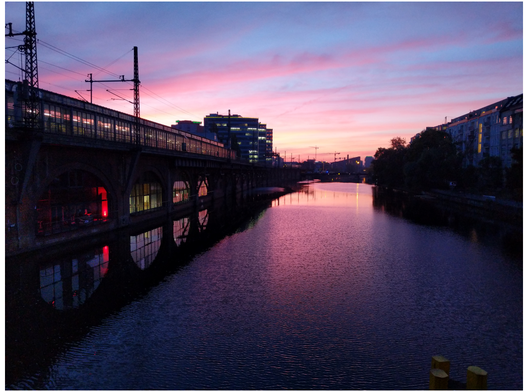
Features

Lecture 06

Computer Vision for Geosciences

2021-03-26

Universidad Nacional Autónoma de México

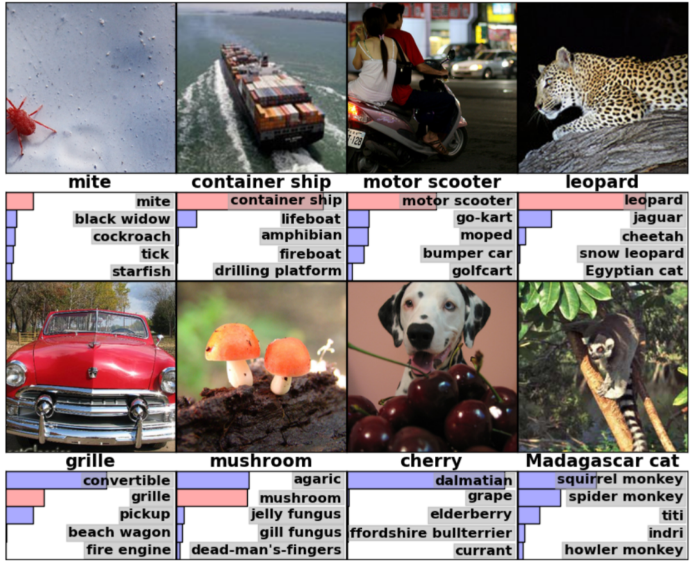
 

- Images in pixel space describe a point in a high dimensional space
- If we ignore spatial relation of pixels it is a point in a space with *wxhxc* dimensions
- Small changes in the real world scene lead to big changes in pixel space
- In this scene we see two times the same location in Berlin from two slightly different angles at different times.

- Two times the same image, with and without noise.

- Simple linear transformations in object space (e.g translation of an object), moves the data point in pixel space to a completely different location (highly non-linear)

Categorize image content



Krizhevsky et al, ImageNet Classification with Deep Convolutional Neural Networks

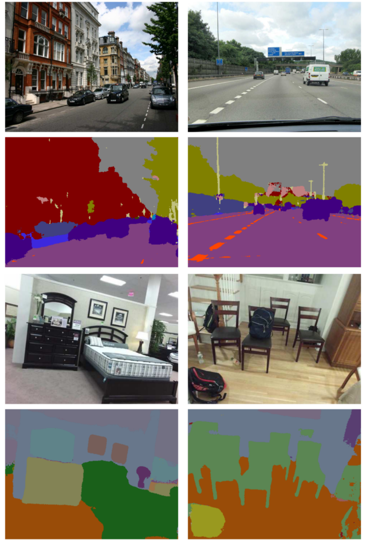
- If we want to interpret what is happening in the image we need a representation in a space that allows a similarity measure
- Two images or image patches that show the same thing with respect to our task should be similar in representation space
- We want to be able to compare images or image patches for various reasons: categorize image content, ...

Finding/tracking objects

- finding and tracking objects, ...

David G.Lowe, Distinctive Image Features from Scale-Invariant Keypoints

Segmentation



- segmentaion of regions, ...

Badrinarayanan et al, SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation

Feature

- Abstract representation of image content
- Encodes relevant properties
- Describes full image or parts of it

- If we want to interpret what is happening in the image we need a representation in a space that allows a similarity measure
- Two images or image patches that show the same thing with respect to our task should be similar in representation space

- We need to find more abstract representations that describe the content of an image invariant to properties that are irrelavant to the task.

Invariance

- Translation
- Rotation
- Scale
- Illumination
- Specification

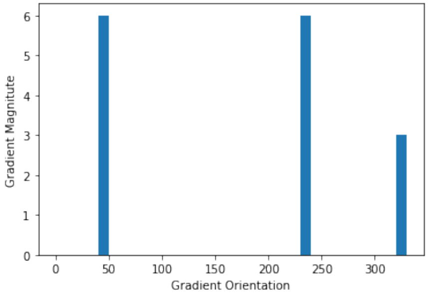
- Potential properties of entities we want to describe.

Properties

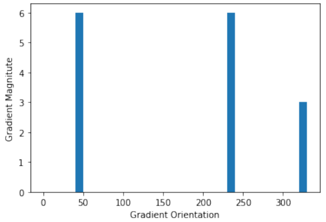
- Shape
- Color
- Texture
- Location
- Motion

Gradients

- Even small lighting changes would make huge differences in pixel space.
- Gradient images alleviate that a little, but magnitude still varies with lighting.
- Contain information about texture and shape.
- Invariant to color.

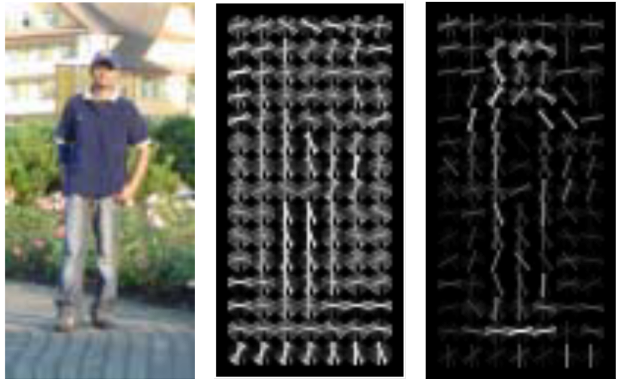


- We use the edge image as basis and partition it into a grid of cells.
- For every cell we count the 'votes' for every pixel in the cell. I.e. add a weighted entry into a histogram with the gradient direction of that pixel.
- Histogram entries are a vector that describes content of the image patch (cell).



- Concatenate cell histograms for description of larger region
- Use 180 degree binning to ignore edge direction
- **Normalize histograms for region (contrast normalization)**

HOG (Histograms of oriented gradients)[1]



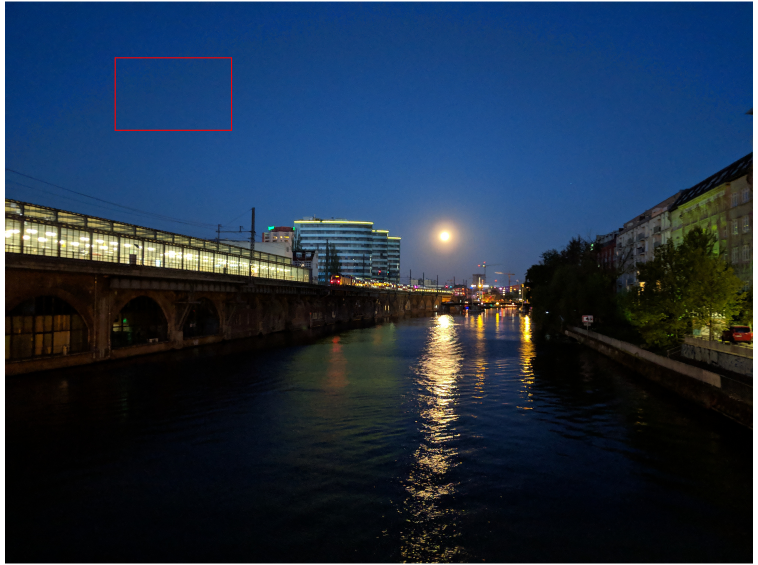
[1]Navneet Dalal and Bill Triggs, Histograms of oriented gradients for human detection

- HOG is an image descriptor based on histograms of gradients
- It is one of many descriptors based on gradient images
- Picture shows a visualization from the paper, which I recommend to read as it is comparably easy to understand but insightful.

Which parts of the image are most descriptive?

- We do not necessarily want to describe the whole image.
- Which parts are most descriptive?

Which parts of the image are most descriptive?



- We do not necessarily want to describe the whole image.
- Which parts are most descriptive?

Which parts of the image are most descriptive?

- We do not necessarily want to describe the whole image.
- Which parts are most descriptive?
- Edges are more interesting than no edges.

Which parts of the image are most descriptive?

- We do not necessarily want to describe the whole image.
- Which parts are most descriptive?
- Corners are even more interesting.

How to compute uniqueness of region?

- Compare patch with all other image patches of all images?
- We can compare the patch with all other patches within the same image (self-difference).
- To compute self-difference for the whole image would be very expensive.

Approximate self-difference: gradients.

- Regions without edges (low-gradients) have low self-difference
- Regions with edges (gradients in one direction) have some self-difference
- Regions with corners (gradients in multiple directions) have high self-difference

Structure tensor

$$A(x,y)=\begin{pmatrix} \sum_{(i,j)\in W} f_x(i,j)^2 & \sum_{(i,j)\in W} f_x(i,j)f_y(i,j) \\ \sum_{(i,j)\in W} f_x(i,j)f_y(i,j) & \sum_{(i,j)\in W} f_y(i,j)^2 \end{pmatrix}$$

- Autocorrelation of gradient image
- f_x is gradient in x direction, f_y is gradient in y direction
- W is a region within the image
- Slide from lecture Digital Image Processing @ Technische Universität Berlin

Structure tensor: edge

$$A(x,y)=\begin{pmatrix} \sum_{(i,j)\in W} f_x(i,j)^2 & \sum_{(i,j)\in W} f_x(i,j) f_y(i,j) \\ \sum_{(i,j)\in W} f_x(i,j) f_y(i,j) & \sum_{(i,j)\in W} f_y(i,j)^2 \end{pmatrix}$$

$$f=\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

Edge

$$f_x=\begin{bmatrix} 0 & 0 & -0.5 & -0.5 & 0 \\ 0 & 0 & -0.5 & -0.5 & 0 \\ 0 & 0 & -0.5 & -0.5 & 0 \\ 0 & 0 & -0.5 & -0.5 & 0 \\ 0 & 0 & -0.5 & -0.5 & 0 \end{bmatrix} \qquad f_y=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A=\begin{bmatrix} 2.5 & 0 \\ 0 & 0 \end{bmatrix}$$

$$det(A)=0 \qquad tr(A)=2.5 \qquad eig(A)=\begin{bmatrix} 0 \\ 2.5 \end{bmatrix}$$

- We do not necessarily want to describe the whole image.
- Which parts are most descriptive?
- Corners are even more interesting.

Structure tensor: edge

$$A(x,y)=\begin{pmatrix} \sum_{(i,j)\in W} f_x(i,j)^2 & \sum_{(i,j)\in W} f_x(i,j) f_y(i,j) \\ \sum_{(i,j)\in W} f_x(i,j) f_y(i,j) & \sum_{(i,j)\in W} f_y(i,j)^2 \end{pmatrix}$$

$$f=\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Edge

$$f_x=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$f_y=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -0.5 & -0.5 & -0.5 & -0.5 & -0.5 \\ -0.5 & -0.5 & -0.5 & -0.5 & -0.5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A=\begin{bmatrix} 0 & 0 \\ 0 & 2.5 \end{bmatrix}$$

$$det(A)=0$$

$$tr(A)=2.5$$

$$eig(A)=\begin{bmatrix} 0 \\ 2.5 \end{bmatrix}$$

- We do not necessarily want to describe the whole image.
- Which parts are most descriptive?
- Corners are even more interesting.

Structure tensor: corner

$$A(x,y)=\begin{pmatrix} \sum_{(i,j)\in W} f_x(i,j)^2 & \sum_{(i,j)\in W} f_x(i,j)f_y(i,j) \\ \sum_{(i,j)\in W} f_x(i,j)f_y(i,j) & \sum_{(i,j)\in W} f_y(i,j)^2 \end{pmatrix}$$

$$f=\begin{bmatrix} 0&0&0&0&0\\ 0&0&1&0&0\\ 0&1&1&1&0\\ 0&0&1&0&0\\ 0&0&0&0&0 \end{bmatrix}$$

Point / Blob

$$f_x=\begin{bmatrix} 0&0&0&0&0\\ 0&0.5&0&-0.5&0\\ 0.5&0.5&0&-0.5&-0.5\\ 0&0.5&0&-0.5&0\\ 0&0&0&0&0 \end{bmatrix}$$

$$f_y=\begin{bmatrix} 0&0&0.5&0&0\\ 0&0.5&0.5&0.5&0\\ 0&0&0&0&0\\ 0&-0.5&-0.5&-0.5&0\\ 0&0&-0.5&0&0 \end{bmatrix}$$

$$A=\begin{bmatrix} 2&0\\ 0&2 \end{bmatrix}$$

$$det(A)=4$$

$$tr(A)=4$$

$$eig(A)=\begin{bmatrix} 2\\ 2 \end{bmatrix}$$

- We do not necessarily want to describe the whole image.
- Which parts are most descriptive?
- Corners are even more interesting.

Structure tensor: eigenvalues

- Both eigenvalues $\approx 0 \rightarrow$ unstructured area
- One eigenvalues $\approx 0 \rightarrow$ edge
- Both eigenvalues $> 0 \rightarrow$ corner

Structure tensor: eigenvalues

$$det(A) - \alpha tr(A)^2 = \lambda_1 \lambda_2 - \alpha(\lambda_1 + \lambda_2)^2 \tag{1}$$

$$\frac{det(A)}{tr(A)} = \frac{\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)} \tag{2}$$

Keypoints using Harris corner detector

- Use weighted sum (Gaussian) for entries in structure tensor
- Non-maximum supression as in Canny edge detector

Which parts of the image are most descriptive?

- Result of Harris corner detector

As good as it gets? No!

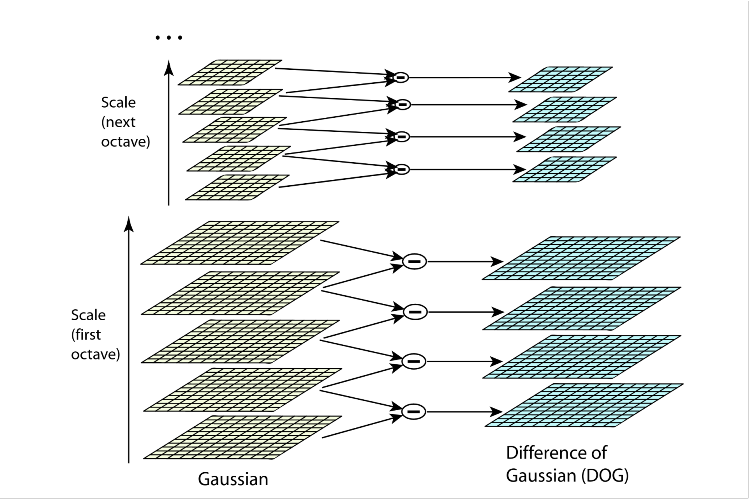
- Harris is not invariant to scale!
- HOG is not invariant to rotation!

As good as it gets? SIFT (Scale-invariant feature transform)[2]

- Keypoint detection and feature descriptors in one method
- Keypoint detection and descriptor based on gradients similar to HOG/Harris
- Uses different scales of the image to achieve scale invariance
- Uses gradient orientation normalization to achieve scale invariance

[2] David G.Lowe, Distinctive Image Features from Scale-Invariant Keypoints

Scale spaces are often used to achieve scale-invariance



- DoG emphasizes edges
- We search for edges on multiple scales
- Non-maximum supression in scale space rejects points with low contrast

- SIFT, HOG and Harris are among the most popular but there are many others
- Metrics for relevance can be defined on higher levels (saliency and attention)