Global-Context Based Salient Region Detection in Nature Images

Hong BAO, Member, De XU, and Yingjun TANG, Nonmembers

SUMMARY Visually saliency detection provides an alternative methodology to image description in many applications such as adaptive content delivery and image retrieval. One of the main aims of visual attention in computer vision is to detect and segment the salient regions in an image. In this paper, we employ matrix decomposition to detect salient objects in nature images. To efficiently eliminate high contrast noise regions in the background, we integrate global context information into saliency detection. Therefore, the most salient region can be easily selected as the one which is globally most isolated. The proposed approach intrinsically provides an alternative methodology to model attention with low implementation complexity. Experiments show that our approach achieves much better performance than that from the existing state-of-art methods.

key words: visual attention, salient map, global color context, matrix decomposition

1. Introduction

A lot of research efforts have been given in modeling visual attention, and the major differences among computational attention models are image features and difference mechanisms of saliency measure. Basically, the major difference among different bottom up visual attention models is the mechanism to measure saliency.

Many of the attention models [1], [2], [6] construct biologically plausible mechanisms based on the findings from psychology or neurobiology. As one of the earliest methods, Itti et al. [1] proposed a center-surround operation as local feature contrast in color, intensity and orientation of an image. Harel et al. [2] created feature maps using Itti’s method but perform their normalization using a graph-based approach. However, as shown in Fig. 1, the local contrast based approach may fail when a large set of similar targets (which are non-salient targets by definition) might be falsely detected as salient because they are associated with a high local contrast. Hence, there exists no global measurement and so these approaches may not produce satisfactory results.

In order to alleviate the negative effect of distractors possessed high local contrast, various global contrast based alternatives have been proposed [3]–[5], [7], [9], [11]. Many of these global contrast based model define bottom-up saliency based on the rarity of features. In this saliency definition form, saliency can be measured by the maximal self-information [8], or the Incremental coding length [9]. Basically, these global contrast based models might check “irregular patterns” in the context of the whole scene. In real-world scenes, however, due to insufficient sampling, the distribution estimation in high dimensional feature spaces is computationally expensive and cannot handle well complex scenes. As shown in Fig. 2, the cluttered backgrounds may yield higher saliency as such backgrounds possess high global exception in the cases with complex scenes. Meanwhile, the object borders are often assigned with higher saliency than the salient regions, even if the neighborhood size parameter is well tuned.

In this paper, the main goal is the automatic detection of visually salient targets in images, where no prior knowledge or object-level information is taken into account. The key idea is to identify targets in an image that are distinct from other targets as perceived by the human visual system during the pre-attention stage.

The remainder of this paper is organized as follows. We propose in Sect. 2 global context based saliency detection method. Experimental results are given in Sect. 3. Finally, we give the conclusive remarks in Sect. 4.
2. The Proposed Computational Attention Model

In this section, we shall introduce the computation of saliency maps and the context extraction for global color information.

2.1 Saliency Measure Based on Robust Matrix Decomposition

By the aforementioned bottom-up saliency definition, salient objects on the visual field have specific visual properties that makes them different than their surroundings. Therefore, the heart of saliency detection is the question of how to select effectively a set of the distinctive paths or regions from the original data with content redundancy. In this sense, the visual saliency detection can be casted as robust matrix decomposition [10]. More precisely, non salient target should have approximately to be low-rank, due to the bust matrix decomposition [10].

To handle this problem, a general rank minimization problem defined as follows:

$$\min_{Z,E} \text{rank}(Z) + \lambda ||E||_1,$$

$$s.t., X = Z + E,$$

(1)

Although the rank function is discrete, fortunately, it can be proven that a globally optimal solution of the above problem can be found by solving the following convex optimization problem:

$$\min_{Z,E} ||Z||_s + \lambda ||E||_{2,1},$$

$$s.t., X = Z + E,$$

(2)

where $|| \cdot ||_s$ denotes the nuclear norm of a matrix, i.e. the sum of the singular values of the matrix. $||E||_{2,1} = \sum_{j=1}^{n} \sqrt{\sum_{i=1}^{n} (|E|_{ij})^2}$ is called as $l_2/l_1 - norm$, and the parameter $\lambda > 0$ is used to balance the effects of the two parts, which could be chose according to properties of the two norms, or tuned empirically. Since $l_2/l_1 - norm$ encourages the columns of $E$ to be zero, the underlying assumption here fits well our saliency definition that salient target generally occupy only a fraction of the image pixels. According to the above framework, salient map can be obtained easily by sparse matrix $E$. The optimization problem can be solved by using the inexact ALM algorithm [14].

2.2 Global Color Context Extraction

A center-weighted spatial variance feature is also introduced in [12] to eliminate regions with small variances close to the boundaries of the image. Though color provides powerful information for saliency analysis, different illuminations can result in different colors of a same object surface. Most existing approaches on global color feature are very sensitive to illumination change. To deal with this case, we introduce a global color feature to describe the saliency of an object. First, all colors in the image are represented by Gaussian Mixture Models (GMMs). Each pixel is assigned to a color component with the probability $\rho(c|I_i)$. Here, considering the illumination change, we adopt the color vector $\{R_x, G_x, B_x\}$ in the 1-order derivative image, so each pixel location is an 6-D vector $[R, G, B, R_x, G_x, B_x]$. Then, we employ the color distribution entropy [13] to describe the spatial information of an image, defined as:

$$\zeta_{c}(P_c) = - \sum_{i=1}^{N_c} P_{c_i} \cdot \log P_{c_i},$$

(3)

where $P_c$ can be computed as annular color histograms. And $N_c$ is the number of histogram bins for the $c$-th color component, $c \in \{R,G,B,R_x,G_x,B_x\}$. In our experiments, we choose $N_c$ to be 8. The Color Distribution Entropy gives the dispersive degree of the pixel patches of a color bin in an image. Based on the Color Distribution Entropy, global color factor $\phi(l_i)$ is obtained as a weighted sum:

$$\phi(l_i) = \sum_{c \in C} \rho(c|l_i) \cdot ||1 - \zeta_c||,$$

(4)

The feature map is also normalized to the range [0, 1]. The color spatial context can be used to describe the saliency of the region. Generally, the wider a color is distributed in the image, the less possible a salient regions contains this color [12]. Figure 3 shows several feature maps.
We propose a method to identify salient parts in images. The saliency maps are not similar in shape and in their spatial distribution. In our experiments, image patches are 8×8 non-overlapping blocks. Similar to [9], a set of 192 color sparse basis functions is used as a sparse representation of natural image patches. For our evaluations, we have labeled all images accurate-to-contour manually. We then compare our model with four state-of-the-art methods. The four saliency detectors are Itti’s model [1], Graph based visual saliency [2], context-aware based global contrast [3], and self-information based method [8], hereby referred to as IT, GBVS, CSD, and SI respectively.

First, we test our model by comparing the saliency map to the Human fixation density maps. The latter is constructed by recording human eye fixations over an image [8]. Qualitative or subjective evaluation provides an insight into the effectiveness of the proposed model. Some of the detection results are illustrated in Fig. 4. Even though the different models work quite well on these examples, the predicted saliency maps are not similar in their shape and in their spatial distribution. Qualitatively speaking, the color spatial feature map described in [12] is also illustrated. Qualitatively speaking, the color spatial context can capture the “object-level” salient region. Especially, when the scenes include one clearly attractive stimuli, the distractors in the background might be suppressed more effectively using our method than the color spatial feature [12]. Thus, if the saliency in an image is dominated by color content, the salient regions are well covered by the global color modeling proposed in this paper.

Base on the model in the above subsection, after the sparse matrix is obtained, each column is denoted as the response of an input patch. A simple quantify the response of sparse matrix is defined as following:

\[
S(l) = \varphi(l) \sum_{j=1}^{d} ||E(j, l)||
\]

where \(d\) denotes the dimension number of the feature vector for each patch \(l\). For a column of the matrix \(E\) corresponding to the \(i\)-th patch, larger (smaller) magnitude implies that the patch is more salient (non-salient). Since sparse matrix \(E\) is obtained by solving a convex optimization problem, the final saliency map is produced by combing global color context in the inference processing.

3. Experimental Results

We conduct several experiments to demonstrate the effectiveness of the proposed method on two public image database for both qualitative and quantitative evaluations. One is the color image database collected by Bruce et al. [8], which usually serves as the benchmark database for comparing visual saliency detection results. Eye fixations are recorded from 20 subjects on the 120 color images. The other one is the SIVAL data set\(^1\). The categories consist of complex objects photographed against 10 different highly diverse backgrounds. In our experiments, image patches are 8×8 non-overlapping blocks. Similar to [9], a set of 192 color sparse basis functions is used as a sparse representation of natural image patches. For our evaluations, we have labeled all images accurate-to-contour manually. We compare our model with four state-of-the-art methods. The four saliency detectors are Itti’s model [1], Graph based visual saliency [2], context-aware based global contrast [3], and self-information based method [8], hereby referred to as IT, GBVS, CSD, and SI respectively.

First, we test our model by comparing the saliency map to the Human fixation density maps. The latter is constructed by recording human eye fixations over an image [8]. Qualitative or subjective evaluation provides an insight into the effectiveness of the proposed model. Some of the detection results are illustrated in Fig. 4. Even though the different models work quite well on these examples, the predicted saliency maps are not similar in their shape and in their spatial distribution. Qualitatively speaking, and in average, the proposed method can succeed in identifying the salient parts of the pictures.

We then compare the ROC curves to the other four methods. The ROC curve results are shown in Fig. 5 and Fig. 6. For the convenience of evaluation, we use the ROC metric proposed in [2] and an averaged ROC area for each ROC curve, denoted as AUC. Especially, the ROC curve is obtained by trying all possible threshold values, and for each value, plotting the true positives rate (TPR) on the Y-axis against the false positive rate (FPR) value on the X-axis. From the ROC curves in Fig. 5, we can see that the both CSD and Our method are reliably in predicting saliency against human fixation map. Further we test the salient region detection with respect to robustness to illumination color variations. In our experiment, we transform each image of the SIVAL database under 11 varying light, the colors of the same object have large difference as shown in Fig. 6. The proposed method is also invariant to illumination change due to the mapped target image obtained by global and local

\(^{1}\)The dataset is available at http://www.cse.wustl.edu/~sg/accio.
color context. We see that the algorithm has accurately determined the attention regions even in transformed images.

As shown in Table 1, there is significant gains in terms of AUC. The proposed model obtains over 3.1% improvement against ICL for SIVIL Dataset, respectively. This demonstrates that the proposed model has better discriminative power than the baselines by taking into account the global context information.

In order to evaluate the robustness of the detection of the most salient node, besides the ROC and AUC, we also calculate the Precision, Recall and F-measure on SIVIL database. As shown in the Fig. 7, the output of our methods are Precision = 0.74, Recall = 0.55, and F-measure = 0.67 respectively for the top 50 percent salient locations.

4. Conclusion

This paper presents a novel visual saliency model based on matrix decomposition and global context. Our algorithm finds pre-attentive and bottom-up saliency. It is inspired by the biological concept of center-surround contrast at the local and global levels. In contrast with existing local contrast based saliency algorithms, we first employ matrix decomposition as a novel saliency measure approach. Further, the effective global color context are extracted and combined to produce the final saliency map. The proposed method detects well in both fixation points and the dominant object. In future work, it would be interesting to include more complex types of constraints into the proposed framework, e.g., constraints involving object-level prediction.

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. Intell., vol.20, no.11, pp.1254–1259, 1998.
[2] J. Harel, C. Koch, and P. Perona, “Graph-based visual saliency,” Proc. NIPS, pp.545–552, 2006.
[3] S. Goferman, L. Zelnikmanor, and A. Tal, Context Aware Saliency Detection. Proceedings of CVPR, pp.2376–2383, 2010.
[4] X. Hou and L. Zhang, “Saliency detection: A spectral residual approach,” Proc. CVPR, pp.1–8, 2007.
[5] V. Gopalakrishnan, Y. Hu, and D. Rajan, “Random walks on graphs for salient object detection in images,” IEEE Trans. Image Process., vol.19, no.12, pp.3232–3242, 2010.
[6] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, “A coherent computational approach to model bottom-up visual attention,” IEEE Trans. Pattern Anal. Mach. Intell., vol.28, no.5, pp.802–817, 2006.
[7] T. Judd, K. Ehinger, F. Durand, and A. Torralba, “Learning to predict where humans look IEEE international conference on computer vision,” Proc. ICCV, 2009.
[8] N. Bruce and J. Tsotsos, “Saliency based on information maximization,” Advances in Neural Information Processing Systems, vol.18, p.155, 2006.
[9] X. Hou and L. Zhang, “Dynamic visual attention: Searching for coding length increments,” Proc. NIPS, pp.681–688, 2008.
[10] E.J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?,” J. ACM, vol.58, no.1, pp.1–37, 2009.
[11] T. Avraham and M. Lindenbaum, “Esaliency (extended saliency): Meaningful attention using stochastic image modeling,” IEEE Trans. Pattern Anal. Mach. Intell., vol.32, no.4, pp.693–708, 2010.
[12] T. Liu, J. Sun, N.N. Zheng, X. Tang, and H.Y. Shum, “Learning to detect a salient object,” IEEE Conference on Computer Vision and Pattern Recognition, pp.1–8, 2007.
[13] J. Sun, X. Zhang, et al. Image Retrieval Based on Color Distribution Entropy, Pattern Recognit. Lett., vol.27, pp.1122–1126, 2006.
[14] Z. Lin, M. Chen, L. Wu, and Y. Ma, “The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices,” UIUC Technical Report, UILU-ENG-09-2215, 2009.