IMAGE CO-SALIENCY DETECTION BY PROPAGATING SUPERPIXEL AFFINITIES

Zhiyu Tan[1], *Liang Wan*[2], *Wei Feng*[1,*], *Chi-Man Pun*[3]

[1] School of Computer Science and Technology, Tianjin University, Tianjin, China
[2] School of Computer Software, Tianjin University, Tianjin, China
[3] Faculty of Science and Technology, University of Macau, Macau, China
{tzy,lwan,wfeng}@tju.edu.cn, cmpun@umac.mo

ABSTRACT

Image co-saliency detection is a valuable technique to highlight perceptually salient regions in image pairs. In this paper, we propose a self-contained co-saliency detection algorithm based on superpixel affinity matrix. We first compute both intra and inter similarities of superpixels of image pairs. Bipartite graph matching is applied to determine most reliable inter similarities. To update the similarity score between every two superpixels, we next employ a GPU-based all-pair SimRank algorithm to do propagation on the affinity matrix. Based on the inter superpixel affinities we derive a co-saliency measure that evaluates the foreground cohesiveness and locality compactness of superpixels within one image. The effectiveness of our method is demonstrated in experimental evaluation.

Index Terms— Image co-saliency detection, superpixel affinity propagation, bipartite graph matching, all-pair SimRank, co-saliency measure

1. INTRODUCTION

Visual saliency modeling simulates the perceptual stimuli of human vision system in detecting highly salient regions from their surroundings. It has been extensively studied in recent years, and successfully used in many applications [1][2][3]. Yet, most existing works focus on saliency map estimation for a single image. Image co-saliency modeling that detects co-salient regions in an image pair is less investigated [4]. Our focus in this paper is proposing a new co-saliency model by globally propagating the intra/inter superpixel affinities and measuring co-saliency maps based on the affinity matrix.

Relation to prior works. Saliency-based visual attention model was early proposed in [5] which evaluates several features such as color, intensity, and orientation, and then integrates them to form a scalar map. This work was further extended by adding other local features [6]. Besides these cognitive models, many computational methods have been developed over the past years, including frequency space methods [7] [8], probabilistic models [9] and learning-based models [10]. Recently, Perazzi et al. [11] developed a new contrast-based model called saliency filter. It decomposes an image into superpixels and computes the saliency map based on two measures of contrast that rate the uniqueness and the spatial distribution of the superpixels.

Image co-saliency detection, on the other hand, is to detect similar salient objects from image pairs. In the work of [4], a co-multilayer graph is constructed with region similarity computed both within and among images. Each region is described by color and texture features. Then single-pair SimRank [12] is adopted to computes the similarity between two regions. In the end, the co-saliency score is computed as a linear combination of the maximal inter similarity value and other three single-image saliency scores [5][13][7].

Our contributions. In this paper, we present a novel co-saliency detection algorithm by fully exploiting the superpixel affinity matrix for an image pair. The first contribution of our method is a self-contained framework, without dependence on single-image saliency maps. Our method actually converts the co-saliency detection problem to saliency detection, thus being able to adapt well-known saliency detection techniques. The second contribution is the new co-saliency model, which evaluates the co-salient map of one image according to its matching in the other image. By this way, we are able to estimate cosaliency maps according to the inter affinity matrix directly. The third contribution is the application of bipartite graph matching on superpixels across image pairs. This produces a good initialization for the followed affinity propagation. Experiments demonstrate that our method achieves a better performance than existing co-saliency and saliency approaches, especially for scenes with complex background.

2. ALGORITHM

The framework of our method is illustrated in Fig. 1. In the following we describe the details of each key component.

* is the corresponding author. This work is supported by the Program for New Century Excellent Talents in University (NCET-11-0365), the National Natural Science Foundation of China (61100121 and 61100122), the research fund for Doctoral Program of Higher Education (20110032120036 and 20110032120041), and the Science and Technology Development Fund of Macau (FDCT 034/2010/A2).

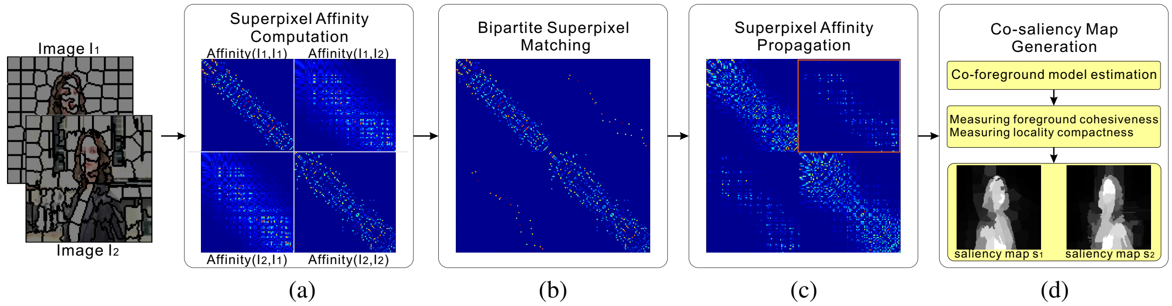


Fig. 1. Our framework. (a) Compute intra and inter superpixel affinity matrices; (2) Conduct bipartite superpixel matching on inter affinity matrix, which produces sparse and reliable affinities; (3) Update the affinity matrix via superpixel affinity propagation; and (4) Generate co-saliency maps from the updated inter affinity matrix.

2.1. Superpixel Affinities

We start by decomposing images into perceptually homogenous regions (superpixels) using SLIC algorithm [14]. A superpixel is described by two types of features, i.e. color distribution and center position. Here, we adopt the color feature from [4], which represents a superpixel as a 100-dimension color histogram by k-meaning clustering color vectors of pixels in the concatenation of RGB, Lab and YCbCr color spaces. We then define the affinity between two superpixels, v_i and v_j, to be a combination of their color feature similarity $C(i,j)$ and position similarity $P(i,j)$, given by

$$\begin{aligned} S_f(i,j) &= \alpha C(i,j) \cdot \beta P(i,j), \\ C(i,j) &= \exp(-\chi^2(c_i, c_j)), \\ P(i,j) &= \exp(-||p_i - p_j||^2), \end{aligned} \tag{1}$$

where c_i and c_j denote color features for v_i and v_j, respectively; $\chi^2(\cdot)$ denotes the Chi-square distance, i.e. $\chi^2 = \sum_{z=1}^{100} \frac{(c_i(z)-c_j(z))^2}{c_i(z)+c_j(z)}$. $P(i,j)$ is exponential function to Euclidean distance between center positions of v_i and v_j. The parameters α and β control the sensitivity of color and position similarities, empirically set as $\alpha = 1$ and $\beta = 0.0005$.

Suppose images I_1 and I_2 are decomposed into M and N superpixels. The overall affinity matrix is composed of the intra affinity matrices for each image in the diagonal direction, and inter affinity matrices across images in the anti-diagonal direction, formulating an $(M+N) \times (M+N)$ matrix (see Fig. 1(a)). For the intra affinity matrices, we compute affinities only for those superpixels being spatially adjacent. For the inter affinity matrices, we consider every superpixel pair, with one superpixel from image I_1 and the other from image I_2. Also note the inter affinity matrix from I_2 to I_1 is transpose to that from I_1 to I_2.

2.2. Bipartite Superpixel Matching

Since our co-saliency measure is derived from the inter affinity matrix, a good initialization is critical to subsequent algorithm. For this purpose, we propose to apply bipartite graph matching on the inter affinities. We first construct a bipartite graph over superpixels from two images, as shown in Fig. 2. Formally, let $G = \{X, Y, B\}$ be a bipartite graph with node set $X \cup Y$, where $X = \hat{X} \cup D_X$ and $Y = \hat{Y} \cup D_Y$. $\hat{X}$ and $\hat{Y}$ are real superpixels in images I_1 and I_2, respectively (see the colored squares in Fig. 2), while D_X and D_Y are dummy nodes (the black squares) to make X and Y have an equal size. For generality, we set the number of dummy nodes $|D_X| = N$ and $|D_Y| = M$. The across-affinity matrix $B = \{b_{ij}\}$ between X and Y is constructed as follows:

$$b_{ij} = \begin{cases} S_f(i,j), \text{ if } v_i \in \hat{X} \text{ and } v_j \in \hat{Y}; \\ 0, \text{ if } v_i \in D_X \text{ and } v_j \in D_Y; \\ \eta \cdot \text{median}_{kl} S_f(k,l), \text{ otherwise}, \end{cases} \tag{2}$$

where η is a weight controlling the effect of dummy nodes.

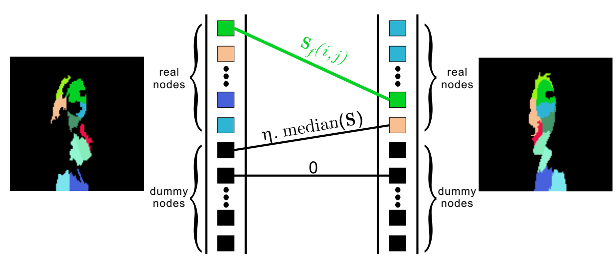


Fig. 2. Bipartite superpixel matching. By the construction of bipartite graph, the optimal matching on images serves as a good initialization for affinity propagation.

Given the bipartite graph, we solve the optimal graph matching in polynomial time using Hungarian algorithm. The bipartite graph matching gives us an optimal one-to-one superpixel matching at the group level, better than searching optimal match for each superpixel independently. After graph matching, we select valid and reliable matches as those connecting with two real nodes and having affinities larger than a threshold, i.e. $\{S_f(i,j) | v_i \in \hat{X}, v_j \in \hat{Y}, S_f(i,j) > \tau\}$. One matching example is shown in Fig. 2.

Our bipartite superpixel matching allows us to rely on the most reliable affinities for the following affinity propagation, which is helpful to guarantee the computing accuracy. Additionally, the generated inter affinity matrix is sparse in nature, leading to high computing efficiency in propagation.

2.3. Superpixel Affinity Propagation

Based on the initialized affinity matrix, we construct a relational graph $G_r = \{V, E\}$. Its node set V is the group of all the superpixels, i.e., $V = \hat{X} \cup \hat{Y}$. The edge set E contains two parts, the spatially neighboring connections in each image and the matching connection due to the bipartite superpixel matching. The edge weights are from Eq. 1.

We now update the superpixel affinities through propagating reliable affinities on graph G_r. Basically, we encourage two nodes to be similar if their neighbors have high similarities. This can be defined as follows,

$$S_p^{(t+1)}(a,b) = \frac{\gamma}{|\mathcal{N}(a)||\mathcal{N}(b)|} \sum_{i \in \mathcal{N}(a)} \sum_{j \in \mathcal{N}(b)} S_p^{(t)}(i,j), \quad (3)$$

where $\gamma \in [0, 1]$ is a decay factor; $|\mathcal{N}(a)|$ and $|\mathcal{N}(b)|$ denote the numbers of neighbors $\mathcal{N}(a)$ and $\mathcal{N}(b)$ for nodes a and b.

The propagation can be achieved by utilizing SimRank algorithm [12] on the graph. As we need to udpate the similarity $S_p(a, b)$ for every two superpixels, we adopt all-pair SimRank algorithm. It is noticed that Eq. 3 is highly parallelizable, since for two node pairs, (a, b) and (c, d), the computation of similarities $S_p(a, b)$ and $S_p(c, d)$ is independent in one iteration. In practice, we do implementation using CUDA architecture, ensuring a very fast timing performance.

2.4. Superpixel Co-saliency Generation

After affinity propagation, we extract co-saliency maps from the inter affinity matrix. The intuition of our approach is evaluating the co-saliency map of one image according to its matching in the other image. By this way, we actually convert co-saliency detection problem to saliency detection. One notable advantage is that we can adapt successful single-image saliency techniques for co-saliency detection. In our work, we borrow the idea of saliency filter [11] and derive a co-saliency score by evaluating two measures, i.e. foreground cohesiveness and locality compactness.

To be specific, we may use row vectors of the inter affinity matrix as feature vectors for the corresponding superpixels. Let us take the inter affinity matrix Affinity(I_1, I_2) from I_2 to I_1 for example. Given superpixels v_i and v_k of image I_1, their row vectors $\mathbf{s}_i$ and $\mathbf{s}_k$ have three possible cases: a) Both have low similarities, indicating v_i and v_k belong to background; b) They have close matching in I_2, indicating v_i and v_k belong to foreground; c) They have different matching in I_2. In this case, the distance between $\mathbf{s}_i$ and $\mathbf{s}_k$ can be very small, may leading to wrong results. To solve this problem, we construct a co-foreground model and based on it measure the saliency of superpixels.

It is noticed that the inter affinity matrix indeed reveals how similar regions distribute in two images. Herein a co-salient foreground model can be defined as the normalized summation of row vectors of the inter affinity matrix, i.e.

$$\mathbf{F}(j) = \frac{1}{Z} \sum_{i=1}^{M} S_p(i,j), j = 1, \ldots, N. \quad (4)$$

where Z is the normalization factor, ensuring $\sum_j F(j) = 1$. $\mathbf{F}$ actually reflects the probability distribution of co-salient foreground in image I_1 with respect to image I_2. We can now evaluate the likelihood that superpixels in image I_1 are produced from the foreground model. This is given by

$$\mathbf{g}(i) = \mathbf{F}^T \mathbf{s}_i, \quad (5)$$

where $\mathbf{s}_i$ is the i-th row vector in Affinity(I_1, I_2) .

Based on the likelihood, we further define two saliency measures. First, the *foreground cohesiveness* is defined as the likelihood of i-th superpixel being in salient region according to appearance cue,

$$A_i = \sum_{k=1}^{M} exp(\mathbf{g}(i)\mathbf{g}(k)) \cdot w_{ik}^p, \quad (6)$$

where $exp(\mathbf{g}(i)\mathbf{g}(k))$ means if v_i and v_k have similar high matching in I_2, they probably belong to salient regions simultaneously. $w_{ik}^p = \frac{1}{Z_i} exp(-\frac{1}{2\sigma_p^2}||\mathbf{p}_i - \mathbf{p}_k||^2)$ controls the influence radius in the spatial domain. Z_i is the normalization factor. Second, we measure the *locality compactness* of foreground objects by means of weighted spatial variance,

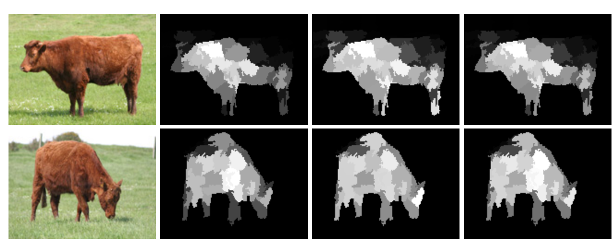
$$L_i = \sum_{k=1}^{M} ||\mathbf{p}_k - \mu_i||^2 \cdot w_{ik}^c \quad (7)$$

where $\mathbf{p}_k$ is the center position of v_k; $\mu_i = \sum_{k=1}^{M} w_{ik}^c \mathbf{p}_k$ is the weighted mean position of v_i, and $w_{ik}^c = \frac{1}{Z_i} exp(\mathbf{g}(i)\mathbf{g}(k))$.

In the end, we combine the above two measures yielding a co-saliency value $S_c(i)$ for each superpixel:

$$S_c(i) = A_i + L_i. \quad (8)$$

The saliency values are normalized finally to the range $[0..1]$. Fig. 3 shows one example.



(a) (b) (c) (d)

Fig. 3. Co-saliency computation. (a) Image pair; (b) Co-foreground cohesiveness; (c) Locality compactness; (d) The resulting co-saliency map.

Fig. 4. Comparative performance of co-saliency detection of the proposed approach and state-of-the-art methods. From top to bottom are input image pairs, our results, results from CM [4], SF [11], HC [15] and RC [15].

3. EXPERIMENTAL RESULTS

In this section, we verify the performance of our co-saliency algorithm on several image pairs, which were used in [4]. Fig. 4 shows the comparison results of our method, existing co-saliency method (eg, CM [4]) and several state-of-art saliency methods, including saliency filter (SF) [11], (HC) [15] and (RC) [15]. As we can see, our method can generate object boundaries more complete and clean than those from CM. For single-image saliency detection, SF is reported to outperform most state-of-art methods [11]. In our experiment, we found that it works well for images with salient objects and relatively simple background, yet may produce obvious artifacts in case of complex background. As shown, SF, HC and RC have sever background artifacts for image pairs “amira” (the second column) and “rimg024a” (the last second column).

We next perform an objective comparison by computing the salience degree between the estimated saliency map and binary ground truth masks. Here, we adopt *F-measure* evaluation metric. It is computed by the weighted mean of *precision* and *recall*, given by $F_\sigma = \frac{(1+\sigma^2)Pre \cdot Rec}{\sigma^2 Pre + Rec}$, where $\sigma^2 = 0.3$ [4]. In order to compute those evaluation metrics, the (co-)saliency map is converted to a binary map using an adaptive threshold. The threshold is determined as two times the mean saliency of a saliency map. Table 1 shows the detailed metric values for all the examples used in this paper. We can see that our method produce results very close to the ground truth in terms of F-measure. As for precision and recall metrics, CM and RC have high precision values yet low recall values in many examples. This means the ratio of false negative is high in some cases. Our method, on the other hand, have much close precision and recall values. As for SF and HC method, their performance is subject to the background complexity, for which SF may have relatively low precision and/or recall values.

Table 1. Performance Evaluation

Image Pair	Our % P - R - F	CM% P - R - F	SF% P - R - F	HC% P - R - F	RC% P - R - F
amira	93 84 91	97 54 82	84 26 56	84 56 76	99 33 68
cdhippoa	92 85 90	90 72 85	76 45 65	77 37 62	98 25 58
rimg050a	95 91 94	87 88 87	87 85 87	85 88 86	90 91 90
areo	89 65 83	55 78 59	61 44 56	33 58 37	64 66 65
rimg024a	96 84 93	94 76 89	82 91 84	76 89 78	82 67 78

4. CONCLUSION

In this paper, we propose a novel algorithm for co-saliency detection from image pairs. After decomposing images into superpixels, our method fully exploits the properties of the intra-/inter- affinity matrix, and develop a co-saliency score based on the inter affinity matrix. Our method actually convert co-saliency detection problem to saliency detection, and can rely on well-established saliency detection techniques. In addition, we propose to apply bipartite graph matching among image pairs to improve the accuracy of superpixel affinity propagation. In the next, we plan to evaluate the performance of our method on a large-volume dataset.

5. REFERENCES

[1] Y. Sugano, Y. Matsushita, and Y. Sato, “Calibration-free gaze sensing using saliency maps,” in *CVPR*, 2010, pp.

2667 –2674.

[2] E. Rahtu, J. Kannala, M. Salo, and J. Heikkil, "Segmenting salient object from images and videos," in *ECCV*, 2010, pp. 366–379.

[3] J. Han, K.N. Ngan, M. Li, and H.-J. Zhang, "Unsupervised extraction of visual attention objects in color images," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 16, no. 1, pp. 141 – 145, 2006.

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

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

[6] R. Valenti, N. Sebe, and T. Gevers, "Image saliency by isocentric curvedness and color," in *ICCV*, 2009, pp. 2185 –2192.

[7] X. Hou and L. Zhang, "Saliency detection: a spectral residual approach," in *CVPR*, 2007.

[8] C. Guo, Q Ma, and L. Zhang, "Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform," in *CVPR*, 2008.

[9] L. Itti and P. Baldi, "Bayesian surprise attracts human visual attention," in *NIPS*, 2005, pp. 547–554.

[10] A. Borji and L. Itti, "Exploiting local and global patch rarities for saliency detection," in *CVPR*, 2012, pp. 478 –485.

[11] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *CVPR*, 2012, vol. 20.

[12] G. Jeh and J. Widom, "Simrank: A measure of structural-context similarity," in *ACM SIGKDD*, 2002.

[13] R. Achanta, S. S. Hemami, F. J. Estrada, and S. Ssstrunk, "Frequency-tuned salient region detection," in *CVPR*, 2009.

[14] R. Achanta, A. Shaji, K. Smith, A. Lucchi, and P. Fua, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274 – 2282, 2012.

[15] 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.