Salient Object Detection via Structured Matrix Decomposition

Houwen Peng, Bing Li, Haibin Ling, Weiming Hu, Weihua Xiong, and Stephen J. Maybank

Abstract—Low-rank recovery models have shown potential for salient object detection, where a matrix is decomposed into a low-rank matrix representing image background and a sparse matrix identifying salient objects. Two deficiencies, however, still exist. First, previous work typically assumes the elements in the sparse matrix are mutually independent, ignoring the spatial and pattern relations of image regions. Second, when the low-rank and sparse matrices are relatively coherent, e.g., when there are similarities between the salient objects and background or when the background is complicated, it is difficult for previous models to disentangle them. To address these problems, we propose a novel structured matrix decomposition model with two structural regularizations: (1) a tree-structured sparsity-inducing regularization that captures the image structure and enforces patches from the same object to have similar saliency values, and (2) a Laplacian regularization that enlarges the gaps between salient objects and the background in feature space. Furthermore, high-level priors are integrated to guide the matrix decomposition and boost the detection. We evaluate our model for salient object detection on five challenging datasets including single object, multiple objects and complex scene images, and show competitive results as compared with 24 state-of-the-art methods in terms of seven performance metrics.

Index Terms—Salient Object Detection, Matrix Decomposition, Low Rank, Structured Sparsity, Subspace Learning.

1 Introduction

Visual saliency has been a fundamental research problem in neuroscience, psychology and vision perception for a long time. It refers to the identification of a subset of vital visual information for further processing. As an important branch of visual saliency, salient object detection is the task of localizing and segmenting the most conspicuous foreground objects from a scene. It has received substantial attention over the last decade due to its wide range of applications in computer vision, such as object detection and recognition [1]–[4], content-based image retrieval [5], [6] and context-aware image resizing [7]–[10].

Many saliency models have been proposed to compute the saliency map of a given image and detect the salient objects. Depending on whether prior knowledge is used or not, current models fall into two categories: bottom-up and top-down. Bottom-up models [7], [11]–[17] are stimulus-driven and essentially based upon local and/or global center-surround difference, using low-level features, such as color, texture and location. The main limitations of these methods are that the detected salient regions may only contain parts of the target objects, or be easily mixed with background. On the other hand, top-down models [18]–[24] are task-driven and usually exploit high-level human perceptual knowledge, such as context, semantics and background priors, to guide the subsequent saliency computation. However, the high diversity of object types limits the generalization and scalability of these models.

A recent trend is to combine bottom-up cues with top-down priors to facilitate detection. A representative series of papers [25]–[28] are based on the *low-rank matrix recovery* (LR) theory [29]. For instance, Shen and Wu [26] propose a *unified LR model* (ULR) with feature transformation to combine traditional low-level features with high-level prior knowledge. Zou et al. [28] introduce the *segmentation priors* derived from image background and boundary cues to assist the *low-rank matrix recovery* (denoted as SLR). Lang et al. [27] present a *low-rank representation* (LRR) [30] based multi-task learning method, in which top-down priors are weighted and combined with multiple features to estimate saliency collaboratively. Generally, these LR-based saliency detection methods assume that an image can be represented as a combination of a highly redundant information part (e.g., visually consistent background regions) and a sparse salient part (e.g., distinctive foreground object regions). The redundant information part usually lies in a low dimensional feature subspace, which can be approximated by a low-rank feature matrix. In contrast, the salient part deviating from the low-rank subspace can be viewed as noise or errors, which are represented by a sparse sensory matrix. Therefore, given the feature matrix $\mathbf{F}$ of an input image, it can be decomposed as a low-rank matrix $\mathbf{L}$ corresponding to the non-salient background and a sparse matrix $\mathbf{S}$ corresponding to the salient foreground objects. Salient object detection can then be formulated as a low-rank matrix recovery problem [29]:

$$\min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \lambda\|\mathbf{S}\|_1 \qquad \text{s.t.} \qquad \mathbf{F} = \mathbf{L} + \mathbf{S}\,, \tag{1}$$

where the nuclear norm $\|\cdot\|_*$ (sum of the singular values of a matrix) is a convex relaxation of the matrix rank

- *H. Peng is with the Institution of Automation, Chinese Academy of Sciences, Beijing 100190, China, and the Department of Computer and Information Sciences, Temple University, Philadelphia, PA. E-mail: houwen.peng@nlpr.ia.ac.cn.*
- *B. Li, W Xiong and W. Hu are with the Institution of Automation, Chinese Academy of Sciences, Beijing 100190, China. E-mail: {bli, wmhu}@nlpr.ia.ac.cn, wallace.xiong@gmail.com.*
- *H. Ling is with the Department of Computer and Information Sciences, Temple University, Philadelphia, PA. E-mail: hbling@temple.edu.*
- *S. Maybank is with the Department of Computer Science and Information Systems, Birkbeck College, London WC1E 7HX, United Kingdom. E-mail: sjmaybank@dcs.bbk.ac.uk.*

Manuscript received XXX XX, 2015.

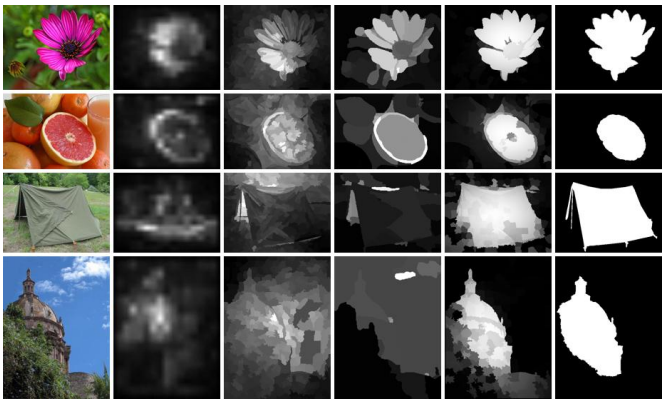

Fig. 1. Typical challenging examples for LR-based salient object detection algorithms. The resulting saliency maps of previous solutions (LRR [27], ULR [26] and SLR [28]) are scattered and incomplete, while our algorithm (SMD) overcomes these difficulties and performs close to the ground truth (GT).

function, $\|\cdot\|_1$ is the ℓ_1-norm which promotes sparsity, and the parameter $\lambda > 0$ controls the tradeoff between the two items.

Though previous LR-based salient object detection algorithms ([26]–[28]) have produced promising results, there still exist several problems:

- Previous studies do not take into account the intercorrelation between elements in $\mathbf{S}$, and thus ignore spatial relations, such as spatial contiguity and pattern consistency, between pixels and patches. Algorithms designed this way may suffer from two limitations: (1) the foreground pixels or patches in the generated saliency map tend to be scattered, as shown in Fig. 1 (LRR and ULR); and (2) the saliency values may be inconsistent within the same object, causing incompleteness of the detected object, as shown in Fig. 1 (LRR, ULR and SLR).
- According to the LR theory (a.k.a robust PCA) [29], the decomposition performance of an observation matrix degrades when there is high coherence between the underlying low-rank and sparse matrices. Therefore, when the background is cluttered or has similar appearance with the salient objects, it is difficult for previous LR-based methods to separate them, as shown in the last two rows of Fig. 1.

To address these issues, we propose a novel *structured matrix decomposition* (SMD) model that treats the (salient) foreground/background separation as a problem of low-rank and structured-sparse matrix decomposition. We enhance the traditional LR model in Eq. (1) with two important components. First, we introduce a *tree-structured sparsity-inducing norm* to constrain $\mathbf{S}$, so that the spatial connectivity and feature similarity of image patches are taken into account in matrix decomposition. This constraint is essentially a hierarchical group sparsity norm over a tree structure, in which an ℓ_∞-norm is employed to enforce within-object patches to share consistent saliency values. Second, we integrate a *Laplacian regularization* to reduce the coherence between the low-rank and structured-sparse matrices. The regularizer takes into account the geometrical structure of the image, encourages local similar patches to share similar representation, and eventually separates the foreground objects from the background as much as possible. These properties enable the proposed SMD model to detect salient objects in jumbled scenes, even when the salient objects have a similar appearance to the background. In addition, SMD enhances object completeness which is sometime hard to achieve by previous solutions.

The main contributions of this work are summarized as follows:

- We develop a novel structured matrix decomposition model, i.e., SMD, for salient object detection. Compared to the classical LR model used in [26]–[28], SMD not only captures the underlying structure of data, but also better handles the challenges arising from coherence of the low-rank and sparse matrices. To the best of our knowledge, this is the first work that explicitly pursues the hierarchical structure of data via structured sparsity in matrix decomposition. Based on the *alternating direction method* (ADM) [31], we derive an effective optimization algorithm to solve the proposed SMD model.
- We present an SMD-based salient object detection framework and evaluate the SMD method on five popular benchmarks involving various scenarios such as single object, multiple objects and complex scenes. Also, we compare our method with 24 state-of-the-art methods using six performance metrics, including the traditional measures, e.g., precision-recall curve and mean absolute error, and the recently proposed weighted F-measure [32]. In the experiments, our SMD-based algorithm achieves competitive results in comparison with other leading methods.

The remainder of this paper is organized as follows. Sec. 2 reviews existing saliency detection models, especially the LR-based methods. Sec. 3 describes the proposed SMD model and derives the ADM-based solution to the model. Sec. 4 presents the SMD-based salient object detection method and extends it to integrate high-level priors. Sec. 5 shows the experimental results, including a thorough comparison with recently proposed salient object detection algorithms and detailed analysis of the components in our algorithm. Finally, Sec. 6 concludes the paper.

2 Related Work

Recent years have witnessed significant advances in saliency detection that includes two major subfields: eye fixation prediction and salient object detection. Recent surveys on eye fixation prediction can be found in [33]–[35], and salient object detection is surveyed in [36], [37]. In this section, we mainly discuss the algorithms belonging to the second subfield, to which our work belongs. But before that, we briefly review some classical early studies that have paved the way to both subfields.

The foundation of most saliency detection algorithms can be traced back to the theories of center-surround difference [38] and multiple feature integration [39]. The most influential model based on the theories is proposed by Itti *et al.* [11], who derive saliency from the difference of Gaussians on multiple feature maps. Another early work is by Harel et al. [40] who define a graph on image and adopt random walks to compute saliency. Some learning-based methods [41], [42] are also proposed to predict saliency by combining multiple feature maps. Latter, researchers refine

the theories by taking account of local [43], [44], regional [45], and global [46] contrast cues, or by searching for saliency cues in the frequency domain [14], [47].

One of the earliest works on *salient object detection* is [48], which formulates saliency detection as a binary segmentation problem. Recent studies can be broadly categorized as either bottom-up or top-down. Bottom-up models are bio-inspired and only use low-level image features. The frequency tuning method [49] detects saliency by computing color deviation from the mean image color at the pixel level. Later, an improved solution [7] is proposed to highlight salient objects with respect to their contexts in terms of low-level feature distinction and global spatial relations. The global contrast method [12] identifies salient regions by estimating dissimilarities between *Lab* color histograms over all image regions. Saliency filters [50] improve the global contrast method [12] by combining color uniqueness and spatial distribution of image regions. Some other bottom-up techniques such as multi-scale modeling [51] and high-dimensional color transformation [17] have been explored for salient object detection. The effectiveness of other complementary cues such as texture [20], depth [52], [53] or surroundedness [54] have also been considered recently.

By contrast, top-down models usually estimate saliency via task-specific learning algorithms or high-level priors. The method in [48] identifies salient objects using a *conditional random field* (CRF) on a multi-scale contrast histogram and spatial distribution features. The latent variable model in [55] estimates saliency by jointly learning a CRF and a specific dictionary. Instead of direct training on image features, saliency aggregation [56] trains a CRF on saliency maps produced by other methods. The random forest model [57] predicts image saliency by training a regressor on discriminative regional features. Most recently, multiple kernel learning [58] and convolutional neural network [59] techniques have been introduced to learn more robust discrimination between salient and non-salient regions.

High-level priors have also been used in top-down models and proved to be effective. For example, a Gaussian fall-off function is frequently recruited to emphasize the center regions (i.e., *center prior*), either directly combined with other cues [19], [21], [60], or used as a spatial feature in learning [48], [57]. The prior belief that image boundary regions are more likely to belong to the background (i.e., *background prior*) is also commonly integrated for saliency computation. A representative work is the geodesic saliency [24], which defines boundary regions as terminal nodes when estimating saliency on an image graph. Alternatively, in [61], [62], boundary regions are used as pseudo-background queries and dictionary templates to facilitate detection. Later, a more robust boundary connectivity prior is introduced in [63]. Besides, the *objectness prior*, which estimates the likelihood of a region being a complete object [2], has been employed in some other saliency models [18], [64], [65].

Our study is related to recent methods that consider the *sparsity prior* in salient object detection. The method in [25] adopts an over-complete dictionary to encode image patches and then feed the coding vector to the LR model to recover salient objects. Later, a supervised method [26] is proposed to leverage feature transformation with the high-level center, color and semantic priors to meet the low-rank and sparse properties. To better fit the LR model, the segmentation prior derived from the connectivity between regions and image borders is exploited to guide matrix recovery [28]. As an extension of the LR model, *low-rank representation* (LRR) [27] introduces a self-representation scheme that reconstructs background regions from the image features themselves rather than by a dictionary. Multi-feature collaborative enhancement and top-down priors obtained from [66] are incorporated into the multi-task extension.

Difference to related LR-based methods. As an LR-based method, our SMD differs from the previous ones [25]–[28] in several aspects. (1) SMD pursues the low matrix rank in a purely unsupervised way, while [25] and [26] respectively resort to supervised sparse representation and feature transformation. The learnt representation or transformation in [25] and [26] is biased toward the training datasets, and therefore suffers from limited adaptability. (2) Our method explicitly encodes information about image structure, i.e., spatial relations and feature similarities of image patches, which are ignored in [25]–[28]. (3) Our method integrates high-level priors into the structured image representation (index tree), while other methods [26], [28] combine such priors by re-weighting the feature.

Discussion with Manifold Ranking (MR) methods. The use of the Laplacian regularization in our method is inspired by, but different from that in the MR algorithm [61]. (1) Our method uses the Laplacian regularization to smooth the feature representation, and to enlarge the difference between foreground objects and background in feature space. By contrast, MR exploits the regularization to enforce continuous saliency values over neighboring patches. (2) MR is built upon the semi-supervised ranking model [67], and defines saliency of an patch as its relevance to the given querying seeds. By contrast, our method uses the low-rank matrix decomposition framework and is purely unsupervised.

Difference with preliminary work. Some preliminary ideas in this paper appeared in the conference version [68]. Compared with [68], the proposed SMD model in this paper is more general, and subsumes the version in [68] as a special case. The new SMD model not only inherits the major advantages of the preliminary model, i.e., it produces a decomposition of an observation matrix into structured parts with respect to image structure, but it is also armed with the new capability to enlarge the separation between salient objects and background in the feature space. The experimental results (Sec. 5) show clearly that the new model is more robust and the resulting saliency maps (Fig. 8) are more visually favorable.

3 Structured Matrix Decomposition Model

3.1 Proposed Model

3.1.1 Basic formulation

Given an input image I, it is first partitioned into N non-overlapping patches $\mathcal{P} = \{P_1, P_2, \cdots, P_N\}$, e.g., superpixels. For each patch P_i, a D-dimension feature vector is extracted and denoted as $\mathbf{f}_i \in \mathbb{R}^D$. The ensemble of feature vectors forms a matrix representation of I, denoted as $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_N] \in \mathbb{R}^{D \times N}$. The problem of salient object detection is to design an effective model to decompose the feature matrix $\mathbf{F}$ into a redundant information part $\mathbf{L}$ (i.e.,

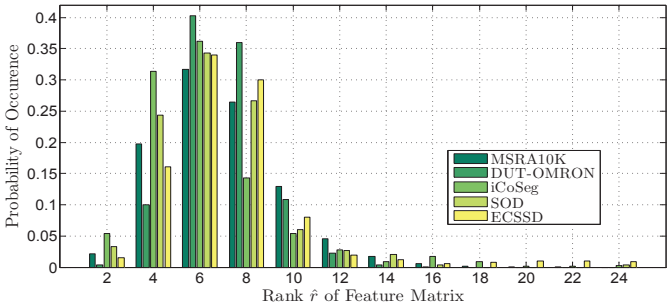

Fig. 2. Rank statistics of feature matrices extracted from image background over five datasets: MSRA10K [48], [69], DUT-OMRON [61], SOD [70], iCoSeg [71] and ECSSD [21].

non-salient background) and a structured distinctive part $\mathbf{S}$ (i.e., salient foreground).

To address the issues discussed in Sec. 1, we propose a novel *structured matrix decomposition* (SMD) model as follows:

$$\min_{\mathbf{L},\mathbf{S}} \Psi(\mathbf{L}) + \alpha\Omega(\mathbf{S}) + \beta\Theta(\mathbf{L},\mathbf{S}) \quad \text{s.t.} \quad \mathbf{F} = \mathbf{L} + \mathbf{S}\ , \qquad (2)$$

where $\Psi(\cdot)$ is a low-rank constraint to allow identification of the intrinsic feature subspace of the redundant background patches, $\Omega(\cdot)$ is a structured sparsity regularization to capture the spatial and feature relations of patches in $\mathbf{S}$, $\Theta(\cdot,\cdot)$ is an interactive regularization term to enlarge the distance between the subspaces drawn from $\mathbf{L}$ and $\mathbf{S}$, and α, β are positive tradeoff parameters.

3.1.2 Low-rank regularization for image background

Having observed that image patches from the background are often similar and approximately lie in a low-dimensional subspace, we apply low-rank regularization on the background feature matrix $\mathbf{L}$ to pursue its intrinsic structure. Since directly minimizing a matrix's rank with affine constraints is an NP-hard problem [30], we instead adopt the nuclear norm as a convex relaxation, i.e.,

$$\Psi(\mathbf{L}) = rank(\mathbf{L}) = \|\mathbf{L}\|_* + \varepsilon\ , \qquad (3)$$

where ε denotes the relaxation error.

To verify the rationality of the low-rank constraint, we evaluate the rank of feature matrices extracted from image background on five salient object datasets (Fig. 2). Specifically, we first divide each image into a regular grid of patches of size 10×10 pixels, excluding those "foreground" patches, which have over 10% pixels from the annotated salient objects. Then, each patch is represented by a feature vector encoding color, edge and texture information (as described in Sec. 4.1). Features from the same image are juxtaposed into a matrix to represent the image background. Finally, we estimate the rank of the feature matrix, denoted by $\widehat{r}$, according to [72], [73]:

$$\widehat{r} = \arg\min_{r} \big(\mathrm{RMSRE}(r-1) - \mathrm{RMSRE}(r)\big) \le \epsilon\ , \qquad (4)$$

where $\mathrm{RMSRE}(r)$ is the *root mean square reconstruction error* between the original matrix and its rank-r approximation estimated by the *singular value decomposition* (SVD), and ϵ is a threshold with value 0.01. Fig. 2 shows the statistics of such estimated ranks of background feature matrices of the images in the five datasets. It shows that about 90% of these matrices can be approximated by a matrix with rank no greater than 10. This confirms our intuition that the image background usually lies in a low-dimensional subspace. Therefore, it encourages us to exploit a low-rank regularization to eliminate redundant information and pursue the intrinsic low-dimensional structure.

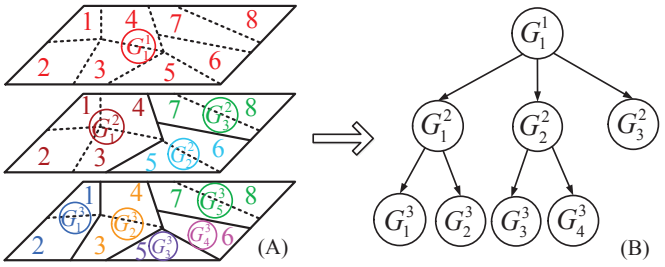

Fig. 3. The construction of an index tree from an image. (A): The hierarchical segmentation of an input image. The digits are the indexes of patches. (B): An index tree constructed over the indices of image patches $\{1,2,\ldots,8\}$. Depth 1 (Root): $G_1^1 = \{1,2,3,4,5,6,7,8\}$. Depth 2: $G_1^2 = \{1,2,3,4\}$, $G_2^2 = \{5,6\}$, $G_3^2 = \{7,8\}$. Depth 3: $G_1^3 = \{1,2\}$, $G_2^3 = \{3,4\}$, $G_3^3 = \{5\}$, $G_4^3 = \{6\}$.

3.1.3 Structured-sparsity regularization for salient objects

In Eq. (1), the ℓ_1-norm regularization treats the columns in $\mathbf{S}$ independently and thus ignores spatial structure information, which can otherwise be used to improve salient object detection (see Fig. 1). In the following, we introduce a novel tree-structured sparsity-inducing norm to model the spatial contiguity and feature similarity among image patches so as to produce more precise and structurally consistent results.

Before detailing the structured regularization, we first give the definition of an *index tree* [74]. An index tree is a hierarchial structure, such that each node contains a set of indices (e.g., corresponding to the superpixels in our task) and the set is the union of the indices of its children. More specifically, for an index tree T with depth d over indices $\{1,2,\ldots,N\}$, let G_j^i be the j-th node at the i-th level. In particular, for the root node, we have $G_1^1 = \{1,2,\ldots,N\}$. The nodes also satisfy two conditions: (1) there is no overlap between the indices of nodes from the same level, i.e., $G_j^i \cap G_k^i = \emptyset$, $\forall 2 \le i \le d$ and $1 \le j < k \le n_i$. Here, n_i denotes the total number of nodes at the i-th level. (2) Let G_{j0}^{i-1} be the parent node of a non-root node G_j^i, then $G_j^i \subseteq G_{j0}^{i-1}$ and $\bigcup_j G_j^i = G_{j0}^{i-1}$. Fig. 3 shows an example tree with $N = 8$ indexes, drawn from hierarchical segmentation of an image.

We use an index tree T to encode the spatial relation of image patches $\mathcal{P}$. Details of index tree construction are postponed to Sec. 4.1. We encode the structurally meaningful tree constraint into a sparsity norm to regularize the matrix decomposition. In this way, we get a general tree-structured sparsity regularization as

$$\Omega(\mathbf{S}) = \sum_{i=1}^{d}\sum_{j=1}^{n_i} v_j^i \|\mathbf{S}_{G_j^i}\|_p\ , \qquad (5)$$

where $v_j^i \ge 0$ is the weight for the node G_j^i, $\mathbf{S}_{G_j^i} \in \mathbb{R}^{D\times|G_j^i|}$ ($|\cdot|$ denotes set cardinality) is the sub-matrix of $\mathbf{S}$ corresponding to the node G_j^i, and $\|\cdot\|_p$ is the ℓ_p-norm[1], $1 \le p \le \infty$. In essence, $\Omega(\cdot)$ is a weighted group sparsity norm defined over a tree structure. It induces the patches within

1. For a matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{m\times n}$, $\|\mathbf{A}\|_p = (\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{i,j}|^p)^{1/p}$.

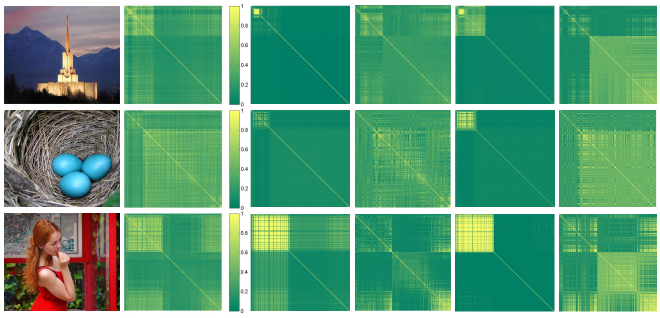

Fig. 4. The pairwise similarity matrices of feature vectors before and after imposing Laplacian regularization. The upper-left block in the matrices represents the similarities of foreground patches, while the bottom-right block indicates the similarities of background patches. Matrices with subscript '$_*$' are the results after imposing Laplacian regularization.

the same group to share a similar representation, and also represents the subordinate or coordinate relations between groups. To enforce the patches from the same group to have identical saliency values, we impose the ℓ_∞-norm on $\mathbf{S}_{G_j^i}$, i.e., $p = \infty$. It uses the maximum saliency value of patches within the group G_j^i to decide whether the group is salient or not [75].

3.1.4 Laplacian regularization

When decomposing the feature matrix $\mathbf{F}$ into a low-rank part $\mathbf{L}$ plus a structured-sparse part $\mathbf{S}$, we also hope to enlarge the distance between the subspaces induced by $\mathbf{L}$ and $\mathbf{S}$, so as to make it easier to separate the salient object from the background. To this end, we introduce a Laplacian regularization based on the local invariance assumption [76]: if two adjacent image patches are similar with respect to their features, their representations should be close to each other in the subspace, and vise versa. Thus motivated, we define the regularization as

$$\Theta(\mathbf{L},\mathbf{S}) = \frac{1}{2}\sum_{i,j=1}^{N} \|\mathbf{s}_i - \mathbf{s}_j\|_2^2 \, w_{i,j} = \mathrm{Tr}(\mathbf{S}\mathbf{M}_\mathbf{F}\mathbf{S}^T) \,, \quad (6)$$

where $\mathbf{s}_i$ denotes the i-th column of $\mathbf{S}$, $w_{i,j}$ is the (i,j)-th entry of an affinity matrix $\mathbf{W} = (w_{i,j}) \in \mathbb{R}^{N\times N}$ and represents the feature similarity of patches (P_i, P_j), $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix, and $\mathbf{M}_\mathbf{F} \in \mathbb{R}^{N\times N}$ is a Laplacian matrix. Specifically, the affinity matrix $\mathbf{W}$ is defined as

$$w_{i,j} = \begin{cases} \exp\big(-\frac{\|\mathbf{f}_i - \mathbf{f}_j\|^2}{2\sigma^2}\big), & \text{if } (P_i,P_j) \in \mathbb{V} \,, \\ 0, & \text{otherwise} \,, \end{cases} \quad (7)$$

where $\mathbb{V}$ denotes the set of adjacent patch pairs which are either neighbors (first-order) or "neighbors of neighbors" (second-order reachable) on the image. The (i,j)-th entry of the Laplacian matrix $\mathbf{M}_\mathbf{F}$ is

$$(\mathbf{M}_\mathbf{F})_{i,j} = \begin{cases} -w_{i,j}, & \text{if } i \neq j \,, \\ \sum_{j\neq i} w_{i,j}, & \text{otherwise} \,. \end{cases} \quad (8)$$

It is interesting to find that the Laplacian regularization in Eq. (6) is explicitly related with $\mathbf{F}$ and $\mathbf{S}$, and can be transferred to be related with $\mathbf{L}$ and $\mathbf{S}$ according to $\Theta(\mathbf{F},\mathbf{S}) = \Theta(\mathbf{L}+\mathbf{S},\mathbf{S}) = \Theta(\mathbf{L},\mathbf{S})$. Essentially, the Laplacian regularization increases the distance between feature subspaces by smoothing the vectors in $\mathbf{S}$ according to the local neighborhood derived from the feature matrix $\mathbf{F}$. It encourages patches within the same semantic region to share similar or identical representation, and patches from heterogeneous regions to have different representation. Fig. 4 shows the pairwise similarity of the elements in $\mathbf{L}$ and $\mathbf{S}$ before and after imposing the Laplacian regularization. It shows that a more distinct block affinity matrix is produced by using the regularization.

Algorithm 1 ADM-SMD.

Input: Feature matrix $\mathbf{F}$, parameters α, β, index tree $T = \{G_j^i\}$ and tree node weight v_j^i (default as 1).
Output: $\mathbf{L}$ and $\mathbf{S}$.

Initialize $\mathbf{L}^0$=$\mathbf{0}$, $\mathbf{S}^0$=$\mathbf{0}$, $\mathbf{H}^0$=$\mathbf{0}$, $\mathbf{Y}_1^0$=$\mathbf{0}$, $\mathbf{Y}_2^0$=$\mathbf{0}$, μ^0=0.1, $\mu_{\max} = 10^{10}$, ρ= 1.1, and $k = 0$.
While not converged **do**
  $\mathbf{L}^{k+1} = \arg\min_{\mathbf{L}} \mathcal{L}(\mathbf{L}, \mathbf{S}^k, \mathbf{H}^k, \mathbf{Y}_1^k, \mathbf{Y}_2^k, \mu^k)$
  $\mathbf{H}^{k+1} = \arg\min_{\mathbf{H}} \mathcal{L}(\mathbf{L}^{k+1}, \mathbf{S}^k, \mathbf{H}, \mathbf{Y}_1^k, \mathbf{Y}_2^k, \mu^k)$
  $\mathbf{S}^{k+1} = \arg\min_{\mathbf{S}} \mathcal{L}(\mathbf{L}^{k+1}, \mathbf{S}, \mathbf{H}^{k+1}, \mathbf{Y}_1^k, \mathbf{Y}_2^k, \mu^k)$
  $\mathbf{Y}_1^{k+1} = \mathbf{Y}_1^k + \mu^k(\mathbf{F} - \mathbf{L}^{k+1} - \mathbf{S}^{k+1})$
  $\mathbf{Y}_2^{k+1} = \mathbf{Y}_2^k + \mu^k(\mathbf{S}^{k+1} - \mathbf{H}^{k+1})$
  $\mu^{k+1} = \min(\rho\mu^k, \mu_{\max})$
  $k = k + 1$
End While
Return $\mathbf{L}^k$ and $\mathbf{S}^k$.

3.2 Optimization

Considering the balance between efficiency and accuracy in practice, we resort to the *alternating direction method* (ADM) [31] to solve the convex problem defined in Eq. (2). We first introduce an auxiliary variable $\mathbf{H}$ to make the objective function separable, i.e., Eq. (2) becomes

$$\begin{aligned} \min_{\mathbf{L},\mathbf{S}} \quad & \|\mathbf{L}\|_* + \alpha\sum_{i=1}^{d}\sum_{j=1}^{n_i} v_j^i \|\mathbf{S}_{G_j^i}\|_p + \beta\mathrm{Tr}(\mathbf{H}\mathbf{M}_\mathbf{F}\mathbf{H}^T) \\ \text{s.t.} \quad & \mathbf{F} = \mathbf{L} + \mathbf{S}, \quad \mathbf{S} = \mathbf{H} \,. \end{aligned} \quad (9)$$

Then, the problem (9) can be solved with ADM, which minimizes the following augmented Lagrangian function $\mathcal{L}$:

$$\begin{aligned} \mathcal{L}(\mathbf{L},\mathbf{S},\mathbf{H},\mathbf{Y}_1,\mathbf{Y}_2,\mu) &= \|\mathbf{L}\|_* \\ &+ \alpha\sum_{i=1}^{d}\sum_{j=1}^{n_i} v_j^i \|\mathbf{S}_{G_j^i}\|_p + \beta\mathrm{Tr}(\mathbf{H}\mathbf{M}_\mathbf{F}\mathbf{H}^T) \\ &+ \mathrm{Tr}(\mathbf{Y}_1^T(\mathbf{F} - \mathbf{L} - \mathbf{S})) + \mathrm{Tr}(\mathbf{Y}_2^T(\mathbf{S} - \mathbf{H})) \\ &+ \frac{\mu}{2}\big(\|\mathbf{F} - \mathbf{L} - \mathbf{S}\|_F^2 + \|\mathbf{S} - \mathbf{H}\|_F^2\big) \,, \end{aligned} \quad (10)$$

where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are the Lagrange multipliers, and $\mu > 0$ controls the penalty for violating the linear constraints. To solve Eq. (10), we search for the optimal $\mathbf{L}$, $\mathbf{S}$ and $\mathbf{H}$ iteratively, and in each iteration the three components are updated alternately. We outline the optimization procedure in Algorithm 1 and call it ADM-SMD. In the following, we provide the details for each iteration.

Updating L: When $\mathbf{S}$ and $\mathbf{H}$ are fixed, the update $\mathbf{L}^{k+1}$ at the $(k+1)$-th iteration is obtained by solving the following

problem:

$$\begin{aligned}\mathbf{L}^{k+1} &= \arg\min_{\mathbf{L}} \mathcal{L}(\mathbf{L}, \mathbf{S}^k, \mathbf{H}^k, \mathbf{Y}_1^k, \mathbf{Y}_2^k, \mu^k) \\ &= \arg\min_{\mathbf{L}} \|\mathbf{L}\|_* + \mathrm{Tr}((\mathbf{Y}_1^k)^T(\mathbf{F}-\mathbf{L}-\mathbf{S}^k)) \\ &\qquad + \frac{\mu^k}{2}\|\mathbf{F}-\mathbf{L}-\mathbf{S}^k\|_F^2 \\ &= \arg\min_{\mathbf{L}} \tau\|\mathbf{L}\|_* + \frac{1}{2}\|\mathbf{L}-\mathbf{X}_\mathbf{L}\|_F^2 ,\end{aligned} \tag{11}$$

where $\tau = 1/\mu^k$ and $\mathbf{X}_\mathbf{L} = \mathbf{F} - \mathbf{S}^k + \mathbf{Y}_1^k/\mu^k$. The solution to Eq. (11) can be derived as

$$\mathbf{L}^{k+1}{=}\mathbf{U}T_\tau[\mathbf{\Sigma}]\mathbf{V}^T, \text{where } (\mathbf{U},\mathbf{\Sigma},\mathbf{V}^T) = \mathcal{SVD}(\mathbf{X}_\mathbf{L}) . \tag{12}$$

Note that $\mathbf{\Sigma}$ is the singular value matrix of $\mathbf{X}_\mathbf{L}$. The operator $T_\tau[\cdot]$ is the *singular value thresholding* (SVT) [77] defined by element-wise τ thresholding of $\mathbf{\Sigma}$. Specifically, let σ_i be the i-th diagonal element of $\mathbf{\Sigma}$, then $T_\tau[\mathbf{\Sigma}]$ is a diagonal matrix defined as

$$T_\tau[\mathbf{\Sigma}] = \mathrm{diag}(\{(\sigma_i - \tau)_+\}), \tag{13}$$

where a_+ is the positive part of a, namely, $a_+ = \max(0, a)$.

Updating H: When $\mathbf{L}$ and $\mathbf{S}$ are fixed, to update $\mathbf{H}^{k+1}$, we derive from Eq. (10) the following problem:

$$\begin{aligned}\mathbf{H}^{k+1} &= \arg\min_{\mathbf{H}} \mathcal{L}(\mathbf{L}^{k+1}, \mathbf{S}^k, \mathbf{H}, \mathbf{Y}_1^k, \mathbf{Y}_2^k, \mu^k) \\ &= \arg\min_{\mathbf{H}} \beta\mathrm{Tr}(\mathbf{H}\mathbf{M}_\mathbf{F}\mathbf{H}^T){+}\mathrm{Tr}((\mathbf{Y}_2^k)^T(\mathbf{S}^k{-}\mathbf{H})) \\ &\qquad + \frac{\mu^k}{2}\|\mathbf{S}^k{-}\mathbf{H}\|_F^2 .\end{aligned} \tag{14}$$

Taking derivative of the objective function in Eq. (14) (the detailed derivation is presented in Appendix A), we have

$$\mathbf{H}^{k+1} = (\mu^k\mathbf{S}^k + \mathbf{Y}_2^k)(2\beta\mathbf{M}_\mathbf{F} + \mu^k\mathbf{I})^{-1} . \tag{15}$$

Updating S: To update $\mathbf{S}^{k+1}$ with fixed $\mathbf{L}$ and $\mathbf{H}$, we get the following tree-structured sparsity optimization problem:

$$\begin{aligned}\mathbf{S}^{k+1} &= \arg\min_{\mathbf{S}} \mathcal{L}(\mathbf{L}^{k+1}, \mathbf{S}, \mathbf{H}^{k+1}, \mathbf{Y}_1^k, \mathbf{Y}_2^k, \mu^k) \\ &= \arg\min_{\mathbf{S}} \alpha\sum_{i=1}^{d}\sum_{j=1}^{n_i} v_j^i\|\mathbf{S}_{G_j^i}\|_p \\ &\quad {+}\mathrm{Tr}((\mathbf{Y}_1^k)^T(\mathbf{F}{-}\mathbf{L}^{k+1}{-}\mathbf{S})){+}\mathrm{Tr}((\mathbf{Y}_2^k)^T(\mathbf{S}{-}\mathbf{H}^{k+1})) \\ &\quad {+}\frac{\mu^k}{2}(\|\mathbf{F}{-}\mathbf{L}^{k+1}{-}\mathbf{S}\|_F^2 + \|\mathbf{S}{-}\mathbf{H}^{k+1}\|_F^2) \\ &= \arg\min_{\mathbf{S}} \lambda\sum_{i=1}^{d}\sum_{j=1}^{n_i} v_j^i\|\mathbf{S}_{G_j^i}\|_p + \frac{1}{2}\|\mathbf{S}-\mathbf{X}_\mathbf{S}\|_F^2 ,\end{aligned} \tag{16}$$

where $\lambda = \alpha/(2\mu^k)$ and $\mathbf{X}_\mathbf{S} = (\mathbf{F}{-}\mathbf{L}^{k+1}{+}\mathbf{H}^{k+1} + (\mathbf{Y}_1^k - \mathbf{Y}_2^k)/\mu^k)/2$. The above problem can be solved by the hierarchical proximal operator [78], which computes a particular sequence of residuals obtained by projecting a matrix onto the unit ball of dual ℓ_p-norm. The detailed procedure when using ℓ_∞-norm is presented in Algorithm 2.

4 SMD-BASED SALIENT OBJECT DETECTION

In this section, we describe our salient object detection algorithm that uses the proposed SMD model. Our algorithm includes two major parts: the first one focuses on low-level features, while the second one incorporates high-level prior knowledge. Fig. 5 shows the framework of SMD-based salient object detection.

Algorithm 2 Solving the tree-structured sparsity.

Input: The index tree T with nodes G_j^i $(i = 1, 2, ..., d; j = 1, 2, ..., n_i)$, weight $v_j^i \geq 0$ (default as 1), the matrix $\mathbf{X}_\mathbf{S}$, parameters α, and set $\lambda = \alpha/(2\mu^k)$.

1: Set $\mathbf{S} = \mathbf{X}_\mathbf{S}$
2: **For** $i = d$ to 1 **do**
3: **For** $j = 1$ to n_i **do**
4: $$\mathbf{S}_{G_j^i}^{k+1} = \begin{cases} \frac{\|\mathbf{s}_{G_j^i}\|_1 - \lambda v_j^i}{\|\mathbf{s}_{G_j^i}\|_1}\mathbf{S}_{G_j^i}, & \text{if } \|\mathbf{S}_{G_j^i}\|_1 > \lambda v_j^i \\ \mathbf{0}, & \text{otherwise} \end{cases}$$
5: **End For**
6: **End For**
7: **Return** $\mathbf{S}^{k+1}$

4.1 Low-level Salient Object Detection

Our framework for low-level salient object detection consists of four steps: image abstraction, index tree construction, matrix decomposition and saliency assignment.

Step 1: Image Abstraction. In this step, an input image is partitioned into compact and perceptually homogeneous elements. Following [26], we first extract the low-level features, including RGB color, steerable pyramids [79] and Gabor filter [80], to construct a 53-dimension feature representation. Then, we perform the *simple linear iterative clustering* (SLIC) algorithm [81] to over-segment the image into N atom patches (superpixels) $\mathcal{P} = \{P_1, P_2, \cdots, P_N\}$. Each patch P_i is represented by a feature vector $\mathbf{f}_i$, and all these feature vectors form the feature matrix as $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_N] \in \mathbb{R}^{D\times N}$ (here $D = 53$).

Step 2: Tree Construction. On top of $\mathcal{P}$, an index tree T is constructed to encode structure information via hierarchical segmentation. To this end, we first compute the affinity of every adjacent patch pair using Eq. (7). Then, we apply a graph-based image segmentation algorithm [82] to merge spatially neighboring patches according to their affinity. The algorithm produces a sequence of granularity-increasing segmentations. In each granularity layer, the segments correspond to the nodes at the corresponding layer in the index tree. Specifically, the granularity is controlled by a affinity threshold $\mathcal{T}$. Finally, we obtain a hierarchical fine-to-coarse segmentation of the input image. Fig. 6 shows a visualized example of hierarchical segmentation, corresponding to a five-layer index tree structure.

Step 3: Matrix Decomposition. When both the feature matrix $\mathbf{F}$ and the index tree T are ready, we apply the proposed SMD model, formulated as Eq. (2) with ℓ_∞-norm, to decompose $\mathbf{F}$ into a low-rank component $\mathbf{L}$ and a structured-sparse component $\mathbf{S}$. As shown in Step 3 of Fig. 5, after jointly imposing the structured-sparsity and Laplacian regularization, the input feature matrix $\mathbf{F}$ is decomposed into structured components $\mathbf{L}$ and $\mathbf{S}$.

Step 4: Saliency Assignment. After decomposing $\mathbf{F}$, we transfer the results from the feature domain to the spatial domain for saliency estimation. Based on the structured matrix $\mathbf{S}$, we define a straightforward saliency estimation function $\mathrm{Sal}(\cdot)$ of each patch in $\mathcal{P}$:

$$\mathrm{Sal}(P_i) = \|\mathbf{s}_i\|_1, \tag{17}$$

where $\mathbf{s}_i$ represents the i-th column of $\mathbf{S}$. A large $\mathrm{Sal}(P_i)$ means a high possibility that P_i belongs to a salient object.

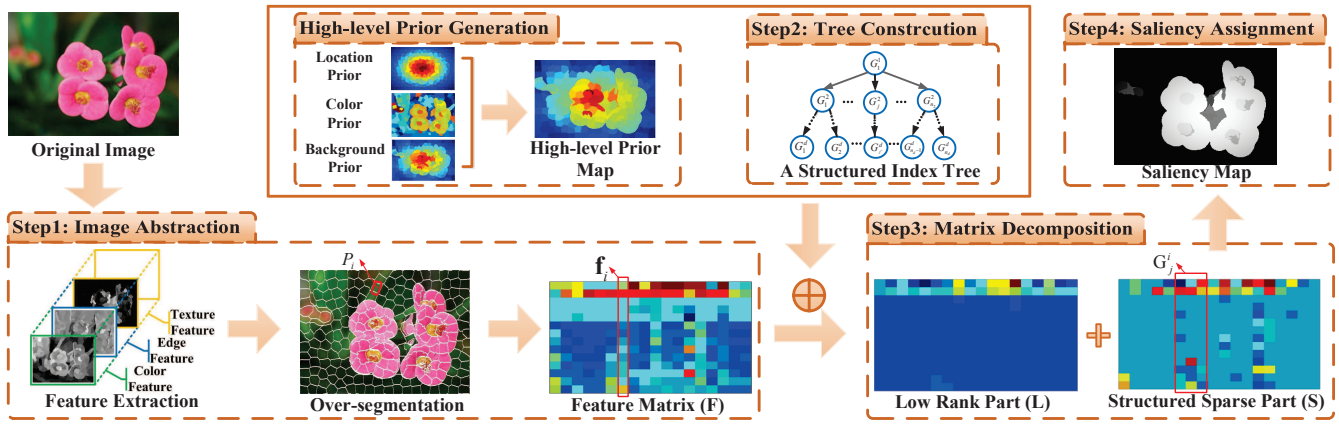

Fig. 5. Framework of the SMD model for salient object detection.

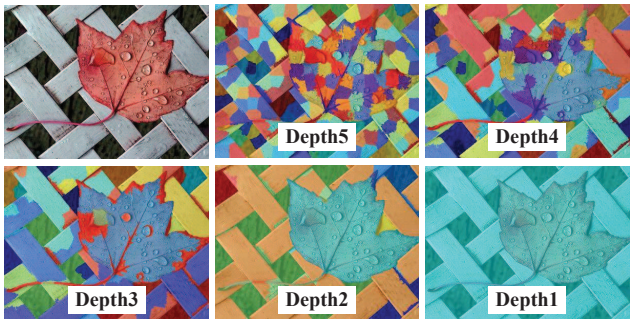

Fig. 6. Illustration of index tree construction based on the graph-based clustering. Each image indicates one layer in the index tree, while each patch represents one node.

After merging all patches together and performing context-based propagation (section 3.2 in [62]), we get the final saliency map of the input image.

4.2 Integrate High-level Priors

We further extend the proposed SMD-based saliency detection to integrate high-level priors. Inspired by the work of Shen and Wu [26], we fuse three types of priors, i.e. location, color and background priors, to generate a high-level prior map. Specifically, the location prior is generated by a Gaussian distribution based on the distances of the pixels from the image center. The color prior used here is the same as [26], which measures human eye sensitivity to red and yellow color. The background prior calculates the probabilities of image regions connected to image boundaries [63]. These three priors are finally multiplied together to produce the high-level prior map (see Fig. 5).

For each patch $P_i \in \mathcal{P}$, its high-level prior, $\pi_i \in [0,1]$, indicates the likelihood that P_i belongs to a salient object based on high-level information. This prior is encoded into the SMD by weighting each component in the tree-structured sparsity-inducing norm differently. In particular, we define v_j^i as

$$v_j^i = 1 - \max\left(\{\pi_k : k \in G_j^i\}\right). \quad (18)$$

Eq. (18) essentially boosts the saliency value of nodes with high prior values by associating them with small penalties v_j^i. This way, the high-level prior knowledge is seamlessly encoded into the SMD model to guide the matrix decomposition and enhance the saliency detection. It is worth noting that if we fix $v_j^i = 1$ for each node G_j^i, the proposed model is degraded to the pure low-level saliency detection model.

TABLE 1
Summary of the benchmark datasets.

Name	Size	Characteristics
MSRA10K [69]	10,000 (imgs)	single object, collected from MSRA [48], simple background, high contrast
DUT-OMRON [61]	5,168	single object, relatively complex background, more challenging
iCoSeg [71]	643	multiple objects, various number of objects with different sizes
SOD [70]	300	multiple objects, various size and location of objects, complex background
ECSSD [50]	1,000	structurally complex natural images, various object categories

5 Experiment

To fully evaluate our algorithm, we conduct a series of experiments using five benchmark datasets involving various scenarios and include 24 recent solutions for comparison.

5.1 Experimental Setup

5.1.1 Datasets

We use five popular benchmark datasets to cover different scenarios. In particular, we use MSRA10K [69] and DUT-OMRON [61] for images with a single salient object, i-CoSeg [71] and SOD [70] for cases with multiple salient objects, and ECSSD [21] for images with complex scenes. The size and detailed characteristic of these benchmark datasets are presented in Tab. 1.

5.1.2 Salient object detection algorithms

The proposed salient object detection algorithm is compared with 24 state-of-the-art solutions, including three LR-based methods (ULR [26], LRR [27] and SLR [28]), four methods ranked the highest according to the survey in [36] (SVO [18], CA [7], CB [19] and RC [12]), and 17 recently developed prominent methods (RBD [63], HCT [17], DSR [62], MC [83], GC [23], DRFI [57], PCA [22], HS [21], TD [20], MR [61], GS [24], SF [50], SS [15], SEG [13], FT [49], LC [16] and SR [14]). Tab. 2 summarizes all the algorithms involved in our experiments.

5.1.3 Parameter settings

The parameters in the implementation of the proposed SMD detector are set as follows. In image abstraction, we set the number of patches N to 200. In tree construction, we set the affinity thresholds as $\mathcal{T} = [100, 400, 2000]$, producing three granularity-increasing segmentations. By adding the initial over-segmentation and the whole image, we build up a five-layer index tree. In matrix decomposition, we empirically

TABLE 2
Summary of all evaluated detection methods.

	Method	Hypothesis		Model
		Uniqueness	Prior[1]	
Top-down methods	**SMD**	global contrast	Sp, Ce, Co, Bg	structured matrix decomposition
	ULR [26]		Sp, Ce, Co, Ob	robust PCA
	LRR [27]		Sp, Ce	low-rank representation
	SLR [28]		Sp, Ob	robust PCA
	GC [23]		Ce	gaussian mixture model
	DRFI [57]	local contrast	Ce, Bg	random forest
	PCA [22]		Ce	principal component analysis
	TD [20]			sparse modeling
	HS [21]			hierarchical energy model
	CB [19]			heuristic context-based model
	RBD [63]		Bg	robust background modeling
	DSR [62]			sparse and dense reconstruction
	MR [61]			manifold ranking
	MC [83]			absorbing Markov chain
	GS [24]			shortest path
	SVO [18]		Ob	energy model
Bottom-up methods	HCT [17]	loc-global contrast	—	least square
	CA [7]			heuristic context-aware model
	RC [12]	global contrast		gaussian modeling
	SF [50]			double Gaussian filters
	FT [49]			frequency domain analysis
	SR [14]			spectral residual
	SS [15]			sparse signal analysis
	SEG [13]	local contrast		Bayesian statistics
	LC [16]			heuristic model

[1] The terms Sp, Ce, Co, Bg and Ob represent sparsity, center, color and objectness priors, respectively.

set the bandwidth parameter δ^2 to 0.05, and the model parameters α and β to 0.35 and 1.1 respectively.

To retain a fair comparison with competing methods, we fix the parameters of our model for all the experiments. It is worth noting that by tuning the parameters on the datasets, our model still has some potential to be improved, as presented in Appendix B. For other algorithms in our comparison, we use the source or binary codes provided by the authors with default parameters. The source code of our method and all experimental results are publicly available at *http://www.dabi.temple.edu/~hbling/SMD/SMDSaliency.html*. Our code is implemented in mixed C++ and Matlab, and its average runtime is 1.217 seconds per image on MSRA10K using a PC of 3.4 GHz and 4GB RAM, with only a single thread used.

5.1.4 Evaluation metrics

For comprehensive evaluation, we use seven metrics including the precision-recall (PR) curve, the F-measure curve, the receiver operating characteristic (ROC) curve, area under the ROC curve (AUC), mean absolute error (MAE), overlapping ratio (OR) and the weighted F-measure (WF) score.

Precision is defined as the percentage of salient pixels correctly assigned, while recall is the ratio of correctly detected salient pixels to all true salient pixels. F-measure is a weighted harmonic mean of precision (P) and recall (R): $F_\beta = (1+\beta^2)P \cdot R/(\beta^2 P + R)$, where β^2 is set to 0.3 to stress precision more than recall [49]. The PR and F-measure curves are created by varying the saliency threshold that determines whether a pixel belongs to the salient object. The ROC curve is generated from true positive rates and false positive rates which are obtained when we calculate the PR curve.

Although commonly used, the above metrics ignore the effects of correct assignment of non-salient pixels and the importance of complete detection. We therefore introduce the MAE and OR metrics to address these issues. Given a continuous saliency map S and the binary ground truth G, MAE is defined as the mean absolute difference between S and G : $\mathrm{MAE} = \mathrm{mean}(|S-G|)$ [50]. OR is defined as the overlapping ratio between the segmented object mask S' and ground truth G: $\mathrm{OR} = |S' \cap G|/|S \cup G|$, where S' is obtained by binarizing S using an adaptive threshold, i.e., twice the mean values of S as in [51]. Finally, we adopt the recently proposed weighted F-measure (WF) metric [32], which is a weighted version of the traditional F-measure. It amends the interpolation, dependency and equal importance flaws of currently-used measures.

5.2 Comparison with State-of-the-Arts

The proposed SMD algorithm is evaluated on the five benchmark datasets and compared with 24 recently proposed algorithms. The results are summarized in Tab. 3 and Fig. 7. Besides, Fig. 8 shows some qualitative comparisons.

The results show that, in most cases, SMD ranks first or second on the five benchmark datasets across different criteria. It is worth noting that, although DRFI [57] is the best performing method, it is a supervised one requiring a large amount of training. In contrast, our method is an unsupervised one, which skips the training process and therefore enjoys more flexibility.

5.2.1 Results on single-object images

The test on images with a single object is conducted on the MSRA10K [69] and DUT-OMRON [61] datasets. The PR and F-measure curves are shown in Fig. 7(A and B), and the WF, OR, AUC and MAE scores in Tab. 3(A and B).

On MSRA10K (Tab. 3(A)), SMD achieves the best performance in terms of WF, OR and MAE, while DRFI [57] obtains the best AUC score. In the PR curves (Fig. 7(A)), DRFI [57] and SMD are the best two among those competitive methods. In the F-measure curves, SMD is superior, as it achieves relatively good results over a large range.

On DUT-OMRON (Tab. 3(B)), all the methods perform worse than on MSRA10K due to the large diversity and complexity of DUT-OMRON. SMD performs the second best in terms of WF and OR, with a very minor margin (0.003) to the best results. The best MAE and AUC scores are achieved by DRFI [57]. This is because DRFI takes advantage of multi-level saliency maps fusion to improve its robustness. The fusion strategy is effective and general, as discussed in Appendix C. In the PR curves (Fig. 7(B)), the precision of SMD is less impressive at low recall rates, but it is competitive at the high recall rates. In terms of F-measure, SMD obtains relatively superior performance, especially when segmenting saliency maps with high thresholds.

5.2.2 Results on multiple-object images

Experiments on images with multiple salient objects are conducted on iCoSeg [71] and SOD [70]. The PR and F-measure curves are shown in Fig. 7(C and D), and the WF, OR, AUC and MAE scores in Tab. 3(C and D).

On iCoSeg (Tab. 3(C)), SMD achieves the best performance in terms of WF, OR and MAE. The AUC score of SMD is a little lower than the best, achieved by DRFI [57]. Fig. 7(C) shows that the PR and F-measure curves of SMD are superior or comparable to other methods. In particular, SMD's F-measure remains high over a wide range, indicating its insensitivity to the selection of a threshold.

On SOD (Tab. 3(D)), SMD performs the best in terms of WF, the second in OR and third in MAE. The PR of SMD

TABLE 3
Results on five datasets in terms of WF, AUC, OR and MAE.

	Metric	**SMD**	DRFI [57]	RBD [63]	HCT [17]	DSR [62]	MC [83]	MR [61]	HS [21]	PCA [22]	TD [20]	GC [23]	RC [12]	SVO [18]
(A) MSRA10K	WF↑[1]	**0.704**[2]	**0.666**	**0.685**	0.582	0.656	0.576	0.642	0.604	0.473	0.561	0.612	0.384	0.339
	OR↑	**0.741**	**0.723**	**0.716**	0.674	0.654	0.694	0.693	0.656	0.576	0.605	0.599	0.434	0.245
	AUC↑	**0.847**	**0.857**	0.834	**0.847**	0.825	0.843	0.824	0.833	0.839	0.815	0.788	0.833	0.844
	MAE↓[1]	**0.104**	**0.114**	**0.108**	0.143	0.121	0.145	0.125	0.149	0.185	0.161	0.139	0.252	0.340
	Metric	**SMD**	ULR [26]	SLR [28]	LRR [27]	GS [24]	SF [50]	CB [19]	CA [7]	SS [15]	SEG [13]	FT [49]	SR [14]	LC [16]
	WF↑	0.704	0.425	0.601	0.448	0.606	0.372	0.466	0.379	0.137	0.349	0.277	0.155	0.345
	OR↑	0.741	0.524	0.691	0.494	0.664	0.440	0.542	0.409	0.148	0.323	0.379	0.256	0.380
	AUC↑	0.847	0.831	0.840	0.801	0.839	0.812	0.821	0.789	0.601	0.795	0.690	0.597	0.690
	MAE↓	0.104	0.224	0.141	0.153	0.139	0.246	0.208	0.237	0.255	0.315	0.231	0.232	0.234
(B) DUT-OMRON	Metric	**SMD**	DRFI [57]	RBD [63]	HCT [17]	DSR [62]	MC [83]	MR [61]	HS [21]	PCA [22]	TD [20]	GC [23]	RC [12]	SVO [18]
	WF↑	**0.424**	**0.424**	**0.427**	0.353	0.419	0.347	0.381	0.350	0.287	0.320	0.358	0.228	0.203
	OR↑	**0.441**	**0.444**	**0.432**	0.393	0.408	0.425	0.420	0.397	0.341	0.337	0.342	0.272	0.151
	AUC↑	0.809	**0.839**	0.814	0.815	0.803	**0.820**	0.779	0.801	**0.827**	0.773	0.719	0.808	0.816
	MAE↓	0.166	**0.138**	**0.144**	0.164	**0.139**	0.186	0.187	0.227	0.207	0.205	0.197	0.290	0.409
	Metric	**SMD**	ULR [26]	SLR [28]	LRR [27]	GS [24]	SF [50]	CB [19]	CA [7]	SS [15]	SEG [13]	FT [49]	SR [14]	LC [16]
	WF↑	0.424	0.254	0.392	0.323	0.363	0.229	0.274	0.222	0.098	0.221	0.159	0.109	0.189
	OR↑	0.441	0.318	0.429	0.353	0.372	0.280	0.338	0.245	0.123	0.239	0.199	0.155	0.183
	AUC↑	0.809	0.805	0.822	0.793	0.814	0.778	0.775	0.771	0.612	0.779	0.636	0.607	0.626
	MAE↓	0.166	0.260	0.161	0.168	0.173	0.272	0.257	0.255	0.199	0.337	0.206	0.181	0.246
(C) iCoSeg	Metric	**SMD**	DRFI [57]	RBD [63]	HCT [17]	DSR [62]	MC [83]	MR [61]	HS [21]	PCA [22]	TD [20]	GC [23]	RC [12]	SVO [18]
	WF↑	**0.611**	**0.592**	**0.599**	0.464	0.548	0.461	0.554	0.536	0.407	0.499	0.522	0.395	0.296
	OR↑	**0.598**	**0.582**	**0.588**	0.519	0.514	0.543	0.573	0.537	0.427	0.506	0.487	0.402	0.293
	AUC↑	0.822	**0.839**	0.827	**0.833**	0.801	0.807	0.795	0.812	0.798	0.817	0.765	**0.829**	0.808
	MAE↓	**0.138**	**0.139**	**0.138**	0.179	0.153	0.179	0.162	0.176	0.201	0.180	0.176	0.234	0.336
	Metric	**SMD**	LR [26]	SLR [28]	LRR [27]	GS [24]	SF [50]	CB [19]	CA [7]	SS [15]	SEG [13]	FT [49]	SR [14]	LC [16]
	WF↑	0.611	0.379	0.473	0.465	0.519	0.347	0.441	0.315	0.126	0.301	0.289	0.152	0.340
	OR↑	0.598	0.443	0.505	0.530	0.520	0.433	0.459	0.297	0.164	0.346	0.387	0.227	0.348
	AUC↑	0.822	0.814	0.805	0.804	0.819	0.812	0.782	0.775	0.630	0.792	0.717	0.632	0.716
	MAE↓	0.138	0.222	0.179	0.170	0.167	0.247	0.201	0.259	0.253	0.326	0.223	0.229	0.227
(D) SOD	Metric	**SMD**	DRFI [57]	RBD [63]	HCT [17]	DSR [62]	MC [83]	MR [61]	HS [21]	PCA [22]	TD [20]	GC [23]	RC [12]	SVO [18]
	WF↑	**0.456**	**0.456**	0.428	0.385	**0.429**	0.390	0.406	0.410	0.343	0.392	0.367	0.335	0.320
	OR↑	**0.419**	**0.447**	**0.406**	0.377	0.398	0.392	0.373	0.325	0.340	0.344	0.281	0.247	0.083
	AUC↑	0.733	**0.742**	0.706	0.720	0.722	**0.746**	0.709	0.731	0.730	0.688	0.631	0.730	**0.734**
	MAE↓	**0.233**	**0.217**	**0.229**	0.243	0.234	0.260	0.261	0.283	0.274	0.279	0.272	0.326	0.413
	Metric	**SMD**	LR [26]	SLR [28]	LRR [27]	GS [24]	SF [50]	CB [19]	CA [7]	SS [15]	SEG [13]	FT [49]	SR [14]	LC [16]
	WF↑	0.456	0.322	0.395	0.382	0.416	0.296	0.362	0.320	0.143	0.306	0.212	0.151	0.247
	OR↑	0.419	0.290	0.400	0.393	0.390	0.212	0.311	0.268	0.114	0.148	0.191	0.197	0.201
	AUC↑	0.733	0.713	0.712	0.723	0.731	0.691	0.685	0.713	0.577	0.677	0.571	0.577	0.580
	MAE↓	0.233	0.308	0.248	0.245	0.251	0.329	0.294	0.312	0.310	0.360	0.316	0.291	0.317
(E) ECSSD	Metric	**SMD**	DRFI [57]	RBD [63]	HCT [17]	DSR [62]	MC [83]	MR [61]	HS [21]	PCA [22]	TD [20]	GC [23]	RC [12]	SVO [18]
	WF↑	**0.517**	**0.517**	**0.490**	0.430	0.489	0.441	0.480	0.449	0.358	0.413	0.437	0.320	0.316
	OR↑	**0.523**	**0.527**	0.494	0.457	0.480	**0.495**	0.491	0.432	0.371	0.398	0.376	0.265	0.084
	AUC↑	**0.775**	**0.780**	0.752	0.755	0.754	**0.779**	0.761	0.766	0.759	0.717	0.685	0.749	0.753
	MAE↓	**0.227**	**0.217**	**0.225**	0.249	0.227	0.251	0.235	0.269	0.291	0.271	0.256	0.334	0.427
	Metric	**SMD**	ULR [26]	SLR [28]	LRR [27]	GS [24]	SF [50]	CB [19]	CA [7]	SS [15]	SEG [13]	FT [49]	SR [14]	LC [16]
	WF↑	0.517	0.351	0.442	0.398	0.436	0.307	0.403	0.304	0.134	0.323	0.199	0.138	0.242
	OR↑	0.523	0.347	0.474	0.442	0.435	0.271	0.419	0.254	0.099	0.206	0.212	0.171	0.206
	AUC↑	0.775	0.755	0.764	0.756	0.758	0.725	0.762	0.702	0.561	0.719	0.600	0.562	0.585
	MAE↓	0.227	0.312	0.252	0.254	0.255	0.329	0.282	0.343	0.320	0.369	0.312	0.308	0.332

[1] The up-arrow ↑ indicates the larger value achieved, the better performance is, while the down-arrow ↓ indicates the smaller, the better.
[2] The best three results are highlighted with **red**, **green** and **blue** fonts, respectively.

is slightly lower than that of DRFI [57], but better than the others. In the F-measure curves, SMD performs the best at higher threshold ranges, while DRFI performs the best at lower ranges. Both are consistently superior to the others.

5.2.3 Results on complex scene images

Our last comparison with the competing methods is conducted on ECSSD [21], which is known to involve complex scenes. As reported in Tab. 3(E), SMD obtains the best performance in terms of WF, the second or third best in OR, AUC and MAE. According to Fig. 7(E), the PR curve of SMD is the second best among those methods, while the area under the F-measure curve is the best. These results validate SMD's strong potential in handling images with complex scenes.

5.2.4 Visual comparison

Fig. 8 shows some visual comparisons of the best methods in the experiments. For single-object images, SMD accurately extracts the entire salient object with few scattered patches, and assigns nearly uniform saliency values to all patches within the salient objects. For images with multiple objects, some methods (e.g., SLR [28], ULR [26] and MR [61]) miss detecting parts of the objects, while some (e.g., HS [21] and HCT [17]) incorrectly include background regions into detection results. By contrast, SMD pops out all the salient objects successfully. For the images with complex scenes, most methods fail to identify the salient objects, while SMD locates them with decent accuracy. Finally, for the images whose foreground and background share similar appearance, SMD successfully separates the salient objects from the background, while other methods often fail. These results illustrate the robustness of the SMD algorithm, and confirm the effectiveness of the proposed structural constraints in separating the coherent low-rank and sparse subspaces.

5.3 Experimental Analysis of the Proposed Method

5.3.1 Analysis of components in the proposed model

To further understand the effects of the components in the proposed SMD algorithm, we test four variations of SMD on the MSRA10K dataset. In particular, each variation

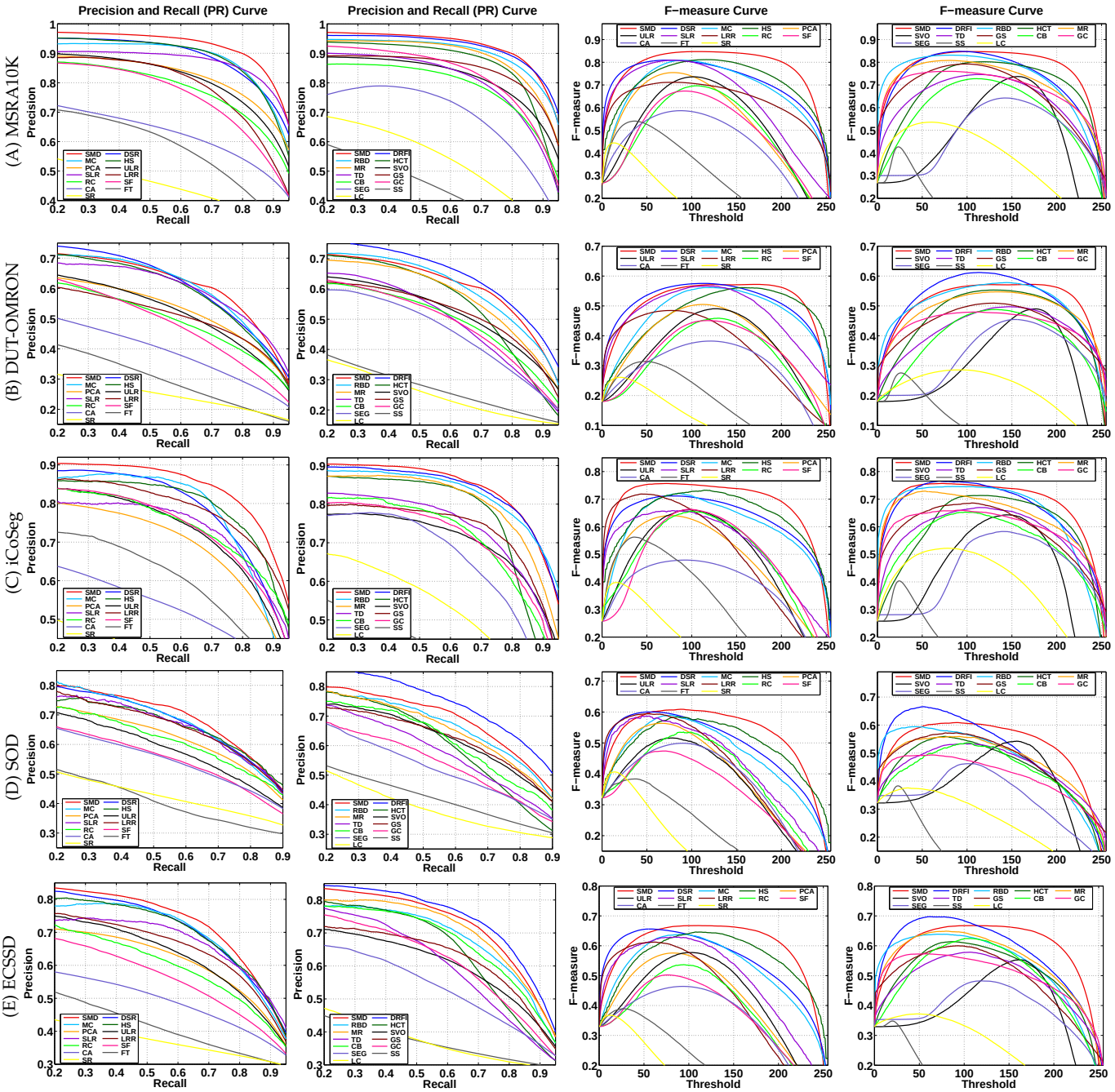

Fig. 7. Quantitative comparison on five datasets in terms of PR and F-measure curves.

corresponds to an objective function listed in Tab. 4, and parameters for each model are tuned separately to obtain optimal results. Furthermore, only low-level features are used to avoid the influence of high-level prior knowledge.

The quantitative results are shown in Fig. 9(left and middle), leading to the following observations. (1) By comparing LR-ℓ_1 with LR-Tree$_1$, we see that encoding tree-structure information gives in average improvement of 4.69% (precision) and 2.76% (true positive rate) over the plain ℓ_1-norm.

TABLE 4
The objective function of different models related to SMD.

Model	Objective Function
LR-ℓ_1	$\min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \alpha\|\mathbf{S}\|_1$
LR-Tree$_1$	$\min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \alpha \sum_{G\in T} \|\mathbf{S}_G\|_1$
LR-Tree$_\infty$	$\min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \alpha \sum_{G\in T} \|\mathbf{S}_G\|_\infty$
SMD	$\min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \alpha \sum_{G\in T} \|\mathbf{S}_G\|_\infty + \beta \mathrm{Tr}(\mathbf{S}\mathbf{M}_\mathbf{F}\mathbf{S}^T)$

(2) The ℓ_∞-norm embedding in the structured sparsity slightly improves the ℓ_1-norm (comparing LR-Tree$_1$ and LR-Tree$_\infty$). (3) The use of the Laplacian regularization significantly improves the LR-Tree$_\infty$ model. These observations indicate that the introduced regularizers are effective and complementary, and, when combined together, lead to excellent performance as reported in the previous subsection.

We further analyze the underlying reasons for the above observed improvements by comparing the saliency maps. As shown in Fig. 11, we observe that: (1) The salient regions identified by LR-Tree$_1$ tend to be connected, whereas the regions identified by LR-ℓ_1 tend to be scattered. This shows that the tree-structured constraint guides matrix decomposition along a structurally meaningful direction. (2) The LR-Tree$_\infty$ model produces smoother saliency maps than LR-Tree$_1$, since the ℓ_∞-norm forces the patches within the

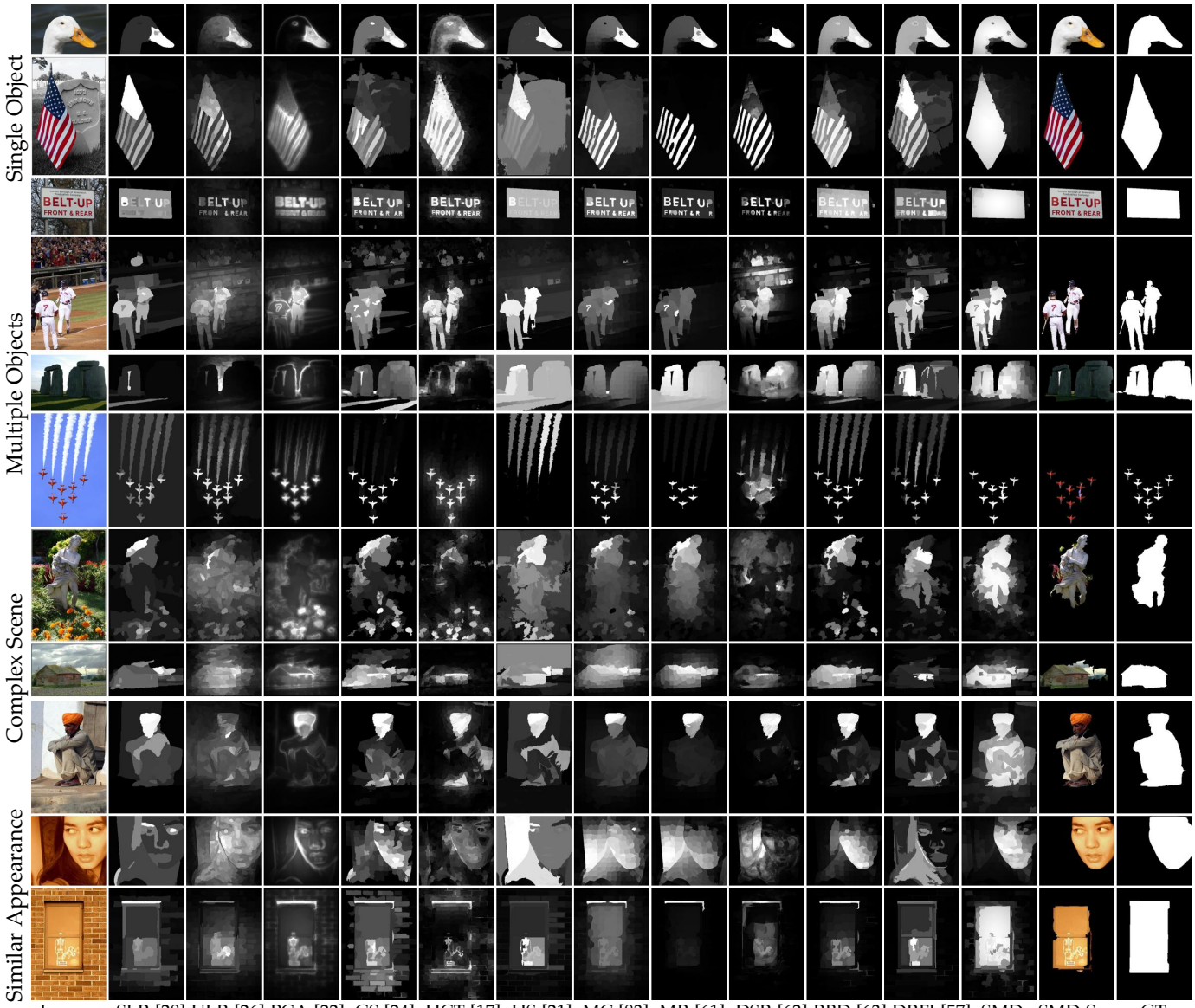

Fig. 8. Visual comparisons of saliency maps of the best methods. Our segmentation results (SMD-Seg), which are produced by simple adaptive thresholding on the saliency maps (SMD), are close to ground truth (GT).

same group to share identical values. (3) The final SMD model produces foreground-background separated maps, whose saliency values are consistent within regions. This is attributed to Laplacian regularization. To make this point clear, we introduce a metric (Sec. 3.2.2 of [84]) to compute the projection distance $d(\cdot,\cdot)$ between the feature subspaces of salient objects ($\mathbf{S}$) and background ($\mathbf{L}$): $d(\mathbf{L},\mathbf{S}) = \|\mathbf{L}\mathbf{L}^T - \mathbf{S}\mathbf{S}^T\|_F^2$. By evaluating the change of $d(\mathbf{L},\mathbf{S})$ before and after imposing the Laplacian regularization, we observe that the projection distance $d(\mathbf{L},\mathbf{S})$ is significantly enlarged, as shown in Fig. 9(right). It shows that the Laplacian regularization boosts the gap between foreground and background.

5.3.2 Analysis of parameters and implementation details

We also analyze the sensitivity of our model to changes of the main parameters α and β. The analysis is conducted by fixing one parameter and tuning the other on MSRA10K. The performance changes are shown in Fig. 12. We observe that, when β is fixed ($\beta = 1.1$), the WF performance initially increases, spikes within a range of α from 0.2 to 0.5, and then decreases. When fixing α to be 0.35 and increasing β, the performance rapidly increases as β approaches 0.6, and then flattens when β crosses 0.8. These observations indicate that our model has only a small sensitivity to changes of the parameters. It works well under a large range of parameter settings, such as α ranging from 0.25 to 0.5, and β ranging from 0.8 to 1.2.

To further analyze the proposed method, we evaluate the effects of some implementation details on the performance. We conduct an comparison experiment to evaluate whether more complex features can affect the model. Specifically, we replace the original 53-dimensional color, edge and texture features with the 93-dimensional discriminative regional features used in DRFI [57], and perform SMD in the same setting. From the experimental results shown in Appendix D, we observe that the complex features perform comparable or slightly superior to the original low-lever features (comparing SMD_regFeat and SMD). It tells us that, to some extend, our method is robust to features. We also evaluate different saliency assignment functions and analyze the effects of the context-based propagation used in our method. The detailed experimental results and analysis are presented

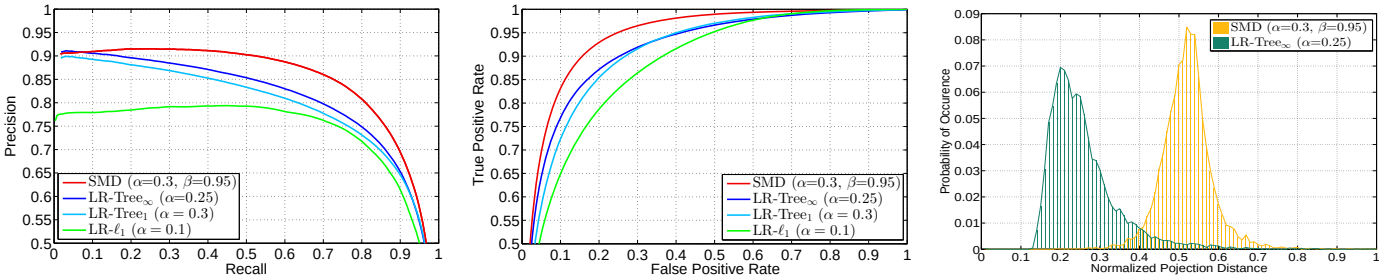

Fig. 9. Left and Middel: The evaluation of performance contribution of each component in our SMD model with respect to PR and ROC metrics. Right: The projection distance distribution of images in MSRA10K dataset before and after imposing the Laplacian regularization.

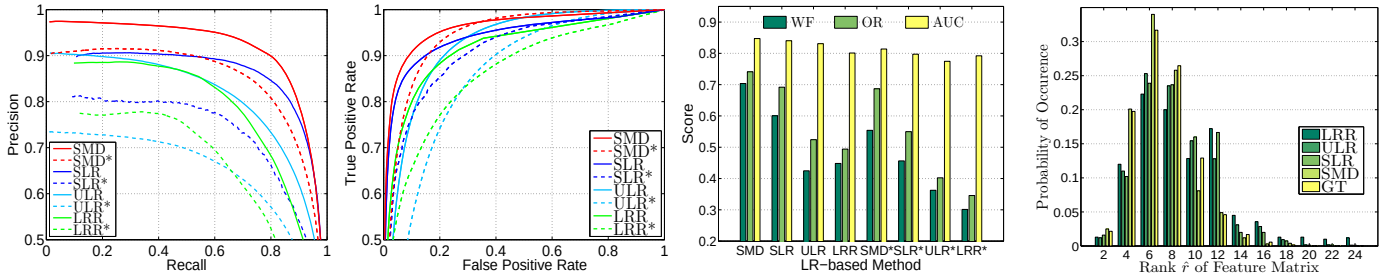

Fig. 10. Left three: comparison of SMD with other LR-based methods (SLR [28], ULR [26] and LRR [27]). The superscript '*' indicates methods without using high-level priors. Rightmost: comparison of rank distribution estimated on the produced background feature matrices.

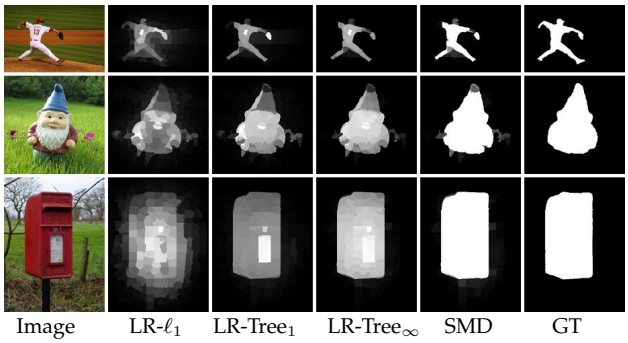

Fig. 11. Saliency maps produced by variations of the SMD model.

in Appendix E and F.

5.3.3 Comparison with LR-based methods

We proceed to compare the proposed SMD method with other LR-based saliency detection methods on MSRA10K under two conditions: with and without high-level priors.

In the case of pure low-level saliency detection (i.e., without high-level priors), Fig. 10 shows that SMD consistently outperforms other LR-based methods in all metrics. In particular, the improvement of SMD over ULR [26] indicates that the integration of image structure information is superior to the learnt feature transformation in matrix decomposition.

When taking high-level priors into account, all the LR models are improved as validated in Fig. 10. SMD again achieves the best performance over all metrics. It indicates that both the structured regularization and high-level priors are beneficial for salient object detection.

Last, rank statistics of the background feature matrix $\mathbf{L}$ are collected for the above LR methods, as summarized in Fig. 10 (rightmost). The results show that the matrices estimated by SMD achieve the lowest ranks among all the LR methods, and their rank distribution is similar to that calculated over the ground truth (GT). This implies that the structured-sparsity and Laplacian regularizations are complementary to the low-rank regularization in matrix decomposition for estimating the intrinsic rank of image features.

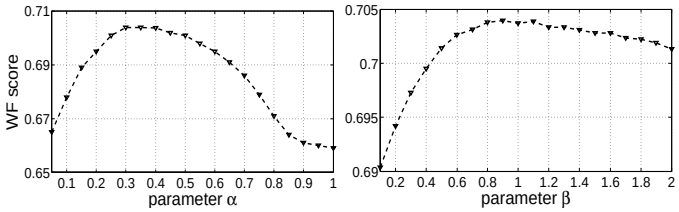

Fig. 12. The sensitivity analysis of parameter α and β.

5.3.4 Failure cases

Our method exploits the low-rank regularization to recover image background, therefore it may be difficult to suppress some small background regions with distinctive appearances, as shown in Fig. 13. The underlying reason is that the feature vectors of those regions are not in the low-dimensional subspace and may be incorrectly highlighted as foreground. Besides, for the salient objects with partial occlusion (see the third column in Fig. 13), SMD fails to consistently highlight the salient object because the constructed index-tree is not precise enough. Exploring more effective region grouping methods, such as [85], may alleviate this problem.

6 Conclusion

In this paper, we have presented a structured matrix decomposition (SMD) model, which formulates the task of salient object detection as a problem of low-rank and structured-sparse matrix decomposition. A hierarchical tree-structured sparsity-inducing norm has been proposed to encode the underlying structure of the image in the feature space, while a Laplacian regularization has been introduced to enlarge the distance between the representation of salient objects and that of the background. High-level prior knowledge has also been integrated into the model to enhance the detection. Experiments on five public datasets have shown that our model achieves encouraging performance compared to the state-of-the-art methods.

Fig. 13. Some failure cases of our method.

For future work, we will consider integrating metric learning or discriminative analysis to explicitly separate the low-rank and structured-sparse matrices in terms of regional difference. In addition, the exploration of more robust and general high-level priors may merit further study.

ACKNOWLEDGMENTS

The authors would like to thank the reviewers and editor for their helpful comments to improve the paper. They thank Dr. Wenbin Zou, Congyan Lang and Rongrong Ji for providing their code, results or helpful suggestions. This work is partly supported by the 973 basic research program of China (Grant No. 2014CB349303), the Natural Science Foundation of China (Grant No. 61472421), the Project Supported by CAS Center for Excellence in Brain Science and Intelligence Technology, and the Project Supported by Guangdong Natural Science Foundation (Grant No. S2012020011081). Ling is supported in part by the US NSF Grants IIS-1218156, 1449860 and 1350521.

REFERENCES

[1] V. Navalpakkam and L. Itti, "An integrated model of top-down and bottom-up attention for optimizing detection speed," in *CVPR*, 2006, pp. 2049–2056.
[2] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," *IEEE PAMI*, vol. 34, no. 11, pp. 2189–2202, 2012.
[3] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition?" in *CVPR*, 2004, pp. 37–44.
[4] D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural Networks*, vol. 19, no. 9, pp. 1395–1407, 2006.
[5] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu, "Sketch2photo: internet image montage," *ACM TOG*, vol. 28, 2009.
[6] P. Wang, J. Wang, G. Zeng, J. Feng, H. Zha, and S. Li, "Salient object detection for searched web images via global saliency," in *CVPR*, 2012, pp. 3194–3201.
[7] S. Goferman, L. Z. Manor, and A. Tal, "Context-aware saliency detection," in *CVPR*, 2010, pp. 1915–1926.
[8] L. Marchesotti, C. Cifarelli, and G. Csurka, "A framework for visual saliency detection with applications to image thumbnailing," in *ICCV*, 2009, pp. 2232–2239.
[9] R. Margolin, L. Zelnik-Manor, and A. Tal, "Saliency for image manipulation," *The Visual Computer*, vol. 29, pp. 381–392, 2013.
[10] J. Sun and H. Ling, "Scale and object aware image thumbnailing," *IJCV*, vol. 104, no. 2, pp. 135–153, 2013.
[11] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE TPAMI*, vol. 20, no. 11, pp. 1254–1259, 1998.
[12] M. Cheng, G. Zhang, N. J. Mitra, X. Huang, and S. Hu, "Global contrast based salient region detection," in *CVPR*, 2011.
[13] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä, "Segmenting salient objects from images and videos," in *ECCV*, 2010, pp. 366–379.
[14] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *CVPR*, 2007, pp. 1–8.
[15] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," *IEEE PAMI*, vol. 34, pp. 194–201, 2012.
[16] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues." in *ACM MM*, 2006.
[17] J. Kim, D. Han, Y. Tai, and J. Kim, "Salient region detection via high-dimensional color transform," in *CVPR*, 2014, pp. 883–890.
[18] K.-Y. Chang, T.-L. Liu, H.-T. Chen, and S.-H. Lai, "Fusing generic objectness and visual saliency for salient object detection," in *ICCV*, 2011, pp. 914–921.
[19] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li, "Automatic salient object segmentation based on context and shape prior," in *BMVC*, 2011, pp. 1–12.
[20] C. Scharfenberger, A. Wong, K. Fergani, J. S. Zelek, and D. A. Clausi, "Statistical textural distinctiveness for salient region detection in natural images," in *CVPR*, 2013, pp. 979–986.
[21] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *CVPR*, 2013, pp. 1155–1162.
[22] R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct?" in *CVPR*, 2013, pp. 1139–1146.
[23] M. Cheng, J. Warrell, W. Lin, S. Zheng, V. Vineet, and N. Crook, "Efficient salient region detection with soft image abstraction," in *ICCV*, 2013, pp. 1529–1536.
[24] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *ECCV (3)*, 2012, pp. 29–42.
[25] J. Yan, M. Zhu, H. Liu, and Y. Liu, "Visual saliency detection via sparsity pursuit," *IEEE SPL*, vol. 17, no. 8, pp. 739–742, 2010.
[26] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *CVPR*, 2012, pp. 2296–2303.
[27] C. Lang, G. Liu, J. Yu, and S. Yan, "Saliency detection by multitask sparsity pursuit," *IEEE TIP*, vol. 21, no. 3, pp. 1327–1338, 2012.
[28] W. Zou, K. Kpalma, Z. Liu, and J. Ronsin, "Segmentation driven low-rank matrix recovery for saliency detection," in *BMVC*, 2013.
[29] E. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–39, 2011.
[30] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE PAMI*, vol. 35, no. 1, pp. 171–184, 2013.
[31] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *NIPS*, 2011, pp. 612–620.
[32] R. Margolin, L. Zelnik-Manor, and A. Tal, "How to evaluate foreground maps." in *CVPR*, 2014, pp. 248–255.
[33] A. Borji, D. N. Sihite, and L. Itti, "Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study," *IEEE TIP*, vol. 22, no. 1, pp. 55–69, 2013.
[34] T. Judd, F. Durand, and A. Torralba, "A benchmark of computational models of saliency to predict human fixations," *MIT tech report, Tech. Rep.*, 2012.
[35] Q. Zhao and C. Koch, "Learning saliency-based visual attention: A review," *Signal Processing*, vol. 93, no. 6, pp. 1401–1407, 2013.
[36] A. Borji, D. N. Sihite, and L. Itti, "Salient object detection: A benchmark," in *ECCV*, 2012, pp. 414–429.
[37] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, "Salient object detection: A survey," *Submmited to IEEE PAMI*.
[38] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, no. 4, pp. 219–227, 1985.
[39] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive psychology*, vol. 12, no. 1, pp. 97–136, 1980.
[40] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *NIPS*, 2006, pp. 545–552.
[41] Q. Zhao and C. Koch, "Learning a saliency map using fixated locations in natural scenes," *Journal of Vision*, vol. 11, no. 3, pp. 1–15, 2011.
[42] Q. Zhao and C. Koch, "Learning visual saliency by combining feature maps in a nonlinear manner using adaboost," *Journal of Vision*, vol. 12, no. 6, pp. 1–15, 2012.
[43] D. Gao, V. Mahadevan, and N. Vasconcelos, "The discriminant center-surround hypothesis for bottom-up saliency," in *NIPS*, 2007.
[44] D. A. Klein and S. Frintrop, "Center-surround divergence of feature statistics for salient object detection," in *ICCV*, 2011, pp. 2214–2219.
[45] F. Liu and M. Gleicher, "Region enhanced scale-invariant saliency detection," in *ICME*, 2006, pp. 1477–1480.
[46] N. D. B. Bruce and J. K. Tsotsos, "Saliency based on information maximization," in *NIPS*, 2005.
[47] J. Li, M. D. Levine, X. An, X. Xu, and H. He, "Visual saliency based on scale-space analysis in the frequency domain," *IEEE PAMI*, vol. 35, no. 4, pp. 996–1010, 2013.

[48] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, and H. Shum, "Learning to detect a salient object," *IEEE PAMI*, vol. 33, 2011.
[49] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, "Frequency-tuned salient region detection," in *CVPR*, 2009, pp. 1597–1604.
[50] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *CVPR*, 2012, pp. 733–740.
[51] X. Li, Y. Li, C. Shen, A. R. Dick, and A. van den Hengel, "Contextual hypergraph modeling for salient object detection," in *ICCV*, 2013, pp. 3328–3335.
[52] Y. Niu, Y. Geng, X. Li, and F. Liu, "Leveraging stereopsis for saliency analysis," in *CVPR*, 2012, pp. 454–461.
[53] H. Peng, B. Li, W. Xiong, W. Hu, and R. Ji, "RGBD salient object detection: A benchmark and algorithms," in *ECCV*, 2014, pp. 1–18.
[54] J. Zhang and S. Sclaroff, "Saliency detection: A boolean map approach," in *ICCV*, 2013, pp. 153–160.
[55] J. Yang and M. Yang, "Top-down visual saliency via joint crf and dictionary learning," in *CVPR*, 2012, pp. 2296–2303.
[56] L. Mai, Y. Niu, and F. Liu, "Saliency aggregation: A data-driven approach," in *CVPR*, 2013, pp. 1131–1138.
[57] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *CVPR*, 2013, pp. 1–8.
[58] M. Jiang, J. Xu, and Q. Zhao, "Saliency in crowd," in *ECCV*, 2014, pp. 17–32.
[59] S. He, R. W. Lau, W. Liu, Z. Huang, and Q. Yang, "Supercnn: A superpixelwise convolutional neural network for salient object detection," *International Journal of Computer Vision*, pp. 1–15.
[60] K. Shi, K. Wang, J. Lu, and L. Lin, "PISA: pixelwise image saliency by aggregating complementary appearance contrast measures with spatial priors," in *CVPR*, 2013, pp. 2115–2122.
[61] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *CVPR*, 2013, pp. 3166–3173.
[62] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *ICCV*, 2013, pp. 2976–2983.
[63] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *CVPR*, 2014, pp. 2814–2821.
[64] P. Jiang, H. Ling, J. Yu, and J. Peng, "Salient region detection by ufo: Uniqueness, focusness and objectness," in *ICCV*, 2013, pp. 1976–1983.
[65] Y. Jia and M. Han, "Category-independent object-level saliency detection," in *ICCV*, 2013, pp. 1761–1768.
[66] T. Judd, K. A. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *ICCV*, 2009, pp. 2106–2113.
[67] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf, "Ranking on data manifolds," in *NIPS*, 2003.
[68] H. Peng, B. Li, R. Ji, W. Hu, W. Xiong, and C. Lang, "Salient object detection via low-rank and structured sparse matrix decomposition," in *AAAI*, 2013.
[69] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE TPAMI*, 2014.
[70] V. Movahedi and J. H. Elder, "Design and perceptual validation of performance measures for salient object segmentation," in *POCV*, 2010.
[71] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen, "Interactively co-segmentating topically related images with intelligent scribble guidance," *IJCV*, vol. 93, no. 3, pp. 273–292, 2011.
[72] J. Ye, "Generalized low rank approximations of matrices," in *ICML*, 2004.
[73] S. Lu, X. Ren, and F. Liu, "Depth enhancement via low-rank matrix completion." in *CVPR*, 2014, pp. 3390–3397.
[74] J. Liu and J. Ye, "Moreau-yosida regularization for grouped tree structure learning," in *NIPS*, 2010, pp. 1459–1467.
[75] K. Jia, T. Chan, and Y. Ma, "Robust and practical face recognition via structured sparsity," in *ECCV*, 2012, pp. 331–344.
[76] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE PAMI*, vol. 33, no. 8, pp. 1548–1560, 2011.
[77] J. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
[78] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach, "Proximal methods for hierarchical sparse coding," *JMLR*, vol. 12, pp. 7–24, 2011.
[79] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *ICIP*, 1995, pp. 444–447.
[80] H. G. Feichtinger and T. Strohmer, *Gabor analysis and algorithms:theory and applications*. Springer, 1998.
[81] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE TPAMI*, vol. 34, no. 11, pp. 2274–2282, 2012.
[82] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *IJCV*, vol. 59, no. 2, pp. 167–181, 2004.
[83] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, "Saliency detection via absorbing markov chain," in *ICCV*, 2013, pp. 65–72.
[84] J. Ham and D. D. Lee, "Grassmann discriminant analysis: a unifying view on subspace-based learning," in *ICML*, 2008, pp. 376–383.
[85] P. A. Arbeláez, J. Pont-Tuset, J. T. Barron, F. Marqués, and J. Malik, "Multiscale combinatorial grouping," in *CVPR*, 2014, pp. 328–335.

Houwen Peng received the BE degree from Dalian University of Technology, China, in 2011. Currently, he is a PhD student jointly training in the Institute of Automation, Chinese Academy of Sciences, and the Department of Computer and Information Science, Temple University. His research interests include pattern recognition, computer vision, and machine learning.

Bing Li received the Ph.D. degree from the Department of Computer Science and Engineering, Beijing Jiaotong University, Beijing, China, in 2009. Currently, he is an associate professor with the Institute of Automation, Chinese Academy of Sciences, Beijing. His research interests include color constancy, visual saliency, and web content mining.

Haibin Ling received the BS and MS degrees from Peking University, China, in 1997 and 2000, respectively, and the PhD degree from the University of Maryland, College Park, in 2006. From 2006 to 2007, he worked as a postdoctoral scientist at UCLA. In 2008, he joined Temple University where he is now an associate professor. His research interests include computer vision and medical image analysis.

Weiming Hu received the PhD degree from the Department of Computer Science and Engineering, Zhejiang University in 1998. From 1998 to 2000, he was a postdoctoral research fellow with the Institute of Computer Science and Technology, Peking University. Currently, he is a professor in the Institute of Automation, Chinese Academy of Sciences. His research interests include visual motion analysis and recognition of web objectionable information.

Weihua Xiong received the Ph.D. degree from the Department of Computer Science, Simon Fraser University, Vancouver, BC, Canada, in 2007. Currently, he is an imaging scientist at OmniVision Technologies Inc.. His research interests include color science, computer vision, color image processing, and stereo vision. He is a reviewer, committee member or session chair for several international conference and journals.

Stephen J. Maybank Stephen J. Maybank received the BA degree in mathematics from Kings College Cambridge in 1976, and the PhD degree in computer science from Birkbeck College, University of London in 1988. He is currently a professor in the Department of Computer Science and Information Systems, Birkbeck College. His research interests include the geometry of multiple images, camera calibration, visual surveillance, etc. He is a fellow of the IEEE.