Image Processing and Computer Graphics

Image Processing

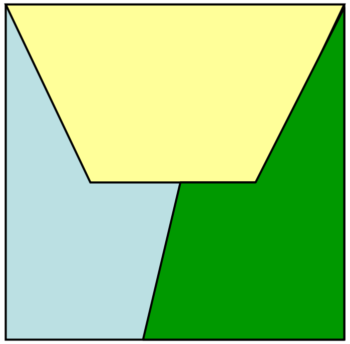
Class 7

Segmentation and Grouping

What is image segmentation?

- Partitioning of the image domain Ω into several (usually disjoint) regions Ω_i

$$\Omega = \bigcup_i \Omega_i \qquad\qquad \Omega_i \cap \Omega_j = \emptyset \quad \forall i \neq j$$

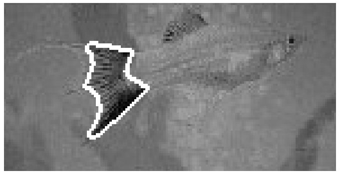
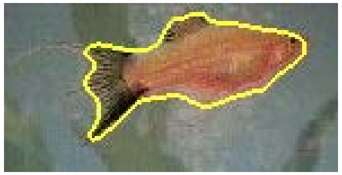


- Frequent special case: two-region segmentation (foreground-background)

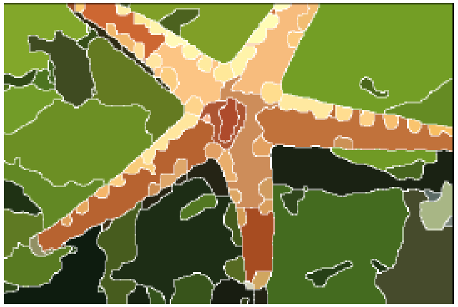
- Important difference to **edge detection**: requires closed contours

- Edge detection is only one part of the solution as it provides pieces of potential contours.

What makes a segmentation a good segmentation?

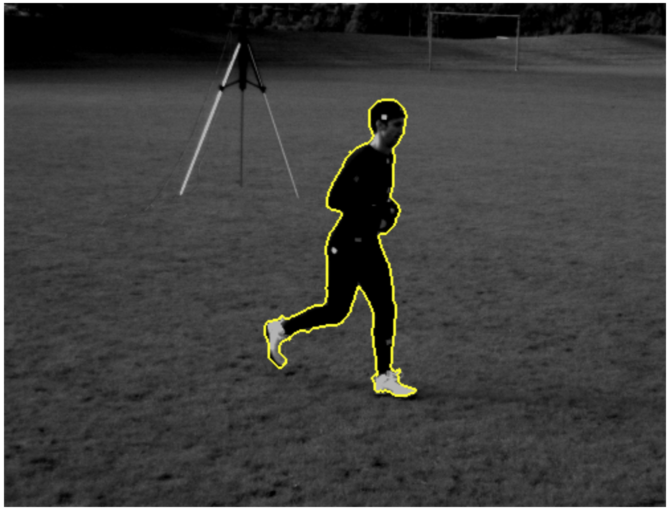
- There are exponentially many possibilities to partition an image.
- Ideally, we wish a hierarchical decomposition of a scene in its objects and their parts → **object segmentation**
- This is impossible from static images without prior knowledge on the appearance of objects
- But image segmentation can also be seen just as the grouping process that combines pixels with similar appearance to regions → **superpixels**



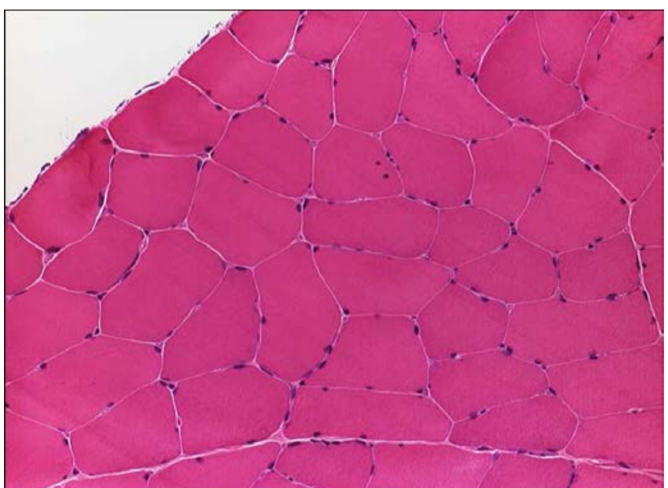
Arbelaez et al. CVPR 09

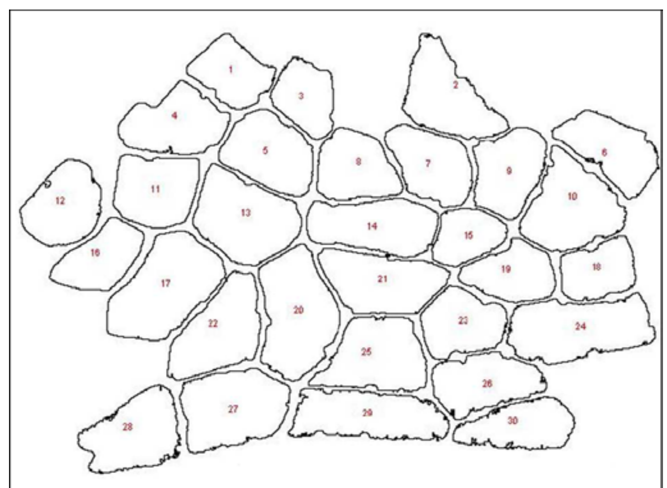
Segmentation with an underlying application

- Track the shape of a human body (Brox et al. 2007)



- Find the rims of muscle fibers (Kim et al. 2007)

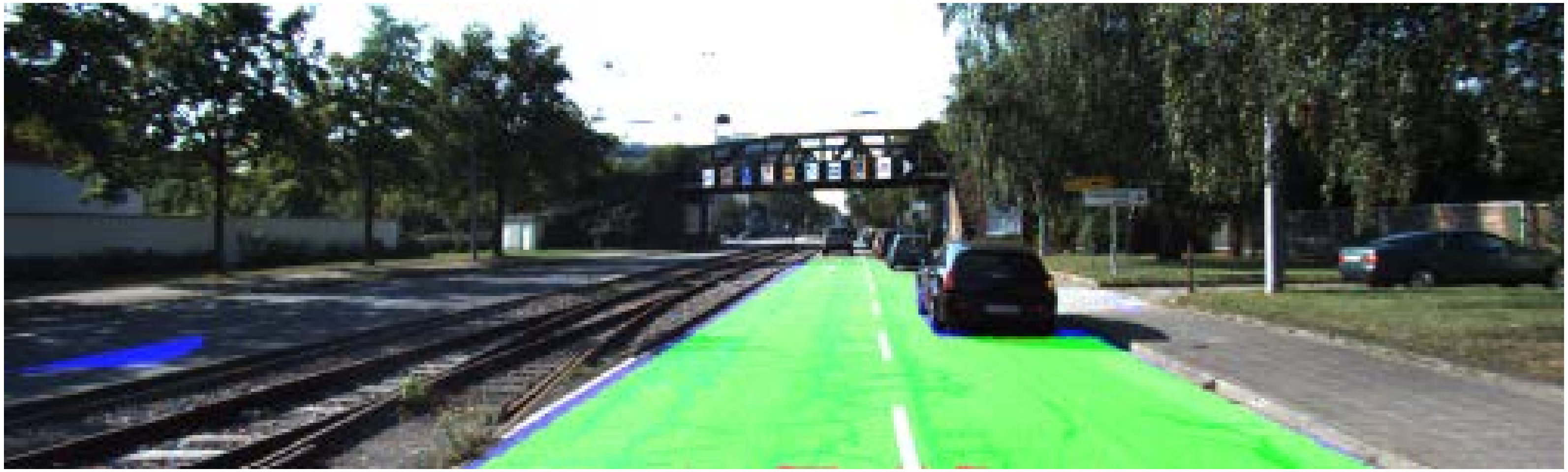




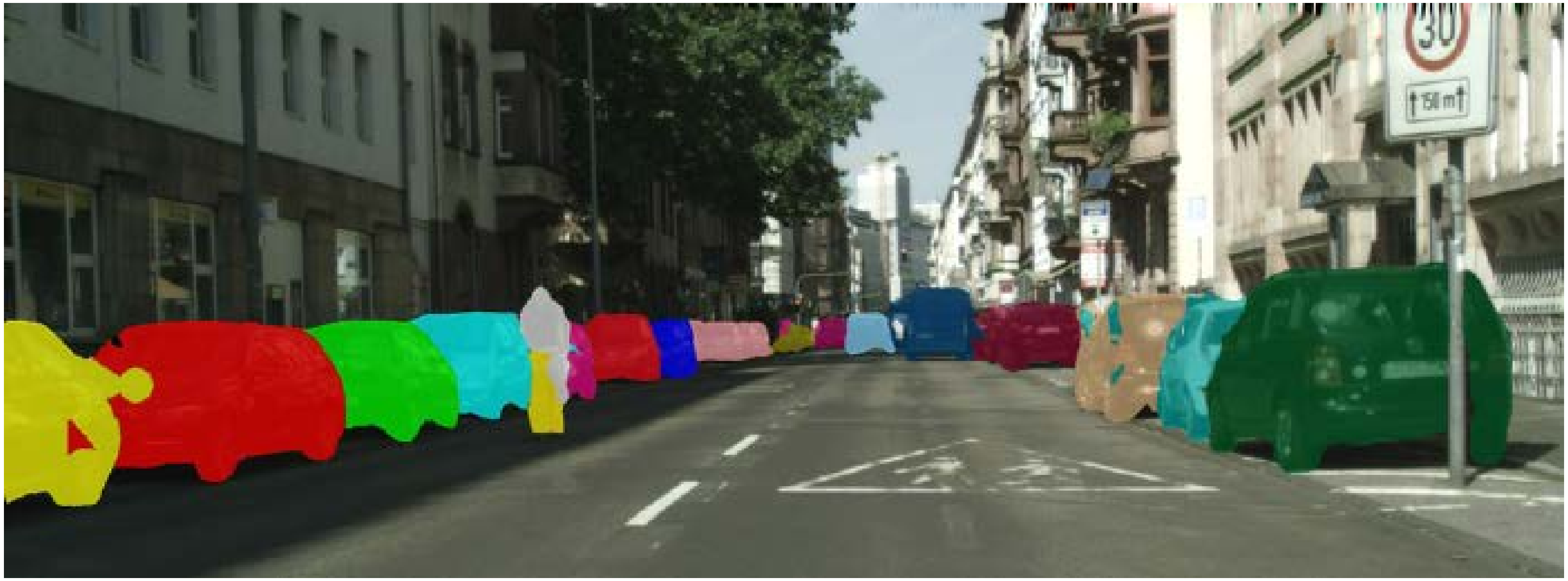
Object class segmentation (semantic segmentation)

Example from Carreira et al. 2014 (the pre-deep-learning era)

Semantic and instance segmentation as learning tasks



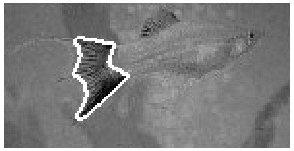
Road segmentation (Oliveira et al. 2016)

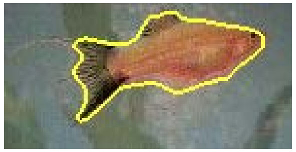


Instance segmentation (Uhrig et al. 2016)

Feature space

- Various features can be used to distinguish one region from another
  - Intensity
  - Color
  - Texture
  - Motion in videos
  - Disparity in stereo images
  - Depth in depth cameras



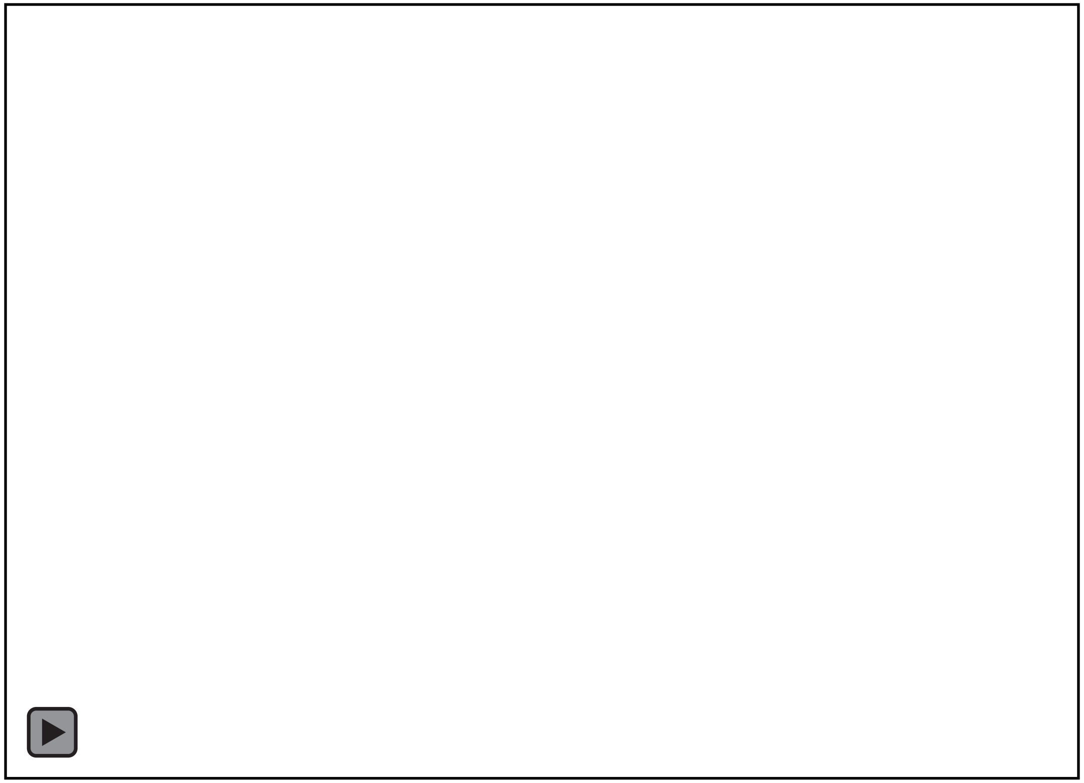


- We can distinguish **first-order features** and **second-order features**.

- First-order features are provided directly by the sensor, while second-order features must be derived from first-order data (texture, motion, disparity).

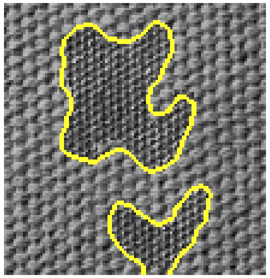
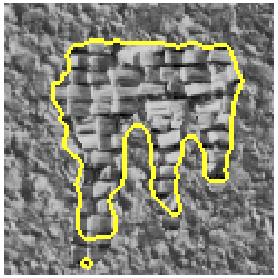
- Second-order features are usually not precisely localized (texture) and/or are not densely available with full confidence (motion, disparity)

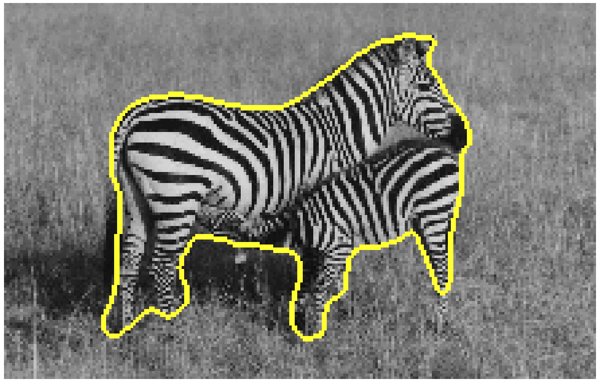
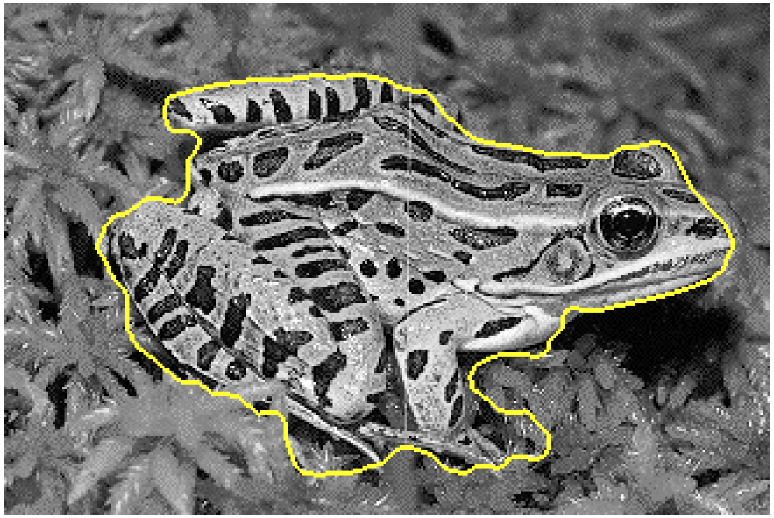
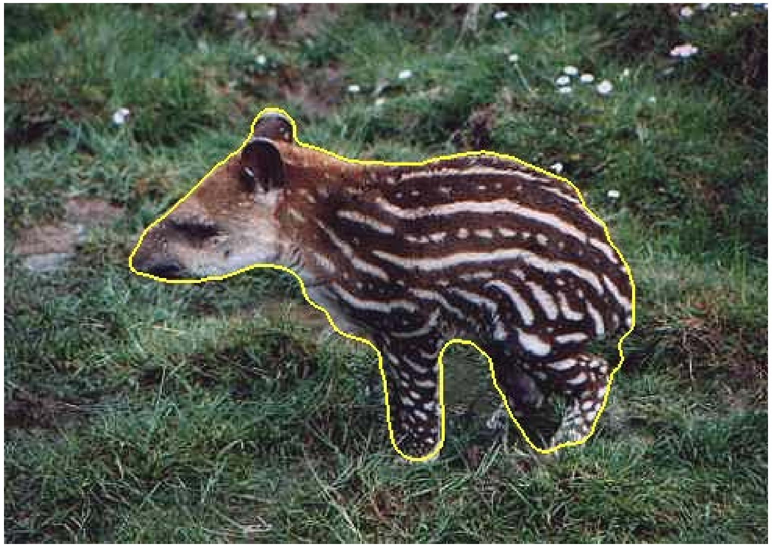
Features for segmentation: intensity/color



Author: Mikaël Rousson

Features for segmentation: texture

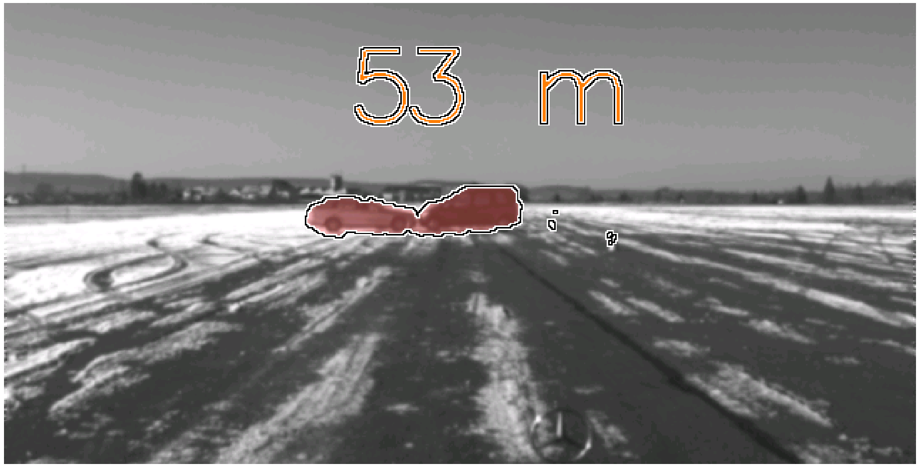




Features for segmentation: motion

Author: Daniel Cremers

Author: Thomas Brox

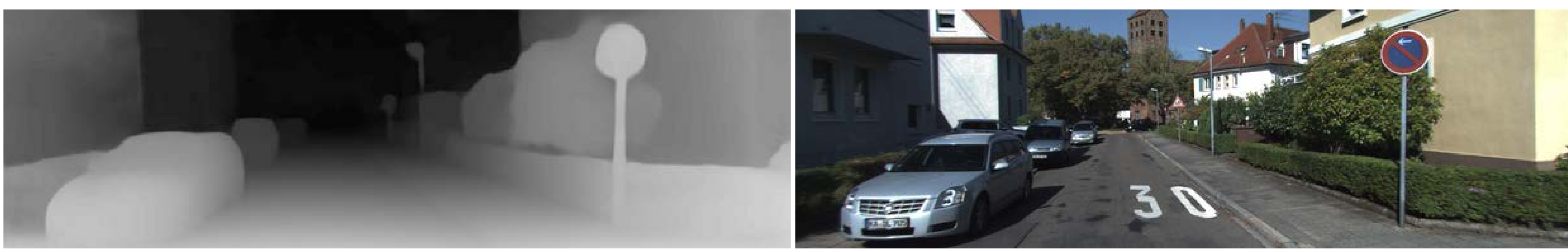
Features for segmentation: motion



Author: Andreas Wedel

Depth images

Depth and color image from the Microsoft Kinect camera



Depth estimate and color image from a stereo camera

Unsupervised vs. supervised methods

- Given the feature space, segmentation methods can be classified into unsupervised techniques (without training data) and supervised techniques (with training data)

- Machine learning counterparts:
  - Unsupervised techniques correspond to clustering
  - Supervised techniques correspond to classification

- Edge-based techniques employ an edge indicator or pairwise similarities between pixels. Then they fit closed contours, such that the contour coincides with strong edges (low similarities).

- Region-based techniques employ a statistical feature model or object model of each region. Regions should fit the region model.

Thresholding

- The most simple way to come to a segmentation is thresholding.
- Converts an intensity image into a binary image:

$$u(x,y) = \begin{cases} 255 & I(x,y) > \theta \\ 0 & I(x,y) \leq \theta \end{cases}$$

- The two states of the binary image assign pixels to the two regions.
- Can be generalized to N regions by introducing $N-1$ thresholds.
- Problems:
  - Often there is no threshold that separates the objects.
  - Point-based operation: the spatial context is ignored.
- Thresholding is a region-based technique with a very simple (and inflexible) region model.

Relationship to clustering

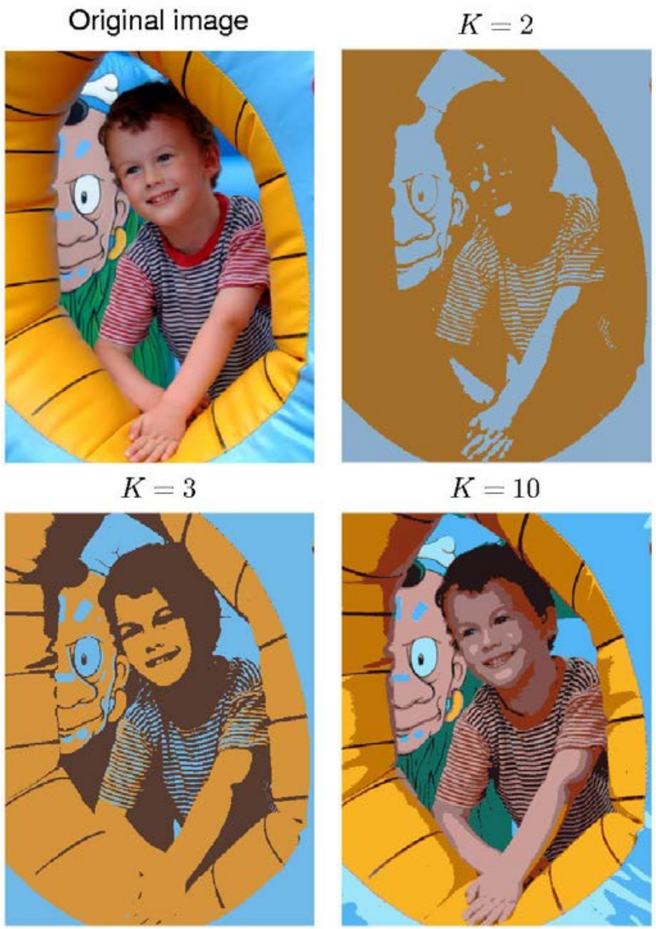
- Image segmentation is similar to clustering: assigning data points (pixels) to clusters (regions)

- There are various clustering methods.

- Most popular: **k-means clustering**
  - Initialize the pixels to belong to a random region
  - Compute the mean feature vector in each region
  - Move a pixel to another region if this decreases the total distance

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2$$

  - Iterate until pixels do not move any longer
  - With $K = 2$, k-means is like thresholding with an automatically determined threshold.

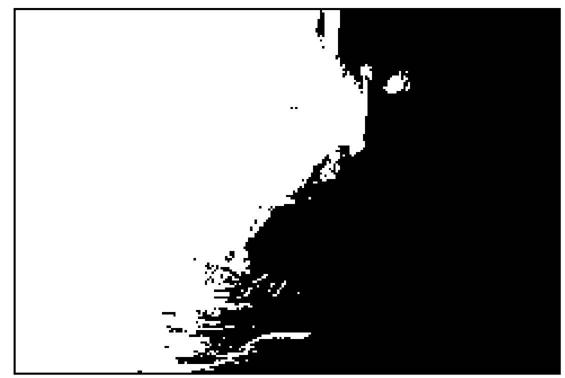


Author: Christopher Bishop

Clustering does not enforce spatial compactness of regions

- Clustering methods ignore spatial context. Only the feature vector determines the assignment of a pixel, not its position in the image.
- We can add the pixel coordinates to the feature vector to enforce compact regions, but this is not equivalent to enforcing smooth region contours.

Intensity+color



Intensity+color+position

Greedy heuristics: region growing and region merging

- **Region growing**
  - **Seed points** represent initial regions
  - For each point on the region boundary: if a neighbor is similar enough, it is assigned to this region (the region grows)
  - Grow until there are no more similar pixels along the region boundary

- **Region merging**
  - Initially all pixels represent their own region.
  - The two most similar regions are successively merged to one larger region.
  - Repeat until a similarity threshold or a given number of regions is reached.

- Some dissimilarity criteria:
  - Euclidean distance of the features' means: $d^2(\mathcal{R}_1, \mathcal{R}_2) = (\mu_1 - \mu_2)^2$
  - Mean Euclidean distance along common boundary
  - (…)

Region merging

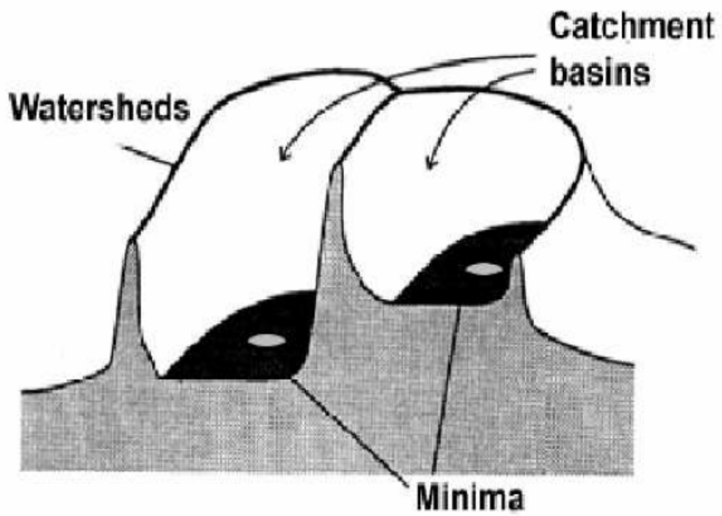
Author: unknown

Watershed segmentation

- Illustrative description: regard the **image gradient magnitude** as mountains and let it rain.

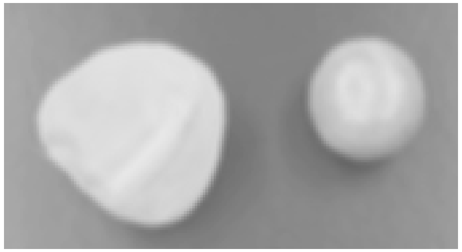
- Water flows downhill and gathers in catchment basins: the regions.

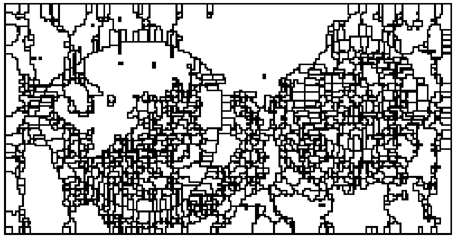
- Regions meet at the watersheds (edges), → region boundaries.



Author: P. Soille

- Tiny gradient fluctuations result in over-segmentations → can be reduced by presmoothing the image





Energy minimization: the snakes model

- Kass, Witkin, and Terzopoulos proposed the following energy functional:

$$E(C) = -\int_0^1 |\nabla I(C(s))|^2 ds + \alpha \int_0^1 |C_s(s)|^2 ds$$

- $C : [0,1] \to \Omega$ is a parametric contour. C_s denotes the first derivative of this contour.

- The first term is also called **external energy**, since it depends on the (external) input image. The second term is called **internal energy**, since it is inherent to the model and independent of the data.

- Minimizing the external energy drives the contour to follow maxima of the gradient.

- Rather than seeking such maxima and to group them to a contour, we consider all possible contours and choose the one that best captures the maxima.

Energy minimization: the snakes model

- The external energy alone is not sufficient. It would lead to a fractal contour with infinite length.

$$E_{ext}(C) = -\int_0^1 |\nabla I(C(s))|^2 ds$$

- This is avoided by the internal energy, which penalizes the length of the contour

$$E_{int}(C) = \alpha \int_0^1 |C_s(s)|^2 ds$$

- The external and internal energy together prefer a compromise of a short contour which captures as much image gradient as possible.

- Energy minimization with the calculus of variation and gradient descent → local minima

Implicit representation of contours

- Introduce an **indicator function** $\phi : \Omega \to [-1, 1]$
- The zero-level line represents the contour

$$C = \{\mathbf{x} \in \Omega |\ \phi(\mathbf{x}) = 0\}$$

- For evolving C evolve ϕ
- Allows for topological changes
- Can be applied in any dimension
- Represents the contour <u>and</u> the enclosed region

Author: Daniel Cremers

Region-based active contours

- Energy minimization based on regions statistics

- The energy states the optimal separation of pixel intensities:

$$E(C) = \int_{\Omega_1} (I - u_1)^2 \mathbf{dx} + \int_{\Omega_2} (I - u_2)^2 \mathbf{dx} + \nu |C|$$

- This is similar to k-means clustering (two-means), <u>but with an additional constraint on the length of the separating contour</u>.

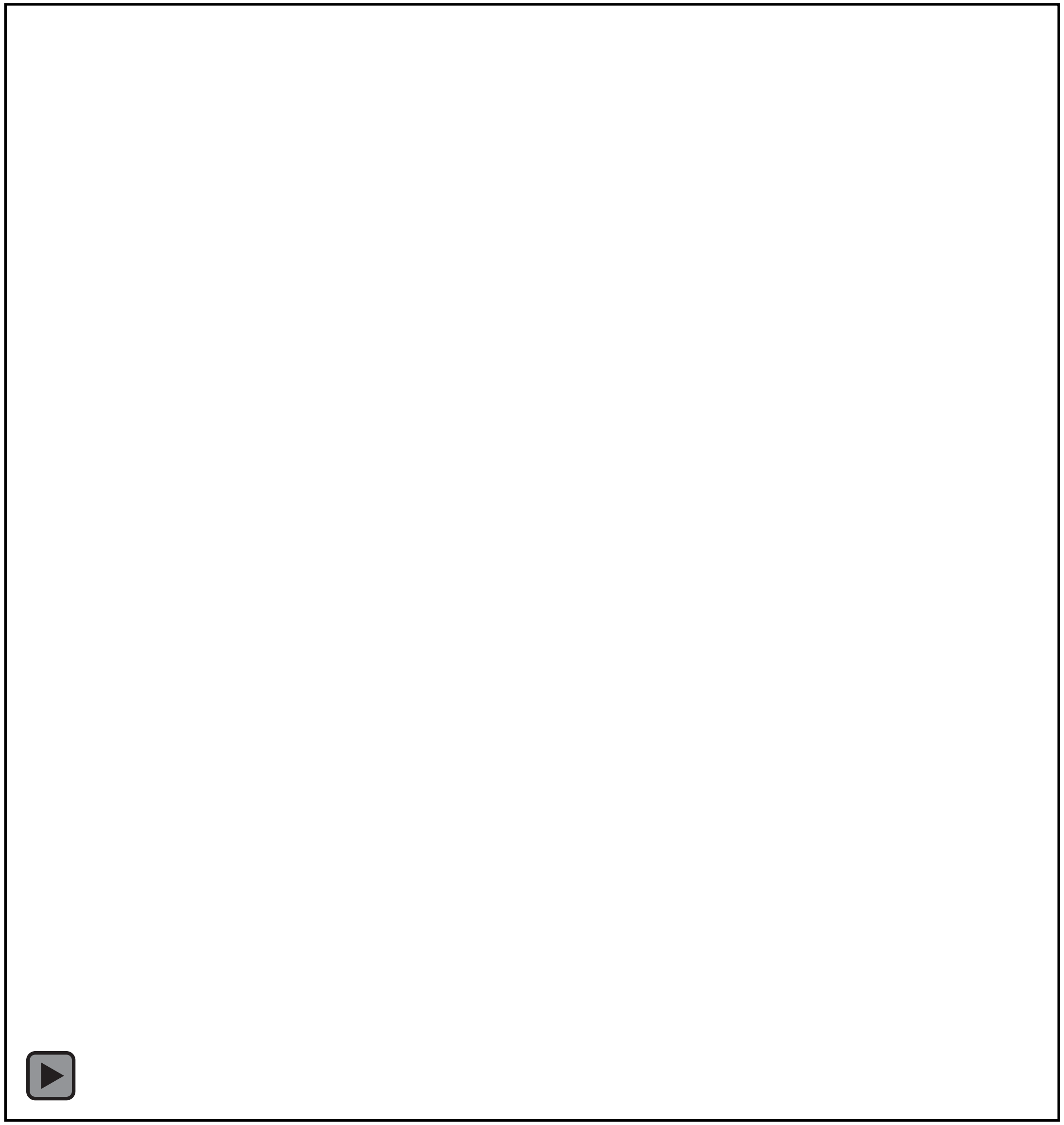
- We can express this using an implicit representation of the contour:

$$E(\phi) = \int \phi \left((I - u_1)^2 - (I - u_2)^2 \right) + \nu |\nabla \phi| \, \mathbf{dx}$$

- Given ϕ, u_1 and u_2 can be found analytically as the mean intensities inside the two regions

$$u_1 = \frac{\int H(\phi) I \, \mathbf{dx}}{\int H(\phi) \, \mathbf{dx}} \qquad u_2 = \frac{\int (1 - H(\phi)) I \, \mathbf{dx}}{\int (1 - H(\phi)) \, \mathbf{dx}}$$

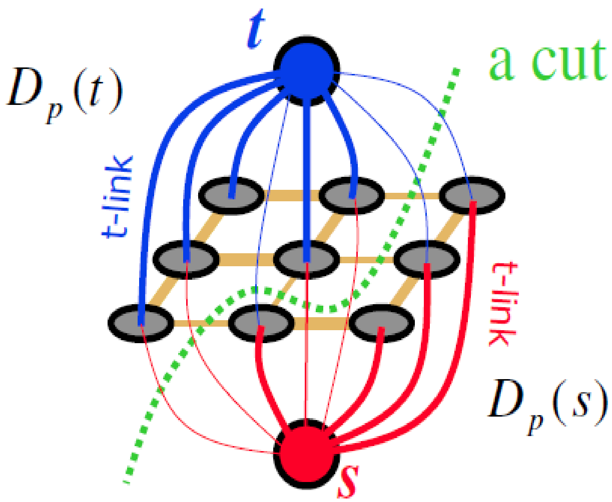
Region-based active contours



Author: Daniel Cremers

Graph cuts for segmentation

- Example of combinatorial optimization



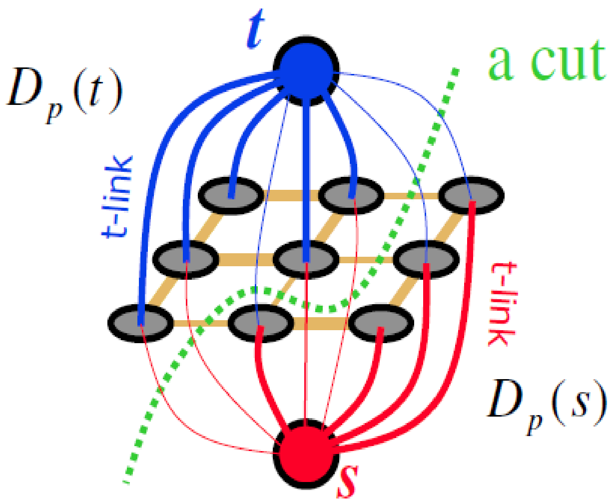
- Graph structure for **min-cut**:
  - Each pixel represented by a **node**. It is interconnected to its neighbors via **edges**.
  - Neighborhoods can have different complexity. Most simple: 4-neighborhood.
  - Two extra nodes (source and target) connected to all pixel nodes

- Goal: find the minimum cut through the graph that separates source and target.

- Source and target nodes correspond to regional models. A pixel is assigned to either of the two regions.

- The connection between neighbors enforces a “regular” labeling.

Energy corresponding to a graph



- Generally the energy reads:

$$E(u) = \sum_i D(u_i) + \sum_{i,j \in \mathcal{N}(i)} w_{ij}\delta(u_i \neq u_j)$$

- First term comprises the links to the source and target nodes, the **t-links**.

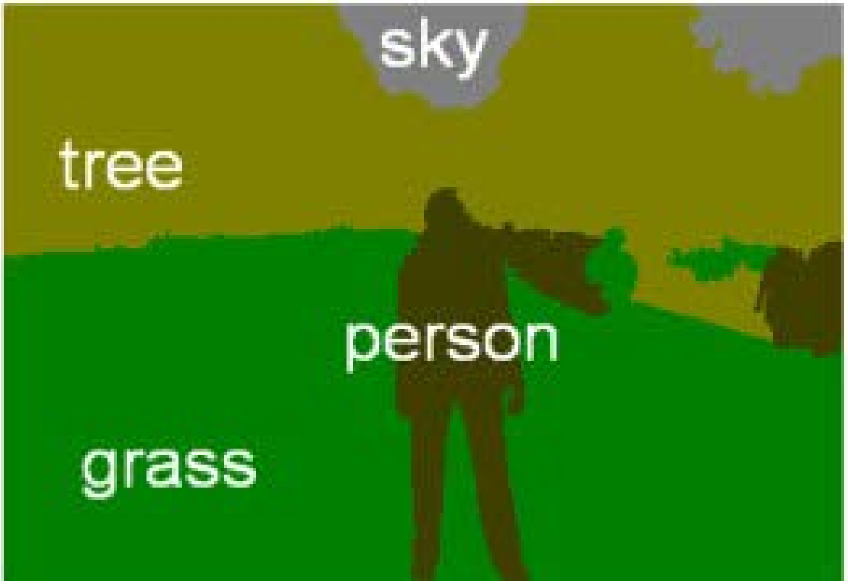
- A simple region model is specified by the mean. This leads to the t-link weights:

$$D_i(0) = (I_i - \mu_1)^2, \quad D_i(1) = (I_i - \mu_2)^2$$

- Second term comprises the **n-links** between neighboring pixels. Usually the weights are fixed, but they can, e.g., depend on the image gradient.

- For fixed means μ_1, μ_2, this combinatorial minimization problem can be solved in polynomial (average case: linear) time.

Semantic segmentation



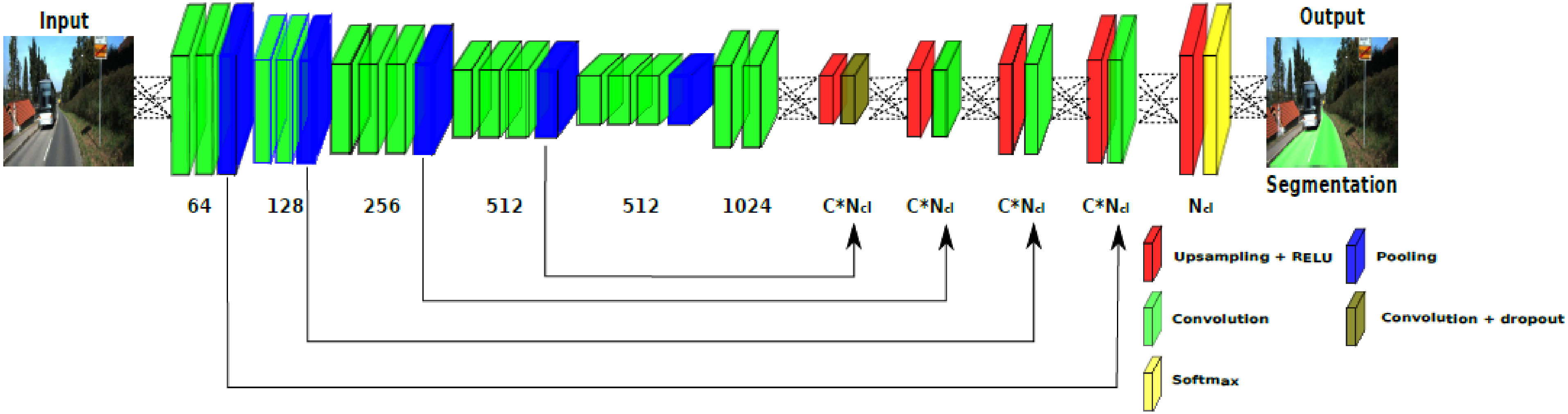
Author: Ladicky et al. 2008

- For each pixel run a classifier, which yields a score $S_i(\mathbf{x})$ for each class i. Here: two classes. In general, multiple classes.

$$E(C) = -\int_{\Omega_1} S_1(\mathbf{x})\mathbf{dx} - \int_{\Omega_2} S_2(\mathbf{x})\mathbf{dx} + \nu|C|$$

- Can be minimized with level sets or graph cuts

Semantic segmentation with deep networks



Oliveira et al. 2016

- Certain network architectures can combine classification and pixel-wise segmentation
  → usage and combination of features is learned
  → both context and detailed features are used

- Sometimes there is a CRF (graph cut or similar) on top

- More in class 10 and in the Computer Vision course

Summary

- The goal of segmentation/grouping depends much on the application.
- We can distinguish image segmentation and object segmentation.
- Object segmentation requires special features (e.g. motion) or top-down knowledge (e.g. shape priors or priors on object appearance → semantic segmentation).
- There are many methods for different scenarios
  - Algorithmic approaches
  - Energy models with contour constraints (contour length, shape prior)
  - Deep learning
- Deep learning based methods are most powerful, but require training data

References

- M. Kass, A. Witkin, D. Terzopoulos: Snakes: active contour models. *International Journal of Computer Vision* 1:321-331, 1988.

- D. Mumford, J. Shah: Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics* 42: 577-685, 1989.

- T. Chan, L. Vese: Active contours without edges. *IEEE Transactions on Image Processing* 10(2):266-277, 2001.

- Y. Boykov, V. Kolmogorov: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision, *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 26(9):1124-1137, 2004*.* Code available: http://pub.ist.ac.at/~vnk/software.html

- L. Ladicky, C. Russell, P. Kohli, P.H. Torr: Graph cut based inference with co-occurrence statistics. *European Conference on Computer Vision*, 2008.

- J. Carreira, R. Caseiro, J. Batista, C. Sminchisescu: Free-Form Region Description with Second-Order Pooling, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014.

- J. Uhrig, M. Cordts, U. Franke, T. Brox: Pixel-level encoding and depth layering for instance-level semantic segmentation, *German Conference on Pattern Recognition,* 2016.

- G. Oliveira, W. Burgard, T. Brox: Efficient deep methods for monocular road segmentation, *International Conference on Intelligent Robots and Systems (IROS)*, 2016.