Bi-Subspace Saliency Detection

Jin Lu
Department of Computer Science and Engineering
University of Connecticut
Storrs, Connecticut
Email: jin.lu@uconn.edu

Fei Dou
Department of Computer Science and Engineering
University of Connecticut
Storrs, Connecticut
Email: fei.dou@uconn.edu

***Abstract*—Relevant to visual attention mechanism, visual saliency are concerns of detecting the salient regions in image and video. Saliency detection is a critical pre-process in many computer vision applications, which requires high-level visual recognition and scene understanding as well as low-level photo and video processing. In this paper, We propose a new bi-subspace bottom-up saliency detection model. Inspired by visually intuitions and bi-subspace priors, we formulate the problem as subspace analysis depicting the background aside from the saliency object in the image. In the meanwhile, another subspace representation via a group-sparse constraint is proposed to depict the structure of the object in our model. Numerical experiment results show that our method is more stable and accurate than other methods.**

I. Introduction

A human's brain process mass of visual information in visible surrounding environment everyday. The various processions conducted in visual system are the concentration of many research fields such as biology [1], neuroscience [2], cognitive science [3] and psychology [4]. Among all the study of those processions, visual attention research which is a determining component concerns about the mechanism of people's eye-movement to visible objects. Born to deal with this problem from the perspective of computer vision and different from the specified purpose of object recognition to detect human face or car, saliency detection aims to detect the most informative and important region of a scene. In the last few years, saliency detection has numerous applications in vision problems including object recognition [5], [6] and video compression [7], [8] as preprocessing steps to focus on the area of interest [9], [10].

Visual saliency detection approaches can be viewed from two main categories - top-down [11] and bottom-up - in general. Bottom-up saliency method considers mainly about pixel-level visual response. Though regardless of any high-level prior knowledge, bottom-up oriented approaches simulate the pattern which people's visual stimulation mainly prefers under good visible environmental condition [3].

In this paper our purpose is to establish a bottom-up saliency detection method through casting this problem into bi-subspace analysis inspired by previous works based on low-rank matrix completion[12]. First of all, we start from elemental results with our common intuitions:

1. The salient object and the un-salient background seemingly belong to "two completely different worlds", namely two subspaces.

2. Different parts in the image belong to "different worlds (subspaces)" as well.

According to the result above, We re-frame group-sparse Robust Principal Component Analysis (R-PCA) [13] recovery model with a simple geometry structural prior to our detection method in this literature. Comparing to former low-rank approaches, the main contribution of our bi-subspace model is that our method captures the information from two subspaces. Moreover, the proposed approach better depicts the structure of the sparsity of saliency map in the image. In the following, we will do a shortly review about some existing mainstream approaches of saliency detection in Section 2 followed by introducing our approach.

II. Related Work

Biological-based works first emerged before the other techniques of saliency detection. According to Koch and Ullman [14], [15], [16], bio-inspired model was first proposed to detect saliency object by Itti et.al. [15]. They proposed a model inspired by biology together with feature integration and the this approach utilized a linear filter on the input image decomposition and constructed Gaussian image pyramid of the brightness map, the color map and the direction map respectively. While in multi-scale space, based on the center-surrounded difference principle, the method also constructed feature maps of brightness, color and direction. This method provides a framework that can quickly detect significant visual points. Ma and Zhang [17] adopt the simulation of visual perception to provide a saliency map. Hael, et.al. [18] modified Itti's model based on the graph structure, which regards all cells (pixels, for example) of the image as nodes in one graph and consequently extracts feature vector for each node, and all these nodes intend to constitute a Markov chain. Besides the biological-based approach, many other kinds of work burst in these years. Zhai and Shah [19] defined saliency of a pixel comparing with the rest of the pixels and proposed a statistical method on frequency histogram. Another influential global contrast based saliency detection algorithm proposed by Hou and Zhang [20] who model the global contrast in frequency domain. In addition, many block-based saliency detection approaches includes Achanta [21], who proposed a

978–1–5090–4228–9/17/$31.00 ©

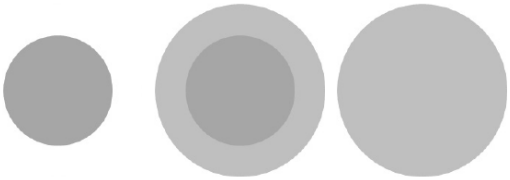

Fig. 1. Color Contrast. The "distance" has an significant effect on humans' eyes to differ the foreground from background.

frequency-tuned approach relying on DoG band pass filters. Cheng et.al. [10] divide the input image by using image segmentation into a plurality of regions and build histogram of colors in each area. As deep learning has recently been brought to computer vision, DL-related methods [22], [23], [24] shed a light upon a new perspective to saliency detection. In contrast to the engineering-handcrafted features for image processing, features extracted via CNN preserve high-level semantic information. However, all the CNNs need heavy parameter-tuning workloads and the parameters in the networks must be typically pre-trained on datasets for visual recognition tasks, which is impossible in some situations. An unignorable example is for medical image datasets, the number of patients is too limited to be well-trained in the network. In this case, deep learning could fail to be constructed due to its high complexity. Furthermore, all DL-related works are time consuming since the problem is non-convex and its gradient-descent-based method can't guarantee its convergence. This drawback limits its application to any fields that couldn't afford to days of model-training and high-end computational devices. For example, training a CNN model for saliency detection takes over 2 GPU days and requires hundreds of gigabytes of storage for the 5000 images in the MSRA-B dataset.

Alternatively, There are also some non-DL approaches trying to consider the problem from other aspects. Shen et.al. brought up a approach based on Unified Low Rank to combine low-level features with high-level priors [12]. Lang et.al. [25] provided a Multi-Task Sparsity Pursuit (MTSP) method to incorporate multi-types of features. Li et.al. [26] build tensor analysis model to search for the most suitable local features and their combinational coefficients. Different from these approaches, we model the problem motivated by more reasonable priors and conduct the structure of saliency map by describing its geometric feature of sparsity.

III. Four Visual Intuitions Based on Bi-Subspace Priors

In this section, we detail the elemental results mentioned in section I and add two more priors for better description about visual saliency.

1. While an image in a specified high-dimensional feature space, its background(the stuff which is not salient) can be located in one low-dimensional subspace. Namely, the background can be represented as "low-rank"; furthermore, the salient object is considerably far away from the subspace. As shown in Fig. 1, humans tend to trust their eyes and believe the left circle is more "salient" than the middle one with the same size due to the different distance between the salient object and the background can shift the feeling of our visual system by its biologic mechanism.

Fig. 2. Convex set supported by interesting points. The points marked are generated by [27], [28]

2. While an image in a specified high-dimensional feature space, every cell of this image can be linearly approximated by other cells; in particular, each cell is located in one small subspace, i.e., pieces contained in any part of the object somehow look alike and therefore we can just represent every piece with other pieces within the same part.

3. The salient spots in an image must be sparse; Meanwhile, it could inlay in a convex set in the image as shown in Fig.2.

4. Each part constituting in the object has substantially the same saliency value. That implies whenever we need to find saliency object, we ought to segment the scenery into parts and then consider it belongs to the salient object or not.

IV. Proposed Method

A. Subspace A: Low-rank subspace of the background

Firstly we over-segment the image into N super-pixels (or cells) by mean-shift method; we notate each super-pixel as $b_i \in \mathbb{R}^D$. Combining all vectors into the matrix we get $\mathbf{B} = [b_1, b_2, ..., b_N] \in \mathbb{R}^{D \times N}$. As we have already shown as Prior 1, $\mathbf{B}$ can be decomposed as a information-redundant matrix $\mathbf{L}$ and sparse matrix $\mathbf{S}$ as follows,

$$\min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \sum_{i=1}^{n} \omega_i \|\mathbf{S}_{p_i}\|_{2,1} \quad s.t. \quad \mathbf{B} = \mathbf{L} + \mathbf{S} \qquad (1)$$

Note that $\sum_{i=1}^{n}\omega_i\|\mathbf{S}_{p_i}\|_{2,1}$ is a sparse-induced term, ω_i is the weight for each group and $\mathbf{S}_{p_i}$ is the ith sub-matrix of $\mathbf{S}$.

Correspondingly in our low-rank model $\mathbf{L}$ is the representation of background regions with significant redundancy added by $\mathbf{S}$ which suggests the salient object outside of the subspace including the background. To measure the saliency of each segments, our approach prefer $l_{2,1}$-norm in $\mathbf{S}$ because in our expectation $\mathbf{S}$ should be a column-sparse matrix, since such a constraint condition suitably annotates our assumed priors that the salient segments in the image is sparse.

To solve the convex problem, extending the Augment Lagrange Multipliers (ALM)[29] algorithm offers a framework for optimizing it. Therefore we transform Eq. (1) into the Lagrangian function:

$$\begin{aligned} L(\mathbf{L},\mathbf{S},\mathbf{G},\mu) = \|\mathbf{L}\|_* + \lambda\sum_{i=1}^{n}\omega_i\|\mathbf{S}_{p_i}\|_{2,1} + \\ < \mathbf{G}, \mathbf{B}-\mathbf{L}-\mathbf{S} > + \frac{\mu}{2}\|\mathbf{B}-\mathbf{L}-\mathbf{S}\|_F^2 \end{aligned} \tag{2}$$

Eq. (2) introduces a Lagrange multiplier $\mathbf{G}$, and $\mu > 0$ is a penalty parameter. To minimize the Lagrange function L in Eq. (2), we alternatively optimize $\mathbf{L}$, $\mathbf{S}$, and $\mathbf{G}$ by iteration. Next we focus on the optimization for the three variables respectively.

Computing $\mathbf{L}$: Firstly let us fix $\mathbf{S}$ and $\mathbf{G}$ to optimize $\mathbf{L}_{k+1}$ at the $(k+1)$th iteration. Then Eq. (2) is equivalent to the minimization below:

$$\begin{aligned} \mathbf{L}^{k+1} &= \arg\min_{\mathbf{L}} L(\mathbf{L},\mathbf{S}^k,\mathbf{G}^k,\mu^k) \\ &= \arg\min_{\mathbf{L}} \|\mathbf{L}\|_* + < \mathbf{G}^k, \mathbf{B}-\mathbf{L}-\mathbf{S}^k > + \frac{\mu^k}{2}\|\mathbf{B}-\mathbf{L}-\mathbf{S}^k\| \\ &= \arg\min_{\mathbf{L}} \sigma\|\mathbf{L}\|_* + \frac{1}{2}\|\mathbf{L}-\mathbf{V}_\mathbf{L}\|_F^2, \end{aligned} \tag{3}$$

in which $\sigma = \frac{1}{\mu^k}$ is a trade-off parameter between the nuclear norm and the l_2 norm. Here $\mathbf{V}_\mathbf{L} = \mathbf{B} - \mathbf{S}^k + \frac{1}{\mu^k}\mathbf{G}^k$.

It is easy to check that the solution to Eq. (3) could be given by

$$\begin{aligned} &\mathbf{L}^{k+1} = \mathbf{U}\Gamma_\sigma(\Sigma)\mathbf{V}^T, \\ &where(\mathbf{U},\Sigma,\mathbf{V}^T) = SVD(\mathbf{V}_\mathbf{L}) \end{aligned} \tag{4}$$

A closed form solution has been proven existing in [] Where Σ is the singular value matrix of $\mathbf{V}_L$. The operator $\Gamma_\sigma[\cdot]$ in Eq (4) is a Singular Value Thresholding (SVT) operator [30], which is defined by element-wise σ threshold of Σ, i.e., $diag(\Gamma_\sigma[\Sigma]) = [t_\sigma[\delta_1], t_\sigma[\delta_2], ..., t_\sigma[\delta_m]]$ and $m = rank(\Sigma)$. Each $t_\sigma[\delta]$ is computed as

$$t_\sigma[\delta] = \begin{cases} \delta-\sigma, & if\delta > \sigma, \\ \delta+\sigma, & if\delta < -\sigma, \\ 0, & otherwise. \end{cases} \tag{5}$$

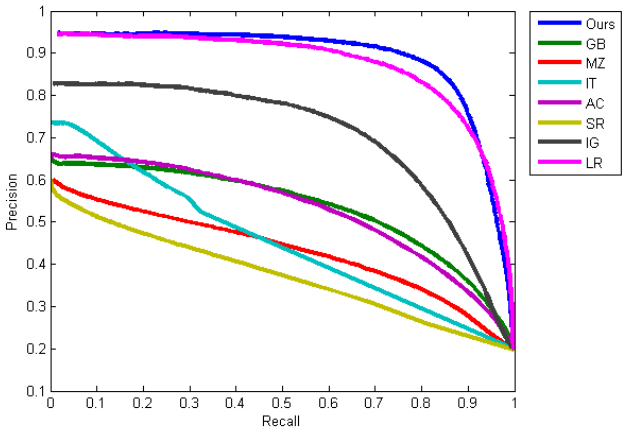

Fig. 3. Precision-recall curves for fixed thresholding of saliency maps.

Computing $\mathbf{S}$: $\mathbf{S}^{k+1}$ is the optimal solution of the following problem.

$$\begin{aligned} \mathbf{S}^{k+1} &= \arg\min_{\mathbf{S}} L(\mathbf{L}^{k+1},\mathbf{S},\mathbf{G}^k,\mu^k) \\ &= \arg\min_{\mathbf{S}} \lambda\sum_{i=1}^{n}\omega_i\|\mathbf{S}_{p_i}\|_{2,1} \\ &+ < \mathbf{G}^k, \mathbf{B}-\mathbf{L}^{k+1}-\mathbf{S} > + \frac{\mu^k}{2}\|\mathbf{B}-\mathbf{L}^{k+1}-\mathbf{S}\|_F^2 \\ &= \arg\min_{\mathbf{S}} \epsilon\sum_{i=1}^{n}\omega_i\|\mathbf{S}_{p_i}\|_{2,1} + \frac{1}{2}\|\mathbf{S}-\mathbf{L}_\mathbf{S}\|_F^2. \end{aligned} \tag{6}$$

in which $\epsilon = \frac{\lambda}{\mu^k}$ and $\mathbf{L}_\mathbf{S} = \mathbf{B} - \mathbf{L}^{k+1} + \frac{1}{\mu^k}\mathbf{G}^k$. According to [31] the closed-form solution of Eq. (6) is

$$\mathbf{S}_{:,i} = sign(\mathbf{L}_{\mathbf{S}:,i})(|\mathbf{L}_{\mathbf{S}:,i}| - \epsilon\omega_i)_+ \tag{7}$$

In Eq. (7) $(\mathbf{A})_+$ is defined as $\mathbf{A}_+ = 0.5(\|\mathbf{A}_{ij}\| + \mathbf{A}_{ij})$.

Algorithm 1 Framework of ALM algorithm to solve the Bi-Subspace Model.

Input:
Image Feature Matrix, $\mathbf{B}$, parameter β and ω_i.
Output:
$\mathbf{L}, \mathbf{S}$;
Intialize $\mathbf{L}^0, \mathbf{S}^0, \mathbf{G}^0, \mu^0 = 0.3, \mu_{max} = 10^7$,and $\beta = 6$;
repeat;
$\mathbf{L}^{k+1} = \arg\min_{\mathbf{L}} L(\mathbf{L},\mathbf{S}^{k+1},\mathbf{G}^k,\mu^k)$
$\mathbf{S}^{k+1} = \arg\min_{\mathbf{S}} L(\mathbf{L}^{k+1},\mathbf{S},\mathbf{G}^k,\mu^k)$
$\mathbf{G}^{k+1} = \mathbf{G}^k + \mu^k(\mathbf{B}-\mathbf{L}^{k+1}-\mathbf{S}^{k+1})$
$\mu^{k+1} = min(\beta\mu^k, \mu_{max})$
$k = k+1$
until converge;
return $\mathbf{L}, \mathbf{S}$;

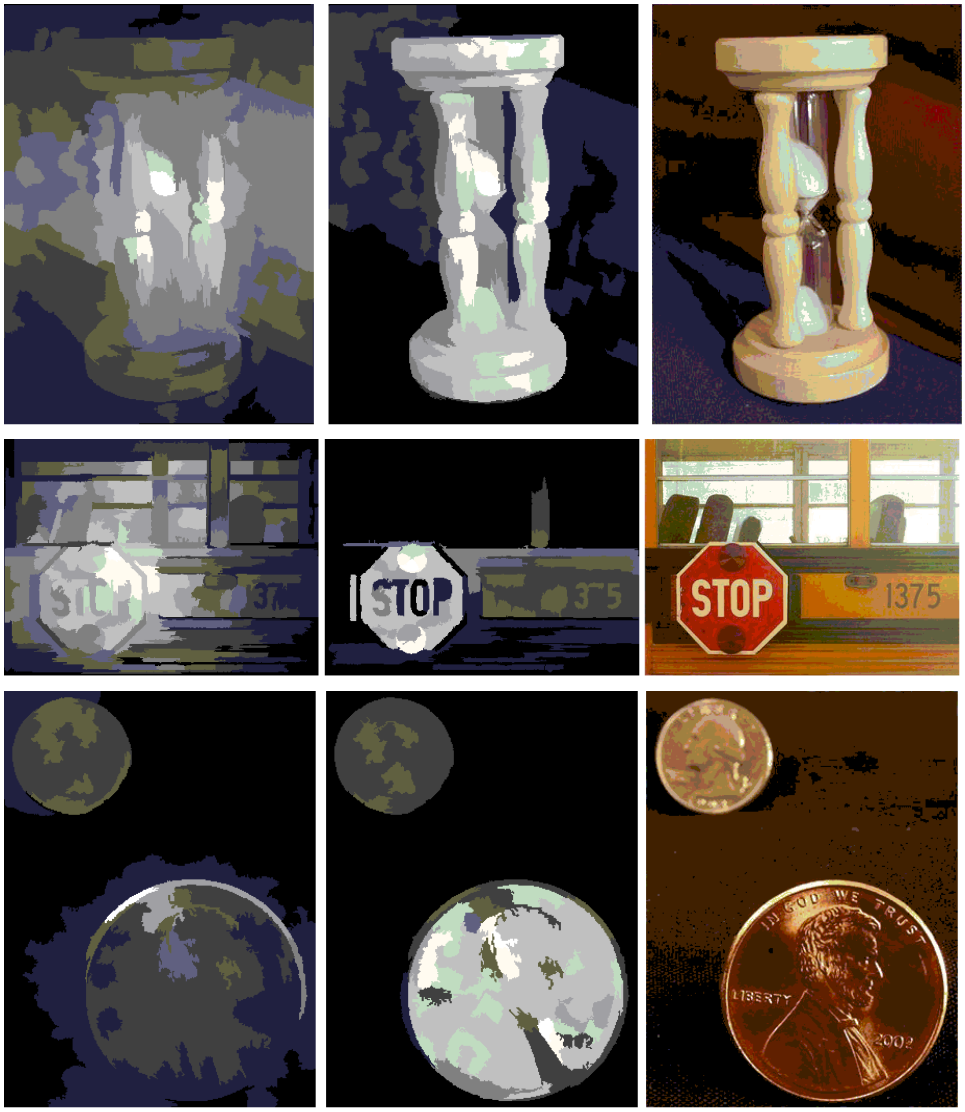

Fig. 4. Examples of saliency maps compared to LR [12] method.

B. Subspace B: Parts in the Image

Although the proposed algorithm in the previous subsection is proper to analysis and model the background subspace, its weakness still exists in practice. In fact, extending the subspace analysis to any part in the image would give benefit more to significant improvement of accuracy of our saliency map prediction. Given a feature data $\mathbf{B} = [b_1, b_2, ..., b_N]$, based on prior knowledge that the pieces in the same part could represent the others within, b_i can be a sparsely linear representation in the redundant dictionary $\mathbf{B}_{i\_}$ ($\mathbf{B}_{i\_}$ is the matrix $\mathbf{B}$ with ith column removed). In another word, b_i is contained in a somehow small-sized subspace of $\mathbf{B}_{i\_}$. Consequently we have

$$\min \|c_i\|_1 + \alpha\|\mathbf{B}_{i\_}c_i - b_i\|_2 \qquad s.t. \quad c_i^T\mathbf{1} = 1. \tag{8}$$

$$\min \|\mathbf{B}_{i\_}c_i - b_i\|_2 + \theta\|c_i\|_1 \qquad s.t. \quad c_i^T\mathbf{1} = 1 \tag{9}$$

where $c_i \in \mathbb{R}^{N-1}$ and θ is a small parameter. Notably in Eq. (8) our second term is an intention to consider the noise in the representation[32].

Consider the dictionary is over-completed, as the case occurs when two data point is very close, then their coefficients of the dictionary c_i may be extremely different. Aim to facilitate our cluster procedure to the data and encode the pieces in the same part in the same small-sized subspace, we approach the method as Zhao et.al. [33] proposed. Our work introduces a Laplacian regularization to intend to enforce the similarity of the sparse coefficients between the similar data

points.

$$\begin{aligned}&\min \|\mathbf{B}_{i\_}c_i - b_i\|_2 + \theta\|c_i\|_1 + \frac{\gamma}{2}\sum_{i,j}\|c_i - c_j\|^2\mathbf{W}_{ij}\\&= \min \|\mathbf{B}_{i\_}c_i - b_i\|_2 + \theta\|c_i\|_1 + \gamma tr(\mathbf{C}\mathbf{L}\mathbf{C}^T)\\&\quad s.t. \quad c_i^T\mathbf{1} = 1.\end{aligned} \tag{10}$$

in which $\mathbf{L} = \mathbf{H} - \mathbf{W}$, $\mathbf{L}$ is a Laplacian matrix, $\mathbf{H} = diag\{\mathbf{H}_{ii}\}_n$, and the diagonal entries are respectively computed as $\mathbf{H}_{ii} = \sum_j \mathbf{W}_{ij}$.

Taking into account of the computational complexity, we specifically select only some most important features to describe each pixel. $\mathbf{B}$ is the average feature values of the pixels in the same super-pixel. As we bring on covariance-matrix between pixels in the same super-pixel, we are capable of describing the relationships between each super-pixel. The covariance matrix $\mathbf{R}$ can be written as

$$\mathbf{R} = \begin{pmatrix} r_{11} & \cdots & r_{1m} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mm} \end{pmatrix} \tag{11}$$

$$r_{ij} = \frac{1}{K-1}\sum_{k=1}^{K}(f_i^k - b_i)(f_j^k - b_j) \tag{12}$$

where in the same super-pixel f_i^k is the ith feature of the kth pixel(K pixels of one super-pixel in all) and b_i is the mean of the ith feature. Note that we have used this quantity to depict the the super-pixel.

Here, we choose

$$d(\mathbf{R}_1, \mathbf{R}_2) = \sqrt{\sum_{i=1}^{m} \ln^2 \lambda_i(\mathbf{R}_1, \mathbf{R}_2)} \tag{13}$$

where $\ln^2 \lambda_i(\mathbf{R}_1, \mathbf{R}_2)$ We propose to employ this distance [] is to compute each element of the matrix $\mathbf{W}$, which shows the difference between two super-pixels by:

$$\mathbf{W}(\mathbf{R}_1, \mathbf{R}_2) = \exp(-\xi d(\mathbf{R}_1, \mathbf{R}_2)) \tag{14}$$

in which ξ is a small parameter we select.

After computing the constraint matrix $\mathbf{W}$ we can directly obtain the Laplacian matrix $\mathbf{L}$ from with the feature matrix $\mathbf{B}$ to optimize Eq. (10). The approach is similar to the method in [34]. We obtain a matrix of sparse-coding coefficients $\mathbf{C} = [c_1, c_2, ..., c_N] \in \mathbb{R}^{N\times N}$. Next we construct a symmetric similarity matrix $\mathbf{C}^L = |\mathbf{C} + \mathbf{C}^T|$. Suppose $\mathbf{C}^L$ is the adjacency matrix of a graph G. The Laplacian matrix $\mathbf{P}$ of the graph G can be written as

$$\begin{aligned}\mathbf{P} &= \mathbf{Q} - \mathbf{C}^L\\ \mathbf{Q}_{ii} &= \sum_j \mathbf{C}_{ij}^L.\end{aligned} \tag{15}$$

Finally, we obtain the parts of an image after applying the K-means algorithm to cluster the eigenvector of the Laplacian matrix $\mathbf{P}$.

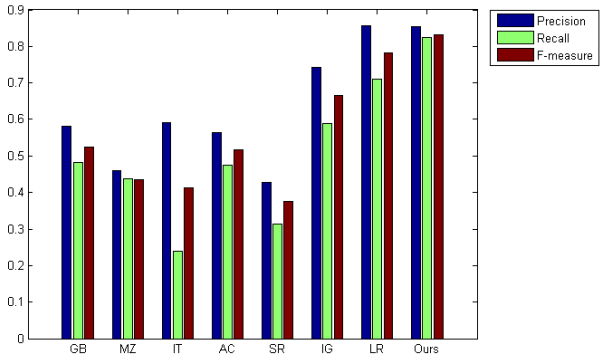

Fig. 5. precision, recall and F-measure with adaptive thresholding on the 1000-image database. Our method performance best on precision, recall and F-measure.

C. Advanced Group Sparsity about Visual Saliency

As our priors show before, The salient object in an image must be sparse [12] and it could inlay in a convex set in the image. Now we prefer to seek for a convex set contain the salient region, we use both the method Weijer et.al. [28] proposed and the color boosted Harris point operator [27] to detect contour points or corners of salient sections. In practice we conduct the equivalent number of interesting points respectively by the first and the second method. The convex set is immediately formed by finding the minimal convex set whose verticals are parts of the interesting point. As we show in Fig. 2, due to the convex region surrounding every interesting point with highly likelihood to be salient, we can always find a bigger convex set to contain all of these regions. Therefore, if any part in the image belong to the convex set partly, it tends to be believed as the sparse salient group of super-pixels and shares the same saliency within the group.

D. The main algorithm

A. **Feature Extraction**. At the first place we abstract the image features including Gabor filter [35], RGB color, CIELab color, first and second order derivatives and steerable pyramids [36] to construct a 60-dimensional feature space. Then, we approach the mean-shift clustering [37] in the feature space to fast over-segment the image into N super-pixels $b_i, i = 1, 2, ..., N$. Then we piece all the feature vector together into a matrix $\mathbf{B} = [b_1; b_2; ...; b_N] \in \mathbb{R}^{D\times N}(D = 60)$.

B. **Salient Parts Estimation**. Following the previous section, this step is firstly to construct a convex set J coarsely containing a majority of salient region, then let

$$\omega_i = 1 - \frac{\|E_i \cap J\|_0}{\|E_i\|_0} \tag{16}$$

in Eq. (16) E_i is the ith cluster, and we roughly define the operator $\|\cdot\|_0$ as the number of pixels in the set. Note that if E_i does not overlap with J, we just let $\omega_i = 0$. It is mentionable that because of faster computing during searching the parts in the image our extraction from super-pixels confines to be

Fig. 6. Visual comparison of saliency maps. We compare our method (Ours) to several state-of-the-art methods.

low-dimensional. This paper selects 7 features (e.g. CIELab, derivatives in each direction) to save the cost of computation.

C. **RPCA Recovery**. This procedure builds on Eq. (4) to recover $\mathbf{L}$ and $\mathbf{S}$ from $\mathbf{B}$ by solving an ALM and group-sparse problem. At the end of the recovery, we compute saliency value for each super-pixel sp_i in a simple way like:

$$Sal(sp_i) = \|\mathbf{S}_i\| \tag{17}$$

After normalization and Gaussian filtering on pixel-level as $Map(x, y)$, where (x, y) is the coordinate of a pixel.

V. Results

In this section, we validate the Bi-Subspace model in Eq. (4) by comparing to the other state-of-the-art models in two evaluation measures. All tests were conducted in Matlab and experiments were performed on an Intel Core i5 2.5GHz and 4GB RAM computer. In the following we will estimate our approach on the database named as MSRA 1000 which is a image sub-database provided by Achanta et.al. [21] together with accurate human-marked labels as ground truth. Numerical comparisons of our method to the classic Low-Rank method and 6 efficient algorithms will be demonstrated as detailed.

Our comparison measures are similar with Achanta et.al. [21] to evaluate the accuracy of saliency detection. The first evaluation fixes the threshold T from 0 to 255 to the saliency values that aims at segmenting an image. The pixels whose saliency values are lower than T are background stuff and the rest of the pixels belong to the salient foreground. Differentiating T from 0 to 255 results with the precision-recall pairs, and a precision-recall figure is shown in Fig. 3. Note that the curve is generated by averaging the results of 1000 test images. The other evaluation is to give an adaptive threshold to the test image in segmentation operation. Above all, calculate the average saliency for each super-pixel and the average saliency over the entire image. Then whenever the saliency of a super-pixel is larger than twice of the mean saliency value, the super-pixel can be viewed as salient. At the final step we compute the precision and recall values and F-measure. where

$$F_\phi = \frac{(1+\phi^2)P \times R}{\phi^2 P \times R}(P = Precision, R = Recall) \tag{18}$$

In Eq.(18) $\phi^2 = 0.32$ is set in the experiment which is the same as in [11], [30].

We make a comparison among our method and other state-of-the-art algorithms besides IT [5], GB [18], AC [21], RA [38], and GC [10] and LR [12]. Softwares or results provided from their paper are utilized for our comparison. As presented in Fig. 3, while shifting the threshold from 0 to 255, the precision-recall curves of all the measures are formed along with ours are presented in Fig. 3, and adopting the adaptive threshold we plot the precision, recall and F-measure which are shown in Fig. 4.

Obviously we could observe that the precision-recall curve of our approach are better than any other approaches and achieves the state-of-the-art. Likewise, adopting the adaptive threshold gives the best performance in Fig. 4. The development of our performance could be visually comprehended in Fig. 5. While the Low-rank approach blinds itself when the structure in the scenery is beyond the rough presumption such like the color assumption or the center assumption proposed in [12], our method grasp the parts of the salient object without

mixing the background interference. The relevant results are shown in Fig. 6.

VI. Conclusion

In this paper, We propose A novel bi-subspace data-driven saliency detection model. Motivated from relatively visually intuitional and bi-subspace assumptions, we consider the problem from subspace analysis to characterize the background and foreground. In the meanwhile, we also utilize the subspace description by group-sparsity for the structure of the object in our model. Just like [12], our model can be equipped with more high-level prior knowledge (e.g. face and car recognition etc.). Numerical experiment results show that our method is more stable and accurate to other methods. For future work, we plan to extend our work from classical RPCA framework to tensor RPCA, by which we may find a more explicit and essential properties of visual saliency detection work.

References

[1] N. Pattyn, X. Neyt, D. Henderickx, and E. Soetens, "Psychophysiological investigation of vigilance decrement: boredom or cognitive fatigue?" *Physiology & Behavior*, vol. 93, no. 1, pp. 369–378, 2008.

[2] R. H. Wurtz and M. E. Goldberg, "The primate superior colliculus and the shift of visual attention," *Invest Ophthalmol*, vol. 11, no. 6, pp. 441–450, 1972.

[3] A. M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive psychology*, vol. 12, no. 1, pp. 97–136, 1980.

[4] N. Lavie, A. Hirst, J. W. De Fockert, and E. Viding, "Load theory of selective attention and cognitive control." *Journal of Experimental Psychology: General*, vol. 133, no. 3, p. 339, 2004.

[5] L. Itti, C. Gold, and C. Koch, "Visual attention and target detection in cluttered natural scenes," *Optical Engineering*, vol. 40, no. 9, pp. 1784–1793, 2001.

[6] V. Navalpakkam and L. Itti, "An integrated model of top-down and bottom-up attention for optimizing detection speed," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2. IEEE, 2006, pp. 2049–2056.

[7] L. Marchesotti, C. Cifarelli, and G. Csurka, "A framework for visual saliency detection with applications to image thumbnailing," in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 2232–2239.

[8] N. Jacobson, Y.-L. Lee, V. Mahadevan, N. Vasconcelos, and T. Q. Nguyen, "A novel approach to fruc using discriminant saliency and frame segmentation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2924–2934, 2010.

[9] T. Judd, K. Ehinger, F. Durand, and A. Torralba, "Learning to predict where humans look," in *2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009, pp. 2106–2113.

[10] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015.

[11] A. Borji, M. N. Ahmadabadi, and B. N. Araabi, "Cost-sensitive learning of top-down modulation for attentional control," *Machine Vision and Applications*, vol. 22, no. 1, pp. 61–76, 2011.

[12] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 853–860.

[13] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, p. 11, 2011.

[14] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," in *Matters of intelligence*. Springer, 1987, pp. 115–141.

[15] L. Itti, C. Koch, E. Niebur *et al.*, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[16] R. Desimone and J. Duncan, "Neural mechanisms of selective visual attention," *Annual review of neuroscience*, vol. 18, no. 1, pp. 193–222, 1995.

[17] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proceedings of the eleventh ACM international conference on Multimedia*. ACM, 2003, pp. 374–381.

[18] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in neural information processing systems*, 2006, pp. 545–552.

[19] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proceedings of the 14th ACM international conference on Multimedia*. ACM, 2006, pp. 815–824.

[20] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–8.

[21] P. W. R. Achanta, F. Estrada, "Salient region detection and segmentation." *In ICVS*, 2008.

[22] G. Li and Y. Yu, "Visual saliency based on multiscale deep features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5455–5463.

[23] L. Wang, H. Lu, X. Ruan, and M.-H. Yang, "Deep networks for saliency detection via local estimation and global search," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3183–3192.

[24] R. Zhao, W. Ouyang, H. Li, and X. Wang, "Saliency detection by multi-context deep learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1265–1274.

[25] C. Lang, G. Liu, J. Yu, and S. Yan, "Saliency detection by multitask sparsity pursuit," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1327–1338, 2012.

[26] B. Li, W. Xiong, W. Hu *et al.*, "Visual saliency map from tensor analysis." in *AAAI*, 2012.

[27] C. Harris and M. Stephens, "A combined corner and edge detector." in *Alvey vision conference*, vol. 15. Citeseer, 1988, p. 50.

[28] J. Van de Weijer, T. Gevers, and A. D. Bagdanov, "Boosting color saliency in image feature detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 1, pp. 150–156, 2006.

[29] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.

[30] C.-F. Chen, C.-P. Wei, and Y.-C. F. Wang, "Low-rank matrix recovery with structural incoherence for robust face recognition," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2618–2625.

[31] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[32] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 2790–2797.

[33] S. Gao, I. W.-H. Tsang, L.-T. Chia, and P. Zhao, "Local features are not lonely–laplacian sparse coding for image classification," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 3555–3561.

[34] W. Förstner and B. Moonen, "A metric for covariance matrices," in *Geodesy-The Challenge of the 3rd Millennium*. Springer, 2003, pp. 299–309.

[35] H. G. Feichtinger and T. Strohmer, *Gabor analysis and algorithms: Theory and applications*. Springer Science & Business Media, 2012.

[36] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: a flexible architecture for multi-scale derivative computation." in *ICIP (3)*, 1995, pp. 444–447.

[37] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[38] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä, "Segmenting salient objects from images and videos," in *European Conference on Computer Vision*. Springer, 2010, pp. 366–379.