Mid Level Image Features : Shapes

Eun Yi Kim

Artificial Inteligence
& Computer Vision
Laboratory

I N D E X

Shapes

Approaches to Shape description
- Region based shape descriptors
- Boundary based descriptors
- Interest Operator + Descriptor

Applications

Interest Operator + Descriptor

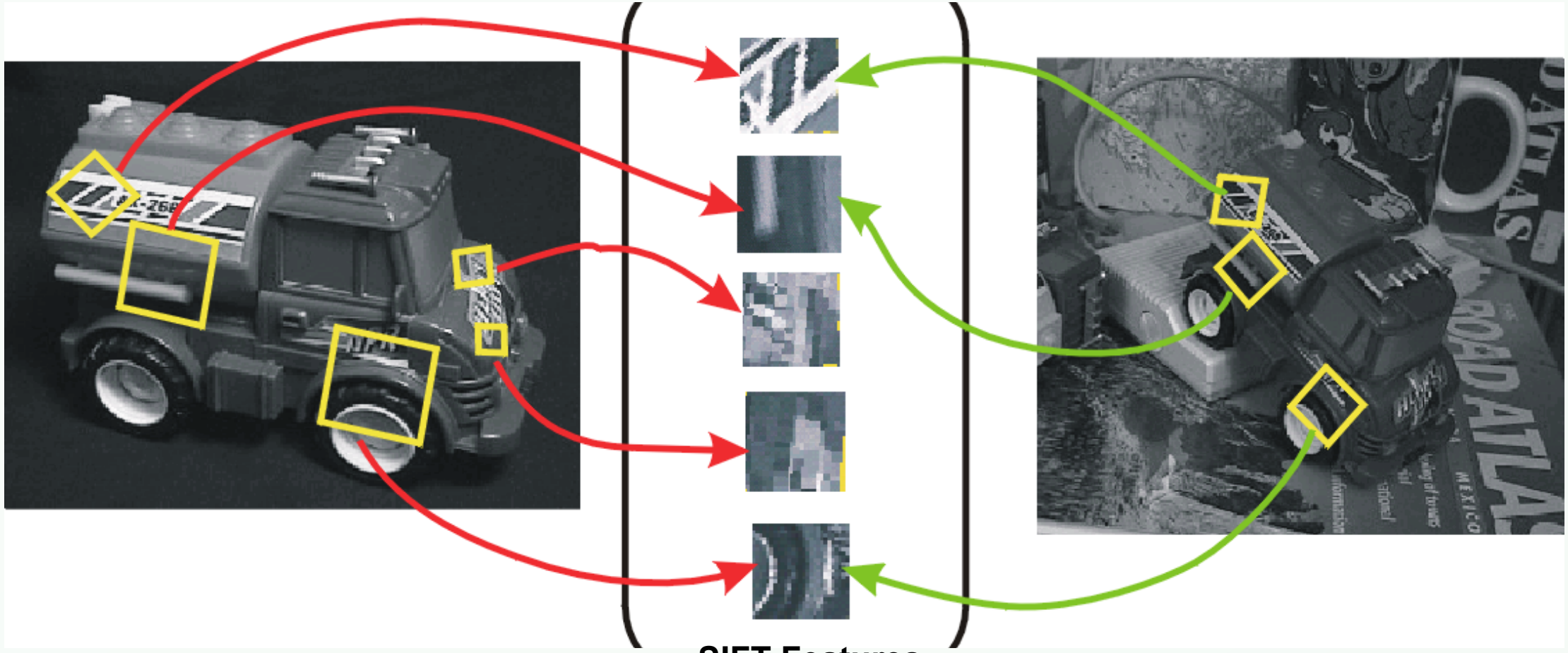
- Harris operator
- Multi-scaled operator
- SIFT (scale invariant feature transform)
- HOG (histogram of oriented gradient)

SIFT: Motivation

- The Harris operator is not invariant to scale and correlation is not invariant to rotation.
- For better image matching, Lowe's goal was to develop an interest operato r that is invariant to scale and rotation.
- Also, Lowe aimed to create a descriptor that was robust to the variations c orresponding to typical viewing conditions. The descriptor is the most-used part of SIFT.

Idea of SIFT

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



Claimed Advantages of SIFT

- **Locality:** features are local, so robust to occlusion and clutter (no prior s egmentation)
- **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 t ypes, with each adding robustness

Overall Procedure at a High Level

1. Scale-space extrema detection
   - Search over multiple scales and image locations.

2. Keypoint localization
   - Fit a model to determine location and scale.
   - Select keypoints based on a measure of stability.

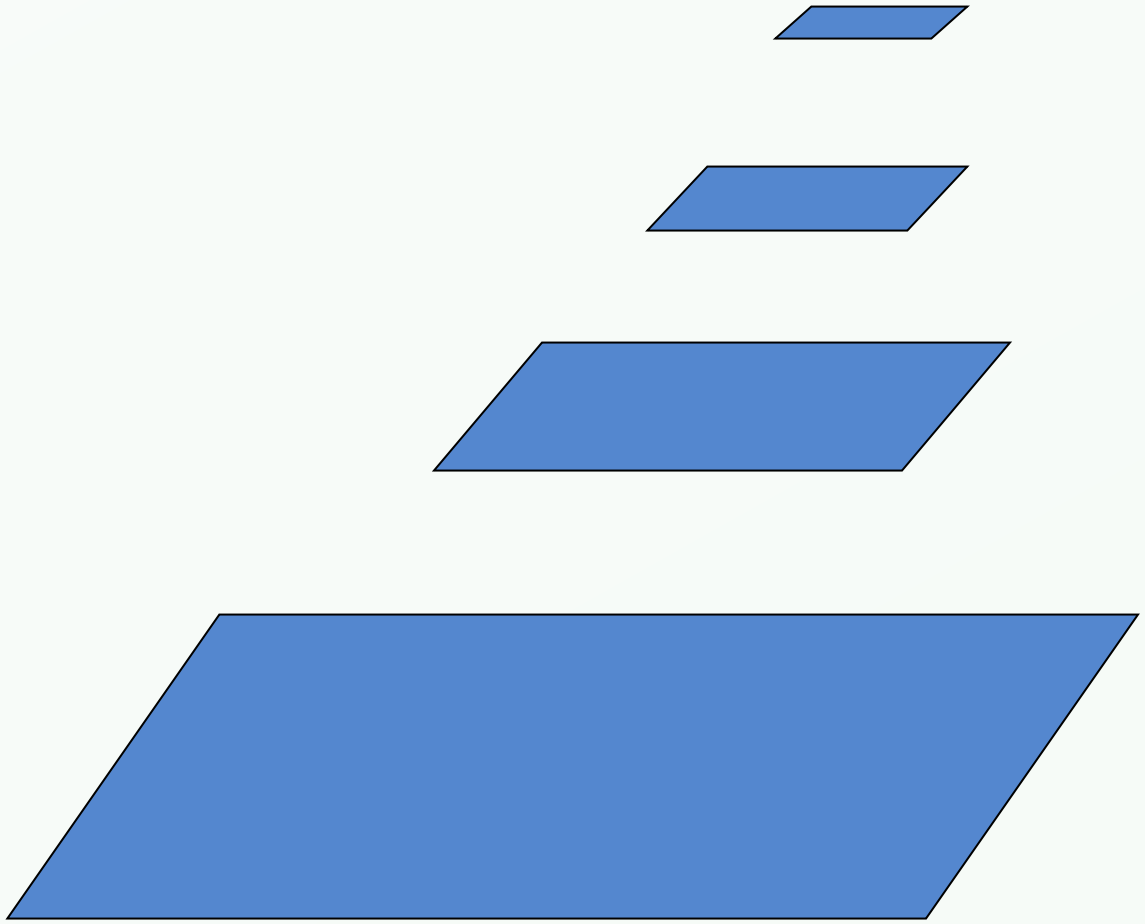
3. Orientation assignment
   - Compute best orientation(s) for each keypoint region.

4. Keypoint description
   - Use local image gradients at selected scale and rotation to describe each keypoint region.

1. Scale-space extrema detection

- Goal: Identify locations and scales that can be repeatedly assigned under diff erent views of the same scene or object.
- Method: search for stable features across multiple scales using a continuous f unction of scale.
- Prior work has shown that under a variety of assumptions, the best function is a Gaussian function.
- The scale space of an image is a function $L(x,y,\sigma)$ that is produced from the c onvolution of a Gaussian kernel (at different scales) with the input image.

Aside: Image Pyramids



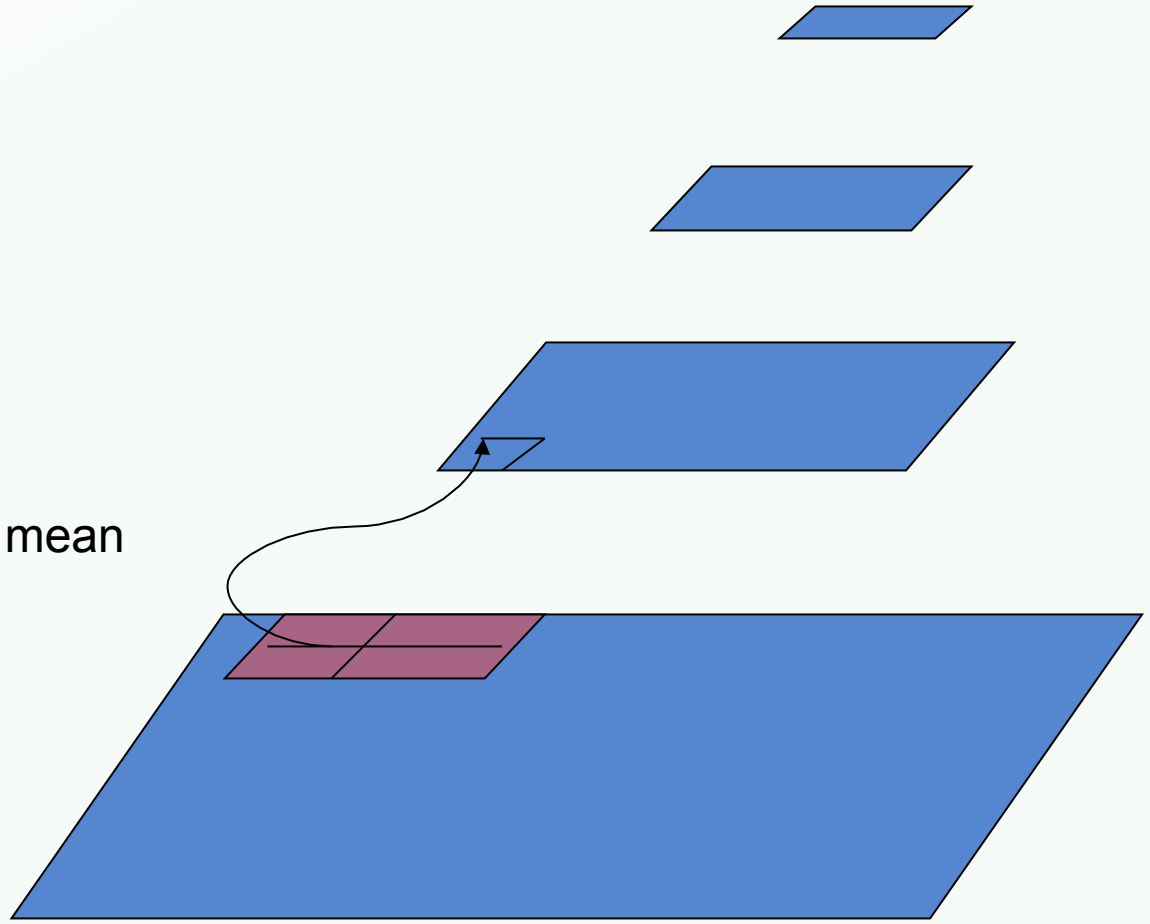
And so on.

3rd level is derived from the 2nd level according to the same funtion

2nd level is derived from the original image according to some function

Bottom level is the original image.

Aside: Mean Pyramid



And so on.

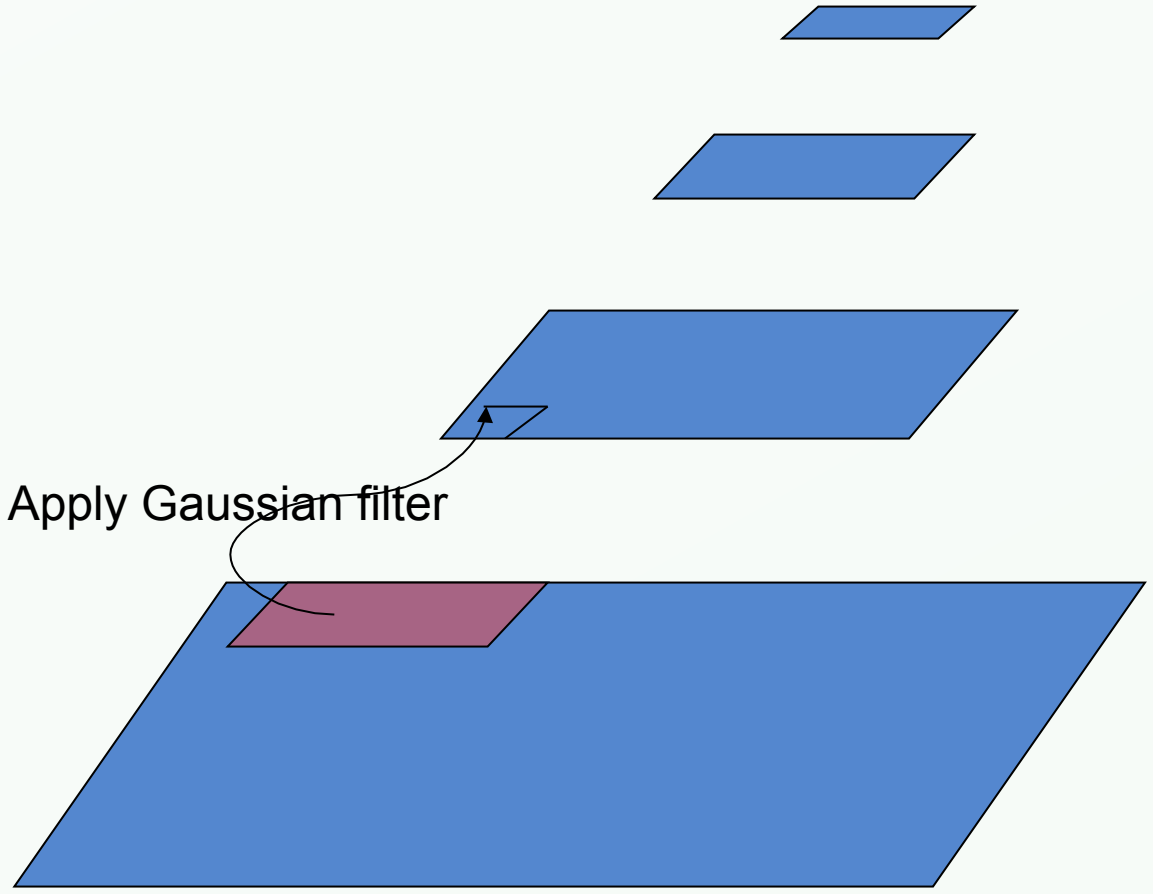
At 3rd level, each pixel is the mean of 4 pixels in the 2nd level.

At 2nd level, each pixel is the mean of 4 pixels in the original image.

Bottom level is the original image.

Aside: Gaussian Pyramid

At each level, image is smoothed and reduced in size.

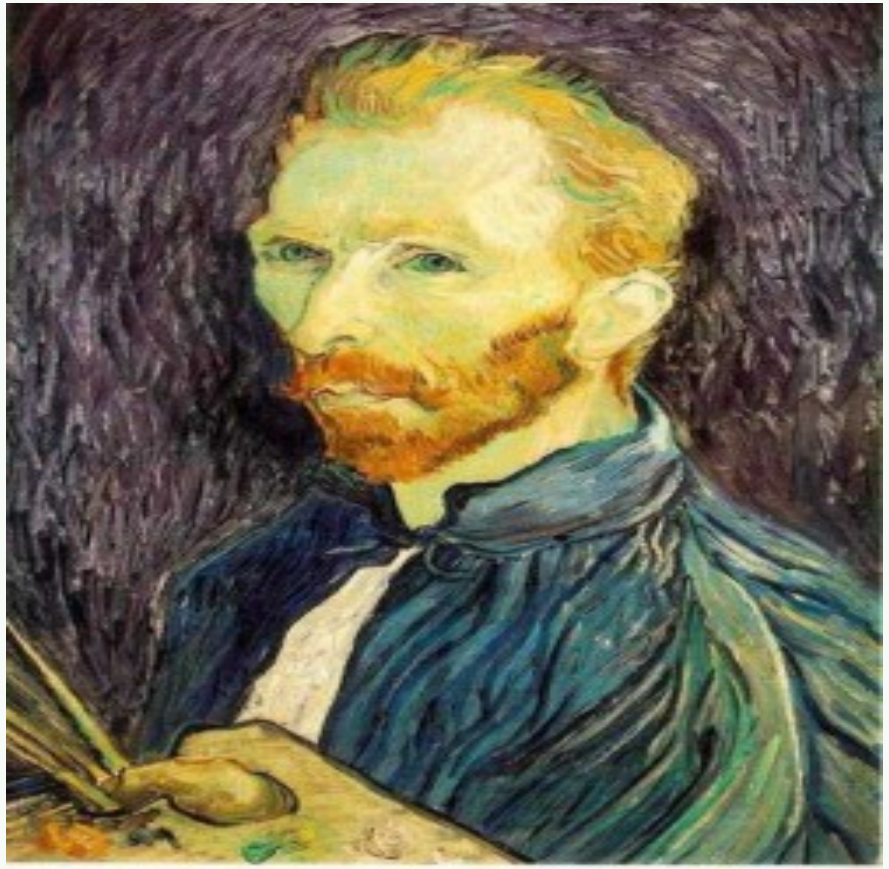


And so on.

At 2nd level, each pixel is the result of applying a Gaussian mask to the first level and then subsampling to reduce the size.

Bottom level is the original image.

Example:
Subsampling with Gaussian pre-filtering



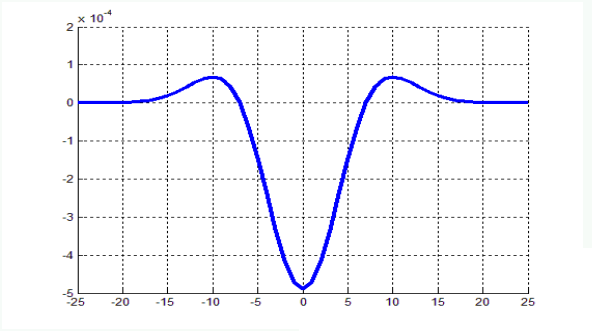
Gaussian 1/2

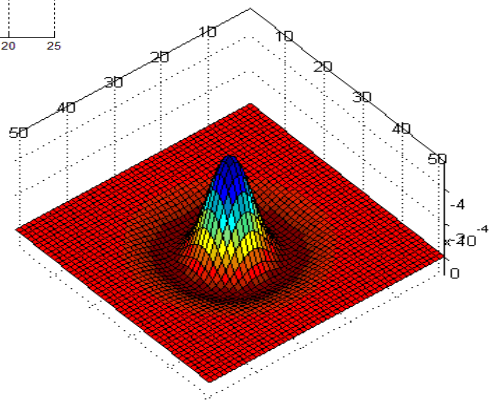
G 1/4

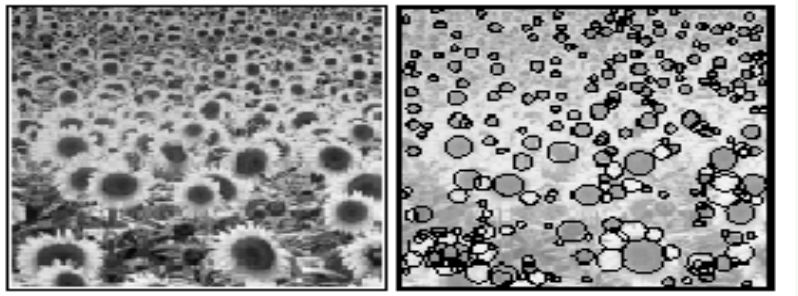
G 1/8

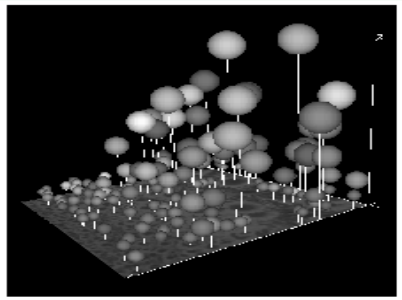
Lowe's Scale-space Interest Points

- **Laplacian of Gaussian** kernel
  - Scale normalised (x by scale^2)
  - Proposed by Lindeberg
- Scale-space detection
  - Find local maxima across scale/space
  - A good "blob" detector







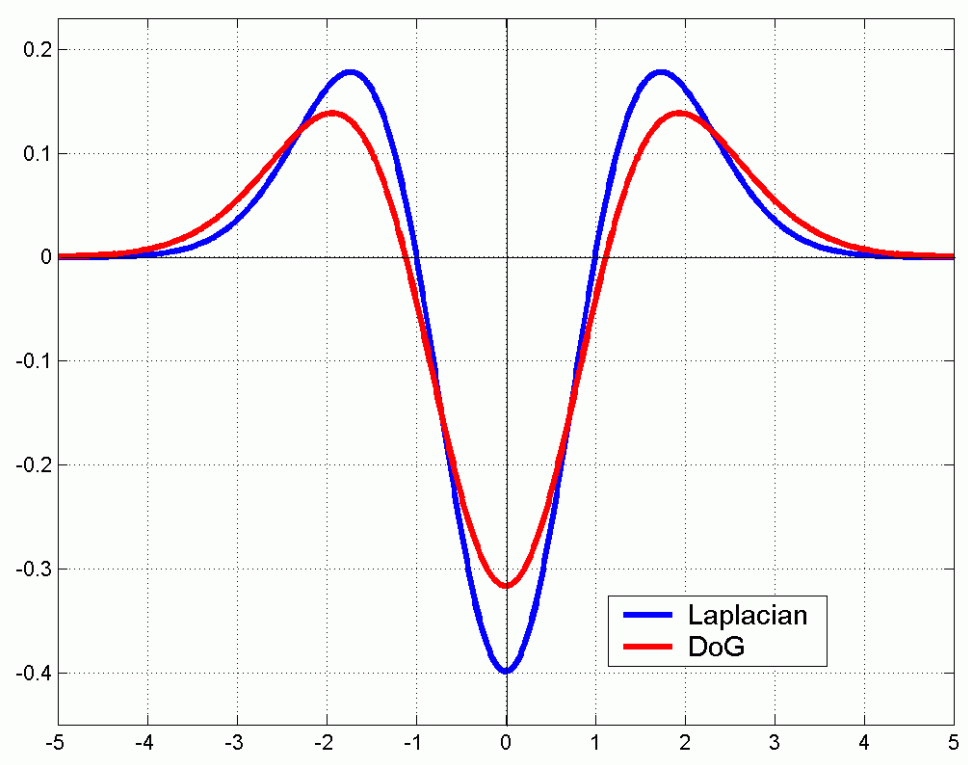


$$G(x, y, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{x^2+y^2}{\sigma^2}}$$

$$\nabla^2 G(x, y, \sigma) = \frac{\partial^2 G}{\partial x^2} + \frac{\partial^2 G}{\partial y^2}$$

[T. Lindeberg IJCV 1998]

Lowe's Scale-space Interest Points: Difference of Gaussians



- Gaussian is an ad hoc solution of heat diffusion equation

$$\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G.$$

- Hence

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k-1)\sigma^2 \nabla^2 G.$$

- k is not necessarily very small i n practice

Lowe's Pyramid Scheme

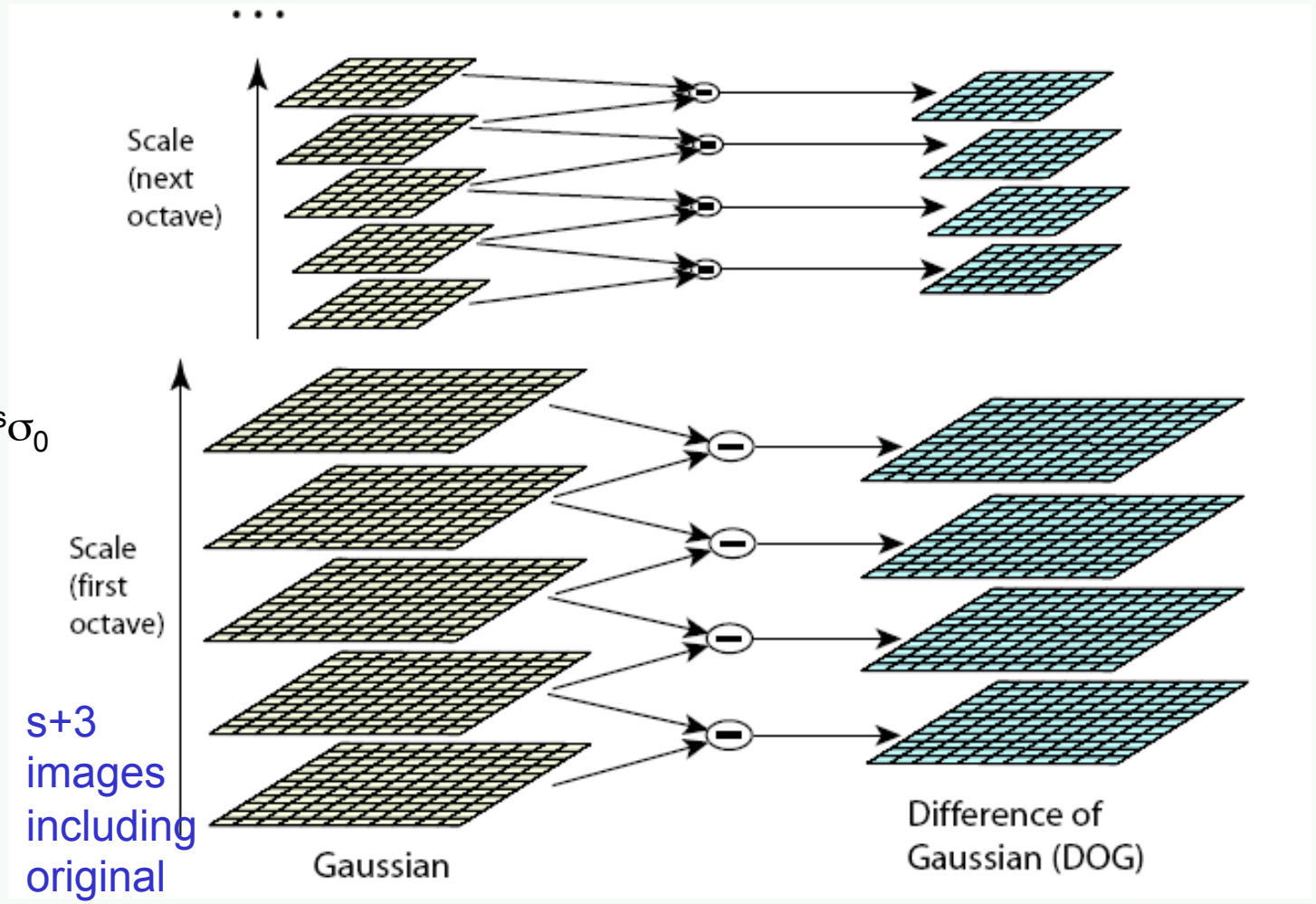
- Scale space is separated into octaves:
  - Octave 1 uses scale σ
  - Octave 2 uses scale 2σ
  - etc.

- In each octave, the initial image is repeatedly convolved with Gaussian s to produce a set of scale space images.

- Adjacent Gaussians are subtracted to produce the DOG

- After each octave, the Gaussian image is down-sampled by a factor of 2 to produce an image ¼ the size to start the next level.

Lowe's Pyramid Scheme



s+2 filters

$\sigma_{s+1}=2^{(s+1)/s}\sigma_0$

.

.

$\sigma_i=2^{i/s}\sigma_0$

.

.

$\sigma_2=2^{2/s}\sigma_0$

$\sigma_1=2^{1/s}\sigma_0$

σ_0

s+3 images including original

s+2 difference images

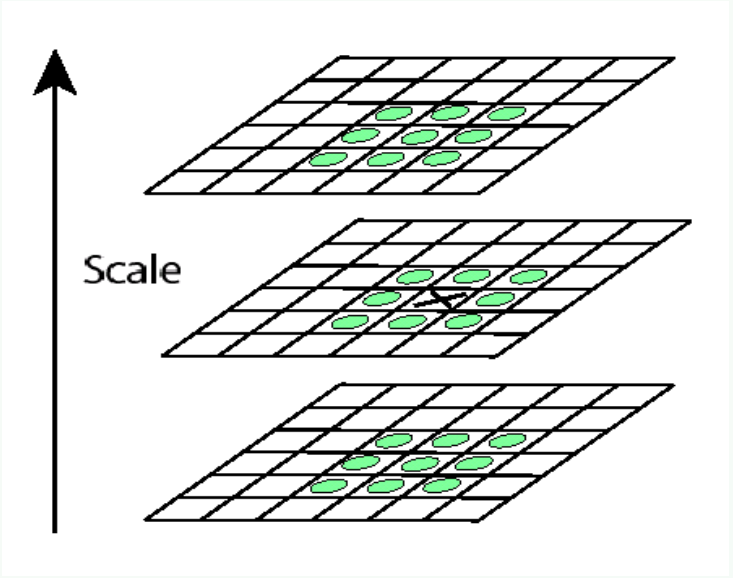
The parameter **s** determines the number of images per octave.

Key point localization

s+2 difference images.
top and bottom ignored.
s planes searched.

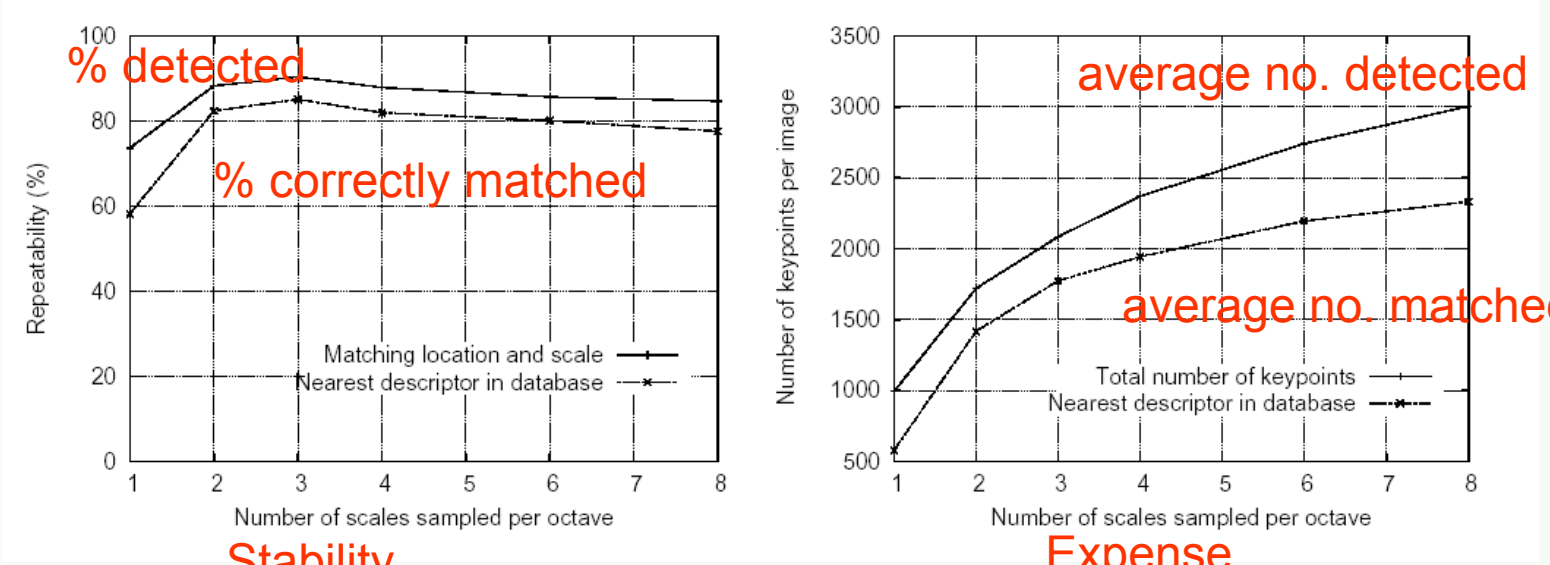
- Detect maxima and minima of difference-of-Gaussian in scale space

- Each point is compared to its 8 neighbors in the current image and 9 neighbors each in the scales above and below



For each max or min found, output is the **location** and the **scale**.

Key point localization



- Sampling in scale for efficiency
  - How many scales should be used per octave? S=?
    - More scales evaluated, more keypoints found
    - S < 3, stable keypoints increased too
    - S > 3, stable keypoints decreased
    - S = 3, maximum stable keypoints found

2. Keypoint localization

- Once a keypoint candidate is found, perform a detailed fit to nearby data to determine
  - location, scale, and ratio of principal curvatures
- In initial work keypoints were found at location and scale of a central sampl e point.
- In newer work, they fit a 3D quadratic function to improve interpolation acc uracy.
- The Hessian matrix was used to eliminate edge responses.

Eliminating the Edge Response

- Reject flats:

$|D(\hat{\mathbf{x}})|$ < 0.03

- Reject edges:

$$\mathbf{H} = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

Let α be the eigenvalue with larger magnitude and β the smaller.

$$\mathrm{Tr}(\mathbf{H}) = D_{xx} + D_{yy} = \alpha + \beta,$$
$$\mathrm{Det}(\mathbf{H}) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta.$$

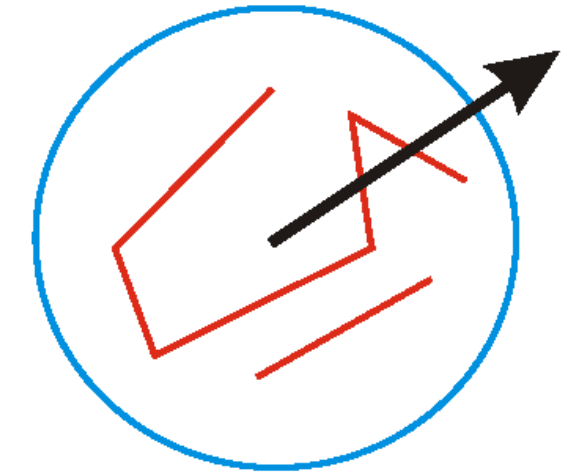
Let $r = \alpha/\beta$.
So $\alpha = r\beta$

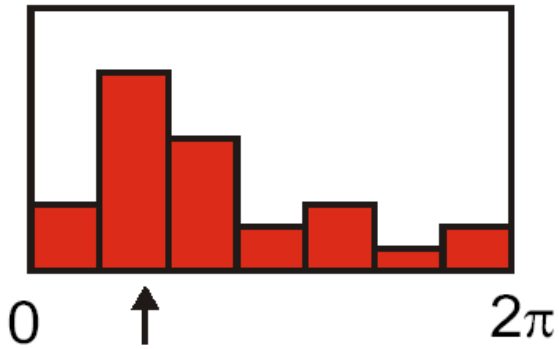
$$\frac{\mathrm{Tr}(\mathbf{H})^2}{\mathrm{Det}(\mathbf{H})} = \frac{(\alpha+\beta)^2}{\alpha\beta} = \frac{(r\beta+\beta)^2}{r\beta^2} = \frac{(r+1)^2}{r},$$

$(r+1)^2/r$ is at a min when the 2 eigenvalues are equal.

- $r < 10$

3. Orientation assignment



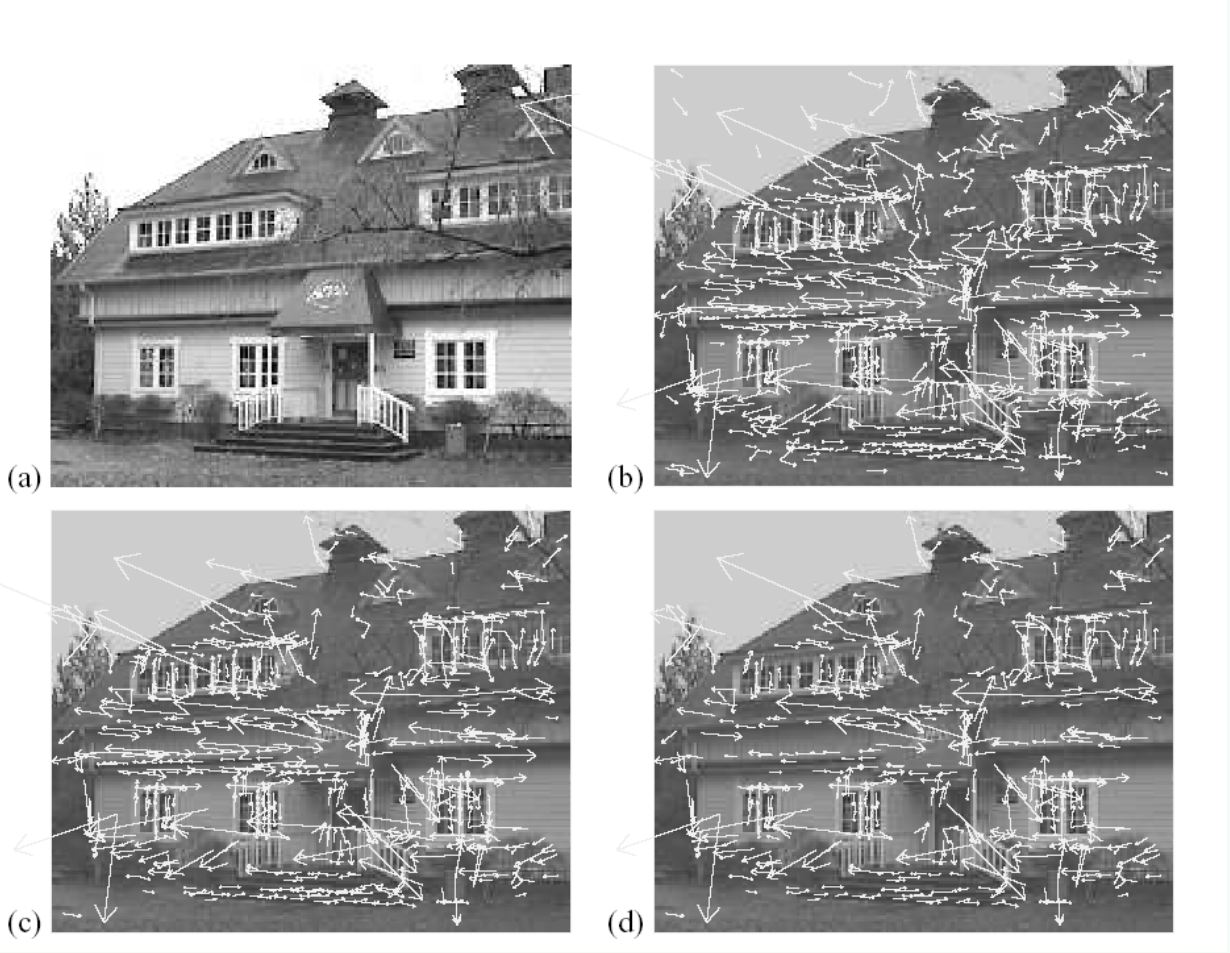


- Create histogram of local gradient directions at selected scale
- Assign canonical orientation at peak of smo othed histogram
- Each key specifies stable 2D coordinates (x, y, scale, orientation)

If 2 major orientations, use both.

Keypoint localization with orientation

233x189



832

initial keypoints

729

keypoints after
gradient threshold

536

keypoints after
ratio threshold

4. Keypoint Descriptors

- At this point, each keypoint has
  - location
  - scale
  - Orientation

- Next is to compute a descriptor for the local image region about each keyp oint that is
  - highly distinctive
  - invariant as possible to variations such as changes in viewpoint and illumination

Normalization

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

Lowe's Keypoint Descriptor (shown with 2 X 2 descriptors over 8 X 8)

gradient magnitude and orientation at each point weighted by a Gaussian

orientation histograms: sum of gradient magnitude at each direction

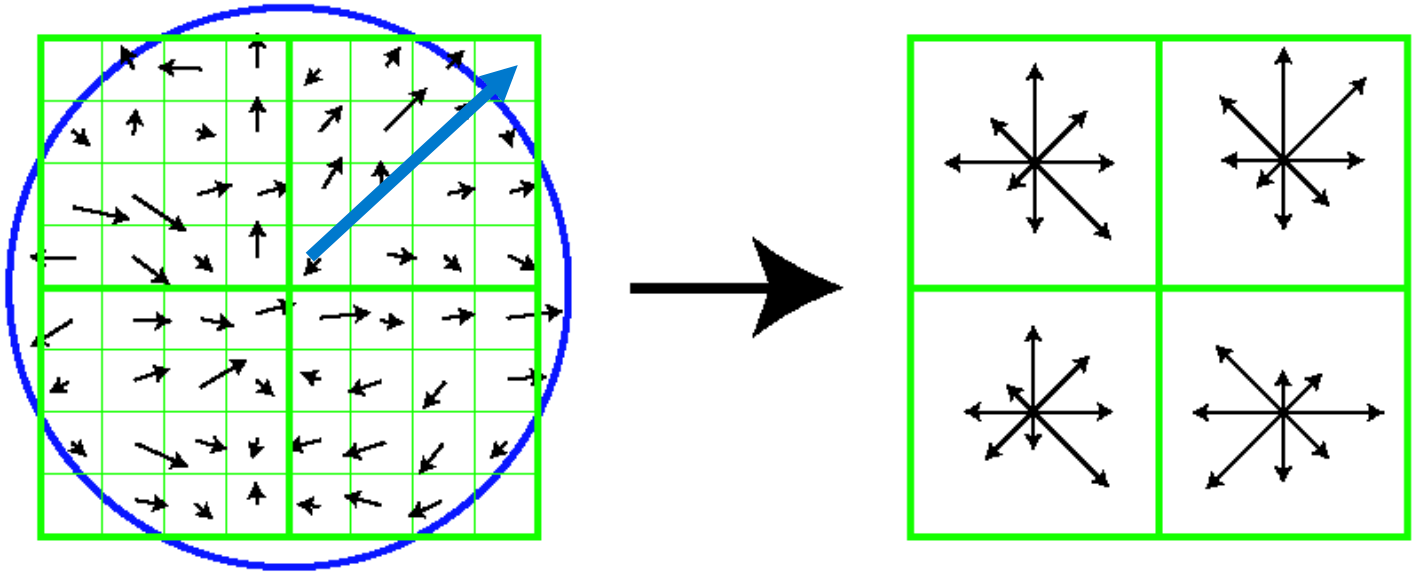


Image gradients

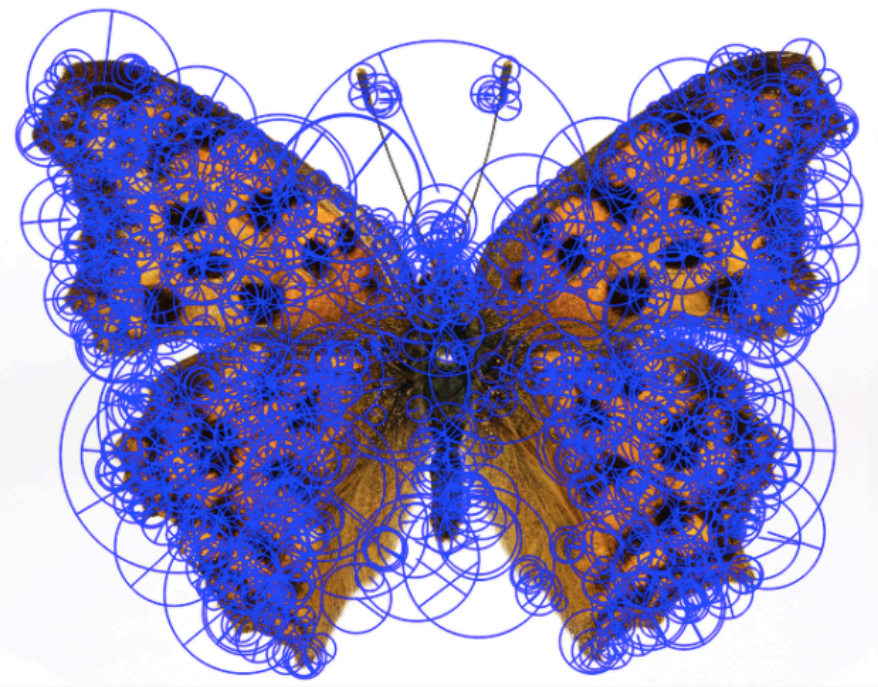
Keypoint descriptor

In experiments, 4x4 arrays of 8 bin histogram is used, a total of 32 features for one keypoint

Lowe's Keypoint Descriptor

- use the normalized region about the keypoint
- compute gradient magnitude and orientation at each point in the region
- weight them by a Gaussian window overlaid on the circle
- create an orientation histogram over the 4 X 4 subregions of the window
- 4 X 4 descriptors over 16 X 16 sample array were used in practice. 4 X 4 times 8 directions gives a vector of **128 values**.

Image Representation with SIFTs

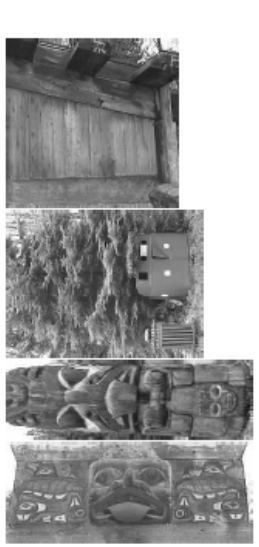


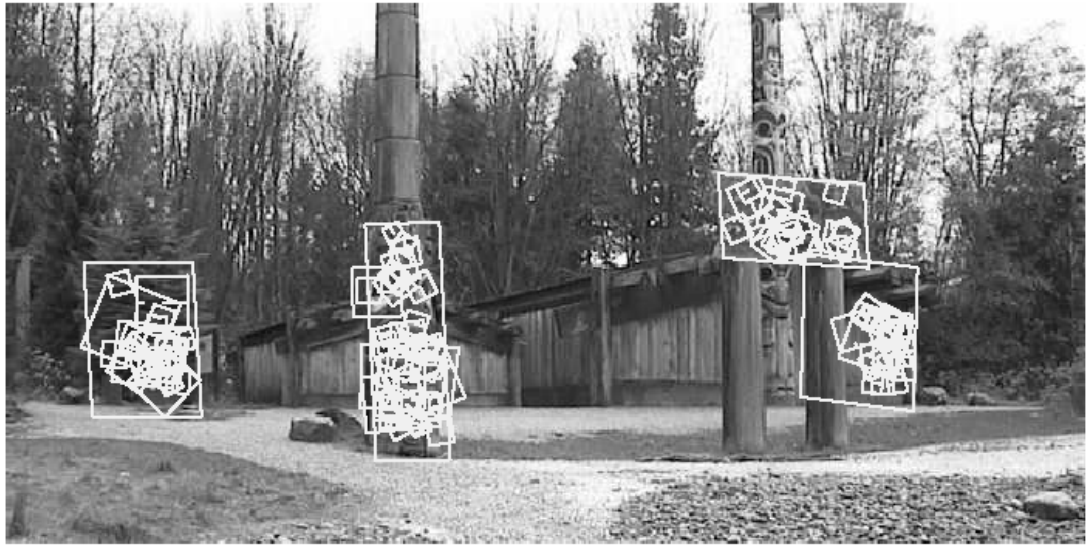
gradient threshold
=1000

gradient threshold
=10000

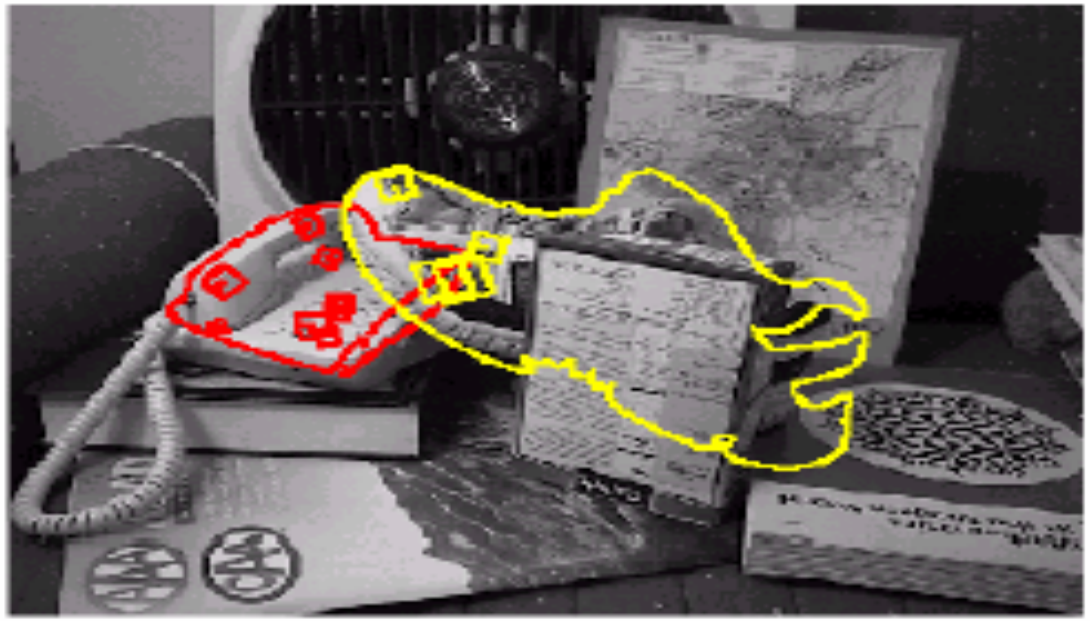
gradient threshold
=15000

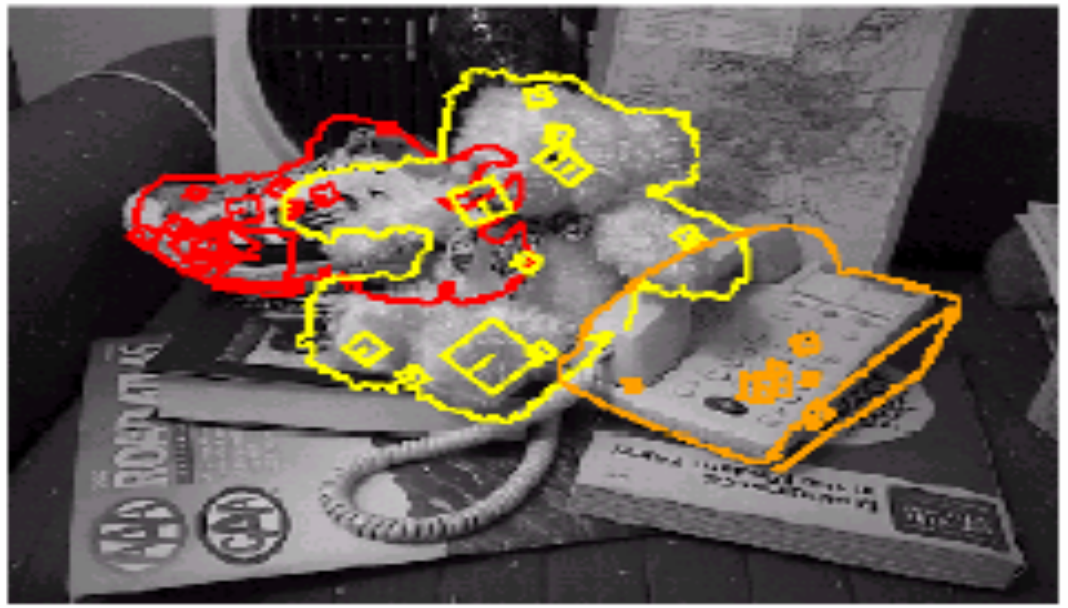
Using SIFT for Matching “Objects”





Using SIFT for Matching “Objects”



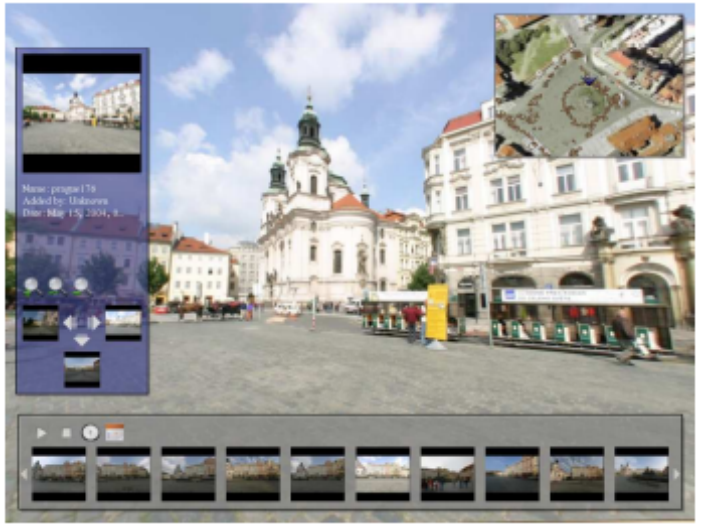


Uses for SIFT

- Feature points are used also for:
  - Image alignment (homography, fundamental matrix)
  - 3D reconstruction (e.g. Photo Tourism)
  - Motion tracking
  - Object recognition
  - Indexing and database retrieval
  - Robot navigation
  - … many others

- Speeding the SIFTs
  - SURF, FAST, BRIEF…

Uses for SIFT



[Photo Tourism:
Snavely et al. SIGGRAPH 2006]

Interest Operator + Descriptor

- Harris operator
- Multi-scaled operator
- SIFT (scale invariant feature transform)
- HOG (histogram of oriented gradient)

Overview

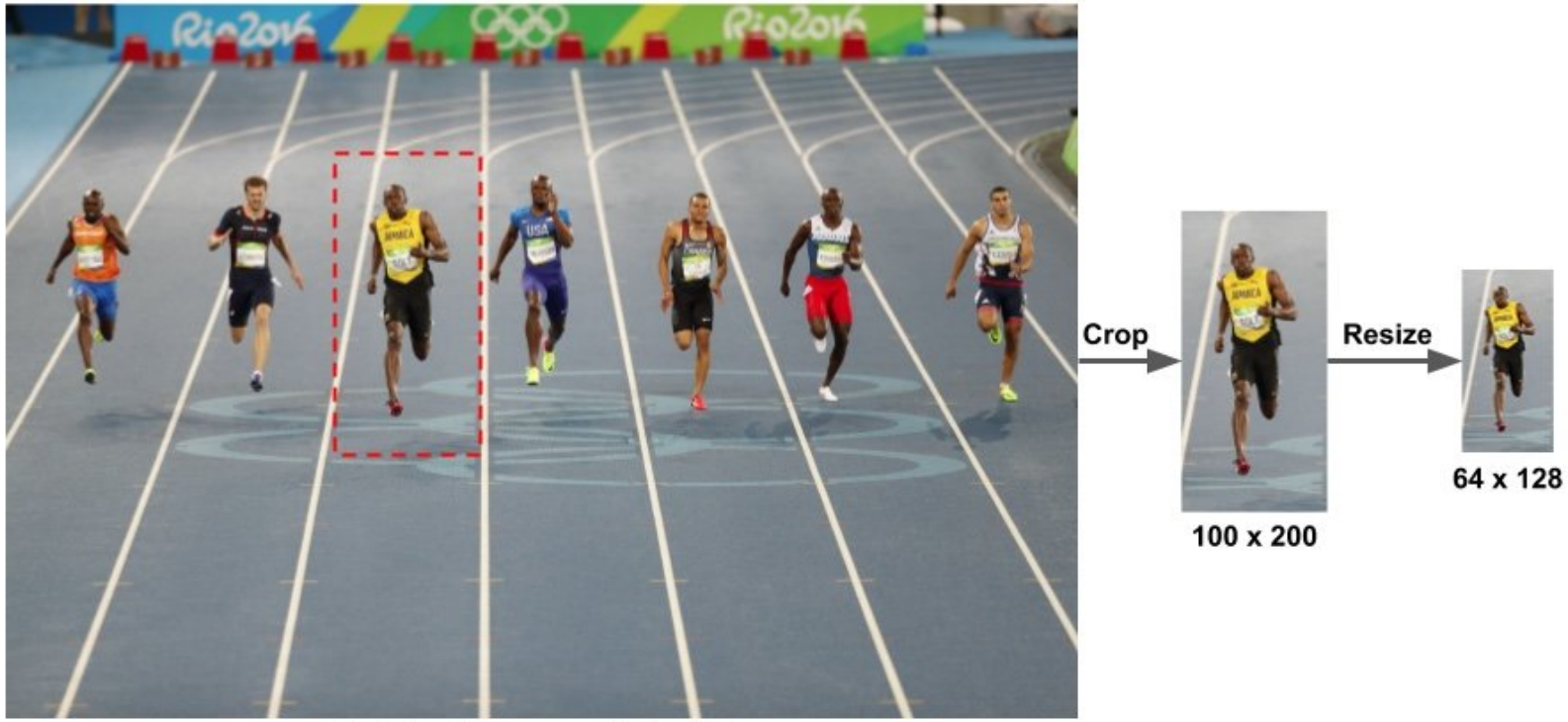
1. Compute gradients in the region to be described
2. Put them in bins according to orientation
3. Group the cells into large blocks
4. Normalize each block
5. Train classifiers to decide if these are parts of a human

R-HOG compared to SIFT Descriptor

- R-HOG blocks appear quite similar to the SIFT descriptors.
- But, R-HOG blocks are computed in dense grids at some single scale without orientation alignment.
- SIFT descriptors are computed at sparse, scale-invariant key image p oints and are rotated to align orientation.

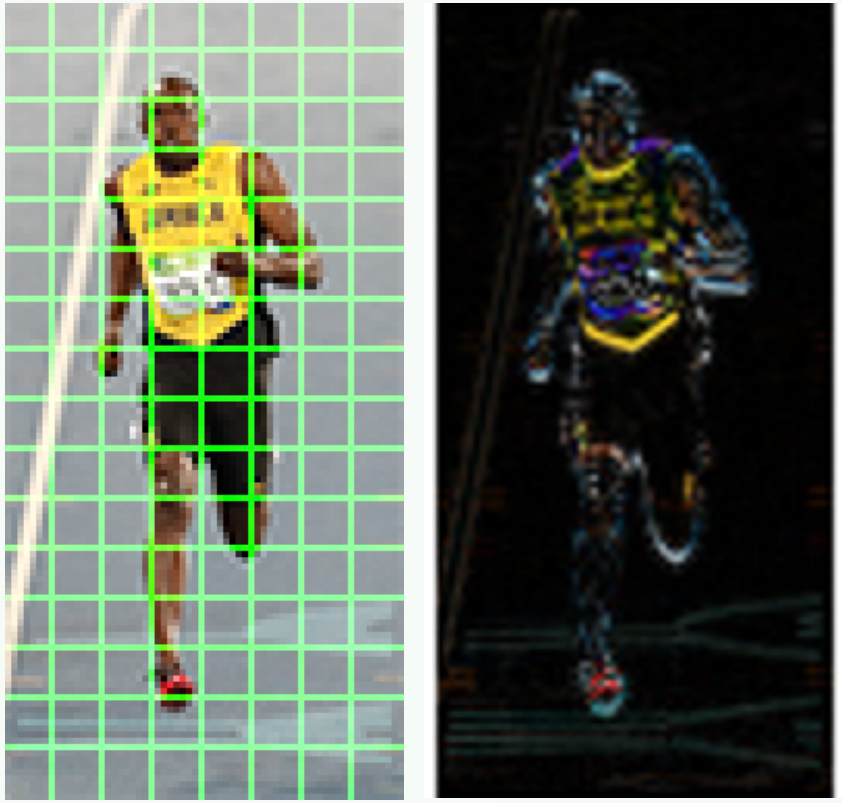
0. Preprocessing

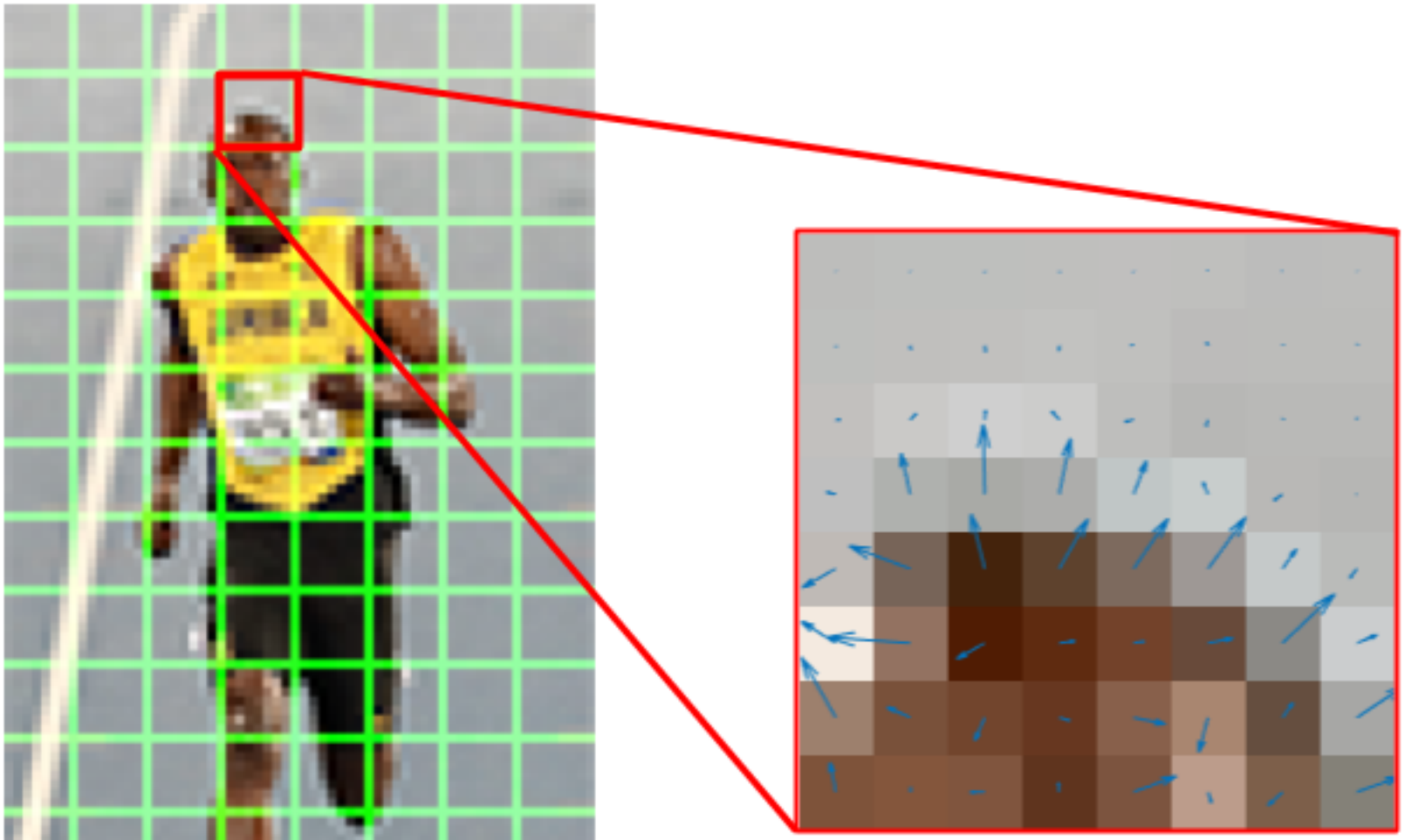
- Crop the patch out of an image
- Resize the patch (e.g. 64×128)



1. Compute Gradients

- To calculate a HOG descriptor, calculate the horizont
  al and vertical gradients. (e.g. sobel, prewitt, etc)
- Find the magnitude and direction of gradients
- Divide the image patch as multiple cells. (e.g. 8x8)





2	3	4	4	3	4	2	2
5	11	17	13	7	9	3	4
11	21	23	27	22	17	4	6
23	99	165	135	85	32	26	2
91	155	133	136	144	152	57	28
98	196	76	38	26	60	170	51
165	60	60	27	77	85	43	136
71	13	34	23	108	27	48	110

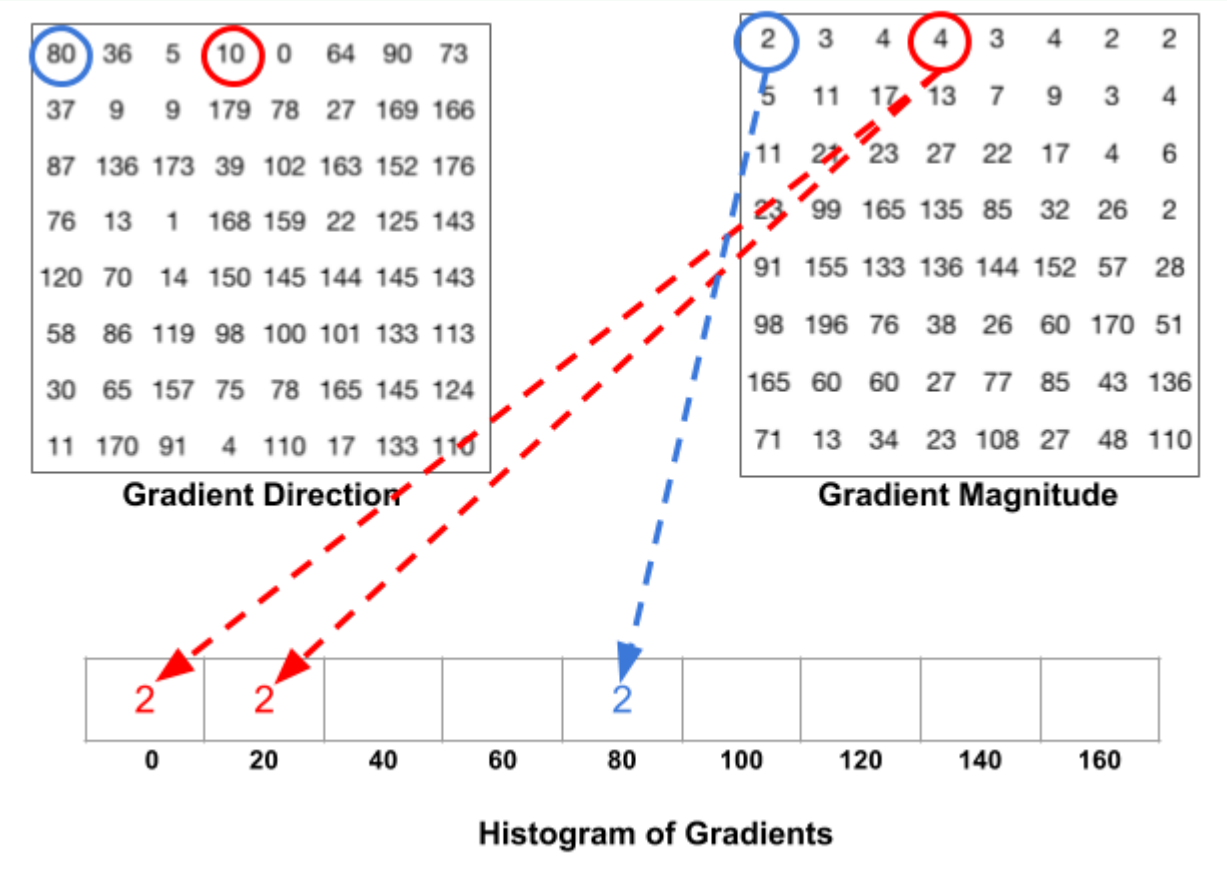
Gradient Magnitude

80	36	5	10	0	64	90	73
37	9	9	179	78	27	169	166
87	136	173	39	102	163	152	176
76	13	1	168	159	22	125	143
120	70	14	150	145	144	145	143
58	86	119	98	100	101	133	113
30	65	157	75	78	165	145	124
11	170	91	4	110	17	133	110

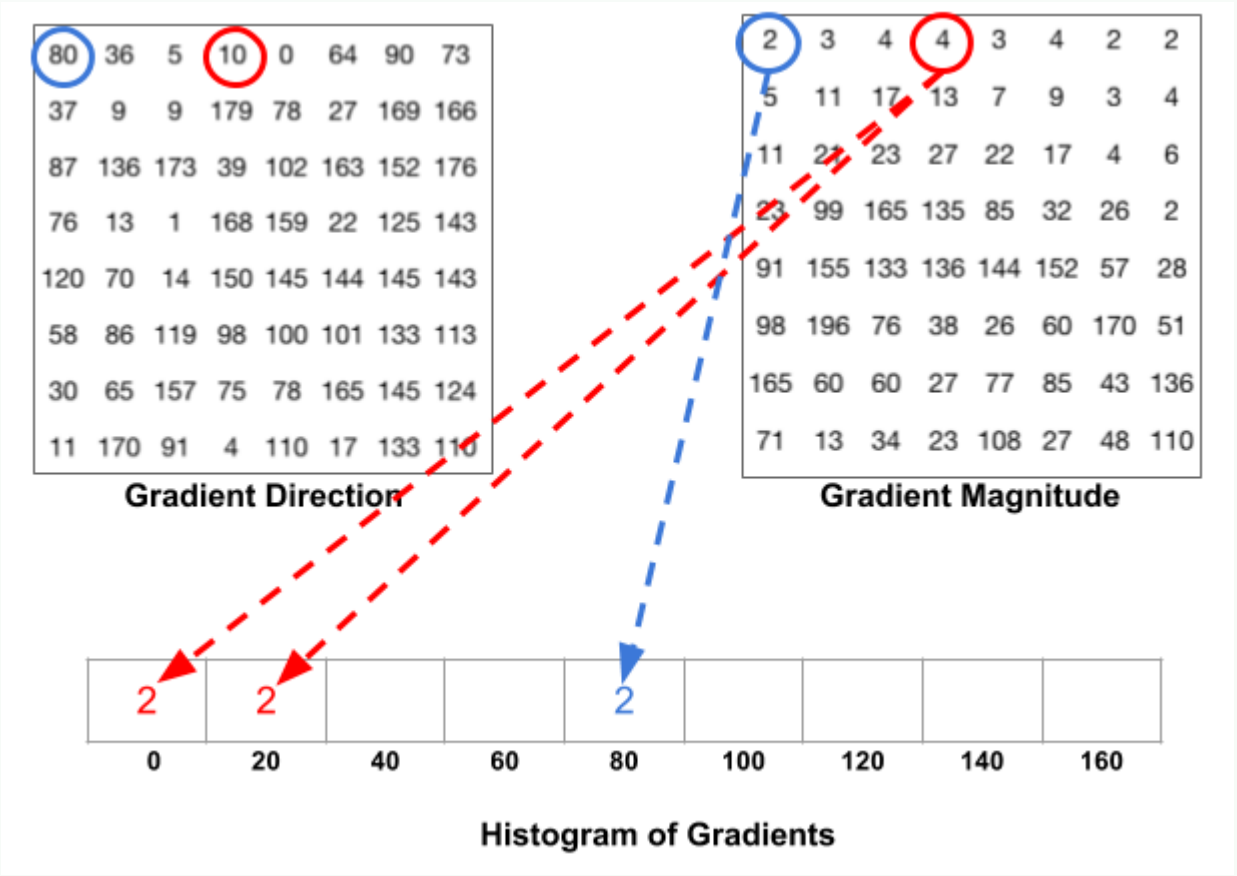
Gradient Direction

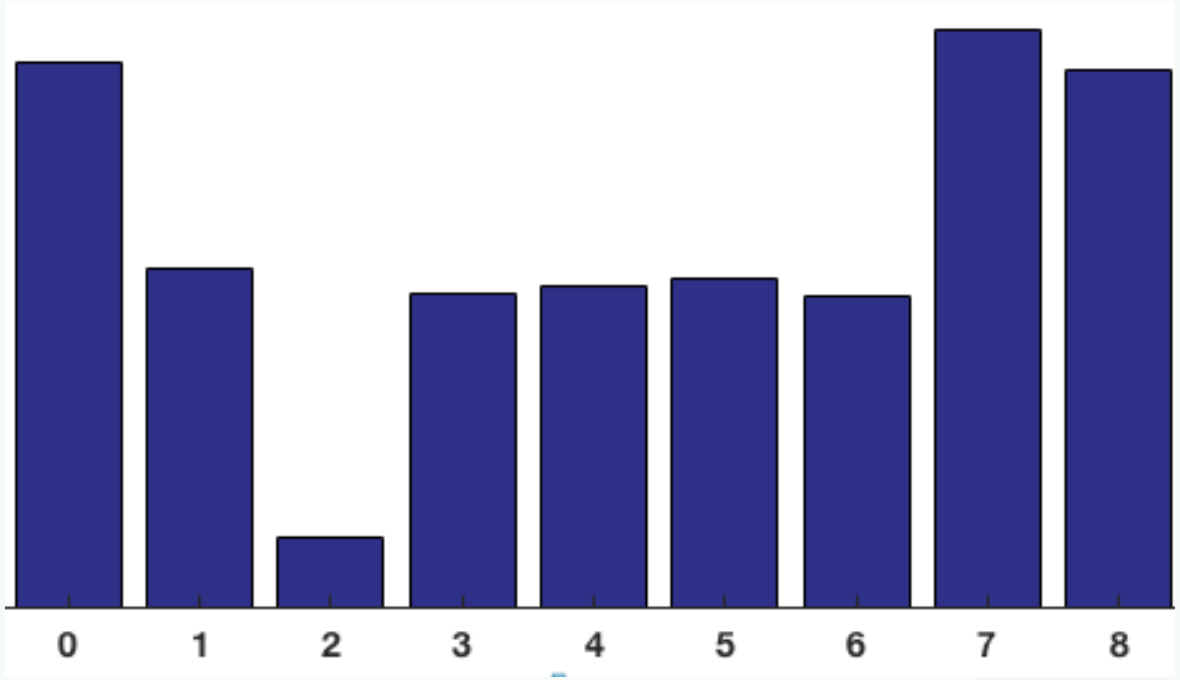
2. Compute HoG in the cells

- Create a histogram of gradients in the se 8×8 cells. The histogram contains 9 bins corresponding to angles 0, 20, 4 0 … 160.
  - The original angles are between 0 and 18 0 degrees instead of 0 to 360 degrees. Th ese are called “unsigned” gradients becau se a gradient and it’s negative are represe nted by the same numbers.



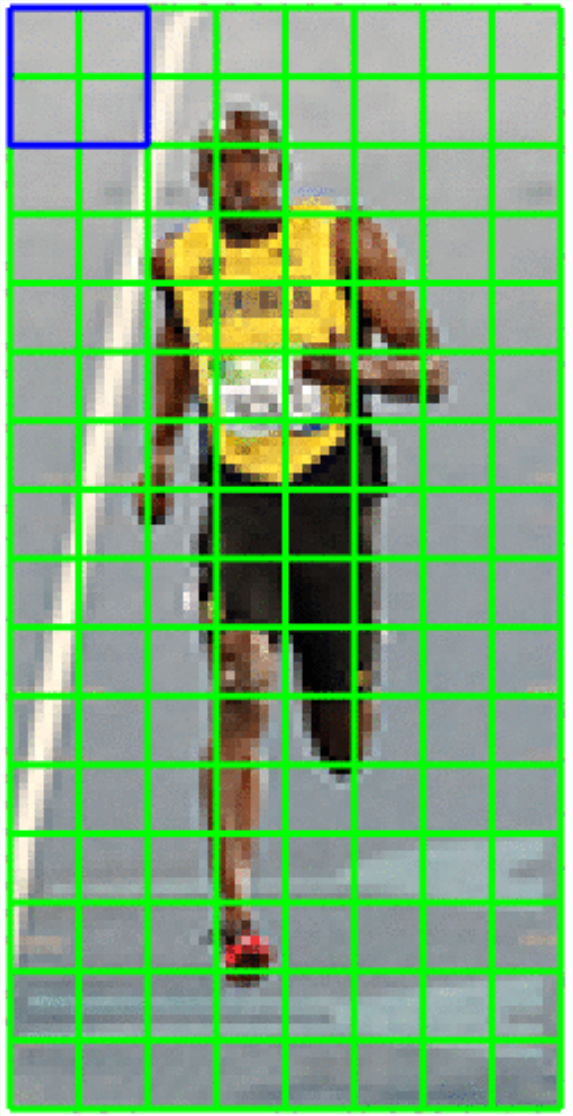
2. Compute HoG in the cells





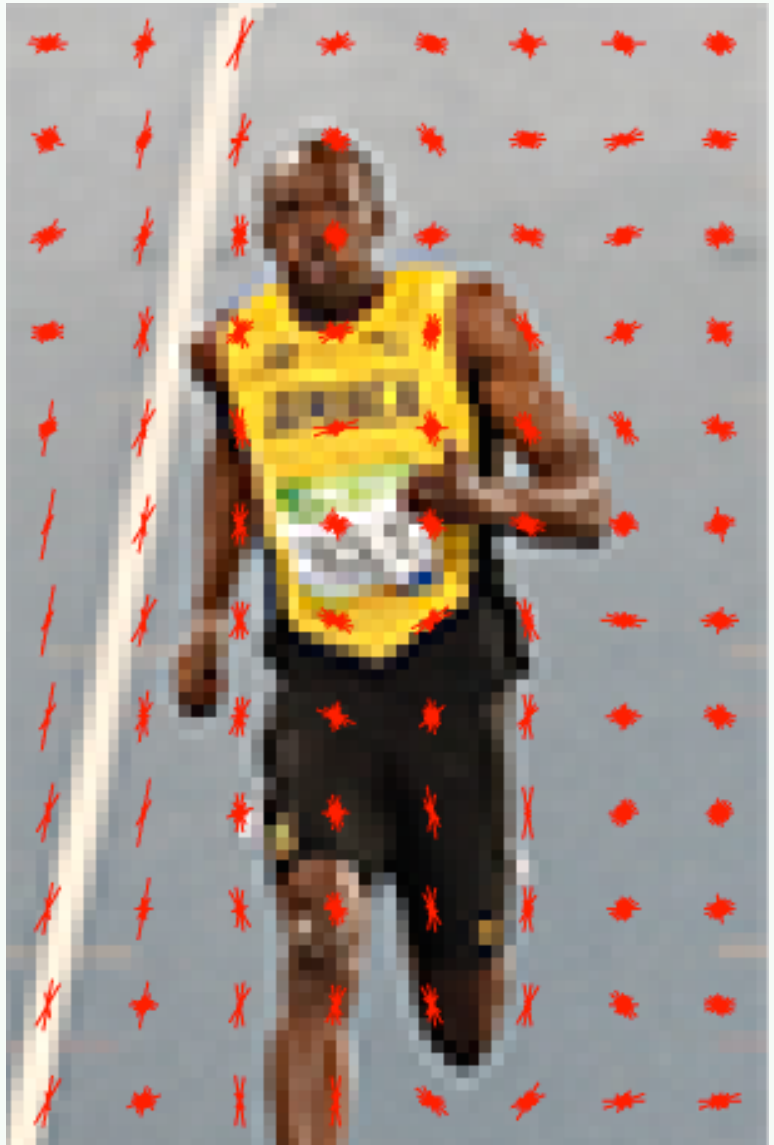
4. 16×16 Block Normalization

- Normalize the histogram vectors of 4 cells (=16x16 pixels)
  - Gradients of an image are sensitive to overall lighting.
  - Ideally, we want our descriptor to be independent of lighting variations. In other words, we would like to “normalize” the histogram, so they are not affected by lighting variations.



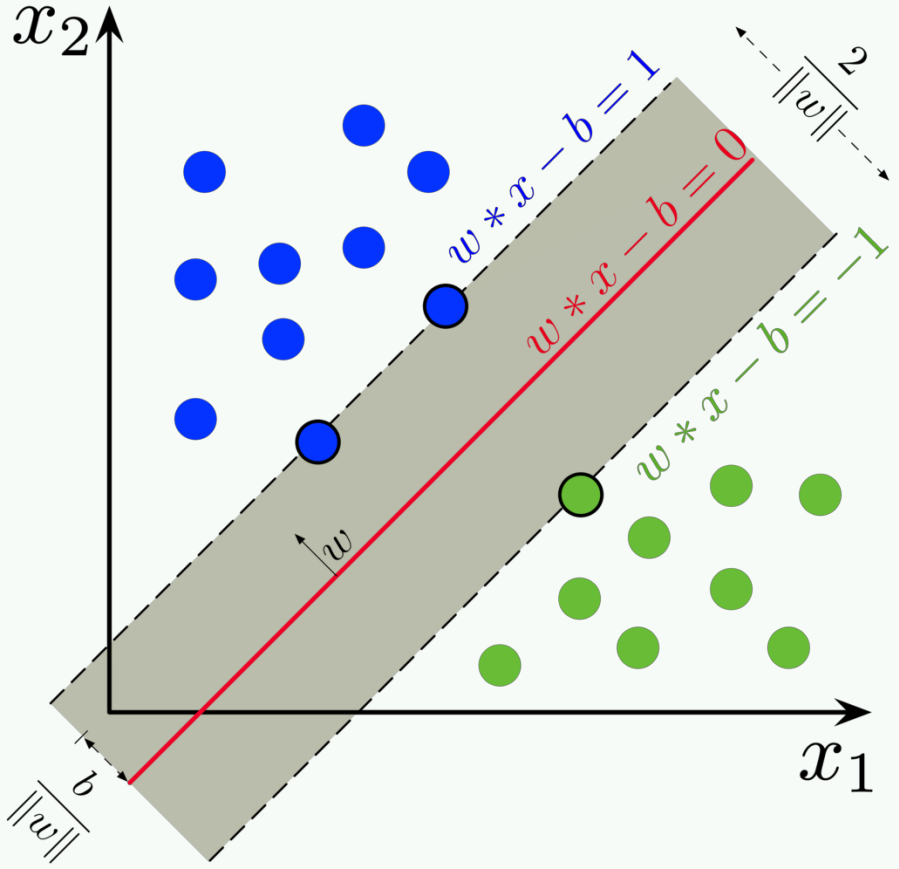
4. 16×16 Block Normalization

- Calculate the HOG feature vector for the entire image patch
  - To calculate the final feature vector for the entire image patch , the 36×1 vectors are concatenated into one giant vector.
  - There are 7 horizontal and 15 vertical positions making a tota l of 7 x 15 = 105 positions.
  - Each 16×16 block is represented by a 36×1 vector. So when we concatenate them all into one giant vector, we obtain a 36 ×105 = 3780-dimensional vector.

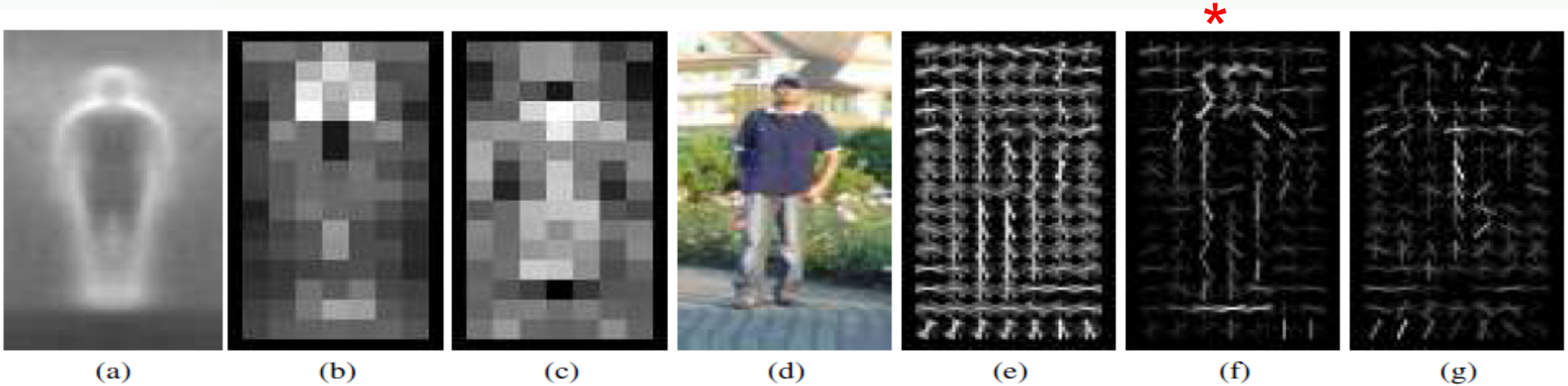


5. SVM Classification

- An SVM is a maximum margin discriminator that maps a feature vector to a real nu mber score whose value is designed to separate targets from clutter.



Pictorial Example



(a) average gradient image over training examples
(b) each “pixel” shows max positive SVM weight in the block centered on that pixel
(c) same as (b) for negative SVM weights
(d) test image
(e) its R-HOG descriptor
(f) R-HOG descriptor weighted by positive SVM weights
(g) R-HOG descriptor weighted by negative SVM weights