ETH Zürich
Chair of Software Engineering

Robotics Programming Laboratory

Bertrand Meyer
Jiwon Shin

Lecture 8: Robot Perception

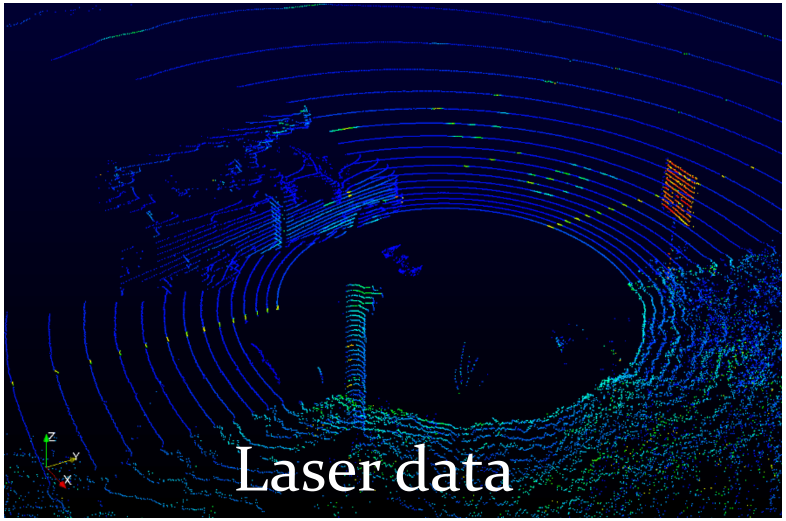
Perception

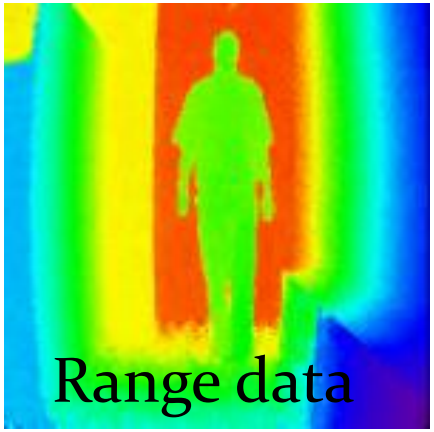
http://pascallin.ecs.soton.ac.uk/challenges/VOC/databases.html#Caltech

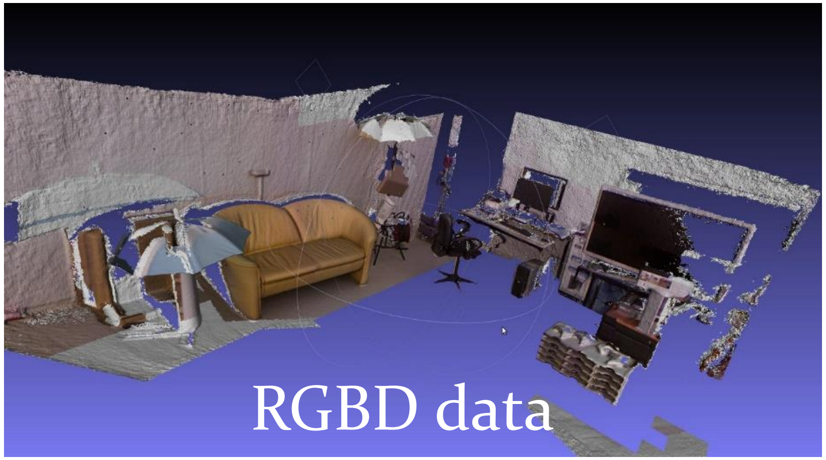
Given visual input, understand the information the input contains

- Object location: object detection
- Type of object: object classification
- Exact object name: object recognition
- Overall scene: scene understanding

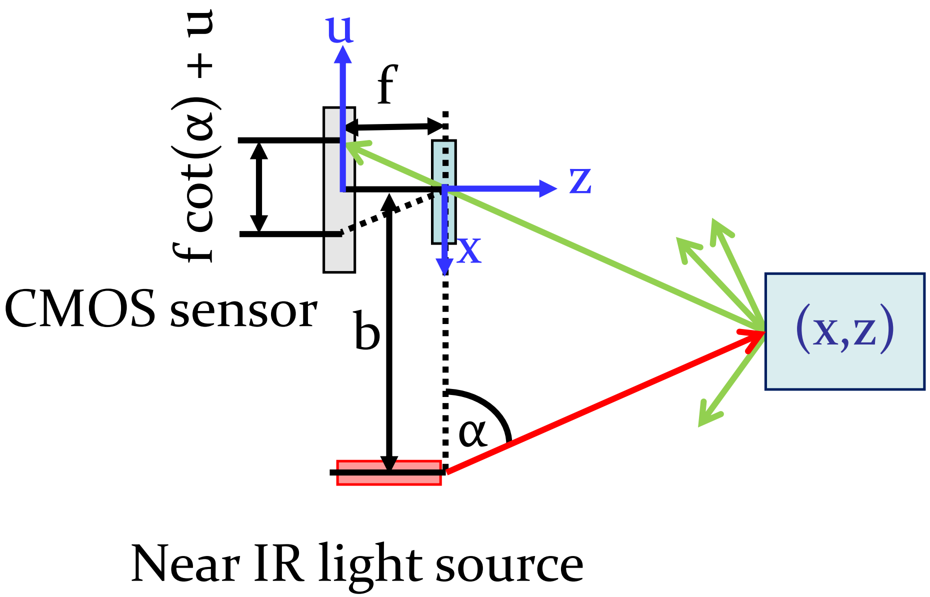
Type of data







Structured light



$$x = \frac{b \cdot u}{f\cot(\alpha) + u} \qquad z = \frac{b \cdot f}{f\cot(\alpha) + u}$$

$$\frac{\partial u}{\partial z} = G_p = \frac{b \cdot f}{z^2}$$

$$\frac{\partial \alpha}{\partial z} = G_\alpha = \frac{b \sin(\alpha)^2}{z^2}$$

Carmine 1.09

- Operating range: 0.35 m – 1.4 m
- Spatial resolution: 0.9 mm at 0.5m
- Depth resolution: 0.1 cm at 0.5m

Structured light

Segmentation

Segmentation: decomposition of an image into consistent regions

- Data that belong to the same region have similar properties
  - Similar color, texture, surface normal, etc.
- Data that belong to different regions have different properties
  - Different color, texture, surface normal, etc.

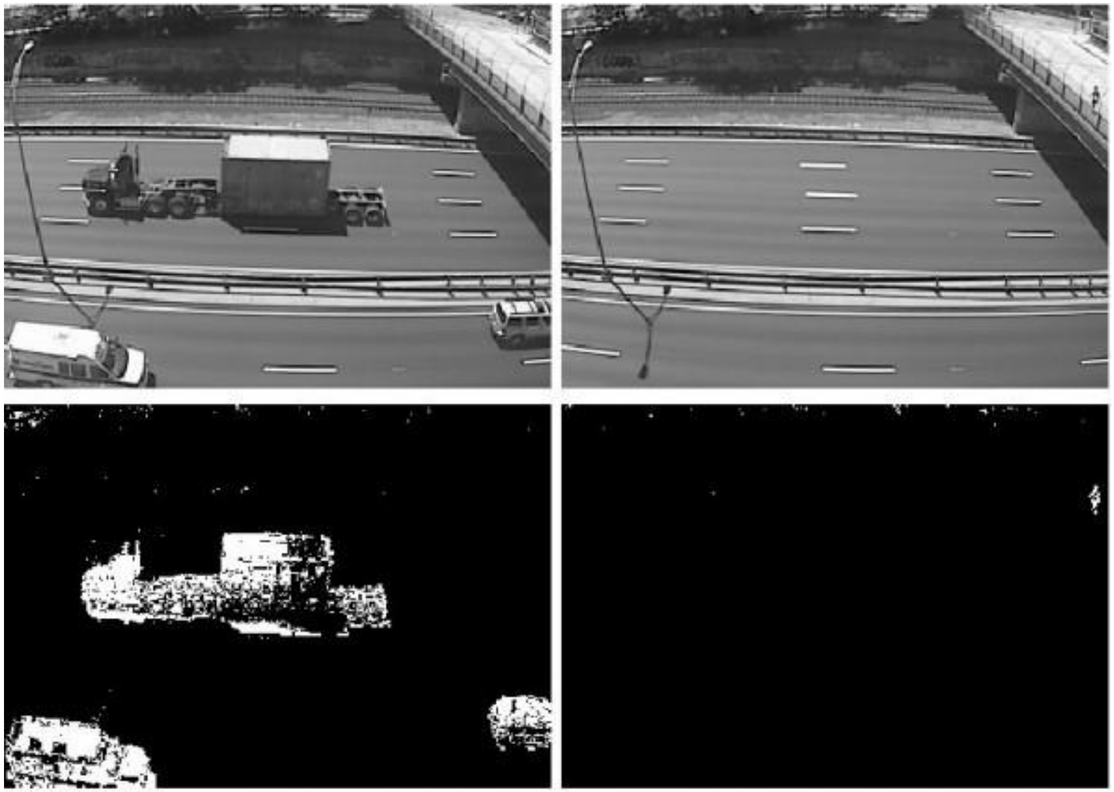
- Segmentation as clustering
  - Partitioning: divide an image into coherent regions
  - Grouping: group together elements of similar properties

Image segmentation

- Divide an image into sensible regions using pixel intensity, color, texture, etc.

- Background subtraction
- Clustering
- Graph-based

Background subtraction



http://vip.bu.edu/files/2010/02/FDR_FPR_control_comparison1-594x636.jpg

Background subtraction

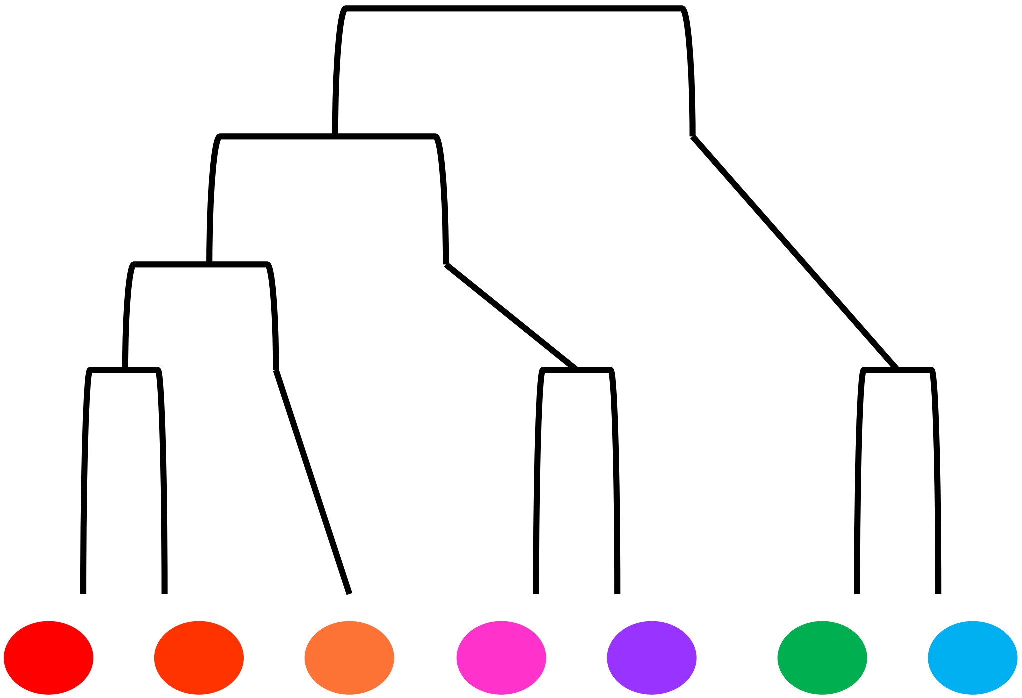
- Subtract an estimate of the appearance of the background from the image
- Consider areas of large absolute difference to be foreground

Issues

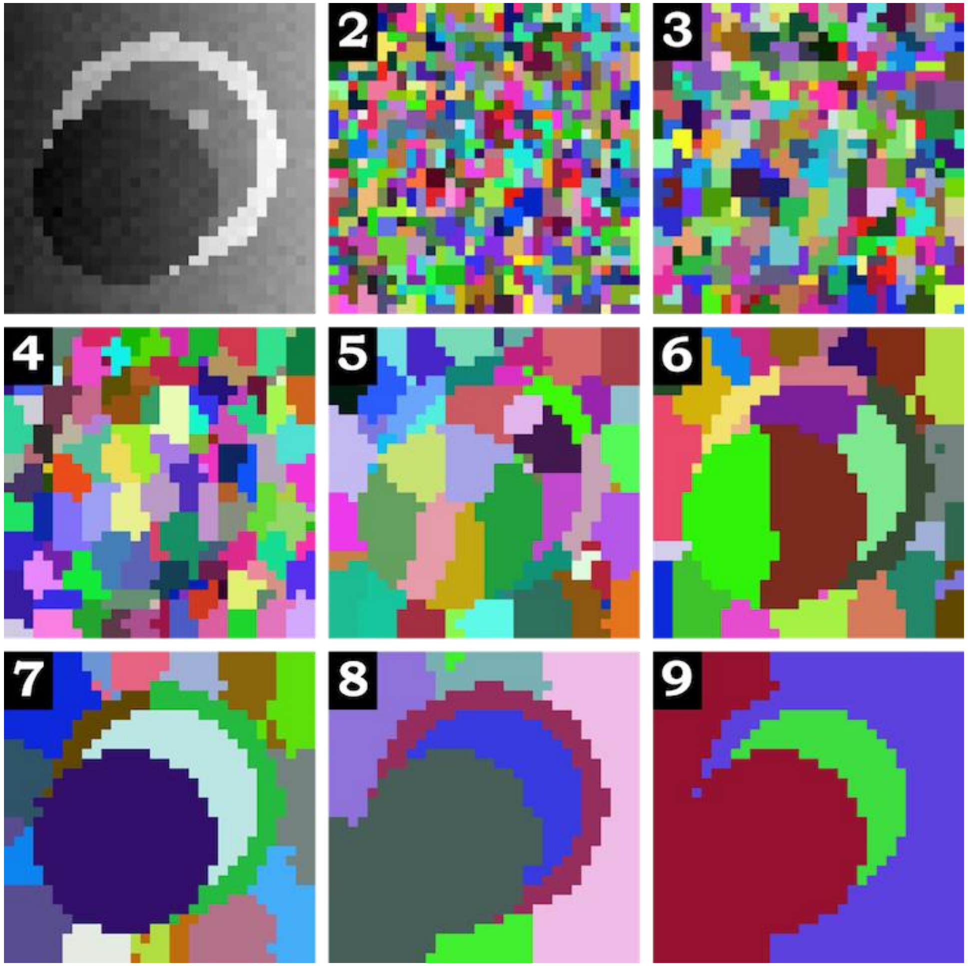
- Obtaining a good estimate of the background is non-trivial
  - Changes in environment, lighting, weather, etc.
  - Use a moving average
- Threshold

Agglomerative clustering

- Consider each data point as a cluster
- Recursively merge the clusters with the smallest inter-cluster distance until the result is satisfactory



Agglomerative clustering



http://www.cse.buffalo.edu/~jcorso/r/files/multilevel_square.png

Agglomerative clustering

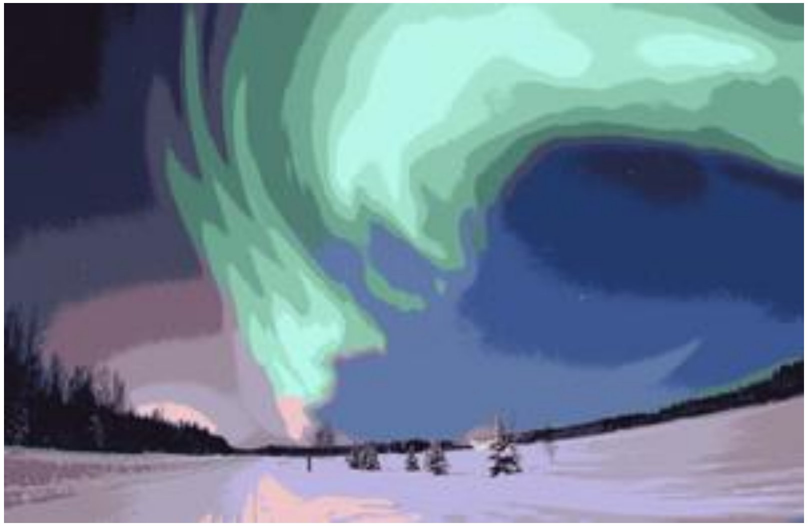
Issues

- Inter-cluster distance
    - Distance between closest elements
    - Distance between farthest elements
    - Average distance between elements
- Number of clusters

K-means clustering

- Choose k data points as seed points
- Recursively assign each data point to the cluster whose center is the closest and recalculate the cluster mean until the center does not change
- Minimize the within cluster sum of squares
    - Tries to produce k clusters of equal size

K-means clustering



http://en.wikipedia.org/wiki/Segmentation_(image_processing)

K-means clustering

Issues

- Segments are not connected in image
    - Using pixel coordinates would break up large regions
- Determining k is non-trivial

Efficient graph-based image segmentation

- Represent image as a graph, each pixel being a node of a graph
- Edges are formed between neighboring pixels
- Merge the nodes such that nodes belonging to the same segment more similar to one another than nodes at the boundary of two segments

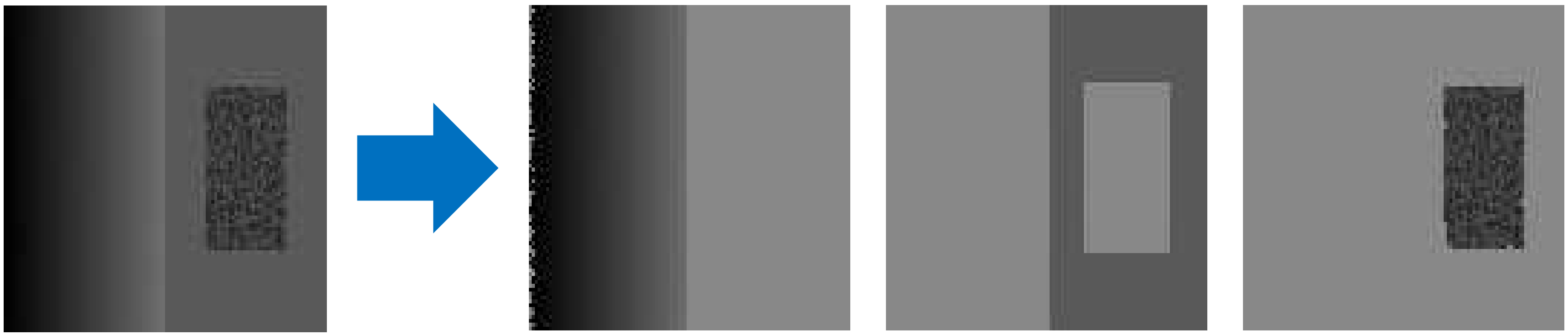
Efficient graph-based image segmentation

- Internal difference of a cluster c:
  - $Int(C) = \max_{e \epsilon MST(C,E)} w(e)$
- Difference between clusters c_1, c_2:
  - $Dif(C_1, C_2) = \min_{v_i \epsilon C_1, v_j \epsilon C_2, (v_i, v_j) \in E} w((v_i,\ v_j))$
- Minimum internal difference:
  - $MInt(C_1, C_2) = \min(Int(C_1) + \tau(C_1),\ Int(C_2) + \tau(C_2))$
  - $\tau(C) = \frac{k}{|C|}$
- A boundary exists between c_1 and c_2 if $Dif(C_1, C_2) > MInt(C_1, C_2)$

Efficient graph-based image segmentation

Felzenszwalb, P. and Huttenlocher, D. 2004. "Efficient Graph-Based Image Segmentation" International Journal of Computer Vision, Volume 59, Number 2.

Efficient graph-based image segmentation



- Regions of consistent properties are grouped together

Issues

- Number and quality of segments depend on the parameter k, smoothing factor , and minimum number of nodes

Range data segmentation

- Generally, we can use image segmentation algorithms by replacing intensity, color, or texture by depth, surface normal, etc.

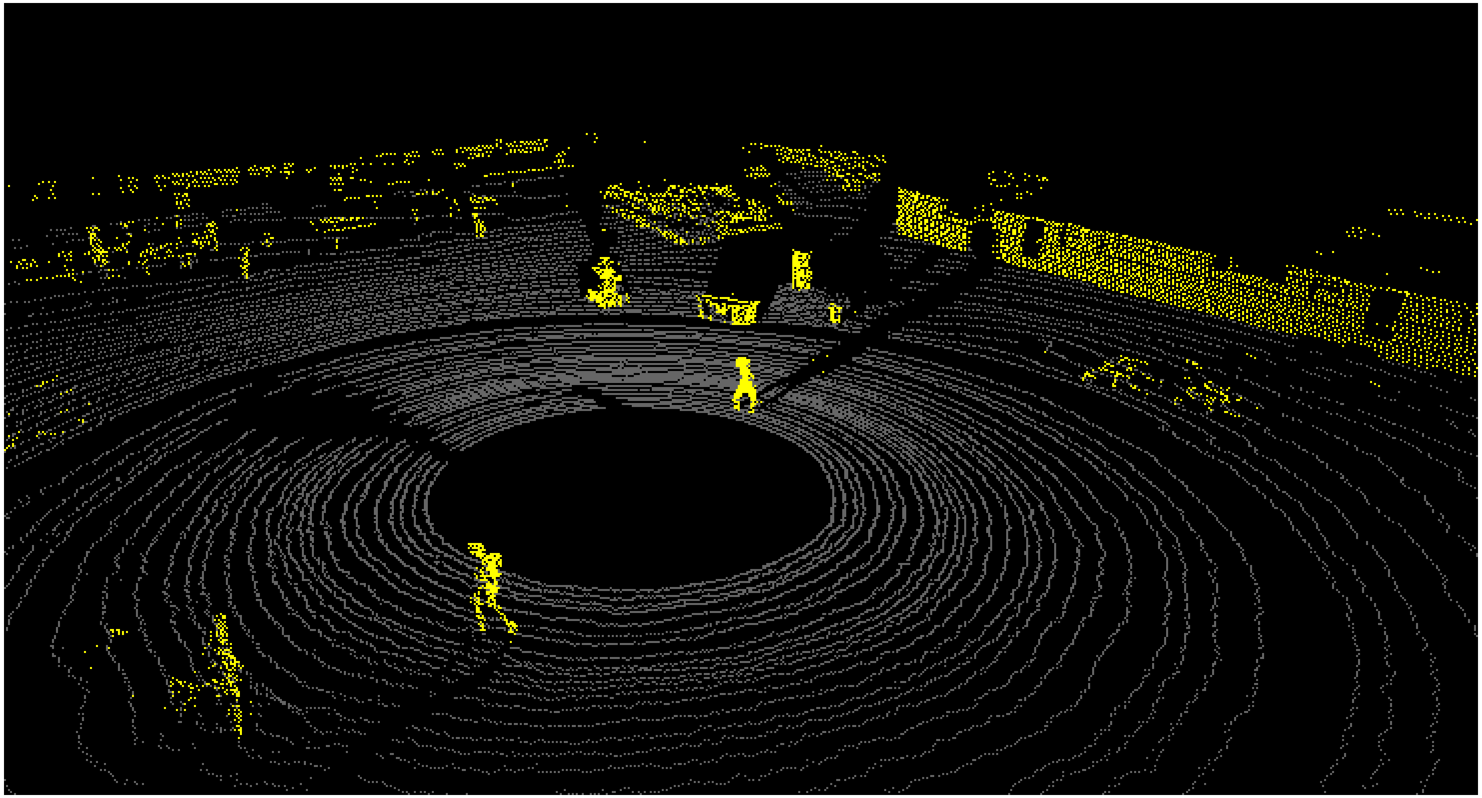
Surface normal computation

$$\boldsymbol{x}_u \equiv {}^{\partial \boldsymbol{x}}\!/_{\partial u}$$

$$\boldsymbol{x}_v \equiv {}^{\partial \boldsymbol{x}}\!/_{\partial v}$$

$$N = \frac{1}{|\boldsymbol{x}_u \times \boldsymbol{x}_v|}(\boldsymbol{x}_u \times \boldsymbol{x}_v)$$

Ground segmentation



http://www-personal.acfr.usyd.edu.au/p.morton/media/img/data_ground.png

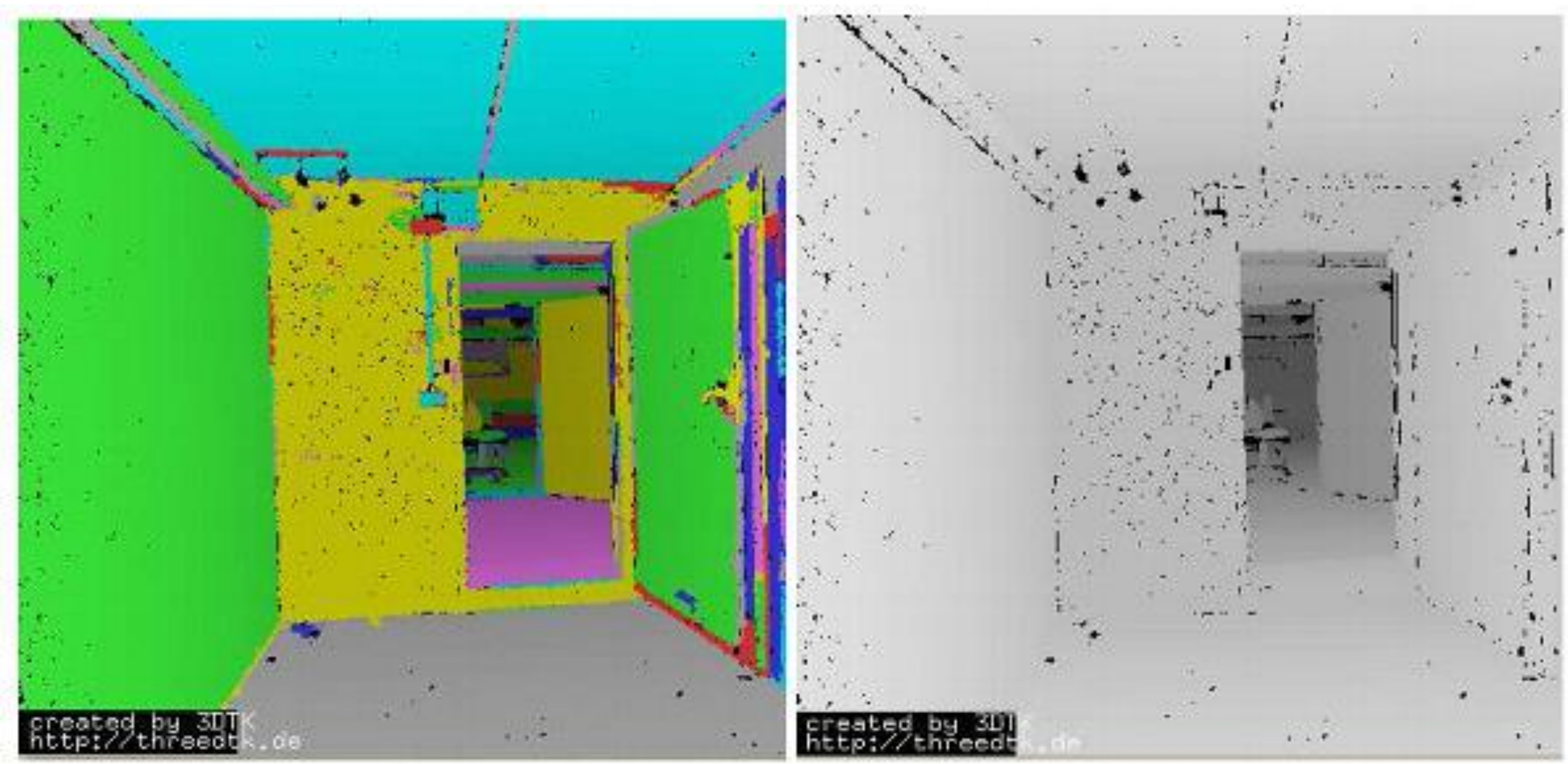
Ground segmentation

- Extract all points below a certain height

Issues

- Data are noisy
    - Objects will also lose information
- Wall cannot be segmented out
- Ground is not always planar

Plane segmentation



http://kos.informatik.uni-osnabrueck.de/icar2013/segmentation.png

Plane segmentation

- Find a plane that minimize the average distance between a set of points and the surface
- Recursively merge the surface patches

Issues

- Not every object is planar
    - Curved objects will be segmented into several segments

Feature extraction

Feature: a piece of information relevant for solving a computational task, e.g., locating an object in an image

- Raw data
- Histogram
- Pyramid of histograms
- Shape

Histogram

- Compute a histogram of intensity or color
- Compute the correlation between example and test

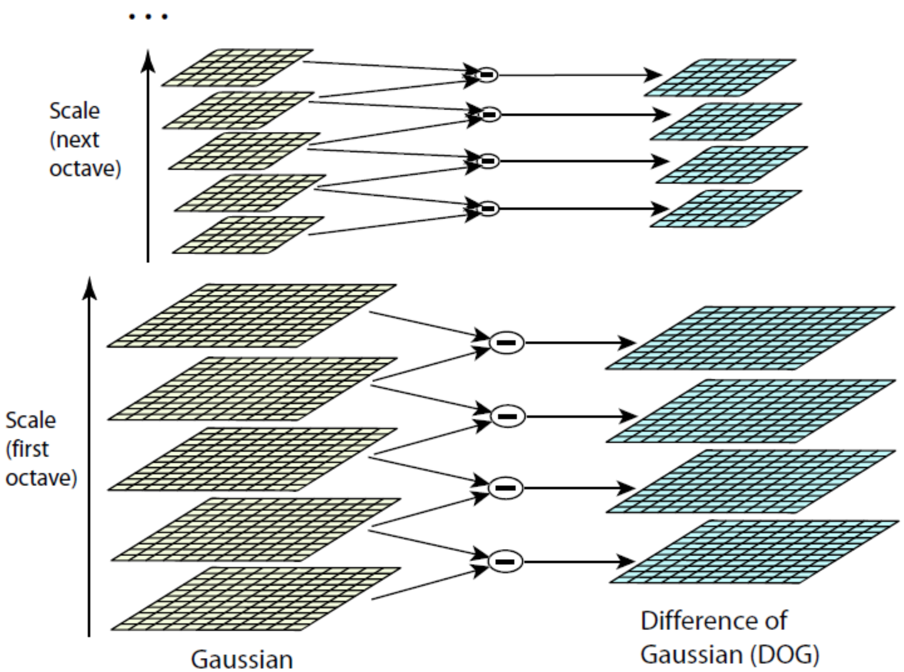
Issues

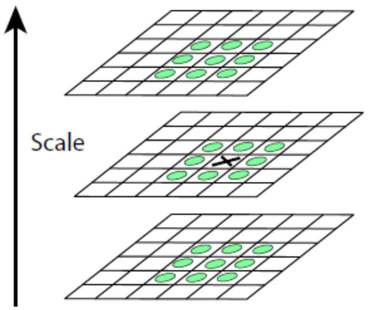
- Loss of the structural information
- Dimensionality

Scale Invariant Feature Transform

Lowe, D. 1999. "Object recognition from local scale-invariant features". Proceedings of ICCV.

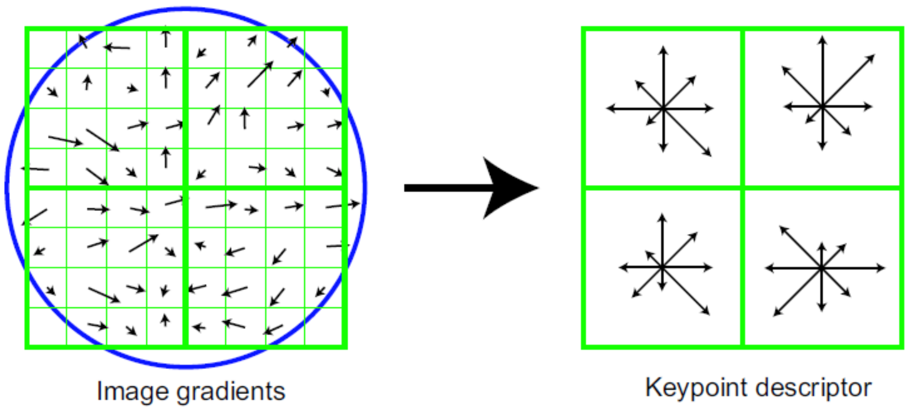
Scale Invariant Feature Transform (SIFT)





- Identify locations and scales that are identifiable from different views of the same object
    - $L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$
    - $D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$
- Detect extrema (local minimum or maximum)

Scale Invariant Feature Transform



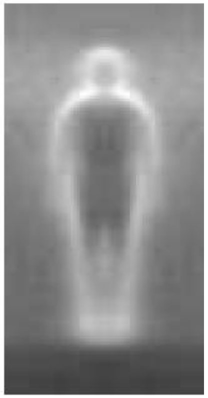
- Remove points of low contrast or poorly localized on an edge
- Orientation assignment

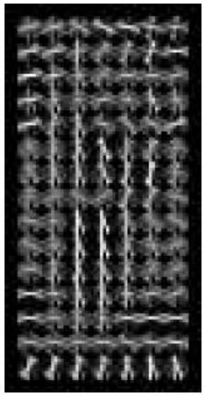
$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1))^2}$$

$$\theta(x,y) = \tan^{-1}\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}$$

- Create a keypoint descriptor: 16 histograms (4x4 grid), each with 8 orientation bins, containing a total of 128 elements.

Histogram of Oriented Gradient





- Divide the image into small rectangular or radial cells
- Each cell accumulates a weighted local 1-D histogram of gradient directions over the pixels of the cell
- Normalize each cell by the energy over larger regions

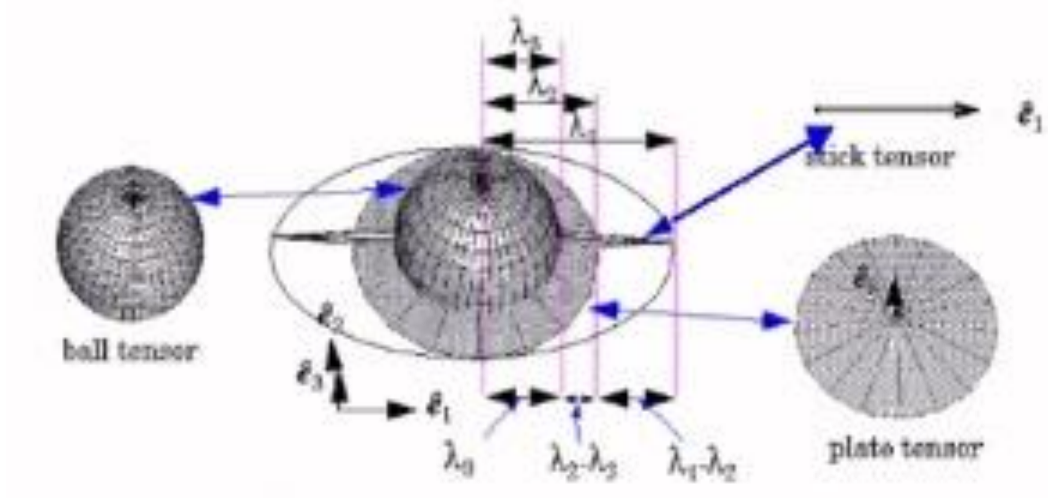
Shape factor

- Compute eigenvectors: λ_1, λ_2, λ_3
    - Point/Spherical: $\lambda_1 \approx \lambda_2 \approx \lambda_3$
    - Planar: $\lambda_1 \approx \lambda_2 \gg \lambda_3$
    - Elongated: $\lambda_1 \gg \lambda_2 \approx \lambda_3$

Issues

- Many different objects have similar shape factor
- Shape factor of an object can depend on the point of view

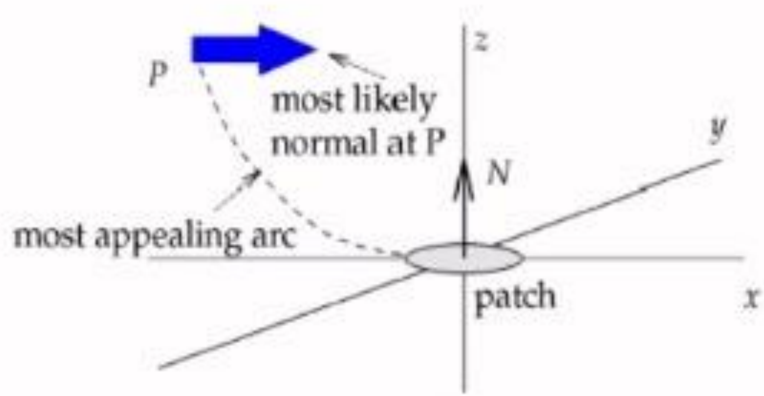
Tensor voting



- 2x2 or 3x3 matrix that captures both the orientation information and its confidence/saliency
  - Shape defines the type of information (point, surface, etc.)
  - Size represents the saliency
- Each token is first decomposed into the basis tensors, and then broadcasts its information to its neighbors.

Medioni, G., Lee, M., Tang. C. 2000. A Computational Framework for Segmentation and Grouping.

Tensor voting

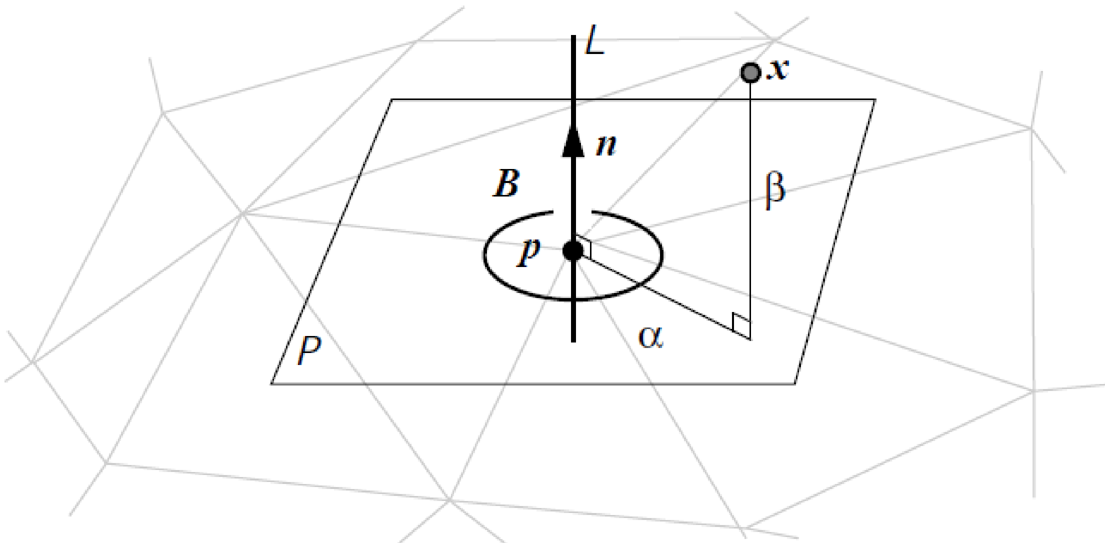


- The magnitude of the vote decays with distance and curvature:

$$V(d,\rho) = e^{-\frac{d^2+c\rho^2}{\sigma^2}}$$

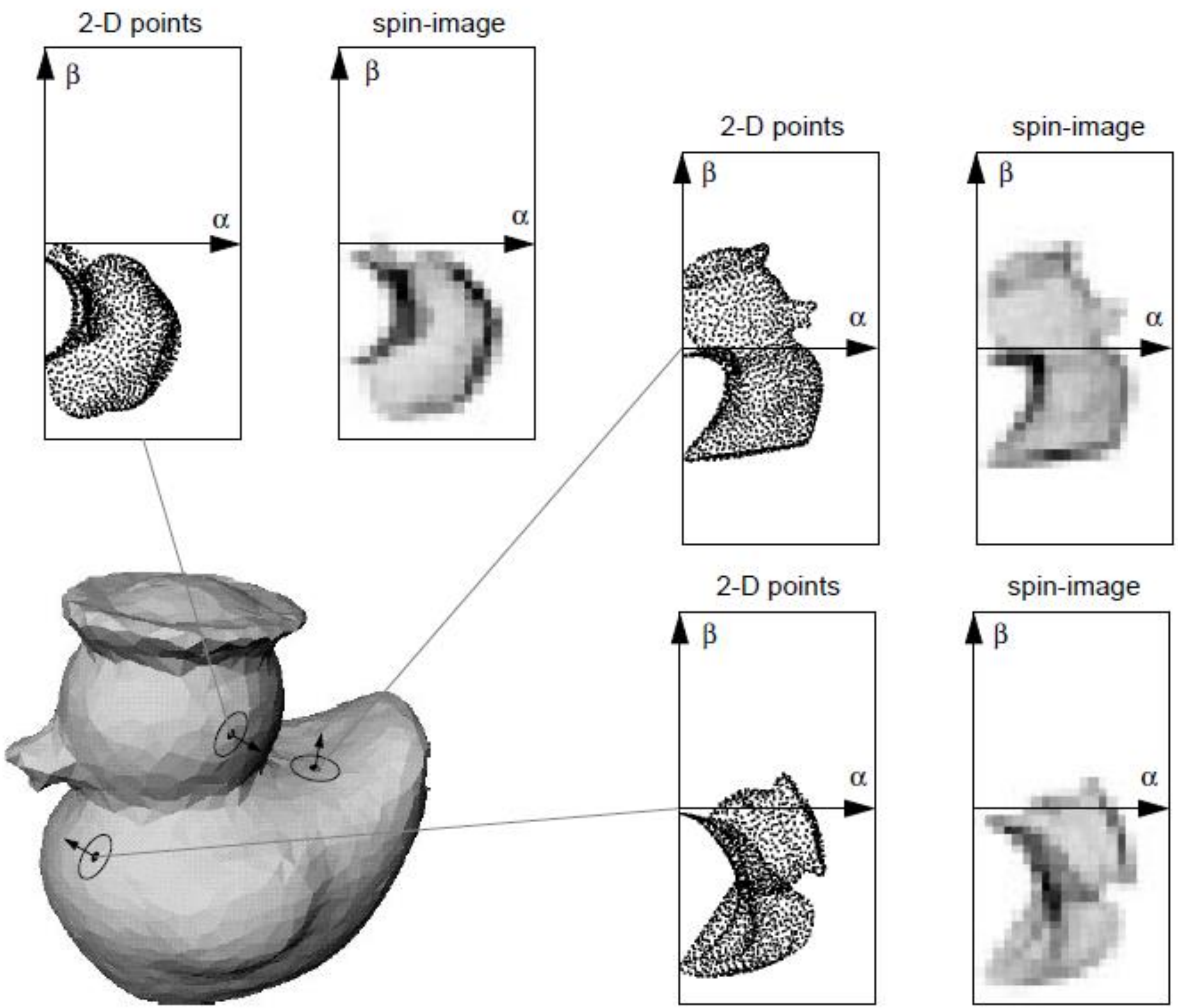
  - d is the distance along the smooth path
  - ρ is the curvature of the path
  - c controls the degree of decay
  - σ controls the size of the voting neighborhood
- Accumulate the votes by adding the matrices
- Analyze the tensor by eigen decomposition

Spin image



Johnson, A., Herbert, M., 1999. "Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes" IEEE Transactions on Pattern Analysis and Machine Intelligence, 21, (5).

Spin image



2-D points
spin-image
β
α
2-D points
spin-image
β
α
2-D points
spin-image
β
α

Spin image

- Collect a histogram of points
  - The resolution of the histogram
  - The size of the histogram
- To compare two spin images P and Q
  - Compute the correlation between two images

$$R(P,Q) = \frac{N\sum p_i q_i - \sum p_i \sum q_i}{\sqrt{(N\sum {p_i}^2 - (\sum p_i)^2)(N\sum {q_i}^2 - (\sum q_i)^2)}}$$

  - Can also apply PCA, remove the mean spin image and compute the Euclidean norm

Classifier

- Take a set of labeled examples
- Determine a rule that assign a label to any new example using the labeled examples

- Training dataset ($\mathbf{x}_i$, y_i)
  - $\mathbf{x}_i$: measurements of the properties of objects
  - y_i: label
- Goal: given a new, plausible $\mathbf{x}$, assign it a label y.

Bayes classifier

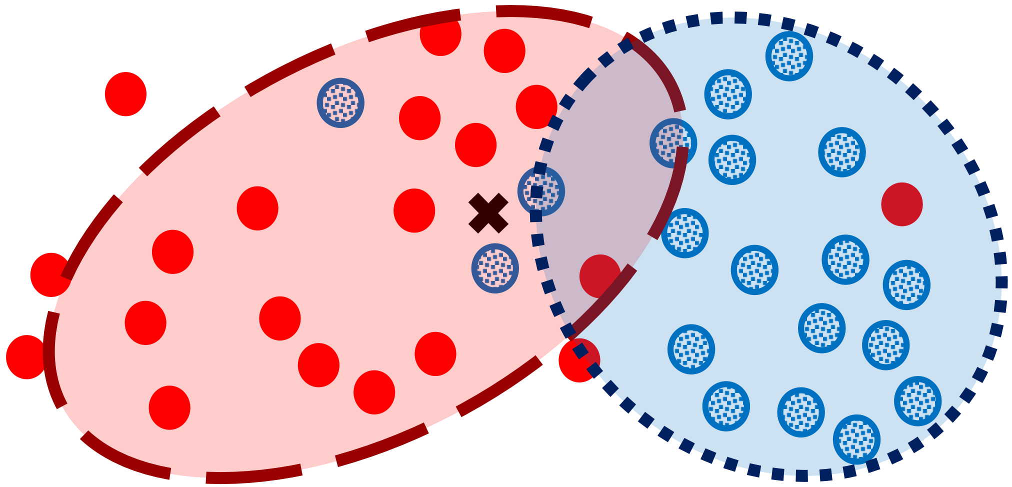
$$p(\, k \mid \mathbf{x}\,) = \frac{p(\,\mathbf{x} \mid k\,)\, p(\,k\,)}{p(\,\mathbf{x}\,)} \propto p(\,\mathbf{x} \mid k\,)\, p(\,k\,)$$

Given $\mathbf{x}$

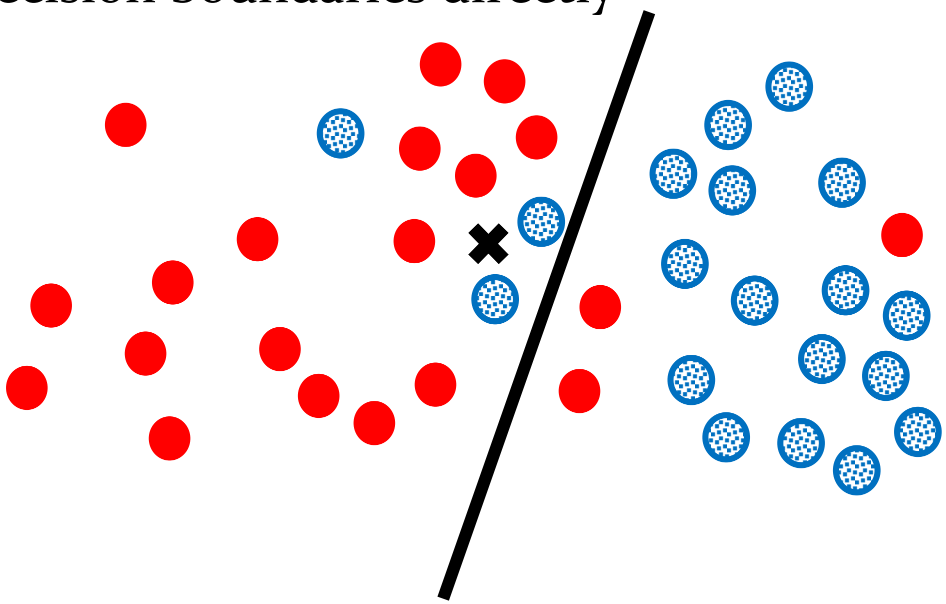
- Assign label k to $\mathbf{x}$ if
  - $p(\,k \mid \mathbf{x}\,) > p(\,i \mid \mathbf{x}\,)$ for all $i \neq k$ and $p(\,k \mid \mathbf{x}\,) >$ threshold
- Assign a random k label between $k_1, \ldots, k_j$ if
  - $p(\,k_1 \mid \mathbf{x}\,) = \ldots = p(\,k_j \mid \mathbf{x}\,) > p(\,i \mid \mathbf{x}\,)$ for all $i \neq k$
- Do not assign a label if
  - $p(\,k \mid \mathbf{x}\,) > p(\,i \mid \mathbf{x}\,)$ for all $i \neq k$ and $p(\,k \mid \mathbf{x}\,) \leq$ threshold

Strategies

- Modeling probability explicitly



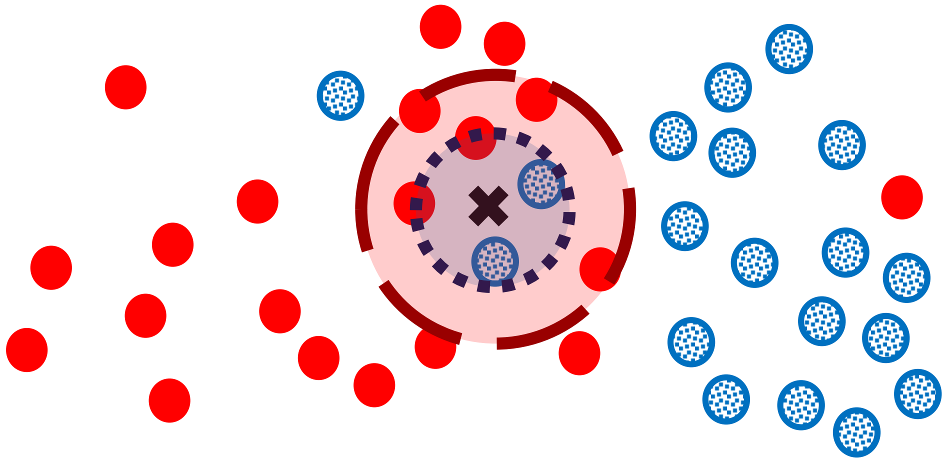
- Determining decision boundaries directly



Nearest neighbor classifier

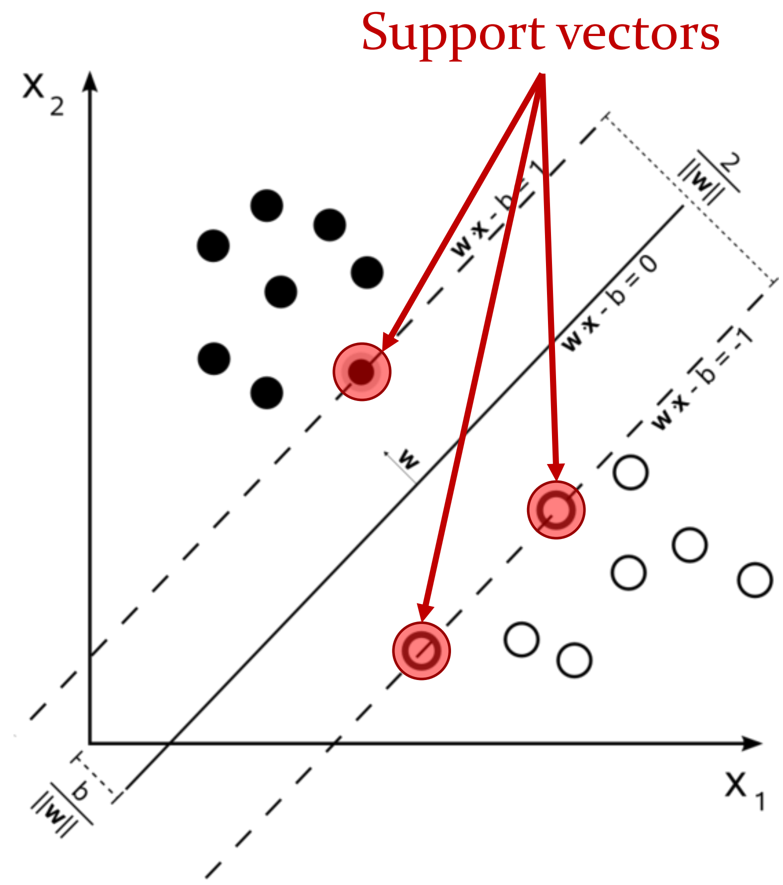
Given $\mathbf{x}$

- Determine M training example that are nearest: $\mathbf{x}_1, \ldots, \mathbf{x}_M$
- Determine class k that has the largest representation n in the set
- Assign label k to $\mathbf{x}$ if n > threshold
- Assign no label otherwise



Support Vector Machine

- Find a hyperplane that maximizes the margin between the positive and negative examples

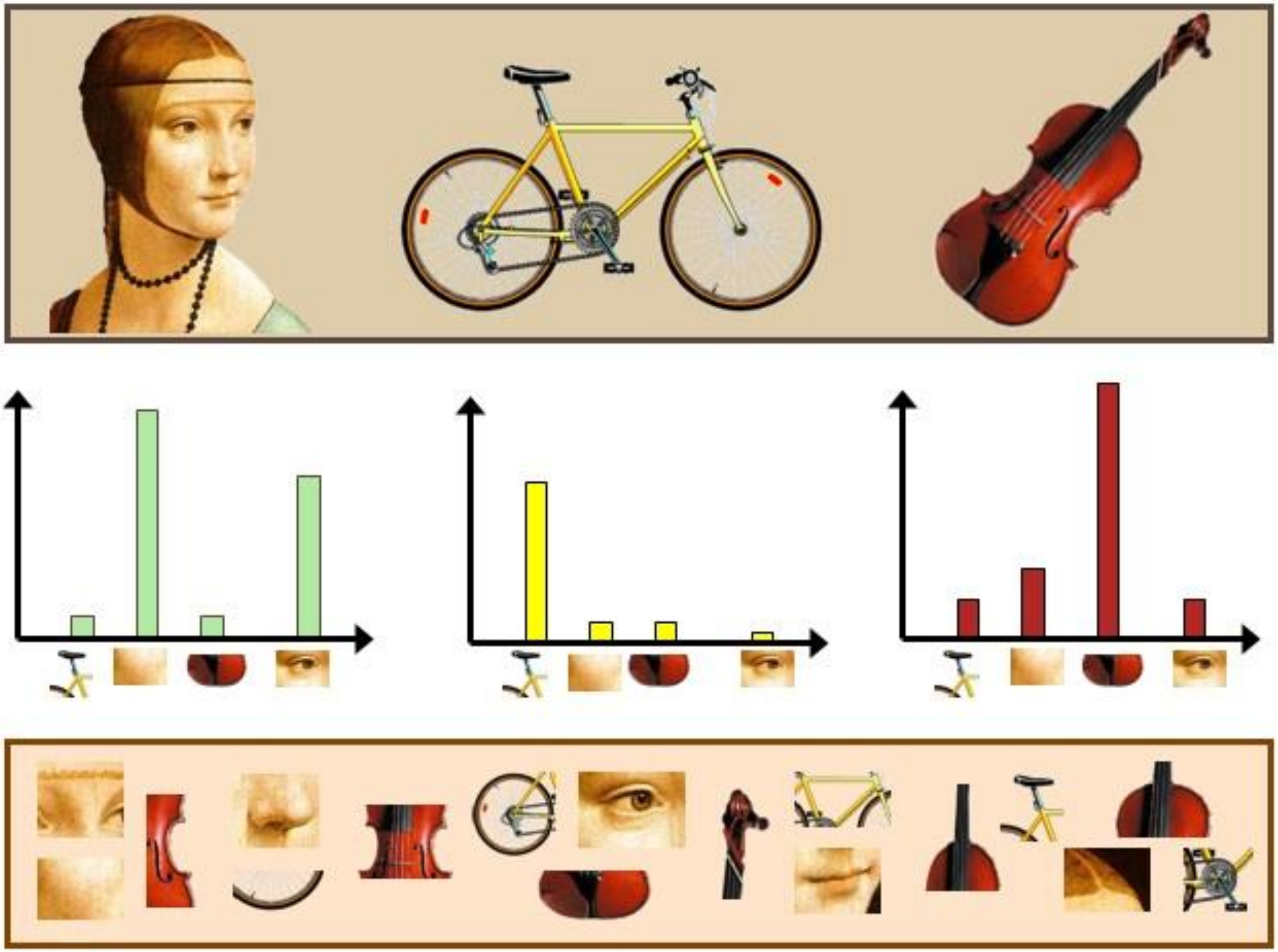


- $\boldsymbol{w} \cdot \boldsymbol{x_i} + b \geq 1$: positive $\boldsymbol{x_i}$
- $\boldsymbol{w} \cdot \boldsymbol{x_i} + b \leq -1$: negative $\boldsymbol{x_i}$
- $\boldsymbol{w} \cdot \boldsymbol{x_i} + b = 1$ or -1 : support vectors

- Classify a point: $f(\boldsymbol{x}) = \text{sign}(\boldsymbol{w} \cdot \boldsymbol{x} + b)$

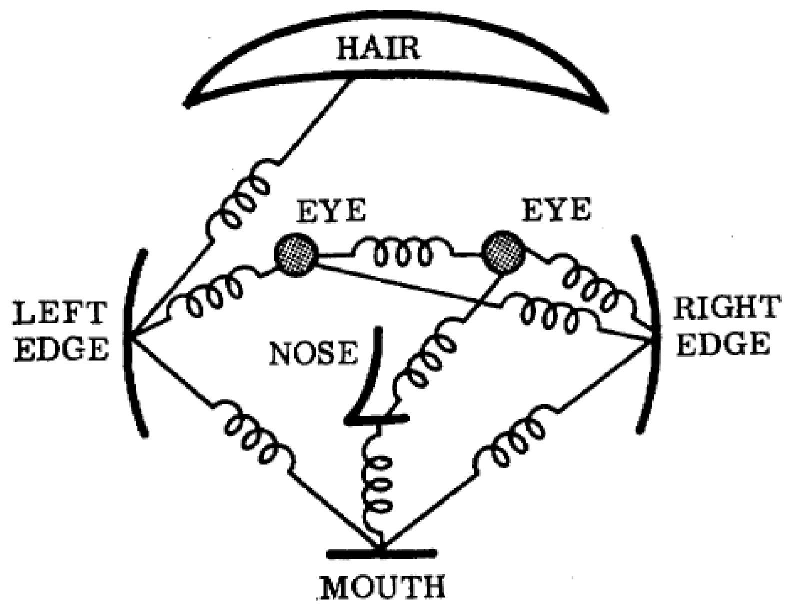
- For multiclass classification, one against all, one against one, etc.

Bag of Words



Spatial relation

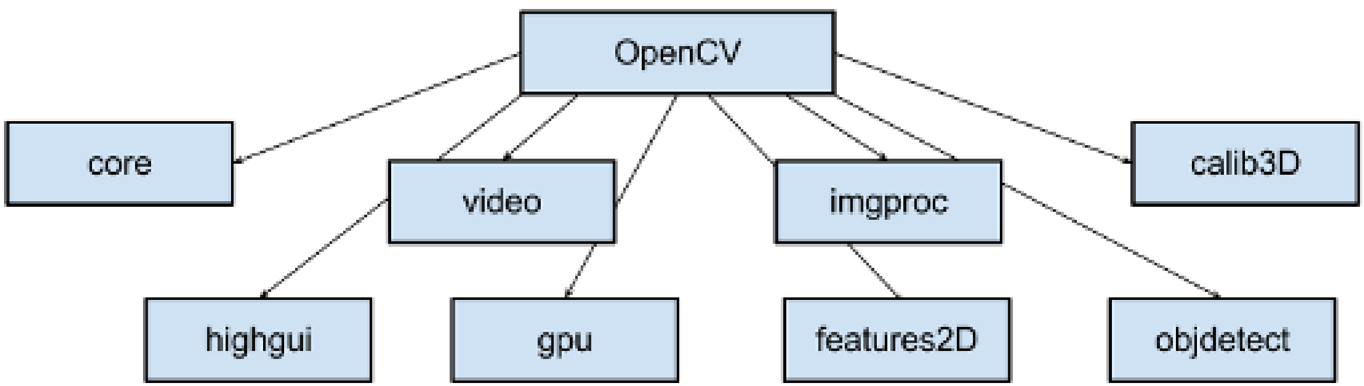
- Appearance
  - How much does a patch of image resemble a known part?
- Spatial relation
  - How well do the parts match the expected shape?



Fischler and Elschlager, IEEE Trans. on Comp., 1973

OpenCV Library

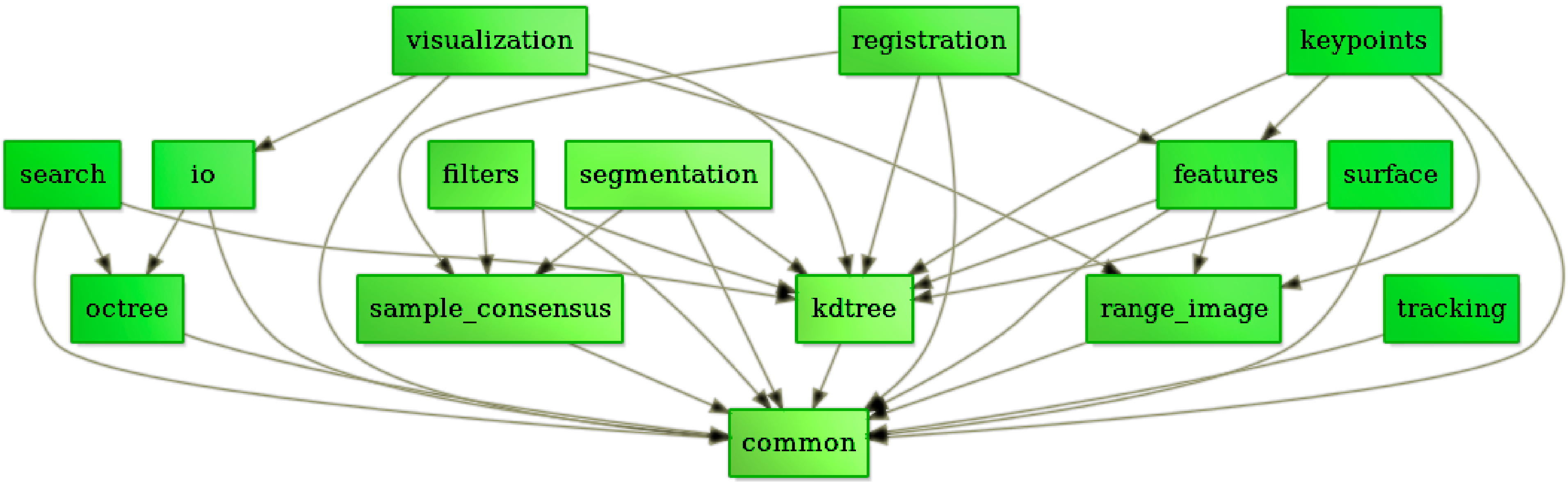
- Open Source Computer Vision Library for image and video processing
- The library has more than 2500 optimized algorithms, including both classic and state-of-the-art computer vision and machine learning algorithms



http://www.opencv.org

Point Cloud Library (PCL)

- Library for 2D/3D image and point cloud processing
- Contains numerous state-of-the art algorithms including filtering, feature estimation, surface reconstruction, registration, model fitting and segmentation



http://www.pointclouds.org