PREATTENTIVE CO-SALIENCY DETECTION

Hwann-Tzong Chen

Department of Computer Science
National Tsing Hua University, Taiwan

ABSTRACT

This paper presents a new algorithm to solve the problem of co-saliency detection, *i.e.*, to find the common salient objects that are present in both of a pair of input images. Unlike most previous approaches, which require correspondence matching, we seek to solve the problem of co-saliency detection under a preattentive scheme. Our algorithm does not need to perform the correspondence matching between the two input images, and is able to achieve co-saliency detection before the focused attention occurs. The joint information provided by the image pair enables our algorithm to inhibit the responses of other salient objects that appear in just one of the images. Through experiments we show that our algorithm is effective in localizing the co-salient objects inside input image pairs.

Index Terms— Saliency Detection

1. INTRODUCTION

Detecting salient or irregular objects in images can be considered as a fast pre-processing step for image analysis, and has been shown to be useful in various applications, for example, human detection [1], defect detection [2], data visualization [3], texture segmentation [4], and visual tracking [5]. Despite the importance and benefit of saliency detection, the problem itself is not well defined, and the validity of a solution to the problem is not easy to evaluate or justify. Generally, the likeliness of an image region being salient or not may depend on the context in the image. Saliency detection could be made more specific if we introduce additional information. In this paper we address the problem of co-saliency detection: Given a pair of input images, we are asked to find the common salient object that appears in both images. Such problems might be solved using some unsupervised algorithms to find automatically the common parts among all possible correspondences between two candidate sets, *e.g.* [6], [7], [8], [9]. Besides co-saliency detection, similar settings that take a pair of images as the input have also been employed in object recognition (*co-recognition* [10]) and image segmentation (*co-segmentation* [11]).

Previous approaches to the co-saliency detection problem usually try to detect local features (*e.g.* SIFT [12]) first, and then perform partial matching between the two sets of features. Unlike previous approaches, we seek to solve the problem of co-saliency detection under a preattentive scheme [3], [13]. Since co-saliency detection can be considered as a pre-processing step for other applications, the computation of co-saliency should be made simple and should be performed before focused attention occurs. The advantage of the proposed algorithm is that it does not need to perform feature matching between the two input images. Furthermore, the computation of our algorithm is straightforward and can be done before focused attention. The experimental results show that our algorithm is effective in singling out the common salient object inside a pair of input images, from many otherwise equally-likely salient objects if only one input image is given.

2. A PROGRESSIVE ALGORITHM FOR PREATTENTIVE CO-SALIENCY DETECTION

To achieve the goal of co-saliency detection, we employ a distribution-based representation that characterizes the sparse features in an image. Based on the feature distributions of a pair of images, we present a progressive algorithm to enhance preattentive responses and can thus identify the common salient objects in both images.

2.1. Sparse Features

We use the sparse feature representation proposed by Hou and Zhang [14] to characterize the input images. A set of basis functions are learned from natural image patches. Totally $120,000$ RGB natural image patches are used for learning the basis functions. The size of an RGB image patch is 8×8 pixels. To model the 8×8 RGB image space, $192 (= 8 \times 8 \times 3)$ basis functions are learned, as shown in Fig. 1. Let $\mathbf{\Phi}$ denote the matrix representation of the set of basis functions, where each column of $\mathbf{\Phi}$ corresponds to a component of the basis. With the set of basis functions, an input RGB image patch $\mathbf{x}$ can be expressed as a linear combination of the basis functions, *i.e.*, $\mathbf{x} = \mathbf{\Phi}\boldsymbol{\alpha}$. The combination coefficients $\boldsymbol{\alpha}$ can be obtained by $\boldsymbol{\alpha} = \mathbf{\Phi}^{-1}\mathbf{x}$, where each row vector of $\mathbf{\Phi}^{-1}$ acts as a filter applied to the RGB image patch. Let $\mathbf{K} = \mathbf{\Phi}^{-1}$, and therefore $\boldsymbol{\alpha}$ can be viewed as the responses to the filter bank

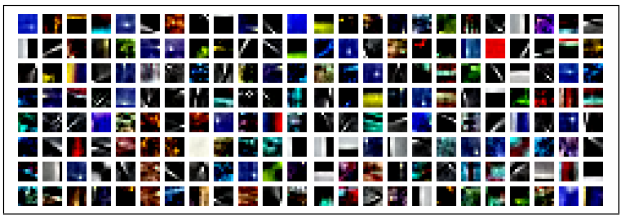

Fig. 1. Sparse feature representation: the 192 basis functions $\mathbf{\Phi}$ for 8×8 RGB natural image patches [14].

$\mathbf{K}$ of 192 filters. For the convenience of later discussion, we denote the ith row vector of $\mathbf{K}$ by $\mathbf{k}_i^{\mathsf{T}}$.

2.2. Building Feature Distributions from Filter Responses

Given a pair of images I_1 and I_2, we can extract the 8×8 image patch at each position inside the images, and obtain two sets of image patches $\mathcal{X} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(n)}, \ldots\}$ and $\mathcal{Y} = \{\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \ldots, \mathbf{y}^{(n)}, \ldots\}$, with respect to I_1 and I_2. In addition, each image patch $\mathbf{x}^{(n)}$ is associated with a saliency weight β_n and each image patch $\mathbf{y}^{(n)}$ is associated with γ_n. Initially we have $\beta_n = 1$ and $\gamma_n = 1$ for all n.

As in [14], we also apply the filter bank $\mathbf{K}$ to the image patches, and compute the absolute values of filter responses. For each filter $\mathbf{k}_i$, we add the absolute responses of the image patches $\{\mathbf{x}^{(n)}\}$ that satisfy $\beta_n > \theta$, where $\theta \in (0, 1)$ is a predefined threshold. Then, for the ith feature we derive a normalized value

$$p_i = \frac{\sum_{\{n|\beta_n > \theta\}} |\mathbf{k}_i^{\mathsf{T}} \mathbf{x}^{(n)}|}{\sum_j \sum_{\{n|\beta_n > \theta\}} |\mathbf{k}_j^{\mathsf{T}} \mathbf{x}^{(n)}|} . \tag{1}$$

We collect the normalized absolute responses and obtain a feature distribution $\mathbf{p} = [p_1, p_2, \ldots]^{\mathsf{T}}$ for image I_1. Similarly, we may build the feature distribution $\mathbf{q} = [q_1, q_2, \ldots]^{\mathsf{T}}$ from $\mathcal{Y}$ for image I_2, where

$$q_i = \frac{\sum_{\{n|\gamma_n > \theta\}} |\mathbf{k}_i^{\mathsf{T}} \mathbf{y}^{(n)}|}{\sum_j \sum_{\{n|\gamma_n > \theta\}} |\mathbf{k}_j^{\mathsf{T}} \mathbf{y}^{(n)}|} . \tag{2}$$

2.3. A Progressive Algorithm

We may compare the distribution $\mathbf{p}$ with the distribution $\mathbf{q}$ by the Kullback-Leibler divergence

$$\mathrm{KL}(\mathbf{p}||\mathbf{q}) = \sum_i p_i \log \frac{p_i}{q_i} . \tag{3}$$

Note that the Kullback-Leibler divergence is asymmetric. To detect co-salient regions, we will need to compute both $\mathrm{KL}(\mathbf{p}||\mathbf{q})$ and $\mathrm{KL}(\mathbf{q}||\mathbf{p})$. Furthermore, it is easy to see that $\mathrm{KL}(\mathbf{p}||\mathbf{q}) \geq 0$ and $\mathrm{KL}(\mathbf{q}||\mathbf{p}) \geq 0$, and the equalities hold if $\mathbf{p}$ and $\mathbf{q}$ are identical.

If the co-salient objects within the two images are found, we expect that the Kullback-Leibler divergence between $\mathbf{p}$ and $\mathbf{q}$ will be small. We compute the derivatives of $\mathrm{KL}(\mathbf{p}||\mathbf{q})$ with respect to p_i and the derivatives of $\mathrm{KL}(\mathbf{q}||\mathbf{p})$ with respect to q_i, and we try to figure out which features are more important to the minimization of the Kullback-Leibler divergence. For brevity's sake, we only list the equations for $\mathbf{p}$. The corresponding equations for $\mathbf{q}$ can be derived likewise. The derivative $\frac{\partial}{\partial p_i}\mathrm{KL}(\mathbf{p}||\mathbf{q})$ is given by

$$\begin{aligned} \frac{\partial}{\partial p_i}\mathrm{KL}(\mathbf{p}||\mathbf{q}) &= \frac{\partial}{\partial p_i} \sum_j (p_j \log p_j - p_j \log q_j) \\ &= p_i + \log p_i + p_i \log p_i - \log q_i - p_i \log q_i - \mathrm{KL}(\mathbf{p}||\mathbf{q}) . \end{aligned} \tag{4}$$

Note that $\sum_i p_i = 1$ should be taken into account. Since we are only interested in the features that cause a decrease of the Kullback-Leibler divergence, we take the negation of the derivatives and ignore those unimportant features. More specifically, we compute

$$\delta\mathrm{KL}(p_i; \mathbf{p}||\mathbf{q}) = \max\left(-\frac{\partial}{\partial p_i}\mathrm{KL}(\mathbf{p}||\mathbf{q}),\, 0\right) . \tag{5}$$

Based on $\delta\mathrm{KL}(p_i; \mathbf{p}||\mathbf{q})$ and $\delta\mathrm{KL}(q_i; \mathbf{q}||\mathbf{p})$ we update the saliency weights β_n and γ_n by

$$\beta_n = \frac{\sum_i \delta\mathrm{KL}(p_i; \mathbf{p}||\mathbf{q}) \; |\mathbf{k}_i^{\mathsf{T}} \mathbf{x}^{(n)}|}{\sum_i \delta\mathrm{KL}(p_i; \mathbf{p}||\mathbf{q})} , \tag{6}$$

and

$$\gamma_n = \frac{\sum_i \delta\mathrm{KL}(q_i; \mathbf{q}||\mathbf{p}) \; |\mathbf{k}_i^{\mathsf{T}} \mathbf{y}^{(n)}|}{\sum_i \delta\mathrm{KL}(q_i; \mathbf{q}||\mathbf{p})} . \tag{7}$$

for all image patches in $\mathcal{X}$ and $\mathcal{Y}$.

With the new saliency weights β_n and γ_n, we recompute $\mathbf{p}$ and $\mathbf{q}$ using Eqs. (1) and (2). We then repeat the previous process and keep updating β_n and γ_n for several iterations. In our experiments we run the algorithm for 15 iterations, and, as a result, the saliency weights will reflect the locations of the co-salient objects within the two input images.

3. EXPERIMENTAL RESULTS

Given a pair of input images, our algorithm is able to find the regions of the common salient object appearing in both images. In the experiments the input images are all down-sampled to 80×120 pixels. We extract every 8×8 RGB image patch from the input images, and use our algorithm to compute the saliency weights β_n and γ_n for each patch $\mathbf{x}^{(n)}$ and $\mathbf{y}^{(n)}$. The saliency weights can be spatially aggregated into saliency maps for visualization. We compare our results with those generated by the *Incremental Coding Length* (ICL) algorithm [14]. The ICL algorithm takes only a single input image, and it identifies salient regions that correspond to seldom activated features. As shown in Fig. 2(a), the saliency map produced by ICL would indicate three salient regions inside the image. Therefore, without additional information,

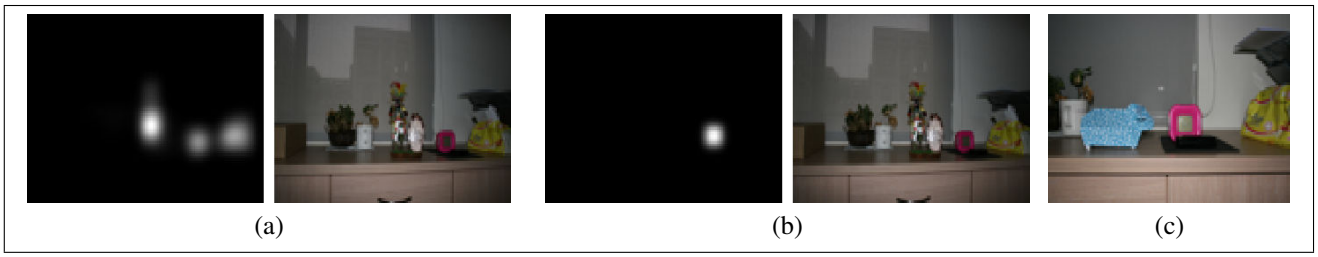

Fig. 2. (a) The saliency map generated by by the *Incremental Coding Length* (ICL) algorithm [14] and the corresponding input image. (b) The saliency map generated by our algorithm for the same input image as in (a), using the image shown in (c) as the reference for the common object.

the salient regions cannot be specifically defined. It is not trivial to single out a specific object based on the saliency map alone, and the region contains the highest saliency values may not be the one of interest. As shown in Fig. 2(b), using the co-saliency information provided by the image in Fig. 2(c), our algorithm successfully finds the location of the common salient object.

More experimental results are shown in Fig. 3. Each set of the results presents a pair of input images and their co-saliency maps. For comparison, Fig. 4 illustrates the saliency maps produced by ICL for the same image data. As can be seen, our algorithm is indeed capable of singling out the common salient objects, which are present in both paired images, using the co-saliency information.

4. CONCLUSION

We have presented a simple progressive algorithm to achieve preattentive co-saliency detection. As shown in the experimental results, our algorithm is effective in identifying the common salient objects inside pairs of input images. If only a single input image is given, it would not be easy to decide which object is the most salient one among many possible candidates. With the additional companion image as a hint, our algorithm can simultaneously locate the co-salient object inside both input images. The computation of preattentive co-saliency detection using our algorithm is straightforward and fast. We expect that our algorithm can also be readily incorporated into other applications such as video analysis as a pre-processing step.

Acknowledgment. This author was supported in part by grants 98-EC-17-A-19-S2-0052, 98-2221-E-007-083-MY3, and NTHU 99N2511E1.

5. REFERENCES

[1] James H. Elder, Simon J. D. Prince, Yuqian Hou, Mikhail Sizintsev, and E. Olevskiy, "Pre-attentive and attentive detection of humans in wide-field scenes," *International Journal of Computer Vision*, vol. 72, no. 1, pp. 47–66, 2007.

[2] Oren Boiman and Michal Irani, "Detecting irregularities in images and in video," *International Journal of Computer Vision*, vol. 74, no. 1, pp. 17–31, 2007.

[3] Christopher G. Healey, Kellogg S. Booth, and James T. Enns, "High-speed visual estimation using preattentive processing," *ACM Transactions on Human Computer Interaction*, vol. 3, no. 2, pp. 107–135, 1996.

[4] Ruth Rosenholz, "Significantly different textures: A computational model of pre-attentive texture segmentation," in *ECCV (2)*, 2000, pp. 197–211.

[5] Ming Yang, Junsong Yuan, and Ying Wu, "Spatial selection for attentional visual tracking," in *CVPR*, 2007.

[6] Hung-Khoon Tan and Chong-Wah Ngo, "Common pattern discovery using earth mover's distance and local flow maximization," in *ICCV*, 2005, pp. 1222–1229.

[7] Alexander Toshev, Jianbo Shi, and Kostas Daniilidis, "Image matching via saliency region correspondences," in *CVPR*, 2007.

[8] Junsong Yuan, Zhu Li, Yun Fu, Ying Wu, and Thomas S. Huang, "Common spatial pattern discovery by efficient candidate pruning," in *ICIP (1)*, 2007, pp. 165–168.

[9] Junsong Yuan and Ying Wu, "Spatial random partition for common visual pattern discovery," in *ICCV*, 2007, pp. 1–8.

[10] Minsu Cho, Young Min Shin, and Kyoung Mu Lee, "Co-recognition of image pairs by data-driven monte carlo image exploration," in *ECCV (4)*, 2008, pp. 144–157.

[11] Carsten Rother, Thomas P. Minka, Andrew Blake, and Vladimir Kolmogorov, "Cosegmentation of image pairs by histogram matching - incorporating a global constraint into mrfs," in *CVPR (1)*, 2006, pp. 993–1000.

[12] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[13] Anne Treisman, "Preattentive processing in vision," *Computer Vision, Graphics, and Image Processing*, vol. 31, no. 2, pp. 156–177, 1985.

[14] Xiaodi Hou and Liqing Zhang, "Dynamic visual attention: searching for coding length increments," in *NIPS*, 2008, pp. 681–688.

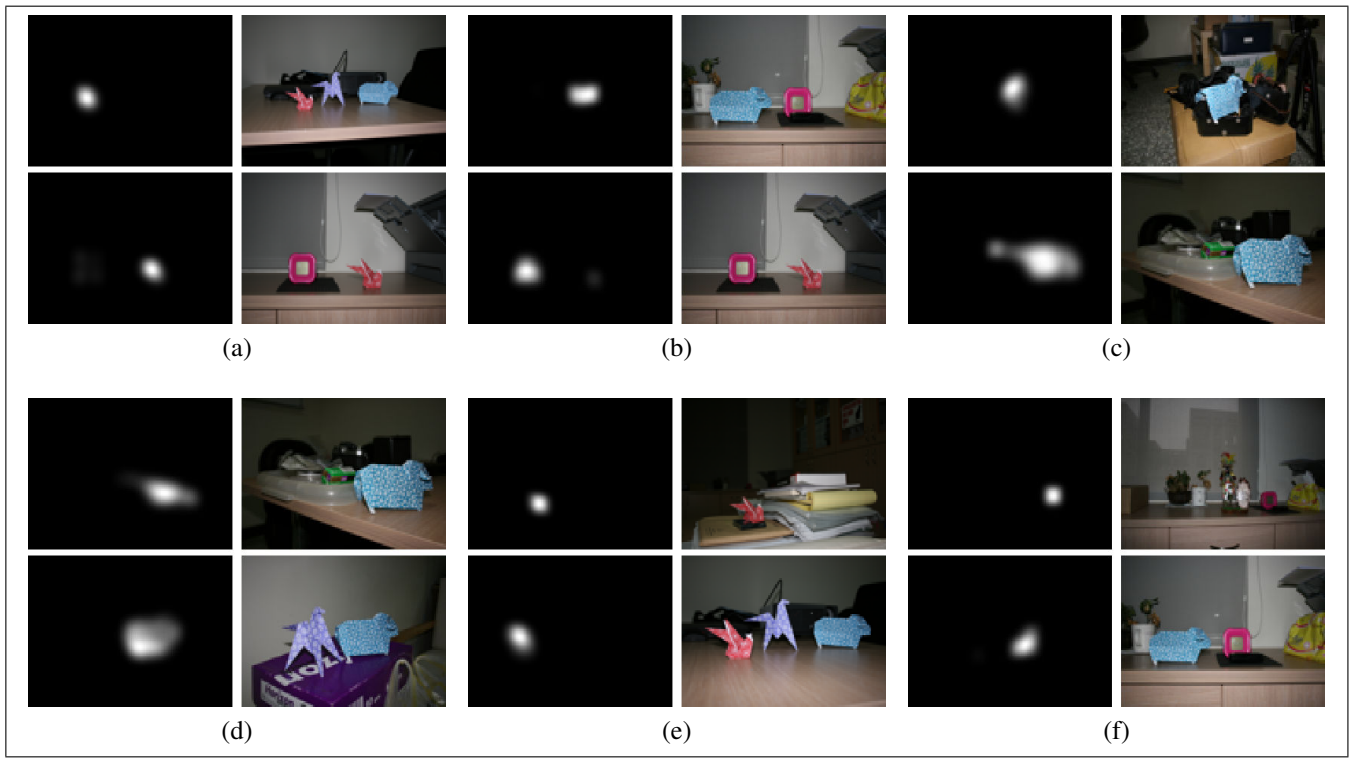

Fig. 3. Results of co-saliency detection using our algorithm. Each set of the results shows a pair of input images and the corresponding co-saliency maps. Note that, if only a single input image is given, it would not be easy to decide which object is the most salient one among many possible candidates, see Fig. 4 for comparison. It is the additional information provided by the companion image that makes the problem more specific. Our algorithm is able to use the joint information to inhibit the responses of other salient objects that appear in only one of the image. As a result, the proposed algorithm is quite effective in singling out the co-salient objects.

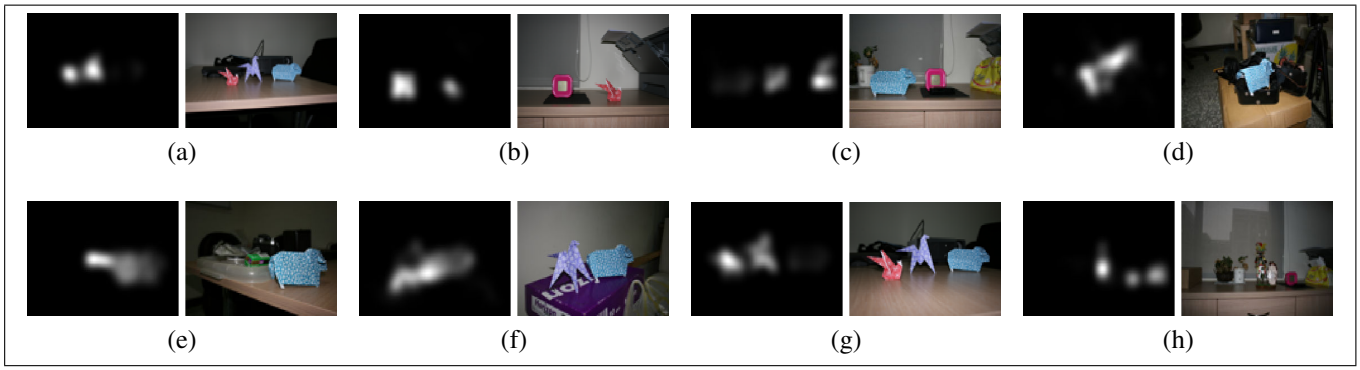

Fig. 4. Results of saliency detection using ICL. Generally a saliency map produced by ICL would present several salient regions. It is not trivial to single out a salient object based on the saliency map.