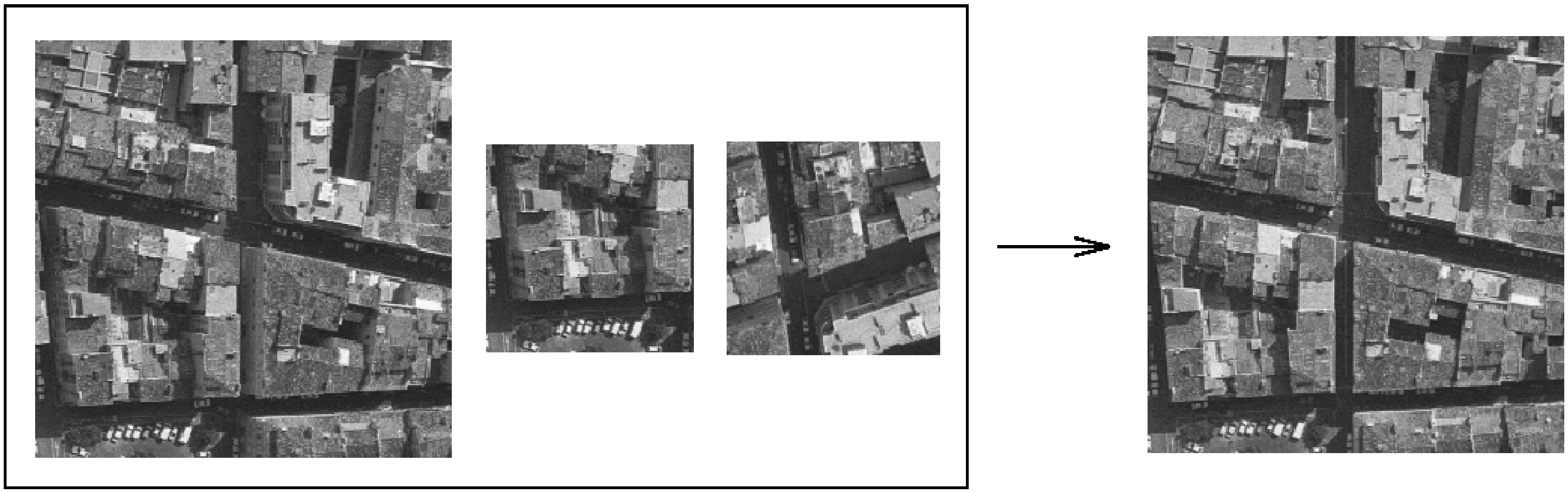
- Semi-local constraints
  - neighboring points should match
  - angles, length ratios should be similar



- Global constraints
- robust estimation of the image transformation (homogaphy, epipolar geometry)

Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Results

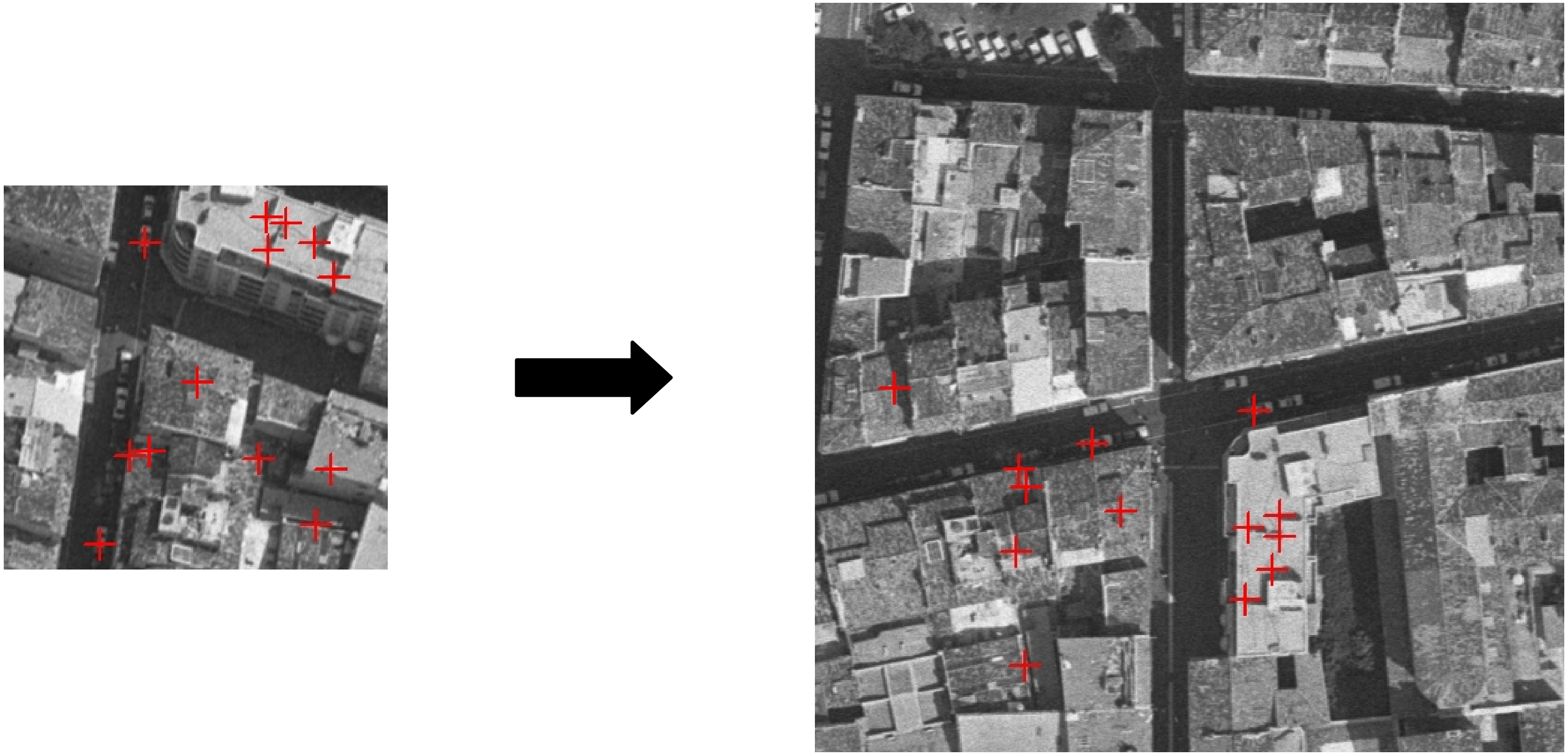


database with ~1000 images

The image on the right is correctly retrieved using any of the images on the left

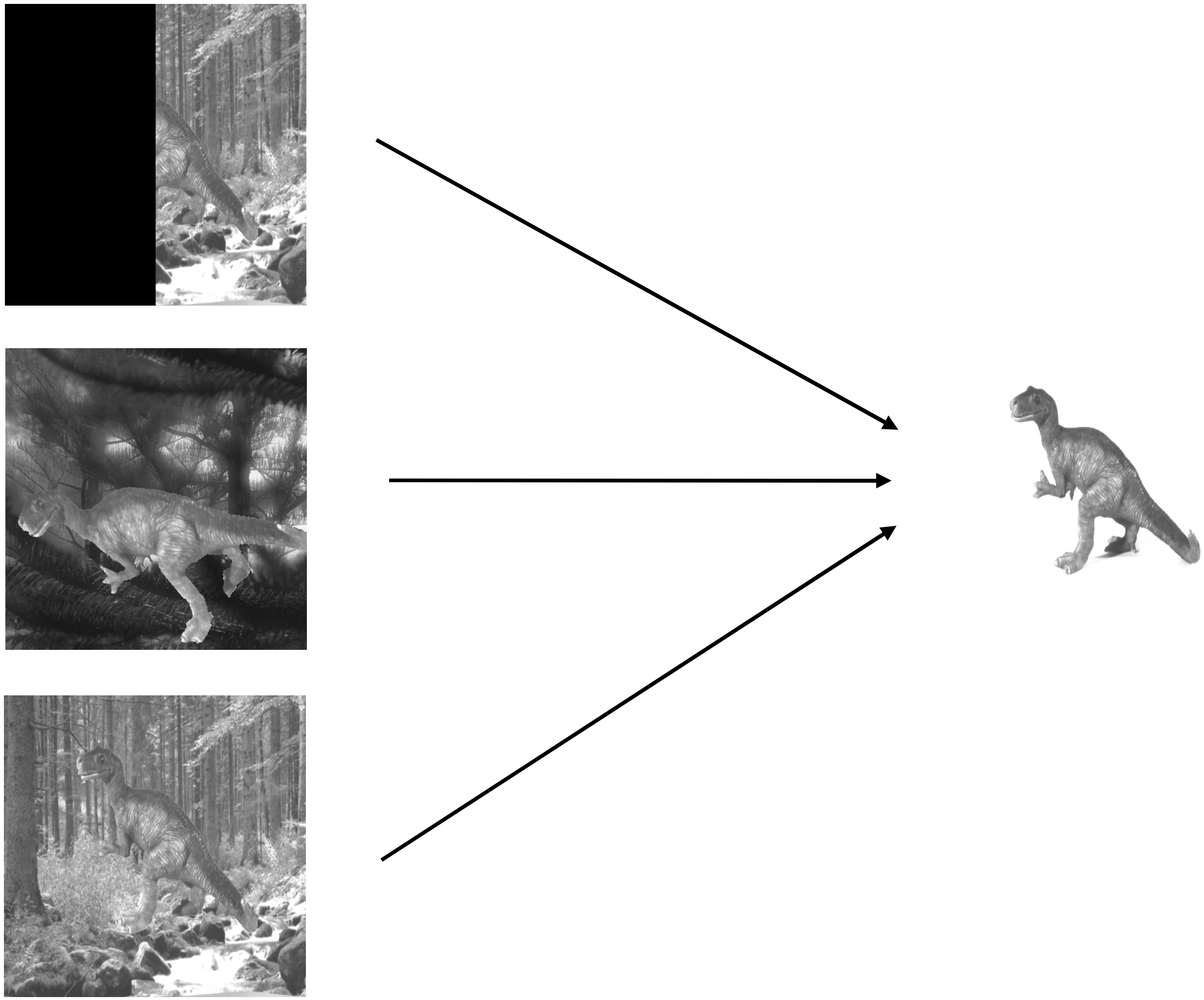
Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Results



Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Results



Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

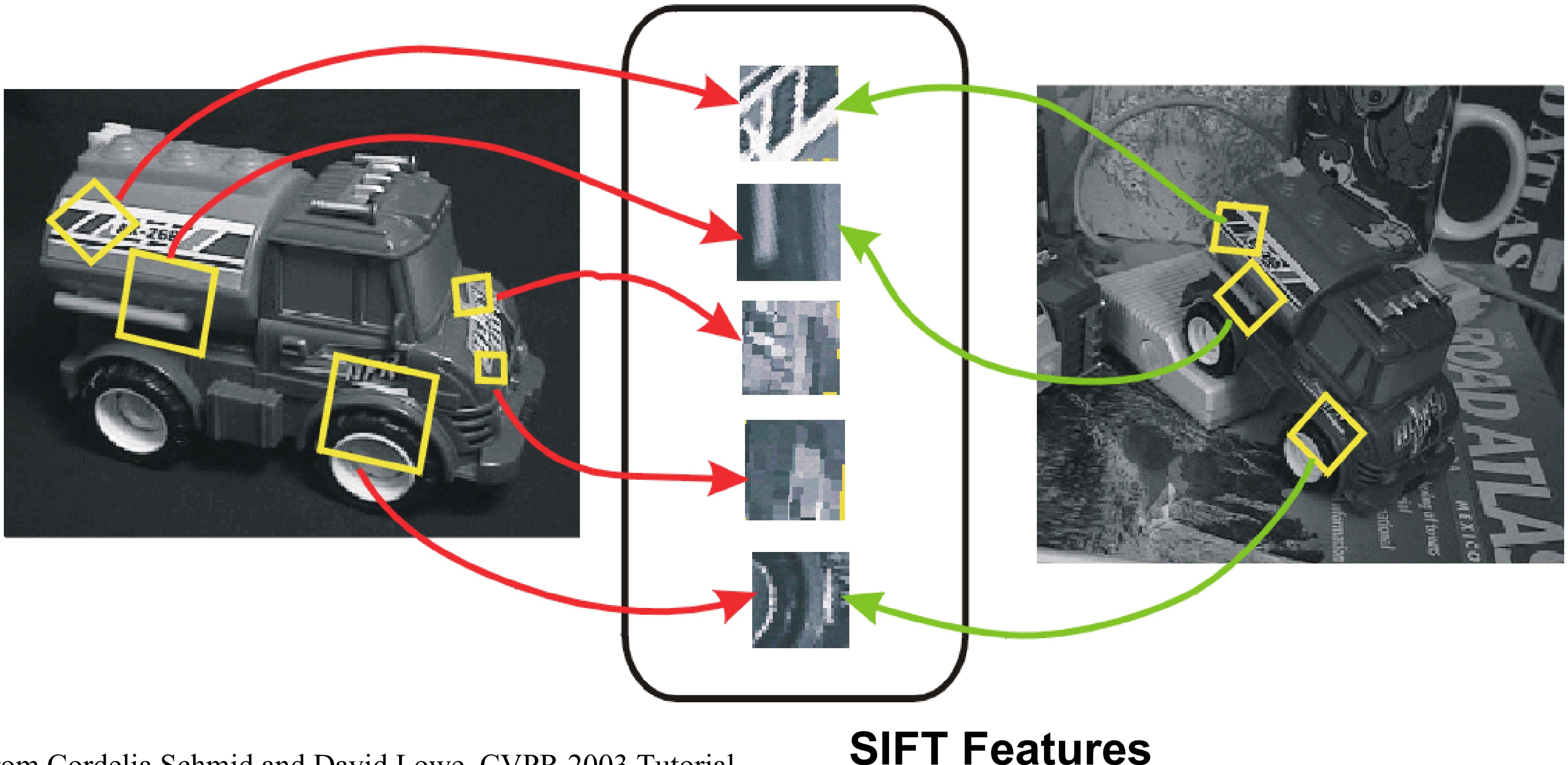
Approach for Matching and Recognition

- Detection of interest points/regions
  - Harris detector ***(extension to scale and affine invariance)***
- Computation of descriptors for each point
  - greyvalue patch, diff. invariants, steerable filter, ***SIFT descriptor***
- Similarity of descriptors
- Semi-local constraints
- Global verification

Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

SIFT (Scale Invariant Feature Transform) –Lowe'04

- Image content is transformed into local feature coordinates that are invariant to translation, rotation, scale, and other imaging parameters



Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Advantages of Invariant Local Features

- **Locality:** features are local, so robust to occlusion and clutter (no prior segmentation)
- **Distinctiveness:** individual features can be matched to a large database of objects
- **Quantity:** many features can be generated for even small objects
- **Efficiency:** close to real-time performance
- **Extensibility:** can easily be extended to wide range of differing feature types, with each adding robustness

Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Approach

- **Scale space extrema detection** : search over all scales and image locations. Efficient implementation by using Difference of Gaussians to identify potential interest points that are invariant to scale and orientation
- **Keypoint localization:** At each candidate location, a detailed model is fit to determine location and scale. Keypoints are selected based on measure of stability
- **Orientation assignment:** One or more orientations are assigned to each keypoint location based on local image gradient directions. All future operations are performed on image data that has been transformed relative to the assigned orientation, scale and location for each feature – invariance
- **Keypoint descriptor :** The local image gradient are measured at the selected scale in the region around each point

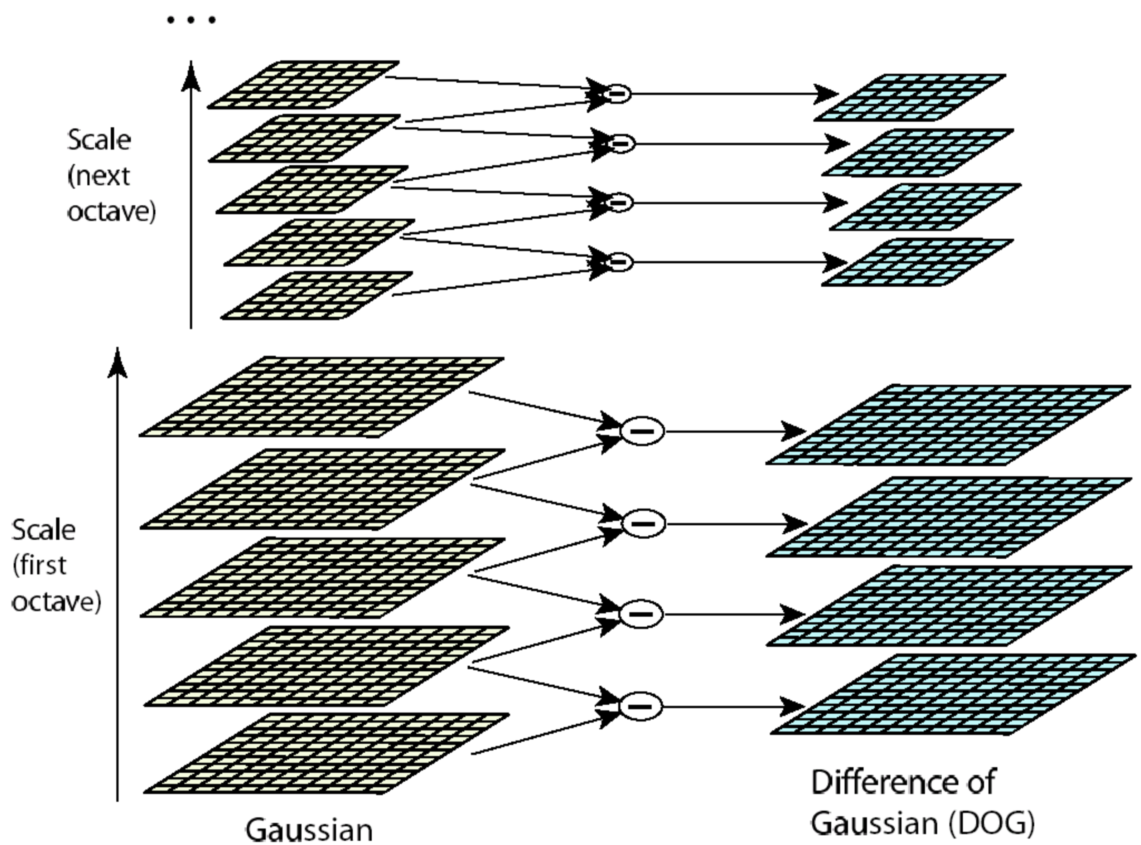
Scale Invariance

Requires a method to repeatably select points in location and scale:

- The only reasonable scale-space kernel is a Gaussian (Koenderink, 1984; Lindeberg, 1994)
- An efficient choice is to detect peaks in the difference of Gaussian pyramid (Burt & Adelson, 1983; Crowley & Parker, 1984 – but examining more scales)
- Difference-of-Gaussian with constant ratio of scales is a close approximation to Lindeberg's scale-normalized Laplacian (can be shown from the heat diffusion equation)
- The extremum of scale-normalized Laplacian of Gaussian produces the most stable image feature compared to Hessian or Harris corner detector (Mikolajczyk. 2002)

Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Scale space processed one octave at a time

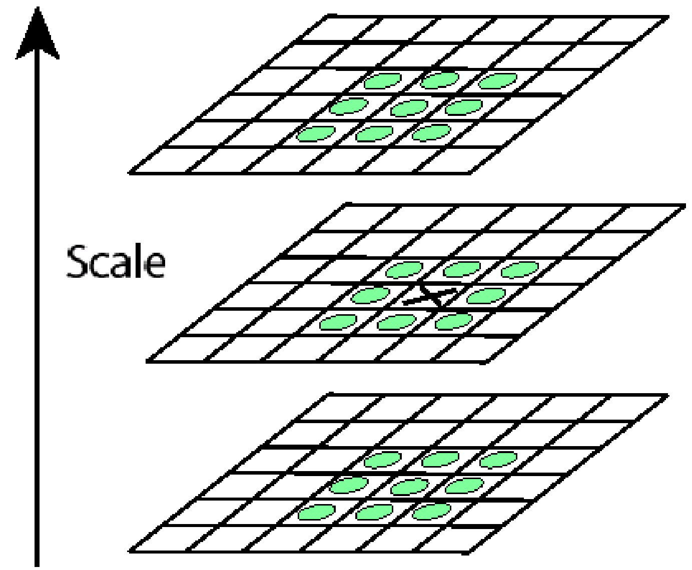


For each octave of scale space, the initial image is repeatedly convolved with Gaussian to produce the set of scale space images (Left). Adjacent Gaussian images are subtracted to produce difference of Gaussian images (Right) After each octave Gaussian image is downsampled by a factor of 2

Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Key point localization

- Detect maxima and minima of difference-of-Gaussian in scale space



Compare each sample point with its eight neighbors in the current image and nine neighbors in the scale above and below
Select only if it is greater or smaller than all the others

Problem : how to determine the frequency of sampling
Consider a white circle on a black background. There is only single scale space maximum where the circular positive central region of DOG function matches the size and location of the circle
There is a trade of between the sampling frequency and rate of detection

Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Key point localization

- To reject the points that have low contrast (sensitive to noise) or poorly localized on the edges

- Fit a quadratic to surrounding values for sub-pixel and sub-scale interpolation (Brown & Lowe, 2002)
- Taylor expansion around point:

$$D(\mathbf{x}) = D + \frac{\partial D}{\partial \mathbf{x}}^T \mathbf{x} + \frac{1}{2}\mathbf{x}^\mathbf{T} \frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x}$$

- The location of extremum is found by

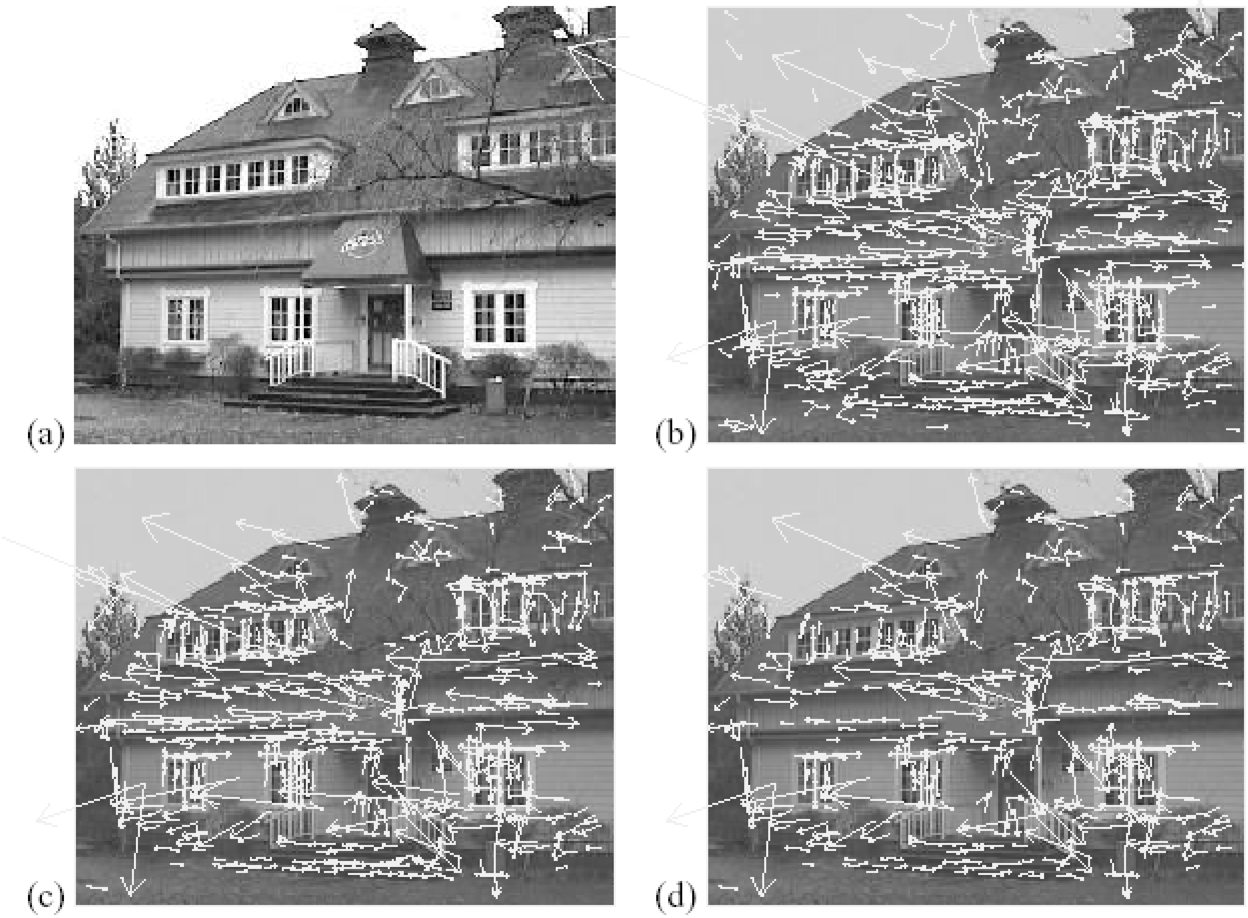
$$\hat{\mathbf{x}} = -\frac{\partial^2 D}{\partial \mathbf{x}^2}^{-1} \frac{\partial D}{\partial \mathbf{x}}$$

The function value at extremum is useful to rejecting unstable extrema with low contrast

Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Example of Keypoint detection

Threshold on value at DOG peak and on ratio of principle curvatures (Harris approach)

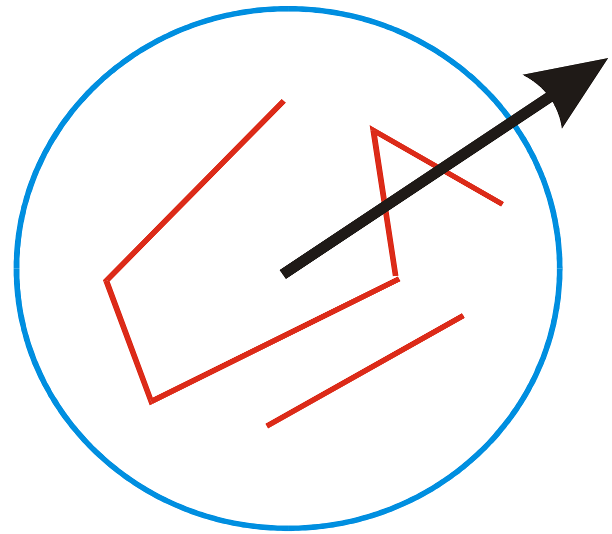


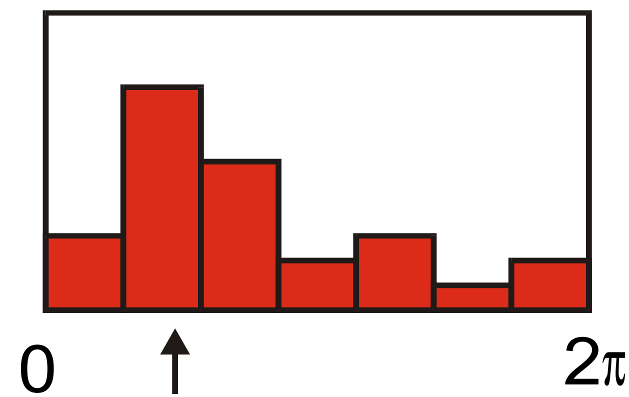
(a) 233x189 image
(b) 832 DOG extrema
(c) 729 left after peak value threshold
(d) 536 left after testing ratio of principle curvatures

Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Select canonical orientation

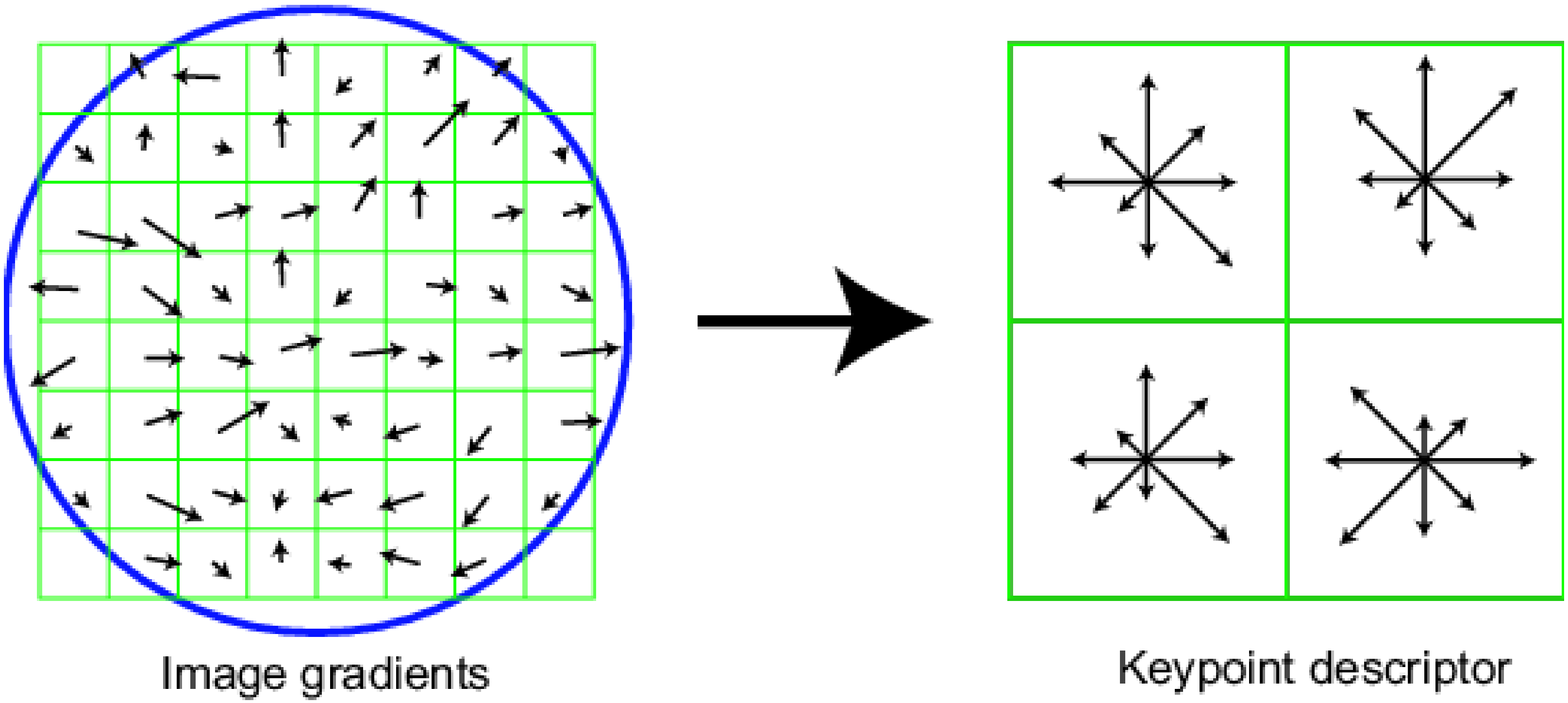
- Create histogram of local gradient directions computed at selected scale
- Assign canonical orientation at peak of smoothed histogram
- Each key specifies stable 2D coordinates (x, y, scale, orientation)





Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

SIFT vector formation



Compute the gradient magnitude and orientation at each sample point in a region around the keypoint
Weight by a Gaussian window. Accumulate into orientation histogram

Adapted from Cordelia Schmid and David Lowe, CVPR 2003 Tutorial

Affine Invariance of Interest Points

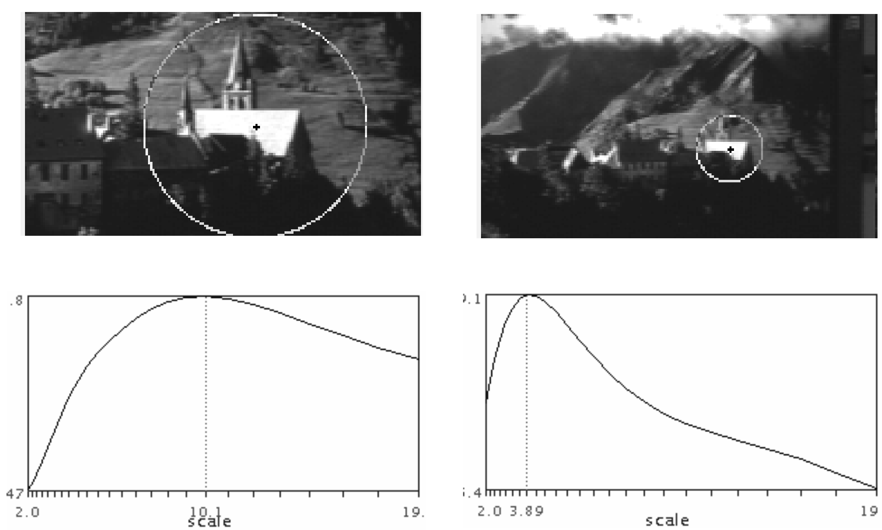
- Scale invariance is not sufficient for large baseline changes

- Affine invariant interest points

Scale invariant Harris points

- Multi-scale extraction of Harris interest points
- Selection of points at characteristic scale in scale space

Laplacian

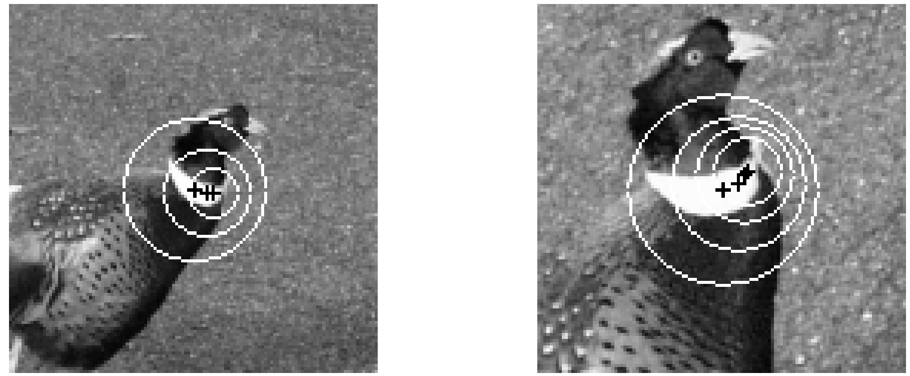


Chacteristic scale :

- maximum in scale space
- scale invariant

Scale invariant interest points

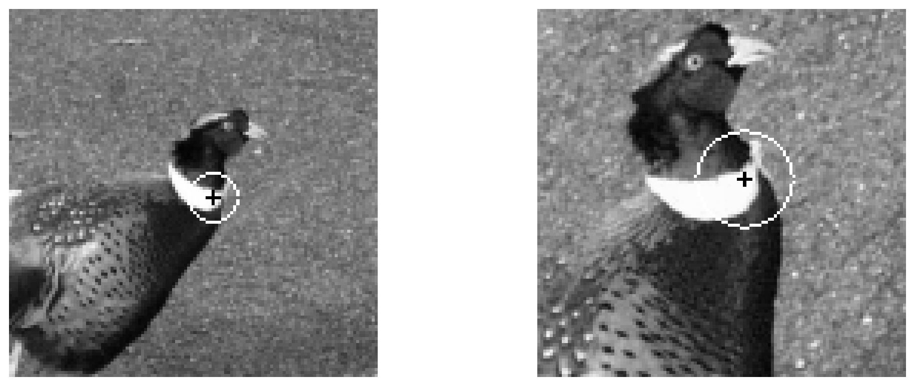
multi-scale Harris points



selection of points

at the characteristic scale

with Laplacian



➡ invariant points + associated regions [Mikolajczyk & Schmid'01]

Viewpoint changes

- Locally approximated by an affine transformation

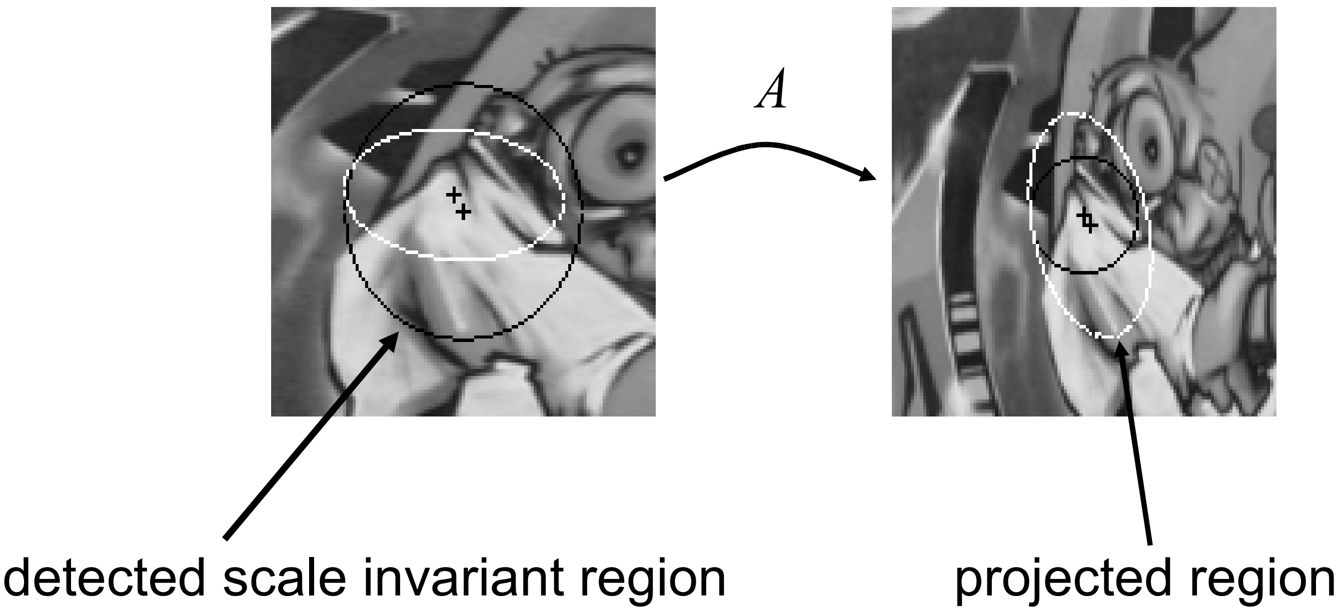


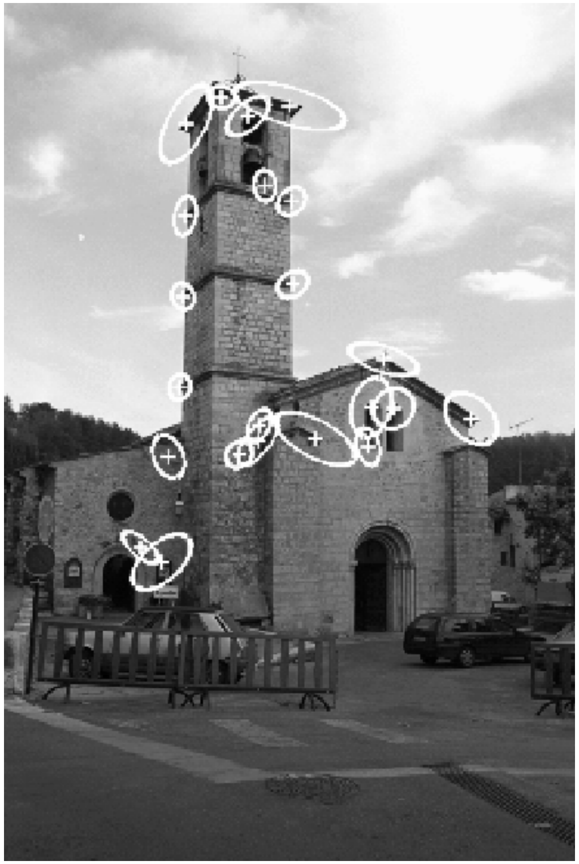
Image retrieval

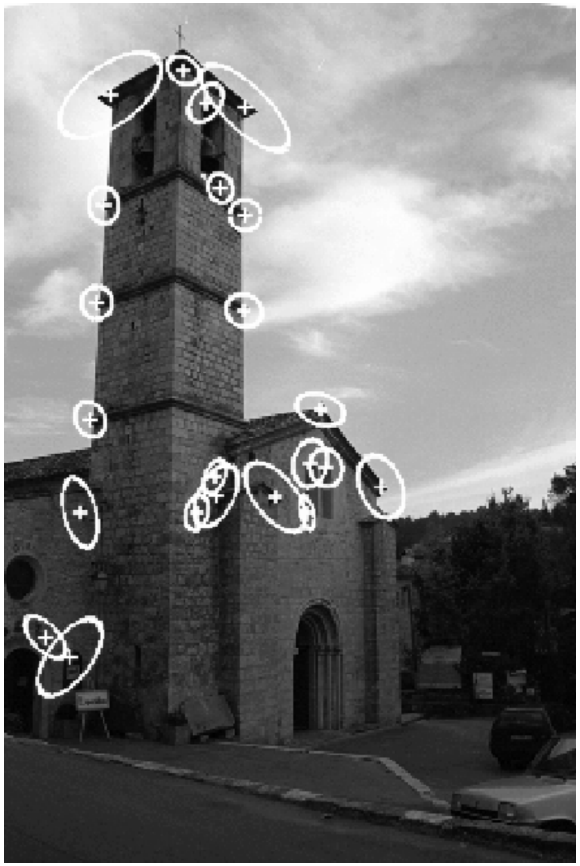
change in viewing angle

• • •

> 5000 images

Matches





22 correct matches

Image retrieval

change in viewing angle

+ scale change

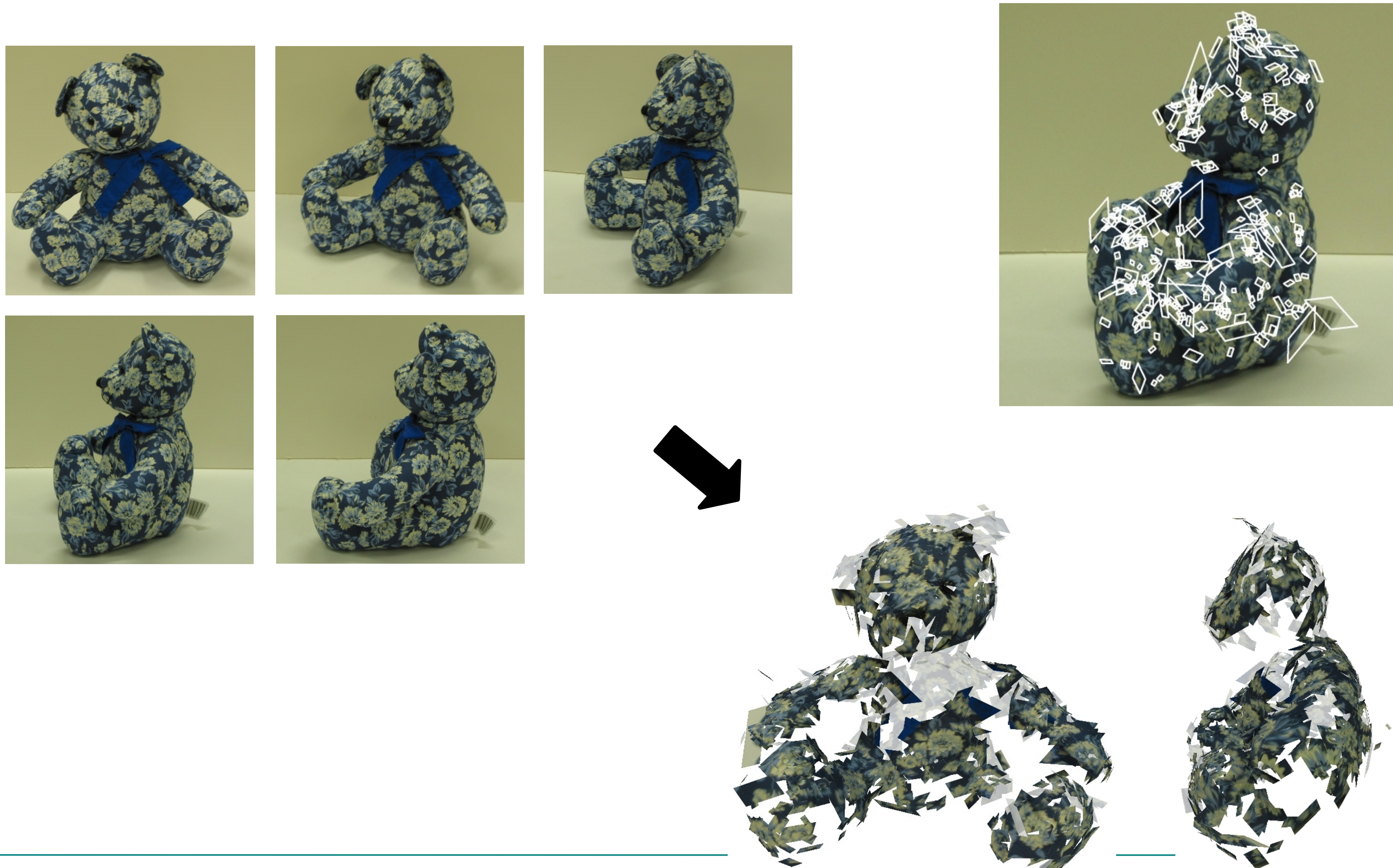
• • •

> 5000 images

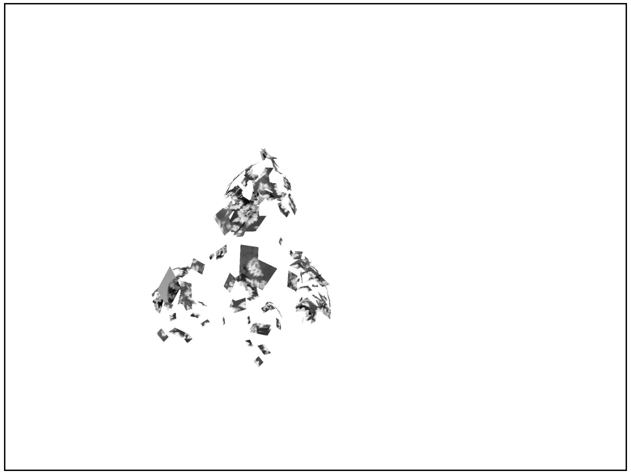
Matches

33 correct matches

3D Recognition



3D Recognition



3D object modeling and recognition using affine-invariant patches and multi-view spatial constraints,

F. Rothganger, S. Lazebnik, C. Schmid, J. Ponce,

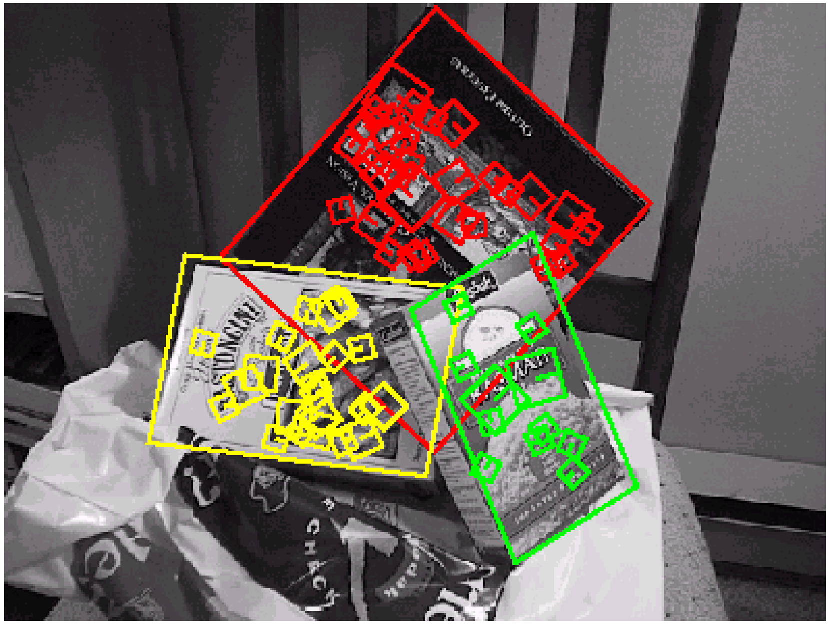
CVPR 2003

Planar texture models

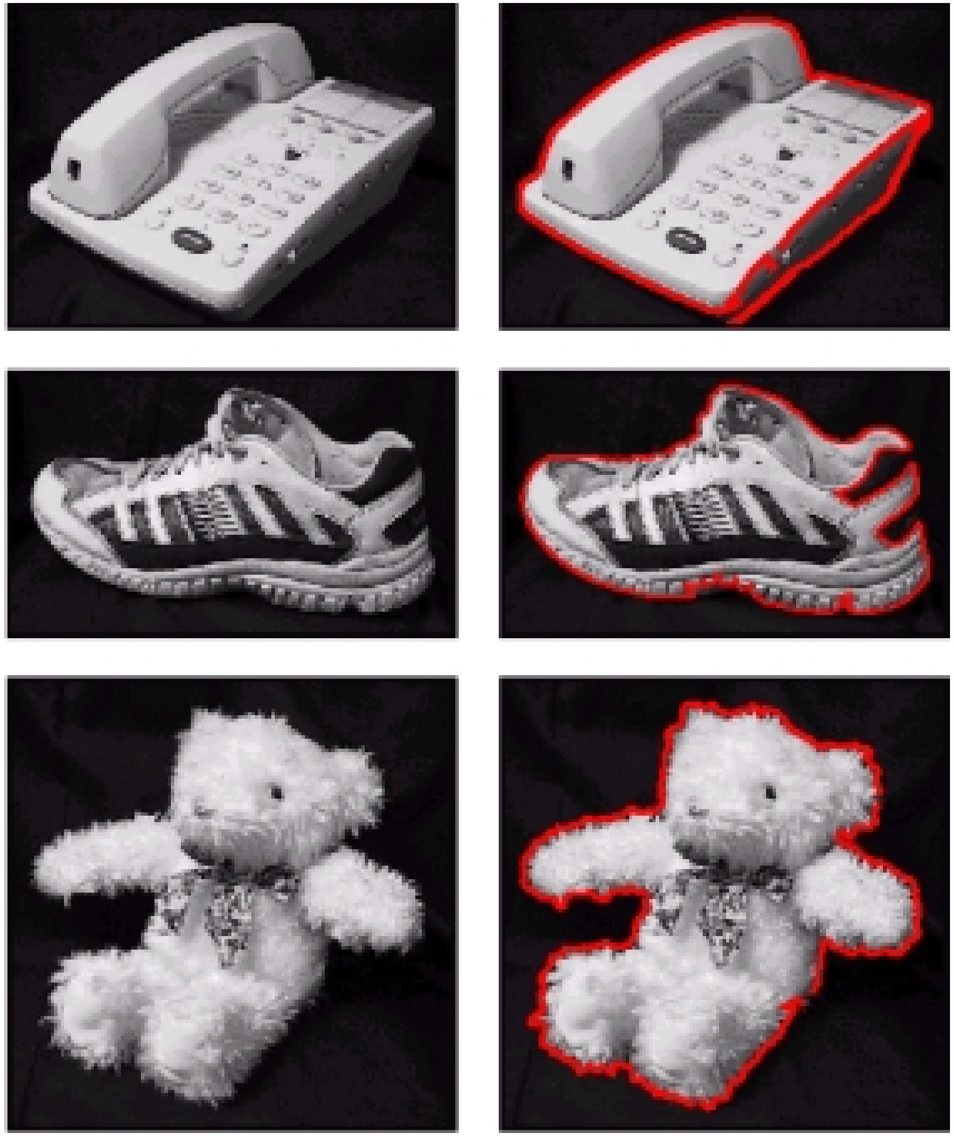
- Models for planar surfaces with SIFT keys

Planar recognition

- Planar surfaces can be reliably recognized at a rotation of 60° away from the camera
- Affine fit approximates perspective projection
- Only 3 points are needed for recognition



3D Object Recognition



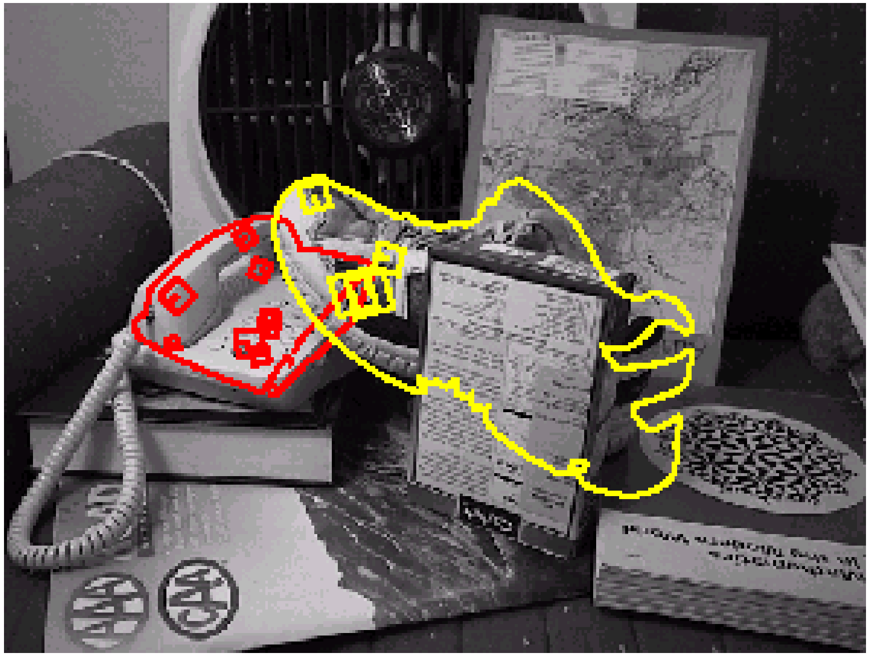
- Extract outlines with background subtraction

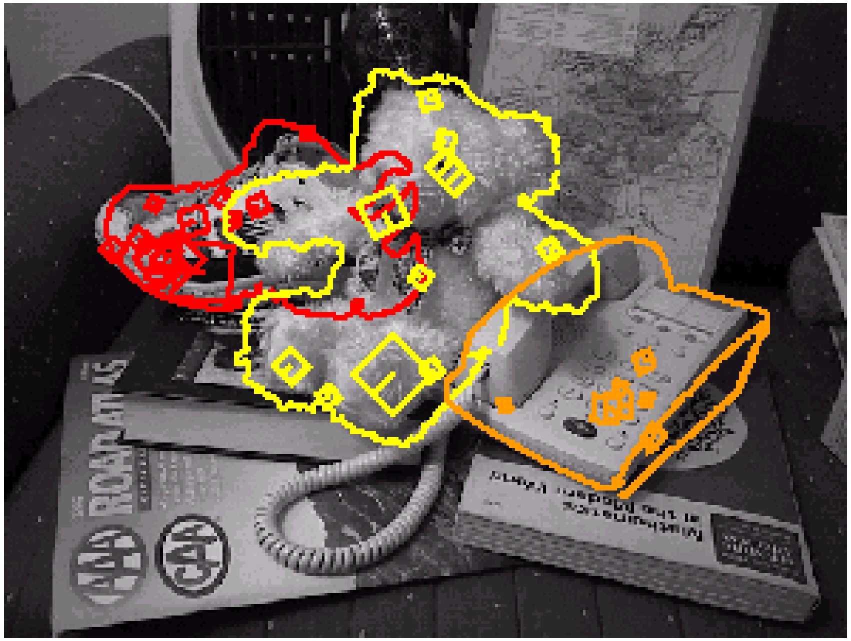
3D Object Recognition

Only 3 keys are needed for recognition, so extra keys provide robustness

Affine model is no longer as accurate

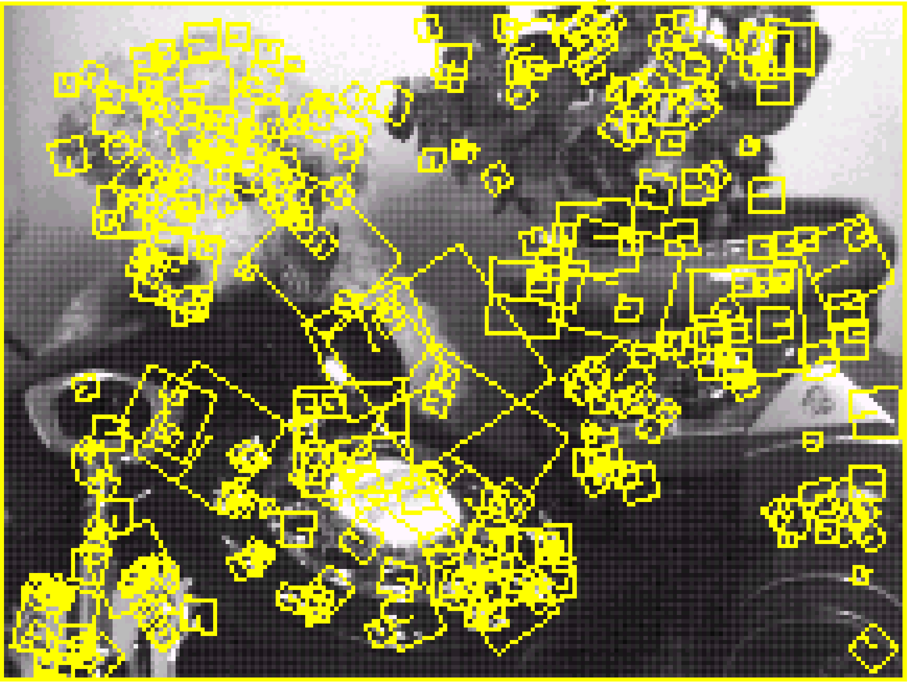
Recognition under occlusion





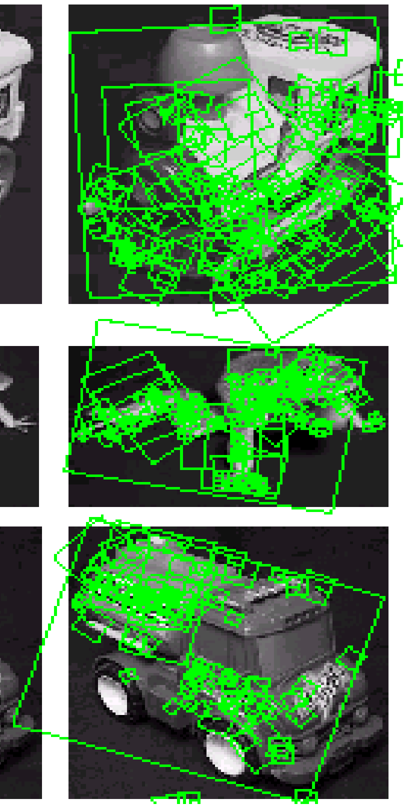
Test of illumination invariance

- Same image under differing illumination

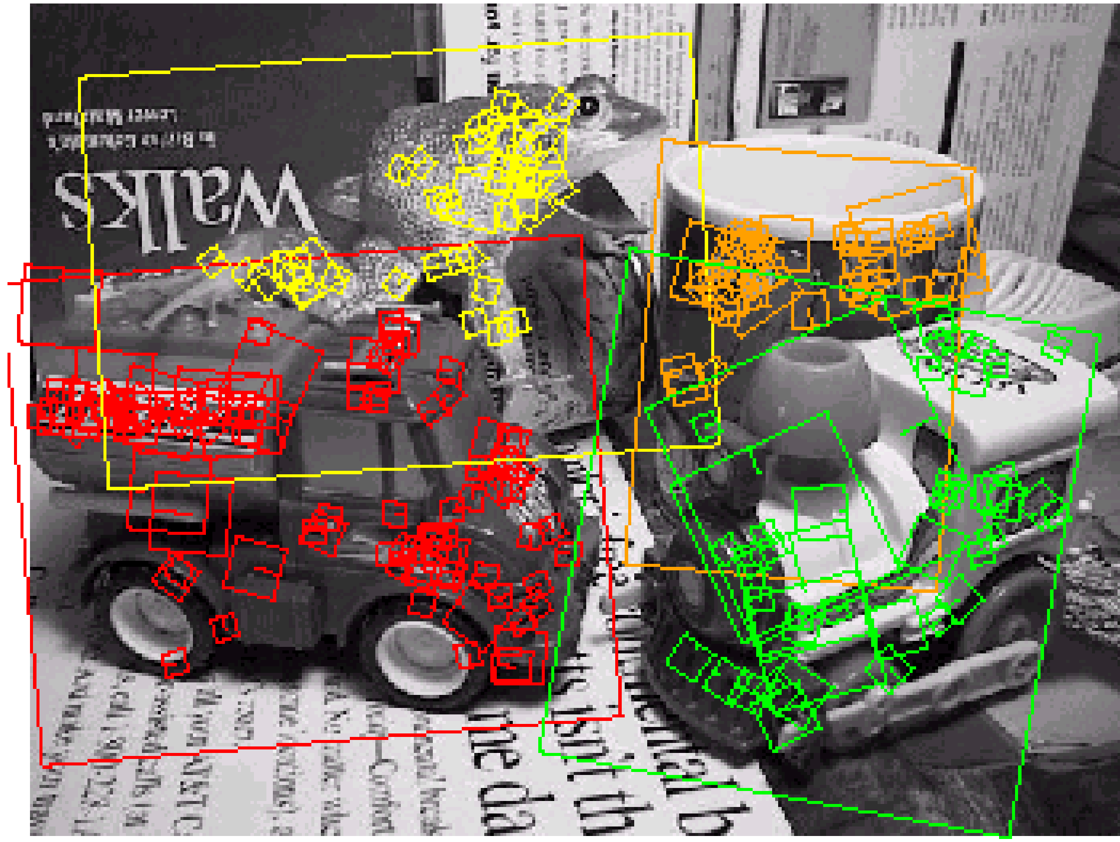


273 keys verified in final match

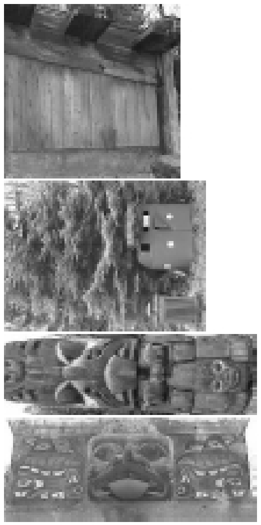
Examples of view interpolation



Recognition using View Interpolation

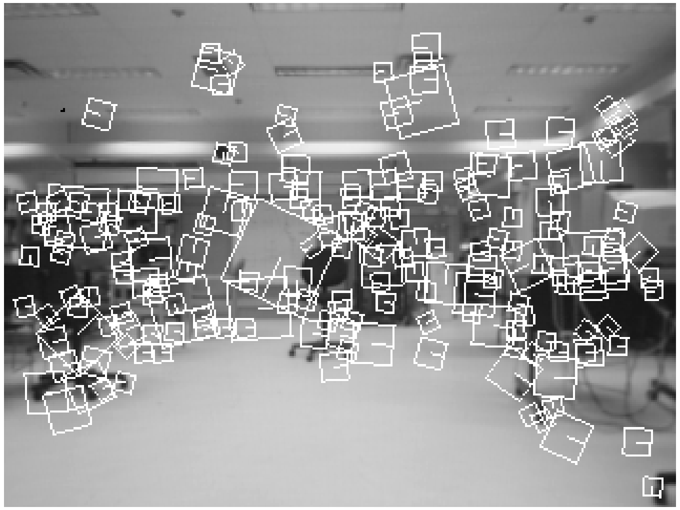


Location recognition

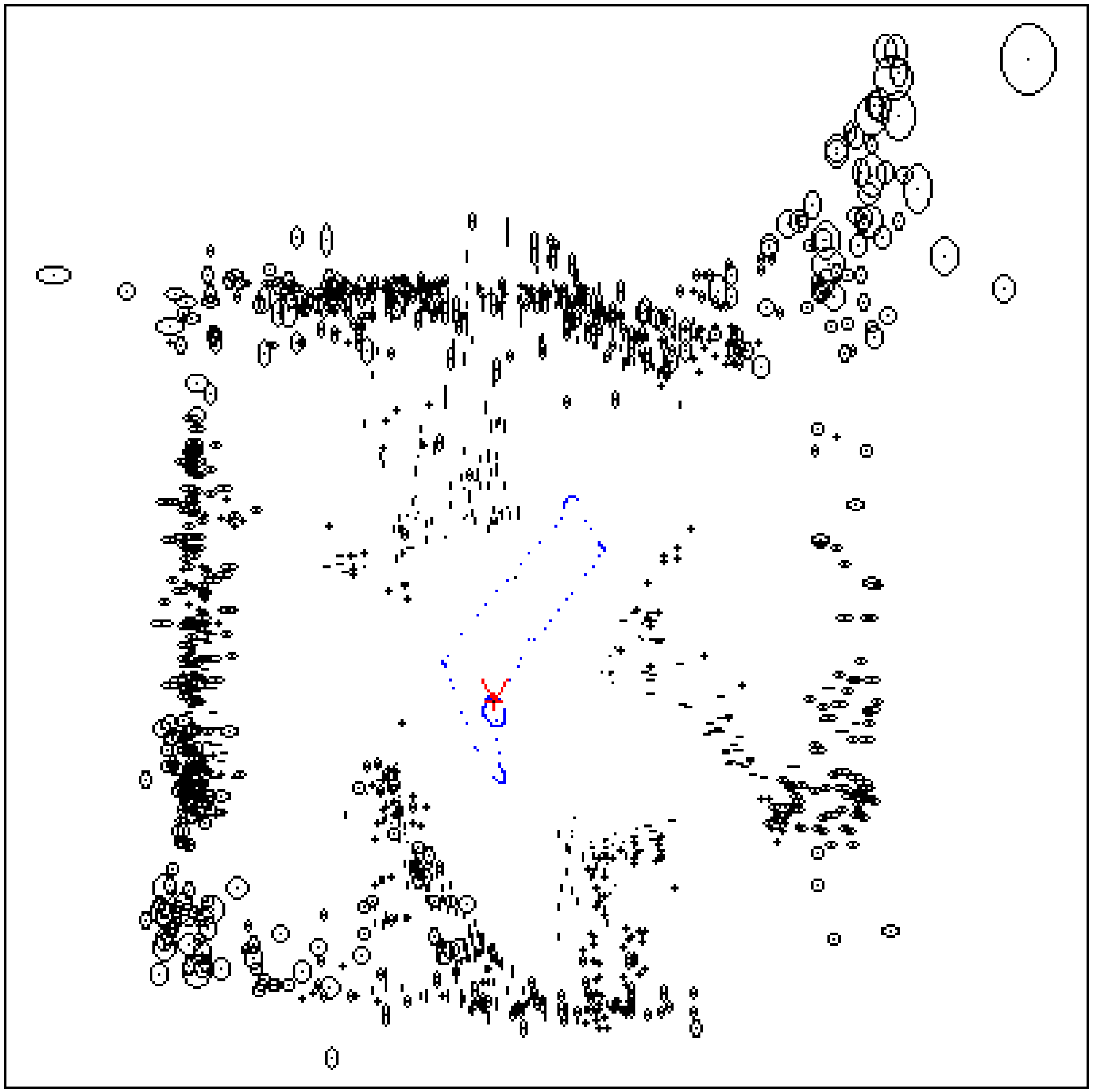


Robot Localization

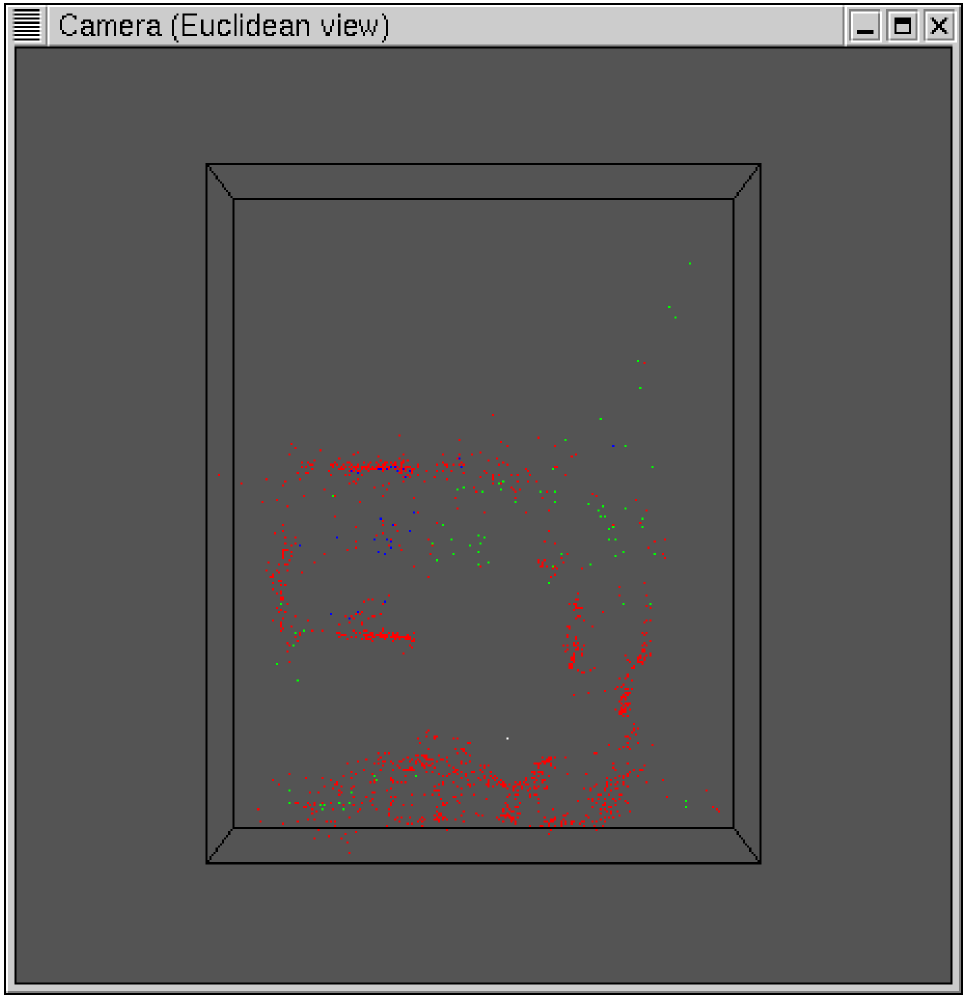
- Joint work with Stephen Se, Jim Little



Map continuously built over time



Locations of map features in 3D



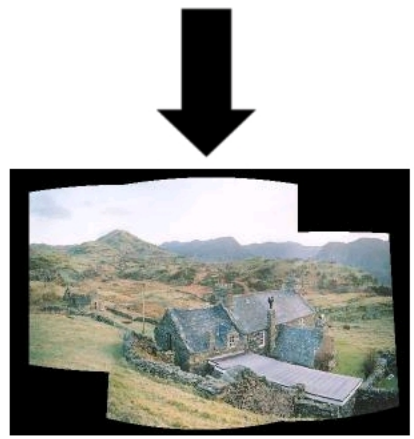
Recognizing Panoramas

- Matthew Brown and David Lowe
- Recognize overlap from an unordered set of images and automatically stitch together
- SIFT features provide initial feature matching

Panorama of our lab automatically assembled from 143 images

Multiple panoramas from an unordered image set

Input images



Output panorama 1

Image registration and blending

(a) *40 of 80 images registered*

(b) *All 80 images registered*

(c) *Rendered with multi-band blending*