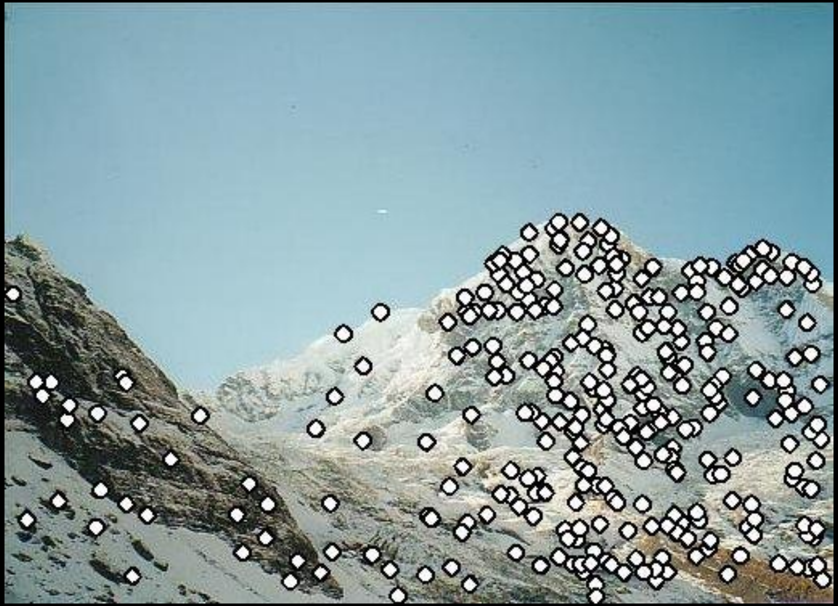
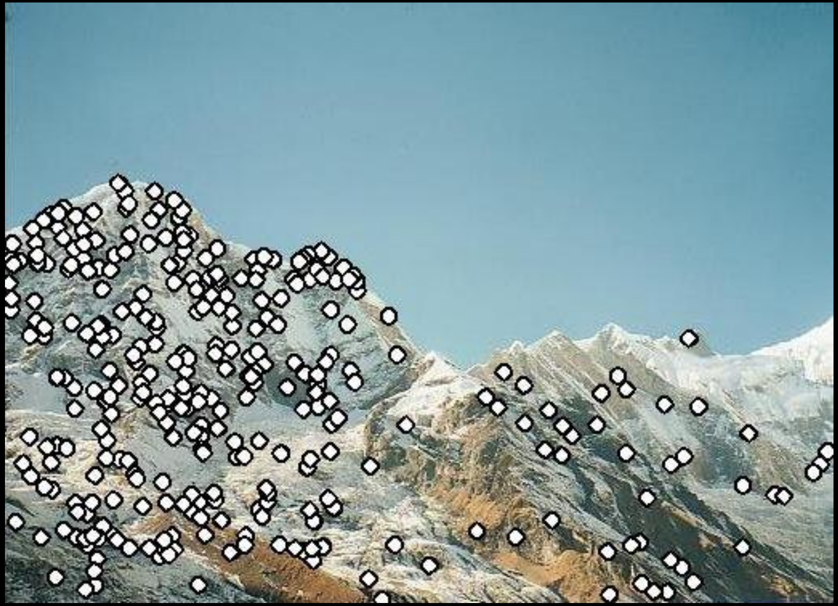
CS4495/6495
Introduction to Computer Vision

4B-L1 *SIFT descriptor*

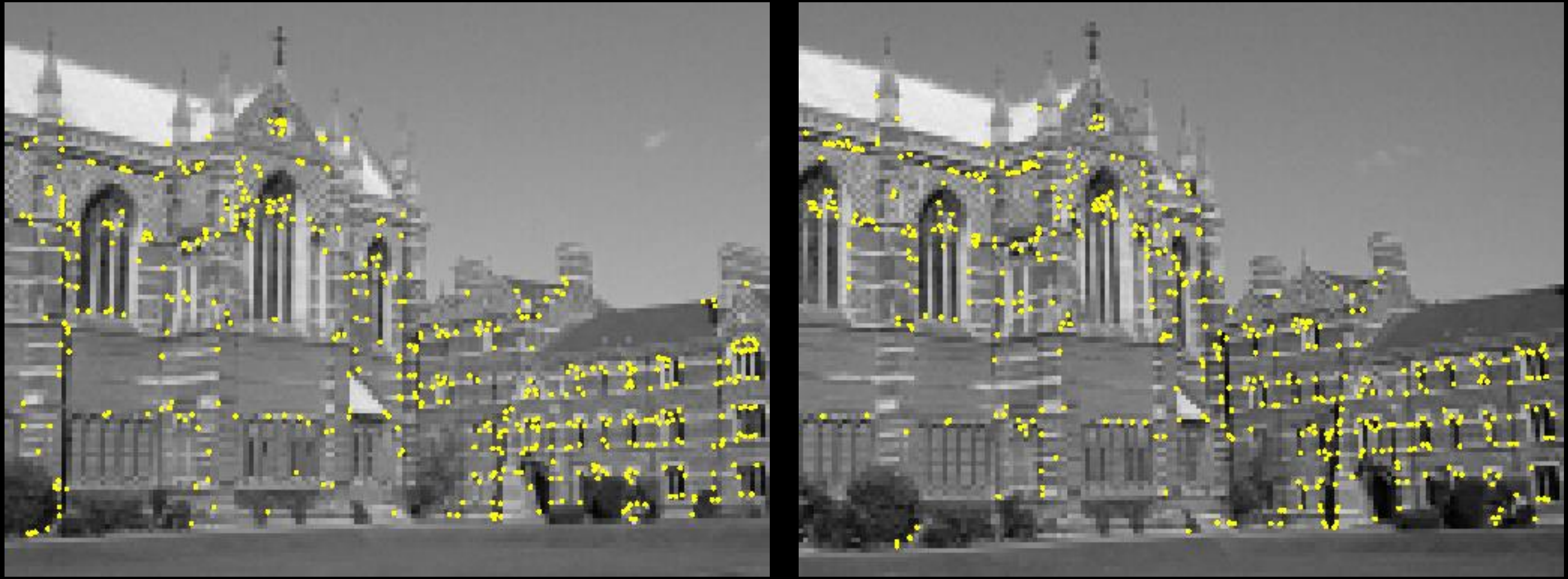
Point Descriptors

- Last time: How to detect interest points





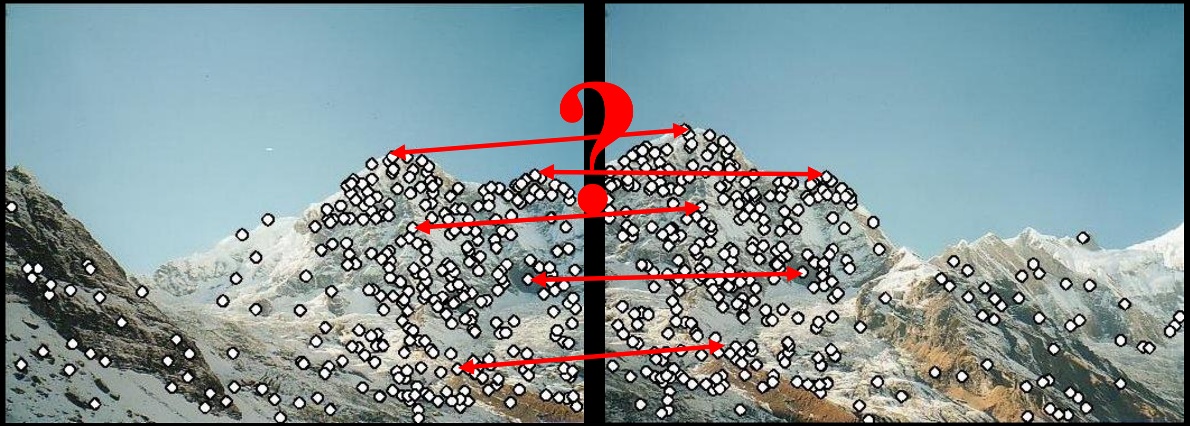
Harris detector



Interest points extracted with Harris (~ 500 points)

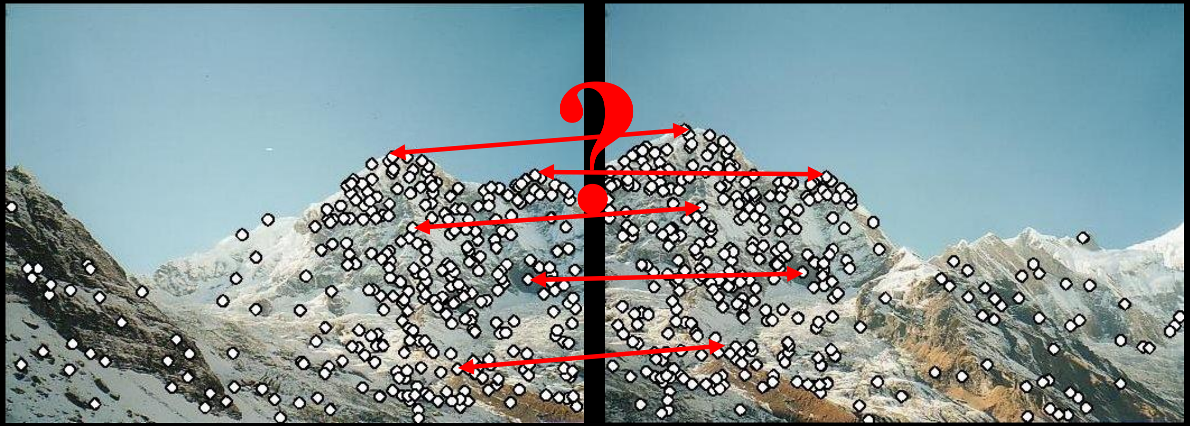
Point Descriptors

- Now: How to match them?



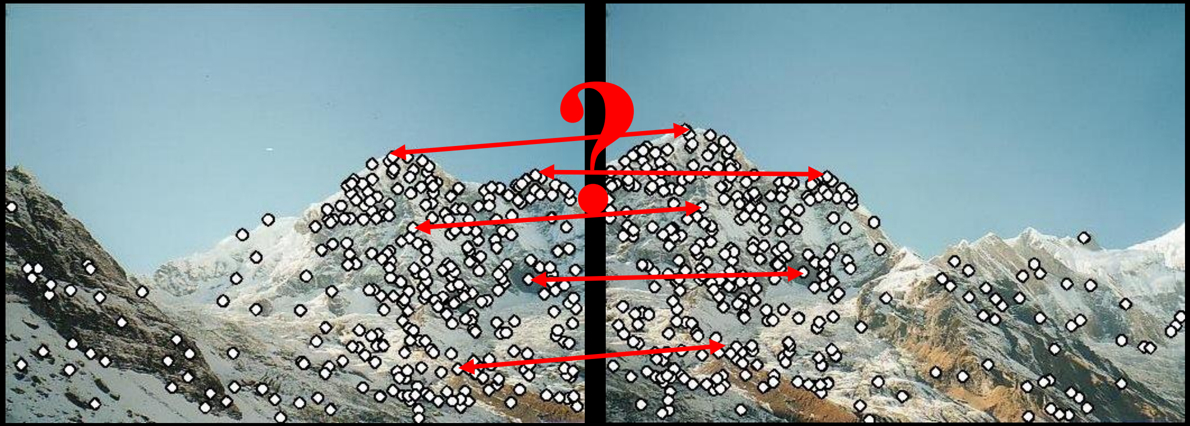
Point Descriptors

- We need to describe them – a "*descriptor*"



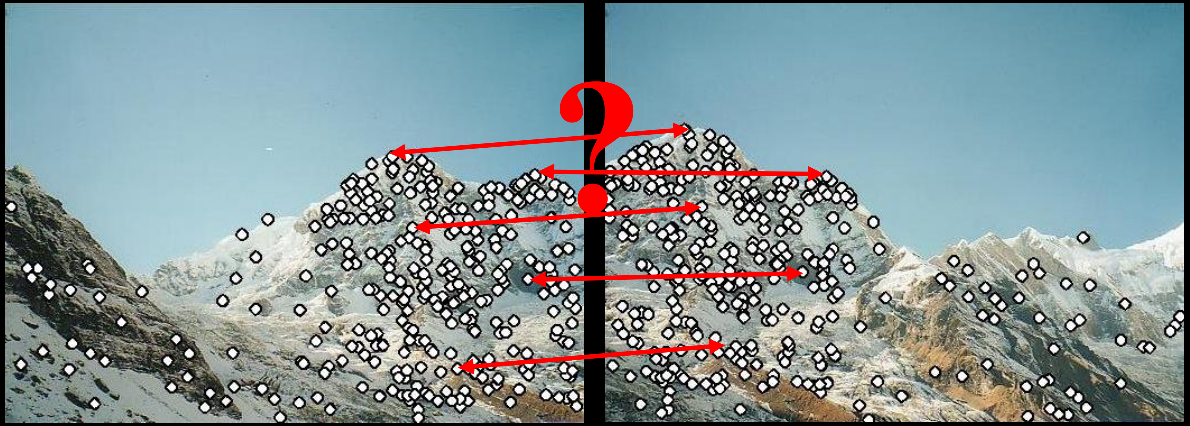
Criteria for Point Descriptors

- We want the descriptors to be the (almost) same in both image – *invariant.*



Criteria for Point Descriptors

- We also need the descriptors to be *distinctive*.



Simple solution?

- Harris gives good detection – and we also know the scale.
- Why not just use correlation to check the match of the window around the feature in image 1 with every feature in image 2?

Simple solution? *Not so good!*

- Not so good because:
  - Correlation is not rotation invariant - why do we want this?
  - Correlation is sensitive to photometric changes.
  - Normalized correlation is sensitive to non-linear photometric changes and even slight geometric ones.
  - Could be slow – check all features against all features

SIFT: Scale Invariant Feature Detection

- Motivation: The Harris operator was not invariant to scale and correlation was not invariant to rotation.

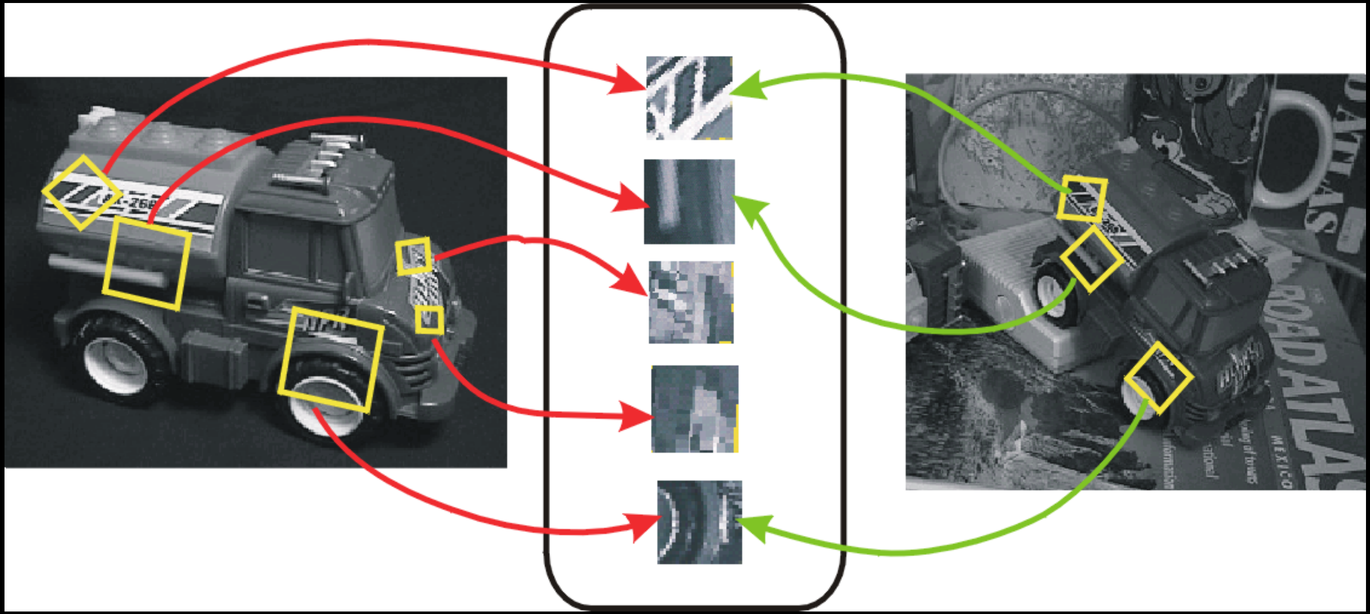
- For better image matching, Lowe's goals were:
  - To develop an interest operator – a *detector* – that is invariant to scale and rotation.
  - Also: create a descriptor that was robust to the variations corresponding to typical viewing conditions. *The descriptor is the most-used part of SIFT.*

Idea of SIFT

- Image content is represented by a constellation of local features that are invariant to translation, rotation, scale, and other imaging parameters

Idea of SIFT

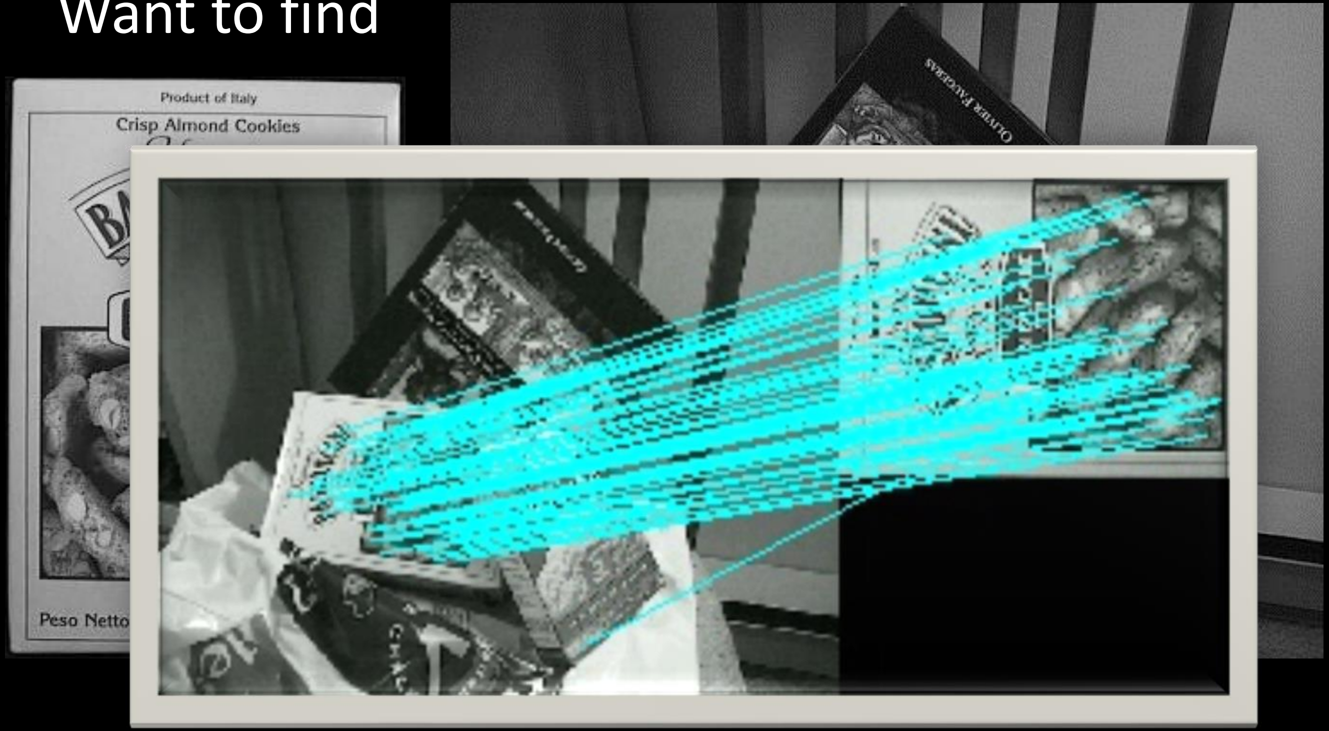
SIFT Features



Another version of the problem…

Want to find

… in here



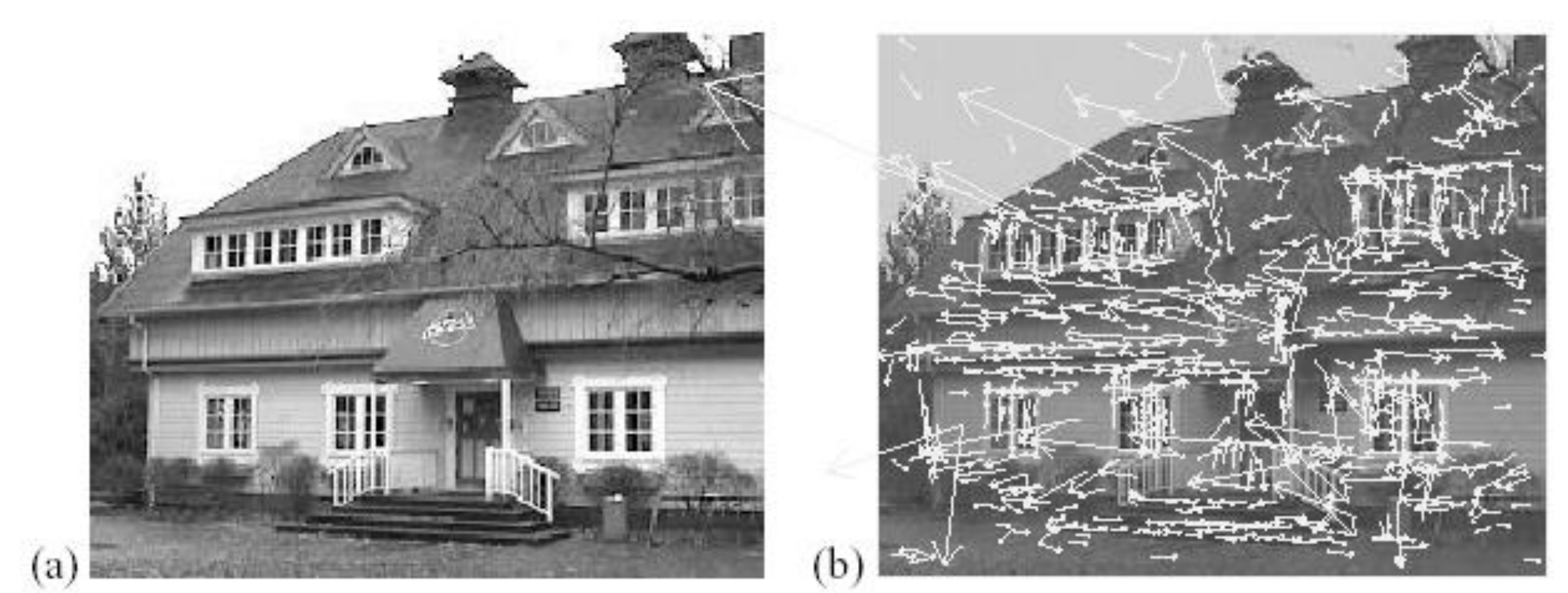
Overall SIFT Procedure

- Scale-space extrema detection
- Keypoint localization

Or use Harris-Laplace or other method

- Orientation assignment
- Keypoint description

Example of keypoint detection



(a) 233x189 image
(b) 832 DOG extrema

Overall SIFT Procedure

1. Scale-space extrema detection
2. Keypoint localization
3. Orientation assignment

   Compute best orientation(s) for each keypoint region.

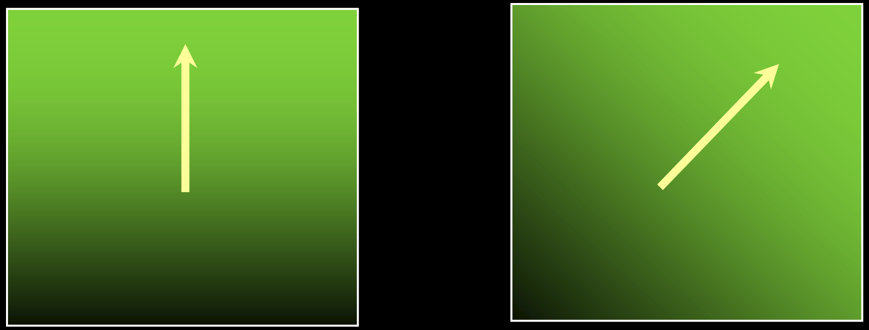
4. Keypoint description

   Use local image gradients at selected scale and rotation

   to describe each keypoint region.

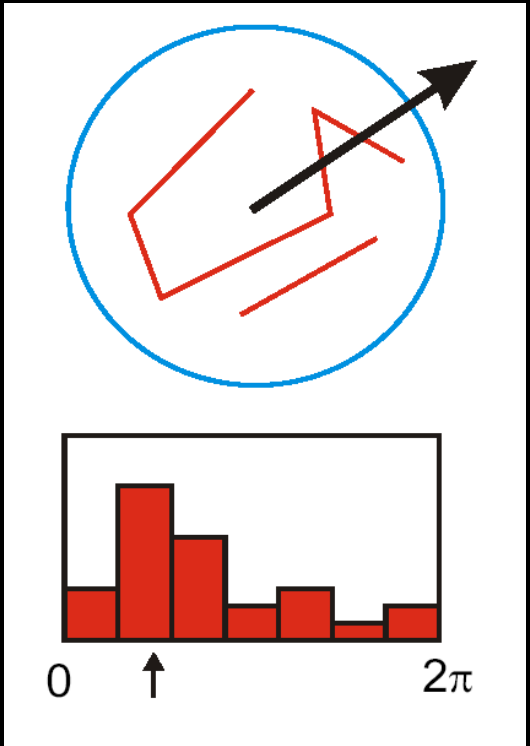
Descriptors Invariant to Rotation

- Find the dominant direction of gradient – that is the base orientation.



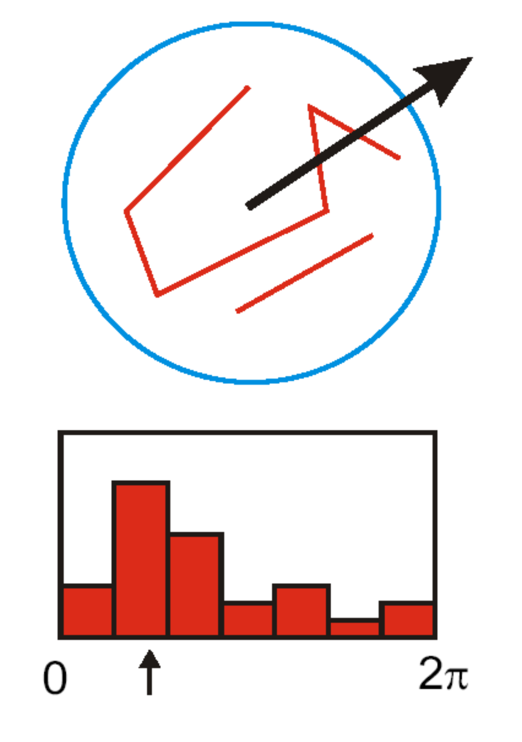
Compute image derivatives relative to this orientation

Orientation assignment



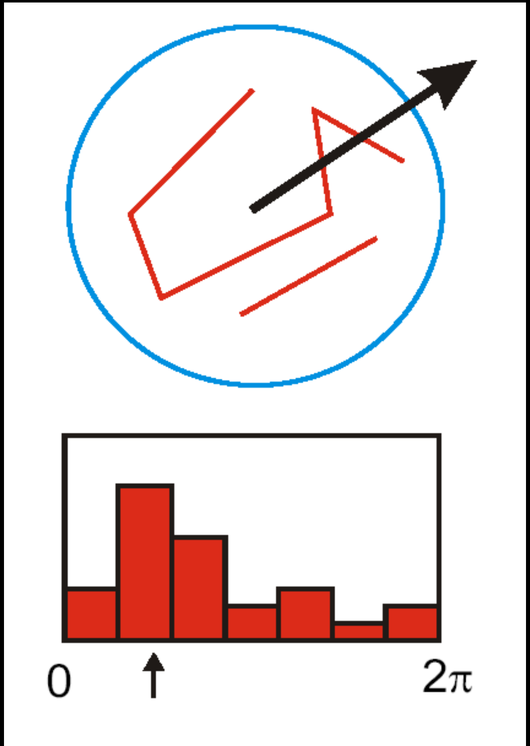
- Create histogram of local gradient directions at *selected* scale – 36 bins

Orientation assignment



- Assign canonical orientation at peak of smoothed histogram

Orientation assignment



- Each *keypoint* now specifies stable 2D coordinates (x, y, scale, orientation) – invariant to those.

4. Keypoint Descriptors

- Next is to compute a descriptor for the local image region about each keypoint that is:
  - Highly ***distinctive***
  - As ***invariant*** as possible to variations such as changes in viewpoint and illumination

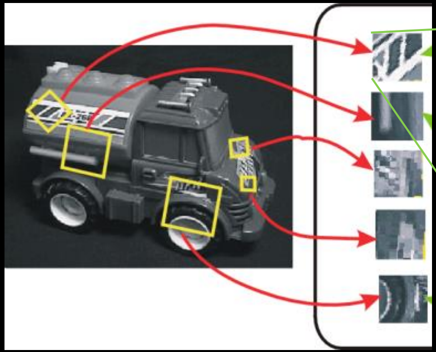
But first… normalization…

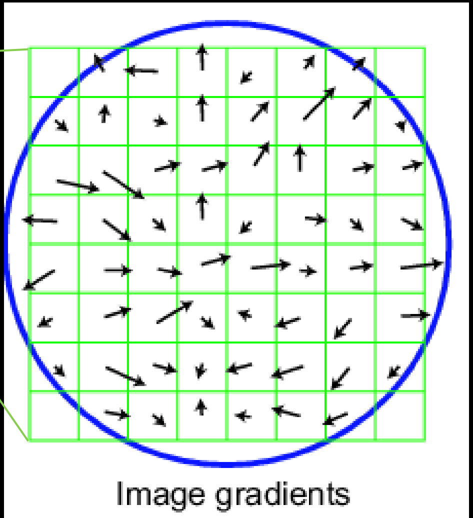
- Rotate the window to standard orientation
- Scale the window size based on the scale at which the point was found.

SIFT vector formation

Compute a feature vector based upon:

- histograms of gradients

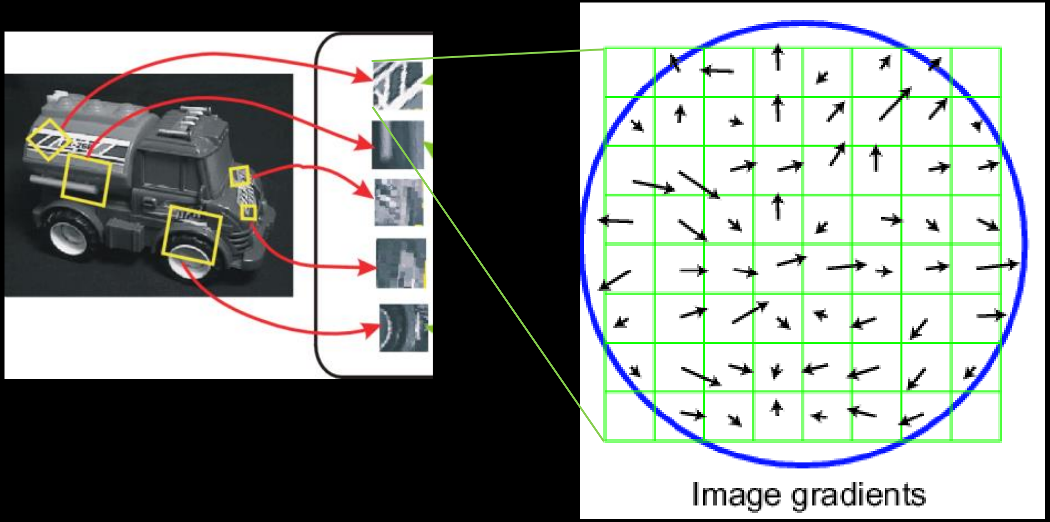




SIFT vector formation

Compute a feature vector based upon:

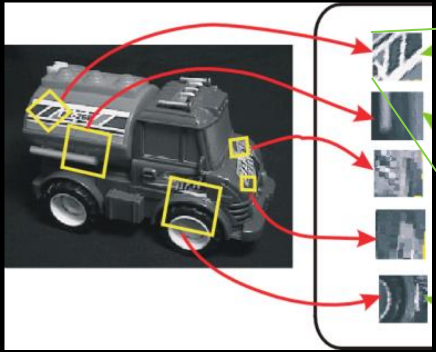
- weighted by a centered Gaussian,

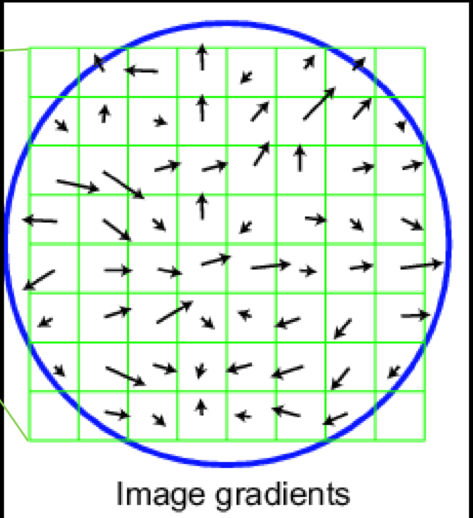


SIFT vector formation

Compute a feature vector based upon:

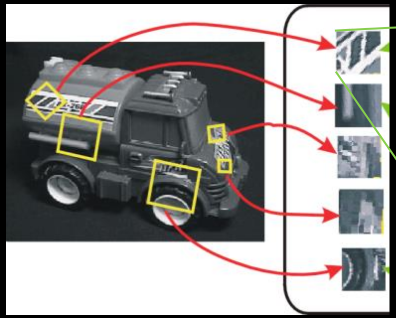
- weighted by the magnitude of the gradient

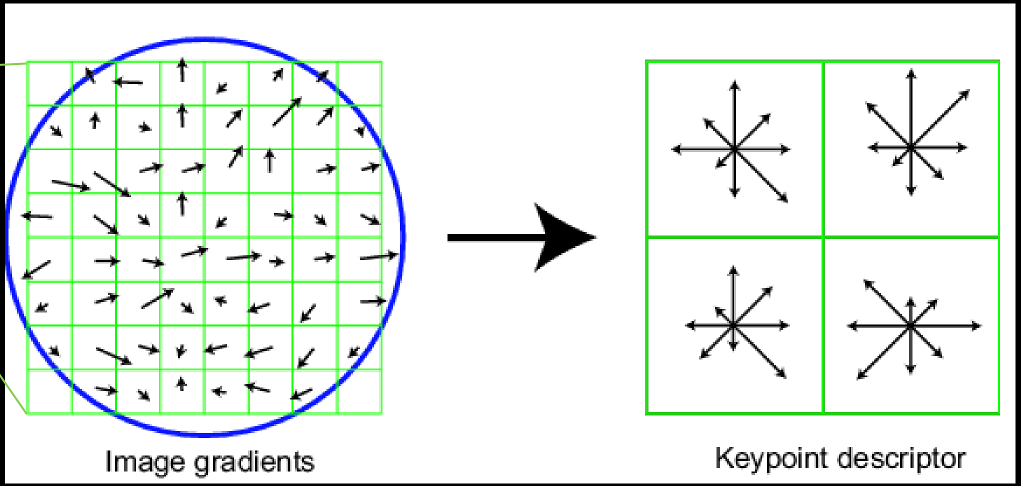




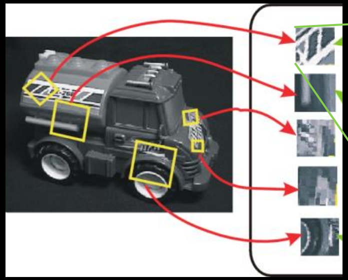
SIFT vector formation

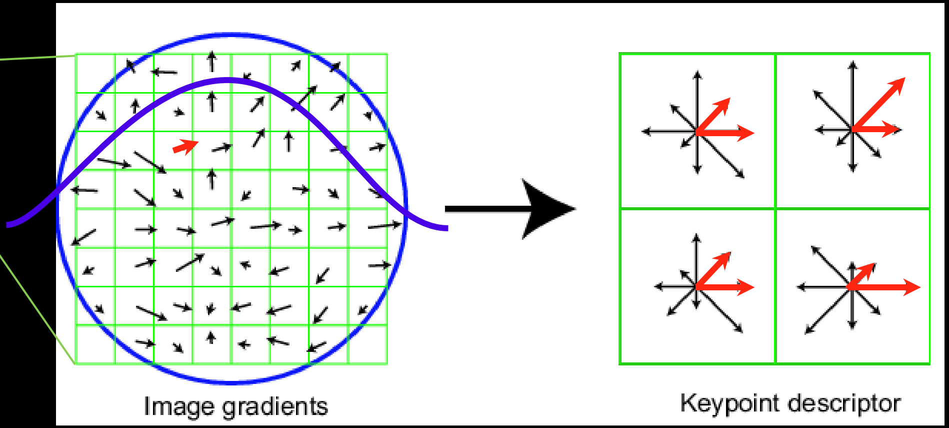
showing only 2x2 here but it really is 4x4





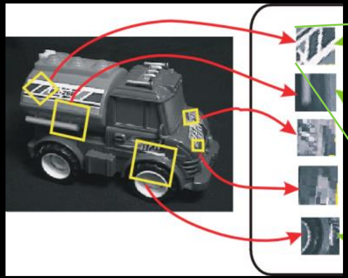
Ensure smoothness

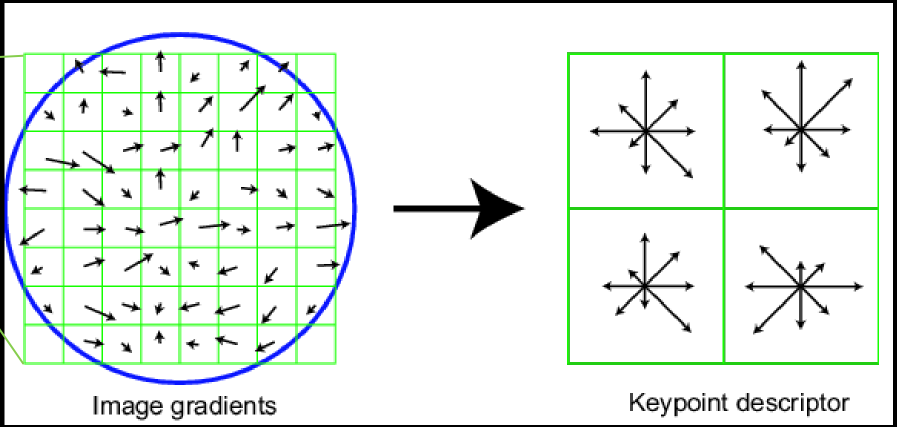




Reduce effect of illumination

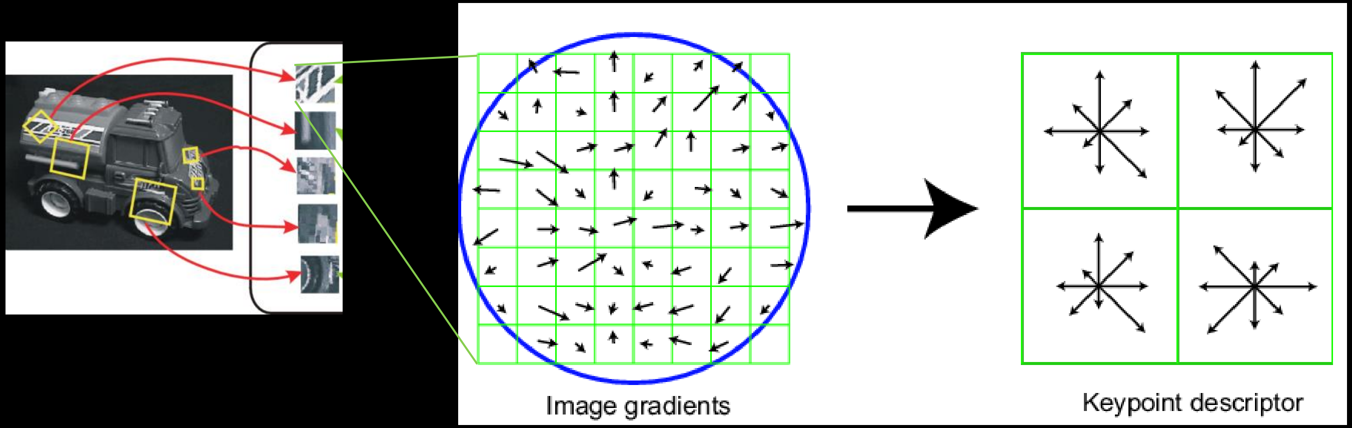
- Clip gradient magnitudes to avoid excessive influence of high gradients
  - after rotation normalization, clamp gradients >0.2





Reduce effect of illumination

- 128-dim vector normalized to magnitude 1.0



Evaluating the SIFT descriptors

- Database images were subjected to rotation, scaling, affine stretch, brightness and contrast changes, and added noise.
- Feature point detectors and descriptors were compared before and after the distortions.
- Mostly looking for stability with respect to change.

Sensitivity to parameters

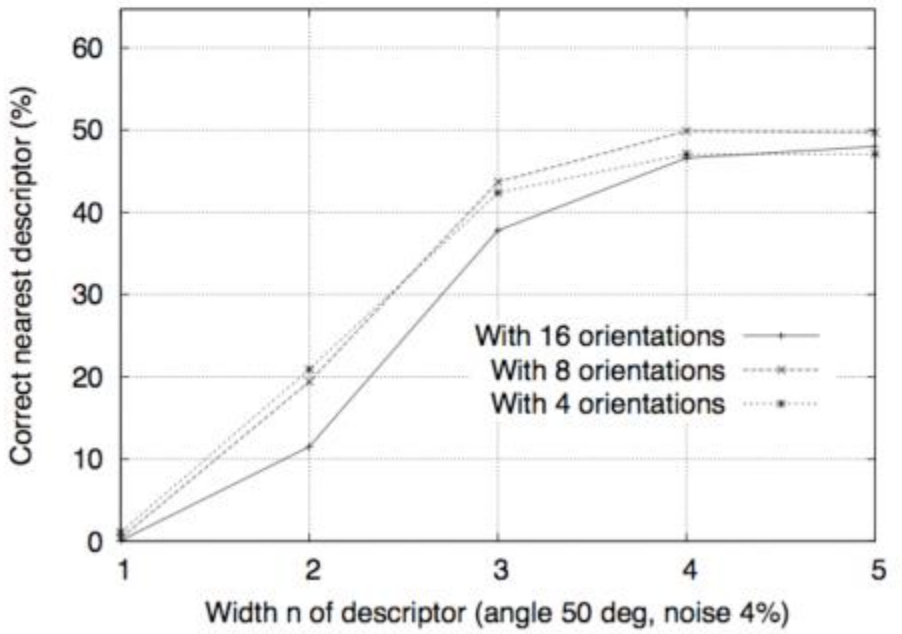
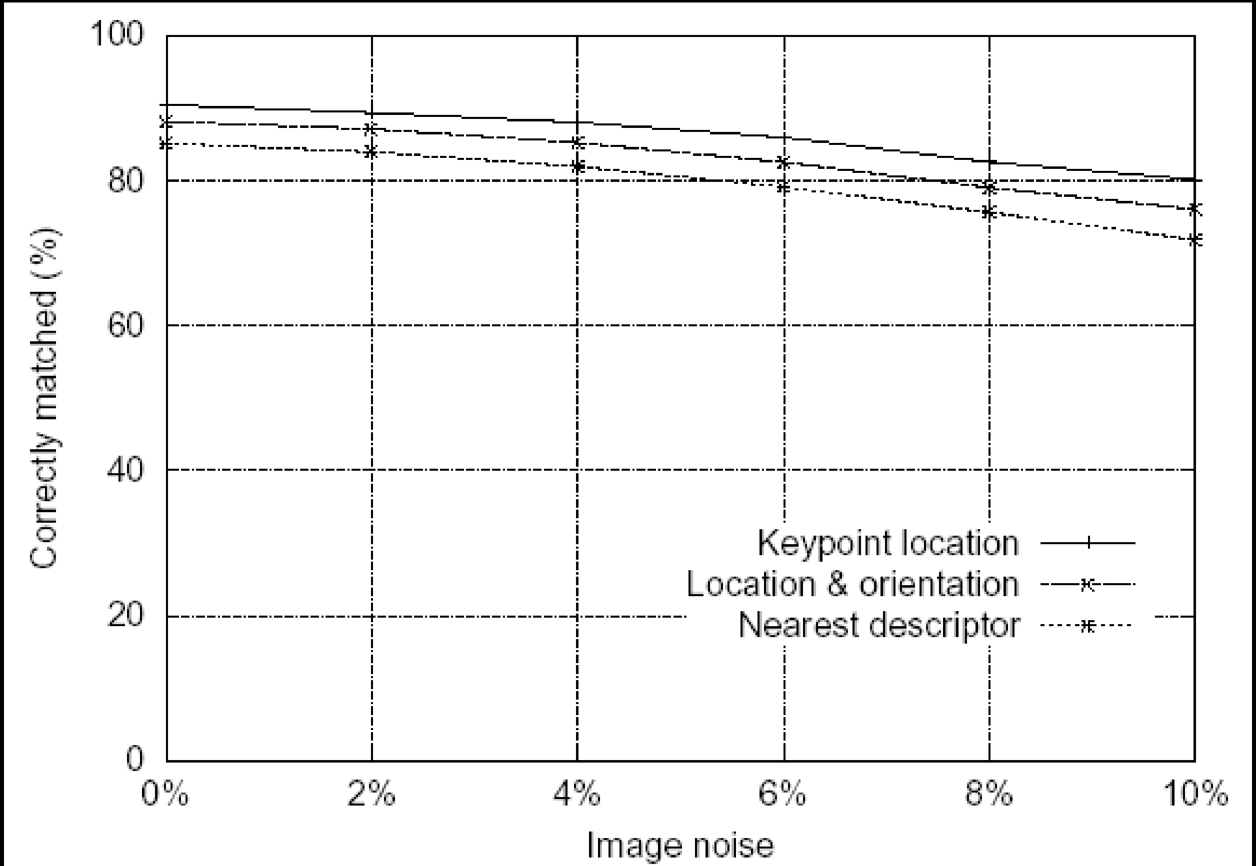


Figure 8: This graph shows the percent of keypoints giving the correct match to a database of 40,000 keypoints as a function of width of the $n \times n$ keypoint descriptor and the number of orientations in each histogram. The graph is computed for images with affine viewpoint change of 50 degrees and addition of 4% noise.

Feature stability to noise

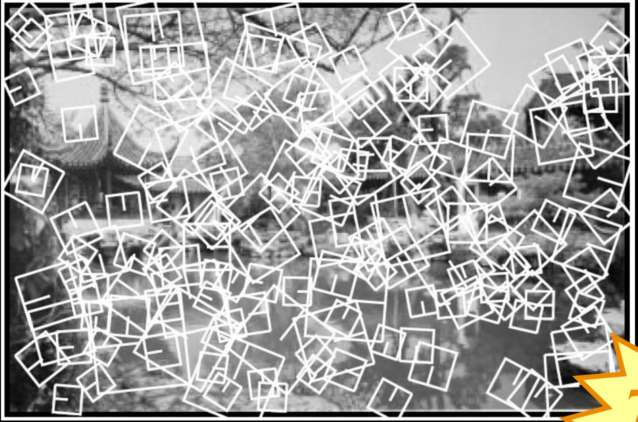


Experimental results - summary

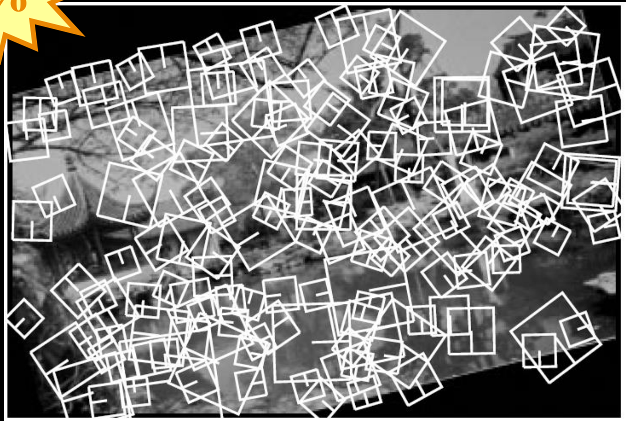
Image transformation	Location and scale match	Orientation match
Decrease constrast by 1.2	89.0 %	86.6 %
Decrease intensity by 0.2	88.5 %	85.9 %
Rotate by 20°	85.4 %	81.0 %
Scale by 0.7	85.1 %	80.3 %
Stretch by 1.2	83.5 %	76.1 %
Stretch by 1.5	77.7 %	65.0 %
Add 10% pixel noise	90.3 %	88.4 %
All previous	**78.6 %**	**71.8 %**

20 different images, around 15,000 keys

Experimental results



Original image



Keypoints on image after rotation (15°), scaling (90%), horizontal stretching (110%), change of brightness (-10%) and contrast (90%), and addition of pixel noise

SIFT matching object pieces (for location)

SIFT matching object pieces (for location)