CO-SALIENCY DETECTION BASED ON REGION-LEVEL FUSION AND PIXEL-LEVEL REFINEMENT

Lina Li*, Zhi Liu†‡, Wenbin Zou§, Xiang Zhang*, Olivier Le Meur†‡

*School of Communication and Information Engineering, Shanghai University, China
†University of Rennes 1, France; ‡IRISA/INRIA-Rennes, France
§College of Information Engineering, Shenzhen University, China
*School of Electronic Engineering, University of Electronic Science and Technology of China
linda9293@163.com, liuzhisjtu@163.com, zouwb2008@gmail.com, uesthero@gmail.com, olemeur@irisa.fr

ABSTRACT

This paper addresses the problem of co-saliency detection, which aims to identify the common salient objects in a set of images and is important for many applications such as object co-segmentation and co-recognition. First, the segmentation driven low-rank matrix recovery model is used for intra saliency detection in each individual image of the image set, to highlight the regions whose features are sparse in each image. Then, a region-level fusion method, which exploits inter-region dissimilarities on color histograms and global consistency of regions over the image set, adjusts the intra saliency maps to obtain the region-level co-saliency maps, which can highlight co-salient object regions and suppress irrelevant regions. Finally, a pixel-level refinement method, which integrates color-spatial similarity between pixel and region with image border connectivity based object prior, generates the pixel-level co-saliency maps with better quality. Extensive experiments on two benchmark datasets demonstrate that the proposed co-saliency model consistently outperforms the state-of-the-art co-saliency models in both subjective and objective evaluation.

Index Terms— Co-saliency detection, saliency model, region-level fusion, pixel-level refinement.

1. INTRODUCTION

The past decade has witnessed the booming research on saliency detection, and a number of saliency models have been proposed and evaluated in a recent benchmark report [1]. The output of saliency model is the saliency map, which indicates the importance of pixels/blocks/regions to human visual attention, and serves as an indispensable cue for a variety of saliency-based applications such as human fixation prediction [2], salient object detection [3], salient object segmentation [4], content-aware image/video retargeting [5], and content-based image compression [6], etc. Note that most saliency models are designed to detect salient objects in a single image, but it is still difficult for the state-of-the-art single-image saliency models to effectively highlight salient objects in an image with complex scene.

Recently, an interesting issue in the research of saliency model is co-saliency detection, which exploits the repeated occurrence of salient objects in a set of images to better highlight such co-salient objects and suppress irrelevant background regions than single-image saliency models. Co-saliency model generates a set of co-saliency maps, which can be exploited for a number of related applications including object co-segmentation [7]-[8], object co-recognition [9], image retrieval [10], image matching [11] and image collage [12].

Some co-saliency models are proposed to detect co-salient objects in a pair of images. In an early work [13], the local structure changes induced by salient objects between the two images, which are captured for nearly the same scene, are used for co-saliency detection. In [14], a preattentive scheme which utilizes the joint information between image pairs is exploited to detect co-salient objects. In [15], the weighted Gist and SIFT flow are used to compute correspondences between images for facilitating the segmentation of common objects in internet images. In [16], with the aid of retrieving similar images using Google, object frequency and foreground likelihood are estimated for segmenting salient objects in the user input image. In [17], co-saliency is modeled as a linear combination of single-image saliency map, which is computed using three available single-image saliency models, and multi-image saliency...
based on a co-multilayer graph. In [18], superpixel affinity matrix for image pair is calculated and propagated using SimRank algorithm to derive co-saliency maps.

For co-saliency detection in an image set, the relevance among different images in the image set is exploited by using various formulations to highlight co-salient object regions. On the basis of single-image saliency map, the salient regions which appear frequently in most images are highlighted as co-salient in [7]. Recently, some co-saliency models leverage the clustering method, region segmentation and object extraction to effectively improve co-saliency detection performance. In [19], the cluster-based co-saliency model first computes the cluster-level co-saliency measures by integrating the contrast cue, spatial cue and corresponding cue among different images, and generates pixel-level co-saliency maps based on the likelihoods of pixels belonging to clusters. In [20], the hierarchical segmentation based co-saliency model measures regional contrast and similarity between regions in different images on the fine segmentation, evaluates object prior of regions on the coarse segmentation, and integrates them to generate the co-saliency map. In [10], the group saliency model estimates co-saliency in a set of images via salient object extraction to maximize between-image similarities and within-image distinctness.

In this paper, we propose a co-saliency model by combining region-level fusion and pixel-level refinement, which effectively exploit similarities between regions and over the image set, and similarities between pixel and region as well as image border connectivity based object prior, respectively. Using the two-level co-saliency measurement, the proposed model achieves a better co-saliency detection performance on two benchmark datasets compared to the state-of-the-art co-saliency models.

The rest of this paper is organized as follows. The proposed co-saliency model is detailed in Section 2. Experimental results are presented and analyzed in Section 3, and conclusion is drawn in Section 4.

2. PROPOSED CO-SALIENCY MODEL

The proposed co-saliency model is illustrated using the co-saliency detection on the example image set shown in Fig. 1. The following three subsections will describe the generation process of intra saliency maps, region-level co-saliency maps, and pixel-level co-saliency maps in turn.

2.1. Intra saliency

In natural images, salient object pixels usually distribute more sparsely than background pixels, which usually have a wider spatial distribution. Therefore, in terms of pixel features such as color and orientation, background pixels usually show a higher correlation than object pixels, and such a difference in the feature space can be effectively exploited to identify salient regions in the image [21]. Note that any single-image saliency model can be used for saliency detection in each individual image. In this paper, we use the segmentation driven low-rank matrix recovery (SLR) model [22] to measure the intra saliency for each image in the image set.

Give the image set \( \{I_m\}_{m=1}^M \), the SLR model is first used to generate for each image the raw saliency map \( S_m^{aw} \). Then for each image \( I_m \), the gPb-owt-ucm method [23] which exploits the globalized probability of boundary based contour detector and the oriented watershed transform to generate the real-valued ultrametric contour map \( U_m \), in which a higher value of a boundary segment indicates a higher boundary strength. By thresholding \( U_m \), a set of boundaries are retained to form a region segmentation result. Specifically, for each \( U_m \) normalized into the range of \([0, 1]\), the threshold is gradually increased from 0.1 until the number of the generated regions is less than 200, to obtain an over-segmentation result for most natural images. Then for each segmented region \( r_m^i (i = 1, ..., N_m) \) in each image \( I_m \), its intra saliency measure is calculated as follows:
\[
S_{\text{int}}(r_{m,j}) = \frac{1}{|r_{m,j}|} \sum_{p \in r_{m,j}} S_{\text{raw}}^m(p)
\]

where \( p \) denotes each pixel in \( r_{m,j} \), and \(|r_{m,j}|\) denotes the number of pixels in \( r_{m,j} \).

In Fig. 1, for the example image set shown in the 1st row, the ultrametric contour maps are shown in the 2nd row, and the region segmentation results are shown in the 3rd row, in which each region is represented using its mean color. Note that in this image set containing 36 images, the co-salient objects are the players of red team. The intra saliency maps are shown in the 4th row, in which most regions of different objects including players of both teams and referees are highlighted, while some background regions in some images are not effectively suppressed.

### 2.2 Region-level co-saliency

All images in the image set \( \{I_m\}_{m=1}^M \) are transformed into the \( Lab \) color space, in which the luminance channel and the two chrominance channels are well decorrelated. Based on the colors of all pixels in the image set, each color channel is uniformly quantized into \( q \) bins, to obtain a color quantization table \( Q \) with \( q^3 \) entries. The parameter \( q \) is set to a moderate value, 16, which is generally sufficient for color quantization of natural images. The quantized color of each entry in \( Q \) is calculated as the mean color of those pixels falling into the entry. By removing those zero-value entries, \( Q \) is updated to have a total of \( K \) (usually \( K << q^3 \)) entries. A mapping function is defined as \( k = Q(c_p) \), where \( c_p \) denotes the color of each pixel \( p \), and \( k \) denotes the entry number in \( Q \).

For each region \( r_{m,i} \), its color histogram \( H_{m,i} \) is calculated based on the quantized colors of all pixels in \( r_{m,i} \), and normalized to have \( \sum_{k=1}^K H_{m,i}(k) = 1 \). Then for each pair of regions, \( r_{m,j} \) and \( r_{n,j} \), the inter-region dissimilarity is defined as the chi-squared distance between \( H_{m,j} \) and \( H_{n,j} \), i.e.,

\[
\phi_{H}(r_{m,j}, r_{n,j}) = \frac{1}{2} \sum_{k=1}^K \frac{(H_{m,j}(k) - H_{n,j}(k))^2}{H_{m,j}(k) + H_{n,j}(k)}
\]

Since a co-salient region in one image can usually find similar regions in other images of the image set, the inter-region dissimilarity between different regions are exploited to find for each region the most similar region in each of the other images, and then used to measure for each region the global consistency over the image set as follows:

\[
C(r_{m,j}) = 1 - \frac{1}{M-1} \sum_{j=1}^M \min_{i=1}^N \left[ \phi_{H}(r_{m,j}, r_{n,j}) \right]
\]

Using Eq. (3), the global consistency measure for each region \( r_{m,j} \) falls into the normalized range of \([0, 1]\), and is used to multiply with the intra saliency measure to serve as an initial estimate of region-level co-saliency measure, i.e.,

\[
S^{\text{co-f}}(r_{m,j}) = S_{\text{int}}(r_{m,j}) \cdot C(r_{m,j})
\]

to suppress the saliency of those regions with a lower global consistency measure. Then by fully exploiting the similarity between different regions over the image set, the final region-level co-saliency measure for each region \( r_{m,j} \) is defined as follows:

\[
S^{\text{co-f}}(r_{m,j}) = \frac{\sum_{n=1}^N \sum_{j=1}^N \left[ 1 - \phi_{H}(r_{m,j}, r_{n,j}) \right] S_{\text{int}}^m(r_{n,j})}{\sum_{n=1}^N \sum_{j=1}^N \left[ 1 - \phi_{H}(r_{m,j}, r_{n,j}) \right]}
\]

where the weight \( [1 - \phi_{H}(r_{m,j}, r_{n,j})] \) with the range of \([0, 1]\) actually represents the similarity between \( r_{m,j} \) and \( r_{n,j} \), accordingly.

On the basis of intra saliency maps shown in the 4th row of Fig. 1, the initial and final region-level co-saliency maps are shown in the 5th and 6th row, respectively. We can observe that the co-salient objects, i.e., the players of red team, are highlighted, while irrelevant regions including the players of white team, referees, and background regions such as playfield and trees are suppressed in the final region-level co-saliency maps.

### 2.3 Pixel-level co-saliency

Based on final region-level co-saliency measures of each image, the pixel-level saliency map is derived by exploiting the color-spatial similarity between each pixel and its neighboring regions, and regulated by the object prior map, which is evaluated based on region's connectivity with image borders. Specifically, for each pixel \( p \in r_{m,j} \), its co-saliency is defined as the weighted sum of co-saliency measures of its neighboring regions with the regulation factor of object prior as follows:

\[
S^{\text{co-p}}_m(p) = \mathcal{O}(p) \cdot \frac{\sum_{r_{m,j} \in N_p} \psi(p, r_{m,j}) \cdot S^{\text{co-f}}(r_{m,j})}{\sum_{r_{m,j} \in N_p} \psi(p, r_{m,j})}
\]

where \( \mathcal{O}(p) \) denotes the object prior for pixel \( p \), \( N_p \) denotes the region-level neighborhood for the pixel \( p \), and \( \psi(p, r_{m,j}) \) denotes the color-spatial similarity between the pixel \( p \) and the region \( r_{m,j} \). Specifically, \( N_p \) includes \( r_{m,j} \).
The color-spatial similarity $\psi(p, r_{m,j})$ is defined as follows:

$$
\psi(p, r_{m,j}) = \begin{cases} 
H_{m,j}[Q(c_p)]_{i=j} = i \\
H_{m,j}[Q(c_p)] \cdot \exp \left( -\frac{\|x_p - y_{m,j}\|_2^2}{\|x_p - y_{m,j}\|_2} \right), & j \neq i
\end{cases}
$$

where $x_p$ denotes the spatial position of pixel $p$, and $y_{m,j}$ denotes the spatial center position of region $r_{m,j}$. Eq. (7) indicates that a higher weight is assigned to the region, which shows a higher color-similarity with the pixel $p$, i.e., a higher occurrence of the quantized color of pixel $p$ in the region's color histogram, and also locates spatially nearer to the pixel $p$. Therefore, the neighboring regions contribute to the saliency calculation of the pixel $p$ according to their color-spatial similarities with the pixel $p$.

The object prior $O(p)$ is defined based on the observation that background regions usually connect with image borders in natural images, while salient object regions do not touch image borders or connect with image borders less than background regions. As suggested in [22], to calculate the object prior based on this observation, a relatively coarse region segmentation, which partitions homogenous background into fewer regions, is suitable to suppress those background regions connecting with image borders more uniformly. Therefore, the ultrametric contour map $U_m$ for each image $I_m$ is thresholded to obtain a coarse region segmentation result with just less than 20 regions, and the object prior for each coarsely segmented region $R_{m,j}$ is defined as follows:

$$
O(R_{m,j}) = 1 - \frac{\left(R_{m,j} \cap B_m \right)^2}{\text{prm}(R_{m,j})}
$$

where $\text{prm}(R_{m,j})$ denotes the outer perimeter of $R_{m,j}$, and $B_m$ denotes the borders of image $I_m$. For each pixel $p$, its object prior is assigned with that of the region containing $p$, i.e., $O(p) = O(R_{m,j})$, $\forall p \in R_{m,j}$.

On the basis of final region-level co-saliency maps in the 6th row of Fig. 1, the generated pixel-level co-saliency maps shown in the bottom row of Fig. 1 further suppress irrelevant regions and preserves well-defined boundaries of co-salient object regions.

3. EXPERIMENTAL RESULTS

We performed experiments on two public datasets with binary ground truths for co-salient objects. Specifically, the Co-saliency Pairs (CP) dataset [17] contains 210 images, i.e., 105 image pairs, and the CMU Cornell iCoseg dataset [24] contains 643 images from 38 object classes, each of which has 5 to 41 images. We compared the co-saliency detection performance with three state-of-the-art co-saliency models, i.e., Li’s model [17], Fu’s model [19] and Liu’s model [20]. Note that Li’s model is only suitable for co-saliency detection on image pairs, and thus only tested on the CP dataset.

For subjective comparison, co-saliency maps generated
for five image pairs in the CP dataset and four image sets in the iCoseg dataset are shown in Fig. 2 and Fig. 3, respectively. For a fair comparison, all co-saliency maps are normalized into the same range of [0, 255]. For some image pairs and image sets with relatively homogenous background, such as the leftmost three image pairs in Fig. 2 and the top-left image set in Fig. 3, co-salient objects are correctly highlighted using most models, while our model can highlight the complete co-salient objects more uniformly compared to other models.

For the other image pairs/sets in Figs. 2-3, which contain non-homogeneous objects (leopards in Fig. 2), large-scale objects (stonehenges in Fig. 3), complex background (the rightmost image pair in Fig. 2 and the two image sets in the right column of Fig. 3), our model highlights the complete co-salient objects and suppresses irrelevant regions more effectively compared to other co-saliency models.

For objective comparison, we performed thresholding with a series of fixed integers from 0 to 255 on co-saliency maps to obtain a set of binary co-salient object masks, and calculated the precision and recall measures using the binary ground truths in the two datasets as reference masks. Then for each co-saliency model, at each threshold, the precision/recall values for all co-saliency maps in each dataset are averaged to plot the precision-recall curve. As show in Fig. 4, the precision-recall curves of our intra, final region-level and pixel-level co-saliency maps show the progressive improvement on co-saliency detection performance, which demonstrates the contribution of region-level fusion and pixel-level refinement, respectively. Fig. 4 shows that on both datasets, the precision-recall curves of our co-saliency maps are consistently higher than other co-saliency maps, and thus can objectively demonstrate that our model outperforms other models on the co-saliency detection performance. Besides, as shown in Fig. 4, the precision-recall curves of saliency maps generated using two state-of-the-art single-image saliency models, i.e., region contrast [25] and hierarchical saliency model [26], are lower than those precision-recall curves for co-saliency models, and this clearly shows the advantage of co-saliency detection in a set of images.

Our model is implemented using Matlab on Ubuntu 12.04 with 3.4GHz CPU. The average processing time per image is 16.48s on CP and 109.27s on iCoseg, respectively, and gPb-owt-ucm occupies 87.3% and 83.8% processing time, respectively.

4. CONCLUSION

This paper has presented an effective co-saliency model for co-saliency detection in a set of images. In our model, the region-level fusion exploits the similarities between different regions and the global consistency measures over the image
set, and the pixel-level refinement utilizes the similarities between region and pixel as well as object priors. Experimental results on two benchmark datasets show the better co-saliency detection performance of our model.

5. REFERENCES