Key concepts

Concepts that are central to our hypothesis:

- Segments: maximally large but roughly homogeneous
- Homogeneity: locally similar statistics (smooth change allowed)
- Detail / Background decomposition
- Structure / Texture organization of the detail

In Computer Vision

Object saliency: propose a segment of the image as a possible object.
Purpose: process a limited number of image regions (typically 150 to 4000 per image).

- Region proposals: CPMC; RIGOR (CVPR 2014)
- Bounding boxes proposals: Selective Search; Edge Boxes by P. Dollar (ECCV 2014). Useful for Deep Networks.

General regions proposal: Outputs regions for "stuff" categories like sky, grass, etc.

- Candidate Regions (ECCV 2014)

Linking bottom-up saliency and object saliency:

- Xiaodi Hou et al (CVPR 2014)

Outline

Introduction

- **Segmentation** simplifies the representation into a structure that should be more meaningful and easier to analyze.
- **Segments** are regions where the image is roughly homogeneous (like superpixels). They may correspond to image structures like sky, sheep, person, a human face.
- As mid-level computer vision, segmentation may be performed without object-specific knowledge.

Region-based Image Segmentation

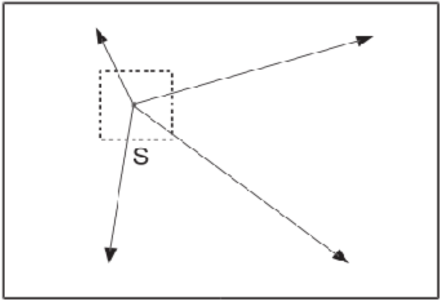
- Edges vs regions
- Segmentation and clustering
- Dividing vs merging
- Greedy algorithms
- How could segmentation be wired in the visual cortex?

Outline

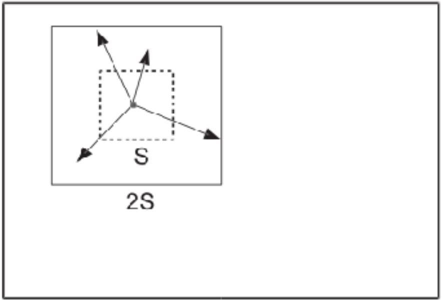
UCLA

SLIC superpixels (R. Achanta et al., PAMI 2012)

- K-means clusters on the 5-dimensional [*labxy*] space.



(*a*) standard *k*-means searches the entire image



(*b*) SLIC searches a limited region

SLIC

Instead of using just an Euclidean distance in the 5D space, a new distance measure D_s is defined in order to control how important is the spatial position x, y with respect to l, a, b.

$$
\begin{aligned}
d_{lab} &= \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2} && (1) \\
d_{xy} &= \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} && (2) \\
D_s &= d_{lab} + \frac{m}{S} d_{xy} && (3)
\end{aligned}
$$

The greater the value of m, the more spatial proximity is emphasized and the more compact the cluster.

SLIC

Algorithm

- Initialize K clusters in grid positions
- Move K clusters to lowest gradient positions (3 pixels vicinity)
  $G(x, y) = ||\mathbf{I}(x+1, y) - \mathbf{I}(x-1, y)||^2 + ||\mathbf{I}(x, y+1) - \mathbf{I}(x, y-1)||^2$
- Assign each pixel to a cluster center (limited to a 2S vicinity).
- Recalculate the centers as the average *labxy* vector of all the pixels belonging to each cluster
- Iterate until convergence
- Fix disconnected segments

Produces an oversegmentation of the image, good starting point for grouping segments.

SLIC

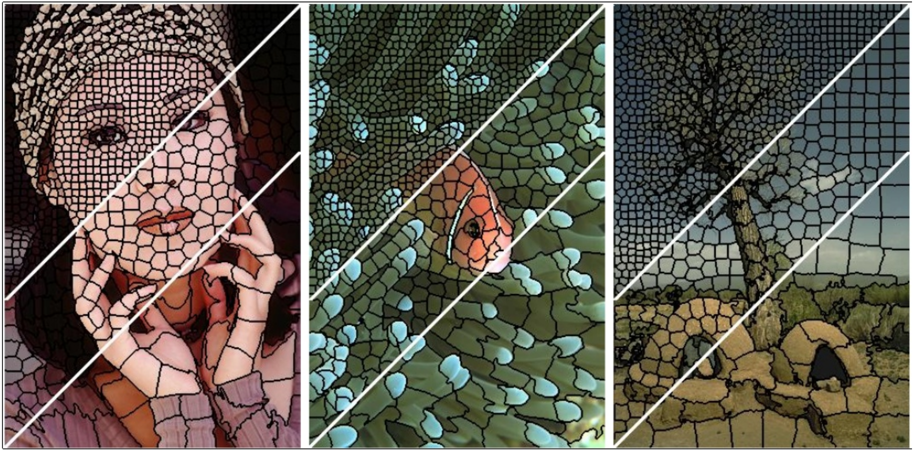
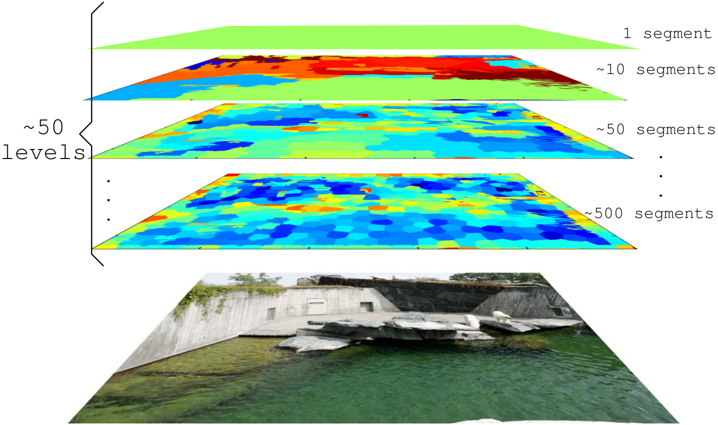


Image segmented into SLIC superpixels of (approximate) size 64, 256, and 1024 pixels. (R. Achanta et al., PAMI 2012)

Outline

Hierarchical grouping



Multiple segmentation levels in a hierarchy. Segments with a good coverage of objects or parts may happen at different levels.

Appearance vector

- The algorithm starts with the basic set of elementary segments $\mathcal{S}_1$ which are produced by SLIC.
- Segmentation hierarchy $\mathcal{S}_{\mathcal{H}} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \cdots \cup \mathcal{S}_L$ of L levels by merging segments as follows.
- Each segment is described by an *appearance vector*

$$V = (\overline{\mu}, \overline{\sigma}, c_x, c_y, w, h), \tag{4}$$

where $\overline{\mu}, \overline{\sigma}$ are the mean and the standard deviation of the Lab color space components and the first and second gradients, $(l, a, b, \nabla_x, \nabla_y, \nabla_x^2, \nabla_y^2)$. Here (c_x, c_y, w, h) are the centroid of the segment and the dimensions of its bounding box.

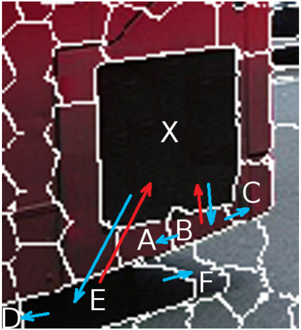
- These appearance vectors are designed so that they can be computed efficiently recursively for new segments composed by merging.

Asymmetric dissimilarity

- Asymmetric dissimilarity function $\Delta^A_{i|j}$ between segments which are 1^{st} or 2^{nd}-order neighbors.
- Appearance term $\Delta^A_{i|j} = ||V_i - V_{i \bigcup j}||_2$
- Modified by an edge-term ($E_{i,j} \in [0, 1]$) that represents the amount of edge-ness on the boundary between two adjacent regions.

$$\begin{aligned} \Delta_{i|j} &= E_{i,j} + \Delta^A_{i|j}, \text{ if } i, j \text{ are } 1\text{−neighbors}, \\ \Delta_{i|j} &= 1 + \Delta^A_{i|j}, \qquad \text{if } i, j \text{ are } 2\text{−neighbors}. \end{aligned} \tag{5}$$

Asymmetric dissimilarity



Asymmetric dissimilarity function. The nodes that are less dissimilar to X are B and E (second neighbor) because they don't modify a lot the statistics of X. However, X modifies a lot the statistics of B and E. Both of these nodes have other nodes to which they are much more similar.

Algorithm

Merging at each level

- Only a fraction (30%) of the segments are merged (top-ranked segments)
- A segment is merged to its least dissimilar 1st or 2nd neighbor: $\arg\min_j \Delta_{i|j},\ j \in 2NN(j)$
- Unless they violate the condition $\Delta_{i|j} > 0.9\Delta_{j|i}$

For each level, evaluate $\Delta_{i|j}$ for the new segments, rank them, and merge the top ranked.

Texture handling

No explicit texture handling mechanism. Fails with coarse textures.

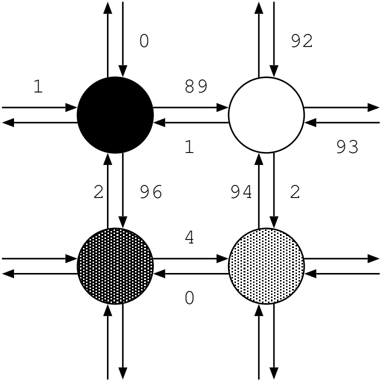
Left: example of vegetation textures captured by a single segment. Right: a coarser texture which failed to be merged in a single segment.

Outline

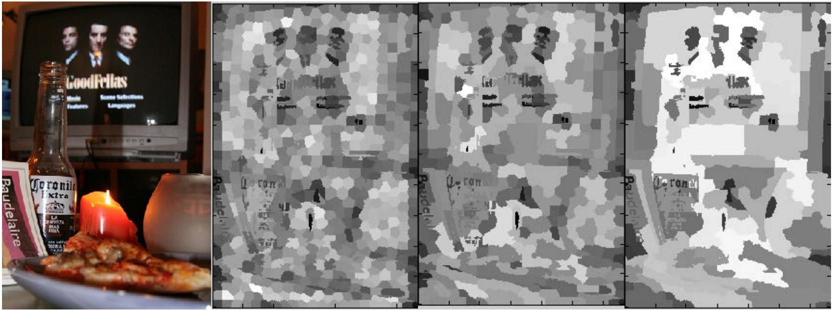
Segments ranking

- Greedy algorithm: needs a mechanism to help capturing *global* properties.
- Start by merging those segments which have similar statistics to their neighbors: homogeneous regions.
- Some segments will grow "faster" across levels.
- To go beyond examining 1st neighbors similarity, use a graph-based approach.
- PageRank (Franceschet, ACM 2011) is suitable for directed graphs and captures global properties.
  - Quantifies the importance of each segment (i.e. graph node) after a sequence of probabilistic transitions over the graph.
  - Probabilistic transitions encoded by a stochastic matrix.
  - Dissimilarity has to be inverted to similarity in the matrix.

PageRank



The PageRank algorithm prioritizes merging segments which are very similar to their neighbors. In this example, the white node (representing a segment) has highest ranking and so is selected first for merging. The edge weights denote (asymmetric) similarity between the segments.



PageRank: at each level we merge first the highest rank (whitest) nodes. Note that the "salient" or different regions have lower rank, while the segments covering homogeneous areas tend to have higher ranks.

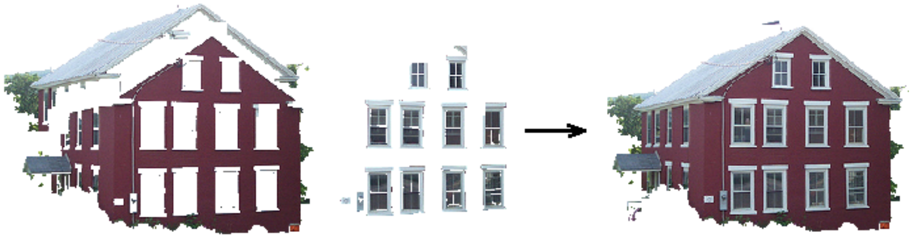
Highlights

- Decisions are taken by local properties ($\Delta_{i|j}$)
- Where decisions are taken is guided by global properties (PageRank).
- Segments can be disconnected (avoids forcing merges between too different segments).
  - w,h,x,y contribute to compactness
  - Compactness can be additionally encouraged in higher levels

UCLA

Outline

Structures at different levels



Candidate regions at different scales. The walls of the house (left panel). Each of the windows (center panel). The whole house (right panel).

Single partition

- Single partition with segments from different levels
- Desirable properties of the segments
  - As big as possible while
  - roughly homogeneous. (Not all objects are roughly homogeneous and this property implies that some objects will be composed by several segments)

Segments A and B are not homogeneous while C is roughly homogeneous because there are few changes between the statistics of the elementary segments that compose it.

Homogeneity criterion

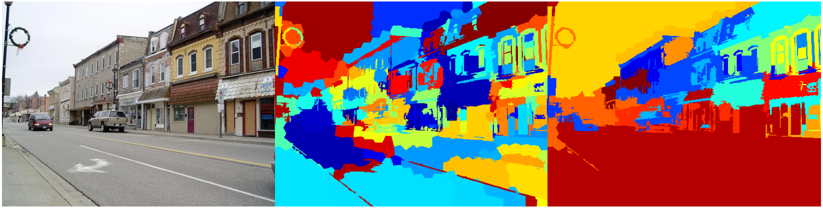
- Appearance vector is $V = (\overline{\mu}, \overline{\sigma})$, where $\overline{\mu}, \overline{\sigma}$ are the mean and the standard deviation of $(I, a, b, \nabla_x, \nabla_y, \nabla^2_x, \nabla^2_y)$ (but the x, y, w, h are not included).
- A maximum difference threshold t between all pairs of segments $||V_i - V_j|| < t, \forall i, j \in S$ is too restrictive and disallows big segments like sky, which may present gradual changes.
- A less restrictive criterion is to threshold the differences between 1st and 2nd-order neighbors within the segment:

$$||V_i - V_j|| < t,\ \forall i, j \in S,\ d_G(i, j) \leq 2, \tag{6}$$

  where $d_G(i, j)$ is the graph distance between i, j and it has to be 1 or 2, that is, 1st or 2nd neighbors.
- Note that this criterion imposes a smooth variation in the statistics of a segment, which makes it possible to describe the segment with linear or polynomial approximations.

UCLA

Single partition



Two partitions resulting from two different threshold values.

UCLA

Outline

Outline

Polynomial approximation

- Each segment is roughly homogeneous, allowing smooth changes
- We represent the appearance of each segment by polynomial models
- Polynomials of orders 0 (constant), 1 (linear), 2 and 3 are chosen based on the error. For similar error, simpler is preferred.
- We refer to this color model as “Background”

Original; polynomial reconstruction; order of polynomial. Dark-blue: 0, light: 1, yellow: 2; red: 3.

Polynomial approximation examples

Outline

1. Bottom-up attention
2. Segmentation by grouping
   - Simple Linear Iterative Clustering
   - Hierarchical grouping algorithm
   - Segments Ranking by PageRank
   - Homogeneity criterion
3. Simple appearance model
   - Polynomial approximation (Background)
   - Residual (Detail)
4. Detail: structure and texture
5. Saliency in complex scenes

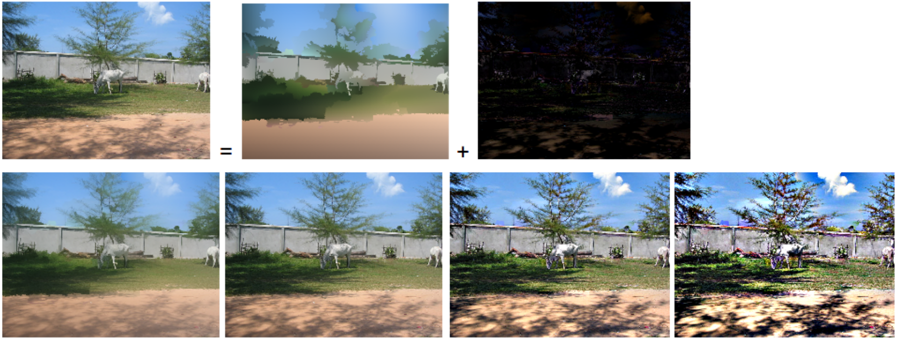
UCLA

Detail

The residual of the segment-wise model is the "detail", or what we can't model by simple polynomial models.

Example: enhancing contrast

We can add back more or less detail to the background.

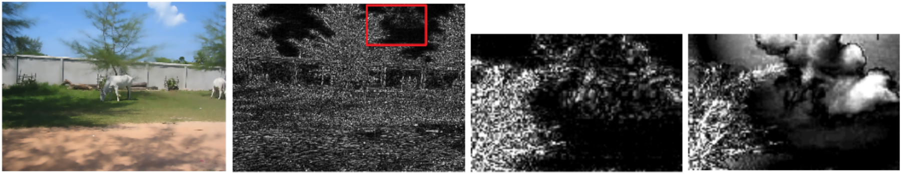


Example: enhancing contrast

We can add back more or less detail to the background. (Artifacts due to segmentation).

Non-local method

Many alternatives smooth locally the image, like Bilateral Filtering. This is different and fails to capture bigger details.



Bilateral filtered image; residual (detail) of the filtering; zoom-in; zoom-in of our detail yielded by the homogeneous segments.

Outline

Detail

decomposed into background and detail:

Different problem

Not to be confused with intrinsic image decomposition:

Is human vision doing background-detail separation?

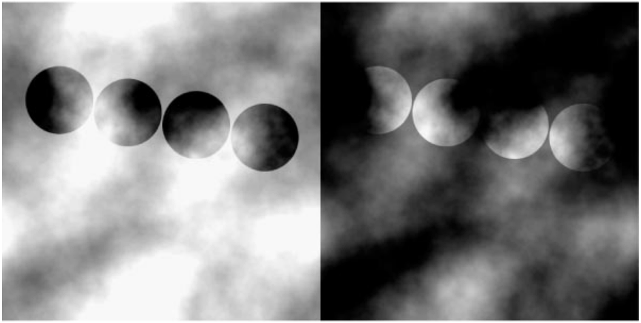


Image by Barton L. Anderson (UNSW) and Jonathan Winawer (MIT). The textured disks on the light and dark surroundings are physically identical. (Consider the cloudy texture as “detail”).

Is the human vision doing background-detail separation?

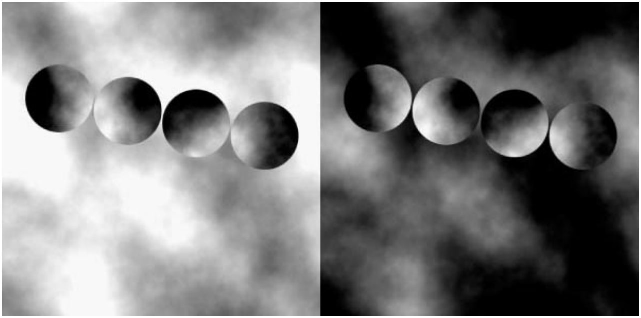
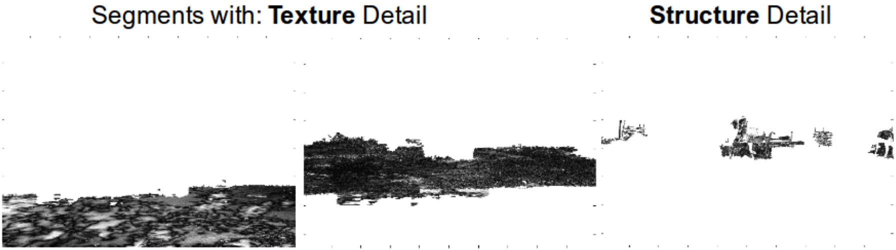


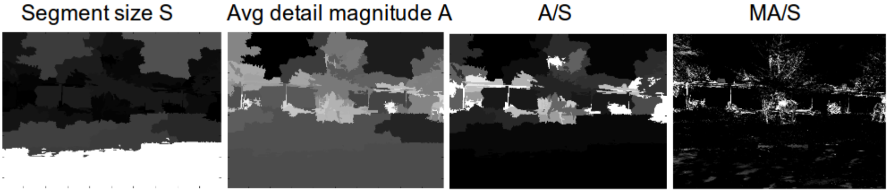
Image by Barton L. Anderson (UNSW) and Jonathan Winawer (MIT). The illusion is broken when the texture of the surroundings is inconsistent with the texture of the disks. The surroundings are simply rotated 90.

Detail can be found in both texture and structure (e.g, the objects).
Scale-dependent.



Detail: texture and structure

The detail in large segments is likely to contain texture. We penalize its saliency.



Detail: texture and structure

We are penalizing texture and respecting detail. Symbolic representation:

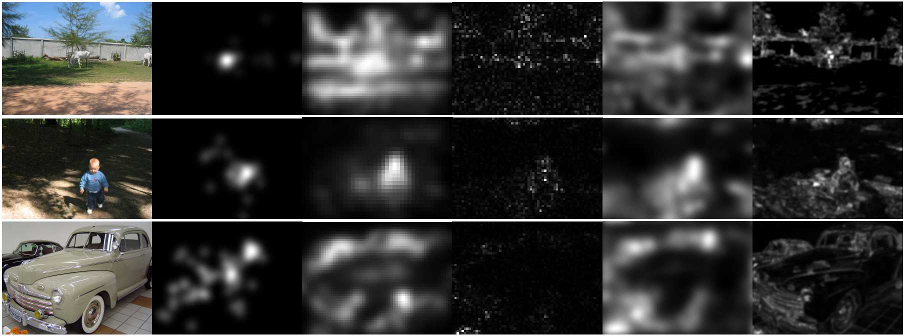
Detail: texture and structure

We are penalizing texture and respecting detail. Symbolic representation:

Detail: texture and structure

We are penalizing texture and respecting detail. Symbolic representation:

Qualitative results



Original; Human; Itti; Spectral signature; AWS; Bonev & Yuille
Human fixations collected on 8 subjects with free-viewing task, first 3 seconds; Eyelink II.

Outline