# Cluster-Based Co-Saliency Detection report

## Paper's Authors

Huazhu Fu, Xiaochun Cao, and Zhuowen Tu

## Advisor

Dr. Maryam Abedi

## Student

Mohammad Shahpouri

**October**
**2022**

# Contents


**List of Figures** ii

**List of Equations** iii

**List of Algorithms** iv

**1 Proposed method** 1

1.1 Cluster-Based Method . . . 2

1.2 Cluster-Based Saliency Cues . . . 2

**2 The Co-Saliency Maps** 3

**3 Experimental results** 3

3.1 Quantitative result . . . 4

3.2 Qualitative result . . . 4

**References** 5

# List of Figures


1 Comparison between the cluster based saliency detection methods. . . . . . . . . . . . . . . . . . . . 4

2 Visual comparison of single image saliency detection on MSRA1000 dataset. (a) Input image. (b) Ground truth. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

## List of Equations


1 Equation 1 . . . . . . . . . . 2

2 Equation 2 . . . . . . . . . . 2

3 Equation 3 . . . . . . . . . . 2

4 Equation 4 . . . . . . . . . . 3

5 Equation 5 . . . . . . . . . . 3

6 Equation 6 . . . . . . . . . . 3

7 Equation 7 . . . . . . . . . . 3

## List of Algorithms

1 Cluster-based Co-Saliency Detection . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 1

# 1 Proposed method

Two-layer clustering method is proposed. One layer groups the pixels on each image (single image), and the other layer associates the pixels on all images (multi-image). The saliency cues for each cluster are computed, and measure the cluster-level saliency. Contrast, spatial, and corresponding cues form the measured features. At last, based on these cluster-level cues, the saliency value for each pixel is calculated, which is used to generate the final saliency map.

---

**Algorithm 1:** Cluster-based Co-Saliency Detection

---

**Input :** The input image set$\{I^j\}_{j=1}^M$, intra cluster number $K_1$, and inter cluster number $K_2$.

**1. Single image saliency detection:**

**for** *each image* **do**

  Clustering image into $K_1$ clusters;

  **for** *each cluster* **do**

    Computing the contrast cue using Equation 1 and spatial cue using Equation 2;

  **end**

  Combining two saliency cues using Equation 5;

  **for** *each pixel* **do**

    Obtaining the final single saliency map using Equation 7

  **end**

**end**

**2. Miltiple image co-saliency detection:**

Clustering all images into $K_2$ clusters;

**for** *each cluster* **do**

  Computing the contrast cue using Equation 1, spatial cue using Equation 2, and corresponding cue using Equation 4;

**end**

Combining the single saliency map and three cluster-based cues using Equation 5;

**for** *each pixel* **do**

  Assigning the final co-saliency map using Equation 7;

**end**

**Output:** The single saliency map and co-saliency map.

---

## 1.1 Cluster-Based Method

K-means and a probability framework are adopted to soft assign the co-saliency value to each pixel.

## 1.2 Cluster-Based Saliency Cues

Contrast, spatial, and corresponding cues are features to compute the saliency map.

**Notations:** The pixel is denoted by $\{p_i^j\}_{i=1}^{N_j}$ with index $i$ in the image $I^i$, where the $N_j$ denotes the $j$th image lattice. $\{z_i^j\}_{i=1}^{N_j}$ denotes the normalized location of the pixel $p_i^j$ in the image $I^j$. $K$ clusters $\{C^k\}_{k=1}^{K}$ are obtained from $M$ images $\{I^j\}_{j=1}^{M}$. The clusters are denoted by a set of D-dimensional vectors $\{\mu^k\}_{k=1}^{K}$, in which $\mu^k$ denotes the prototype (cluster center) associated with the cluster $C^k$. And the function $b\colon \mathbb{R}^2 \to \{1 \dots K\}$ associates the pixel $p_i^j$ and the cluster index $b(p_i^j)$.

*1) Contrast Cue:* Contrast cue represents the visual feature uniqueness on the single or multiple images [1] [2] [3]. The contrast operator simulates the human visual receptive fields. The contrast cue $w^c(k)$ of cluster $C^k$ is defined using its feature contrast to all other clusters:

$$w^c(k) = \sum_{i=1, i\neq k}^{K} \left( \frac{n^i}{N} \|\mu^k - \mu^i\|_2 \right) \tag{1}$$

where a $L_2$ norm is used to compute the distance on the feature space, $n^i$ represents the pixel number of cluster $C^i$, and $N$ denotes the pixel number of all images.

*2) Spatial Cue:* In human visual system, the regions near the image center draw more attention than the other regions [4]-[5]. When the distance between the object and the image center increases, the attention gain is depreciating.

$$w^s(k) = \frac{1}{n^k} \sum_{j=1}^{M} \sum_{i=1}^{N_j} \left[ \mathcal{N}\left( \|z_i^j - o^j\|^2 | 0, \sigma^2 \right) \cdot \delta[b(p_i^j) - C^k] \right] \tag{2}$$

where $\delta(\cdot)$ is the Kronecker delta function, $o^j$ denotes the center of image $I^j$, and Gaussian kernel $\mathcal{N}(\cdot)$ computes the Euclidean distance between pixel $z_i^j$ and the image center $o^j$, the variance $\sigma^2$ is the normalized radius of images. And the normalization coefficient $n^k$ is the pixel number of cluster $C^k$.

*3) Corresponding Cue:* is presented to measure how the cluster distribute on the multiple images. A $M$-bin histogram $\hat{\mathbf{q}}^k = \{\hat{q}_j^k\}_{j=1}^{M}$ is adopted to describe the distribution of cluster $C$ in $M$ images:

$$\hat{q}_j^k = \tfrac{1}{n^k} \textstyle\sum_{i=1}^{N_j} \delta[b(p_i^j) - C^k], \quad j = [1 \dots M] \tag{3}$$

where $n^k$ is the pixel number of cluster $C^k$, which enforces the condition $\sum_{j=1}^{M} \hat{q}_j^k = 1$. Then, the corresponding cue $w^d(k)$ is defined as:

$$w^d(k) = \frac{1}{\mathrm{var}(\hat{\mathbf{q}}^k) + 1} \tag{4}$$

where $\mathrm{var}(\hat{\mathbf{q}}^k)$ denotes the variance of histogram $\mathbf{q}^k$ of the cluster $C^k$

# 2 The Co-Saliency Maps

Multiplication operation is employed to integrate the saliency cues.

Before combining saliency cues, each cue map normalized to standard Gaussian using the distribution of scores across all clusters. Then the cluster-level co-saliency probability $p(k)$ of cluster $k$ is defined as:

$$p(C^k) = \prod_i w_i(k) \tag{5}$$

where $w_i(k)$ denotes saliency cues.

The saliency likelihood of the pixel $x$ belonging to the cluster $C^k$ satisfies a Gaussian distribution $\mathcal{N}$ as:

$$p(x|C^k) = \mathcal{N}\left(\|v_x, \mu^k\|_2 | 0, \sigma_k^2\right) \tag{6}$$

where $v_x$ denotes the feature vector of pixel $x$, and the variance $\sigma_k$ of Gaussian uses the variance of cluster $C^k$. Hence, the marginal saliency probability $p(x)$ is obtained by summing the joint saliency $p(C^k)p(x|C^k)$ over all clusters:

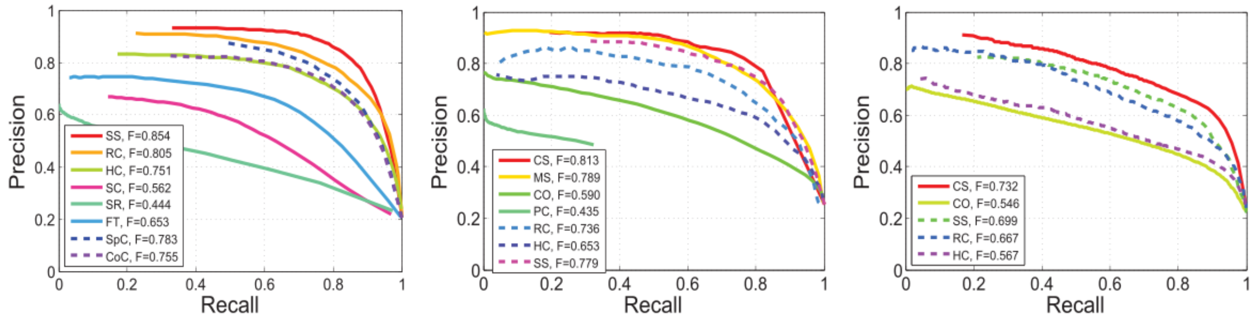
$$p(x) = \sum_{k=1}^{K} p(x, C^k) = \sum_{k=1}^{K} p(x|C^k)p(C^k) \tag{7}$$

Finally, the pixel-level co-saliency is obtained.

# 3 Experimental results

MSRA1000 [6] dataset and Co-saliency Pairs dataset in [7] are used to evalute the performance of proposed method.

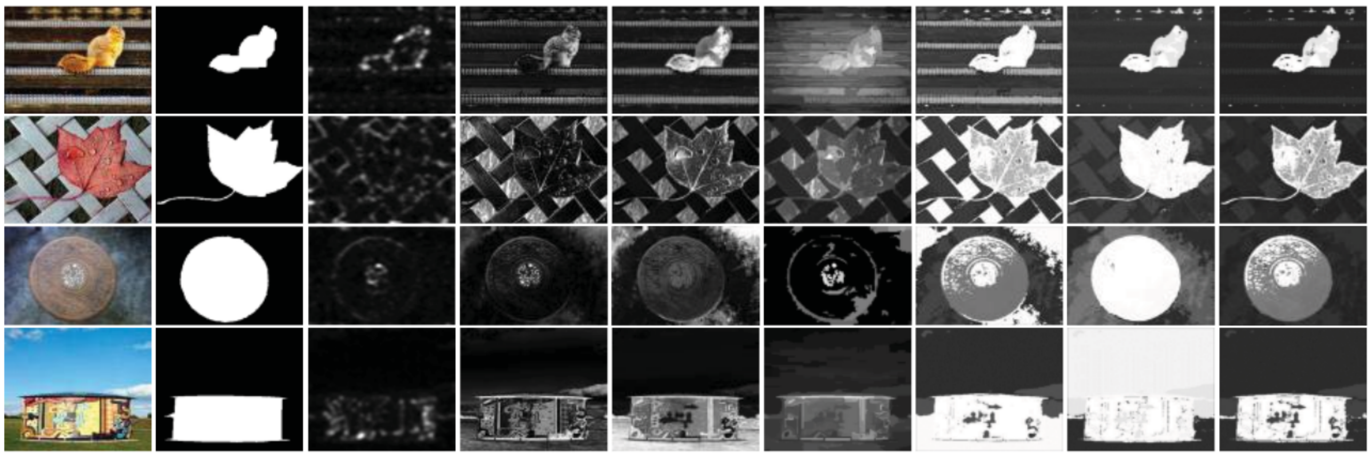
## 3.1 Quantitative result




(a) Saliency on single image (b) Co-saliency on image pair (c) Co-saliency on multi-image

**Figure 1.** Comparison between the cluster based saliency detection methods.

## 3.2 Qualitative result



(a) Image (b) Ground truth (c) SR (d) SC (e) HC (f) RC (g) Contrast cue (h) Spatial cue (i) Our method

**Figure 2.** Visual comparison of single image saliency detection on MSRA1000 dataset. (a) Input image. (b) Ground truth.

## References


[1] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998. 2

[2] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *CVPR 2011*, pp. 409–416, 2011. 2

[3] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 2, pp. 353–367, 2011. 2

[4] L. Duan, C. Wu, J. Miao, L. Qing, and Y. Fu, "Visual saliency detection by spatially weighted dissimilarity," in *CVPR 2011*, pp. 473–480, 2011. 2

[5] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *2009 IEEE 12th International Conference on Computer Vision*, pp. 2106–2113, 2009. 2

[6] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1597–1604, 2009. 3

[7] H. Li and K. N. Ngan, "A co-saliency model of image pairs," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3365–3375, 2011. 3