Saliency Detection via Nonlocal L_0 Minimization

Yiyang Wang[1], Risheng Liu[2], Xiaoliang Song[1], and Zhixun Su[1]

[1] School of Mathematical Sciences, Dalian University of Technology, China
[2] School of Software Technology, Dalian University of Technology, China
{yywerica,ericsong}@gmail.com {rsliu,zxsu}@dlut.edu.cn

Abstract. In this paper, by observing the intrinsic sparsity of saliency map for the image, we propose a novel nonlocal L_0 minimization framework to extract the sparse geometric structure of the saliency maps for the natural images. Specifically, we first propose to use the k-nearest neighbors of superpixels to construct a graph in the feature space. The novel L_0-regularized nonlocal minimization model is then developed on the proposed graph to describe the sparsity of saliency maps. Finally, we develop a first order optimization scheme to solve the proposed nonconvex and discrete variational problem. Experimental results on four publicly available data sets validate that the proposed approach yields significant improvement compared with state-of-the-art saliency detection methods.

1 Introduction

The recent years have witnessed significant advances in saliency detection [1–6]. Visual saliency is making the most informative scene stand out from their neighbors and grabbing immediate attention. It is originally a task of eye fixation prediction [7–10], and recently has been extended to salient region detection [11, 12, 6, 2, 13]. Both of them can be categorized as either bottom-up [14] or top-down [15] approaches in general. The former is fast, pre-attentive, data-driven saliency extraction while the latter is slower, task dependent, goal driven saliency extraction [6]. In this paper, we focus on the bottom-up salient region detection.

Most bottom-up saliency methods rely on predefined assumptions about salient objects and backgrounds. One of the early bottom-up saliency detection methods is proposed by Itti et al. [8], which focuses on the role of color and orientation priors. Goferman et al. [1] propose a context-aware algorithm that represents the scene based on four assumptions of human visual attention. However, using the predefined assumptions only cannot generate maps that uniformly cover the whole object and suppress the background well. With the emergence of superpixels [16, 17], an increasing number of image saliency detection approaches are proposed on region level to reduce calculation amount and receive uniform foreground and suppressed background. One significant class among them is graph based approaches [7, 18–24]. Yang et al. [19] propose to detect salient regions in images through manifold ranking on a graph which incorporates local grouping cues and boundary priors. Liu et al. [24] provide a diffusion viewpoint

to model saliency detection which shows its well performance on salient region detection. However, all the previous work start from the perspective of improving the prior and achieving good results statistically on public benchmarks, but few of them take the characteristic of the ideal result into consideration.

We in this paper note that all of the ground truth saliency maps are binary images which contain a few intensity changes among neighboring pixels/superpixels. Namely, the intensity changes are very sparse. Due to the fact that the L_0 norm of a vector is the number of non-zero entries which directly measures sparsity, we propose a graph based nonlocal L_0 (NLL_0 for short) minimization for visual saliency detection. The proposed NLL_0 method is able to capture the sparse properties of saliency maps, thereby leading to reliable results. Furthermore, the graph construction plays a key role in graph based methods. Different from local graphs that connect neighbor regions in the spatial space (e.g. 1-ring and 2-ring graph), we construct a nonlocal graph based on the k-nearest neighbors (k-NN for short) in the feature space. This k-NN graph better extracts the structure of the image, and can generate a more uniform forward and suppressed background. Due to the discrete nature and the non-convexity of the proposed model, conventional convex optimization algorithm cannot be directly used to solve it. By extending alternating direction method (ADM) [25–27] to non-convex discrete optimization problem, this paper proposes an efficient numerical scheme to solve our NLL_0 saliency detection model. Overall, the main contributions in this paper are summarized as follows:

- We propose a novel NLL_0 model on image graph for saliency detection. As the L_0 norm is the naive metric used to describe sparsity of the gradients, the saliency maps of our proposed model can exactly highlight the foreground and suppress the background.
- A nonlocal graph is constructed in the feature space composed by color and spatial position features, and each vertex in this graph is connected to its k-NN.
- As a non-trivial by product, a first order numerical scheme is introduced to efficiently solve the non-convex discrete NLL_0 optimization model.

2 The Proposed Method

In this section, an L_0 norm based minimization is introduced on image graph to model saliency detection. For a given image, we first over segment it into superpixels by SLIC method [16] and construct a graph based on the k-NN of superpixels in the feature space. Then, our NLL_0 minimization is proposed on this graph incorporating with a saliency control map that is generated by contrast and object priors.

2.1 Graph Construction

It is worth mentioning that graph construction plays a key role in graph based methods [7, 18–21], and the connected relationship is top priority. As two classic

graphs, 1-ring and 2-ring graph are widely used in superpixel based saliency detection methods [18, 19]. However, 1-ring and 2-ring graph are two typical graphs that connect neighbor elements in the spatial space (see Fig. 1(c) and Fig. 1(d) to have an intuitive sense), which is premised on the assumption that neighboring superpixels in the spatial space are likely to share similar saliency values. This assumption is absolutely incorrect near the edges between the salient and indistinctive regions thus may generate redundant regions. The example shown in Fig. 2 also shows that with these two graphs, the background and forward are not differentiated effectively. In this paper, we propose to construct graph in the feature space to take nonlocal relationship into consideration (Fig. 1(e)).

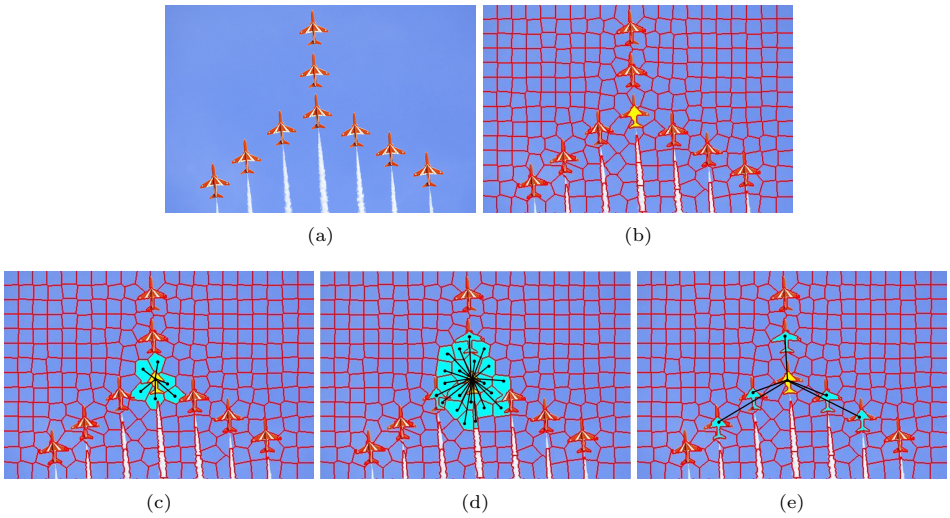


Fig. 1. The comparisons of 1-ring, 2-ring and k-NN graph. (a) The input image. (b) The superpixels of the input image, and a yellow patch is specified. (c) The connected relationship in 1-ring graph. (d) The connected relationship in 2-ring graph. (e) The connected relationship in k-NN graph.

Given an image, we generate superpixels by SLIC method [16] to be the elements of saliency estimation, denoted as a vector of N elements with corresponding values $\mathbf{v} = (v_1, v_2, \cdots, v_N)$. It is not specific that any other edge preserving methods to generate superpixels can be used in this place. Then we construct an undirected, symmetric and weighted graph G. It consists a finite set E of edges and is associated with a weight function $\omega : E \to \mathbb{R}_+$ satisfying $\omega_{pq} = \omega_{qp}$, for all $pq \in E$.

The weight between superpixels p and q is expressed as:

$$\omega_{pq} = \exp(-\frac{\|\mathbf{f}_p - \mathbf{f}_q\|^2}{2\sigma^2}), \tag{1}$$

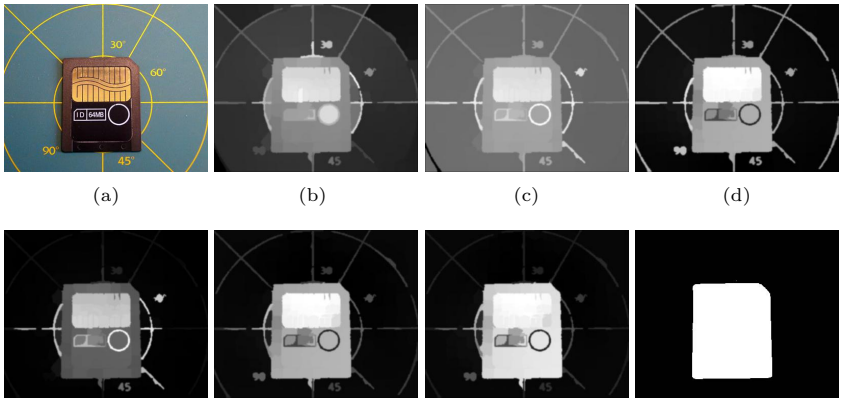
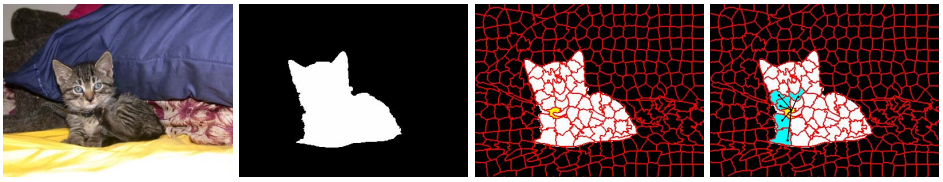


Fig. 2. Well performance of our graph from quality analysis. (a) Input image. (b) Result of 1-ring graph. (c) Result of 2-ring graph. (d) Result of k-NN graph. (e) Result of 2-ring graph with four boundaries of the image connected. (f) Result of k-NN graph with four boundaries of the image connected. (g) Result of our graph. (h) Ground truth.

where $\mathbf{f}_p = (\alpha \mathbf{c}_p, \mathbf{l}_p)$ is a feature vector at superpixel p comprising of its appearance $\mathbf{c}_p$ (i.e.,the mean of the superpixels in CIE LAB color space) and location $\mathbf{l}_p$ (i.e., the mean of the coordinates of superpixels in spatial space). Parameter α is used to control the balance between these two features, and $\|\cdot\|$ denotes the L_2 norm of a vector. For each superpixel p, we calculate its weight with every other q in image domain, and find its k-NN to exploit local relationships in the feature space. Keep these k values which indicate the most similar superpixels of p and set all the others to zero. The connected relationship between p and q is indicated by the nonzero ω_{pq}. Connecting neighbors in feature space performs well than other traditional graph construction methods from the results in the top line of Fig. 2.

With the purpose of reducing the geodesic distance of similar superpixels that have large spatial distance, Yang et al. [19] propose to connect the four boundaries of the image together in the 2-ring graph which performs well on generating suppressed background. We further enforce that the most indistinctive regions in saliency control map $\mathbf{v}^c$ (is discussed in Section 2.3) and the four boundaries of the image should be connected to each other to get a better performance. We choose the latter 25% of the superpixels of $\mathbf{v}^c$ in descending order as the most indistinctive regions in our approach. The results shown in the second line of Fig. 2 indicate that the strategy of connecting boundaries and indistinctive regions together does improve the saliency result, and our proposed method have the best performance from both quality analysis (Fig. 2) and quantity analysis (Fig. 4(c)).



(a) (b) (c) (d)

Fig. 3. The nonlocal gradient and its sparsity. (a) An input image. (b) The ground truth of the input image. (c) The superpixels of the ground truth, and a specified patch coloring in yellow. (d) The related patch (colored in cyan) when computing the nonlocal gradient of the specified yellow patch.

2.2 NLL_0 for Saliency Detection

As discussed before, the ideal result in salient region detection, i.e. the ground truth (Fig. 3(b)), has extremely sparse gradient on superpixels. To obtain the same property of our saliency result, we propose to minimize a variational formulation based on the L_0 norm of gradient. Though it is applied in many low-level image processing problems [28–30], we first extend it for graph based visual saliency detection. To begin with, we will introduce the basic definitions and notations which are borrowed from Bougleux et al. in [31] and Gilboa-Osher in [32], regarding local differential geometry operators that will be useful in the rest of the paper.

Different from the definition of gradient on a regular image grid, the gradient of each vertex p on the graph G is defined for pair points $pq \in E$ as:

$$\nabla_\omega \mathbf{v}_p = (\partial_q v_p : pq \in E). \tag{2}$$

Specifically, $\nabla_\omega \mathbf{v}_p$ is a vector of all partial derivatives, i.e. $\partial_q v_p = (v_q - v_p)\sqrt{\omega_{pq}}$, where ω_{pq} is the weight between p and q as discussed in Section 2.1.

A visual example is shown in Fig. 3(d) to explain the nonlocal gradient on graph. Suppose that p denotes the yellow region in Fig. 3(c). According to the definition of our k-NN graph, its k-NN are denoted by the cyan patches (see Fig. 3(d)). Regarding the definition of $\nabla_\omega \mathbf{v}_p$, we can conclude that most values of $\nabla_\omega \mathbf{v}_p$ are zero. Thus, a natural idea that depicts this property is to employ L_0 norm on $\nabla_\omega \mathbf{v}_p$. It can also be seen that only few superpixels has nonzero gradient in the ground truth, and that truly confirm the reliability of minimizing the L_0 norm of gradient.

Based on considerations above, the proposed NLL_0 model on image graph is defined as follows:

$$\min_{\mathbf{v}} \sum_p \|\nabla_\omega \mathbf{v}_p\|_0 + \frac{\lambda}{2}\|\mathbf{v} - \mathbf{v}^c\|^2, \tag{3}$$

where $\|\nabla_\omega \mathbf{v}_p\|_0$ is defined as $\|\nabla_\omega \mathbf{v}_p\|_0 = \sharp\{q|\partial_q v_p \neq 0; pq \in E\}$[3], λ is a positive constant that controls the trade-off between the sparsity and the fidelity term.

[3] The notation '$\sharp$' is a mathematical representation which stands for the cardinality of a set.

The new symbol $\mathbf{v}^c$ in the fidelity term is the saliency control map which will be introduced in the next section.

2.3 Saliency Control Map

We in this paper use the contrast prior and object prior to compute the salient control map $\mathbf{v}^c$.

We use the contrast prior for the reason that research from perceptual analysis [33] and relative works [18, 6, 34] have indicated the effectiveness of the contrast measure. Given the mean color value $\mathbf{c}_p$ and the weight $\omega_{pq}^{(l)}$ between superpixel v_p and v_q in spatial space, the contrast measure of v_p is defined as:

$$v_p^{con} = \sum_{q \neq p} \omega_{pq}^{(l)} \|\mathbf{c}_p - \mathbf{c}_q\|^2, \tag{4}$$

where $\omega_{pq}^{(l)} = \exp(-\frac{1}{2\sigma_l^2}\|\mathbf{l}_p - \mathbf{l}_q\|^2)$ with normalized average coordinates $\mathbf{l}_p$ and $\mathbf{l}_q$.

On the other hand, object prior provides an assumption on the most likely location of salient region. Though there are many approaches on generating either high-level [23, 5] or low level [1, 13, 18, 35] object prior, we choose to use low-level object prior in this paper:

$$v_p^{obj} = \exp(-\frac{\|\mathbf{l}_p - \bar{\mathbf{l}}\|^2}{2\sigma_c^2}), \tag{5}$$

where $\bar{\mathbf{l}}$ is the central coordinate position of the interest.

Then we combine the above two priors together for each superpixel v_p to generate the saliency control map $\mathbf{v}^c$ in a simple way:

$$v_p^c = v_p^{con} \times v_p^{obj}. \tag{6}$$

Based on the experimental results, we use the convex-hull based center prior [35] as the object prior $\mathbf{v}^{obj}$ in this paper. More details and experimental analysis on the influence of the object prior are conducted in Section 4.3.

3 Optimization

Our NLL_0 minimization is indeed a very challenging problem due to its discreteness and non-convexity. For the failure of using the traditional gradient descent or other discrete optimization methods to optimize the problem, we propose a problem solving strategy based on ADM which is now very popular in solving large scale sparse representation problems [25–27].

By introducing an auxiliary variable $\mathbf{d}$, our graph based NLL_0 variational model can be written in an equivalent form:

$$\begin{aligned} \min_{\mathbf{d},\mathbf{v}} \quad & \sum_p \|\mathbf{d}_p\|_0 + \frac{\lambda}{2}(v_p - v_p^c)^2, \\ \text{s.t.} \quad & \mathbf{d}_p = \nabla_\omega \mathbf{v}_p, \qquad p = 1, \cdots, N. \end{aligned} \tag{7}$$

We employ the ADM to solve problem (7). In each iteration, we alternatively solve

$$\begin{aligned} \mathbf{v}^{k+1} &= \arg\min_{\mathbf{v}} \sum_p \frac{\lambda}{2}(v_p - v_p^c)^2 + \frac{\rho}{2}\|\mathbf{d}_p^k - \nabla_\omega \mathbf{v}_p + \frac{1}{\rho}\mathbf{y}_p^k\|^2, \\ \mathbf{d}^{k+1} &= \arg\min_{\mathbf{d}} \sum_p \|\mathbf{d}_p\|_0 + \frac{\rho}{2}\|\mathbf{d}_p - \nabla_\omega \mathbf{v}_p^{k+1} + \frac{1}{\rho}\mathbf{y}_p^k\|^2, \\ \mathbf{y}_p^{k+1} &= \mathbf{y}_p^k + \rho(\mathbf{d}_p^{k+1} - \nabla_\omega \mathbf{v}_p^{k+1}). \end{aligned} \tag{8}$$

The solution of the subproblem on v_p is characterized by its first order optimality condition:

$$\lambda(v_p - v_p^c) + \rho \mathrm{div}_\omega(\mathbf{d}_p^k - \nabla_\omega \mathbf{v}_p + \frac{1}{\rho}\mathbf{y}_p^k) = 0. \tag{9}$$

Here, $\mathrm{div}_\omega \mathbf{u}$ is defined as the divergence of $\mathbf{u}$, and its discretization at p can be deduced

$$\mathrm{div}_\omega \mathbf{u}_p = \sum_q (u_{pq} - u_{qp})\sqrt{\omega_{pq}}, \quad pq \in E. \tag{10}$$

where u_{pq} is the vector element corresponding to q [32].

According to Eq. (10), the solution v_p^{k+1} of the subproblem can be explicitly written as:

$$\begin{aligned} v_p^{k+1} = & \frac{1}{\lambda + 2\rho\sum_q \omega_{pq}}(\lambda v_p^c \\ & - \rho \sum_q \sqrt{\omega_{pq}}(d_{pq}^k - d_{qp}^k + \frac{1}{\rho}y_{pq}^k - \frac{1}{\rho}y_{qp}^k) \\ & + 2\rho \sum_q \omega_{pq} v_q^{k+1}). \end{aligned} \tag{11}$$

Though $\mathbf{v}^{k+1}$ can be directly solved by a linear system, a fast approximated solution $v_p^{k+1,n+1}$ is provided by a Gauss-Seidel iterative scheme, and 2 iterations ($n = 2$) are enough [36] to determine a good approximation of the minimizer in experiments:

$$\begin{aligned} v_p^{k+1,n+1} = & \frac{1}{\lambda + 2\rho\sum_q \omega_{pq}}(\lambda v_p^c \\ & - \rho \sum_q \sqrt{\omega_{pq}}(d_{pq}^k - d_{qp}^k + \frac{1}{\rho}y_{pq}^k - \frac{1}{\rho}y_{qp}^k) \\ & + 2\rho \sum_q \omega_{pq} v_q^{k+1,n}), \mathbf{v}^{k+1,n=0} = \mathbf{v}^k, \end{aligned} \tag{12}$$

Algorithm 1 Salient Region Detection via NLL_0

Input: Given image I and necessary parameters.
Postprocessing: Generate superpixels by SLIC [16].
Output: Saliency score $\mathbf{v}$ of the given image. $\mathbf{v} = \{v_p\}$.
1: Construct the k-NN graph on superpixels.
2: Calculate saliency control map $\mathbf{v}^c$ by (6).
3: Initialization: $v_p \leftarrow 0$, $\mathbf{d}_p \leftarrow \mathbf{0}$.
4: **repeat**
5: **while** n < maxIter **do**
6: Compute $v_p^{k+1,n+1}$ by (12).
7: **end while**
8: Compute $\mathbf{d}_p^{k+1}$ through the thresholding function (15).
9: Update the Lagrangian multiplier $\mathbf{y}_p^{k+1}$ by (8).
10: **until** max iterations reached.

The subproblem on $\mathbf{d}$ is apparently sophisticated subproblem due to the non-convexity and discontinuity of L_0 norm. We can obtain its solution by the following lemma which can be proved in the same way as [28].

Lemma 1. *The optimal solution* $\mathbf{x}^\star$ *of the following problem:*

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 + \frac{\beta}{2}\|\mathbf{x} - \mathbf{z}\|^2, \tag{13}$$

is defined as: for every component x_i *of* $\mathbf{x}$*,*

$$x_i^\star = \begin{cases} 0, & |z_i| \le \sqrt{\frac{2}{\beta}}, \\ z_i, & \text{otherwise}, \end{cases} \tag{14}$$

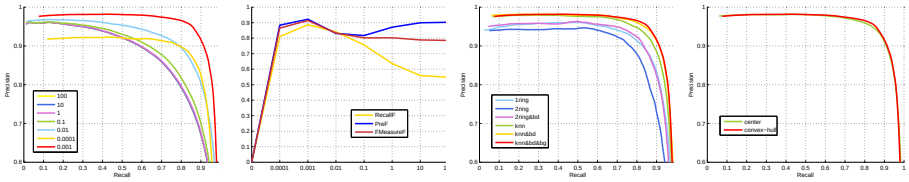
According to Lemma 1, the solution $\mathbf{d}_p^{k+1}$ of the subproblem on $\mathbf{d}$ is given as follows:

$$d_{pq}^{k+1} = \begin{cases} 0, & |\partial_q v_p^{k+1} - \frac{1}{\rho} y_{pq}^k| \le \sqrt{\frac{2}{\rho}}, \\ \partial_q v_p^{k+1} - \frac{1}{\rho} y_{pq}^k, & \text{otherwise}. \end{cases} \tag{15}$$

Finally, the main steps of the proposed saliency detection method are summarized in Algorithm 1.

4 Experiments

Experiments taken in this paper can be divided into four parts: parameter evaluation in Section 4.1, comparison of graph in Section 4.2, analysis of object prior in Section 4.3 and comparison with state-of-the-art methods in Section 4.4.



(a) (b) (c) (d)

Fig. 4. (a) Precision and recall rate for the regularization parameter λ on the ASD data set. (b) Average precision, recall and F-measure curve with different λ parameter. (c) Effectiveness of our graph by PR-curve. (d) Comparison of using image center prior and convex-hull-based center prior as object control map respectively in our model.

4.1 Parameter Evaluation

We set the number of superpixels $N = 300$ in all the experiments. There are 5 parameters in our NLL_0 method: σ and α in Eq.(1), the number of nearest neighbors k, λ and ρ in Eq.(7). The parameters are empirically set to $\sigma^2 = 0.05$, $\alpha = 0.9$, $k = 5$, $\lambda = 0.001$ and $\rho = 0.0001$.

The parameter λ is a weight directly controlling the balance between regularization and fidelity, which makes an important role in our approach. We assign λ as 0, 0.1, 0.01, 0.001, 0.0001, 1, 10 and 100 in our approach respectively. And draw the corresponding PR-curves (Fig. 4(a)) and the average precision, recall and F-measure curves (Fig. 4(b)) with different λ.

In particular, $\lambda = 0$ corresponds to an extreme case that only regular term is contained in the NLL_0 model. The gradient of each superpixel in the optimal solution of this case must be zero, which corresponds to an all-black color image. The PR-curve of $\lambda = 0$ is not drawn in Fig. 4(a) since it is a straight line from 0 to 0.2. With the increase of λ, the saliency control map $\mathbf{v}^c$ takes a bigger role in the NLL_0 model, and more edge information is remained. But it changes smoothly when $\lambda > 1$ (see Fig. 4(b)) and the PR-curves almost overlapped when $\lambda = 1, 10$ and 100.

Whether according to the PR-curves (Fig. 4(a)) or the average precision, recall and F-measure curves (Fig. 4(b)), $\lambda = 0.001$ is the best choice in our approach, which is fixed and applied in all the experiments in this paper.

4.2 Comparison of Graph

In Section 2.1, we give an example to illustrate the effectiveness of our k-NN graph. However, this is not sufficient. In this subsection we extensively test their validity by using public data set. We use the precision-recall curve to show the well performance of our graph.

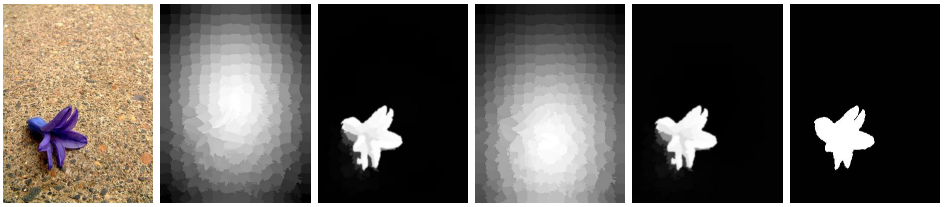
Fig. 4(c) shows the comparison result of 6 different strategies to construct graph. From the Fig. 4(c), the 2-ring graph is the worst choice for our NLL_0 method, the 1-ring graph and the 2-ring graph with boundaries connected (is used in [19]) perform almost the same. Leaving out the strategy of connecting

the four boundaries and the most indistinctive regions together, the k-NN graph itself performs better than the graphs constructed in the spatial space, which furthermore indicates that the strategy of constructing graph in the feature space can contribute to the NLL_0 model. Also, the strategy that connects all the boundaries and background together does have a better effect.

4.3 Analysis of Object Prior

In this paper, we use low-level object prior $\mathbf{v}^{obj}$ in the saliency control map $\mathbf{v}^c$. Certainly, different coordinate positions $\bar{\mathbf{l}}$ of the interest will generate different prior maps by Eq. (5). In this section, we conduct an experiment on two different strategies for obtaining $\bar{\mathbf{l}}$, and the results show that our NLL_0 method is robust to these two low-level object priors.

Regarding the image center as $\bar{\mathbf{l}}$ is a common assumption in saliency detection [1, 13]. It hypothesizes that people taking photographs generally frame the focus and assign higher value to the image elements near the image center. But there always exist some special cases that image center assumption may bring wrong instruction. Xie et al. [35] propose a object prior which calculates the convex hull of the object and lets its center as $\bar{\mathbf{l}}$, which obtains a more reasonable and robust object measure map. We use the two object priors in the saliency control map respectively, and the discrepancy between them is quite small through quantitative analysis (see Fig. 4(d)). An intuitive example is also given in Fig. 5. Though the prior maps generated by this two strategies are different (Fig. 5(b) and (d)), the final results (Fig. 5(c) and (e)) are nearly the same. The experimental results show the robustness of our NLL_0 method to the object prior.



(a) (b) (c) (d) (e) (f)

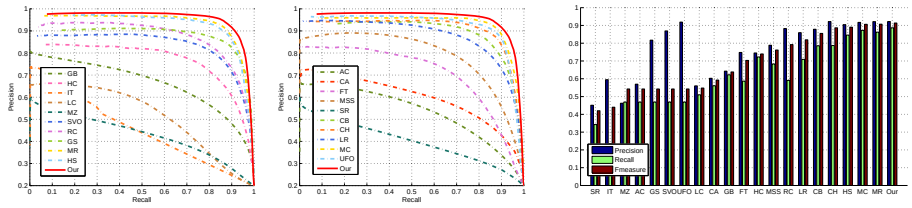
Fig. 5. Robustness of the object prior. (a)Input image. (b) Object prior map based on image center [1, 13, 21]. (c) Result of using image center as the object prior. (d) Object prior map based on convex-hull [18]. (e) Result of using convex-hull center as the object prior. (f) Ground truth.

4.4 Comparison with State-of-the-art Methods

Our experiments are conducted on ASD, MSRA, ECSSD and iCoseg data sets. The ASD data set contains 1000 images, is a subset of MSRA which contains a

(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (l) (m)

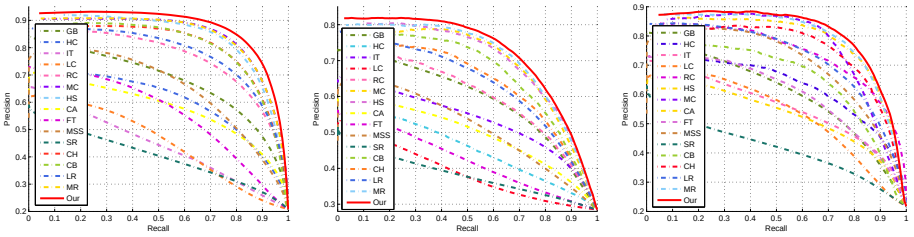
Fig. 6. Visual comparison on MSRA, ECSSD and iCoseg data sets. The top three rows are selected from MSRA data set, the middle three are from ECSSD data set, and the last two rows are chosen from iCoseg data set. We compare 9 state-of-the-art methods, from (b)-(j): CA[1], HC[6], RC[6], CB[13], LR[4], CH[18], MR[19], HS[37] and MC[20]. Saliency maps generated by our algorithm (l) is the closest to the ground truth (m).



(a) (b) (c)

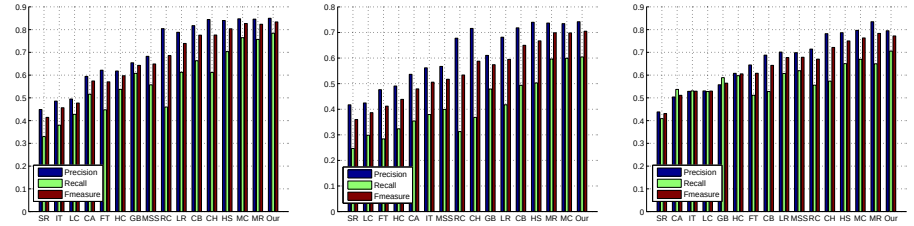
Fig. 7. Performance of the proposed algorithm compared with 20 state-of-the-art methods on the ASD database. (a), (b) Average precision recall curve by segmenting saliency maps using fixed thresholds. (c) Mean precision, recall, and F-measure values of the evaluated methods.

large variation among 5000 images of natural scenes, animals, people, etc. ASD is publicly available in salient object detection and is evaluated in almost every saliency paper. ECSSD is a data set with 1000 images, which includes many semantically meaningful but structurally complex images for evaluation. We use this data set to show the well-performance of difficult images with complex backgrounds. The iCoseg data set including 38 groups of 643 images, which is a publicly available co-segmentation data set. The images in iCoseg may contain



(a) (b) (c)

Fig. 8. Performance of the proposed algorithm compared with 15 state-of-the-art methods on the MSRA database through average precision recall curve. (a) Comparison on MSRA data set. (b) Comparison on ECSSD data set. (c) Comparison on iCoseg data set.



(a) (b) (c)

Fig. 9. Performance of the proposed algorithm compared with 15 state-of-the-art methods on the MSRA database through mean precision, recall, and F-measure values. (a) Comparison on MSRA data set. (b) Comparison on ECSSD data set. (c) Comparison on iCoseg data set.

one or multiple salient objects. We use this data set to evaluate the performance of multiple salient object detection.

We compare our NLL_0 approach with 20 previous methods on ASD, including 12 earlier but classical algorithm: AC[11], FT[12], GB[7], HC[6], IT[8], LC[38], MSS[39], MZ[9], SR[10], SVO[2], CB[13], RC[6], and recent state-of-the-art algorithms: CA[1], CH[18], GS[3], HS[37], LR[4], MR[19], UFO[5], MC[20].

Visual comparison of selected 9 state-of-the-art approaches is shown in Fig. 6. As is shown in Fig. 6, our approach can deal with the challenging images with complex background and also works well on image contains multi-salient objects (the sixth row in Fig. 6).

Quantitative comparisons on ASD data set are shown in Fig. 7. Fig. 7 (a) and (b) are the PR-curves of the 20 previous methods applied on ASD. All the methods are separated into two groups for the purpose of clear visual effect. It is shown that our approach favorable outperforms other methods, while achieves similar performance as HS, MR and MC in terms of PR-curve. However, as is discussed in [34], neither the precision nor recall measure considers the true neg-

ative counts. These measures respond the ability of assigning saliency to salient regions, but fail to react the ability of detecting the opposite. Our approach not only successfully assigns saliency to salient regions, but also successfully does the opposite. Comparison of the eighth to eleventh columns in Fig. 6 demonstrates the well performance of our approach on generating a well suppressed background. The time complexity of our algorithm mainly focuses on Gauss-Seidel iteration in the subproblem of u, and its running time is proportional to the square of N at each iteration. Since few steps are enough for Gauss-Seidel strategy to get the desired result in our experiments, so the time complexity of our algorithm is totally $O(N^2)$[4].

We also compare 15 state-of-the-art approaches on MSRA, ECSSD and iCoseg data sets, for the reason that the relevant codes of some previous approaches are not publicly available. Fig. 8 and Fig. 9 show the quantitative comparison through both PR-curve and F-measure. The difference between our method and others is clear, manifesting that our NLL_0 approach can be widely used in different types of images.

5 Conclusion and Future Work

This paper develops a energy minimization based on L_0 norm for salient region detection. We construct a graph based on the k-NN of superpixels in the feature space. Then, NLL_0 is proposed on this graph with saliency control map that is generated by contrast and object prior. We solve this non-convex minimization problem by an iterative strategy based on ADM. Our NLL_0 approach is evaluated on various challenging image sets with comparison to state-of-the-art techniques to show its superiority for saliency detection. In the future, we plan to focus on generating semantic features from the image that can better describe the characteristic of the salient region.

Acknowledgement. Risheng Liu is supported by the National Natural Science Foundation of China (No. 61300086), the China Postdoctoral Science Foundation (2013M530917, 2014T70249), the Fundamental Research Funds for the Central Universities (No. DUT12RC(3)67) and the Open Project Program of the State Key Laboratory of CAD&CG, Zhejiang University, Zhejiang, China (No. A1404). Zhixun Su is supported by National Natural Science Foundation of China (Nos. 61173103, 91230103) and National Science and Technology Major Project (No. 2013ZX04005021).

References

1. Goferman, S., Zelnik-Manor, L., Tal, A.: Context-aware saliency detection. IEEE T. PAMI **34** (2012) 1915–1926

[4] We would like to list the running time (second per image) for the compared methods in our paper: Our 0.97s, CA 48.65s, CB 1.97s, LR 14.89s, CH 1.07s, MR 1.02s, MC 0.25s. It can be observed that our detector is comparable among all the MATLAB implementation based saliency detectors in our paper.

2. Chang, K.Y., Liu, T.L., Chen, H.T., Lai, S.H.: Fusing generic objectness and visual saliency for salient object detection. In: ICCV. (2011)
3. Wei, Y., Wen, F., Zhu, W., Sun, J.: Geodesic saliency using background priors. In: ECCV. (2012)
4. Shen, X., Wu, Y.: A unified approach to salient object detection via low rank matrix recovery. In: CVPR. (2012)
5. Jiang, P., Ling, H., Yu, J., Peng, J.: Salient region detection by ufo: Uniqueness, focusness and objectness. In: ICCV. (2013)
6. Cheng, M.M., Zhang, G.X., Mitra, N.J., Huang, X., Hu, S.M.: Global contrast based salient region detection. In: CVPR. (2011)
7. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: NIPS. (2006)
8. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE T. PAMI **20** (1998) 1254–1259
9. Ma, Y.F., Zhang, H.J.: Contrast-based image attention analysis by using fuzzy growing. In: ACM Multimedia. (2003)
10. Hou, X., Zhang, L.: Saliency detection: A spectral residual approach. In: CVPR. (2007)
11. Achanta, R., Estrada, F., Wils, P., Süsstrunk, S.: Salient region detection and segmentation. In: ICVS. (2008)
12. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: CVPR. (2009)
13. Jiang, H., Wang, J., Yuan, Z., Liu, T., Zheng, N., Li, S.: Automatic salient object segmentation based on context and shape prior. In: BMVC. (2011)
14. Mai, L., Niu, Y., Liu, F.: Saliency aggregation: A data-driven approach. In: CVPR. (2013)
15. Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.Y.: Learning to detect a salient object. IEEE T. PAMI **33** (2011) 353–367
16. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: SLIC superpixels compared to state-of-the-art superpixel methods. IEEE T. PAMI **34** (2012) 2274–2282
17. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE T. PAMI **22** (2000) 888–905
18. Yang, C., Zhang, L., Lu, H.: Graph-regularized saliency detection with convex-hull-based center prior. IEEE Signal Processing Letters **20** (2013) 637–640
19. Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M.H.: Saliency detection via graph-based manifold ranking. In: CVPR. (2013)
20. Jiang, B., Zhang, L., Lu, H., Yang, C., Yang, M.H.: Saliency detection via absorbing markov chain. In: ICCV. (2013)
21. Jiang, Z., Davis, L.S.: Submodular salient region detection. In: CVPR. (2013)
22. Gopalakrishnan, V., Hu, Y., Rajan, D.: Random walks on graphs to model saliency in images. In: CVPR. (2009)
23. Jia, Y., Han, M.: Category-independent object-level saliency detection. In: ICCV. (2013)
24. Liu, R., Lin, Z., Shan, S.: Adaptive partial differential equation learning for visual saliency detection. In: CVPR. (2014)
25. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning **3** (2011) 1–122
26. Lin, Z., Liu, R., Su, Z.: Linearized alternating direction method with adaptive penalty for low-rank representation. In: NIPS. (2011)

27. Liu, R., Lin, Z., Su, Z.: Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning. In: ACML. (2013)
28. Xu, L., Lu, C., Xu, Y., Jia, J.: Image smoothing via l_0 gradient minimization. ACM TOG **30** (2011) 174
29. Xu, L., Zheng, S., Jia, J.: Unnatural l_0 sparse representation for natural image deblurring. In: CVPR. (2013)
30. Pan, J., Su, Z.: Fast l_0-regularized kernel estimation for robust motion deblurring. IEEE Signal Processing Letters (2013)
31. Bougleux, S., Elmoataz, A., Melkemi, M.: Discrete regularization on weighted graphs for image and mesh filtering. In: Scale Space and Variational Methods in Computer Vision. Springer (2007) 128–139
32. Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. Multiscale Modeling and Simulation **7** (2008) 1005–1028
33. Einhäuser, W., König, P.: Does luminance-contrast contribute to a saliency map for overt visual attention? European Journal of Neuroscience **17** (2003) 1089–1097
34. Perazzi, F., Krahenbuhl, P., Pritch, Y., Hornung, A.: Saliency filters: Contrast based filtering for salient region detection. In: CVPR. (2012)
35. Xie, Y., Lu, H.: Visual saliency detection based on bayesian model. In: ICIP. (2011)
36. Bresson, X.: A short note for nonlocal tv minimization. (2009)
37. Yan, Q., Xu, L., Shi, J., Jia, J.: Hierarchical saliency detection. In: CVPR. (2013)
38. Zhai, Y., Shah, M.: Visual attention detection in video sequences using spatiotemporal cues. In: ACM Multimedia. (2006)
39. Achanta, R., Susstrunk, S.: Saliency detection using maximum symmetric surround. In: ICIP. (2010)