Structure-Sensitive Saliency Detection via Multilevel Rank Analysis in Intrinsic Feature Space

Chenglizhao Chen, Shuai Li, Hong Qin, *Senior Member, IEEE*, and Aimin Hao

***Abstract*— This paper advocates a novel multiscale, structure-sensitive saliency detection method, which can distinguish multilevel, reliable saliency from various natural pictures in a robust and versatile way. One key challenge for saliency detection is to guarantee the entire salient object being characterized differently from nonsalient background. To tackle this, our strategy is to design a structure-aware descriptor based on the intrinsic biharmonic distance metric. One benefit of introducing this descriptor is its ability to simultaneously integrate local and global structure information, which is extremely valuable for separating the salient object from nonsalient background in a multiscale sense. Upon devising such powerful shape descriptor, the remaining challenge is to capture the saliency to make sure that salient subparts actually stand out among all possible candidates. Toward this goal, we conduct multilevel low-rank and sparse analysis in the intrinsic feature space spanned by the shape descriptors defined on over-segmented super-pixels. Since the low-rank property emphasizes much more on stronger similarities among super-pixels, we naturally obtain a scale space along the rank dimension in this way. Multiscale saliency can be obtained by simply computing differences among the low-rank components across the rank scale. We conduct extensive experiments on some public benchmarks, and make comprehensive, quantitative evaluation between our method and existing state-of-the-art techniques. All the results demonstrate the superiority of our method in accuracy, reliability, robustness, and versatility.**

***Index Terms*— Structure-sensitive descriptor, multi-level low-rank decomposition, salient object detection, visual saliency.**

I. Introduction and Motivation

SALIENCY can be considered as certain state/attribute of a region that can be utilized to clearly distinguish itself from its vicinity. It has become a standard out-of-the-box toolkit in many low-level computational vision tasks such as image resizing [7], [8], image retrieval [9], [10], image stylization [11], [12], object detection [13]–[15], etc.

Manuscript received October 14, 2013; revised October 3, 2014 and January 26, 2015; accepted February 9, 2015. Date of publication February 12, 2015; date of current version April 15, 2015. This work was supported in part by the U.S. National Science Foundation under Grant IIS-0949467, Grant IIS-1047715, and Grant IIS-1049448 and in part by the National Natural Science Foundation of China under Grant 61190120, Grant 61190121, Grant 61190125, and Grant 61300067. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Anuj Srivastava. *(Corresponding author: Shuai Li.)*

C. Chen, S. Li, and A. Hao are with the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China (e-mail: lishuai@buaa.edu.cn).

H. Qin is with the Department of Computer Science, Stony Brook University, Stony Brook, NY 11790 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2015.2403232

From the point of view of human visual system, saliency means the most distinctive object that could be easily captured by our human vision system from the scene. The human visual system usually judges the image importance by focusing attention either on distinctive, unpredictable, rare, and surprise regions, or on distinctive texture patterns. Humans can routinely and effortlessly perceive saliency out of the abundant inter-twined information. Specially, earlier research about the visual system and human brain conducted by Dale Purves [16] has indicated: (1) Only parts of human cerebral cortex neuron cells will become active once the vision system capturing some distinctive objects; (2) A particular group of neuron cells, called Orientation Selective Neuron Cell, will become more active when the object captured by the vision system has direction-specific closed boundary. Therefore, highly-effective saliency detection algorithms should resort to meaningful and powerful feature metrics to facilitate uniqueness measurement and warrant sufficiently discriminative power. Accordingly, saliency detection methods can be roughly classified into two categories: local-level contrast based methods and global-level uniqueness based methods.

For the local-level contrast based methods, the commonly-used feature attributes include color, gradient, edge, contour, frequency spectra/coefficient, and even their combinations [17]–[21]. These methods usually employ rarity statistics [22]–[24], mutation degree analysis [25], [26], and prior knowledge learning [4], [20] to further boost the saliency detection. However, due to their overly-emphasized local significance or global rarity, the saliency detection quality of such methods tends to solely depend on the original image contents. In principle, they still suffer from the following problems: (1) Naive rarity statistics on local-level attributes gives rise to much tiny false-positive saliency with messy distribution; (2) The mutation degree analysis within local regions tends to over-emphasize the object boundaries, while leaving the inner regions of the object being undetected; and (3) Prior knowledge based learning/regression significantly depends on the quality and scale of the training samples as well as the sophisticated tuning of the underlying classifier parameters.

In contrast, as for the global-level uniqueness based methods, their central ideas are to first describe the sub-part by considering the attributes in a relatively larger neighboring region with histogram-like statistics, and then determine the saliency region by globally comparing the local statistics. Despite the improved success of such methods,

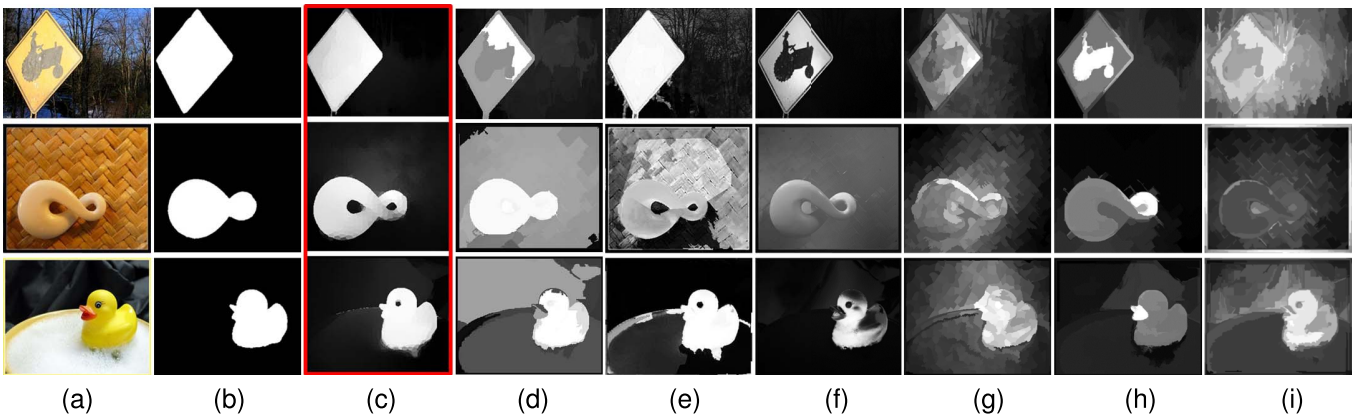


Fig. 1. Saliency detection result comparison among different methods. (a) The source images **SRC**, (b) The ground truth **GT**, (c) Our method, (d) Hierarchical saliency detection method **HS13** [1], (e) Mid-level cues based method **MC13** [2], (f) Saliency filter method **SF12** [3], (g) Low-rank matrix recovery based method **LK12** [4], (h) Contour based method **CB11** [5], and (i) Region contrast method **RC11** [6].

certain difficulties still prevail for texture-pattern saliency detection [27], [28], since the repeated patterns of such sub-parts in a global scope will greatly deteriorate their rarity and uniqueness measurement.

To ameliorate, with the ever-improving capability to globally analyze the linear correlation of the image sub-regions, in most recent years the low-rank decomposition [28], [29] has been employed for saliency detection and achieved great success with growing interest. Generally speaking, low-rank decomposition based saliency detection methods [4] usually employ learning based feature transformation to reduce the correlations between feature representations of non-salient background. The principle behind it is that, there may exist strong correlation among the feature representations of non-salient background, thus feature representations of salient objects can be easily regarded as the sparse part [23], [28] by applying the low-rank decomposition.

However, as shown in Fig. 1f, directly using low-rank decomposition in color space or other accompanying attribute space of images still encounters some difficulties for the task of meaningful and reliable saliency detection. On the one hand, they can not distinguish the sub-parts that locally have similar structures but globally distinct topological relationship from the real salient candidates, since such sub-parts may have completely linear correlations in the aforementioned attribute space. On the other hand, it is hard to separate the spurious small salient parts from noise, since noise is usually independent of small-scale salient parts and thus will be left out as sparse components during the low-rank decomposition.

Based on the above rationales, we summarize the technical challenges existed in most of the state-of-the-art methods as follows. First, it lacks of an intrinsic and informative attribute descriptor to serve as the structural feature carrier for saliency measurement. A new descriptor is required to simultaneously encode the local structural feature and global topological information for the cross-scale / cross-range uniqueness and rarity measurement. Meanwhile, it should have robust and local-transform / isometric-deformation invariant characteristics for similar texture pattern recognition and denoising capability. Second, more intelligent and more versatile global relationship analysis model deserves to be further explored for the robust saliency detection that has scale-aware semantic meanings. Third, considering the topology-free property of the image contents and the extra computational cost caused by the less-significant features, a more efficient image-content-driven geometrical analysis method with explicit physical meanings is urgently needed.

To tackle the aforementioned challenges, we focus on scale-aware, structure-sensitive, robust, and versatile saliency detection (see Fig. 1c) by introducing multi-level low-rank decomposition analysis in the intrinsic feature space. And the intrinsic feature representation is expected to possess the following attributes: (1) Salient objects should have totally different feature representation compared with non-salient background; (2) Non-salient background should have similar feature representation (to be easily regarded as the low-rank part); (3) Inner regions of salient objects should have distinctive feature representation (to be easily regarded as the sparse part). Towards this goal, we first over-segment image into regular super-pixels to remove trivial details, while the original image topology is well maintained. Then, based on the topology of super-pixels, we define a physics-based diffusion metric in 2D image space, and design a new feature representation based on iso-line shape measurement for each super-pixel. Since this metric is structure-sensitive, the feature representation of the object's inner-regions can be easily distinguished from the background surroundings. Specifically, the contributions of our work can be summarized as follows:

- We formulate a physics-based anisotropic geometrical analysis model to automatically represent objects and their surroundings respectively in a total different way. Meanwhile, this model also naturally integrates both local and global information in a multi-scale manner which provides the possibility to capture the saliency in a multi-level solution.
- We propose a novel scale-aware saliency detection method by integrating the intrinsic super-pixel descriptor into a newly designed framework via multi-level low-lank decomposition. As a result, we are able to capture the top-down saliency based on the transformation of

similarity level in our intrinsic feature space spanning from local to global scales.

II. Related Work

A. Local-Level Contrast Based Method

The central idea of the local-level contrast based methods originates from the definition of saliency: being significantly distinctive from non-salient surroundings. To distinguish the salient sub-parts from the less-dramatic background, local comparisons are always helpful. For example, Itti et al. [18] proposed to perform image saliency detection by using mutation degree analysis for center-surrounding operators, Einhauser et al. [30] proposed to perform image saliency detection by considering the luminance contrast, and other similar methods include pixel color comparison based method [17], local structural feature based method [26], and biological feature based local-contrast approach [31]. Although specific local features concentrate on different saliency aspects in such methods, purely performing contrast comparison over local features may lead to false-positive sub-parts, since some scattered tiny sub-parts may be easily deemed as saliency due to their high contrast. To alleviate, Goferman et al. [32] proposed to incorporate global constraints and semantic priors into local contrast analysis. Similarly, instead of semantic priors, Perazzi et al. [3] integrated the local contrast with the global distribution as the saliency criterion. Nevertheless, both methods produce side effects, which further deteriorate the discriminative power of the intrinsically unreliable features and thus lead to the missing of certain significant saliency. Jiang et al. [5] proposed to use the local contrast as clue for the computation of the globally optimal object contour. However, since their optimal object contour heavily depends on the result of local contrast based saliency detection, incorrect saliency from local contrast comparison or the existence of multiple salient objects definitely deteriorates their performance.

B. Global-Level Uniqueness Based Method

Taking the global uniqueness as a major consideration, researchers have started to pay more attention to the detection of the most distinctive sub-objects in a global scope [6], [33]. For example, Hou et al. [34] proposed to conduct saliency measurement by taking into account the log-spectrum deviation of the image patch within one or more images, while Sun et al. [23], [35] resorted to the uniform sparse coding representation of the color related attributes. Similarly, Xie et al. [2] integrated the smoothing constraints into the framework of sparse coding, and mid-level cues originating from varying super-pixel size are also taken into consideration. However, the key-point based convex hull is indispensable to drive their integration framework to capture the saliency, and poor performance occurs when detected key-points differs from the true saliency. Besides, Yan et al. [28] measured the global uniqueness by further extending sparse coding based models to low-rank matrix recovery based ones. However, such methods usually fail to detect the relatively small-scale salient elements and the repetitive texture-pattern saliency, because the descriptions of the similar salient sub-parts in uniform feature space still have strong linear correlation. Then, Shen et al. [4] alleviated the problem of repeatable feature description by additionally introducing a learning based linear transform. Although such improvement does lead to better results, it never fundamentally solves the prior problem, because the introduced linear transform may damage the feature discrimination. Hence, it outputs faulty saliency detection results and blurs the boundaries of the salient object, wherein the flaws in these details inevitably result in new difficulties for downstream applications. Most recently, Yang et al. [36] proposed to construct multi-scale saliency space by exploring the uniqueness of four boundary nodes of the constructed image graph. Similar to [36]. Yan et al. [1] proposed to detect multi-scale saliency via integrating local and global uniqueness clues based on super-pixels with varying size. In addition, Li et al. [37] proposed a multi-scale saliency detection method by defining sparse representation and PCA based reconstructing errors as saliency measurements, together with a possible integration of pixel-level error propagation based refinement and Bayesian framework based measurement, in such a way they can obtain robust saliency detection results.

C. Brief Summary

Although most of the state-of-the-art saliency detection methods are competitive, they are rather operating in isolation, and they are not very well integrated. As a result, they are still struggling to make trade-off between the three major indicators: local mutation, global uniqueness, and the rarity scope. Our observation is that, both additional specific constraints and scale-free mathematical models are difficult to simultaneously conform to the definition of saliency, which will inevitably produce unpredictable side effects. Therefore, strongly inspired by the above observations, this paper focuses on the comprehensive exploration and integration of intrinsic multi-scale feature descriptor (in Section III) with the multi-level low-rank analysis model (in Section IV) for robust scale-aware salient object detection.

III. Structure-Sensitive Descriptor

Consider the insights revealed by Dale Purves [16], the saliency detection should follow the sparse mechanism of human vision system as much as possible by designing specific feature representation to make the salient object automatically distinctive from the scene background. This requires the feature descriptor should have following advantages: (1) It should be affine transformation invariant and isometric deformation invariant, which is expected to suppress the saliency value of tiny object with multiple occurrences; (2) It should very well integrate both local and global information, which is expected to facilitate the saliency scope determination; (3) It should be robust to noisy corruption; and (4) It should be easy to enable parallel computation. However, because of the messy color distribution and the unpredictable global shape correlation of the objects embedded in the natural image, the commonly-used color-based feature is hard to achieve the distinctive representation,

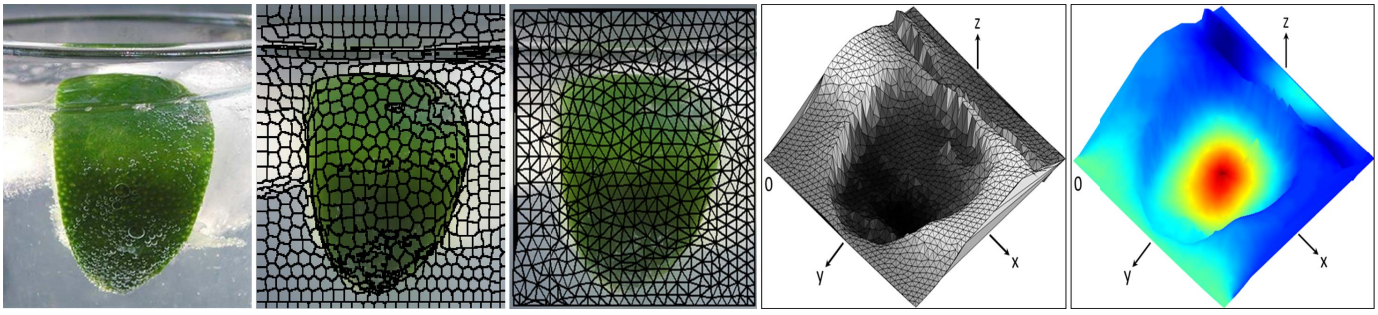


Fig. 2. Pipeline of our bi-harmonic distance metric computation. (a) Original image, (b) Super-pixel segmentation, (c) Delaunay triangulation, (d) Manifold mesh construction, (e) Bi-harmonic distance field.

especially for the statistics-based features derived from single image. Therefore, we propose to use mid-level super-pixels to eliminate some tiny color differences inside an object. To robustly measure the global shape correlation, motivated by the heat diffusion based metric definition in our prior works [38], [39], we further define super-pixel bi-harmonic distance field to depict the topology of super-pixels. In this section, we will focus on the details of our structure-sensitive descriptor defined in local-information and global-correlation integrated intrinsic feature space.

A. Bi-Harmonic Distance Metrics on Manifold Spanned by Super-Pixels

The bi-harmonic distance metric [40] has achieved great success in geometry processing, because it has many built-in advantages, such as the natural integration of local and global information, being structure-sensitive and parameter-free. Specifically, the feature representation based on this intrinsic metric exhibits high discriminative power, which has great potential to facilitate the salient object separation from its non-salient background. However, it remains difficult to directly define this powerful intrinsic distance metric over the 2D images due to the following reasons: (1) 2D images comprising regular pixels are both topology-free and boundary-free without any intuitive geometric meaning. This unavoidably hinders the rigorous and reliable Laplacian differential analysis required in the definition of bi-harmonic distance based on heat diffusion; (2) It is impractical to directly employ the pixel as a basic unit towards meaningful differential analysis, since the pixel-level Laplacian matrix does not support multi-scale functionality that is highly demanded in any novel shape descriptors. Strongly motivated by the need of novel suitable descriptors for salient object detection, we elaborate our new intrinsic shape descriptor on a manifold space enabled by the construction of super-pixels.

As shown in Fig. 2, in the interest of maintaining global topology information (which is essential for bi-harmonic metric measurement), while omitting unnecessary details (it may be noted that, trivial details occasionally influence the global topology), we first decompose the 2D image into over-completely segmented super-pixels (the number of super-pixel is 900), and then convert the super-pixels to a 2D manifold embedded in 3D space. In order to guarantee the regularity of the succeedingly constructed manifold, here we employ the SLIC method [41] to conduct relatively uniform segmentation for super-pixels (The reasons and details are detailed in Section VI-A).

Meanwhile, to respect the anisotropic property exhibited in original color space of the image, we inherit the anisotropy by taking the average intensity (I_s) of each super-pixel as its third dimensional coordinate in 3D space. Therefore, with each super-pixel's geometric center serving as a 3D vertex, the manifold mesh corresponding to each image can be constructed by Delaunay triangulation (Fig. 2(c)), wherein the 3D coordinate of each vertex is represented as (x, y, I_s). With the vertex set of the manifold mesh (Fig. 2(d)) denoted by $P = \{p_1, p_2, ..., p_n\}$, we now define the bi-harmonic distance metric via discrete Laplacian-matrix $L = A^{-1}M$ based anisotropic heat diffusion [42], [43], where A is a diagonal matrix and A_{ii} is proportional to the average area of the triangles sharing vertex p_i. And M is formulated as

$$M_{ij} = \begin{cases} \sum_k m_{i,j} & \text{if } i = j \\ -m_{ij} & \text{if } p_i \text{ and } p_j \text{ are adjacent} \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where $m_{ij} = \cot \alpha_{ij} + \cot \beta_{ij}$, α_{ij} and β_{ij} are the opposite angles of two adjacent triangles sharing edge $p_i p_j$. To better respect the color image, we use $|r_i - r_j| + |g_i - g_j| + |b_i - b_j|$ to calculate the distance of the color component that is embedded in the edge length, where (r, g, b) denotes the average color value of the super-pixel p.

So far, we can formulate the bi-harmonic distance between super-pixel p_i and p_j as

$$D(i, j)^2 = \sum_{k=1}^{K} \frac{(\phi_k(i) - \phi_k(j))^2}{\lambda_k^2}, \tag{2}$$

where $\phi_k(i)$ and λ_k^2 respectively denote the k-th eigenvector and eigenvalue among K adopted smallest eigenvalues.

Therefore, we can compute a corresponding bi-harmonic distance field for each super-pixel. Take the super-pixel located at the image center as an anchor vertex, Fig. 2(e) demonstrates the bi-harmonic distance distribution, wherein the color ranging from red to blue means that the bi-harmonic distance goes from the near to the distant.

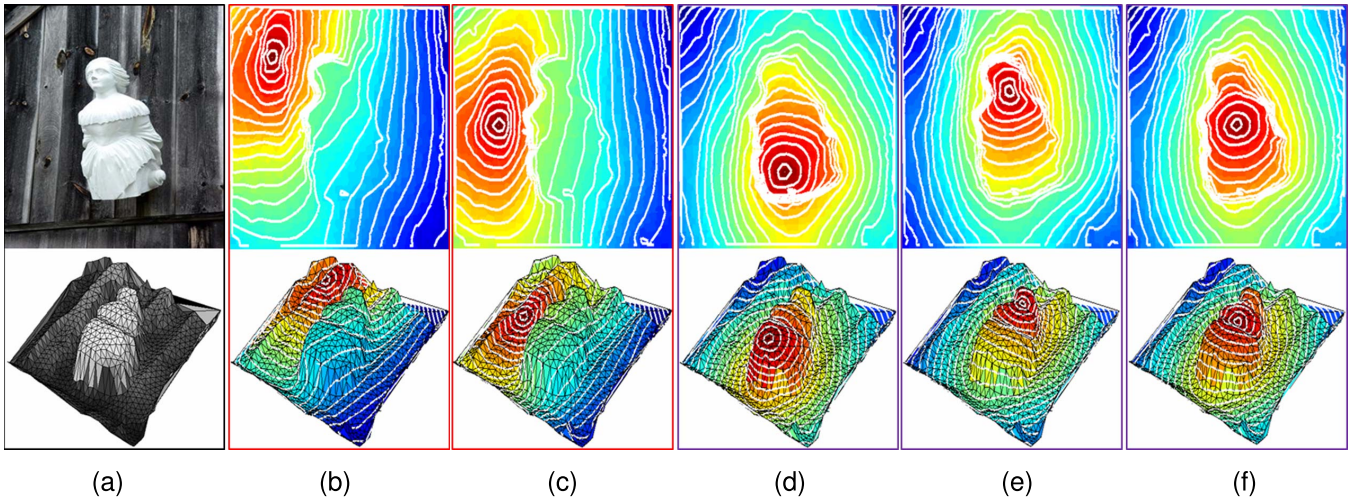


Fig. 3. The 2D (top row) and 3D (bottom row) illustration of bi-harmonic iso-line shape context based descriptor. (a) Original image and its corresponding manifold mesh, (b-c) The iso-lines with non-salient super-pixels as anchor points, (d-f) The iso-lines with salient super-pixels as anchor points.

The biggest advantage of introducing bi-harmonic distance metric into 2D image space is its characteristic of being structure-sensitive, which can be easily observed in Fig. 2(e) where the major bi-harmonic distance distribution is located inside the object, and large gap exists along the object boundaries.

B. Iso-Line Shape Context Based Descriptor

Given the bi-harmonic distance field corresponding to a specific super-pixel, it encodes both the local geometrical structure and the global topological relationships with other super-pixels in an elegant way. However, the obtained bi-harmonic distance field is super-pixel-wise and discrete, it needs to be further exploited for the saliency-centered feature representation. In fact, the local-to-global principal diffusion tendency encoded in bi-harmonic distance metric can be represented by considering scope-increasing sub-groups of key super-pixels that possess identical bi-harmonic distance values, and the iso-line contour shape of such sub-group of super-pixels can sufficiently represent the bi-harmonic diffusion patterns at different scales (from local to global). Therefore, to fully respect the aforementioned insights noted by Dale Purves [16], as shown in Fig. 3, we propose a bi-harmonic iso-line shape context based descriptor, whose central idea is to compute and integrate the probability distribution of the iso-lines ranging from inner-ring to outer-ring.

First, to obtain a bi-harmonic iso-line over the triangular mesh, we interpolate the bi-harmonic distance values according to the distance values of the triangular vertices w.r.t. a specific iso-value, and gather the interpolated points for further filtering in a triangle-wise fashion. Since different iso-lines should not cut across each other, the following interpolated points should be eliminated: (1) If there exist more than two points (belonging to different triangle edges respectively) which have equal distance values (or similar values), and please refer to Section V for more details; (2) Those points which can not match any other points with equal distance values. After this filtering process, the remaining points are used to construct different iso-lines according to their distance level (see details of the formulation in Section V).

Fig. 4. Illustration of iso-line shape context computation. (a) The source image **SRC**, (b) The bi-harmonic iso-line distribution of a given initial start vertex. For each iso-line, e.g., the one marked in black color, the Euclidean distances of arbitrary iso-points pair (tiny purple cycles) are computed for further histogram analysis.

Next, for each iso-line, we compute the Euclidean distance for each iso-point pair (see Fig. 4), and further compute probability distribution statistics for such normalized Euclidean distances. Therefore, the distribution shape of each iso-line can be represented in the form of histogram. Finally, we form the super-pixel-specific descriptor by concatenating the histograms of multiple normalized iso-lines into a high dimensional vector. Since the most outer-ring (far away from the start vertex) of iso-lines are not reliable due to the image boundary effects, we empirically take the 15 inner-ring iso-lines over total 30 ones into account in our implementation (details of parameter selection can be found in Section VI-A).

It should be noted that our descriptor has many unique advantages as follows. (1) Our descriptor has multi-scale discriminative power because of the natural integration of local (the inner-ring iso-lines) and global (the outer-ring iso-lines) information. In other words, either from the global perspective or from the local perspective, the descriptions of the super-pixels inside the salient object (Fig. 3(d-f)) are totally

different from those non-salient regions (Fig. 3(b-c)). Directly benefitting from this property, the overlapping representation in traditional feature space is perfectly avoided. (2) With the help of the local structure-aware characteristics, element description for sub-parts belonging to the identical object exhibit obvious differences (see inner iso-lines in Fig. 3(d-f)). Therefore, regions inside the salient object can be easily regarded as sparse parts by applying low-rank decomposition on the entire description matrix (see details in Section IV). (3) Our descriptor is rotation and scale invariant because of the utility of the normalized probability distribution statistics, and this property contributes tremendously in suppressing duplicated non-salient patterns. These attractive advantages collectively facilitate the robust scale-aware saliency detection, which will be discussed in details in the following sections.

IV. Scale-Aware Salient Object Detection

Based on our structure sensitive descriptor, we primarily concentrate on the scale-aware saliency detection by exploiting the powerful capability of low-rank decomposition in this section.

A. Saliency Capture Based on Low-Rank Decomposition

Since saliency detection aims to detect the most distinctive things from their surroundings, our purpose is to measure the rareness of each super-pixel according to its cross-scope bi-harmonic iso-line shapes. Consider the sparsity-related insight noted by Dale Purves [16], we employ low-rank decomposition over the feature matrix F to obtain its low-rank component and sparse component, and then use the sparse matrix S as the saliency indicator, because the low-rank component means commonly-occurring object while the sparse component represents the rare/distinctive object in some sense.

Since each super-pixel is represented as a high-dimensional descriptor f_i, we can reorganize the 2D image in the form of descriptor matrix as $F = [f_1, f_2, \ldots, f_n]$ (see F Matrix in Fig. 6). From the viewpoint of the matrix decomposition, the matrix F can be divided into a low-rank component and a sparse component according $F = L + S$, and L and S respectively corresponds to the correlated elements and independent elements. The traditional low-rank decomposition is usually defined as

$$(L^*, S^*) = arg \min_{L,S} (rank(L) + \lambda \|S\|_0). \tag{3}$$

However, this problem is NP-hard, but this problem can be approximated by the nuclear norm $||L||_*$ and L_1-norm $||S||_1$ by the following formulation:

$$(L^*, S^*) = arg \min_{L,S} (||L||_* + \lambda \|S\|_1). \tag{4}$$

And there are several mature methods that can be utilized to solve the aforementioned problem, such as the bilateral random projection (BRP) based method [44], [45] and the robust principal component analysis (RPCA) [46]. As shown in the middle row of Fig. 5, for salient super-pixels, the diffusion patterns of inner-ring iso-lines exhibit strong anisotropic property. In contrast, as shown in the bottom row of Fig. 5, the diffusion patterns of both inner-ring and outer-ring iso-lines exhibit strong correlation, wherein (**B1**) is similar to (**B2**), (**B2**) is similar to (**B3**), (**B3**) is similar to (**B4**), and so on. Therefore, when considering the rareness attribute of salient object, sparse matrix S obtained from Eq. (4) can very well indicate the saliency level of its corresponding super pixel. More residuals in matrix S mean higher saliency value, and vice versa. Thus, we employe the L_1-norm of each column in matrix S for the saliency assignment of each super-pixel, and please refer to the sub-figure (**T2**) in the top row of Fig. 5 for details. Furthermore, we conduct sparsity propagation in the neighborhood via the following equation:

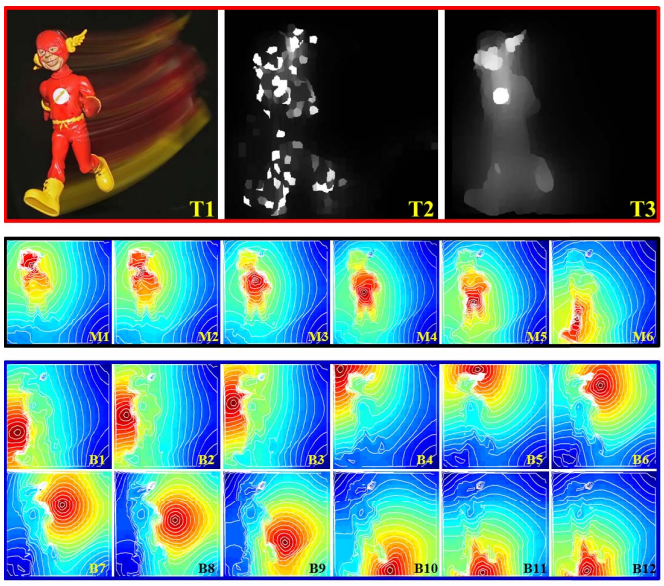


Fig. 5. Illustration of the effect of Sparse Matrix in saliency detection. Top row: **T1** is the source image, **T2** is the illustration of the Sparse Matrix entries' values, and **T3** is the illustration of the propagated values of the Sparse Matrix. Middle row (**M1-M6**) demonstrates the bi-harmonic diffusion patterns of **salient** super-pixels. The bottom row (**B1-B12**) demonstrates the bi-harmonic diffusion patterns of **non-salient** background.

$$S_i = \frac{1}{Z} \sum_{j \in D} S_j \times \omega_j. \tag{5}$$

Here $D = \frac{1}{6} \times min(W, H)$ controls the propagating distance, W, H respectively represents the width and height of the input source image, $Z = \sum_{j \in D} \omega_j$, $\omega_j = exp(-\gamma \times (|R_i - R_j| + |G_i - G_j| + |B_i - B_j|))$, R_i, G_i, B_i denote the average color of the i-th super-pixel, and we empirically set $\gamma = 4$. The propagated values of the sparse matrix are illustrated in the sub-figure (**T3**) in the top row of Fig. 5.

Being represented by our novel descriptor, super-pixels at the very center of salient object (i.e., the minority with uniqueness topology position) tend to have highest anisotropic strength. The overall anisotropic distribution has the following property (see Fig. 3 and Fig. 5):

$$A_{inner-S} > A_{outer-S} > A_{non-S}, \tag{6}$$

where $A_{inner-S}$ indicates the anisotropic strength of the most centering super-pixels of salient object, $A_{outer-S}$ indicates the

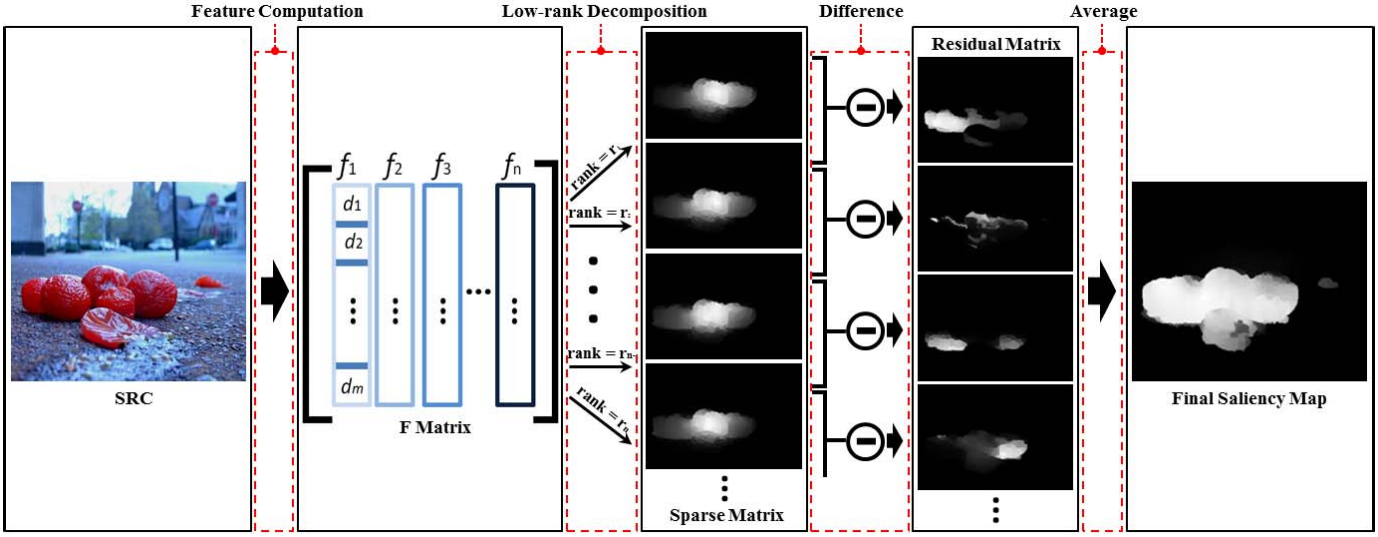


Fig. 6. Pipeline of multi-level low-rank decomposition based saliency detection. We first use our novel descriptor to represent the source image **SRC** (***Feature computation***). Then, we apply the ***Low-rank decomposition*** to get the **Sparse Matrix** at different rank level, and seek the ***Difference*** among these **Sparse Matrices** to obtain **Residual Matrices**. Finally, the **Final Result** is computed based on the ***Average*** of all residual matrices.

anisotropic strength of the rest super-pixels inside the salient object but near the object boundaries, and A_{non-S} represents the non-salient background. The anisotropic strength of our intrinsic feature space gradually declines from inner regions to boundary regions of the salient object, and finally becomes isotropic for non-salient background. Obviously, this elegant property guarantees that the most inner regions of salient object can be easily identified as the sparse parts by low-rank decomposition and have high precision rate. However, due to various sizes and structures of salient objects, there should exist no magic rank level to fully recover the entire salient object, and poor recall rate occasionally arises if directly applying the low-rank decomposition as the saliency criterion (Fig. 9c). Therefore, we propose multi-level low-rank decomposition to overcome this limitation, which will be discussed in the next section.

B. Multi-Level Low-Rank Components and Their Operations

As mentioned in Section III, apart from the highly discriminative power for salient object and non-salient background, another advantage of our descriptor is its simultaneously integration of both local and global information. From the perspective of single super-pixel inside the salient object, its anisotropic strength also exhibits a sharp change from local anisotropy (determined by topology position) to global isotropy (sharing the identical contour of the salient object with the others). Although non-salient super-pixels also enjoy anisotropic strength change, considering that non-salient background is boundary-free and salient object has closed contour, the anisotropic strength changing rate of salient super-pixels is much higher than that of non-salient background. With respect to this characteristic, we apply the "extent" rank strategy, called multi-level low-rank decomposition, instead of the traditional "slice" rank strategy [4], [28]. Hence, we can capture the saliency based on anisotropic strength changing rate.

Traditional methods tend to solve the aforementioned low-rank decomposition problem by minimizing the sum of the kernel norm and the L_1-norm without the explicit control on the rank level. In sharp contrast, we define our multi-level low-rank decomposition model as:

$$F = L + S + G, \quad \text{s.t. } rank(L) \le r,\ card(S) \le c, \tag{7}$$

where G is the error matrix. To solve this problem approximately, we use the following formulation to settle the low-rank decomposition problem:

$$\begin{aligned} &\min_{L,S} ||F - L - S||_F^2, \\ &\text{s.t. } rank(L) \le r,\ card(S) \le c, \end{aligned} \tag{8}$$

where $||\cdot||_F$ is the Frobenius norm, the rank constraint r and the cardinality constraint c are used to explicitly control the low-rank degree and the sparse degree respectively. The optimization problem of Eq. (7) can be solved by employing the one-fix-another-solved strategy iteratively as follows:

$$\begin{cases} L_t = \underset{rank(L)\le r}{\arg\min}\ ||F - L - S_{t-1}||_F^2 \\ S_t = \underset{card(S)\le k}{\arg\min}\ ||F - L_t - S||_F^2 \end{cases}, \tag{9}$$

where the lower rank constraint r intrinsically emphasizes the stronger similarities among super-pixels, meanwhile, the sparse constraint k provides a flexible way to control the uniqueness degree of the saliency to be detected in the global setting. Since the motivation of our multi-level low-rank decomposition is to explore the saliency in the perspective of the changing rate of anisotropic strength, we naturally obtain a scale space along the rank dimension by gradually varying the rank constraint r and setting sparse constraint k with a hard threshold to alleviate the complexity. As for computation, we employ the GoDec method [45] to efficiently accelerate this optimization process.

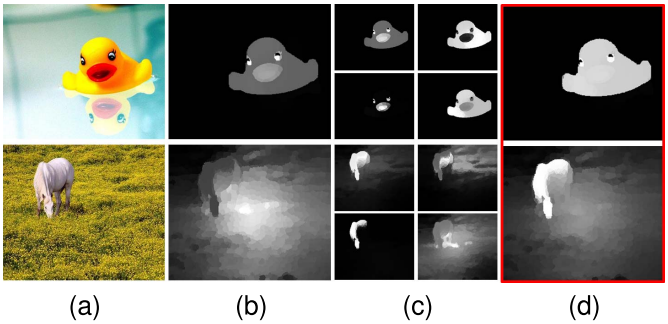


Fig. 7. Demonstration of the performance improvement benefitting from our local and global features tightly coupled with the multi-level low-rank analysis framework. (a) Original image, (b) Saliency map resulted from the RPCA based low-rank method [4], (c) Different-level saliency map resulted from the BRP based low-rank method, (d) Saliency map resulted from our method.

After the low-rank decomposition, we use the following equation to assign saliency value for each super-pixel:

$$S_i^r = \sum_{j=1}^{D} |S^r(j,i)|, \tag{10}$$

where S_i^r indicates the saliency value of the i-th super-pixel at rank level r, D is the feature dimension, and S^r is the sparse matrix with rank level r. Then, the residual sparse matrices at different rank level can be computed by:

$$R_i^{(r1,r2)} = |S_i^{r_1} - S_i^{r_2}|/Z, \tag{11}$$

where Z is the normalization factor, $S_i^{r_1}$ and $S_i^{r_2}$ are saliency value of the i-th super-pixel at rank level r_1 and r_2 respectively. It is apparent that, the residuals of sparse matrices under different rank level r (see Fig. 7c) natural indicate the changing rate of anisotropic strength, and we utilize the average of multiple residual matrices for robust scale-aware saliency (S_f) detection with the following formulation:

$$S_f = 1/N_D \cdot \sum_{p=1}^{N_D} R^p, \tag{12}$$

where S_f is the final saliency map, N_D is the total number of residual sparse matrices, and R^p is the p-th residual sparse matrix. Fig. 6 intuitively shows the pipeline of our scale-aware saliency detection based on our multi-level low-rank decomposition.

In our implementation, we set 12000 as the hard sparse threshold k, and define the variable rank constraint r to be between 5 to 15 (see Fig. 9b). Therefore, 10 residual matrices are obtained for multi-scale saliency measurement. Fig. 7 demonstrates the superiority of our multi-level low-rank decomposition (Fig. 7d) over the traditional solution (Fig. 7b).

V. CUDA Implementation

Since our structure-sensitive descriptor enables parallel computation, we have fully implemented it on CUDA. We first invoke one CUDA thread for each super-pixel (see Fig. 8), of which we interpolate nodes for triangle edges (Step.1 in Algorithm 1), formulate iso-line points (Step.2 in Algorithm 1), and finally describe the context of iso-line shapes (Step.3 in Algorithm 1). Then, the feature

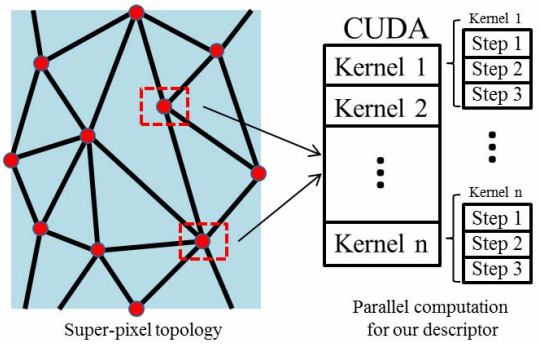


Fig. 8. Demonstration of using CUDA to parallelly compute our descriptor, and super-pixels are denoted by red points.

Algorithm 1 Kernel Function

Input: $T = \{t_1, t_2, ..., t_N\}$, where N is the total super pixel number, and $t_i = \{(V_\alpha, B_\alpha), (V_\beta, B_\beta), (V_\gamma, B_\gamma)\}$ represents triangle whose nodes V is initialized by bi-harmonic value B

Initialization: $L_{num} = 30, L_{int} = 1/L_{num}, \varepsilon = L_{int}/1000$

Output: f_i

Step 1. Interpolate nodes for triangle edges

for each *edge* $[V_m, V_n]$ *in triangle* t_i

interpolate k *nodes* ξ,
where $k = \lfloor \|V_m, V_n\|_2^2 / L_{int} \rfloor$,
$\xi = \{(v_1, b_1), (v_2, b_2), ..., (v_k, b_k)\}$,
$v \in (V_m, V_n)$ *and* $b \in (B_m, B_n)$;

$t_i\{end+1\} = \xi$;

end for

Step 2. Formulate iso-line points

for each $nodes(V, B)$ *in* t_i

if $\mathcal{A}.\ (B_m - B_n) < \varepsilon$, *and*
$\mathcal{B}.\ V_m \in edge[V_\alpha, V_\beta], V_n \in edge[V_\alpha, V_\gamma]$,
but $\beta \neq \gamma$, *and*
$\mathcal{C}.$ *no* V_q *exists, where*
$(B_q - B_m) < \varepsilon$ *or* $(B_q - B_n) < \varepsilon$;

then

$id = \lfloor (B + \varepsilon)/L_{int} \rfloor$;

$ISO_{id}\{end+2\} = [V_m, V_n]$;

end if

end for

Step 3. Describe the context of iso-line shape

for each $ISO_{i=1:15}$

compute L_2 *distance of arbitrary iso-point pairs, and obtaining distance pool* $P\{i\}$;

$H = histogram\ analysis\ for\ P\{i\}$;

$f_i\{end+1\} = H$;

end for

matrix $F = \{f_1, f_2, \ldots, f_n\}$ is obtained by collecting the outputs of each graphic processing unit (GPU). Details of CUDA kernel function is documented in the following algorithm.

The first step of our CUDA kernel function is to interpolate bi-harmonic values for the triangle edges, and these

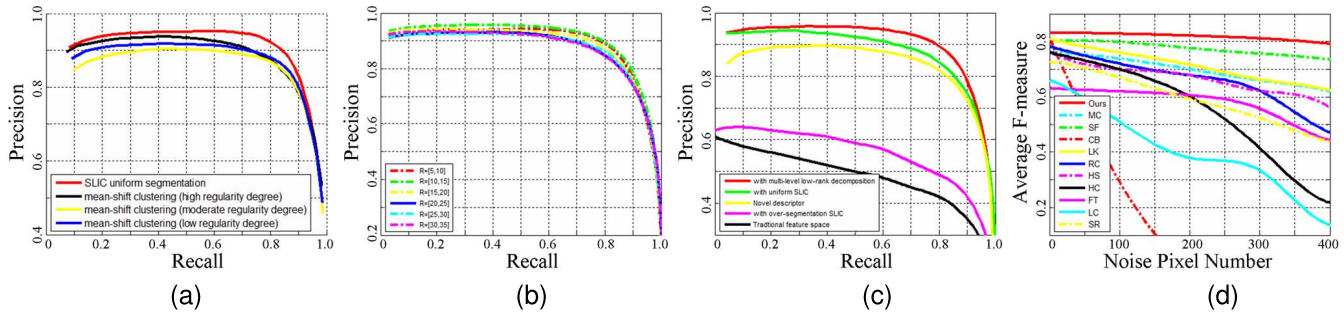


Fig. 9. (a) Average Precision-Recall performance comparison of different super-pixel methods over **Achanta** dataset (1000 images) [19], (b) Precision-recall curves using different extents of rank level for multi-level low-rank decomposition on **Achanta** dataset [19] and **SED1** dataset [47], (c) Precision-recall curves of our method combined with different saliency detection solutions on **Achanta** dataset [19] and **SED1** datasets [47]. The method of using traditional feature is proposed in [28], (d) Saliency detection results and their comparisons over noise-corrupted images. Results from **HS13** [1], **MC13** [2], **SF12** [3], **LK12** [4], **CB11** [5], **RC11** [6], **HC11** [6], **LC06** [33], **FT09** [19], and **SR07** [34] are documented here.

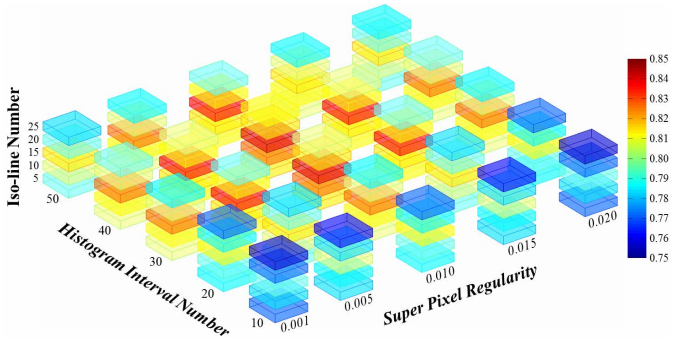


Fig. 10. Quantitative analysis for the parameters' selection, including the interval number used in histogram analysis, the regularity level of super-pixel decomposition, and the iso-line number to represent the bi-harmonic diffusion pattern. The color from blue to red indicates the value of F-measure changes from low to high.

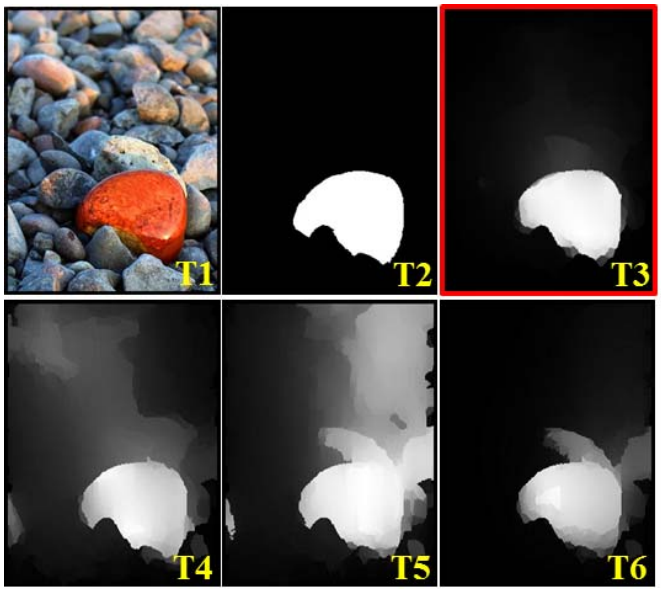


Fig. 11. Demonstration of the saliency maps produced with different-regularity super-pixel configurations, wherein the sub-figure (**T1**) is the input image, (**T2**) is the ground truth, (**T3**) is the saliency map produced by our method via SLIC segmentation [41], while (**T4-T6**) are the results produced by our method via mean-shift clustering based super-pixel segmentation [48], with low/moderate/high regularity respectively.

newly interpolated points will facilitate the construction of bi-harmonic iso-lines. The second step is to eliminate the points which are not belonging to any iso-line, and then formulate iso-lines. Given a specific iso-line, the final step is designed to represent the bi-harmonic diffusion context based on Euclidean distance of arbitrary bi-harmonic point pairs.

VI. Experimental Results and Evaluation

A. Parameter Selection

In principle, there are four parameters influencing the performance of our salient object detection framework: (1) The regularity of super-pixels (Section III-A); (2) The interval number used in histogram analysis (Section III-B); (3) The number of actually used iso-lines (Section III-B); and (4) The rank level range for multi-level low-rank decomposition (Section IV-B). As the first three parameters simultaneously affect the performance of using low-rank decomposition for saliency estimation, we comprehensively test their effects as a whole on the overall performance in order to obtain the optimal solution at the beginning, and later deal with the optimal extent of rank level for multi-level low rank decomposition.

1) The Regularity of Super-Pixels: To better analyze the influence of irregular super-pixel segmentation, we have conducted extensive experiments based on SLIC [41] and mean-shift clustering for super-pixel segmentation [48], and the quantitative comparison results can be found in Fig. 9a. According to our analysis, we adopt a relatively-uniform super-pixel segmentation strategy, because our diffusion geometry based descriptor is heavily dependent on the underlying manifold mesh quality of the original 2D image. And the irregular manifold mesh also has negative impact on the low-rank hypothesis of non-salient background. Yet, the SLIC decomposition with high-regularity strength can also make the contours of salient objects obscure and deteriorate the distinguishing power of our structure-sensitive descriptor. Extensive testing results shown in Fig. 10 suggest that setting the optimal SLIC super-pixel regularity level to be 0.01 can make the best tradeoff. It is apparent that, the SLIC based uniform super-pixel segmentation gives rise to better

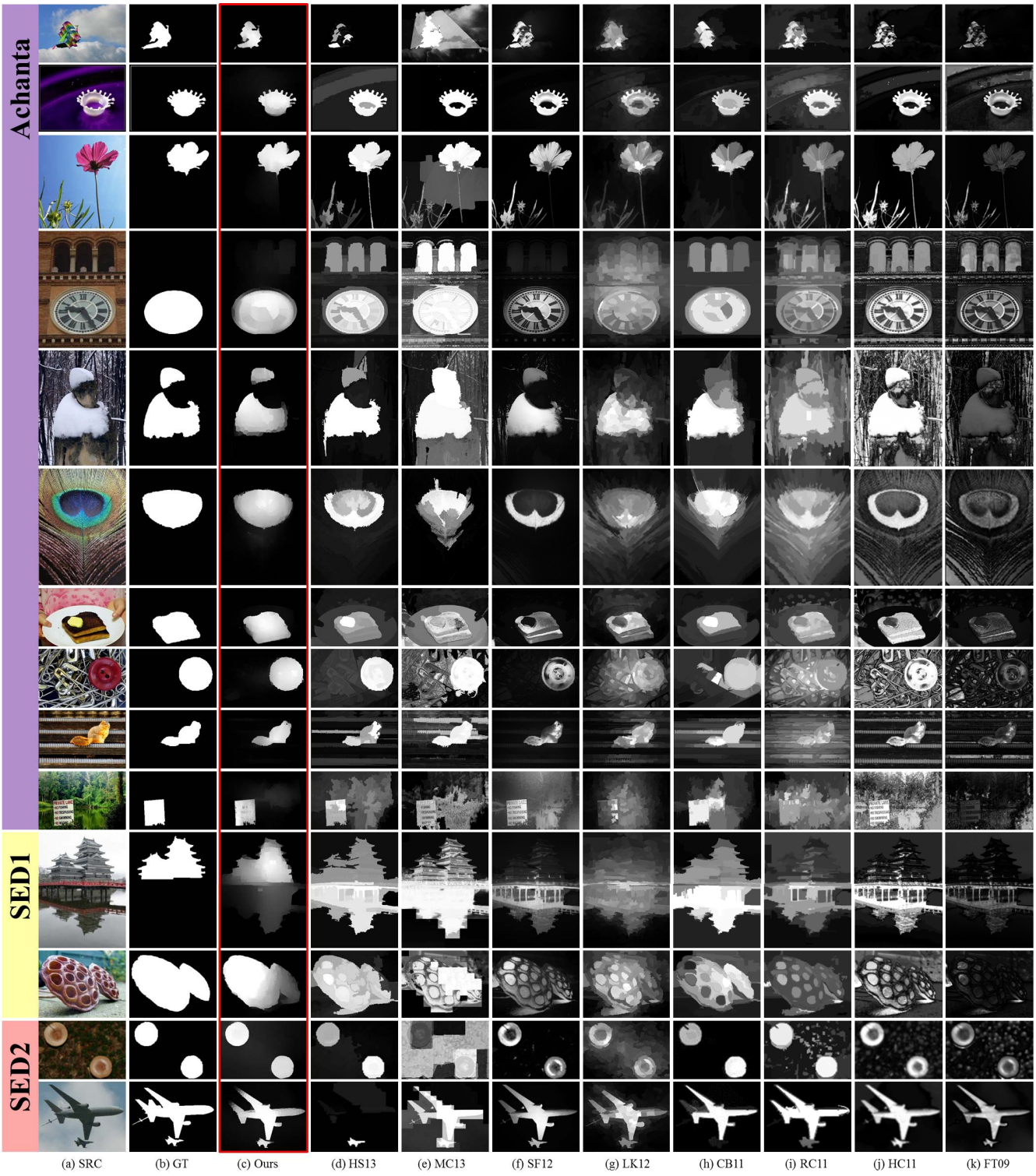


Fig. 12. More saliency detection results and their comparisons among our method, ground truth (**GT**), and current state-of-the-art methods, including **HS13** [1], **MC13** [2], **SF12** [3], **LK12** [4], **CB11** [5], **RC11** [6], **HC11** [6], **FT09** [19].

performance (with 900 super-pixels), which outperforms mean-shift based super-pixel segmentation because of the relative regularity of SLIC. And the average performance testing over **Achanta** dataset indicates that the performance will deteriorate when the super-pixel's regularity decreases (it may be noted that, in Fig. 9a, high regularity case has almost 900 super-pixels, moderate regularity case has almost 700 super-pixels, and low regularity case has almost 500 super-pixels). Correspondingly, Fig. 11 demonstrates the different saliency maps produced by our method with different-regularity super-pixel configurations.

2) The Number of Actually Used Iso-Lines: Based on our observation, dense iso-lines can precisely represent the bi-harmonic diffusion pattern, but too many iso-lines will

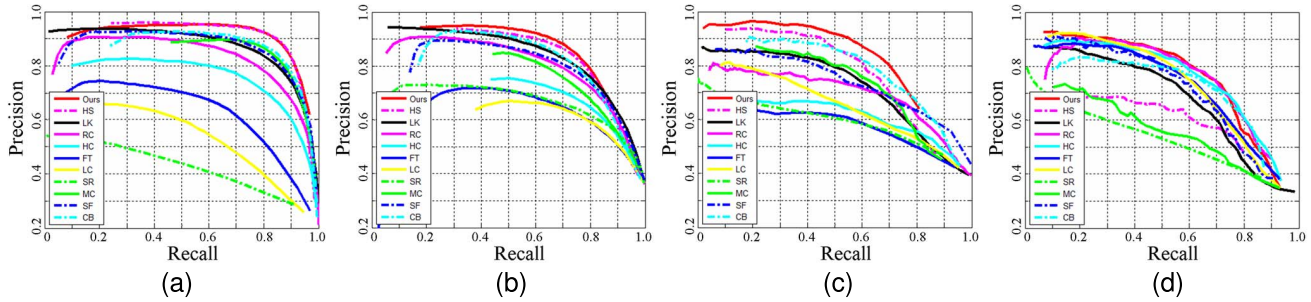


Fig. 13. Precision-recall curves for dynamic thresholding of saliency maps. Different options of our method are compared with **HS13** [1], **MC13** [2], **SF12** [3], **LK12** [4], **CB11** [5], **RC11** [6], **HC11** [6], **LC06** [33], **FT09** [19], and **SR07** [34]. (a) Precision-recall comparisons of different methods based on **Achanta** dataset [19], (b) Precision-recall comparisons of different methods based on **MSRA** dataset [20], (c) Precision-recall comparisons of different methods based on **SED1** dataset [47], (d) Precision-recall comparisons of different methods based on **SED2** dataset [47].

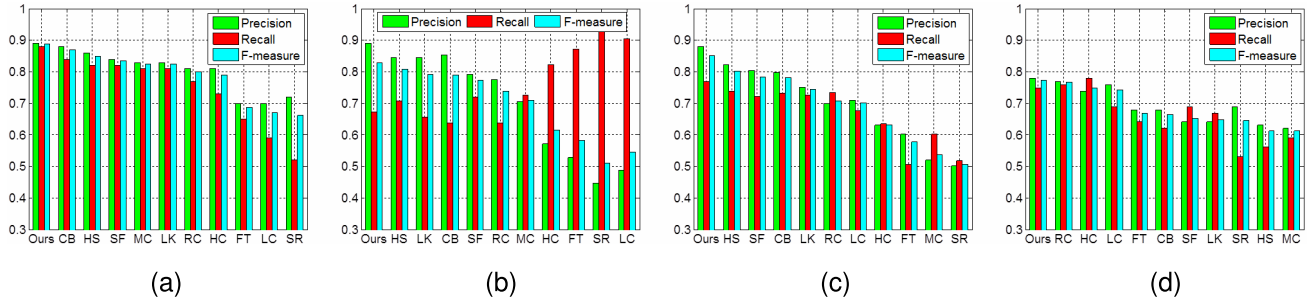


Fig. 14. Average precision, recall and F-measure comparison of different methods on (a) **Achanta** dataset [19], (b) **MSRA** dataset [20], (c) **SED1** dataset [47], and (d) **SED2** dataset [47]. The results of our method are compared with **HS13** [1], **MC13** [2], **SF12** [3], **LK12** [4], **CB11** [5], **RC11** [6], **HC11** [6], **LC06** [33], **FT09** [19], and **SR07** [34].

inevitably increase the computational burden. Therefore, we choose to compute total 30 bi-harmonic iso-lines in our implementation. Since iso-lines near the image boundaries tend to offer meaningless diffusion patterns, only the inner-ring iso-lines are used to represent the bi-harmonic diffusion pattern, and results in Fig. 10 suggest 15 as the optimal choice.

3) The Interval Number for Histogram Analysis: Obviously, for histogram analysis, large interval number leads to strong discriminative power, which can increase the probability of treating non-salient background as sparse part by the low-rank decomposition. Furthermore, large interval number also increases the feature dimension which heavily affects the computational cost. However, small interval number deteriorates the discriminative power. According to our test results in Fig. 10, we select 40 as the optimal interval number for histogram analysis.

4) The Rank Level Range for Multi-Level Low-Rank Decomposition: After the above three parameters are determined, Fig. 9b demonstrates the precision-recall curves using different rank level range, and the results suggest that better results tend to arise with the rank level range [5, 15]. Therefore, we formulate our multi-level low-rank decomposition as follows:

$$S_f = \sum_{r=5}^{14} |S_{r+1} - S_r|, \tag{13}$$

where subscript r indicates the specific rank level, and S_f denotes the final sparse matrix after multi-level low-rank decomposition.

With all parameters being selected, Fig. 9c demonstrates the overall performance of our method combining with different components. From Fig. 9c, it is obvious that both our novel descriptor and multi-level low-rank decomposition can remarkably improve the performance of salient object detection. As shown in Fig. 12, our method can produce less false-alarm and intact saliency.

B. Comparison With Other Methods

In this paper, we evaluate the performance of our method on four public available datasets recommended by recent benchmark [49], including the Achanta [19], MSRA [20], SED1, and SED2 [47]. The Achanta dataset contains 1000 images, wherein each image has only one salient object with ground truth accurately marked. The MSRA dataset contains 5000 images with rectangle ground truth labeled by nine different users. Similar to the Achanta dataset, the SED1 dataset also contains 100 images with one salient object in each image, while the surroundings of the salient object are more complex than those in the Achanta dadaset. Also with complex surroundings, the SED2 dataset contains 100 images with two salient objects in each image.

We conduct extensive experiments and make quantitative comparison with ten current state-of-the-art methods, including spectral residual method (SR07) [34], frequency tuned method (FT09) [19], low level contrast method (LC06) [33], histogram contrast method (HC11) [6], region contrast method (RC11) [6], contour based method (CB11) [5], low-rank matrix recovery based method (LK12) [4], saliency filter method (SF12) [3], mid-level clue based method (MC13) [2] and hierarchical saliency detection method (HS13) [1] to verify and validate our method.

We adopt the precision-recall indicator to conduct evaluation. To facilitate the precision-recall computation, in the

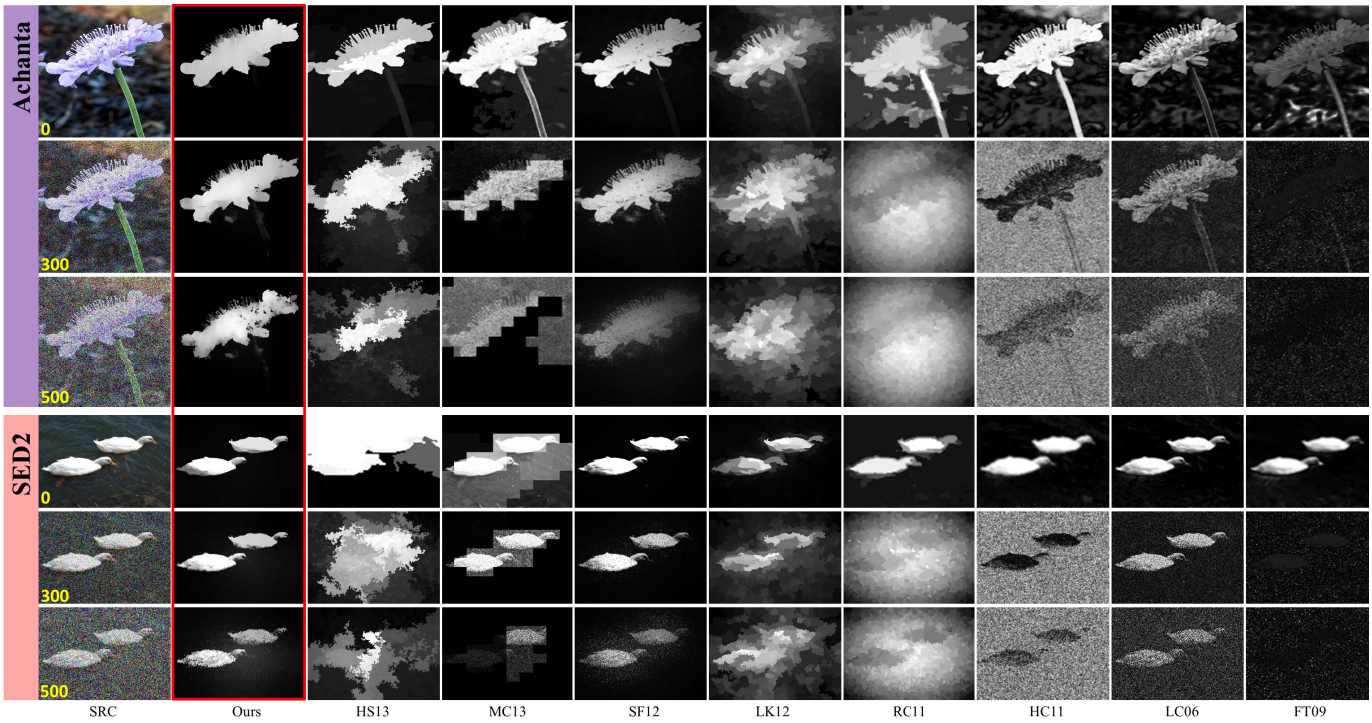


Fig. 15. Saliency detection results and their comparisons over noise-corrupted images. The numbers (colored in yellow) in the first column images indicate the noise pixel number. **SRC** is the source image, and **HS13** [1], **MC13** [2], **SF12** [3], **LK12** [4], **RC11** [6], **HC11** [6], **LC06** [33], and **FT09** [19] are shown here. Specifically, we do not demonstrate the result of **CB11** [5] here, because it is very sensitive to noise, and it becomes unavailable with no saliency value assigned when existing more than 100 noise pixels.

first set of experiments, all the images are further segmented according to the saliency maps of different methods with the same threshold $T \in [0, 1]$, wherein pixels with saliency value larger than T are labeled as foreground. If the obtained foreground pixel is consistent with that in the ground truth mask, it is deemed as real saliency. Specifically, since the ground truth of MSRA dataset are represented with bounding boxes, we follow the validation setting adopted by [49]: given the computed saliency map, find a bounding box that can cover at least 95% saliency pixels, and then calculate the precision-recall scores based on these bounding boxes.

The final precision-recall curves are obtained using the average computation results by varying T from 0 to 1. As the recall rate is inversely proportional to the precision, the tendency of the trade-off between precision and recall can truly reflect the performance. Precision-recall curves in Fig. 13a, Fig. 13b and Fig. 13c quantitatively indicate that our method is much better than other methods in terms of both precision and recall criteria. For multiple salient object detection, methods including HS13, CB11, MC13, SF12, and LK12 have good precision-recall curves on single salient object datasets (Fig. 13a), while their performance tends to deteriorate rapidly in the SED2 dataset (Fig. 13d). In contrast, our method's performance remains to be good. Actually, because of the existence of multiple salient objects, the low-rank hypothesis become invalid occasionally. This might have a negative impact on the claimed advantage of our method (see details in Section VI-C).

In the second set of experiments, we first adopt the public implementation [4] to over-segment all the images. Then, for each segment, we label it as foreground if its average saliency score is greater than the adaptive threshold $T = 2\times$ (the average saliency of all the segments). Fig. 14a, Fig. 14b, Fig. 14c and Fig. 14d document the statistics of all the methods in terms of average precision, recall, and F-measure, where $F = ((\beta^2+1)P*R)/(\beta^2 P+R)$ (P = Average Precision, R = Average Recall) and $\beta^2 = 0.3$. It shows that our method apparently outperforms other state-of-the-art methods. Specifically, even though both our method and that in [4] employ low-rank decomposition for saliency detection, our method can achieve better results.

Besides, Fig. 15 shows the performance comparison with noise-corrupted images, and Fig. 9d demonstrates the variation tendency of average F-measure over varying noise number. Since the relevant feature description in other methods is sensitive to noise, their performance tends to deteriorate rapidly when increasing the noise level. In sharp contrast, it still has little negative impact on our method even for large-scale noises, which demonstrates the superiority of our method in robustness.

C. Limitation

Because our method is based on the assumption that a salient object should have distinctive diffusion pattern from non-salient background, but in practice, this assumption may not always be true. For images containing multiple salient objects (see Fig. 16a), both salient objects (region A and C in Fig. 16d) tend to have similar diffusion pattern. Therefore, salient objects are no longer regarded as the low-rank part

TABLE I
AVERAGE TIME CONSUMPTION OF SINGLE IMAGE IN ACHANTA DATASET [19]

Method	Ours	HS [1]	LK12 [4]	MC13 [2]	SF12 [3]	CB12 [5]	RC11 [6]	HC11 [6]	LC06 [33]	FT09 [19]	SR07 [34]
Time (s)	3.76	0.475	6.83	52.9	0.235	0.592	0.225	0.039	0.024	0.018	0.089
Code	Matlab	C++	Matlab	Matlab	C++	Matlab	C++	C++	C++	C++	Matlab

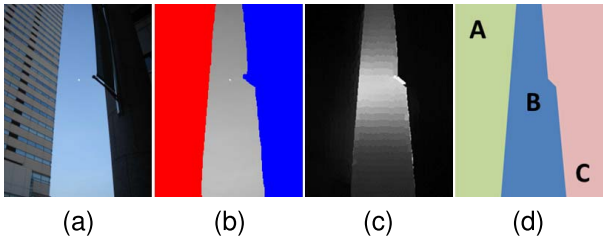


Fig. 16. Limitations of our method. (a) The source image is containing two salient objects, (b) The ground truth with each salient object marked with red and blue color respectively, (c) The saliency map computed by our method, (d) The visual explanation why our method fails in such a case.

by the low-rank decomposition process, and non-salient background (region B in Fig. 16d) is assigned with high saliency value (Fig. 16c). One way to overcome this limitation is to incorporate high-level semantic knowledge which can automatically assign higher saliency value to the meaningful buildings than to the background sky in this example, and this research issue deserves our future investigation.

Another limitation of our method is that, our method tends to be time-consuming in general. Table I documents the average time cost of each method. All of these methods are run on a computer with Quad Core i7-3770 3.4 GHz, 8GB RAM and NVIDIA GeForce GTX 660 Ti. In fact, the high accuracy is actually somehow less desirable in the interest of efficiency, and the computation of our novel descriptor is the major bottle neck in terms of time consumption (costing about 2.5 seconds). For single 500*500 image, let us consider our CUDA based parallel implementation as an example, our method costs about 3.7 seconds to complete the entire computation, which is still much less than the low-rank decomposition based saliency detection method [4].

VII. CONCLUSION

In this paper, we have presented a novel and versatile method to address a suite of research challenges in multi-scale, structure-sensitive saliency detection of natural images. The critical novel technical elements include: the bi-harmonic distance metric based intrinsic shape descriptor, multi-level low-rank decomposition based optimization model, and their elegant integration for the natural trade-off among local mutation, global uniqueness, and rarity scope towards a brand new saliency measurement, all of which contribute to physics-based vision, shape representation, image optimization, and pattern recognition. Our comprehensive experiments and extensive comparisons with other state-of-the-art methods have demonstrated our method's obvious advantages in terms of accuracy, reliability, robustness, and versatility.

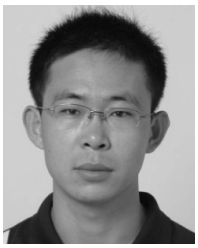
Our ongoing research efforts are concentrated on extending our key ideas to handle feature driven non-rigid registration, image co-segmentation, self-learning based image annotation, and image retrieval.

REFERENCES

[1] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1155–1162.
[2] Y. Xie, H. Lu, and M.-H. Yang, "Bayesian saliency via low and mid level cues," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1689–1698, May 2013.
[3] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 733–740.
[4] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 853–860.
[5] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li, "Automatic salient object segmentation based on context and shape prior," in *Proc. Brit. Mach. Vis. Conf.*, 2011, pp. 1–12.
[6] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 409–416.
[7] M. Ding and R. Tong, "Content-aware copying and pasting in images," *Vis. Comput.*, vol. 26, nos. 6–8, pp. 721–729, 2010.
[8] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 185–198, Jan. 2010.
[9] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu, "Sketch2Photo: Internet image montage," *ACM Trans. Graph.*, vol. 28, no. 5, 2009, Art. ID 124.
[10] C. Siagian and L. Itti, "Rapid biologically-inspired scene classification using features shared with visual attention," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 300–312, Feb. 2007.
[11] L. Marchesotti, C. Cifarelli, and G. Csurka, "A framework for visual saliency detection with applications to image thumbnailing," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 2232–2239.
[12] S. Bhattacharya, R. Sukthankar, and M. Shah, "A framework for photo-quality assessment and enhancement based on visual aesthetics," in *Proc. ACM Int. Conf. Multimedia*, 2010, pp. 271–280.
[13] V. Navalpakkam and L. Itti, "An integrated model of top-down and bottom-up attention for optimizing detection speed," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2006, pp. 2049–2056.
[14] C. M. Privitera and L. W. Stark, "Algorithms for defining visual regions-of-interest: Comparison with eye fixations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 9, pp. 970–982, Sep. 2000.
[15] D. Gao, S. Han, and N. Vasconcelos, "Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 6, pp. 989–1005, Jun. 2009.
[16] D. Purves, *Brains: How They Seem to Work*. Upper Saddle River, NJ, USA: FT Press Science, 2010.
[17] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proc. ACM Int. Conf. Multimedia*, 2003, pp. 374–381.
[18] 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.
[19] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1597–1604.
[20] T. Liu *et al.*, "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 353–367, Feb. 2011.
[21] J. Feng, Y. Wei, L. Tao, C. Zhang, and J. Sun, "Salient object detection by composition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1028–1035.

[22] O. Boiman and M. Irani, "Detecting irregularities in images and in video," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2005, pp. 462–469.
[23] X. Sun, H. Yao, R. Ji, P. Xu, X. Liu, and S. Liu, "Saliency detection based on short-term sparse representation," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 1101–1104.
[24] L. Duan, C. Wu, J. Miao, L. Qing, and Y. Fu, "Visual saliency detection by spatially weighted dissimilarity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 473–480.
[25] X. Hou and L. Zhang, "Dynamic visual attention: Searching for coding length increments," in *Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, Inc., 2008, pp. 681–688.
[26] R. Valenti, N. Sebe, and T. Gevers, "Image saliency by isocentric curvedness and color," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 2185–2192.
[27] X. Chen, H. Huo, T. Fang, and D. Li, "New approach to texture saliency based on intrinsic relationship among texture features," *Proc. SPIE*, vol. 6790, pp. 67902V-1–67902V-8, Nov. 2007.
[28] J. Yan, M. Zhu, H. Liu, and Y. Liu, "Visual saliency detection via sparsity pursuit," *IEEE Signal Process. Lett.*, vol. 17, no. 8, pp. 739–742, Aug. 2010.
[29] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 612–620.
[30] W. Einhauser and P. Konig, "Does luminance-contrast contribute to a saliency map for overt visual attention?" *Eur. J. Neurosci.*, vol. 17, no. 5, pp. 1089–1097, 2003.
[31] D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural Netw.*, vol. 19, no. 9, pp. 1395–1407, 2006.
[32] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2376–2383.
[33] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proc. ACM Int. Conf. Multimedia*, 2006, pp. 815–824.
[34] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[35] B. Han, H. Zhu, and Y. Ding, "Bottom-up saliency based on weighted sparse coding residual," in *Proc. ACM Int. Conf. Multimedia*, 2011, pp. 1117–1120.
[36] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-S. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3166–3173.
[37] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2976–2983.
[38] S. Li, Q. Zhao, S. Wang, T. Hou, A. Hao, and H. Qin, "A novel material-aware feature descriptor for volumetric image registration in diffusion tensor space," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 502–515.
[39] S. Li, H. Qin, and A. Hao, "Multi-scale local features based on anisotropic heat diffusion and global eigen-structure," *Sci. China Inf. Sci.*, vol. 56, no. 11, pp. 1–10, 2013.
[40] Y. Lipman, R. M. Rustamov, and T. A. Funkhouser, "Biharmonic distance," *ACM Trans. Graph.*, vol. 29, no. 3, 2008, Art. ID 27.
[41] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels," École Polytechn. Fédérale Lausanne, Lausanne, Switzerland, Tech. Rep. 149300, 2010.
[42] J. Sun, M. Ovsjanikov, and L. Guibas, "A concise and provably informative multi-scale signature based on heat diffusion," in *Proc. Symp. Geometry Process.*, 2009, pp. 1383–1392.
[43] F. Moreno-Noguer, "Deformation and illumination invariant feature point descriptor," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 1593–1600.
[44] M. Fazel, E. Candes, B. Recht, and P. Parrilo, "Compressed sensing and robust recovery of low rank matrices," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Oct. 2008, pp. 1043–1047.
[45] T. Zhou and D. Tao, "GoDec: Randomized low-rank & sparse matrix decomposition in noisy case," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 33–40.
[46] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Robust, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, Inc., 2009, pp. 2080–2088.
[47] S. Alpert, M. Galun, R. Basri, and A. Brandt, "Image segmentation by probabilistic bottom-up aggregation and cue integration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
[48] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
[49] A. Borji, D. N. Sihite, and L. Itti, "Salient object detection: A benchmark," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 414–429.

Chenglizhao Chen received the M.S. degree in computer science from the Beijing University of Chemical Technology, in 2012. He is currently pursuing the Ph.D. degree in technology of computer application from Beihang University, Beijing, China. His research interests include pattern recognition, computer vision, and machine learning.



Shuai Li received the Ph.D. degree in computer science from Beihang University, where he is currently an Assistant Professor with the State Key Laboratory of Virtual Reality Technology and Systems. His research interests include computer graphics, pattern recognition, computer vision, physics-based modeling and simulation, and medical image processing.

Hong Qin (SM'93) received the B.S. and M.S. degrees in computer science from Peking University, and the Ph.D. degree in computer science from the University of Toronto. He is currently a Professor of Computer Science with the Department of Computer Science, Stony Brook University. His research interests include geometric and solid modeling, graphics, physics-based modeling and simulation, computer-aided geometric design, visualization, and scientific computing.

Aimin Hao received the B.S., M.S., and Ph.D. degrees in computer science from Beihang University. He is currently a Professor with the Computer Science School, Beihang University, where he is also the Associate Director of the State Key Laboratory of Virtual Reality Technology and Systems. His research interests are on virtual reality, computer simulation, computer graphics, geometric modeling, image processing, and computer vision.