ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 5, May 2013

DETECTION OF IMAGE PAIRS USING CO-SALIENCY MODEL

N S Sandhya Rani [1], Dr. S. Bhargavi[2]

4th sem MTech, Signal Processing, S. J. C. Institute of Technology, Chickballapur, Karnataka, India[1]

Professor, Dept of ECE, S. J. C. Institute of Technology, Chickballapur, Karnataka, India[2]

Abstract-In this paper a method is presented to identify co-attention objects from an image pair. This method provides an effective way to predict human fixations within multi-images, and robustly highlight co-salient regions. This method generates the SISM by computing three visual saliency maps within each image. For the MISM computation, a co-multilayer graph is introduced using a spatial pyramid representation for the image pair. Two types of descriptors (i.e., color and texture visual descriptors) are extracted for each region node, which are then used to compute the similarity between a node-pair. Finally, a fast single-pair SimRank algorithm is employed to measure the similarity based on the normalized SimRank score.

Key words—Attention model, co-saliency, similarity, Sim-Rank.

I. INTRODUCTION

Most of the existing saliency models focus on detecting salient objects from an image rather than an image pairs.Similar object detection from multiple images has become one of the most important and challenging task in multimedia applications.So a method is introduced to detect Co-saliency from an image pair that may have some common objects.The Co-saliency is modeled as a linear combination of the single image saliency map and multi image saliency map.

A. *SISM (Single Image Saliency Map)*

There is no method that can detect the saliency accurately for all images.In order to achieve robust saliency detection a weighted saliency detection method is proposed.The goal of SISM is used to identify the salient regions within each image.If a pixel is identified as a salient pixel then it will have a high single image saliency value.Else it can be regarded as a background pixel.

SISM has three types of saliency maps

- Itti’s model saliency
- Frequency tuned saliency
- Spectral residual saliency

B. *MISM (Multi Image Saliency Map)*

The goal of the MISM is to extract the multi image saliency information from multi images .If the two images contain a similar object; the object region in each image should be assigned high value saliency values. It means that more visual attention will be attracted by this object. Else low multi image saliency values should be considered for the dissimilar regions.

MISM mainly consists of four stages

- Pyramid Decomposition of an Image Pair
- Region Feature Extraction
- The Co-Multilayer Graph Representation
- Normalized Simrank Similarity Computation

C. CO-SALIENCY MAP

The goal of the proposed method is to extract the co-saliency map from an image pair. The co-saliency map is build by a linear combination of the SISM and MISM. The goal of SISM is used to identify the salient regions within each image. The goal of the MISM is to extract the multi image saliency information from multi images. It means that a region with high co-saliency value will not only exhibit strong single-image but also multi image saliency.

II. ARCHITECTURAL DESIGN FOR PROPOSED METHOD

A method is introduced to detect Co-saliency from an image pair that may have some common objects.The proposed method is modeled as a linear combination of the single image saliency map and multi image saliency map. Architectural design for proposed system for multi-image saliency detection is given in Fig. 1, which mainly consists of the single image saliency map, multi image saliency map and co-saliency map.

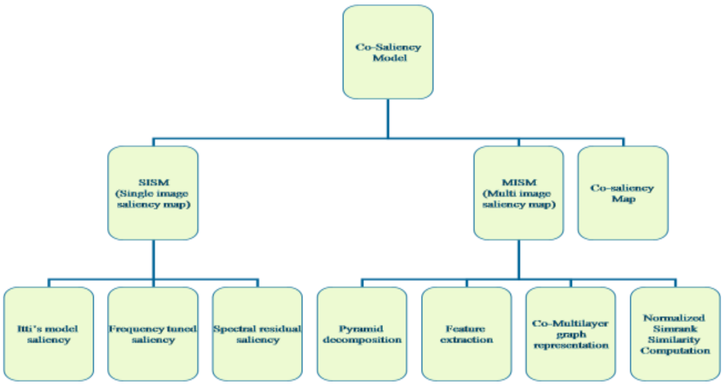

Co-Saliency Model
SISM (Single image saliency map)
MISM (Multi image saliency map)
Co-saliency Map
Itti's model saliency
Frequency tuned saliency
Spectral residual saliency
Pyramid decomposition
Feature extraction
Co-Multilayer graph representation
Normalized Simrank Similarity Computation

Fig.1. Architectural design for proposed system

III. ANALYSIS OF THEPROPOSED METHOD

The co-saliency defined in our paper is obtained by computing the single-image saliency and multi-image saliency maps. The first is used to identify the salient regions within each image. The second aims to measure the saliency for a pair of images.

A. SINGLE IMAGE SALIENCY (SISM)

We calculate three types of saliency maps, namely

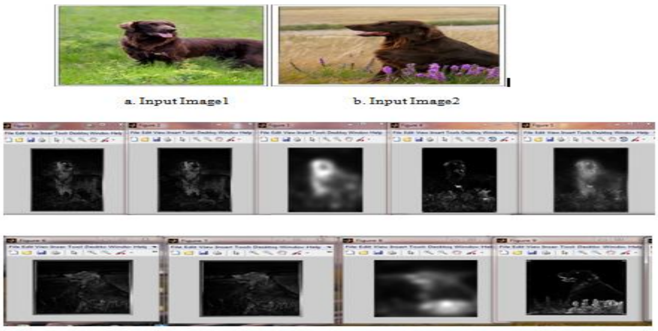
- *Itti's model saliency:* is the well-known saliency model which mimics the visual search process of human. The saliency map is computed using multi scale image features in a bottom-up manner.
- *Frequency-tuned saliency:* This estimates the center-surround contrast using color and luminance features based on a frequency-tuned approach.
- *Spectral residual saliency:* This saliency model employs the log-spectrum of an input image, and extracts the spectral residual of an image in the spectral domain.

In order to achieve robust saliency detection, a weighted saliency detection method is proposed in our work, which aims to improve detection performance by combining several saliency maps linearly. Assume I denote an input image, while S_l represents the corresponding single-image saliency map. We have

$$\sum_{j=1}^{J} W_j \mathcal{N}(s_{lj})\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

Where $\mathcal{N}\left(s_{lj}\right)$ denotes the nth normalized saliency map where each pixel has the salient value in the range [0,1] . Here, W_j denotes the weight with $\sum_{j=1}^{J} W_j = 1$. From (1), we can see that if a pixel is identified as a salient pixel by most of algorithms, it will have a high single-image saliency value. Otherwise, it can be regarded as a background pixel.

An example of the single-image saliency map is illustrated in Fig. 2, where the original image *dog* is shown in Fig. 2 (a)- (b) . Fig. 2(c) show the saliency maps extracted by the methods Itti's method [1] , Frequency tuned method [2] , Spectral residual method [3] respectively.



a. Input Image1 b. Input Image2

(c) Saliency map by Itti's method , Frequency tuned method , Saliency map by Spectral residual and Our single-image saliency map.

Fig. 2.1. Example of the single-image saliency map. (a)-(b) Original image pairs (c) Saliency map by Itti's method , Frequency tuned method , Saliency map by Spectral residual method.

B. Multi-Image Saliency Map(MISM)

Unlike the single-ima2qge saliency map that is used to describe the region saliency within an image, the goal of MISM is to extract the multi-image saliency information from multiple images. Given a pair of images, the multi-image saliency is defined as the inter-image correspondence, which can be obtained by feature matches. the multi-image saliency map of the image is defined as

$$s_g(I_i(p)) = \max_{q \epsilon I_j} \text{sim}\,(I_i(p), I_j(q)) \ldots\ldots\ldots.(2)$$

Where *p* and *q* denote entities (e.g., pixels or regions) in images $I_i(p) and\ I_j(q)$, respectively. sim ( ) represents a function that measures the similarity between two entities.

The block diagram of our proposed multi-image saliency detection is given in Fig. 3, which mainly consists of four stages, namely pyramid decomposition, feature extraction, SimRank optimization, and multi-image saliency computation.

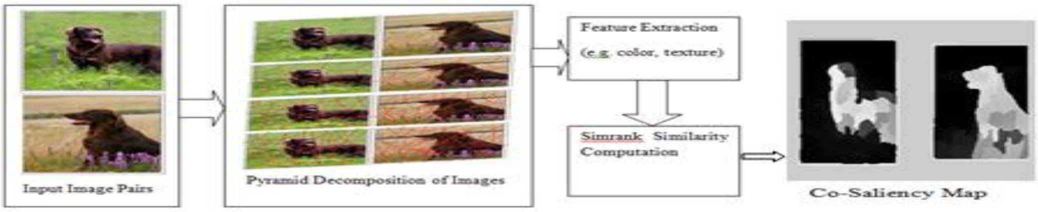


Fig. 3. Block diagram of the multi-image saliency extraction.

These are the following steps used in MISM

- *PYRAMID DECOMPOSITION OF IMAGE*

This stage is used to obtain a pyramid of images with decreasing resolutions. As a first step of the MISM computation described in Fig. 3, an initial over-segmentation is performed by partitioning an image into multiple regions is as shown in Fig.4. Each image is divided into a sequence of increasingly finer spatial regions by repeatedly partitioning the regions at each level of resolution. This is pyramid decomposition because each region at one level may be divided into several sub regions at the next level.

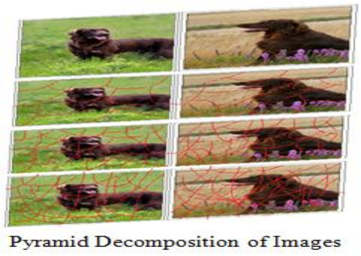


Fig. 4. Block diagram of Pyramid Decomposition of an Image Pair.

- *REGION FEATURE EXTRACTION*

Two properties are used as descriptors of regions, i.e., color and texture descriptors. The color descriptor is used to describe the region appearance from the aspect of color variations, while the texture descriptor is designed to describe the region appearance in terms of texture property. The block diagram of region feature extraction is illustrated in Fig.5.

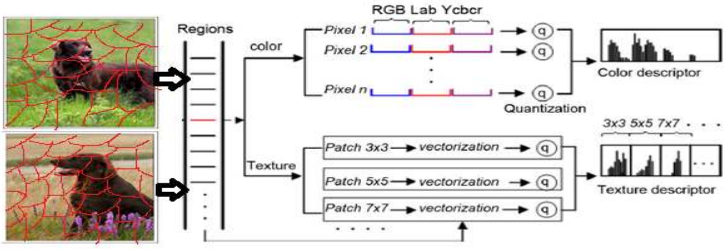


Fig.5. Block diagram of region feature extraction (e.g., the region with yellow color)

a. COLOR FEATURE EXTRACTION

In the proposed method RGB, L*a*b* and YCbCr color spaces are used together to represent the color feature. Each color space is adjusted to range from 0 to 1. To create the color visual descriptor of a region, we first represent a pixel as a 9-dimensional (9-D) color vector by combining components of RGB, L*a*b* and YCbCr color spaces. Then all pixels in the image pair are quantized into clusters by using the k-means clustering algorithm. Each cluster center is called a codeword. For each region, we simply compute the histogram by counting the number of code words at each bin (i.e., cluster). The color descriptor for a region is represented by the bins of the histogram. It is noticed that three color spaces are used to build the color histogram. By concatenating three color spaces.

b. TEXTURE FEATURE EXTRACTION

The texture descriptor is created only from RGB color space. Given an image pair, we first extract $p \times p$ patches from color images, then perform the k-means clustering over all vectors to generate M clusters. Patch words are then defined as the centers of the clusters. We measure the frequency of patch words and create the texture descriptor by combining a series of histograms of patch words. Each component histogram represents the probability of occurrence of each patch type (one bin per patch words. We concatenate the component histograms to generate the final texture descriptor.

$$f^t(k) = [H_{3X3}(k), H_{5X5}(k), H_{7X7}(k), \dots] \dots\dots(3)$$

Where $H_{iXi}(k)$ denotes the histogram computed for the k th region of size iXi . The descriptor dimension is the sum of all patch words. The texture descriptor is normalized to sum to unity.

➢ *THE CO-MULTI LAYER GRAPH REPRESENTATION*

After feature extraction, we are ready to measure the similarity so as to infer the co-salient object from a pair of images. We begin by designing a co-multilayer graph $G = (V, E)$ with $v\epsilon V \; and \; e\epsilon E$ nodes and edges , where the $\{V^1 U V^{11}\}$ nodes denote a set of regions. Two $v_i \; and \; v_j$ nodes and are connected by the directed links $e_{ij} \; and \; e_{ji}$ and , which have weights $w(e_{ij}) and \; w(e_{ji})$ and , respectively. An example of our co-multilayer graph model is shown in Fig.6, which contains three-level pyramid decomposition for a pair of images. Each region is represented as a node, which connects with other nodes by the directed edges. Note, each node not only has links with the neighboring layers within an image, it but also connects with other image nodes.

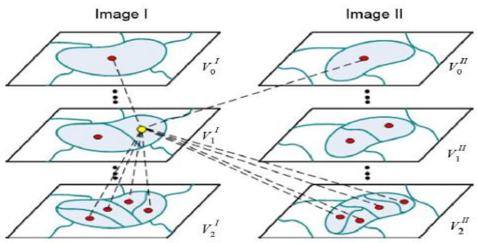


Fig.6. co-multilayer graph model.

Let $i \; and \; j$ denote two nodes and l_i and l_j represent their level numbers then the weight w_{ij} for the edge e_{ij} is defined as

$$w_{ij} = \{\exp(-\Theta_f d(f_i, f_j)), \; if \; l_i - l_j = -1 \; or \; l_i - l_j = 0\} \quad .....(4)$$

With,

$$d(f_i, f_j) = X^2(f_i, f_j) = \sum_{z=1}^{z_f} \left[\frac{((f_i(z) - f_j(z))^2)}{((f_i(z) + f_j(z)))}\right] ......(5)$$

where $f_i \; and \; f_j$ denote the color or texture descriptor for regions $i \; and \; j$,respectively. z_f denotes the dimensional number of the descriptor. Θ_f is a constant that controls the strength of the weight. $X^2(\,)$ denotes the chi-square distance.

➢ *NORMALIZED SIM-RANK SIMILARITY COMPUTATION*

SimRank,is a link-based similarity measure, is used to compute the similarity score of two region nodes. the basic intuition of SimRank is “two objects are similar if they are referenced by similar objects”, Let denote the similarity score between objects and , which is defined as

$$s(a, b) = \frac{c}{|I_n||I_n|} \sum_{i=1}^{|I_n(a)|} \sum_{j=1}^{|I_n(b)|} s(I_{ni}(a), I_{nj}(b)|) \quad ...(6)$$

Where c is a decay factor between 0 and 1, $|I_n(a)| \; and \; |I_n(b)|$ denote the numbers of in-neighbors and for nodes $and \; b$, respectively.

The normalization of the SimRank score to measure the similarity, i.e.

$$s^*(a, b) = \frac{s(a,b)}{\max(s(a,a), s(b,b))} ....(7)$$

From (7), we have$s^*(a,b) = 1$ when the nodes and share the same sub-region nodes. Substituting (7) into (2), the multi-image saliency map can be rewritten as

$$s_g\big(I_i(p)\big) = \max_{q\epsilon I_j} s^*(I_i(p), I_j(q))\text{........(8)}$$

Where $p \ and \ q$ and denote the region nodes in an image pair $I_i(p) \ and \ I_j(q)$.

C. CO-SALIENCY MAP

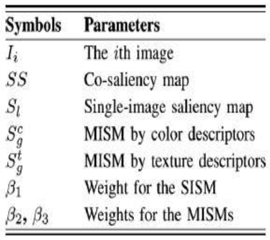
Our goal is to extract the co-saliency map from an image pair. We have presented the methods for computing single-image and multi-image saliency maps. Now we are ready to extract the co-saliency from an image pair (I_i, I_j). Let ss_i and ss_j denote the co-saliency maps for the image pair (I_i, I_j). $R\{I\}$ represents a set of regions in the image I. By combining two saliency maps (1) and (8), we have

$$\begin{aligned} ss(I_i(p)) &= \alpha_1.s_l\big(I_i(p)\big) + \alpha_2.s_g\big(I_i(p)\big) \\ &= \alpha_1.s_l(I_i(p)) + \alpha_2.(\alpha_3.s_g^c\,(I_i(p)) + \alpha_4.s_g^t\,(I_i(p))) \\ &= \beta_1.s_l\big(I_i(p)\big) + \beta_2.s_g^c\,\big(I_i(p)\big) + \beta_3.s_g^t\,\big(I_i(p)\big) \quad , \\ &\qquad for\ all \qquad p\boldsymbol{\epsilon}R\{I_i\}\text{......} \end{aligned}$$

...(9)

Where β_j is a constant with $\beta_1 + \beta_2 + \beta_3 = 1$ that is used to control the impact of the SISM and MISM on the image co-saliency. s_g^c and s_g^t denote the MISMs obtained by color and texture descriptors, respectively. The detailed parameter descriptions can be found in Table I. From (9), we can see that the co-saliency map is built by a linear combination of the SISM and MISM. It means that a region with high co-saliency value will not only exhibit strong single-image saliency but also multi-image saliency. The contributions of the SISM and MISM are controlled by the weighted coefficients β_j . From our empirical study, good performance can be achieved when β_1 takes value between 0.5 and 0.8.

TABLE I
PARAMETER DESCRIPTION

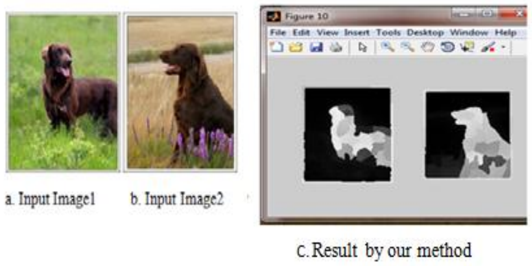


Symbols	Parameters
I_i	The *i*th image
SS	Co-saliency map
S_l	Single-image saliency map
S_g^c	MISM by color descriptors
S_g^t	MISM by texture descriptors
β_1	Weight for the SISM
β_2, β_3	Weights for the MISMs

VI. RESULTS

- Single Object Detection

We evaluate our method on a set of complex image pairs, which contain foreground objects with higher appearance variations or backgrounds with complex scenes. Some original image pairs are shown in Figs.7. (a)- (b). The corresponding results are represented in Figs. 7 (c) respectively. Experimental results show that good performance for detecting co-salient objects can be achieved by our method



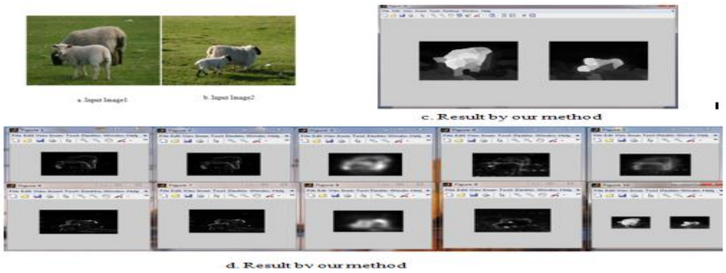
a. Input Image1 b. Input Image2

c.Result by our method

Fig.6.1. Saliency Detection for Single Object (a)-(b): Original images. (c): Results by our method.

- Multiple Objects Detection

We also evaluate our method on a set of image pairs containing objects. Some example images are shown in the in Fig.8.. The results by our methods are illustrated in the Fig. 8.(c)-(d). Experimental results show that our method can detect co-salient multiple objects from an image pair. For example, dog with different colors are shown in the first row of Fig. 8, which exhibit different poses. Both of them are identified as salient objects by our method.



a. Input Image1 b. Input Image2

c. Result by our method

d. Result by our method

Fig .6.2 Saliency Detection for multiple objects, (a)-(b): Original images pairs. (c)- (d): Results by our method.

VII. CONCLUSION

In conclusion, a method is presented to identify co-attention objects from an image pair. This method provides an effective way to predict human fixations within multi-images, and robustly highlight co-salient regions. This method generates the SISM by computing three visual saliency maps within each image. For the MISM computation, a co-multilayer graph is introduced using a spatial pyramid representation for the image pair. Two types of descriptors (i.e., color and texture visual descriptors) are extracted for each region node, which are then used to compute the similarity between a node-pair. Finally, a fast single-pair SimRank algorithm is employed to measure the similarity based on the normalized SimRank score. Experimental results were obtained by applying the proposed method to several image pairs. It has been shown that our method achieves good performance for the co-salient objects detection. In the future, we hope to incorporate more visual features (e.g., shape and contour features) to further improve the performance. Also extensions to many potential applications such as the image retrieval, semantic object discovery and co-recognition will be investigated.

REFERENCES

[1] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Trans. Pattern Anal. Mach. In-tell,* vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[2] R. Achanta, S. S. Hemami, F. J. Estrada, and S. Süsstrunk, “Frequencytuned salient region detection,” in *IEEEComp. Soc. Conf.Comput. Vis. Pattern Recognit. (CVPR)*, 2009, pp. 1597–1604.

[3]X. Hou and L. Zhang, “Saliency detection: A spectral residual approach,” in *Proc. IEEE Comp. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2009.

[4] W.Wang, Y.Wang,Q.Huang, and W. Gao, “Measuring visual saliency by site entropy rate,” in *Proc. IEEE Comp. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2010, pp. 2368– 2375.

[5] S. Goferman and L. Zelnik-Manor, “Context-aware saliency detection”, in *Proc. IEEE Comp. Soc. Conf. Comput. Vis.Pattern Recognit. (CVPR)*, 2010, pp. 2368–2375.

[6] V. Mahadevan, “Saliency-based discriminant tracking,” in *Proc. IEEE Comp. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2009, pp. 1007–1013.

[7] J. You, G. Liu, L. Sun, and H. Li, “A multiple visual models based perceptive framework for multilevel video summarization,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 3, pp. 273–285, Mar. 2007.

[8] H. Li and K. N. Ngan, “Automatic video segmentation and tracking for content-based applications,” *IEEE Commun. Mag.*, vol. 45, no. 1, pp. 27–33, 2007.

[9] A. Berengolts and M. Lindenbaum, “On the distribution of saliency,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 1973–1990, Dec. 2006.

[10] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

AUTHOR’S PROFILE

SANDHYA RANI N S perceiving MTech degree in Signal Processing in the department of Electronics and Communication engineering, SJCIT, from VTU University, Belgaum, KARNATAKA, INDIA.

Dr.S.Bhargavi is presently working as a Professor in the department of Electronics and Communication engineering, SJCIT, Chikballapur, Karnataka, India. She is having 13 years of teaching experience. Her areas of interest are Communication systems, Networksecurity, Robotics, Embedded Systems, Protocol Engineering, Image Processing, Low Power VLSI, Wireless communication, ASIC and Cryptography.