FAST AND ACCURATE MULTITASK SALIENCY DETECTION BY CONSIDERING THE FEATURES OF COLOUR, TEXTURE, ORIENTATION AND LUMINANCE

S.Hambert Solomon raja,*

S.Manikandan**

Abstract: .

Detection of salient regions of an image is very essential in the fields of image processing applications such as image segmentation, image compression, and image retrieval. The major challenges in this area are clarity, speed of operation etc. Till now lot of methods have been evolved, but there is a need in the detection process to improve the clarity and speed. Multitask saliency detection aims at considering colour, texture, and orientation, as feature descriptors and ended with an appropriate result. Here in this proposed design the features of colour, texture, orientation and luminance are going to be considered to get an accurate and fast operation. Since luminance plays a vital role in an image, by including luminance the detection process will become more efficient. Also the overall process is fast, easy to implement and generates high quality saliency maps of the same size and resolution as the input image

***Index Terms-** saliency detection, multitask learning, Multifeature modelling, sparse and low rank.*

* PG Scholar, *Department of Electronics and Communication Engineering*, PSN College of Engineering &Technology, Tirunelveli, India

** Assistant Professor, *Department of Electronics and Communication Engineering*, PSN College of Engineering &Technology, Tirunelveli, India.

I. INTRODUCTION

Visual attention is crucial in determining visual experience, leading to the challenging problem of saliency detection that is an important function for image processing and understanding. Saliency detection is related to many applications, such as automatic image cropping [5], image classification [14]. Summarization of a photo collection [17], retargeting [19] and image thumbnailing [21].Therefore, the saliency detection problem has been extensively studied in signal processing, computer vision, machine learning, and even biological literature.

According to whether the detection procedure requires human interaction or not, existing methods are divided into two categories: bottom-up (unsupervised) and top-down (supervised). In this paper, for ease of presentation, we shall first study the problem under the first setting, i.e., no learning process from labeled images is taken into account. Then, it will be shown that the proposed model can be generalized naturally environment.

Saliency detection is used to select automatically the sensory information that is notable to a human vision system. From the perspective of computer vision, the goal is to find the image regions where one or more of their features differ from those in the surroundings. As a comprehensive task, it contains several issues such as how to extract effective features and what is the optimal criterion for measuring saliency. Through lots of efforts, researchers have found several effective feature descriptors, mainly including colour, texture, orientation and luminance, as surveyed. For a certain feature schema, many computational methods have been established for measuring and detecting saliency However, the salient regions can be seldom well described by only a single feature, and generally, it is hard for such methods to handle well a wide range of images. This is because a single-feature descriptor usually only captures one aspect of the visual information. For example, the colour-based descriptors may not handle well the images with rich textures. Therefore, in this paper, we study a basic problem as follows:

• Provided that each image is described by several types of features, how can we integrate these multiple features for accurate and reliable saliency detection?

In fact, it is generally accepted that saliency detection may benefit from the integration of multiple visual features. Unfortunately, most existing literature on this direction focuses on the "naive" combination frameworks. Typically, after the saliency maps are computed for each of the features individually, they are normalized and then combined in a linear or nonlinear fashion for producing a final saliency map. The cross-feature information is not well utilized in the inference process, and it is often difficult for such naive approaches to produce reliable results.

To make effective use of multiple features, in this paper, we propose a multitask saliency model for saliency detection. Fig. 1 outlines the proposed method, which differs significantly from the previous methods in its motivation and methodology. We treat saliency detection as a sparsity pursuit

Problem and integrate multiple types of features for detecting

Saliency collaboratively. Since the cross-feature information has been well considered, such a joint inference schema can produce more accurate and reliable results than the models of producing the saliency maps individually. The inference process is formulated as a constrained nuclear norm and an $l_{2,1}$-norm minimization problem, which is convex and can be solved efficiently with augmented Lagrange multiplier (ALM) method.

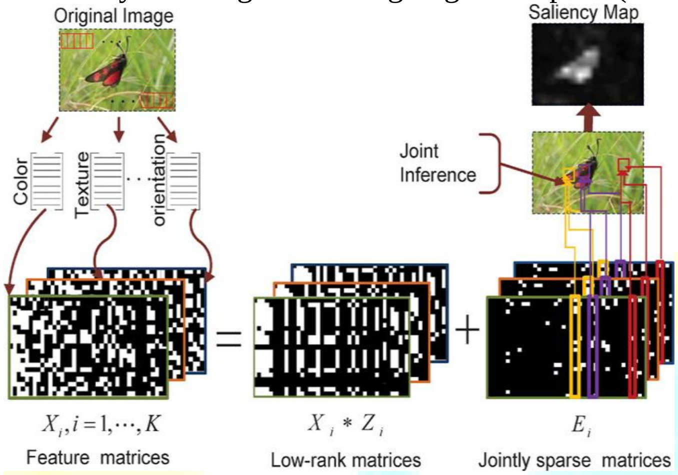


Fig. 1. For a given image, first, we extract K types of features, resulting in k types of feature matrices $X_1, X_2 \ldots X_k$ with each X_i corresponding to a certain type of feature. Second, its saliency map is inferred by seeking the consistently sparse elements E from the joint decompositions of multiple-feature matrices X_i into pairs of low-rank and sparse matrices. Note here that our method can also handle the saliency detection problem based on a single feature (i.e., k=1)

In addition to the ability of modeling multiple features, as will be seen, another advantage of multitask saliency is that it can be naturally generalized to incorporate the top-down priors so as to produce more accurate results. In summary, the contributions of this paper mainly include the following.

1) We propose a multitask saliency detection. Compared with existing models, the proposed framework seamlessly integrates multiple types of features into a unified inference procedure, which is formulated as a convex optimization problem. With some mild modifications, the proposed model can also handle the top-down priors from supervised environment.

2) Based on the proposed framework, we establish effective algorithms for saliency detection. Experimental results show that our algorithms remarkably outperform other state-of-the-art algorithms. Our algorithms are also computationally efficient.

3) The proposed multitask saliency is a general multitask method for achieving sparsity jointly. It may be useful for other related problems.

Please take a look at the images on the top row of Figure2. How would you describe them? Probably you'd say "a smiling girl", "a figure in a yellow flower field", and "a weight lifter in the Olympic games" (or something similar). Each title describes the essence of the corresponding image – what most people think is important or salient.

A profound challenge in computer vision is the detection of the salient regions of an image. The numerous applications (e.g., [1, 21, 17, 20]) that make use of these regions have led to different definitions and interesting detection algorithms. Classically, algorithms for saliency detection focused on identifying the fixation points that a human viewer would focus on at the first glance [9, 8, 24, 3, 6, 12]. This type of saliency is important for understanding human attention as well as for specific applications such as auto focusing. Others have concentrated on detecting a single dominant object of an image [13, 7, 5]. For instance, in Figure 2, such methods aim to extract the "girl", the "figure", and the "athlete" (third row). This type of saliency is useful for several high-level tasks, such as object recognition [20] or segmentation [18].

Input			
Descriptions	happy girl smiling kid cute girl	man in flower field in the fields spring blossom	Olympic weight lifter Olympic victory Olympic achievement
Salient object			
Our saliency			

Figure 2. Our multitask saliency detection results (bottom) comply with the descriptions that people provided (samples in the second row) for the input images (top). People tend to describe the scene rather than the dominant object. Classical saliency extraction algorithms aim at the third row, which might miss the essence of the scene. Conversely, we maintain all the essential regions of the image.

This calls for introducing a new type of saliency – multitask saliency. Here, the goal is to identify the pixels that correspond to the bottom row (and to the titles). According to this concept, the salient regions should contain not only the prominent objects but also the parts of the background that convey the context.

We differentiate between three types of images, as illustrated in Figure 2. In the girl's case, the background is not interesting; hence, we expect the extracted salient region to coincide with the salient object. In the flower-field's case, the texture of the flowers is essential for understanding the content. However, only a small portion of it – the portion surrounding the figure – suffices. In the weight lifter's case, some of the contextual background is vital for conveying the scene. This is not necessarily the portion surrounding the athlete, but rather a unique part of the background (the weights and the Olympic logo). Therefore, detecting the prominent object together with naive addition of its immediate surrounding will not suffice.

This paper proposes a novel algorithm for multitask saliency detection. The underlying idea is that salient regions are distinctive with respect to both their local and global surroundings. Hence, the unique parts of the background, and not only the dominant objects, would be marked salient by our algorithm (e.g., the Olympics logo in Figure 2). Moreover, to comply with the Gestalt laws, we prioritize regions close to the foci of attention. This maintains the background texture, when it is interesting, such as in the case of the flower field in Figure 2.

We demonstrate the utility of our multitask saliency in two applications. The first is retargeting [1, 19, 16], where we show that our saliency can successfully mark the regions that should be kept untouched. The second is summarization [17, 25, 2, 15, 4], where we demonstrate that saliency based collages are informative, compact, and eye-pleasing. The contribution of this paper is hence threefold. First, we introduce principles for multitask saliency (Section 2). Second, we propose an algorithm that detects this saliency (Section 3) and present results on images of various types (Section 4).

2. Principles of multitask saliency

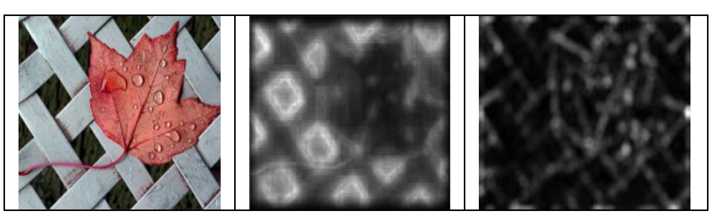
Our multitask saliency follows four basic principles of human visual attention, which are supported by psychological evidence [22, 26, 10, 11]:

1. Local low-level considerations, including factors such as contrast and color.

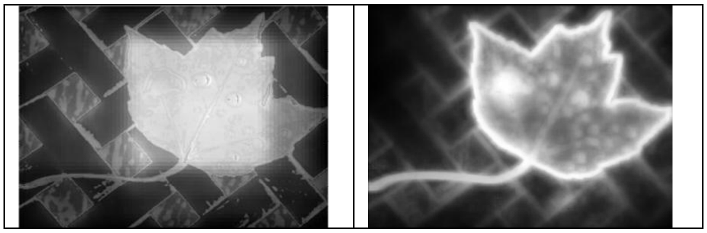
2. Global considerations, which suppress frequently occurring
Features, while maintaining features that deviate from the norm.

3. Visual organization rules, which state that visual forms may possess one or several centers of gravity about which the form is organized.

4. High-level factors, such as human faces.



(a) Input (b) Local [24] (c) Global [7]



(d) Local-global [13] (e) our multitask saliency

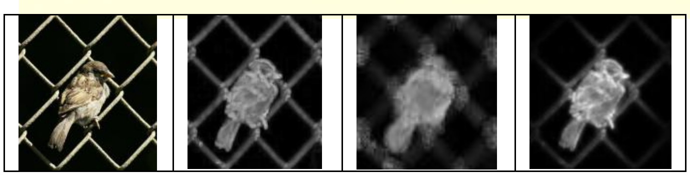
Figure 3. Comparing different approaches to saliency

Related work typically follows only some of these principles and hence might not provide the results we desire. The biologically-motivated algorithms for saliency estimation [9, 8, 24, 3, 6, 12] are based on principle (1). Therefore, in Figure 3(b), they detect mostly the intersections on the fence. The approaches of [7, 5] focus on principle (2). Therefore, in Figure 3(c), they detect mostly the drops on the leaf. In [13] an algorithm was proposed for extracting rectangular bounding boxes of a single object of interest. This was achieved by combining local saliency with global image segmentation, thus can be viewed as incorporating principles (1) and (2). In Figure 3(d) they detect as salient both the fence and the leaf, with higher importance assigned to the leaf.

We wish to extract the salient objects together with the parts of the discourse that surrounds them and can throw light on the meaning of the image. To achieve this we propose a novel method for realizing the four principles. This method defines a novel measure of distinctiveness that combines principles (1),(2),(3). As illustrated in Figure 3(e) our algorithm detects as salient the leaf, the water-drops and just enough of the fence to convey the context. Principle (4) is added as post-processing.

3. Detection of multitask saliency

In this section we propose an algorithm for realizing principles (1)–(4). In accordance with principle (1), areas that have distinctive colors or patterns should obtain high saliency. Conversely, homogeneous or blurred areas should obtain low saliency values. In agreement with principle (2), frequently-occurring features should be suppressed. According to principle (3), the salient pixels should be grouped together, and not spread all over the image.



(a) Input (b) Scale 1 (c) Scale 4 (d) Final

Figure 4. The steps of our saliency estimation algorithm

This section is structured as follows (Figure 4). We first define single-scale local-global saliency based on principles (1)–(3). Then, we further enhance the saliency by using multiple scales. Next, we modify the saliency to further accommodate principle (3). Finally, principle (4) is implemented as post-processing.

3.1 Local-global single-scale saliency: There are two challenges in defining our saliency. The first is how to define distinctiveness both locally and globally. The second is how to incorporate positional information.

3.2 Multi-scale saliency enhancement: Background pixels (patches) are likely to have similar patches at multiple scales, e.g., in large homogeneous or blurred regions. This is in contrast to more salient pixels that could have similar patches at a few scales but not at all of them. Therefore, we incorporate multiple scales to further decrease the saliency of background pixels, improving the contrast between salient and non-salient regions.

3.3 Including the immediate context: According to Gestalt laws, visual forms may possess one or several centers of gravity about which the form is organized [11] (principle (3)). This suggests that areas that are close to the foci of attention should be explored significantly more than faraway regions. When the regions surrounding the foci convey the context, they draw our attention and thus are salient.

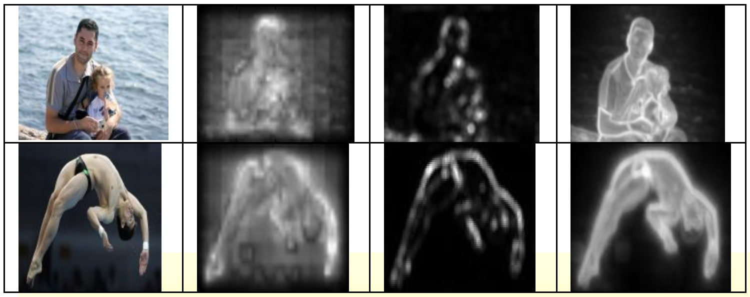
3.4 High-level factors: Finally, the saliency map should be further enhanced using some high-level factors, such as recognized objects or face detection. In our implementation, we incorporated the face detection algorithm of [23], which generates 1 for face pixels and 0 otherwise. The saliency map of Equation (5) is modified by taking the maximum value of the saliency map and the face map.

4. Results

This section evaluates the results of our approach. Figures 5–7 compare our results with the biologically-inspired local contrast approach of [24] and the spectral residual global approach of [7]. Later on in Figure 9 we compare our results with the single-object detection of [13].

As will be shown next, the method of [24] detects as salient many non-interesting background pixels since it does not consider any global features. The approach of [7] fails to detect many pixels on the prominent objects since it does not incorporate local saliency. Our approach consistently detects with higher accuracy the pixels on the dominant objects and their contextual surroundings. In all the results presented here, our saliency maps were computed without face detection for a fair comparison.

We distinguish between three cases. The first case (Figure 5) includes images that show a single salient object over an uninteresting background. For such images, we expect that only the object's pixels will be identified as salient. In [24], some pixels on the objects are very salient, while other pixels – both on the object and on the background – are partially salient as well. In [7] the background is nicely excluded, however, many pixels on the salient objects aren't detected as salient. Our algorithm manages to detect the pixels on the salient objects and only them.

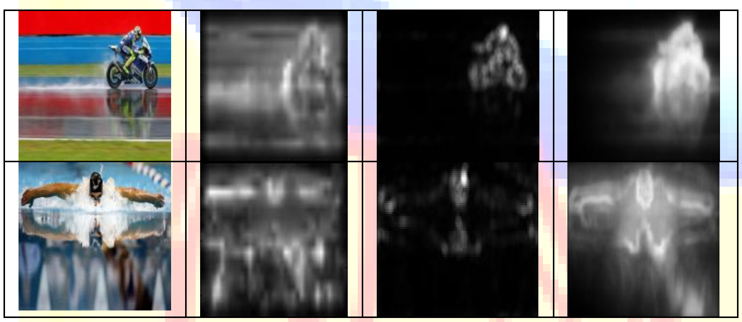


(a) Input (b) Local method (c) Global method (d) Our approach

Figure 5. Comparing saliency results on images of a single object over an uninteresting background

The second case (Figure 6) includes images where the immediate surroundings of the salient object shed light on the

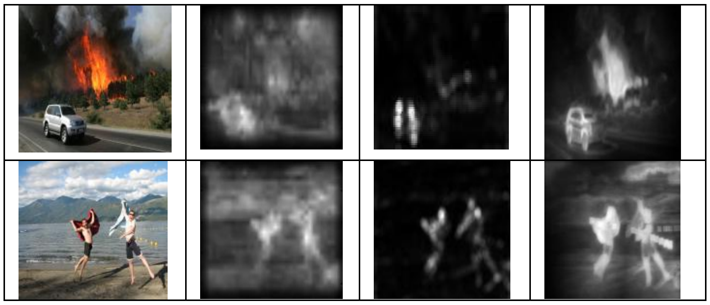
Story the image tells. In other words, the surroundings are also salient. Unlike the other approaches, our results capture the salient parts of the background, which convey the context. For example the motor-cyclist is detected together with his reflection and part of the race track and the swimmer is detected together with the foam he generates.



(a) Input (b) Local method (c) Global method (d) Our approach

Figure 6. Comparing saliency results on images in which the immediate surroundings of the salient object is also salient

The third case includes images of complex scenes. For instance, Figure 7 shows an image of a car in a fire scene and an image of two cheering guys by the lake and mountains. It can be observed that our approach detects as salient both the vehicle and the fire in the first scene and the guys with part of the scenery in the other one.



(a) Input (b) Local method (c) Global method (d) Our approach

Figure 7. Comparing saliency results on images of complex scenes

To obtain a quantitative evaluation we compare ROC curves on the database presented in [7]. This database includes 62 images of different scenes where ground-truth was obtained by asking people to "select regions where objects are presented". In part of the images only the dominant object was marked while in others also a part of the essential context was selected. Even-though this database is not perfectly suited for our task Figure 8 shows that our algorithm outperforms both [7] and [24].

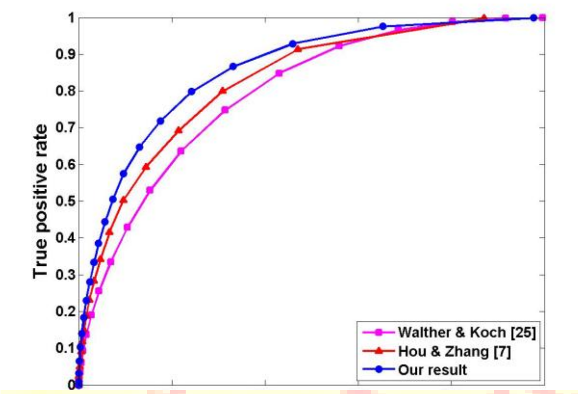


Figure 8. ROC curves for the database of [7].

Methods like [13] are not designed for such complex scenes, but rather for single dominant-object images. We do not have access to their code; hence we cannot show their results on Figures 6-7. Instead, comparisons are shown on images from their paper (Figure 9). In [13], a large database of single-object images is presented with impressive extraction results. In the left two images of Figure 9, they successfully extract

Figure 9. Comparing our saliency results with [13]. Top: Input images. Middle: The bounding boxes obtained by [13] capture a single main object. Bottom: Our saliency maps convey the story.

the”man” and the”bird”.Conversely, our saliency maps indicate that the images show “two men talking” (as both are marked salient) and a “bird on a branch feeding its fledglings”, hence providing the context. The image of the woman demonstrates another feature of our algorithm. While [13] detect the upper body of the woman (the black dress is captured due to its salient color), our algorithm marks as salient the entire woman as well as some of the stone wall, thus capturing her posing to the camera.

5. CONCLUSION

This paper proposes a new type of saliency – multitask saliency – which detects the important parts of the scene. This saliency is based on four principles observed in the psychological literature: local low-level considerations, global considerations, visual organizational rules, and high level factors. The paper further presents an algorithm for computing this saliency.

There exists a variety of applications where the context of the dominant objects is just as essential as the objects themselves.

This paper evaluated the contribution of multitask saliency in two such applications – retargeting and summarization. In the future we intend to learn the benefits of this saliency in more applications, such as image/video compressing and image collection browsing. The proposed method seamlessly integrates multiple features to produce jointly the saliency map within a single inference step and thus produces more accurate and reliable results. The proposed method may have general appealing for multitask learning.

REFERENCES

[1] S. Avidan and A. Shamir. Seam carving for content-aware image resizing. ACM Trans. on Graphics, 26(3), 2007.

[2] S. Battiato, G. Ciocca, F. Gasparini, G. Puglisi, and R. Schettini. Smart photo sticking. In Adaptive Multimedia Retrieval. IEEE, 2007.

[3] N. Bruce and J. Tsotsos. Saliency based on information maximization. Advances in neural information processing systems, 18:155, 2006.

[4] S. Goferman, A. Tal, and L. Zelnik-Manor. Puzzle-like collage. Computer Graphics Forum (EUROGRAPHICS), 29, 2010.

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

[6] J. Harel, C. Koch, and P. Perona. Graph-based visual saliency. Advances in neural information processing systems,
19:545, 2007.

[7] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. In CVPR, pages 1–8, 2007.

[8] L. Itti and C. Koch. Computational modelling of visual attention. Nature Reviews Neuroscience, 2(3):194–204, 2001.

[9] L. Itti, C. Koch, and E. Niebur. A Model of Saliency Based Visual Attention for Rapid Scene Analysis. PAMI, pages 1254–1259, 1998.

[10] C. Koch and T. Poggio. Predicting the visual world: silence is golden. Nature Neuroscience, 2:9–10, 1999.

[11] K. Koffka. Principles of Gestalt Psychology. Routledge & Kegan Paul, 1955.

[12] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau. A coherent computational approach to model bottom-up visual attention. PAMI, 28(5):802–817, 2006.

[13] T. Liu, J. Sun, N. Zheng, X. Tang, and H. Shum. Learning to Detect A Salient Object. In CVPR, 2007.

[14] E. Nowak, F. Jurie, and B. Triggs. Sampling strategies for bag-of-features image classification. Lecture Notes in Computer Science, 3954:490, 2006.

[15] Picasa. http://picasa.google.com.

[16] Y. Pritch, E. Kav-Venaki, and S. Peleg. Shift map image editing. In ICCV, 2009.

[17] C. Rother, L. Bordeaux, Y. Hamadi, and A. Blake. Autocollage. ACM Trans. Graph., 25(3):847–852, 2006.

[18] C. Rother, V. Kolmogorov, and A. Blake. "GrabCut": interactive foreground extraction using iterated graph cuts. ACM Trans. on Graphics, 23(3):309–314, 2004.

[19] M. Rubinstein, A. Shamir, and S. Avidan. Improved seam carving for video retargeting. ACM Trans. on Graphics,
27(3), 2008.

[20] U. Rutishauser, D. Walther, C. Koch, and P. Perona. Is Bottom-Up Attention Useful for Object Recognition? In CVPR, volume 2, 2004.

[21] B. Suh, H. Ling, B. B. Bederson, and D. W. Jacobs. Automatic thumbnail cropping and its effectiveness. In UIST,
pages 95–104, 2003.

[22] A. Treisman and G. Gelade. A feature-integration theory of attention. Cognitive Psychology, 12(1):97–136, 1980.

[23] P. Viola and M. Jones. Rapid Object Detection Using a Boosted Cascade of Simple Features. In CVPR, 2001.

[24] D.Walther and C. Koch. Modeling attention to salient protoobjects. Neural Networks, 19(9):1395–1407, 2006.

[25] J. Wang, J. Sun, L. Quan, X. Tang, and H. Shum. Picture collage. In CVPR, pages 347–354, 2006.

[26] J. Wolfe. Guided search 2. 0. A revised model of visual search. Psychonomic bulletin & review, 1(2):202–238, 1994.