MAIA

SIFT

Distinctive Image Features from Scale-Invariant Keypoints (SIFT)

Lowe, D.G. (2004). *Distinctive Image Features from Scale – Invariant Keypoints*. International Journal of Computer Vision, 60, 2 (2004), pp. 91-110. http://www.cs.ubc.ca/~lowe/pubs.html

SIFT Features

- SIFT = Scale Invariant Feature Transform

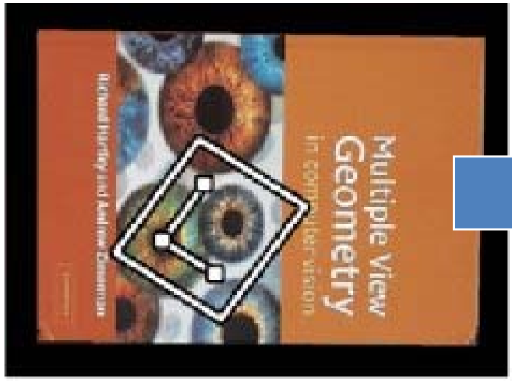
  Extract image features

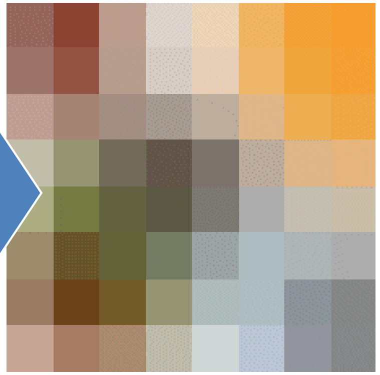
  - Invariant to scale and rotation
  - Partially invariant to change of illumination and change of 3D viewpoint

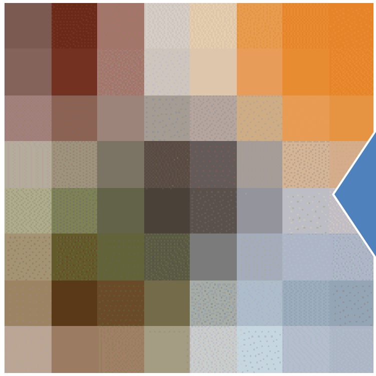
- Match features in a database

  Object recognition

SIFT Features



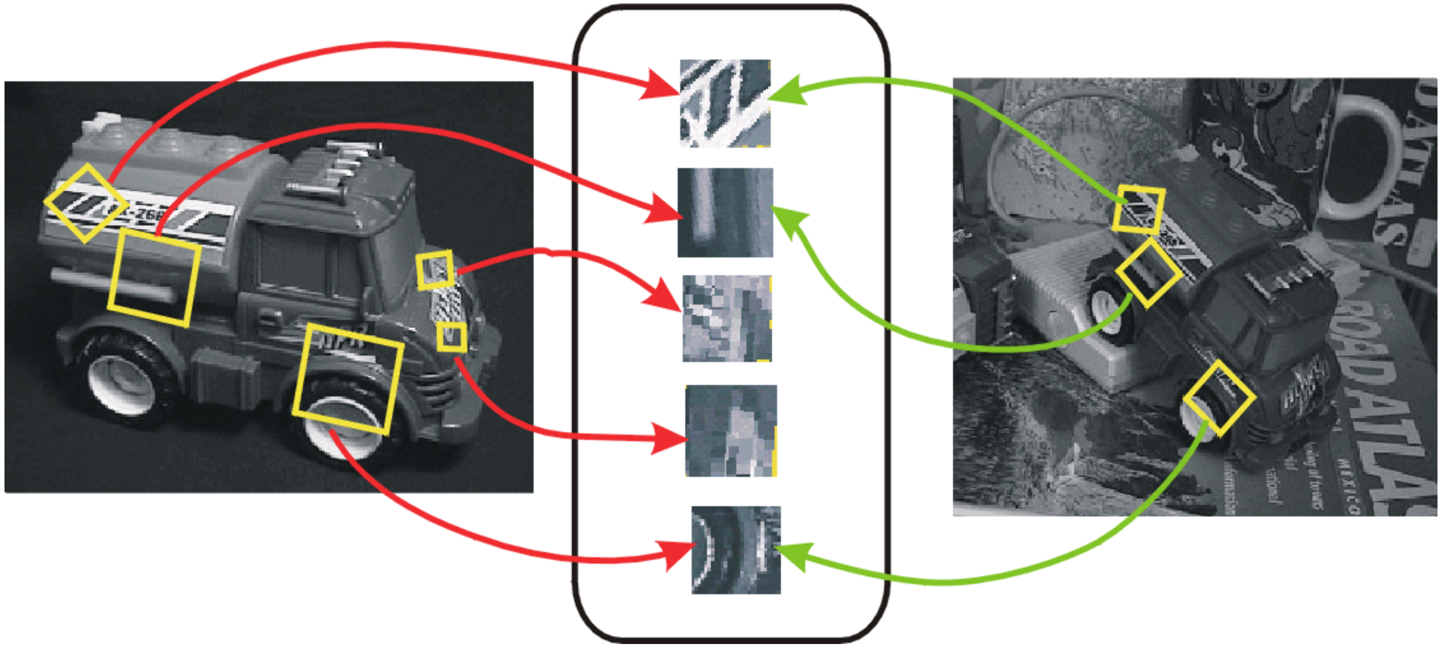




Slides extracted from D.G.
Lowe

SIFT Features

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



SIFT Features

SIFT Features

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

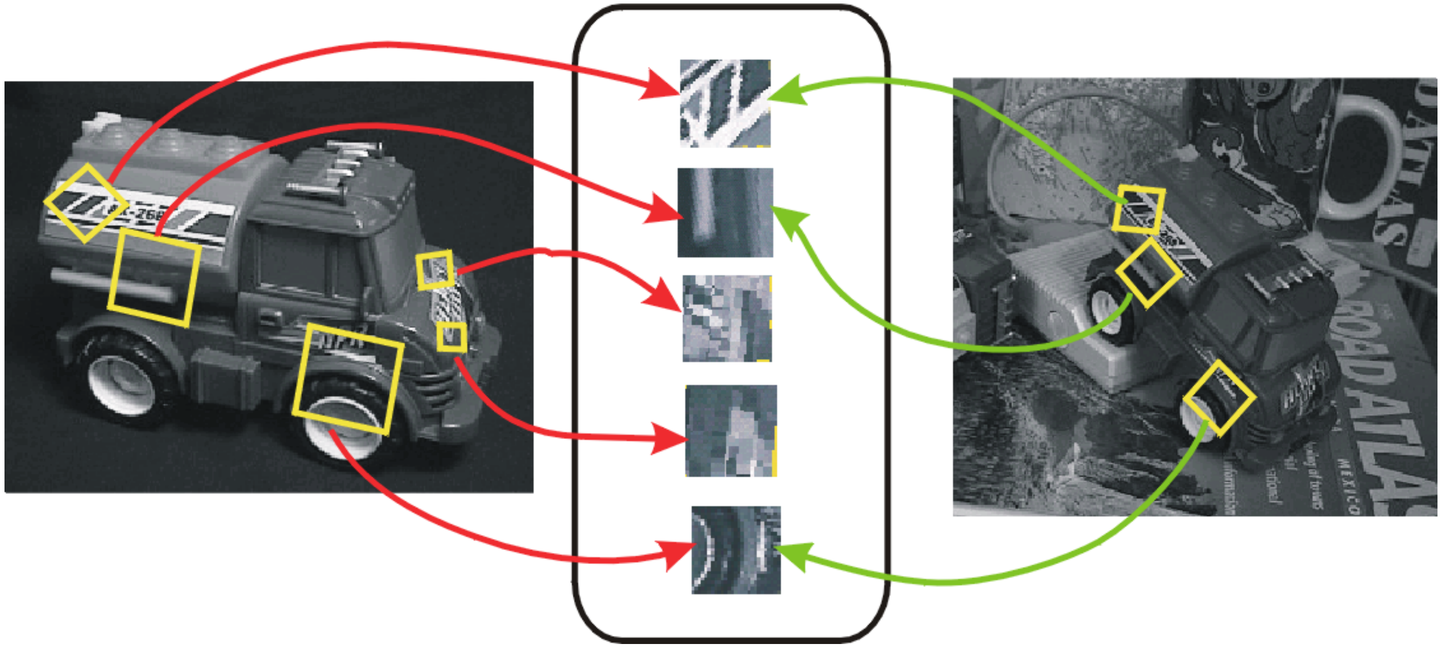
Universitat de Girona

SIFT Features

Slides extracted from D.G. Lowe

SIFT Features

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



SIFT Features

SIFT Features

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

SIFT overview

Scale extrema detection

Search for candidate points that are invariant to scale and rotation (scale-space pyramid)

↓

Keypoint localization

Reject candidates that have low contrast or are localized along an edge

↓

Orientation assignment

One or more orientations are assigned to the keypoints based on the local image gradient directions

↓

Keypoint descriptor

Assign a vector of data to each keypoint based on the local image gradient directions

↓

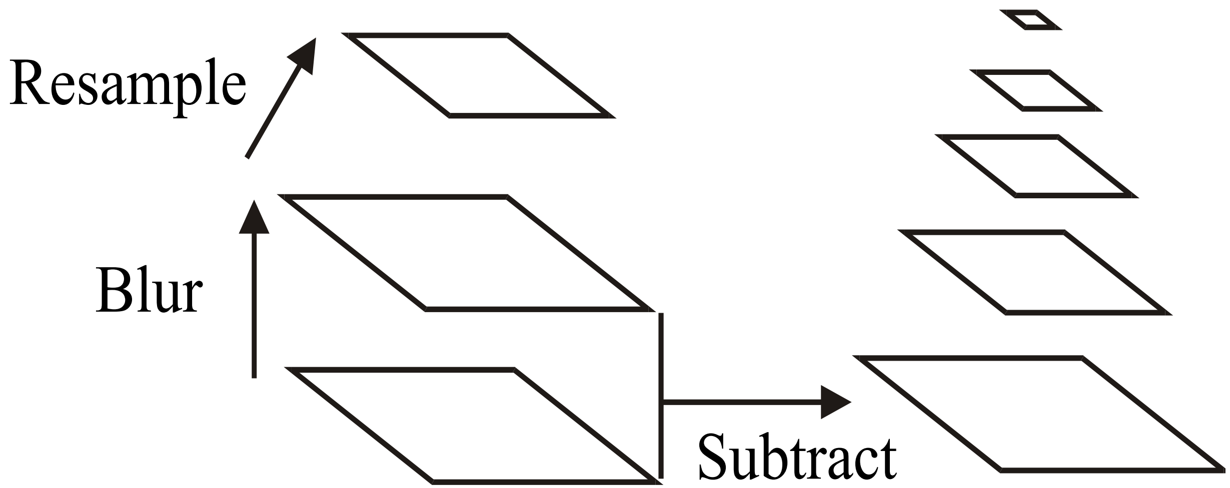
SIFT KEYS

SIFT Features

- **Build Scale-Space Pyramid**

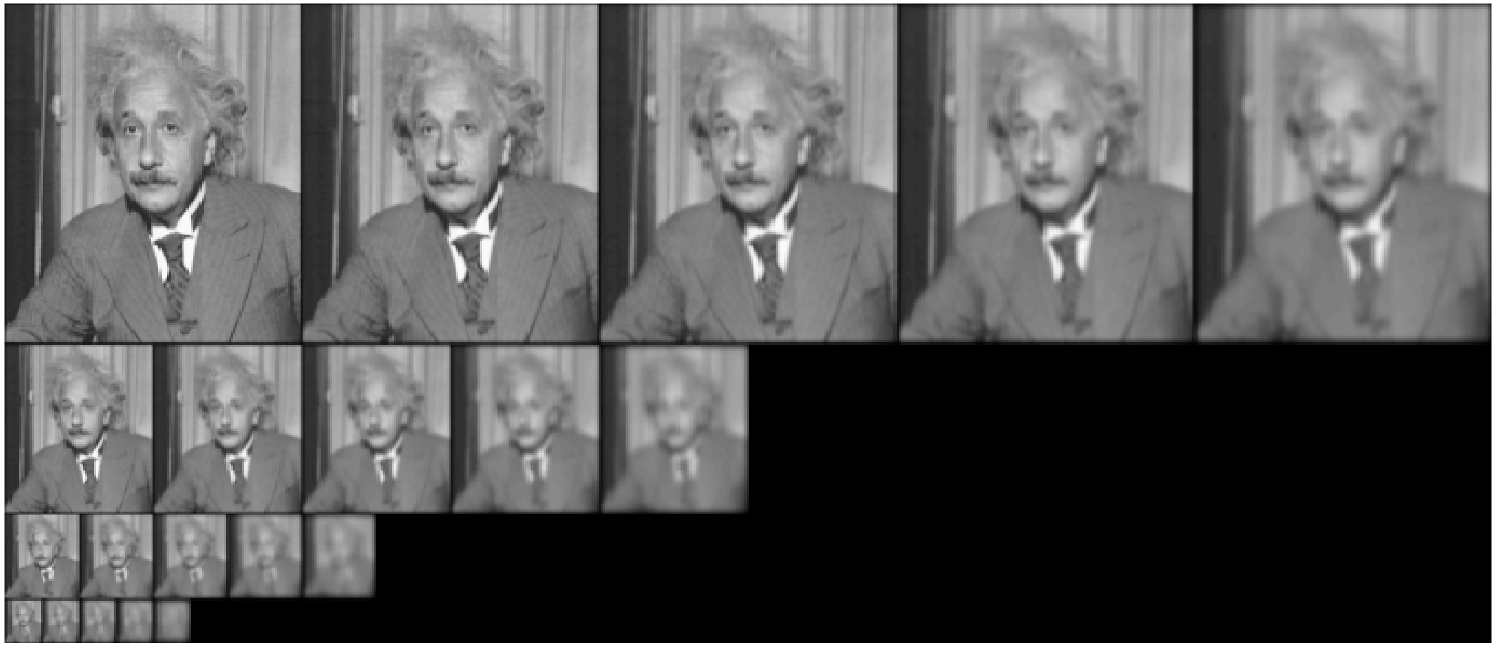
  - All scales must be examined to identify scale-invariant features

  - An efficient function is to compute the Difference of Gaussian (DoG) pyramid (Burt & Adelson, 1983)



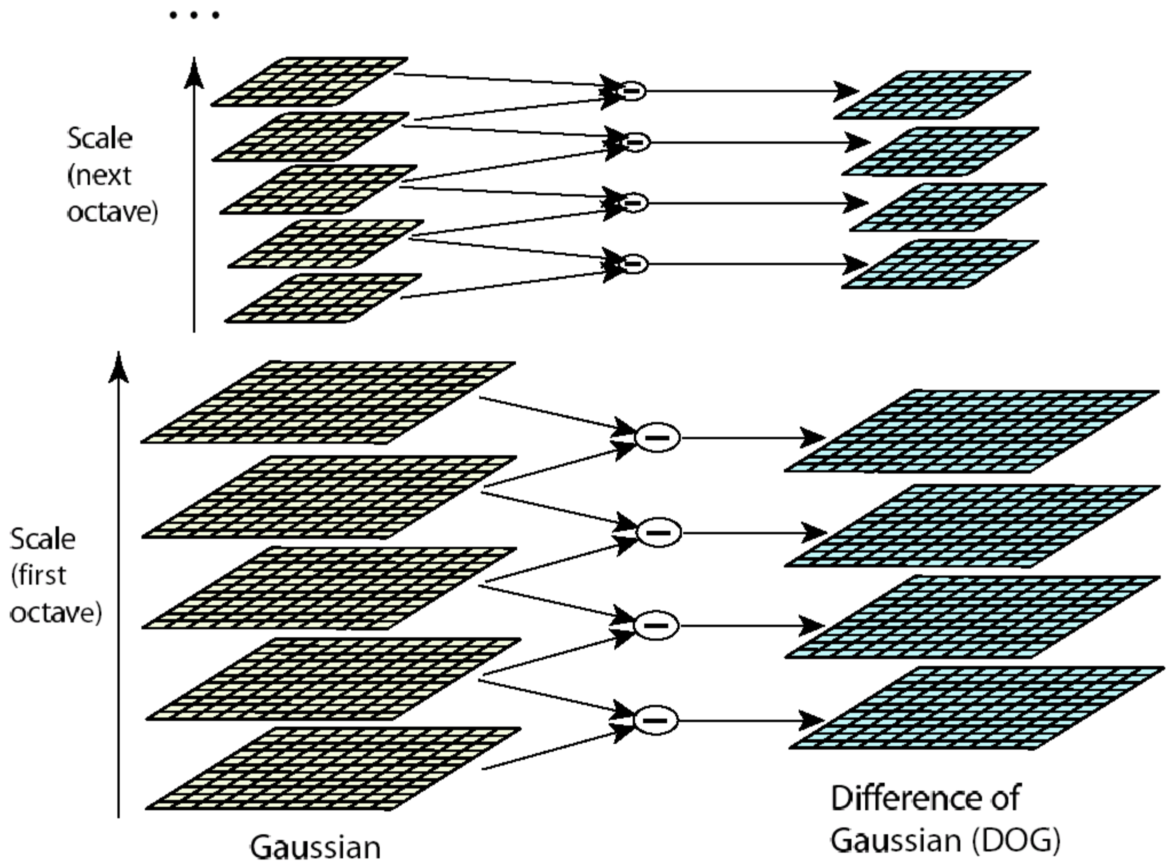
SIFT Features

The original image is convolved with incremental Gaussian to produce images separated by a constant value



SIFT Features

- **Scale space processed one octave at a time**



SIFT Features

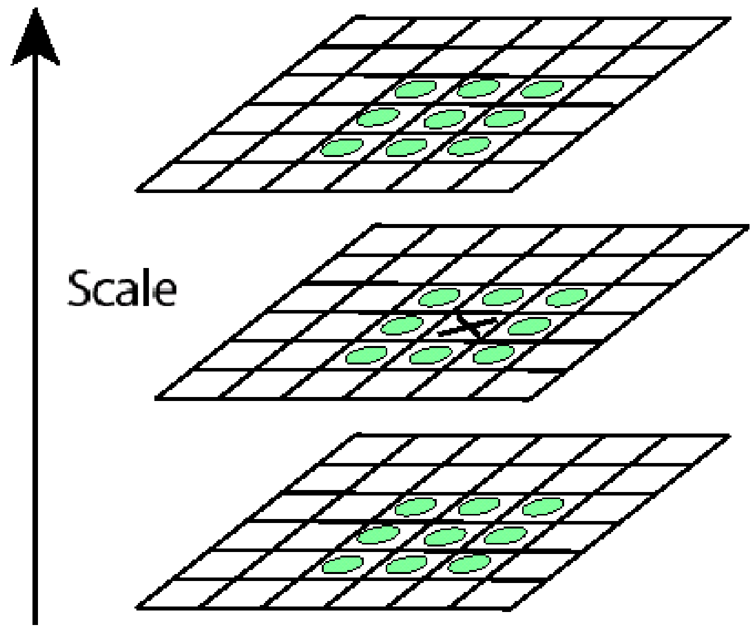
Example of Difference of Gaussian (DoG)

SIFT Features

- **Key point localization**

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

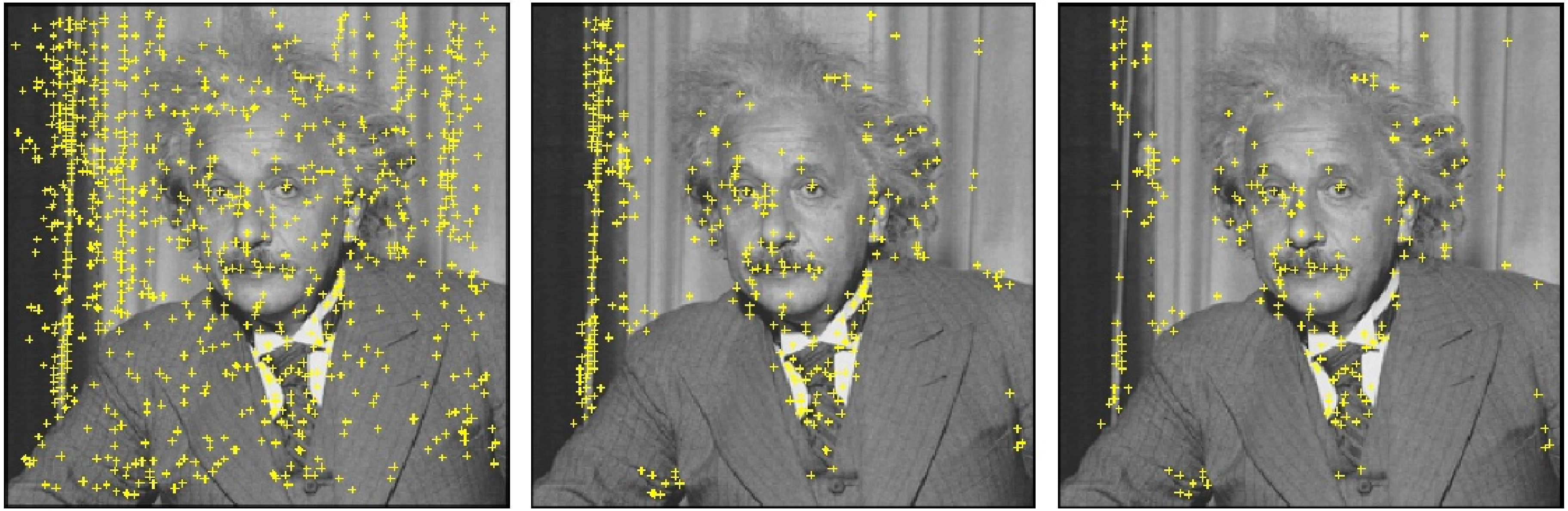
- A point is selected as candidate if it is smaller or greater than its 26 neighbors



Universitat de Girona

SIFT Features

- **Example Key point localization**



Candidate points **Without low contrast** **Without edge points**

SIFT Features

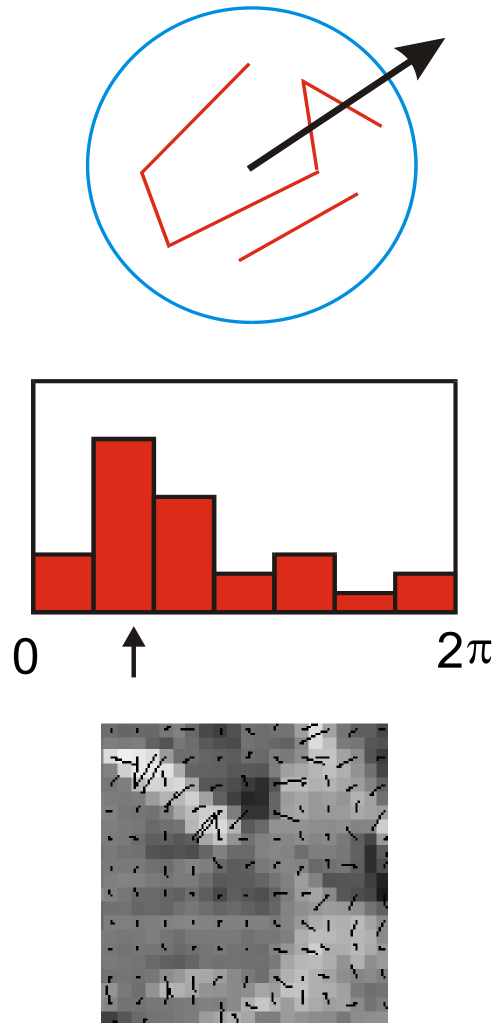
- **Select canonical orientation**

- 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)

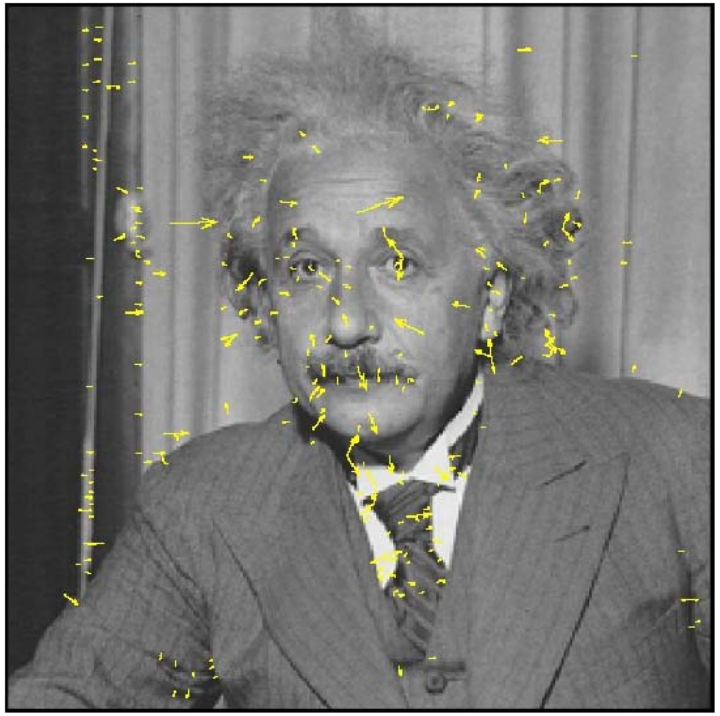
- Create a separate keypoint at any other orientation within 80% of the maximum value



Universitat de Girona

SIFT Features

- **Select canonical orientation**

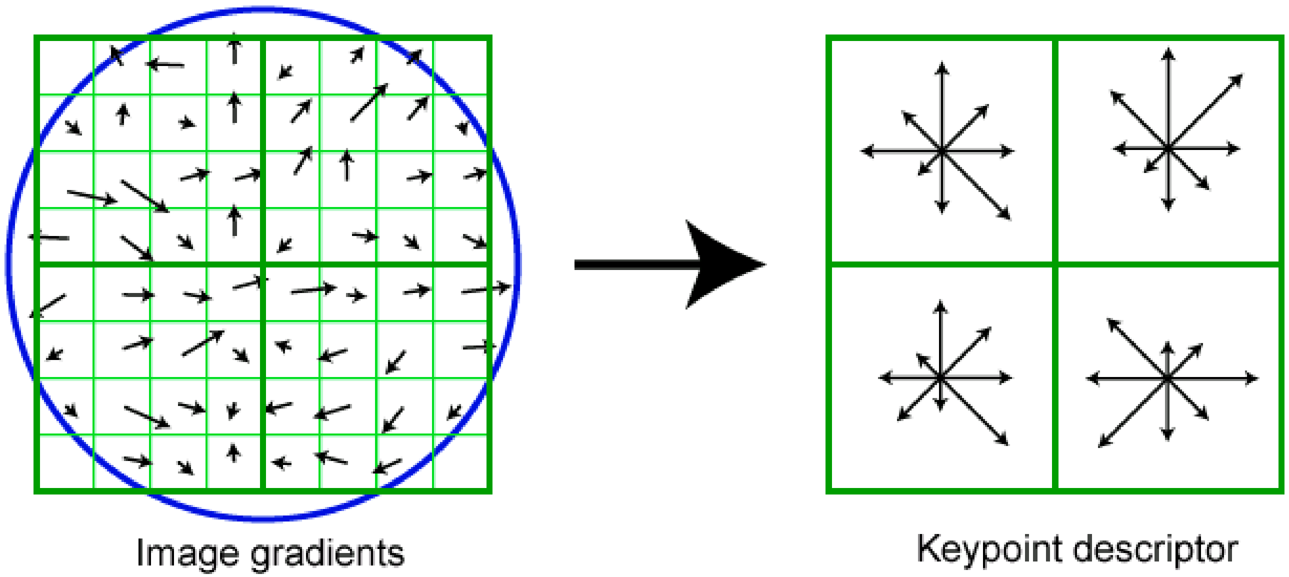


SIFT Features

- **Keypoint description**

  - A histogram of gradients is build

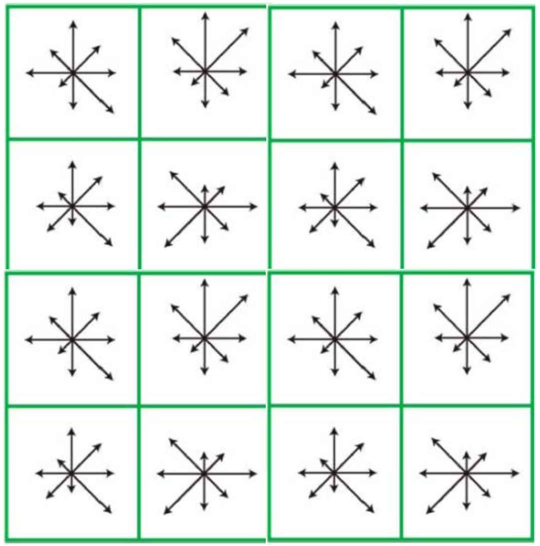
  - Each entry is calculated as the sum of all gradient magnitudes from the corresponding subwindow



SIFT Features

- **Keypoint description**

The image descriptor is a VECTOR that contains the values of the gradient orientation histogram entries.



In the Lowe paper:

- Window of 16x16 pixels.

- Region divided into 4x4 subwindows

- Gradient orientation is discretized to angles of 45º → histogram has 8 entries

- **Descriptor** = 4 x 4 x 8 = 128 elements

SIFT Features. Object Recognition

SIFT KEY Matching

↓

Clustering of matched keys

↓

Fitting a geometric model

↓

OBJECT RECOGNIZED IN IMAGE

- Keypoints are matched in a database using the nearest neighbour algorithm. The database is formed by keypoints of training images.

- Clusters of at least 3 keypoints that agree on an object and its pose are identified. These are interpreted as an object.

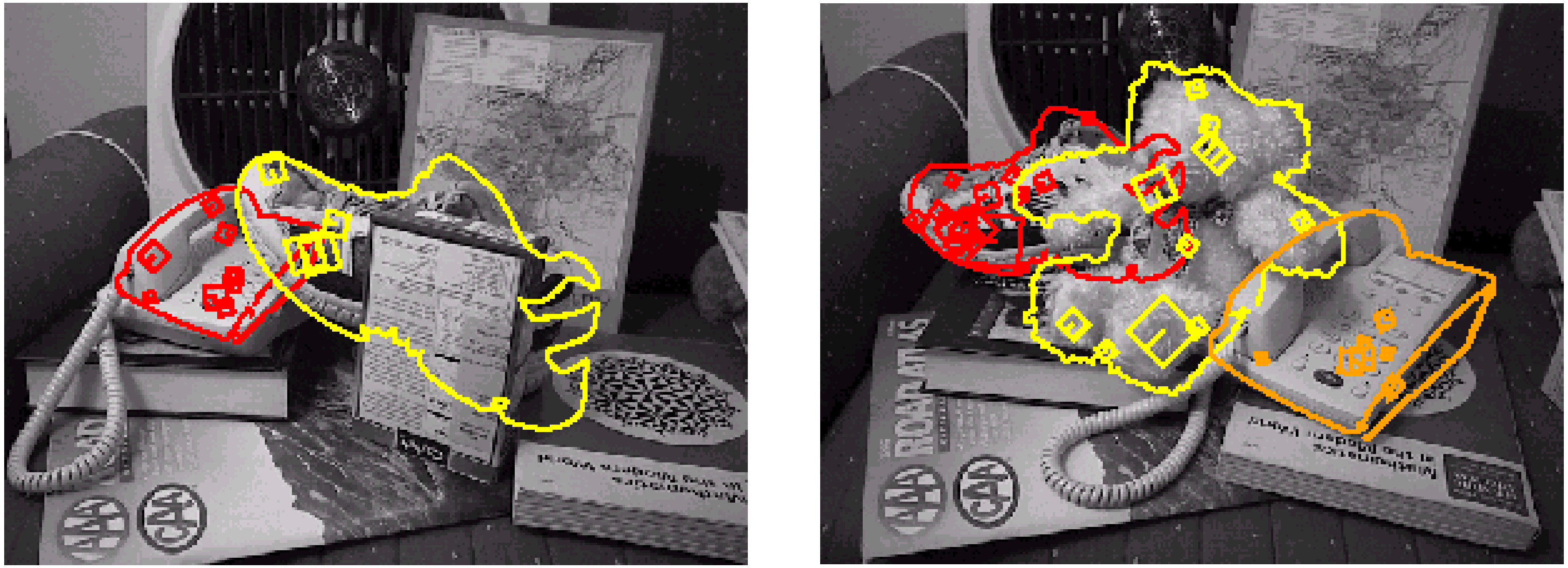
- Each cluster is checked by performing a detailed fit to the model. The quality of the fit is used to accept or reject the interpretation.

Object Recognition. Results

Training images are shown on the left. The 3 bigger squares indicate the recognized objects region. The smaller squares are the matched keypoints. The line inside indicates the keypoint orientation

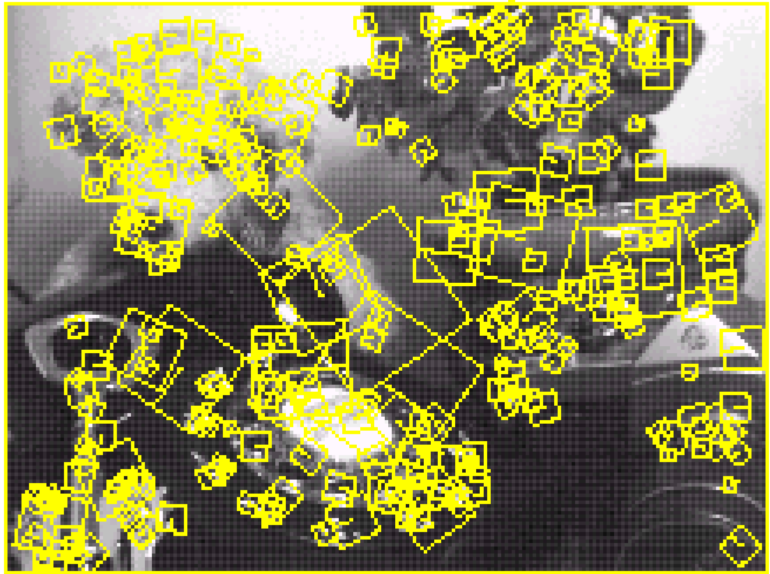
The objects in the images are highly occluded

Object Recognition. Results



Recognition under occlusion

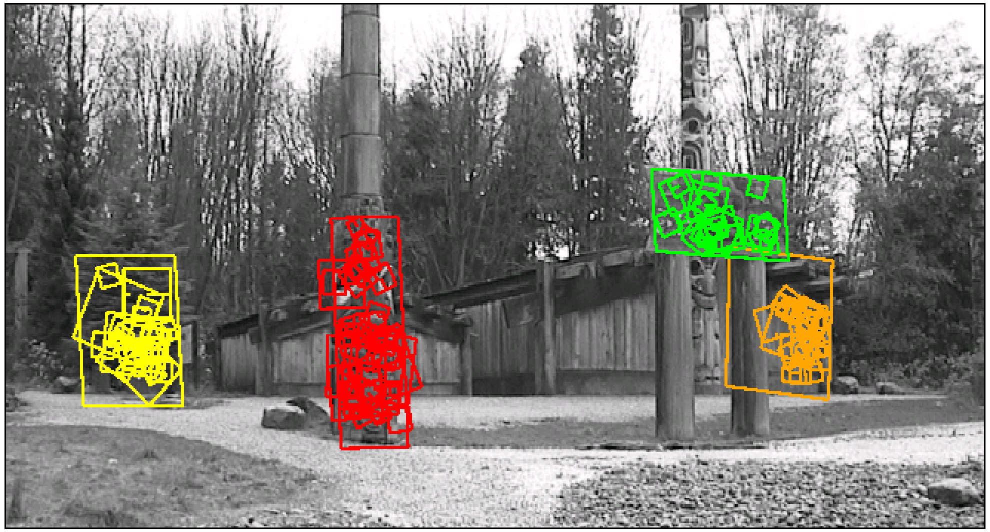
Object Recognition. Results



273 keys verified in final match

Same image under different illumination

Object Recognition. Results



Recognition under rotation and occlusion

SIFT features

- **Conclusions**

- SIFT keypoints are invariant to image rotation, scale, and robust to substantial range of affine distortion, addition of noise, and change in illumination

- We can use the sift descriptors (128 feature vector) to perform object recognition

- All the constants used are computed experimentally with one set of images → this can be a weak point