Saliency: What do you look at in an image?

Students Reading Group
2016
by
Avik Hati

Department of Electrical Engineering
Indian Institute of Technology Bombay

Outline

- Introduction
- Motivation
- Superpixel based approach
- Results & Discussions

Introduction

- What stands out in every image?

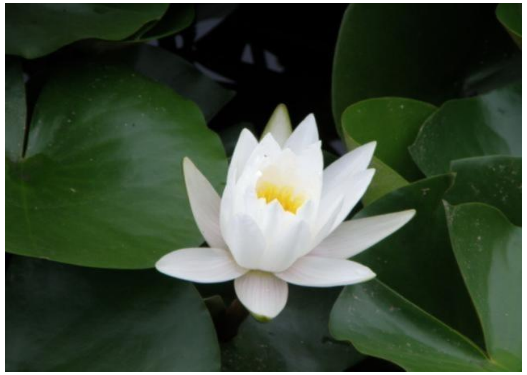
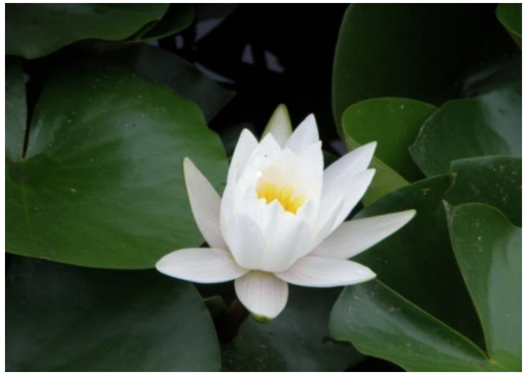



Image courtesy: HFT dataset

Introduction

- What stands out in every image?

Butterfly



Flower

Green apple

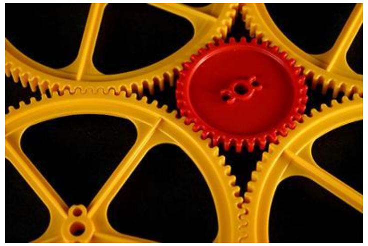
Red apple

Image courtesy: HFT dataset

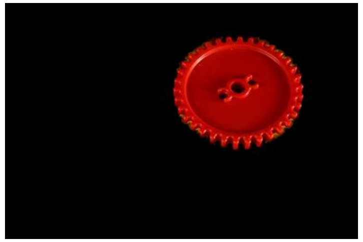
Definition

- Saliency: measure of feature distinctness of regions or objects in images

- Salient objects
  - capture attention
  - distinctive in color or spatial features



Input image



Salient object

Image courtesy: MSRA dataset

Saliency map

- Saliency map: a saliency value assigned to each pixel

Input image

Saliency map

Image courtesy: MSRA dataset

Saliency map

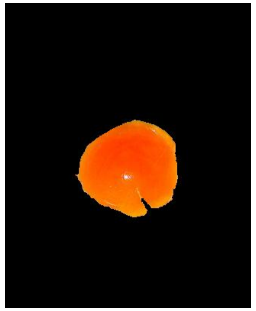
- Saliency map: a saliency value assigned to each pixel

Input image

Saliency map

- Salient objects: saliency map is thresholded to obtain salient objects

Thresholded saliency (binary)



Salient object

Result from Zhu *et al.*, CVPR 2014

Image courtesy: MSRA dataset

Motivation

- Applications
  - Object segmentation
  - Video summarization
  - Region-of-interest image compression
  - Image and video quality assessment
- Large image databases and long videos
- Only the important parts are processed
- Reduction in complexity

Typical approaches

- Local approach (center-surround window)
  - Pixel based method [… initial works …]

- Global approach (global comparison)
  - Superpixel based method [… several works …]
  - Patch based method [… several works …]

- Frequency domain method [… a small volume of works ...]

Typical approaches

- Local approach (center-surround window)
  - Pixel based method [… initial works …]

- **Global** approach (global comparison)
  - **Superpixel based method** [… several works …]
  - Patch based method [… several works …]

- Frequency domain method [… a small volume of works ...]

Zhu, Wangjiang, Shuang Liang, Yichen Wei, and Jian Sun. "Saliency optimization from robust background detection." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2814-2821. 2014.

Superpixel segmentation

- Segment the image into superpixels

Input image

Image courtesy: Li *et al.* 2011

Superpixel segmentation

- Segment the image into superpixels

Superpixel segmentation

Superpixel boundary is shown in green

Superpixel segmentation

- Segment the image into superpixels
  - group of pixels
  - contains pixels of uniform color
  - retain object boundary unlike rectangular patches

Superpixel segmentation

Superpixel boundary is shown in green

Superpixel segmentation

- Segment the image into superpixels
  - group of pixels
  - contains pixels of uniform color
  - retain object boundary unlike rectangular patches

- Compute features (vectors) for every superpixel
  - color mean
  - color histogram

Superpixel segmentation

Superpixel boundary is shown in green

Saliency of superpixels

- Compute saliency value of every superpixel and assign it to pixels inside it

Superpixel segmentation

Saliency of superpixels

- Saliency value of this superpixel should be ***large***

Superpixel segmentation

Saliency of superpixels

- Saliency value of this superpixel should be ***large***

Superpixel segmentation

Saliency of superpixels

- Saliency value of this superpixel should be ***small***

Superpixel segmentation

Saliency of superpixels

- Saliency value of this superpixel should be ***small***

Superpixel segmentation

Notations

- A superpixel v_i
- **Spatial** distance between two superpixels v_i and v_j is $d_s(v_i, v_j)$

v_i

v_j

Notations

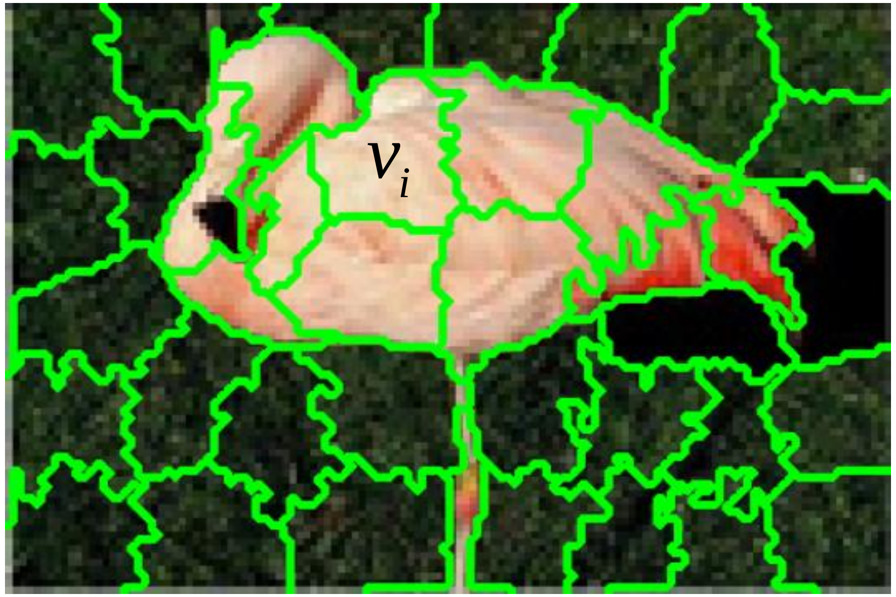
- A superpixel v_i
- **Feature** distance between two superpixels v_i and v_j is $d_f\left(v_i, v_j\right)$

Notations

- A superpixel v_i
- **Spatial** distance between two superpixels v_i and v_j is $d_s(v_i, v_j)$
- **Feature** distance between two superpixels v_i and v_j is $d_f(v_i, v_j)$
- **Geodesic** distance between two superpixels v_i and v_j is $d_g(v_i, v_j)$
- **Saliency** value of a superpixel v_i is $\mathcal{S}(v_i)$

Saliency computation

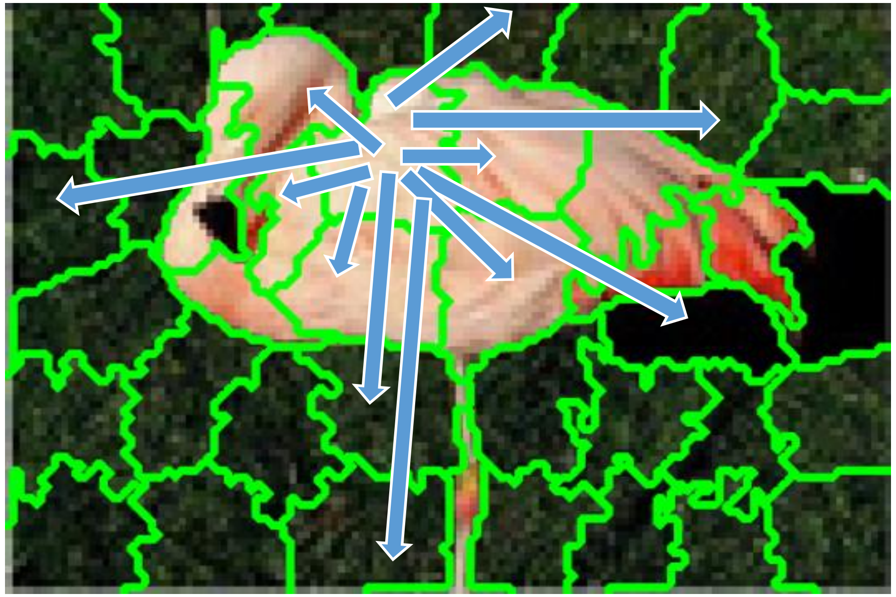
- 'Global' method
- Saliency value of a superpixel



Saliency computation

- ‘Global’ method
- Saliency value of a superpixel is the sum of its feature distances with the rest of the superpixels in the image

$$\mathcal{S}(v_i) = \sum_{j=1}^{N} d_f(v_i, v_j)$$

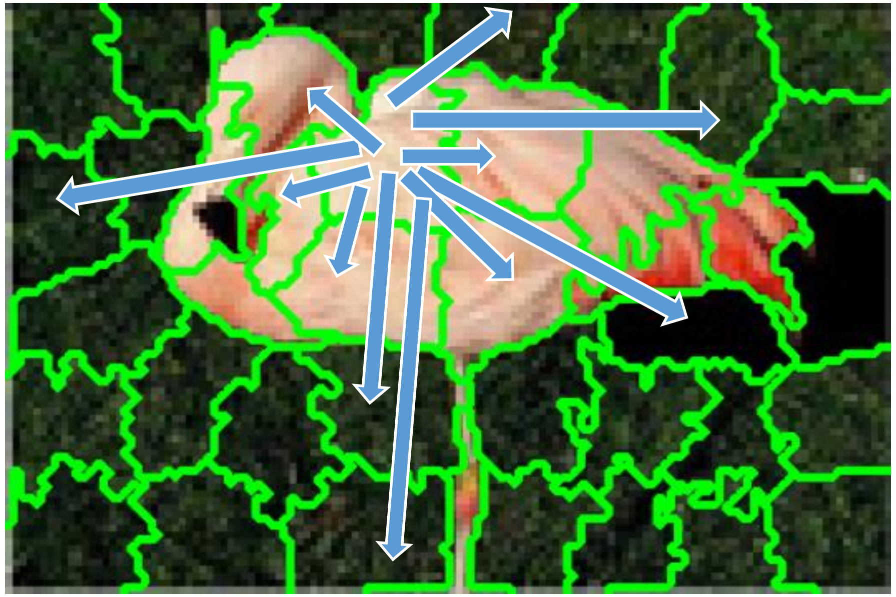


Saliency computation

- Incorporate 'local' contrast
- We want to give more importance to nearer superpixels

Feature distance

$$\mathcal{S}(v_i) = \sum_{j=1}^{N} d_f\left(v_i, v_j\right)$$



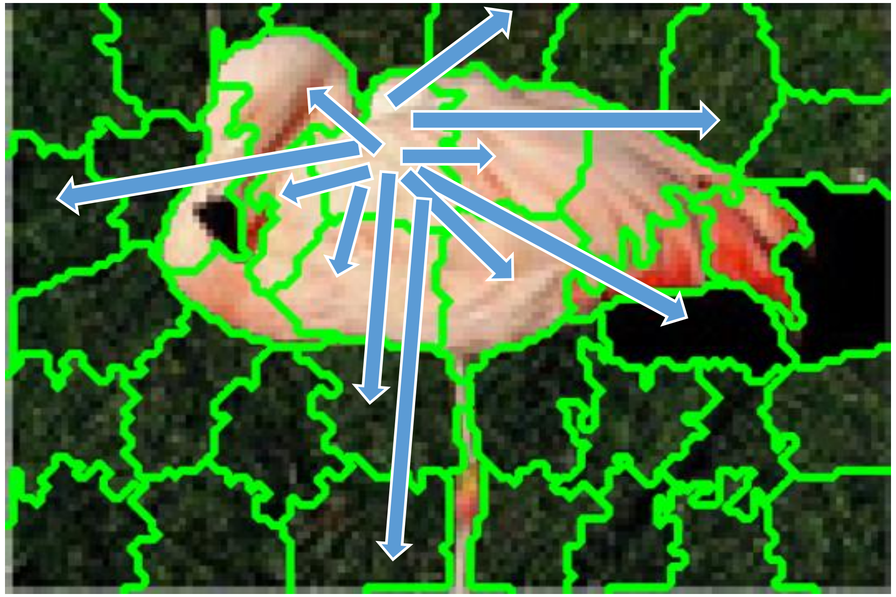
Saliency computation

- Incorporate 'local' contrast
- We want to give more importance to nearer superpixels

Feature distance

Inverse of spatial distance

$$\mathcal{S}(v_i) = \sum_{j=1}^{N} d_f(v_i, v_j) \exp\left(-\frac{1}{2\sigma_1^2} \left(d_s(v_i, v_j) \right)^2 \right)$$



Background information

- Add semi-supervision
- Use background information
- For every superpixel, find its probability of belonging to background
  - i.e. not being salient

- Background probability of a superpixel v_j is $\mathcal{P}_{\mathcal{B}}\left(v_j\right)$

$$\mathcal{S}\left(v_i\right)=\sum_{j=1}^{N} \underbrace{d_f\left(v_i,v_j\right)}_{\text{Feature distance}} \underbrace{\exp\left(-\frac{1}{2\sigma_1^2}\left(d_s\left(v_i,v_j\right)\right)^2\right)}_{\text{Inverse of spatial distance}}$$

Background information

- Add semi-supervision
- Use background information
- For every superpixel, find its probability of belonging to background
  - i.e. not being salient

- Background probability of a superpixel v_j is $\mathcal{P}_{\mathcal{B}}\left(v_j\right)$
- Modified saliency equation

$$\mathcal{S}\left(v_i\right)=\sum_{j=1}^{N} \underbrace{d_f\left(v_i,v_j\right)}_{\text{Feature distance}} \underbrace{\exp\left(-\frac{1}{2\sigma_1^2}\left(d_s\left(v_i,v_j\right)\right)^2\right)}_{\text{Inverse of spatial distance}} \underbrace{\mathcal{P}_{\mathcal{B}}\left(v_j\right)}_{\textbf{Background probability}}$$

Background superpixels

- Boundary superpixels may belong to background

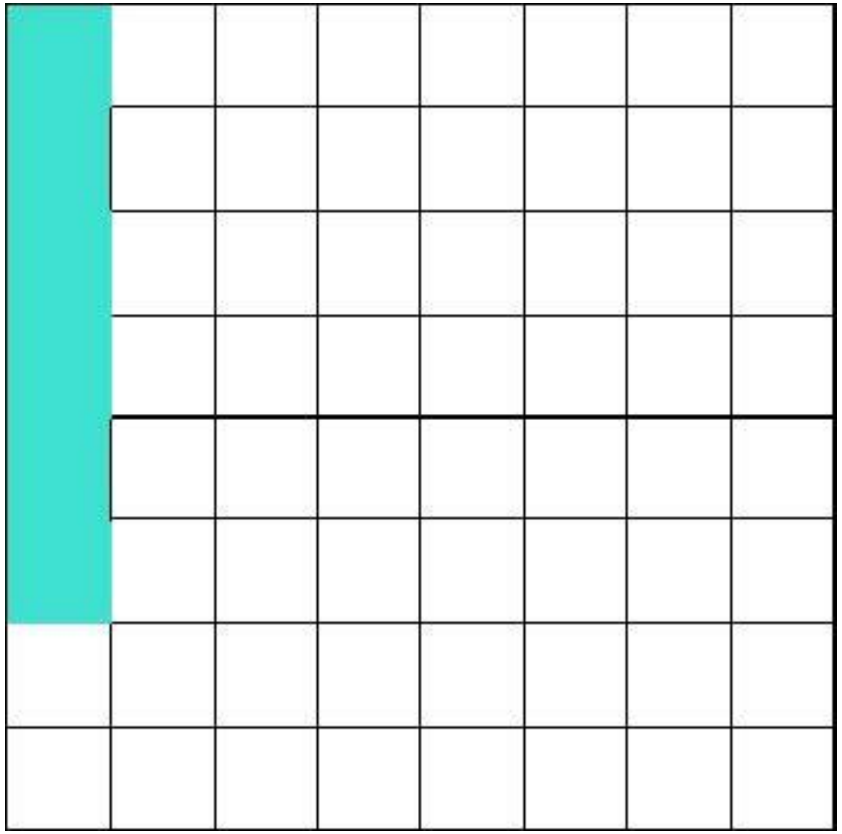


Image superpixel similarity. Same color indicates high feature similarity

Continued ...

- Non-boundary superpixel with high background probability

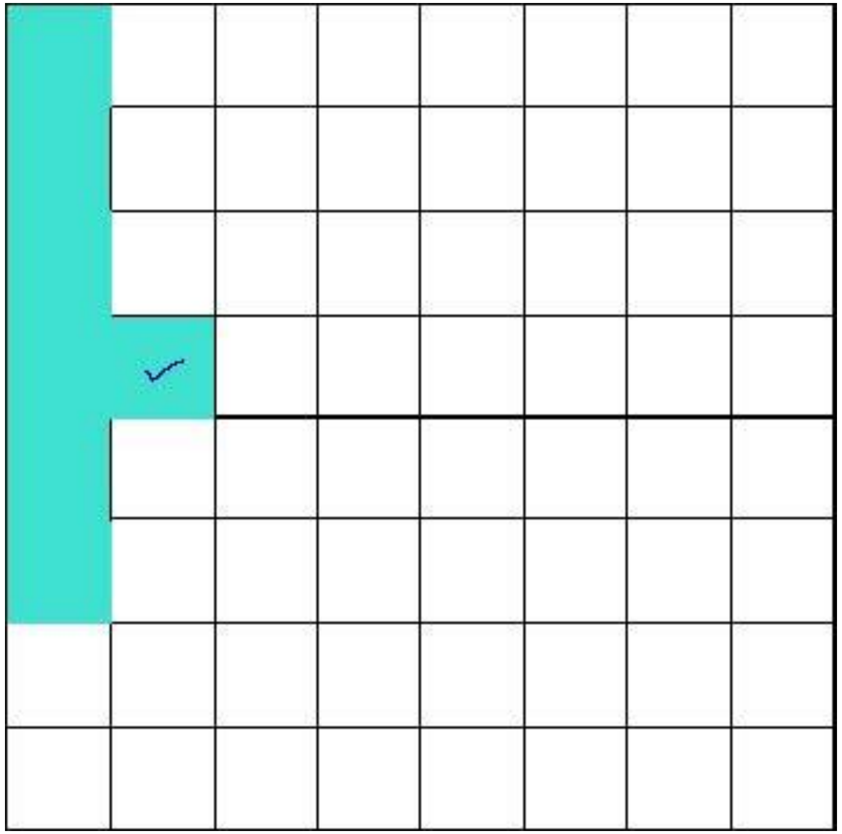


Image superpixel similarity. Same color indicates high feature similarity

Continued …

- Boundary superpixel similar to non-boundary superpixels

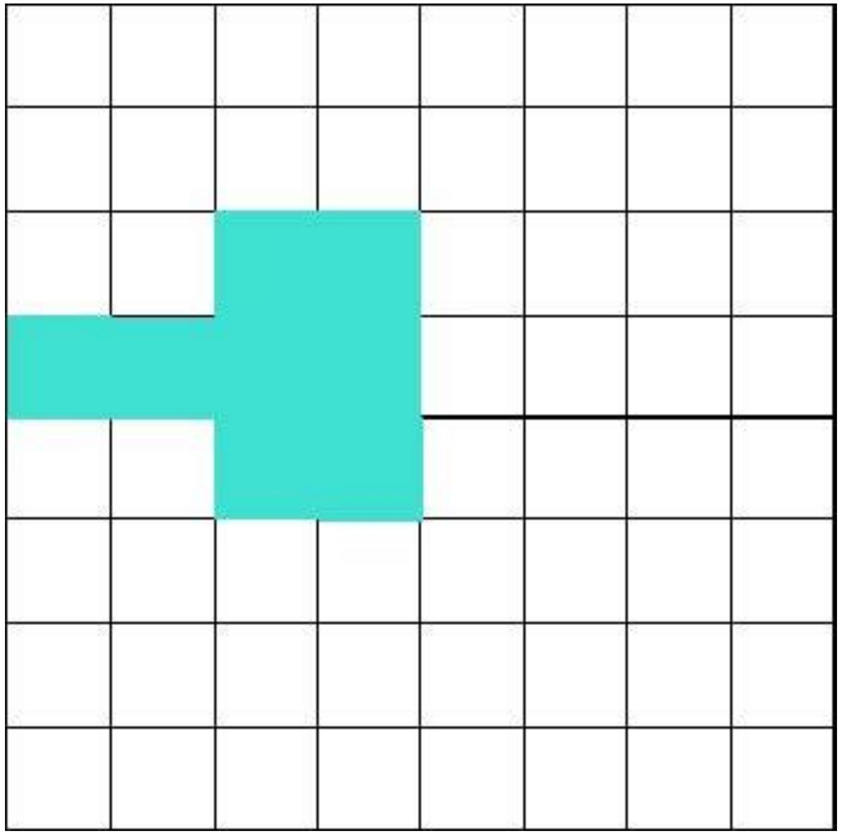


Image superpixel similarity. Same color indicates high feature similarity

Continued …

- Boundary superpixel with small background probability

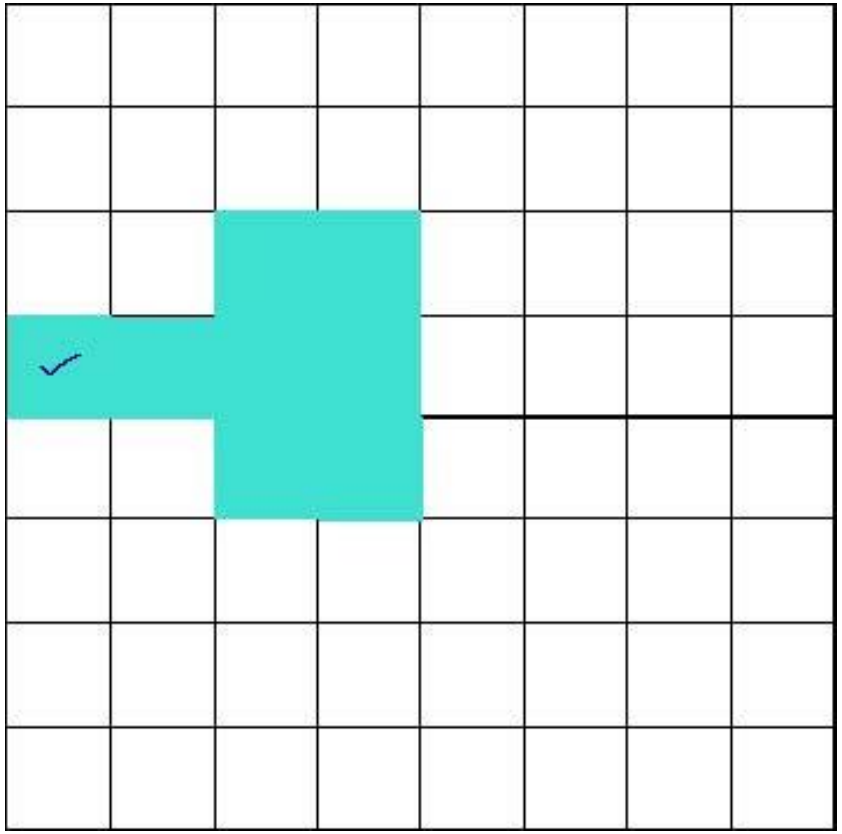


Image superpixel similarity. Same color indicates high feature similarity

Continued …

- Introduce geodesic distance that combines
  - Feature distance
  - Spatial location

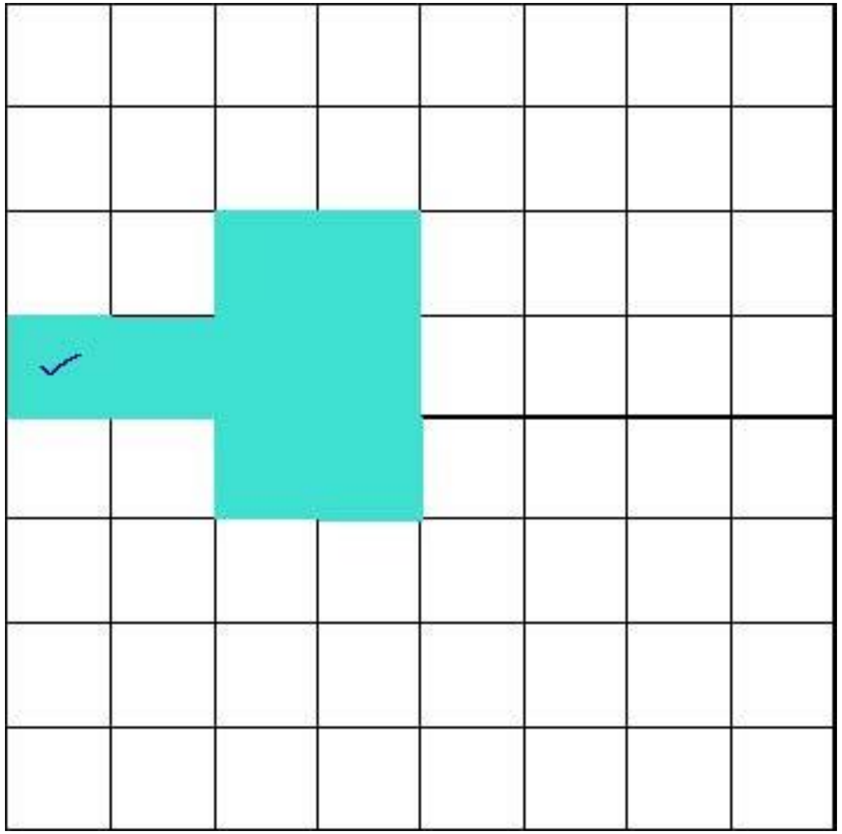
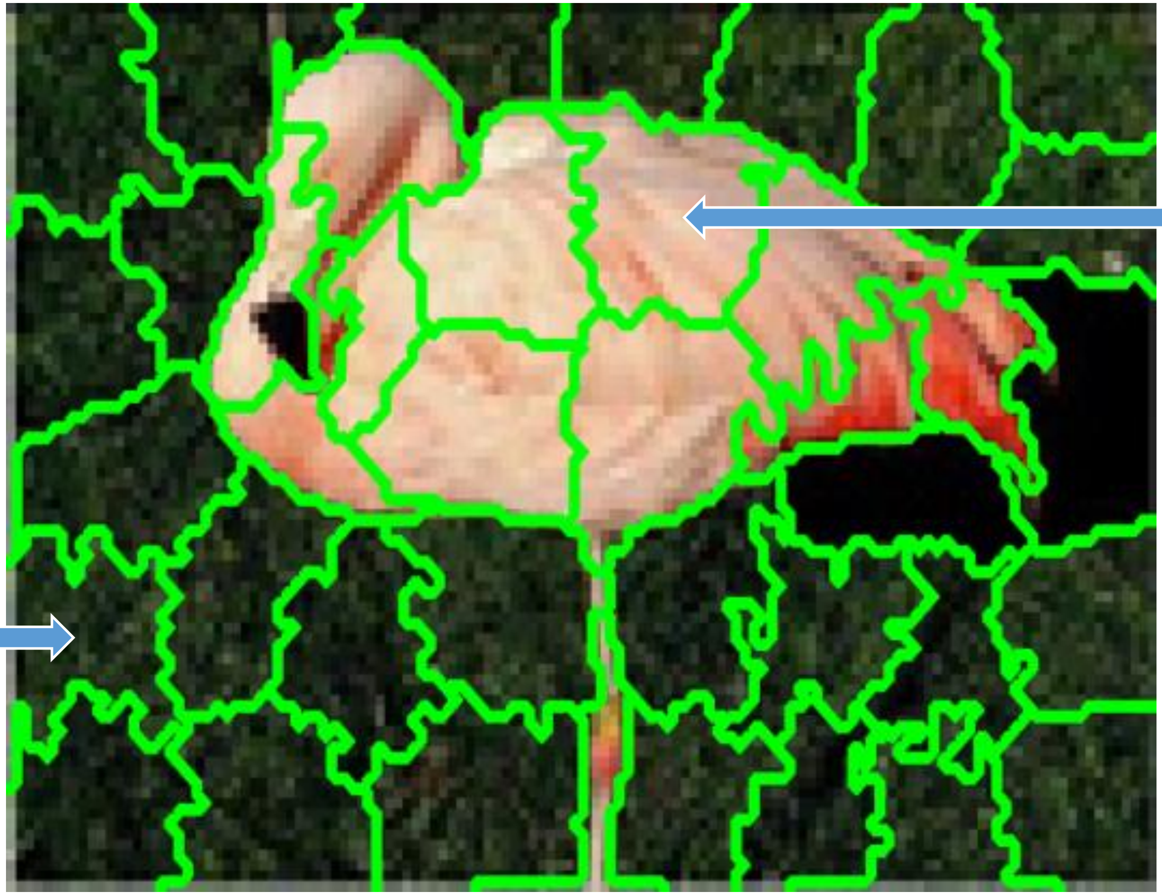


Image superpixel similarity. Same color indicates high feature similarity

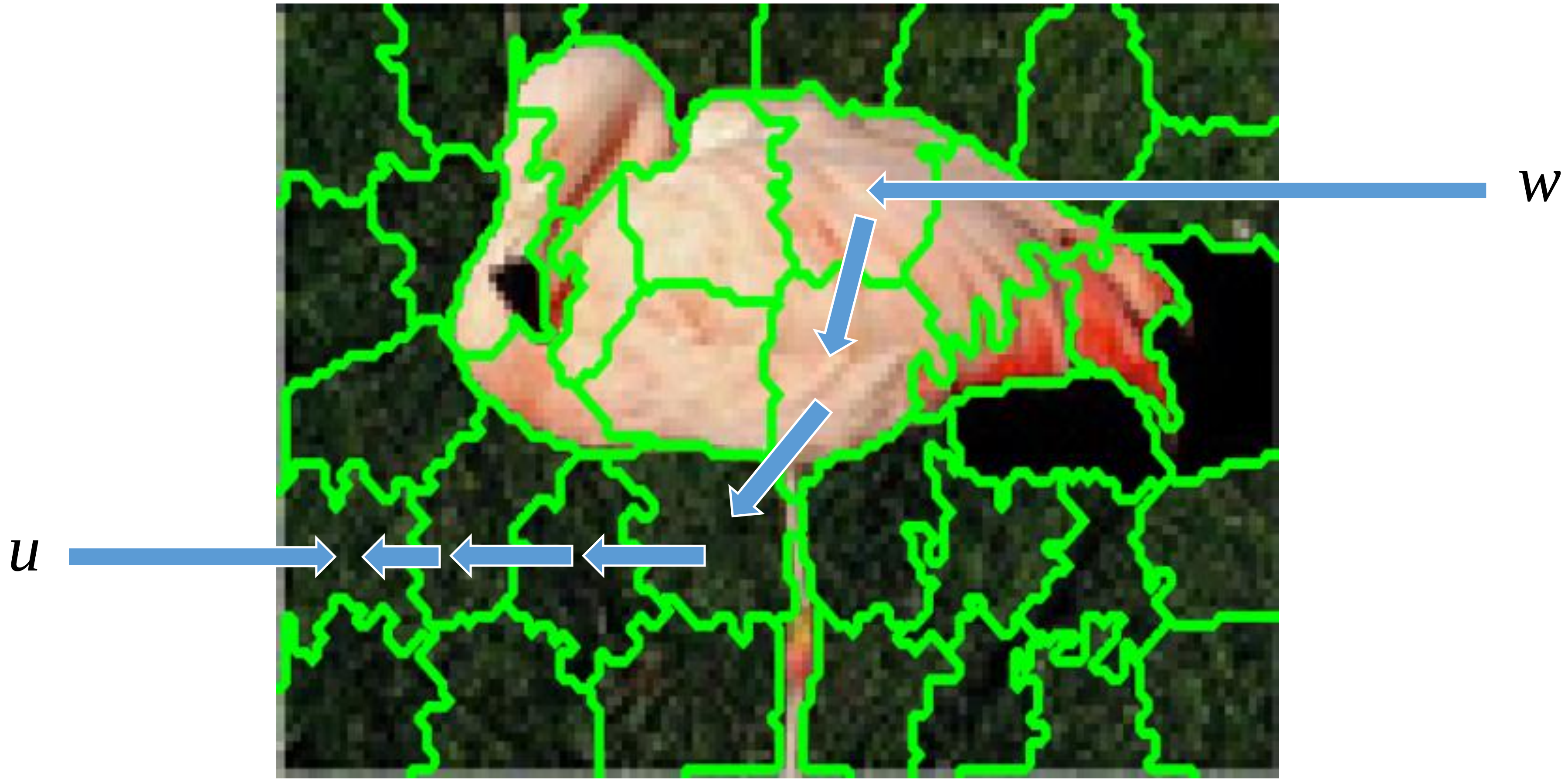
Geodesic distance



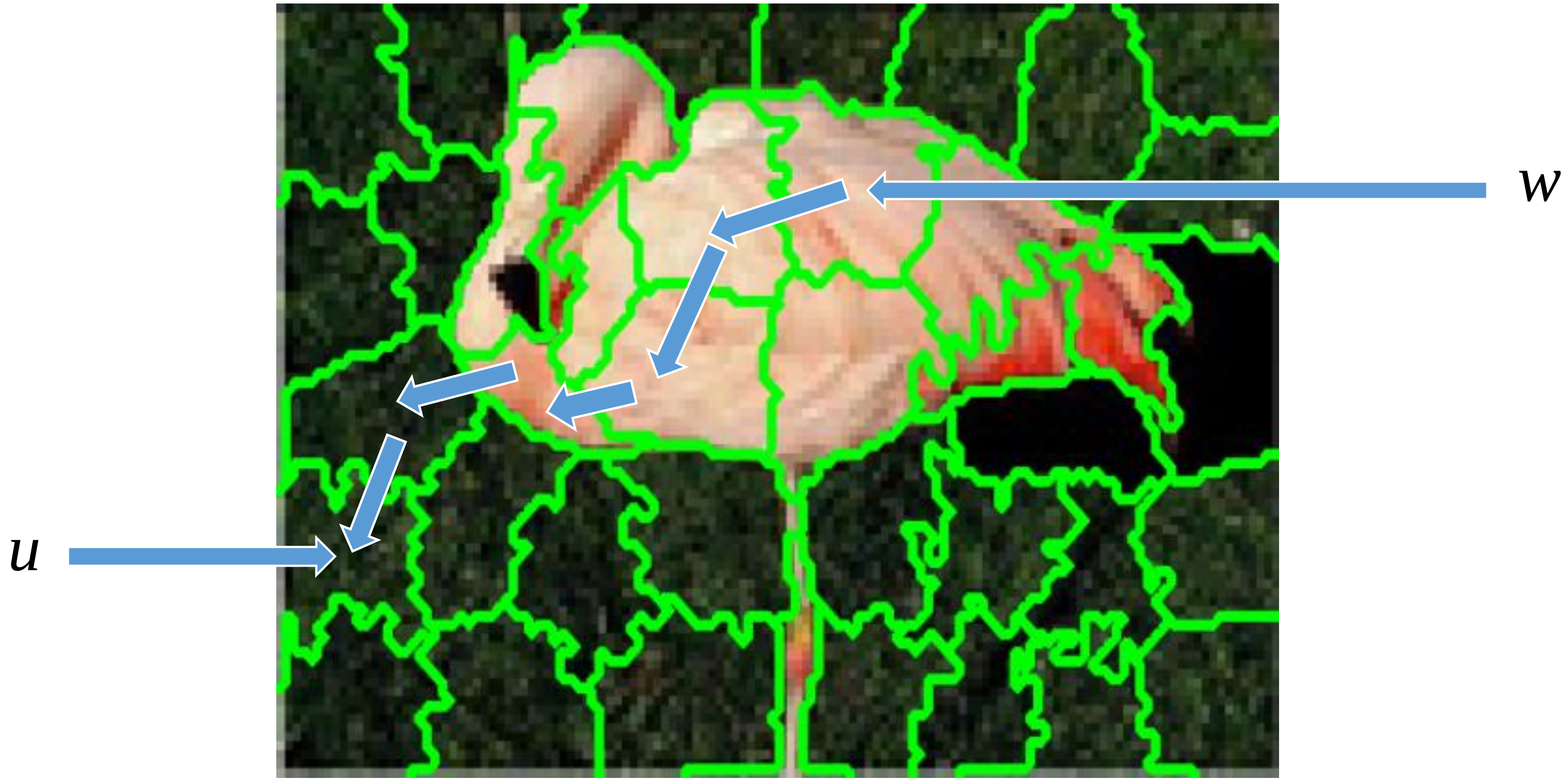
w

u

Geodesic distance



Geodesic distance



Geodesic distance

- Geodesic distance between two superpixels u and w is
  - the sum of feature distances of neighboring superpixels
  - along the *shortest path* $u = v_1, v_2, v_3, \ldots, v_{L-1}, v_L = w$
  - between u and w in terms of feature distance

$$d_g(u,w) = \min_{\text{shortest path}} \sum_{i=1}^{L-1} d_f(v_i, v_{i+1})$$

- *L* is different for different u and w

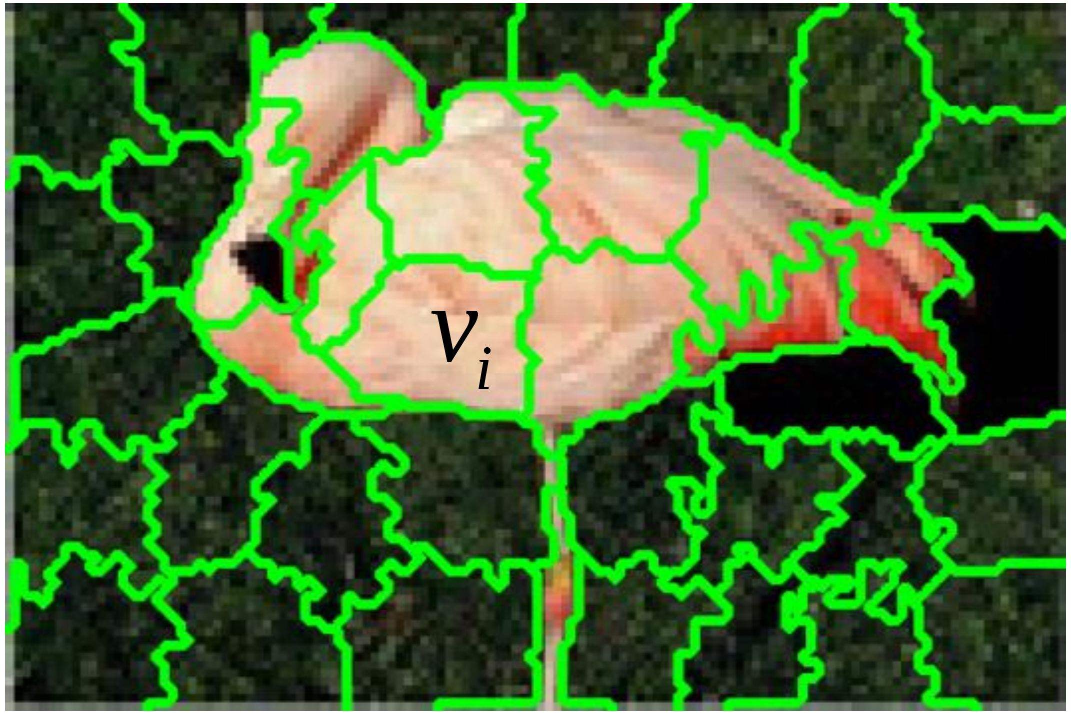
Background probability

- Geodesic similarity $d_g^*(v_i, v_j) = \exp\left(-\frac{1}{2\sigma_2^2}\left(d_g(v_i, v_j)\right)^2\right)$

Background probability

- Geodesic similarity $d_g^*(v_i, v_j) = \exp\left(-\frac{1}{2\sigma_2^2}\left(d_g(v_i, v_j)\right)^2\right)$

- Background contrast

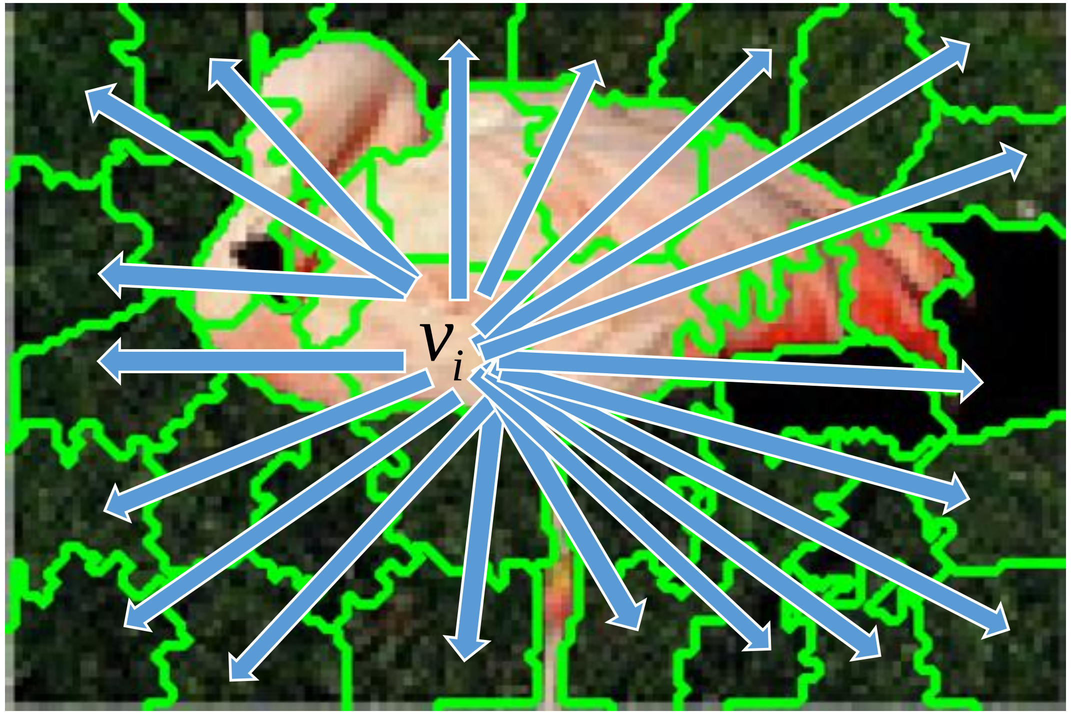


Background probability

- Geodesic similarity $d_g^*\left(v_i, v_j\right) = \exp\left(-\frac{1}{2\sigma_2^2}\left(d_g\left(v_i, v_j\right)\right)^2\right)$

- Background contrast

$$\mathcal{C}_{\mathcal{B}}\left(v_i\right) = \frac{\sum_{j=1}^{N} d_g^*\left(v_i, v_j\right)\delta\left(v_j \in \mathcal{B}\right)}{\sqrt{\sum_{j=1}^{N} d_g^*\left(v_i, v_j\right)}}$$

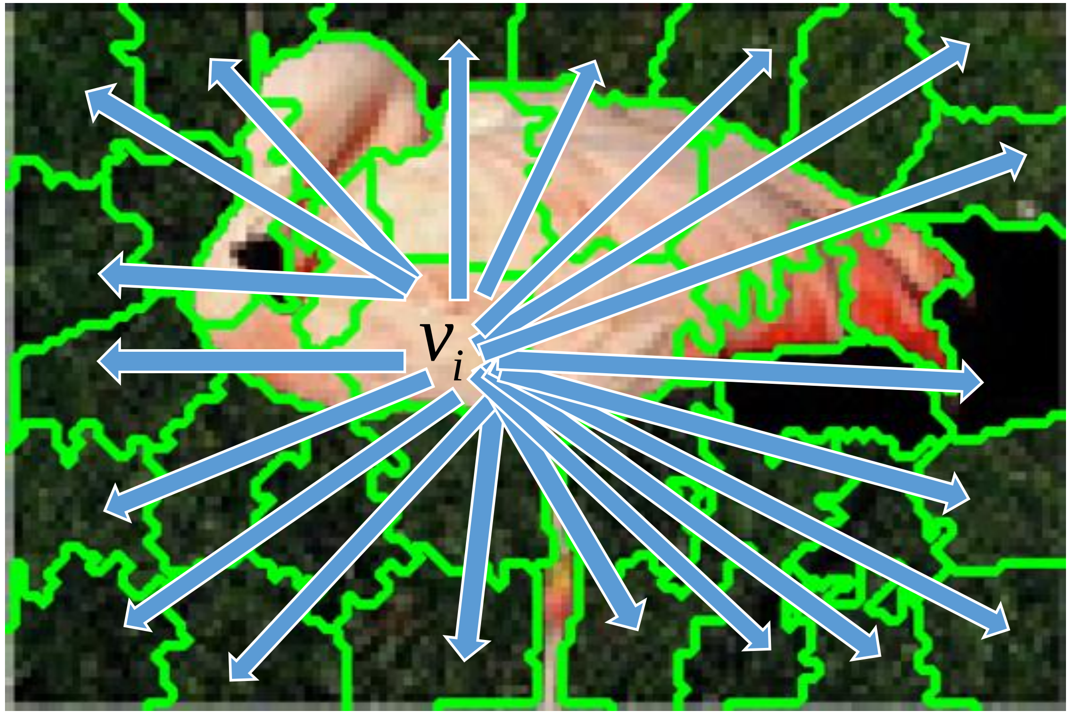


Background probability

- Geodesic similarity $d_g^*(v_i, v_j) = \exp\left(-\frac{1}{2\sigma_2^2}\left(d_g(v_i, v_j)\right)^2\right)$

- Background contrast

$$\mathcal{C}_{\mathcal{B}}(v_i) = \frac{\sum_{j=1}^{N} d_g^*(v_i, v_j)\,\delta(v_j \in \mathcal{B})}{\sqrt{\sum_{j=1}^{N} d_g^*(v_i, v_j)}}$$



- Background probability of a superpixel

$$\mathcal{P}_{\mathcal{B}}(v_i) = 1 - \exp\left(-\frac{1}{2\sigma_3^2}\left(\mathcal{C}_{\mathcal{B}}(v_i)\right)^2\right)$$

Saliency equation re-visited

- Saliency value of a superpixel v_i

$$\mathcal{S}(v_i) = \sum_{j=1}^{N} d_f(v_i, v_j)$$

$$\mathcal{S}(v_i) = \sum_{j=1}^{N} d_f(v_i, v_j) \exp\left(-\frac{1}{2\sigma_1^2}\left(d_s(v_i, v_j)\right)^2\right)$$

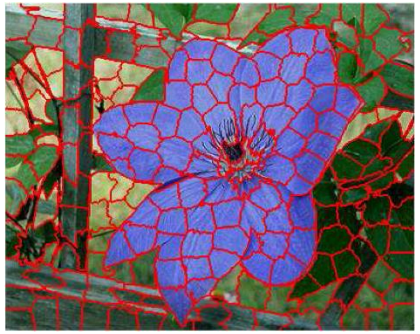
$$\mathcal{S}(v_i) = \sum_{j=1}^{N} d_f(v_i, v_j) \exp\left(-\frac{1}{2\sigma_1^2}\left(d_s(v_i, v_j)\right)^2\right) \mathcal{P}_{\mathcal{B}}(v_j)$$

Feature distance

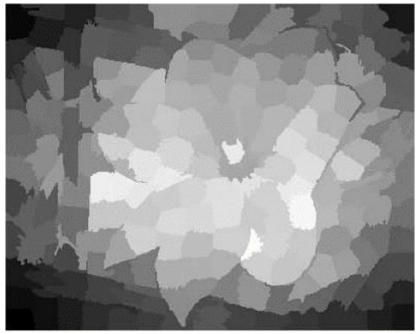
Inverse of spatial distance

Background probability

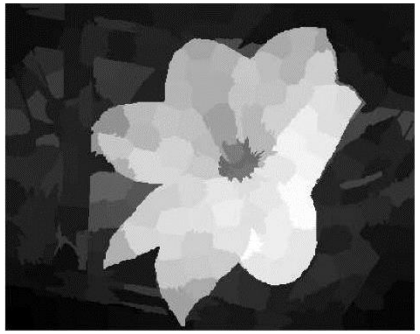
Results



Input image



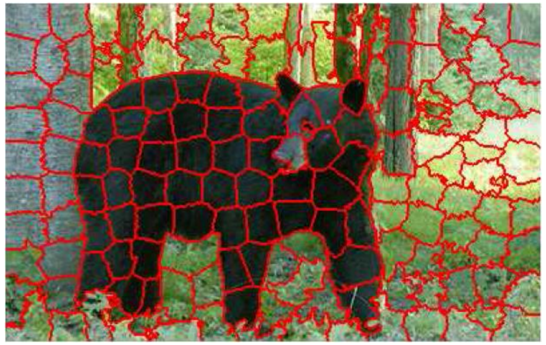
Saliency map
with
Feature distance
+
Inverse of
spatial distance



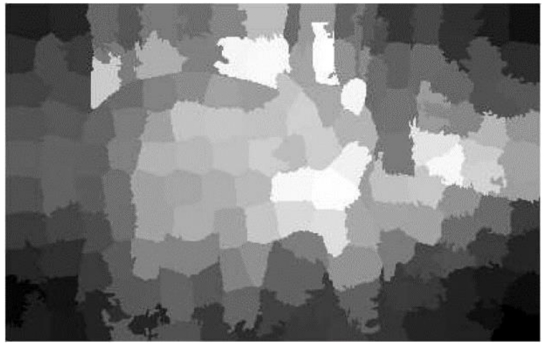
Saliency map
with
Feature distance
+
Inverse of
spatial distance
+
Background
probability

Result from Zhu *et al.*, CVPR 2014

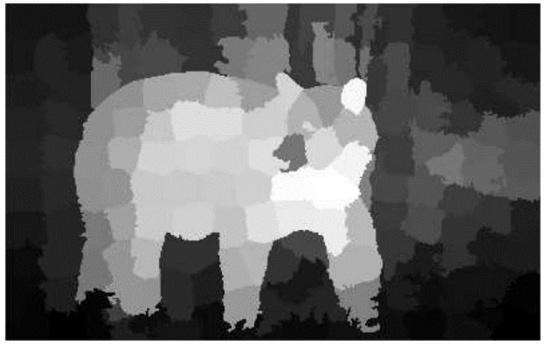
Results



Input image



Saliency map
with
Feature distance
+
Inverse of
spatial distance

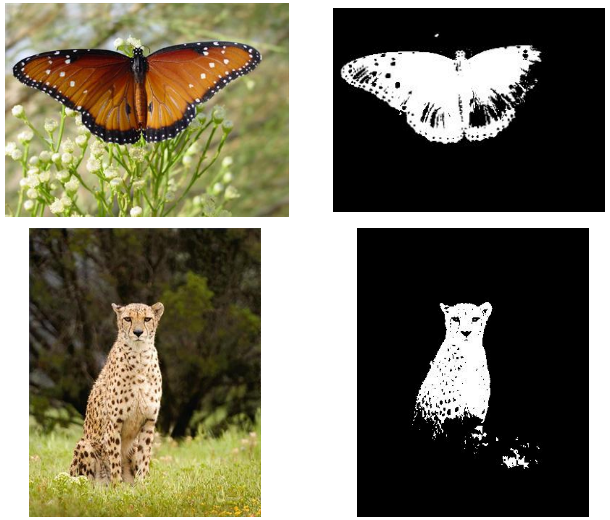


Saliency map
with
Feature distance
+
Inverse of
spatial distance
+
Background
probability

Result from Zhu *et al.*, CVPR 2014

Challenges

- Local approach (center-surround window)
  - Pixel based
  - Boundary of large smooth salient object detected
- Global approach (global comparison)
  - Superpixel based and patch based
  - Salient objects are not uniformly highlighted
  - Upon thresholding,
    - textures, holes, spurious points, incomplete objects
- Frequency domain
  - Saliency map is blurred



Input image   Thresholded saliency

Result from Perazzi *et al.*, CVPR 2012

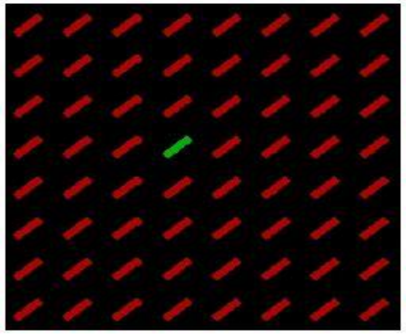
Image courtesy: MSRA dataset

Saliency in synthetic data

- Use local approach (center-surround window)

Based on color

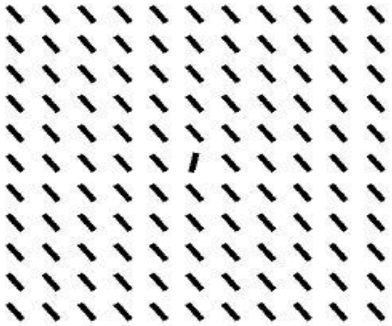
Input image



Saliency map

Based on orientation

Input image



Saliency map

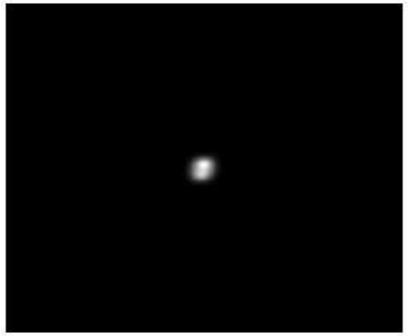
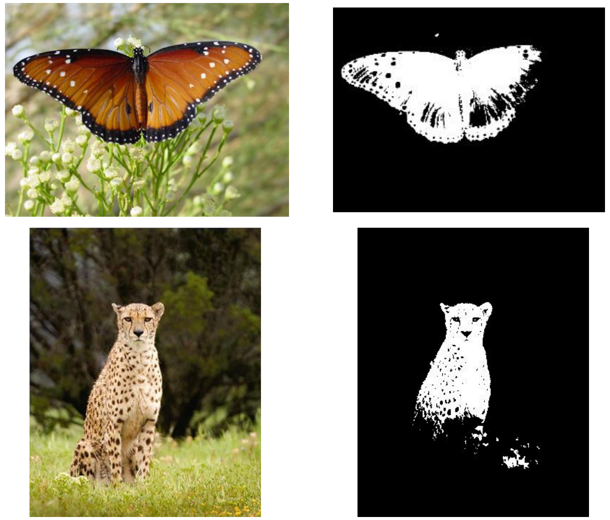


Image courtesy: Gao *et al*.

Challenges

- Local approach (center-surround window)
  - Pixel based
  - Boundary of large smooth salient object detected
- Global approach (global comparison)
  - Superpixel based and patch based
  - Salient objects are not uniformly highlighted
  - Upon thresholding,
    - textures, holes, spurious points, incomplete objects
- Frequency domain
  - Saliency map is blurred



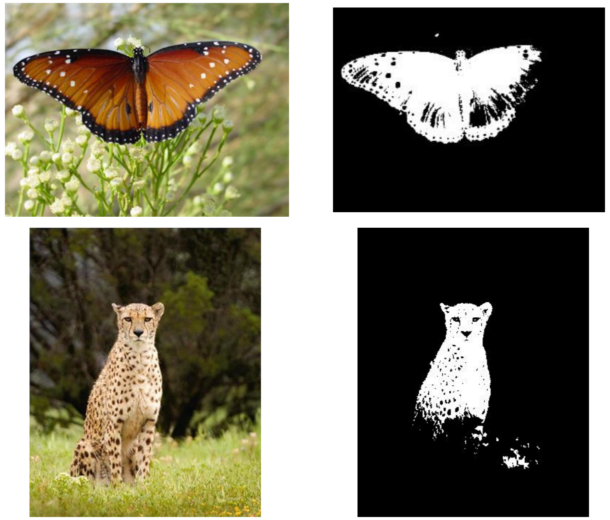
Input image Thresholded saliency

Result from Perazzi *et al.*, CVPR 2012

Image courtesy: MSRA dataset

Challenges

- Local approach (center-surround window)
  - Pixel based
  - Boundary of large smooth salient object detected
- Global approach (global comparison)
  - Superpixel based and patch based
  - Salient objects are not uniformly highlighted
  - Upon thresholding,
    - **textures, holes, spurious points, incomplete objects**
- Frequency domain
  - Saliency map is blurred



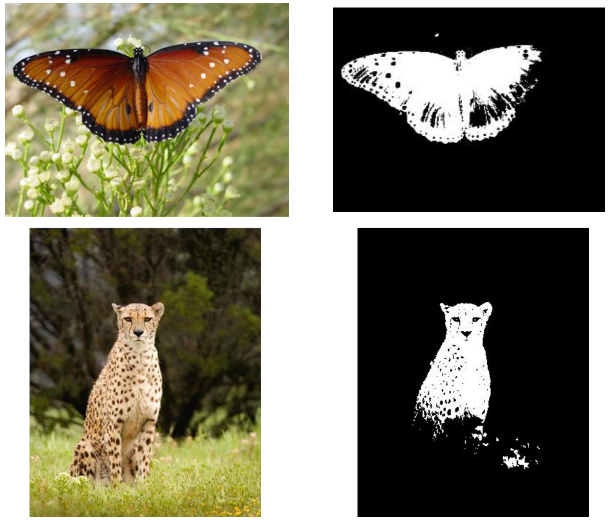
Input image | Thresholded saliency

Result from Perazzi *et al.*, CVPR 2012

Image courtesy: MSRA dataset

Challenges

- Local approach (center-surround window)
  - Pixel based
  - Boundary of large smooth salient object detected
- Global approach (global comparison)
  - Superpixel based and patch based
  - Salient objects are not uniformly highlighted
  - Upon thresholding,
    - **textures, holes, spurious points, incomplete objects**
      - **possible solution1: neighboring superpixels – similar saliency value**
- Frequency domain
  - Saliency map is blurred



Input image Thresholded saliency

Result from Perazzi *et al.*, CVPR 2012

Image courtesy: MSRA dataset

Saliency optimization

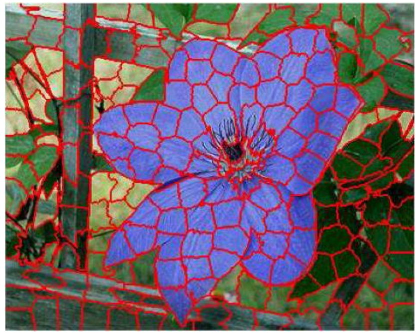
- Minimize the cost function

$$\min_{\mathbf{s}} \left(\left(\sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \left(s_i - s_j \right)^2 d_f \left(v_i, v_j \right) \right) + \left(\sum_{i=1}^{N} \left(1 - s_i \right)^2 \mathcal{S}\left(v_i \right) \right) + \left(\sum_{i=1}^{N} \left(s_i \right)^2 \mathcal{P}_{\mathcal{B}} \left(v_i \right) \right) \right)$$

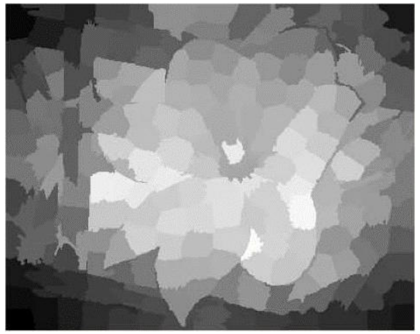
- s_i is the updated saliency value of superpixel v_i after optimization

- First term: data fitting term
- Second term: foreground (salient region) prior
- Third term: background (non-salient region) prior

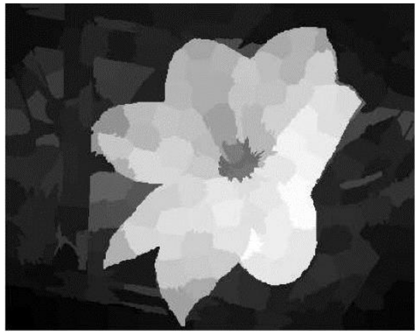
Results



Input image



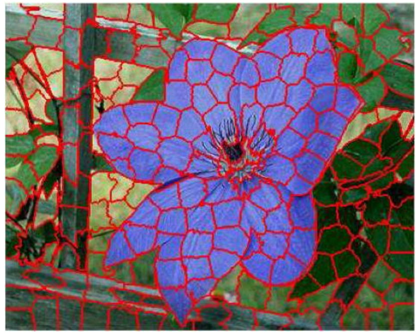
Saliency map
with
Feature distance
+
Inverse of
spatial distance



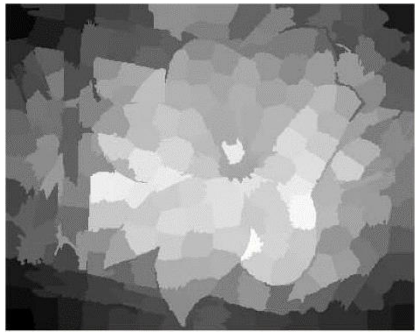
Saliency map
with
Feature distance
+
Inverse of
spatial distance
+
Background
probability

Result from Zhu *et al.*, CVPR 2014

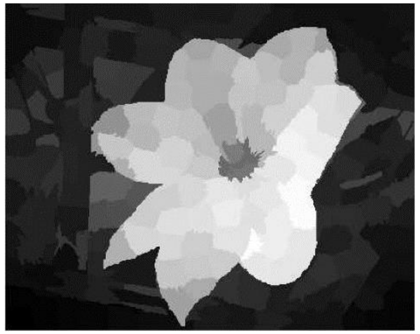
Results



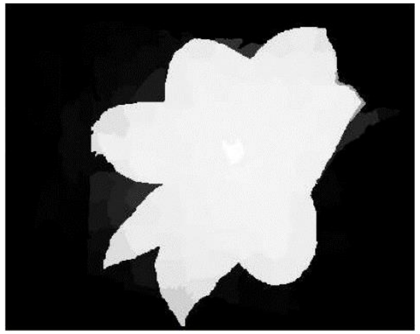
Input image



Saliency map
with
Feature distance
+
Inverse of
spatial distance



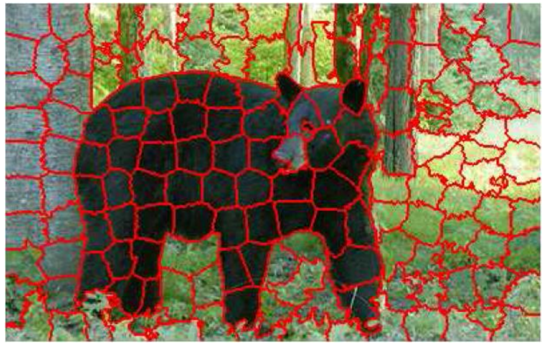
Saliency map
with
Feature distance
+
Inverse of
spatial distance
+
Background
probability



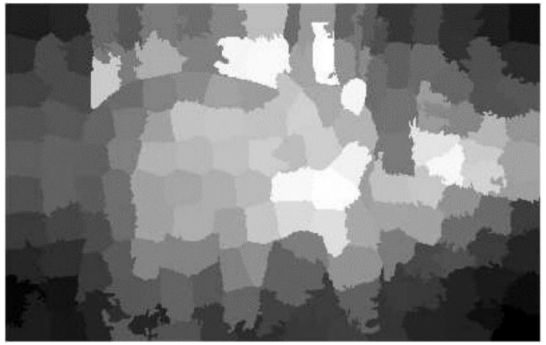
Saliency map
after
optimization

Result from Zhu *et al.*, CVPR 2014

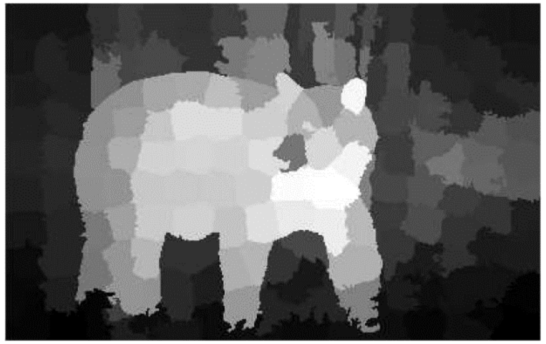
Results



Input image



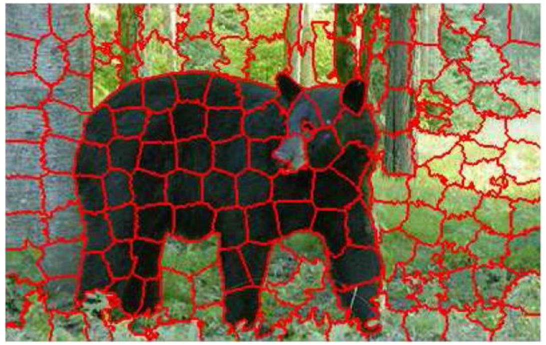
Saliency map
with
Feature distance
+
Inverse of
spatial distance



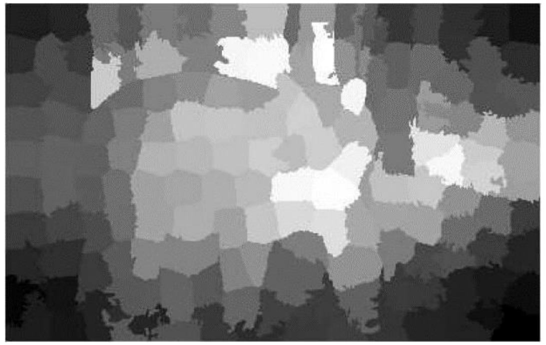
Saliency map
with
Feature distance
+
Inverse of
spatial distance
+
Background
probability

Result from Zhu *et al.*, CVPR 2014

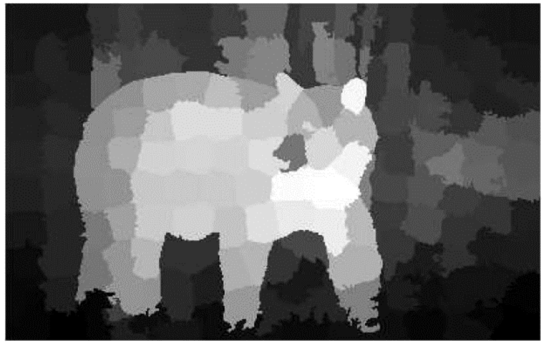
Results



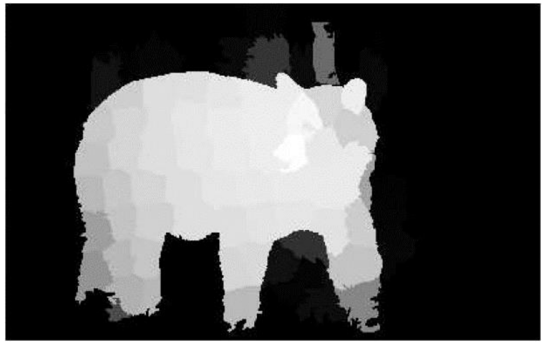
Input image



Saliency map
with
Feature distance
+
Inverse of
spatial distance



Saliency map
with
Feature distance
+
Inverse of
spatial distance
+
Background
probability

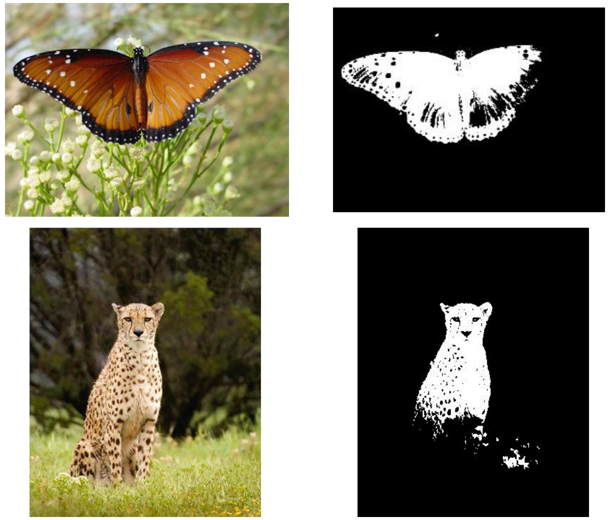


Saliency map
after
optimization

Result from Zhu *et al.*, CVPR 2014

Challenges

- Local approach (center-surround window)
  - Pixel based
  - Boundary of large smooth salient object detected
- Global approach (global comparison)
  - Superpixel based and patch based
  - Salient objects are not uniformly highlighted
  - Upon thresholding,
    - **textures, holes, spurious points, incomplete objects**
- Frequency domain
  - Saliency map is blurred



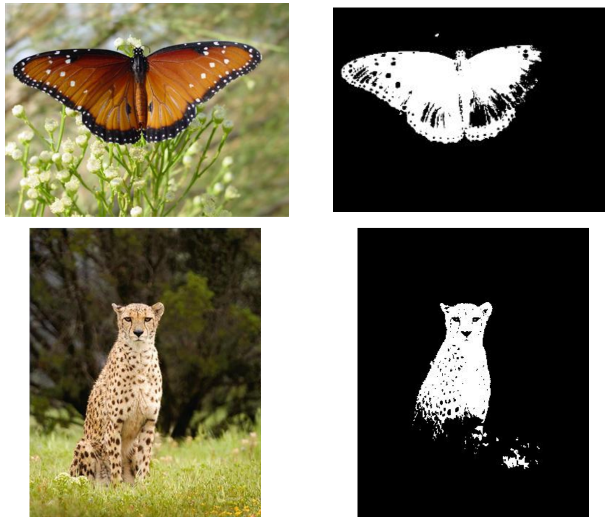
Input image   Thresholded saliency

Result from Perazzi *et al.*, CVPR 2012

Image courtesy: MSRA dataset

Challenges

- Local approach (center-surround window)
  - Pixel based
  - Boundary of large smooth salient object detected
- Global approach (global comparison)
  - Superpixel based and patch based
  - Salient objects are not uniformly highlighted
  - Upon thresholding,
    - **textures, holes, spurious points, incomplete objects**
      - **possible solution2: obtain larger regions before saliency computation**
- Frequency domain
  - Saliency map is blurred



Input image   Thresholded saliency

Result from Perazzi *et al.*, CVPR 2012

Image courtesy: MSRA dataset

Thank You