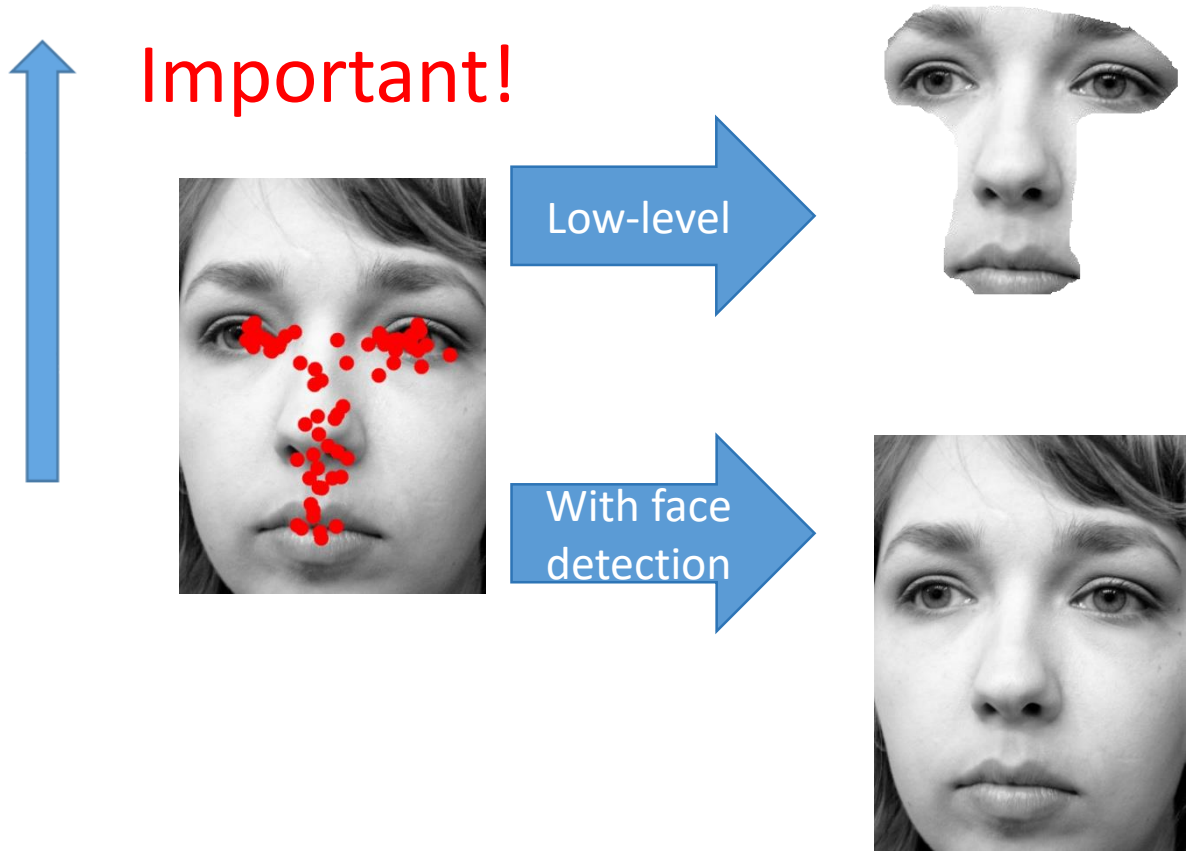
Some Representative Methods for Saliency Detection-II

Shenghua Gao

ShanghaiTech University

Some Interesting phenomenon in saliency detection

- Low-level(contrast)
  - Color
  - Orientation
  - Size
  - Motion
  - Depth
- High-level
  - People
  - Context

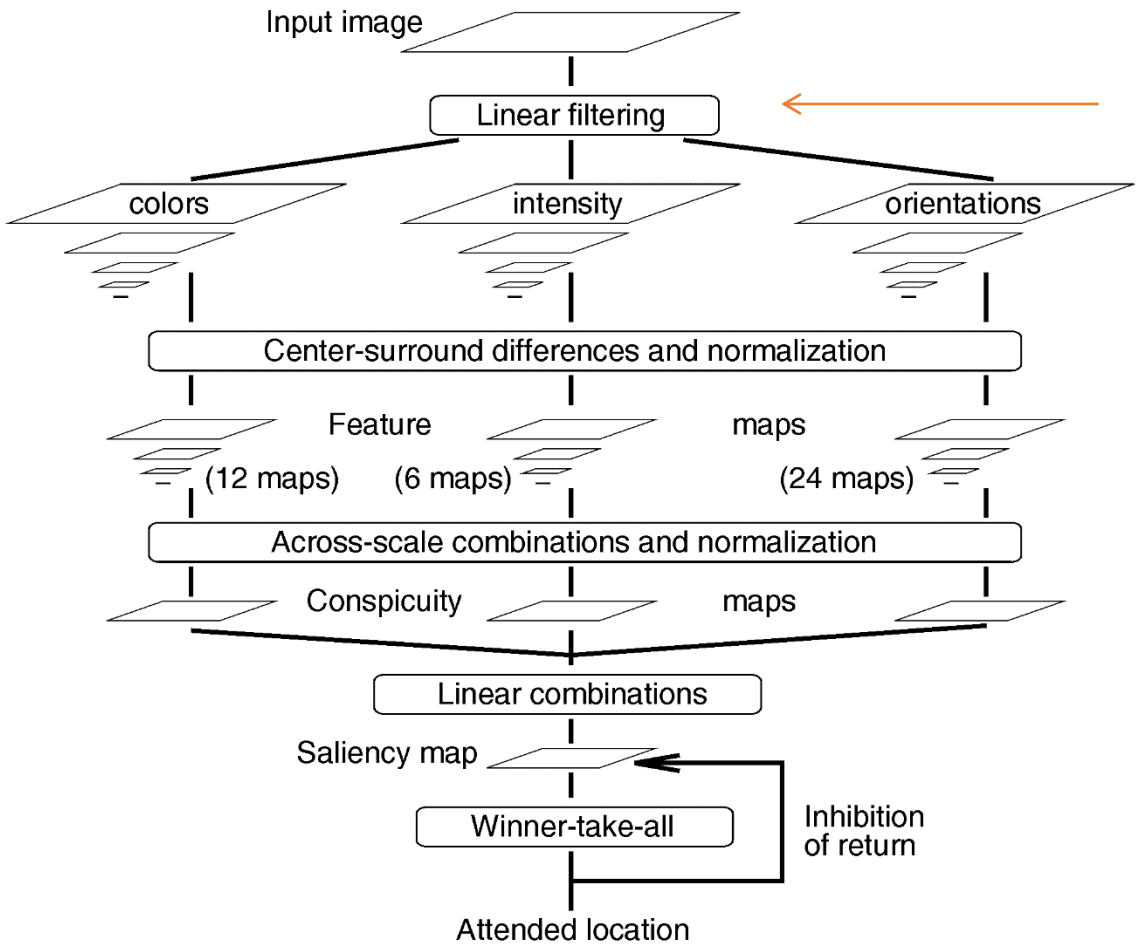


Judd et al, 2009

Outline

- Button-up approach
  - L. Itti's approach
  - Frequency-tuned
  - Multi-scale contrast
  - Depth of field
  - Spectral Residual approach
  - Global contrast based
- **Top-down approach**
  - Context-aware
- Information Maximization

A revisit of Itti's method



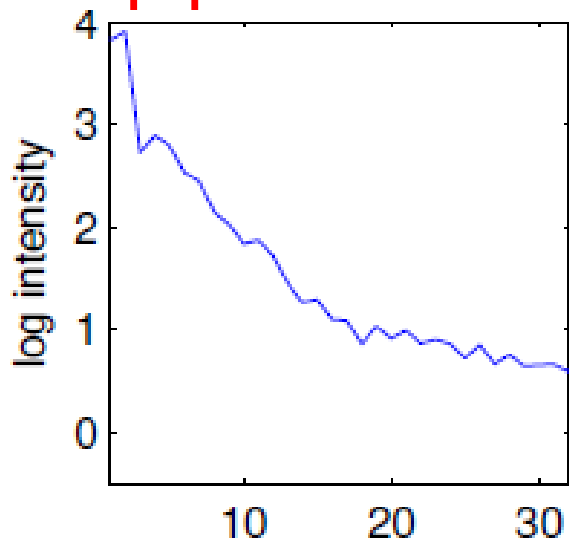
R,G,B,Y

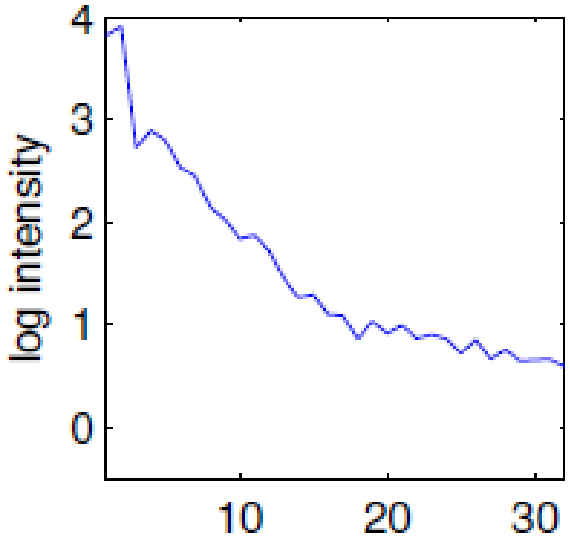
Gaussian Pyramids

Gabor pyramids for θ = {0º, 45º, 90º, 135º}

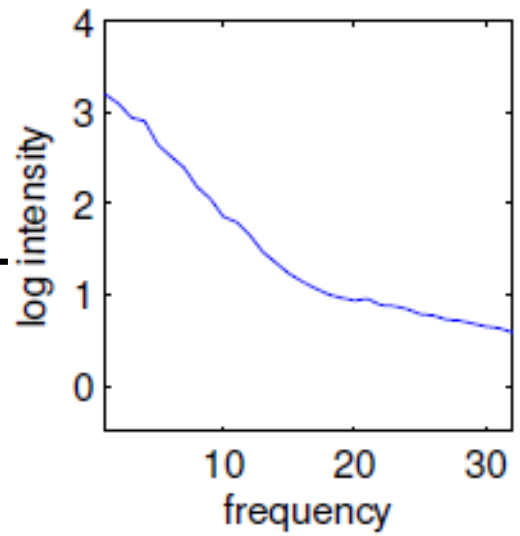
A revisit of Spectral Residual approach

$$\Rightarrow F \Rightarrow$$

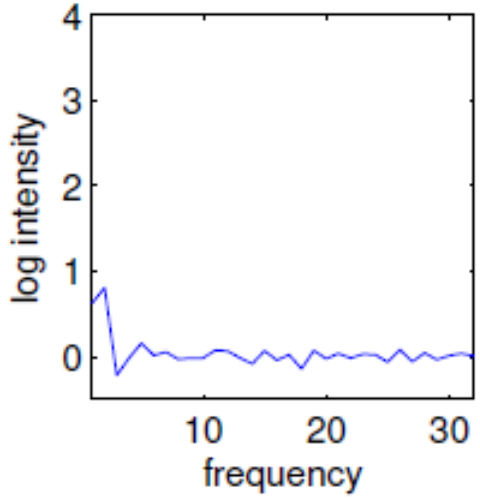




—



=



Context-Aware Saliency Detection

Stas Goferman, Lihi Zelnik-Manor, Ayellet Tal

CVPR,2010

Algorithms

- First glance: human attention.

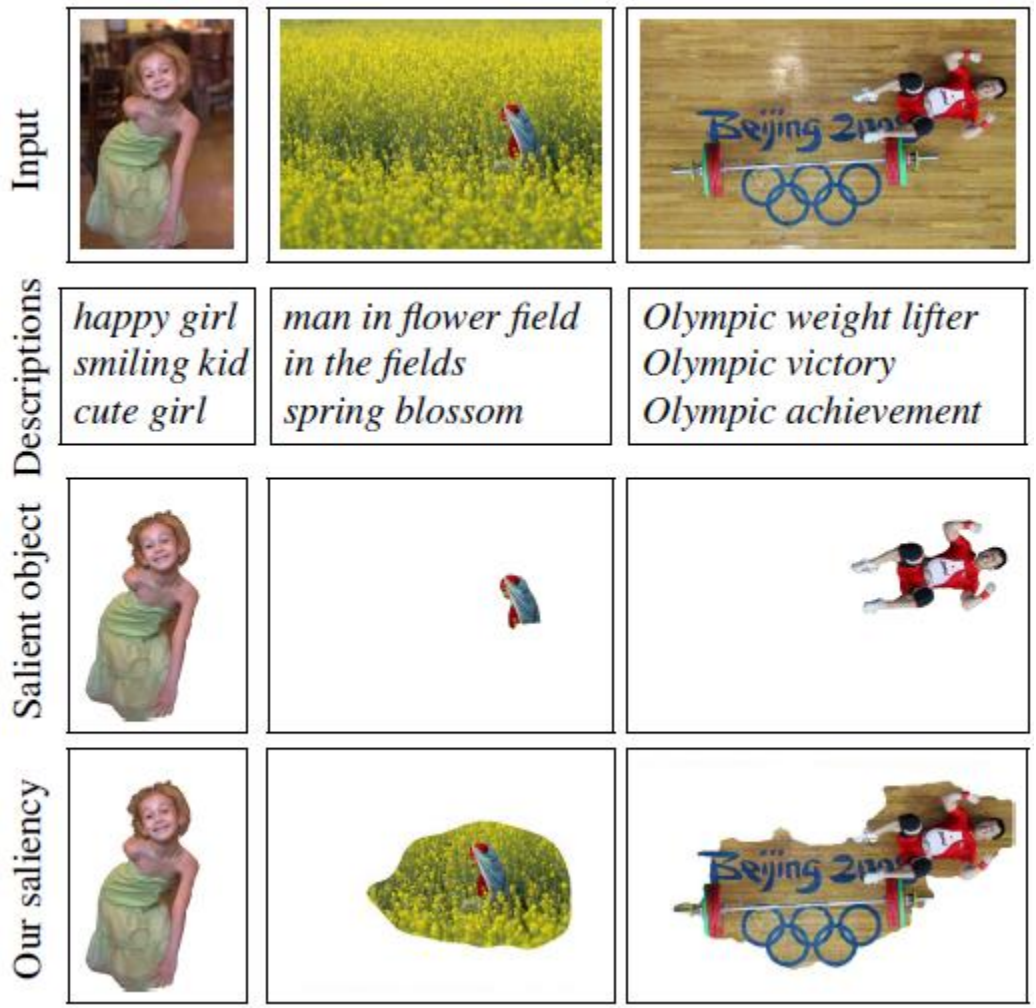
  EX: auto focusing.

- Dominant object

  EX: object recognition/segmentation.

- context of the dominant objects:

image classification, summarization of a photo collection, thumb nailing, and retargeting.


Input
Descriptions
Salient object
Our saliency
happy girl
smiling kid
cute girl
man in flower field
in the fields
spring blossom
Olympic weight lifter
Olympic victory
Olympic achievement

Principles of context-aware saliency

- 1. Local low-level considerations, including factors such as contrast and color.

  the distinctive color and other features area should obtain high attention.

- 2. Global considerations, which suppress frequently occurring features, while maintaining features that deviate from the norm.

  redundant information should be suppressed and popping up the novelty part.

Principles of context-aware saliency

- 3. Visual organization rules, which state that visual forms may possess one or several centers of gravity about which the form is organized.

  the salient pixels should be grouped together, and not spread all over the image.

- 4. High-level factors, such as human faces.

  implemented as post-processing.

Detection of context-aware saliency

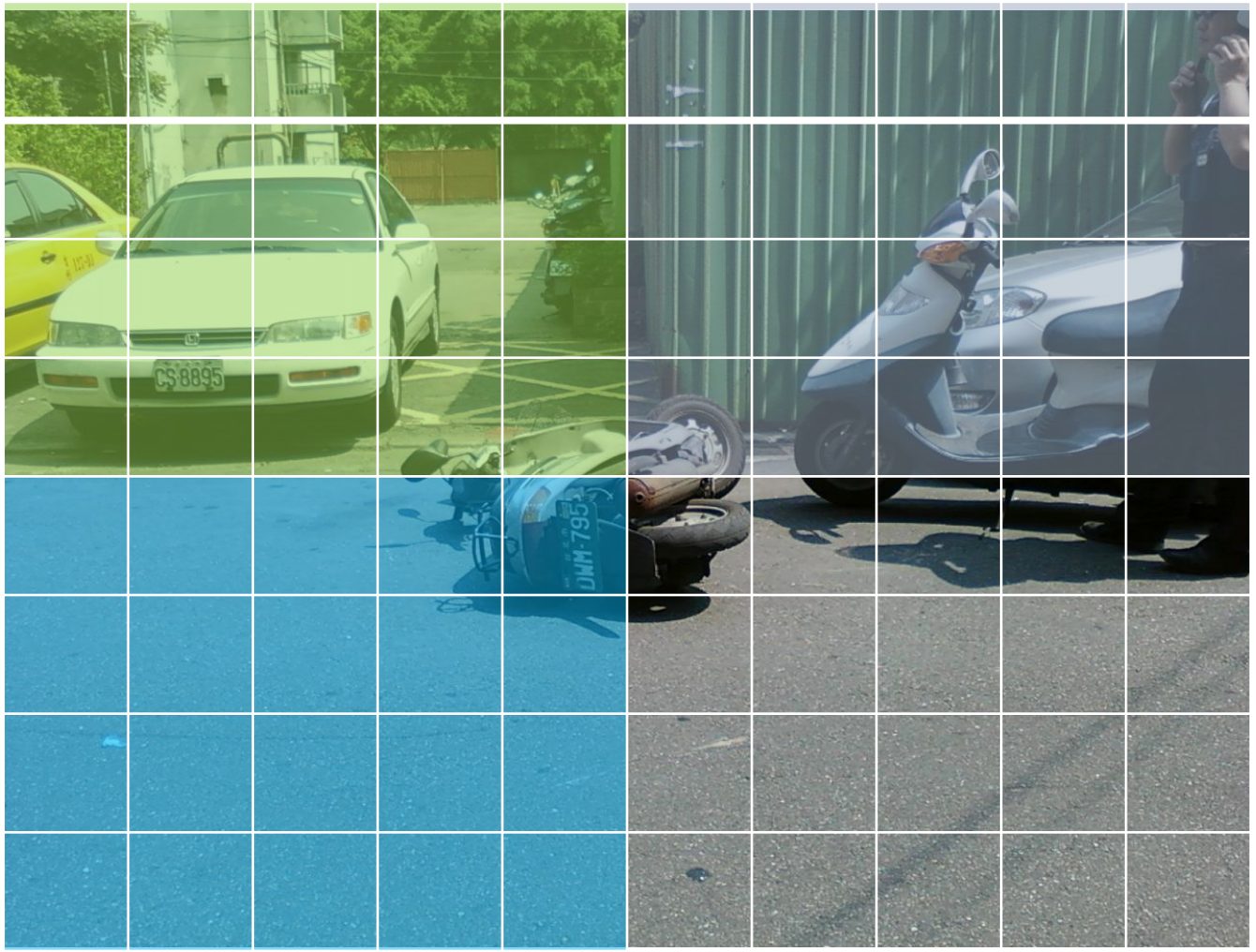
- Areas that have distinctive colors or patterns should obtain high saliency.
- frequently-occurring features should be suppressed.

- The salient pixels should be grouped together, and not spread all over the image.
- Faces can attract our attention more.

Local-global single-scale saliency

- Consider a single patch of scale *r* at each pixel. Thus, a pixel *i* is considered salient if the appearance of the patch pi centered at pixel *i* is distinctive with respect to all other image patches.

Patch image

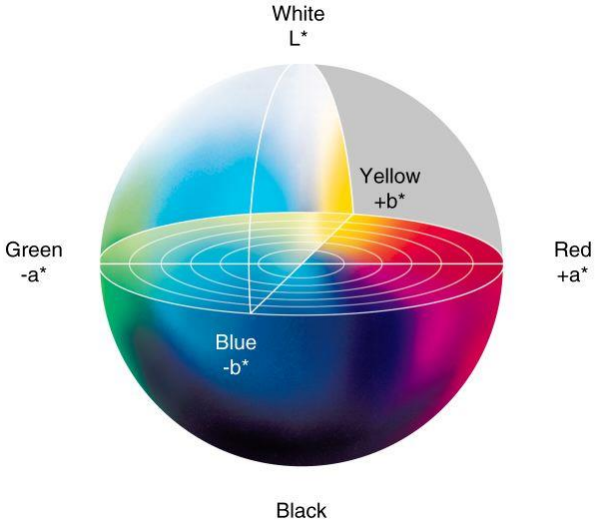


$\boldsymbol{D_{color}(p_i, p_j)}$

- Euclidean distance between the vectorized patches p_i and p_j in CIE L*a*b* color space.

$$d_c(p_i, p_j) = \sqrt{(L_i - L_j)^2 + (a_i - a_j)^2 + (b_i - b_j)^2}$$

Color hist



$D_{position}(p_i, p_j)$

- Euclidean distance between the positions of patches p_i and p_j.
  - a patch *pi* is salient when the patches similar to it are nearby;
  - it is less salient when the resembling patches are far away.

Dissimilarity

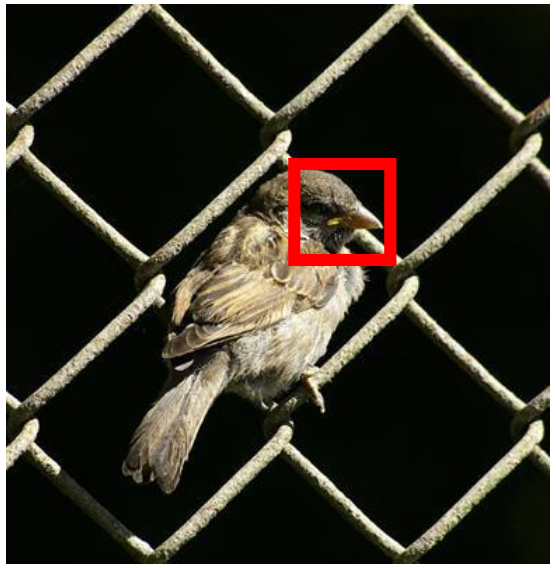
- *This dissimilarity measure is proportional to the difference in appearance and inverse proportional to the positional distance.*

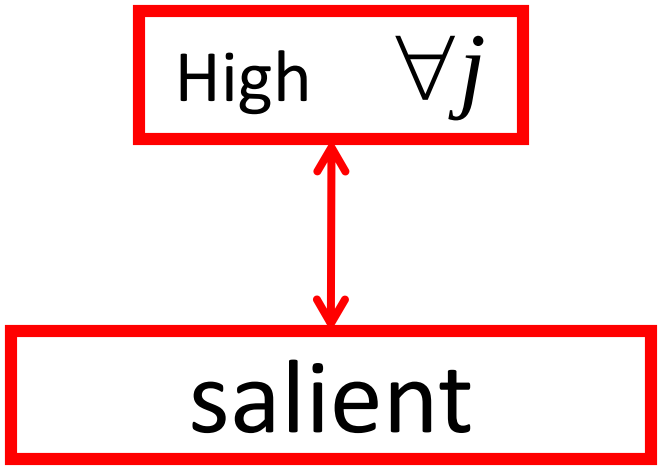
$$d(p_i, p_j) = \frac{d_{color}(p_i, p_j)}{1 + c \cdot d_{position}(p_i, p_j)}$$

Context-Aware

- Distance between a pair of patches:

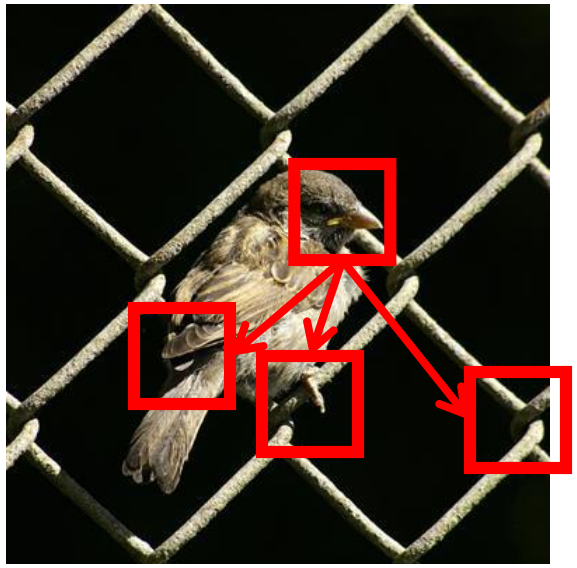
$$d(p_i, p_j) = \frac{d_{color}(p_i, p_j)}{1 + c \cdot d_{position}(p_i, p_j)}$$





- For every patch pi, we search for the K most similar patches $\{q_k\}_{k=1}^{K}$ in the image.

- $S_i^r = 1 - \exp\{-\frac{1}{K}\sum_{k=1}^{K} d(p_i^r, q_k^r)\}$



High for K most similar ones

Salient

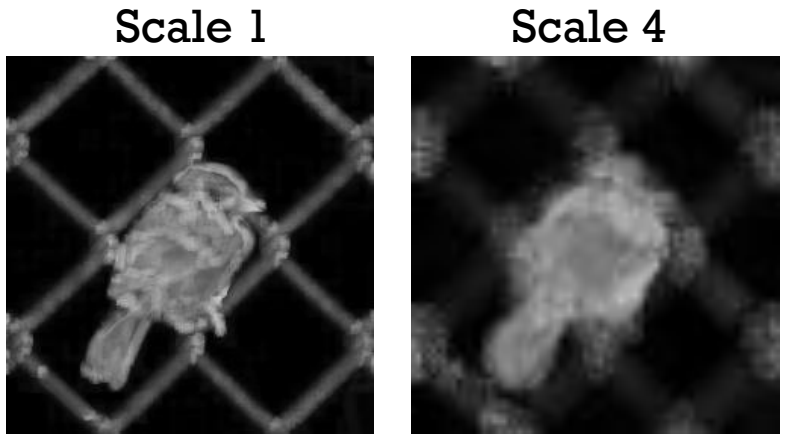
Multi-scale saliency enhancement

- Background pixels (patches) are likely to have similar patches at multiple scales.

- Incorporating multiple scales to further decrease the saliency of background pixels, improving the contrast between salient and non-salient regions.

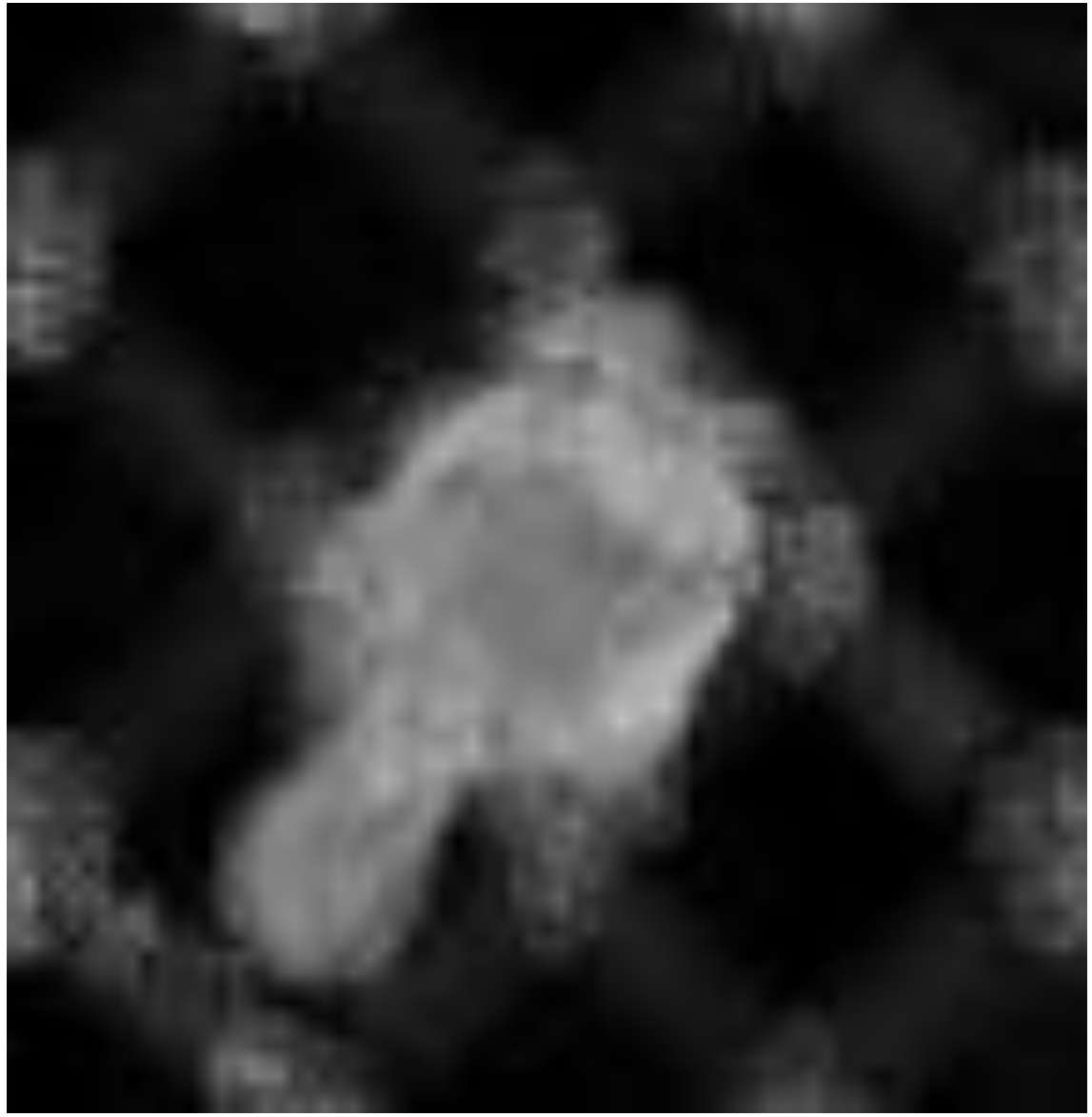
- For a patch pi of scale r, we consider as candidate neighbors all the patches in the image.
- We choose the K most similar patches to compute saliency.

$$S_i^r = 1 - \exp\{-\frac{1}{K}\sum_{k=1}^{K} d(p_i^r, q_k^{r_k})\}$$



- The saliency at pixel i is taken as the mean of its saliency at different scales:

$$\overline{S_i} = \frac{1}{M} \sum_{r \in R} S_i^r$$



Including the immediate context

- Gestalt laws:

visual forms may possess one or several centers of gravity about which the form is organized.

Simulate the visual contextual effect

- 1. A pixel is considered attended if its saliency value exceeds a certain threshold.

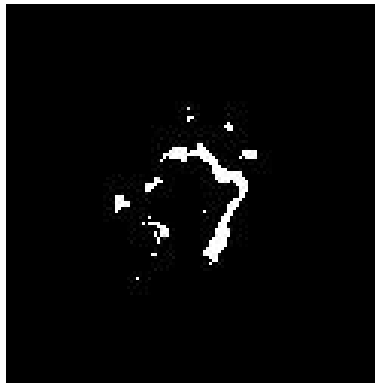
- 2. each pixel outside the attended areas is weighted according to its Euclidean distance to the closest attended pixel.

$$\hat{S}_i = \bar{S}_i(1 - d_{foci}(i))$$

Let $d_{foci}(i)$ be the Euclidean positional distance between pixel i and the closest focus of attention pixel, normalized to the range [0,1]

Context-Aware

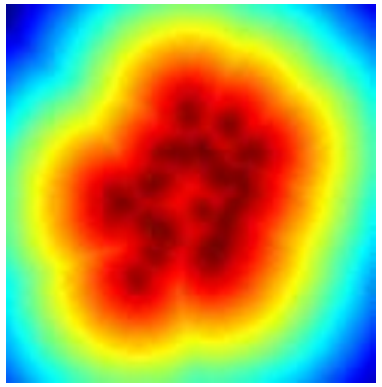
- Foci = $\bar{S}_i > 0.8$



- Include distance map

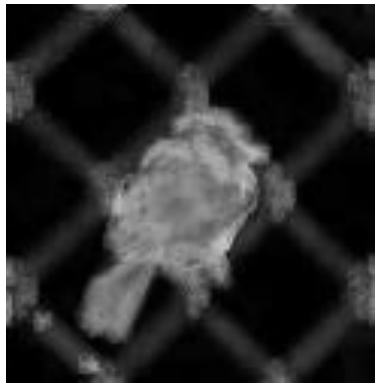
$$\hat{S}_i = \bar{S}_i \left(1 - d_{foci}(i)\right)$$

$1 - d_{foci}(i)$



X

$\bar{S}_i$



High-level factors

- Enhancement factors:

1. Recognized objects
2. Face detection

High-level factors

- face detection algorithm

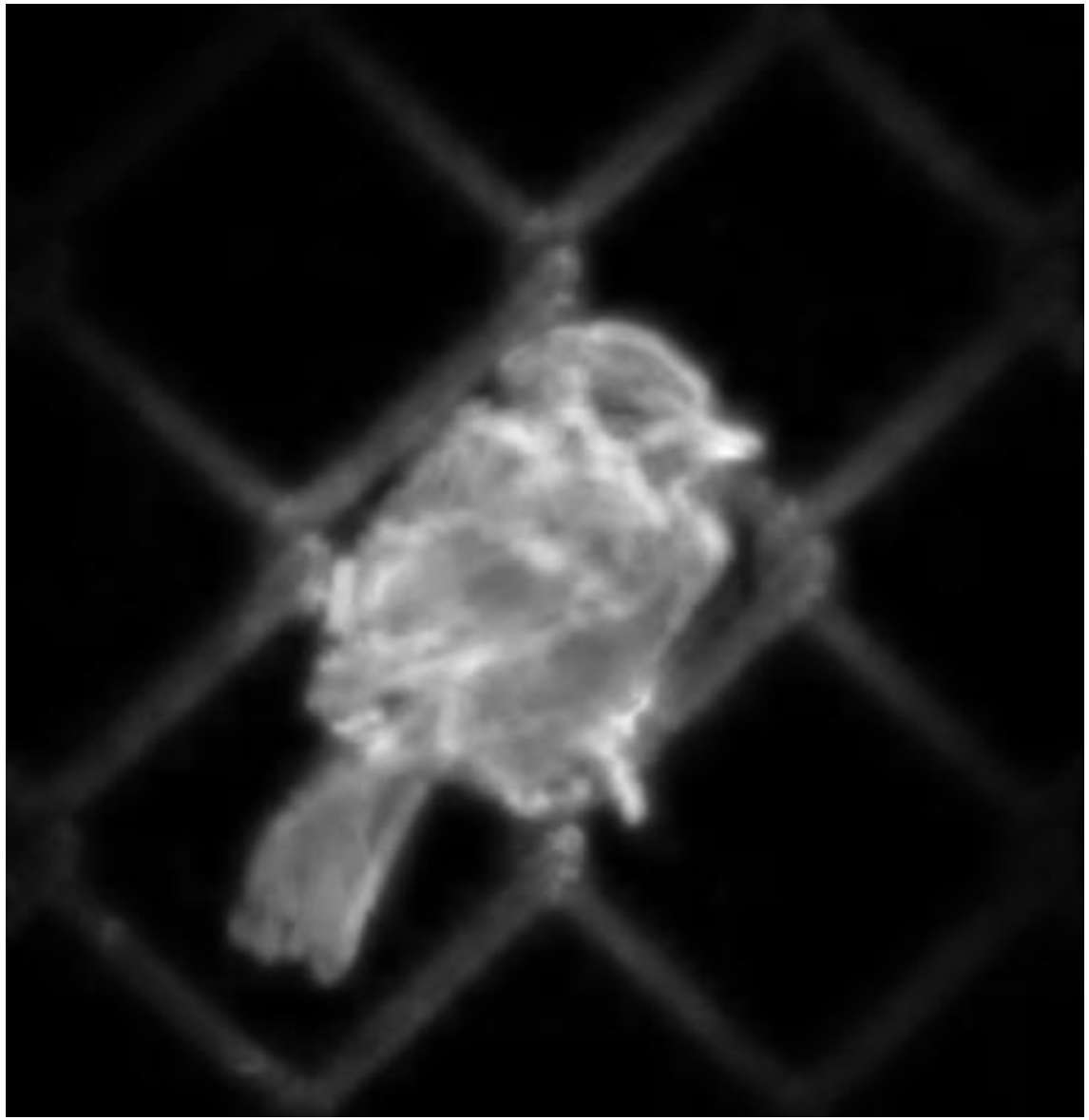
$$\hat{S}_i = \bar{S}_i(1 - d_{foci}(i))$$

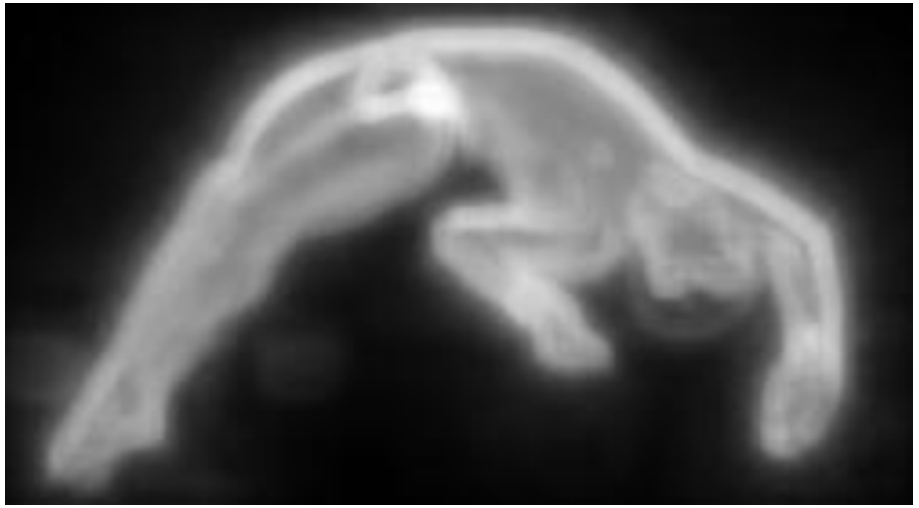
modified

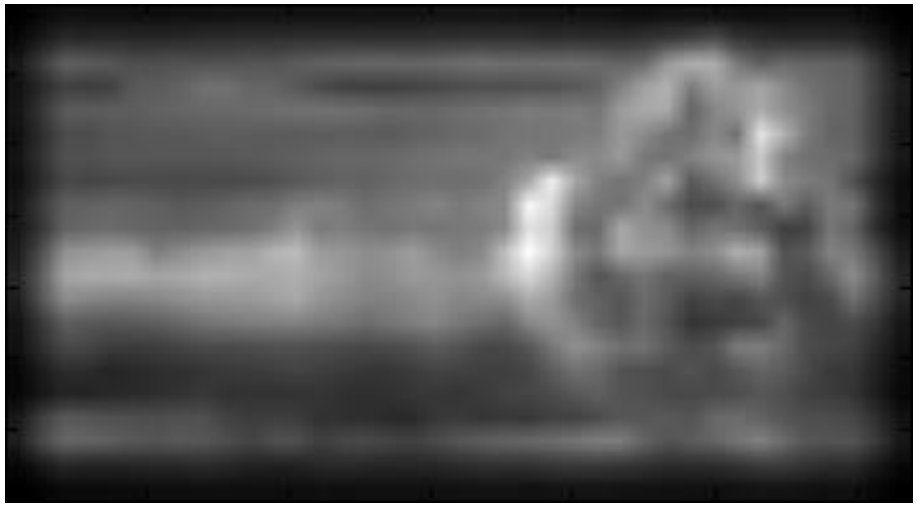
$$S_i = \begin{cases} Si & \text{, if } Si > face(i) \\ face(i) & \text{, otherwise} \end{cases}$$

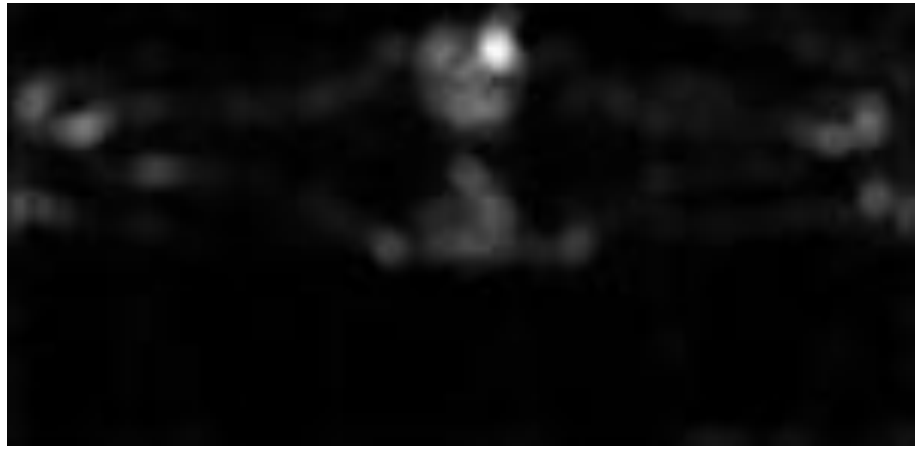
face detection algorithm of [23], which generates 1 for face pixels and 0 otherwise.

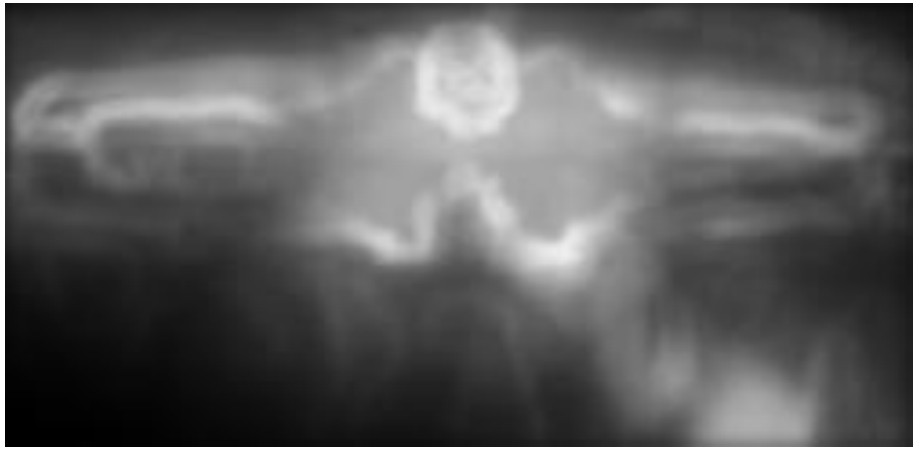
[23] P. Viola and M. Jones. Rapid Object Detection Using a Boosted Cascade of Simple Features. In CVPR, 2001.

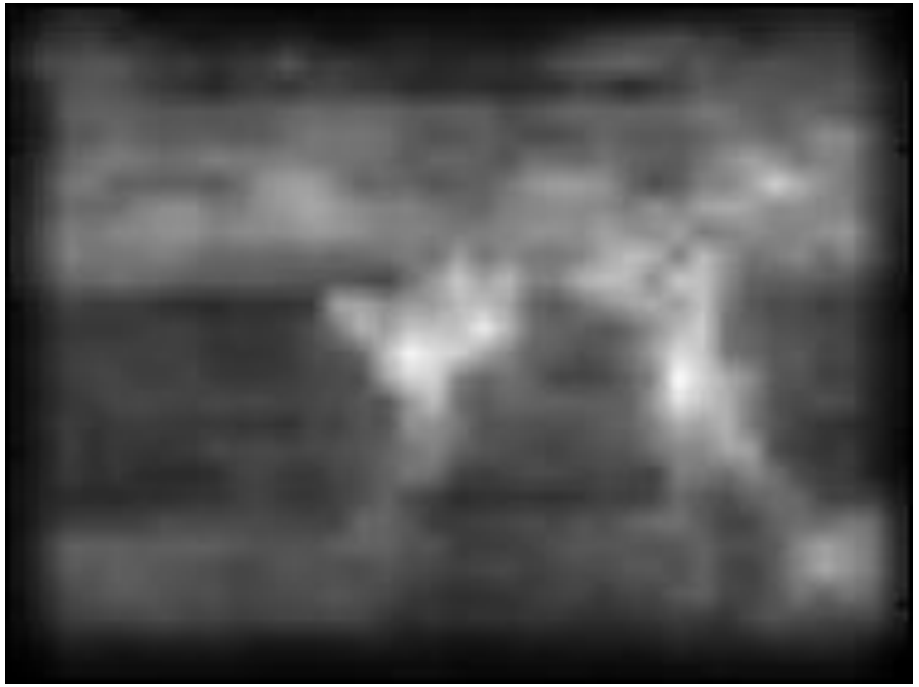


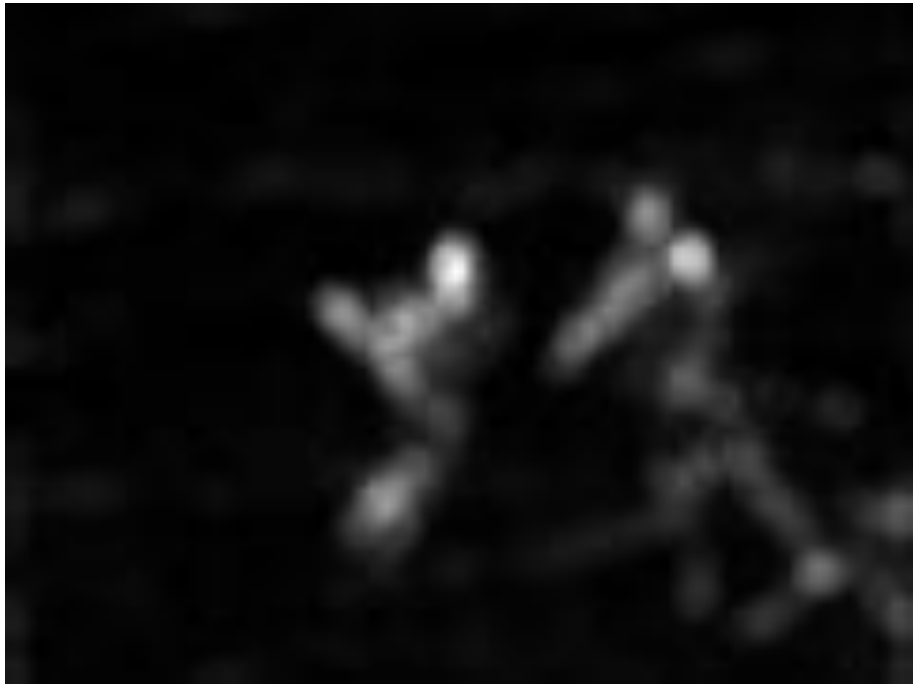


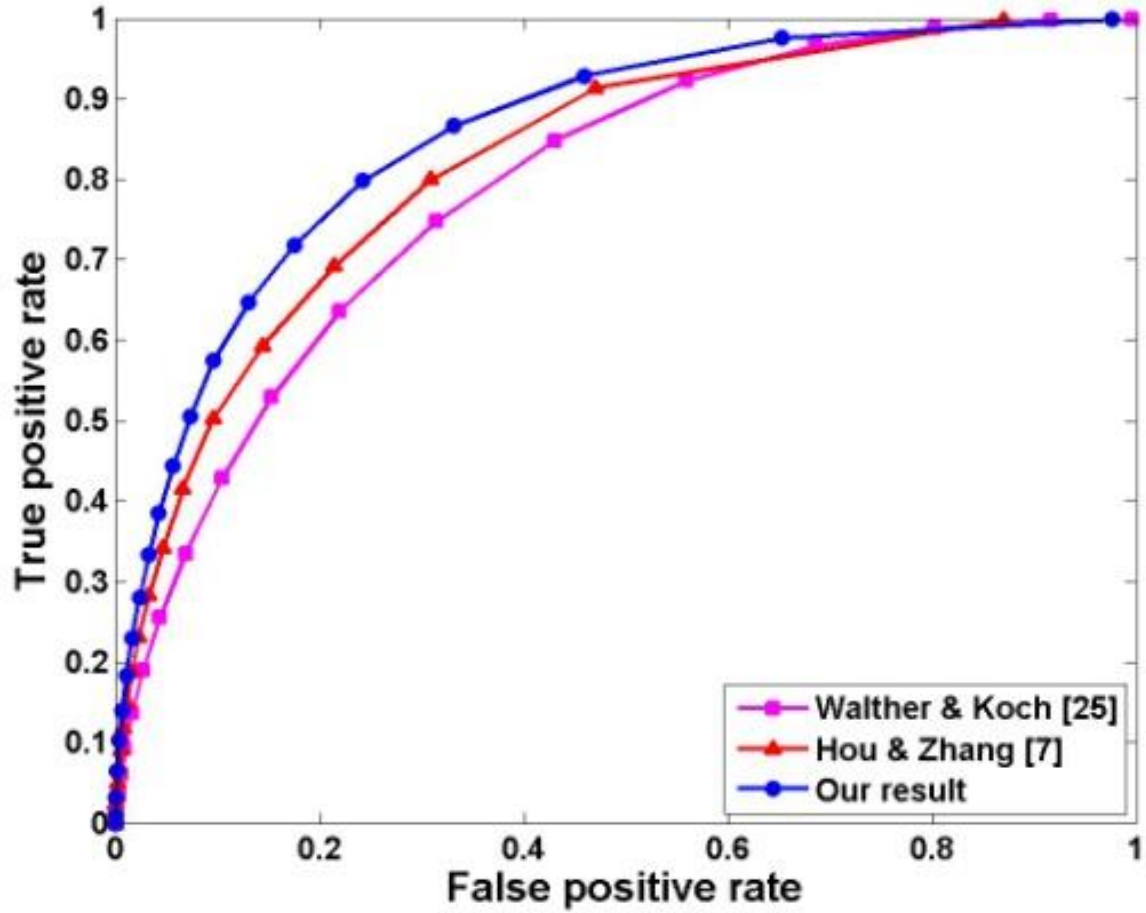











True positive rate
False positive rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0
0.2
0.4
0.6
0.8
1
Walther & Koch [25]
Hou & Zhang [7]
Our result

Applications

- Image retargeting

- Summarization through collage creation

Image retargeting

- Resizing an image by expanding or shrinking the non-informative regions.

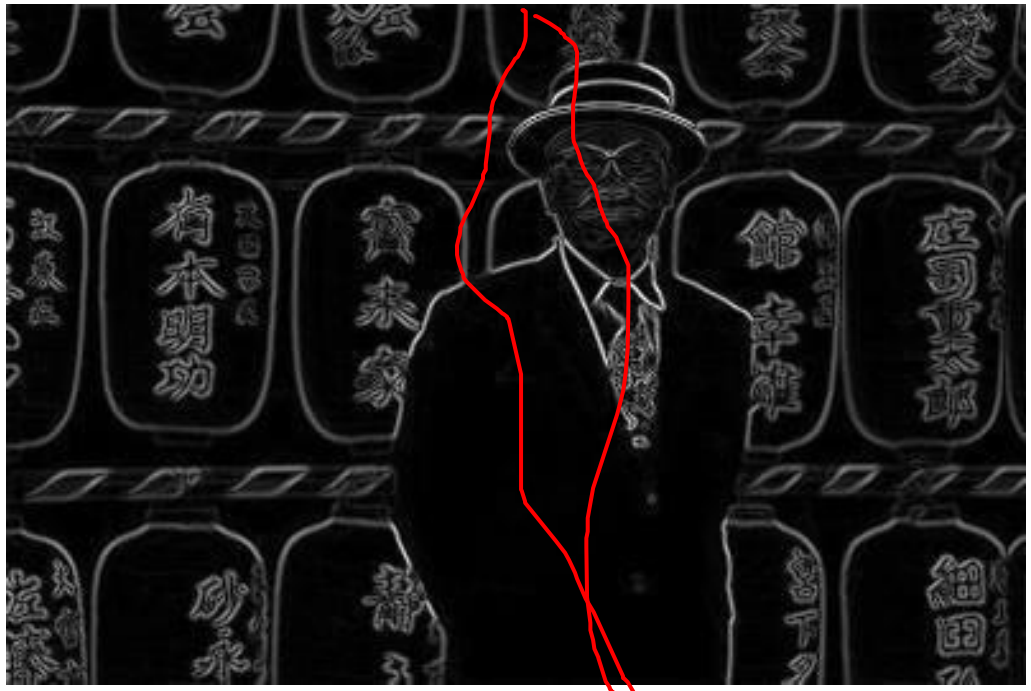
- Seam carving

M. Rubinstein, A. Shamir, and S. Avidan.

Improved seam carving for video retargeting. [2008]

- Context-aware saliency


有本明功
館幸雄
庄司重太郎
細田
砂永
静
宮下
左藤




庄司重太郎
館幸雄
細田
宮下タ


寶来家
館幸雄
庄司重太郎
宮下タ

Summarization through collage creation

- The salient objects as well as informative pieces of the background should be maintained in summaries.

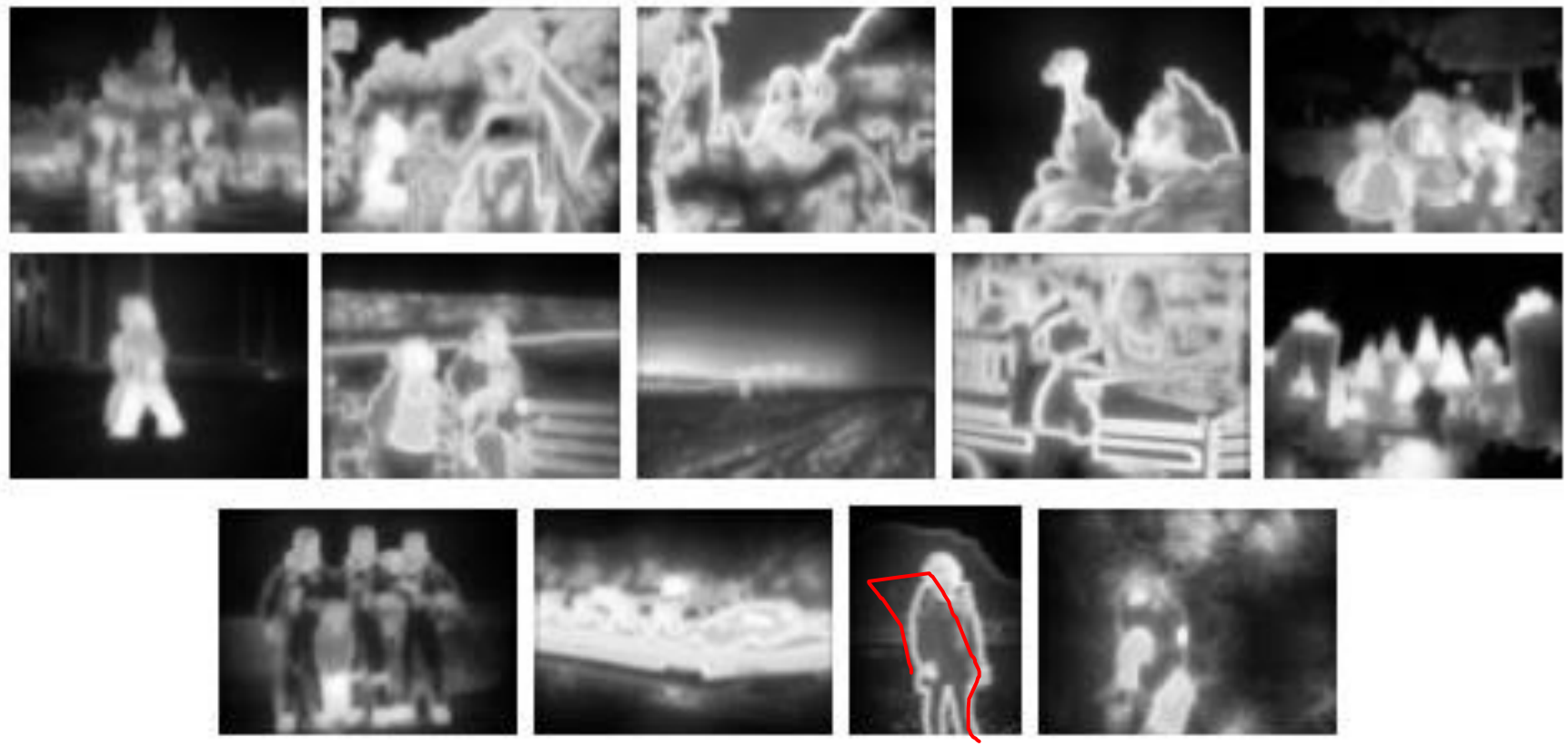
Automating collage creation

S. Goferman, A. Tal, and L. Zelnik-Manor.

Puzzle-like collage. [2010]

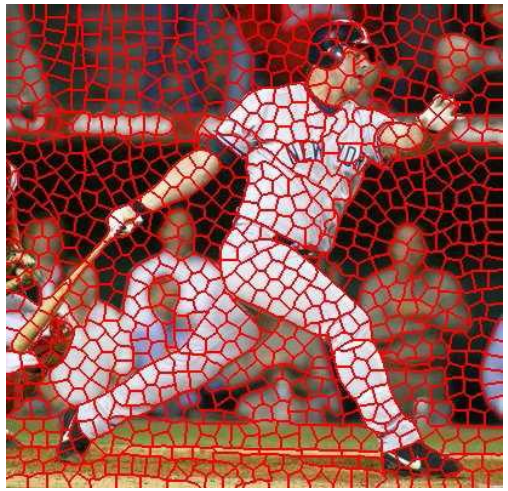
- 3 stages:

- Compute the saliency maps for images.
- Extracts regions-of-interest by considering both saliency and image edge information.
- Assemble non-rectangular ROIs.



How to improve saliency detection?

- Segmentation: over-segmentation (superpixel)



- How to evaluate the similarity between patches?
  - Sparse coding

$$\min_{\alpha} \|y - A\alpha\|_2^2 + \lambda\|\alpha\|_1$$

Homework 2

- What's the relationship/difference between eye fixation detection and salient object detection?
- How to get the ground-truth for salient object detection and eye fixation detection.
- List the main steps for spectral residual approach for saliency detection.
- Implement context-aware saliency detection and spectral residual method
  - Test them with some images
  - Submit the codes, report (including the details of your implementation, the problems you find, the tricks that help improve saliency and your reasons, and results. In English. CVPR format)

- Email:cv2_hw@sina.com
- Due date: March 15