



# Segmentation Driven Low-rank Matrix Recovery for Saliency Detection

Wenbin Zou, Kidiyo Kpalma, Zhi Liu, Joseph Ronsin

## ► To cite this version:

Wenbin Zou, Kidiyo Kpalma, Zhi Liu, Joseph Ronsin. Segmentation Driven Low-rank Matrix Recovery for Saliency Detection. 24th British Machine Vision Conference (BMVC), Sep 2013, Bristol, United Kingdom. pp.1-13, 2013. <hal-00853385>

**HAL Id: hal-00853385**

**<https://hal.archives-ouvertes.fr/hal-00853385>**

Submitted on 22 Aug 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Segmentation Driven Low-rank Matrix Recovery for Saliency Detection

Wenbin Zou<sup>1</sup>

[zouwb2008@gmail.com](mailto:zouwb2008@gmail.com)

Kidiyo Kpalma<sup>1</sup>

[kidiyo.kpalma@insa-rennes.fr](mailto:kidiyo.kpalma@insa-rennes.fr)

Zhi Liu<sup>2</sup>

[liuzhisjtu@163.com](mailto:liuzhisjtu@163.com)

Joseph Ronsin<sup>1</sup>

[joseph.ronsin@insa-rennes.fr](mailto:joseph.ronsin@insa-rennes.fr)

<sup>1</sup> IETR / UMR CNRS 6164,

INSA Rennes,

Université Européenne de Bretagne  
Rennes, France

<sup>2</sup> School of Communication and

Information Engineering,

Shanghai University  
Shanghai, P.R. China

---

## Abstract

Low-rank matrix recovery (LRMR) model, aiming at decomposing a matrix into a low-rank matrix and a sparse one, has shown the potential to address the problem of saliency detection, where the decomposed low-rank matrix naturally corresponds to the background, and the sparse one captures salient objects. This is under the assumption that the background is consistent and objects are obviously distinctive. Unfortunately, in real images, the background may be cluttered and may have low contrast with objects. Thus directly applying the LRMR model to the saliency detection has limited robustness. This paper proposes a novel approach that exploits bottom-up segmentation as a guidance cue of the matrix recovery. This method is fully unsupervised, yet obtains higher performance than the supervised LRMR model. A new challenging dataset PASCAL-1500 is also introduced to validate the saliency detection performance. Extensive evaluations on the widely used MSRA-1000 dataset and also on the new PASCAL-1500 dataset demonstrate that the proposed saliency model outperforms the state-of-the-art models.

## 1 Introduction

Visual saliency characterizes what captures the perceptual attention of human. Saliency detection has recently become an active research topic. This mainly lies in its promising applications, e.g., image classification [17, 30], object recognition [28, 29], object segmentation [16, 23], object detection [2], picture collage [13, 35], image summarization [14, 32], etc. Existing saliency models can be broadly classified into two categories: biological models and computational ones.

The biological model is pioneered by Koch and Ullman [22] who derived visual saliency from a set of topographical maps of elementary features like orientation of edges, color and luminance. The biological model is usually implemented using the center-surround scheme with different formulations on a set of features [12, 20, 34]. As the objective of biological models is to find some points that mostly catch human attention, the resulting saliency maps

are typically sparse and blurry, and limit their applications mainly for prediction of eye fixations.

Instead, the computational models, inspired by the biological models, aim at discovering objects standing out from surrounding regions. A number of computational models measure the saliency based on global, local and regional contrasts with different forms [1, 8, 14, 21], and a variety of theories and methods, including information theory [6, 37], graph theory [4, 18], machine learning [26], statistical model [27], frequency domain analysis [15, 19], have been exploited to build saliency models. These models may work well for objects within consistent scenes, thus, most of them are validated on relatively simple dataset like MSRA-1000 [1]. However, they still lack robustness to detect objects in complex images with cluttered background and/or low contrast between objects and background. Recently, a new trend is to formulate the problem of saliency detection with low-rank matrix recovery (LRMR) model, in which an image is decomposed into a low-rank matrix which corresponds to the background, and a sparse one which links to salient objects. In [36], sparse coding is used as an intermediate representation of image features and then fits to LRMR model to recover salient objects. As pointed out in [31], the sparse coding can not guarantee that, in the entire image representation, the sparse codes of salient objects are sparse and those of the background are of low-rank. Therefore, Shen et Wu [31] proposed to modulate the image features with learnt transform matrix and high-level priors to meet the low-rank and sparse properties. This sounds reasonable and remarkable experiment results have been demonstrated. Unfortunately, training is required and the learnt transform matrix is somewhat biased toward the training dataset, therefore it suffers from limited adaptability.

Based on the aforementioned issues, this study aims at an unsupervised LRMR model for saliency detection. The key idea is to derive a bottom-up prior to constrain image features so that they can fit well to the LRMR model. For this purpose, we propose a novel generic prior named as *segmentation prior* which is created from bottom-up segmentation. The segmentation prior softly separates objects from background so that the objects are highlighted and the background is highly redundant in the feature domain. Traditionally, the saliency is used as a guidance cue for the application of object segmentation, however, the proposed approach proceeds in the opposite direction that investigates benefits of bottom-up segmentation for saliency detection.

For evaluation, a new dataset PASCAL-1500, created from PASCAL VOC 2012 segmentation challenge [10], is also introduced to validate saliency detection performance. Most previous saliency models are evaluated on relatively simple dataset, such as the widely used MSRA-1000 [1], in which most images contain only a single object, typically located in the center of the image, and the appearance contrast between object and background is generally obvious. In contrast, the PASCAL-1500 dataset is much more challenging: images contain multiple objects appearing at a variety of locations and scales, and background regions are much more cluttered. Extensive evaluation has been carried out on both MSRA-1000 dataset and PASCAL-1500 dataset. Experiments show that the proposed unsupervised LRMR model outperforms the state-of-the-art models, even the supervised one [31] which is also based on the LRMR model.

## 2 Overview

Figure 1 presents the framework of the proposed segmentation driven low-rank matrix recovery model. An input image is firstly segmented into three-level segmentations: *fine-grained*

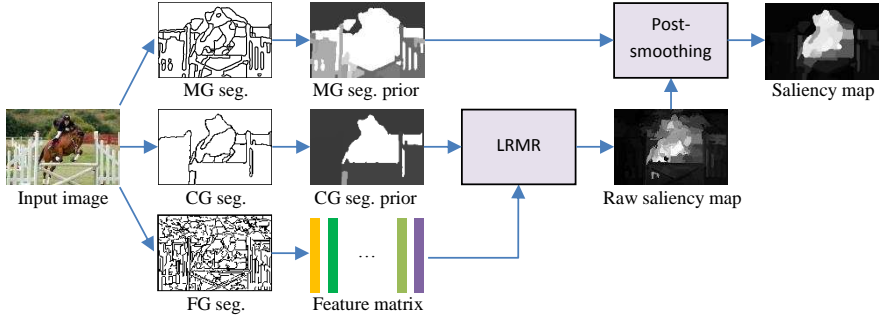


Figure 1: Framework of the proposed saliency model. Input image is firstly segmented into three levels. Feature descriptors are accumulated within superpixels of fine-grained (FG) segmentation. Segmentation priors are derived from the medium-grained (MG) segmentation and coarse-grained (CG) segmentation, respectively. The final saliency map is obtained by smoothing the raw saliency map generated by LRMR model with the MG segmentation prior.

(FG), *medium-grained* (MG) and *coarse-grained* (CG). The FG segmentation significantly over-segments the image into a number of superpixels (to avoid confusion, the segments of FG segmentation are called "superpixels" rather than "regions" used in MG and CG segmentations). The MG segmentation also over-segments the image but generates regions as few as possible. The CG segmentation aims at maximally separating objects from the background, thus the image may be over-segmented or under-segmented. Based on these three-level segmentations, image features are extracted from the superpixels, and segmentation priors are derived from the MG and CG segmentations. Then, the low-rank matrix recovery (LRMR) model is applied to generate the raw saliency map. Finally, the raw saliency map is smoothed by using the MG segmentation prior to generate an optimal saliency map.

### 3 Low-rank matrix recovery model

Given an input image  $\mathbf{I}$ , let  $\mathbf{P} = \{p_1, p_2, \dots, p_N\}$  be a set of  $N$  superpixels created by FG segmentation, and  $\mathbf{a}_n \in \mathbb{R}^{d \times 1}$  be the feature vector of the superpixel  $p_n$ , where  $d$  is the dimension of feature descriptor. The image  $\mathbf{I}$  is represented by a feature matrix  $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N] \in \mathbb{R}^{d \times N}$ .

In real images, the background pixels generally show similar appearance, and have strong correlations in the feature space. This suggests that the feature matrix  $\mathbf{A}$  might have low-rank property, and it can be decomposed into two parts, a low-rank matrix  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N] \in \mathbb{R}^{d \times N}$ , and a sparse one  $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N] \in \mathbb{R}^{d \times N}$

$$\mathbf{A} = \mathbf{U} + \mathbf{E} \quad (1)$$

Applying this model to saliency detection, the background is naturally represented by the low-rank matrix  $\mathbf{U}$ , and the objects might be captured by the sparse matrix  $\mathbf{E}$ .

To recover the matrix  $\mathbf{U}$  and  $\mathbf{E}$ , the problem can be formulated with the Lagrangian

representation

$$\begin{aligned} \min \quad & \text{rank}(\mathbf{U}) + \lambda \|\mathbf{E}\|_0 \\ \text{s.t.} \quad & \mathbf{A} = \mathbf{U} + \mathbf{E} \end{aligned} \quad (2)$$

where  $\lambda$  is a coefficient to balance  $\mathbf{U}$  and  $\mathbf{E}$ , and  $\|\cdot\|_0$  indicates  $l_0$ -norm. Unfortunately, this is a NP-hard problem as the matrix rank and  $l_0$ -norm are not convex. Recent theoretic analysis in [7] shows that, under rather weak assumptions, the low-rank matrix  $\mathbf{U}$  and the sparse matrix  $\mathbf{E}$  can be exactly recovered by

$$\begin{aligned} \min \quad & \|\mathbf{U}\|_* + \lambda \|\mathbf{E}\|_1 \\ \text{s.t.} \quad & \mathbf{A} = \mathbf{U} + \mathbf{E} \end{aligned} \quad (3)$$

where  $\|\cdot\|_*$  is the nuclear norm of matrix  $\mathbf{U}$  (the sum of singular values of  $\mathbf{U}$ ), and  $\|\cdot\|_1$  indicates  $l_1$ -norm. The regularization of  $l_1$ -norm ensures to produce a sparse matrix  $\mathbf{E}$ . The optimal matrices  $\mathbf{U}$  and  $\mathbf{E}$  can be obtained by alternatively minimizing (3) over one while keeping the other one fixed.

With the optimal sparse matrix  $\mathbf{E}$ , the saliency value of superpixel  $p_n$  is given by the  $l_1$  energy of the vector  $\mathbf{e}_n$

$$s_n = \sum_{i=1}^d |\mathbf{e}_n(i)| \quad (4)$$

The saliency value  $s_n$  represents the probability of superpixel  $p_n$  belonging to an object, i.e., a larger value indicates a higher probability, and vice versa. The saliency map of image  $\mathbf{I}$  is then generated by assigning the saliency value of each superpixel to all pixels in the superpixel.

## 4 LRMR with segmentation prior

Directly fitting the LRMR model to the problem of saliency detection is under the assumption that background is homogeneous and has high contrast with objects. In reality, however, many backgrounds are cluttered and objects may be similar to part of the background regions. This results in false positive detection results. To improve the robustness of saliency detection, a feasible method is to adopt high-level priors to modulate input features [24, 31], so that the feature matrix has a lower rank. The underlying idea of the modulation is to give small weights to feature vectors of those superpixels which are more likely to be background and large weights to those corresponding to objects.

Many priors have been proposed for saliency detection, such as center prior, color prior and learnt transform prior [31]. The main drawback of these priors is the lack of adaptability, since they are either obtained by empirical statistics or trained from the annotated images. For example, center prior assumes that objects always appear in the center of image, and color prior believes that the objects are in warm colors. Obviously, these assumptions are not always valid in practice. In addition, the learnt transform prior tends to fail when the test image has a high difference with the training images.

Here we introduce a bottom-up segmentation driven prior, named as *segmentation prior*. Firstly, let us take a look at the images and their CG segmentations in Figure 2. Salient objects locate at diversity of positions: center, bottom, left, right and corner. Both background and objects are typically segmented into several regions, thus, the bottom-up segmentation

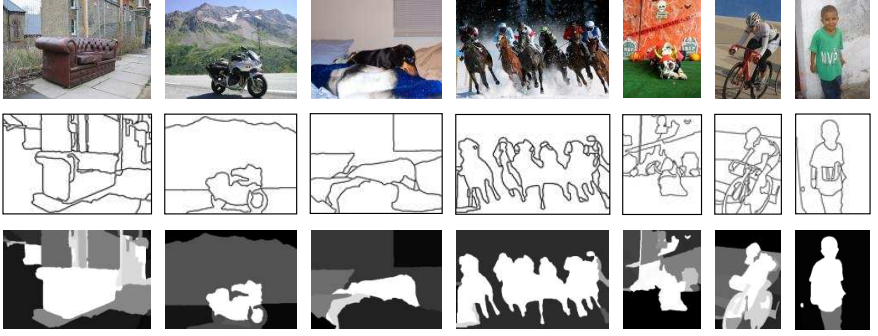


Figure 2: Examples of segmentation prior. First row: input images; second row: bottom-up segmentation results; last row: segmentation prior where white indicates a higher weight and black represent a lower weight.

can not be expected to totally separate objects from background. However, the segmented regions of background have very high probability of connecting with the border of the image, while very few object regions link to it. Even if an object is truncated on the border, like the bike and the child of the two right-most images, border regions of object are small compared to the whole object in the image. In contrast, the border regions of the background are usually large, as background appears more uniform, like sky, road, tree, wall, etc. This observation implies that objects can be roughly separated from the background by the bottom-up segmentation. Therefore, we propose the segmentation prior according to the connectivity between each region and image border. Let  $r_m$  be a segmented region of image  $\mathbf{I}$ , the segmentation prior of region  $r_m$  is defined as

$$h_m = \exp\left(-\frac{\|r_m \cap C\|}{\sigma \psi_m}\right) \quad (5)$$

where  $\|\cdot\|$  denotes the length of intersection,  $C$  is the border of image  $\mathbf{I}$ ,  $\psi_m$  is the outer perimeter of region  $r_m$ , and  $\sigma$  is a balance parameter which is set to 0.3 in our experiments. Clearly, if a region touches the image border, its prior value is in the range of  $(0, 1)$ , otherwise it is equal to 1. In other words, the segmentation prior gives a small weight to the region touching the image border. By using (5), segmentation priors of all regions can be computed, and form the prior of the input image.

In Figure 2, one might observe that, on one hand, there are still some regions of the background without connection with the image border, on the other hand, some regions of objects are inevitably merged with the background. Indeed, such a strategy can not perfectly separate the objects from the background. However, the segmentation prior derived from CG segmentation can serve as a guidance cue for LRMR model to address the task of saliency detection.

Suppose an input image  $\mathbf{I}$  is segmented into  $N$  superpixels by FG segmentation, and represented by a feature matrix  $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N]$ . Let  $\mathbf{H}_c = [h_1^c, h_2^c, \dots, h_N^c]$  denote a set of CG segmentation prior values of the superpixels. In order to recover salient objects well with the LRMR model, the feature matrix  $\mathbf{A}$  is firstly modulated by the CG segmentation prior  $\mathbf{H}_c$

$$\mathbf{B} = [h_1^c \mathbf{a}_1, h_2^c \mathbf{a}_2, \dots, h_N^c \mathbf{a}_N]. \quad (6)$$

Then, the modulated feature matrix  $\mathbf{B}$  is used as the input of the standard LRMR model

$$\begin{aligned} \min \quad & \|\mathbf{U}\|_* + \lambda \|\mathbf{E}\|_1 \\ \text{s.t.} \quad & \mathbf{B} = \mathbf{U} + \mathbf{E} \end{aligned} \quad (7)$$

As the segmentation prior assigns small weights to most of background feature vectors in  $\mathbf{B}$ , the  $l_1$  energies of the corresponding vectors in the recovered matrix  $\mathbf{E}$  are inclined to be small. Therefore, objects are highlighted more effectively in the matrix  $\mathbf{E}$ .

## 5 Post-smoothing

Raw saliency map generated by the LRMR model might still contain some noises: some large saliency values in background area or small values in objects. There are mainly two reasons for this phenomenon: on one hand, some superpixels of background might be strongly similar to those of objects in the feature domain; on the other hand, the LRMR model decomposes the feature matrix without considering spatial constraint. To remove the noises, the raw saliency map is smoothed at two scales: FG and MG levels.

Let  $\mathbf{S} = \{s_1, s_2, \dots, s_N\}$  denote the saliency values of all superpixels  $\mathbf{P} = \{p_1, p_2, \dots, p_N\}$  in the image. At the FG level, the saliency of each superpixel  $p_n$  is penalized by its adjacent superpixels

$$s'_n = s_n + \alpha \sum_{j \in \mathcal{N}} s_j \cdot \exp(-\|\mathbf{a}_n - \mathbf{a}_j\|_2^2) \quad (8)$$

where  $\mathcal{N}$  is a set of adjacent neighbours of superpixel  $p_n$ ,  $\|\cdot\|_2$  denotes  $l_2$ -norm, and  $\alpha$  is the weight to balance the impact of neighbours on the current superpixel, which is set to 0.5 in our experiments. Obviously, neighbours with appearance more similar to the current superpixel are considered to give more contributions to compute the saliency, and vice versa. Therefore, the FG level smoothing ensures the saliency coherent with its neighbours belonging to the same category (object or background).

The FG level smoothing might be still far from labelling saliency at object level. We also perform a MG level processing on the FG smoothed saliency map  $\mathbf{S}'$ . To do this, segmentation prior of the MG segmentation is also computed.

Let  $\mathbf{S}' = \{s'_1, s'_2, \dots, s'_K\}$ ,  $\mathbf{H}_m = \{h_1^m, h_2^m, \dots, h_K^m\}$  and  $\mathbf{R} = \{r_1, r_2, \dots, r_L\}$  denote the FG smoothed saliency map, MG segmentation prior and MG segmentation of the input image respectively, where  $K$  is the number of pixels,  $L$  is the number of regions. The saliency value of region  $r_l$  is computed by

$$\bar{s}_l = \frac{1}{T_l} \sum_{k \in r_l} s_k'' \quad (9)$$

where  $T_l$  is the number of pixels in the region  $r_l$ ,  $s_k''$  is the weighted saliency value of pixel  $k$

$$s_k'' = h_k^m s'_k. \quad (10)$$

Therefore, the final saliency map of image is obtained by distributing saliency values of all regions into corresponding pixels. Notice that, the parameters of MG segmentation are set to ensure an image is over-segmented as few regions as possible. Thus, this process generates more smooth saliency map than assigning saliency values based on superpixels; in addition, object contours are also preserved well.

## 6 Experiments

The proposed saliency model is evaluated on two datasets: MSRA-1000 [1] and PASCAL-1500. The MSRA-1000 is one of the most commonly used datasets for evaluation of saliency detection. It contains 1000 natural images with pixel-wise manually labelled ground-truth. This is a relatively simple dataset, as most images are with high appearance contrast between object and background which is typically consistent; in addition, most images only contain a single object in the center.

In order to objectively validate the saliency detection, we introduce a more challenging dataset, PASCAL-1500<sup>1</sup>. This dataset contains 1500 real-world images from PASCAL VOC 2012 segmentation challenging [10]. Since not all images of PASCAL VOC are suitable for the evaluation of saliency detection, only images intuitively deemed to have some salient objects are selected. In the PASCAL-1500, many images have highly cluttered background; about 40% images contain multiple objects (on average 3 objects); objects appear at a variety of locations and scales. To create binary ground truths for evaluating saliency detection performance, pixel-wise annotated segmentation ground truths in PASCAL VOC are modified by labelling object pixels as "1" and other pixels as "0".

In the rest of subsections, we give the relevant implementation details and performance evaluation metrics, and then we discuss the results obtained on both datasets.

### 6.1 Implementations

For image description, we use (i) RGB color components along with the saturation and hue components, (ii) 36 Gabor filters [11] with 12 orientations and 3 scales, (iii) 12 steerable pyramid filters [33] with 4 orientations and 3 scales. These three types of features are accumulated within the superpixel and stacked together to form a 53-dimensional feature vector to represent the superpixel. This image description is exactly the same as that used in the training based low-rank matrix recovery (TLR) model [31]. However, the TLR model is supervised, which needs to learn a transformation matrix  $\mathbf{T}$  integrating to the LRMR model. In contrast, our method is fully unsupervised.

MG and CG segmentations are generated by hierarchical segmentation of gPb [3] (globalized probability of boundary). The output of gPb is a valued ultrametric contour map (UCM). The MG and CG segmentations are created by thresholding the UCM, which is normalized from 0 to 1, at 0.125 and 0.25 respectively. The gPb generally preserves global contours of objects, and it fits well to the MG and CG segmentations. However, it can not apply to the FG segmentation very well, as it tends to group uniform areas into a large region. This makes feature descriptors extracted from superpixels of background to be insufficiently redundant, and thus they lack low-rank property which is essential for LRMR model. Therefore, segment size controllable Mean-shift [9] is used to obtain FG segmentation, where the minimum segment area is set to 200 pixels.

The low-rank matrix recovery model is solved by the augmented Lagrange multiplier method proposed in [25]. The balance parameter  $\lambda$  of the model is set to 0.05.

---

<sup>1</sup><http://wzou.perso.insa-rennes.fr>



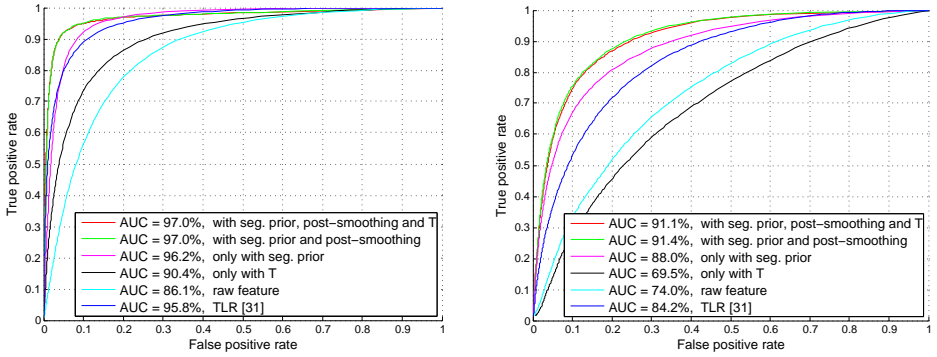


Figure 3: ROC curves and AUC scores with different components on MSRA-1000 (left) and PASCAL-1500 (right) datasets. **T** is the learnt transformation prior proposed in TLR model [31]. Performance of the TLR model is also presented.

## 6.2 Evaluation Metrics

To objectively evaluate the performance of saliency detection, we adopt the widely used metrics: receiver operator characteristic (ROC) curve to measure the similarity between the saliency map and the ground-truth, and the area under the curve (AUC) for quantitative comparison between different models. To obtain the ROC curve, saliency maps are normalized from 0 to 255 and thresholded using integer values within  $[0, 255]$ . Then for each thresholding, the true positive rate and the false positive rate over all test images are computed. Finally, The ROC curve is generated by plotting the true positive rate values on the y-axis against false positive rate values on the x-axis.

## 6.3 Baselines

As baselines, we use eight recently proposed saliency models, of which most obtained state-of-the-art performance on several datasets. The baselines include training based low-rank matrix recovery (TLR) model [31], region contrast based model (RC) [8], kernel density estimation based (KDE) model [27], context and shape prior based (CBS) model [21], context-aware (CA) model [14], frequency-tuned (FT) model [1], graph based (GB) model [18] and spectral residual (SR) model [19]. For evaluation of other recent models, one can refer to the benchmarking report [5].

## 6.4 Results and discussion

Firstly, we validate the contributions of the CG segmentation prior and post-smoothing module of our model. Figure 3 summarizes the evaluation. For comparison, the performance of TLR model [31], which is the most relative to our model, is also presented in Figure 3. In the TLR model, transformation prior learnt from MSRA-B dataset containing 5000 annotated images, semantic prior (face detection), color prior and center prior are integrated to LRMR model for saliency detection. In contrast, our model only uses bottom-up segmentation priors, without using any supervisory information.

As shown in Figure 3, if raw feature extracted from image is directly used as the input of LRMR model, the saliency detection performance is low. With the CG segmentation prior,

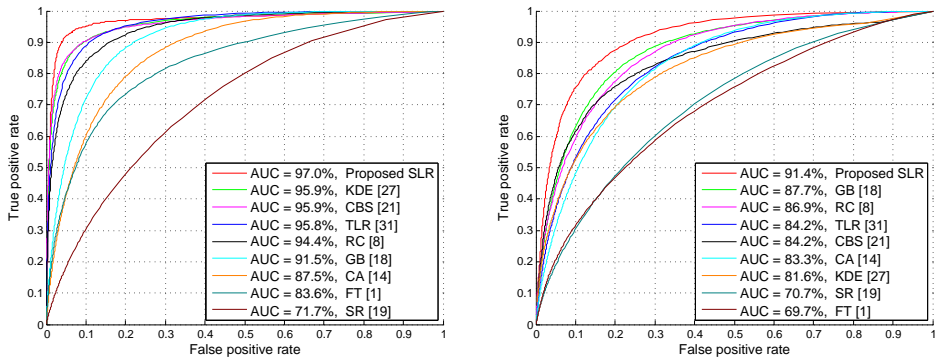


Figure 4: ROC curves and AUC scores of different models on MSRA-1000 (left) and PASCAL-1500 (right) datasets. The models are ranked by AUC scores in the legend.

the AUC scores increase significantly: from 86.1% to 96.2% on MSRA-1000, and from 74.0% to 88.0% on PASCAL-1500. By integrating the post-smoothing component, the AUC score on MSRA-1000 increases with only 0.8%; however, on PASCAL-1500, it increases with 3.1%. This is mainly because the LRMR model with the CG segmentation prior has obtained pretty good results on the relatively simple MSRA-1000, and there is no more room for improvement. In contrast, PASCAL-1500 is much more challenging, and more increase is obtained from a relatively low AUC score.

To see if the learnt transformation prior  $\mathbf{T}$  of TLR model [31] can help to improve our system performance or not, we also incorporate it to our model. As shown in Figure 3, the transformation prior does not give any improvement on both datasets. In addition, integrating the transformation prior (without the CG segmentation prior and post-smoothing) to LRMR model, on the PASCAL-1500, even decreases the performance compared to only using raw feature for matrix recovery. This suggests that the transformation prior is somewhat biased toward images similar to those in the training dataset, and lacks the adaptability to the PASCAL-1500. Figure 3 also reveals that, on the MSRA-1000, the bottom-up segmentation prior is superior to the learnt transformation prior (96.2% versus 90.4% in terms of AUC score).

In Figure 4, the proposed segmentation driven low-rank matrix recovery (SLR) model is compared with the eight baselines. The ROC curves and AUC scores of these baselines are computed using authors' publicly available codes or their results. Clearly, the proposed model outperforms other models on both datasets, with 1.1% and 3.7% improvement in terms of AUC score on MSRA-1000 and PASCAL-1500 respectively, compared to the best one among these baselines.

Figure 5 shows some examples of saliency maps generated using the top 5 models with the best performance in the baselines and the proposed SLR model. At least two observations can be derived from these examples. On one hand, all models obtain pretty good results when the input image is with high contrast between object and background, like the image in the first column. On the other hand, the proposed model is able to detect objects with occlusion and cluttered scene, while the baselines suffer from limited robustness. For example, in the last three columns, the person, horse and the boats are within cluttered backgrounds, and the car is significantly occluded by the tree trunk. The baselines either fail or only partly highlight these objects; however, our model shows its ability to discover them.

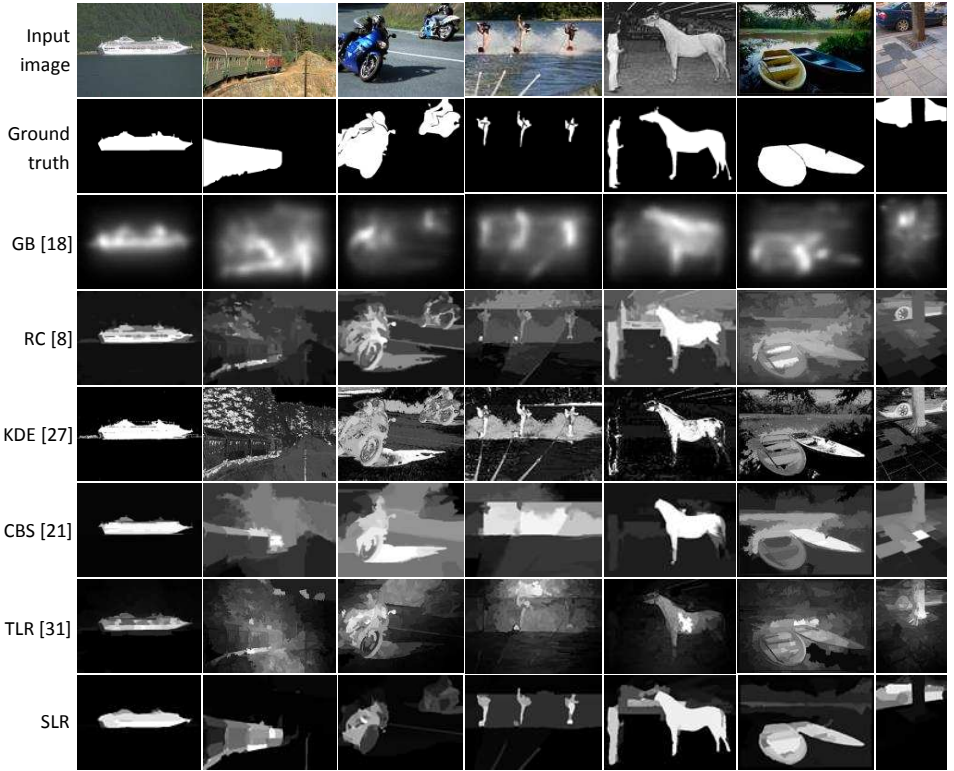


Figure 5: Examples of saliency maps generated using the top 5 models with the best performance in the baselines and the proposed SLR model.

## 7 Conclusion

This paper has presented a novel saliency detection model by leveraging low-rank matrix recovery (LRMR) and bottom-up segmentations. In order to fit the LRMR model to visual saliency detection well, a generic segmentation prior is proposed to guide the matrix recovery, so that image background is represented by a low-rank matrix and salient objects are recovered by a sparse matrix. A new challenging dataset PASCAL-1500 has been introduced for the evaluation of saliency detection. Experiments, on the widely used MSRA-1000 and the newly introduced PASCAL-1500, show that the proposed saliency model outperforms the state-of-the-art models.

## Acknowledgements

This work was partially supported by the French research agency ANR through the Caiman project, the National Natural Science Foundation of China under Grant No. 61171144, the Key (Key grant) Project of Chinese Ministry of Education (No. 212053) and Innovation Program of Shanghai Municipal Education Commission (No. 12ZZ086).

## References

- [1] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk. Frequency-tuned salient region detection. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, 2009.
- [2] Bogdan Alexe, Thomas Deselaers, and Vittorio Ferrari. What is an object? In *Proc. IEEE Conference Computer Vision and Pattern Recognition (CVPR)*, 2010.
- [3] Pablo Arbeláez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5):898–916, 2011.
- [4] Tamar Avraham and Michael Lindenbaum. Esaliency (extended saliency): Meaningful attention using stochastic image modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4):693–708, 2010.
- [5] Ali Borji, Dicky N Sihite, and Laurent Itti. Salient object detection: A benchmark. In *Proc. European Conference on Computer Vision (ECCV)*. 2012.
- [6] Neil Bruce and John Tsotsos. Saliency based on information maximization. *Advances in neural information processing systems*, 18:155, 2006.
- [7] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *arXiv preprint arXiv:0912.3599*, 2009.
- [8] Ming-Ming Cheng, Guo-Xin Zhang, N.J. Mitra, Xiaolei Huang, and Shi-Min Hu. Global contrast based salient region detection. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, 2011.
- [9] Dorin Comaniciu and Peter Meer. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.
- [10] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338, 2010.
- [11] Hans Georg Feichtinger. *Gabor Analysis and Algorithms: Theory and Applications*. Birkhauser Boston, 1997.
- [12] Dashan Gao, Vijay Mahadevan, and Nuno Vasconcelos. The discriminant center-surround hypothesis for bottom-up saliency. *Advances in neural information processing systems*, 20:1–8, 2007.
- [13] Stas Goferman, Ayellet Tal, and Lihi Zelnik-Manor. Puzzle-like collage. In *Computer Graphics Forum*, volume 29, pages 459–468, 2010.
- [14] Stas Goferman, Lihi Zelnik-Manor, and Ayellet Tal. Context-aware saliency detection. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, 2010.
- [15] Chenlei Guo, Qi Ma, and Liming Zhang. Spatio-temporal saliency detection using phase spectrum of quaternion fourier transform. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, 2008.

- [16] Junwei Han, King N Ngan, Mingjing Li, and Hong-Jiang Zhang. Unsupervised extraction of visual attention objects in color images. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(1):141–145, 2006.
- [17] Tatsuya Harada, Yoshitaka Ushiku, Yuya Yamashita, and Yasuo Kuniyoshi. Discriminative spatial pyramid. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, 2011.
- [18] Jonathan Harel, Christof Koch, and Pietro Perona. Graph-based visual saliency. pages 545–552. MIT Press, 2007.
- [19] Xiaodi Hou and Liqing Zhang. Saliency detection: A spectral residual approach. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, 2007.
- [20] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998.
- [21] Huaizu Jiang, Jingdong Wang, Zejian Yuan, Tie Liu, Nanning Zheng, and Shipeng Li. Automatic salient object segmentation based on context and shape prior. In *Proc. British Machine Vision Conference (BMVC)*, 2011.
- [22] Christof Koch and Shimon Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. pages 115–141. Springer, 1987.
- [23] Daniel Kuettel, Matthieu Guillaumin, and Vittorio Ferrari. Segmentation propagation in imagenet. In *Proc. European Conference on Computer Vision (ECCV)*, 2012.
- [24] Congyan Lang, Guangcan Liu, Jian Yu, and Shuicheng Yan. Saliency detection by multitask sparsity pursuit. *IEEE Transactions on Image Processing*, 21(3):1327–1338, 2012.
- [25] Zhouchen Lin, Minming Chen, and Yi Ma. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*, 2010.
- [26] Tie Liu, Zejian Yuan, Jian Sun, Jingdong Wang, Nanning Zheng, Xiaouu Tang, and Heung-Yeung Shum. Learning to detect a salient object. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(2):353–367, 2011.
- [27] Zhi Liu, Ran Shi, Liquan Shen, Yinzhu Xue, King Ngi Ngan, and Zhaoyang Zhang. Unsupervised salient object segmentation based on kernel density estimation and two-phase graph cut. *IEEE Transactions on Multimedia*, 14(4):1275–1289, 2012.
- [28] Ueli Rutishauser, Dirk Walther, Christof Koch, and Pietro Perona. Is bottom-up attention useful for object recognition? In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, 2004.
- [29] F Shahbaz Khan, Joost van de Weijer, and Maria Vanrell. Top-down color attention for object recognition. In *Proc. IEEE International Conference Computer Vision (ICCV)*, 2009.

- [30] Gaurav Sharma, Frédéric Jurie, and Cordelia Schmid. Discriminative spatial saliency for image classification. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [31] Xiaohui Shen and Ying Wu. A unified approach to salient object detection via low rank matrix recovery. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, 2012.
- [32] Denis Simakov, Yaron Caspi, Eli Shechtman, and Michal Irani. Summarizing visual data using bidirectional similarity. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, 2008.
- [33] Eero P. Simoncelli and William T. Freeman. The steerable pyramid: a flexible architecture for multi-scale derivative computation. In *Proc. IEEE International Conference Image Processing*, 1995.
- [34] Dirk Walther and Christof Koch. Modeling attention to salient proto-objects. *Neural Networks*, 19(9):1395–1407, 2006.
- [35] Jingdong Wang, Long Quan, Jian Sun, Xiaoou Tang, and Heung-Yeung Shum. Picture collage. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, 2006.
- [36] Junchi Yan, Mengyuan Zhu, Huanxi Liu, and Yuncai Liu. Visual saliency detection via sparsity pursuit. *IEEE Signal Processing Letters*, 17(8):739–742, 2010.
- [37] Lingyun Zhang, Matthew H Tong, Tim K Marks, Honghao Shan, and Garrison W Cottrell. Sun: A bayesian framework for saliency using natural statistics. *Journal of Vision*, 8(7), 2008.