Dense and Sparse Reconstruction Error Based Saliency Descriptor

Huchuan Lu, *Senior Member, IEEE*, Xiaohui Li, Lihe Zhang, Xiang Ruan, and Ming-Hsuan Yang, *Senior Member, IEEE*

***Abstract*— In this paper, we propose a visual saliency detection algorithm from the perspective of reconstruction error. The image boundaries are first extracted via superpixels as likely cues for background templates, from which dense and sparse appearance models are constructed. First, we compute dense and sparse reconstruction errors on the background templates for each image region. Second, the reconstruction errors are propagated based on the contexts obtained from K-means clustering. Third, the pixel-level reconstruction error is computed by the integration of multi-scale reconstruction errors. Both the pixel-level dense and sparse reconstruction errors are then weighted by image compactness, which could more accurately detect saliency. In addition, we introduce a novel Bayesian integration method to combine saliency maps, which is applied to integrate the two saliency measures based on dense and sparse reconstruction errors. Experimental results show that the proposed algorithm performs favorably against 24 state-of-the-art methods in terms of precision, recall, and F-measure on three public standard salient object detection databases.**

***Index Terms*— Saliency detection, dense/sparse reconstruction error, sparse representation, context-based propagation, region compactness, Bayesian integration.**

I. Introduction

VISUAL saliency is concerned with the distinct perceptual quality of biological systems which makes certain regions of a scene stand out from their neighbors and catch immediate attention. Numerous biologically plausible models have been developed to explain the cognitive process of humans and animals [1], [2]. In computer vision, more emphasis is paid to detect salient objects in images based on features with generative and discriminative algorithms. Due to its advantage of reducing the complexity of scene analysis, saliency detection plays an important preprocessing role in many computer vision tasks, including image segmentation [3], categorization [4], detection [5], recognition [6], [7], thumbnailing [8] and compression [9], [10], to name a few.

Recently many saliency detection approaches have been proposed, which can be generally divided into two categories: slow, knowledge-driven, top-down models and fast, data-driven, bottom-up models. Contrast-based saliency has been widely investigated in recent years and has become one of the most active sub-topics in bottom-up saliency. The research on contrast-based saliency mainly focuses on two aspects, which can be summarized as "how to contrast" and "what to contrast with".

Motivated by the neuronal architecture of the early primate vision system, Itti *et al.* [11] propose a saliency detection model based on multi-scale image features and define visual attention as the local center-surround difference, which is the early answer to "what to contrast with." Klein and Frintrop [12] regard the center-surround difference as the multi-scale contrast of the center and surround feature distributions, inspired by the information theory Kullback-Leibler divergence. While center-surround contrast-based measures are able to detect salient objects, existing bottom-up approaches are less effective in suppressing background pixels. Different from the center-surround contrast, local contrast is measured by comparing a region only with its relevant contexts (defined as a set of region neighbors in the spatial or feature space) [13]–[15].

Despite local contrast accords with the neuroscience principle global contrast should also be taken into account when one region is similar to its surrounds but still distinct in the whole scene. In other words, global contrast aims to capture the holistic rarity or uniqueness from an image. Achanta *et al.* [16] use global Gaussian blur to suppress noise and high frequency patterns but do not account for spatial relationship, which may lead to highlighted background. Recent methods [17], [18] measure global contrast-based saliency based on spatially weighted feature dissimilarities. Perazzi *et al.* [19] formulate saliency estimation using two Gaussian filters by which color and position are respectively exploited to measure region uniqueness and distribution.

Recent years, saliency detection algorithms based on learning [20], [21] or deep learning [22], [23] techniques have attracted more and more attention. Lu *et al.* [20] first learn optimal saliency seeds, and then propagates saliency information under a diffusion framework. Liu *et al.* [21] learn a novel Partial Differential Equation system adaptively to model the

Manuscript received May 13, 2015; revised October 1, 2015 and December 18, 2015; accepted January 21, 2016. Date of publication February 2, 2016; date of current version February 23, 2016. The work of H. Lu and L. Zhang were supported by the National Natural Science Foundation of China under Grant 61528101, Grant 61472060, and Grant 61371157. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yongyi Yang.

H. Lu, X. Li, and L. Zhang are with the Faculty of Electronic Information and Electrical Engineering, School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China (e-mail: lhchuan@dlut.edu.cn; lxh1988924@gmail.com; zhanglihe@dlut.edu.cn).

X. Ruan is with OMRON Corporation, Kusatsu 525-0035, Japan (e-mail: ruanxiang@gmail.com).

M.-H. Yang is with the Department of Electrical Engineering and Computer Science, University of California at Merced, Merced, CA 95344 USA (e-mail: mhyang@ucmerced.edu).

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

Digital Object Identifier 10.1109/TIP.2016.2524198

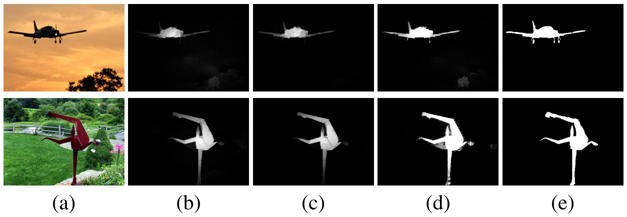


Fig. 1. Saliency maps based on dense and sparse reconstruction errors. Brighter pixels indicate higher saliency values. (a) Original images. (b) (c) Saliency maps from dense and sparse reconstruction, respectively. (d) The Bayesian integrated saliency map of (b) and (c). (e) Ground truth.

learned saliency seeds propagation process. Zhao *et al.* [22] employ deep Convolutional Neural Networks (CNNs) for visual saliency detection which simultaneously take global and local context into consideration. In [23], multi-scale features are extracted by CNNs to model a high-quality saliency detection framework.

In this paper, we exploit image boundaries as the likely background regions from which templates are extracted. Motivated by the observations that dense representation from background template is able to capture the intrinsic properties of background characteristic but is sensitive to noise, while sparse representation manages to model the uniqueness and compactness of regions but is less robust when background templates incorporate foreground regions, we use their combination as indication of saliency which will work in a complementary way to compensate for the defects of each other. We exploit a context-based propagation mechanism to obtain more uniform reconstruction errors over the image. The reconstruction error of each pixel is then assigned by an integration of multi-scale reconstruction errors. Furthermore, we incorporate region compactness into the dense and sparse reconstruction errors to further improve their accuracy. In addition, we present an effective Bayesian integration method to combine saliency maps constructed from dense and sparse reconstruction (see Figure 1).

The main contributions of this work are as follows:

1. We propose an algorithm to detect salient objects by dense and sparse reconstruction for each individual image, which computes more effective bottom-up contrast-based saliency.
2. A context-based propagation mechanism is proposed for region-based saliency detection, which uniformly highlights the salient objects and smooths the region saliency.
3. A compactness weighted reconstruction error is proposed by incorporating spatial compactness into the dense and sparse reconstruction errors.
4. We present a Bayesian integration method to combine saliency maps, which achieves more favorable results than the existing integration strategy.

II. Related Works

A. Boundary Prior for Saliency Detection

According to the basic rule of photographic composition that human photographers tend to place objects of interest in the center of photographs [24], Wei *et al.* [25] define geodesic saliency by regarding image boundaries as background priors, which is validated reasonable and effective. Based on the fact that there is a strong center bias in some saliency detection databases [26], we also consider image boundary regions as background priors for saliency detection. However different from [25], we extract image boundary regions as likely cues for background templates, by which dense and sparse reconstruction errors are computed for each image region.

B. Image Representation for Saliency Detection

To describe patch features in a relatively low dimensional space, Duan *et al.* [18] adopt an equivalent method to Principle Component Analysis (PCA) to reduce data dimension and use the reduced dimensional feature to calculate patch dissimilarities. Sparse representation is similarly employed as a way of image feature representation in [15], [27], and [28], with a dictionary trained from a large set of natural image features. Since each image patch is represented by a dictionary or basis functions learned from a set of natural image patches rather than the remaining other patches of its corresponding image, the most relevant visual information of each individual image is not fully extracted and exhaustively used in saliency estimation. Therefore, these methods do not uniformly detect salient objects or suppress the background in a scene.

To address the above mentioned issues, we make full use of image visual information by constructing background bases from the extracted background templates for each individual image. From the view of reconstruction, we assume that the image background can be better reconstructed than the foreground by a linear combination of the background bases. Therefore, the contrast-based saliency of an image region is indicated by its reconstruction error, which implies its difference from the background information. In other words, larger reconstruction error based on the background templates indicates larger saliency value for an image region.

C. Bayesian Inference for Saliency Detection

Bayes formula is used in many saliency detection models such as [24], [29], and [30]. Zhang *et al.* [24] propose a saliency detection algorithm based on a Bayesian framework from which bottom-up saliency emerges naturally as the self-information of visual features. Xie and Lu [30] develop a prior map to replace the constant value in [29] as Bayes prior probability, which largely improves the accuracy of Bayes posterior probability. Both [29] and [30] compute the likelihood probability through CIELab color. However, the noise in color space may be introduced again despite it has been removed by the prior, which may have negative impact on the posterior, sometimes even make the posterior more inaccurate than the prior. Therefore, if a saliency map which better expresses image saliency than color information acts as the observation likelihood, a more accurate Bayes posterior probability could be achieved, as done in this work.

III. Dense and Sparse Reconstruction Errors

We use both dense and sparse reconstruction errors to measure the saliency of each image region. We note that

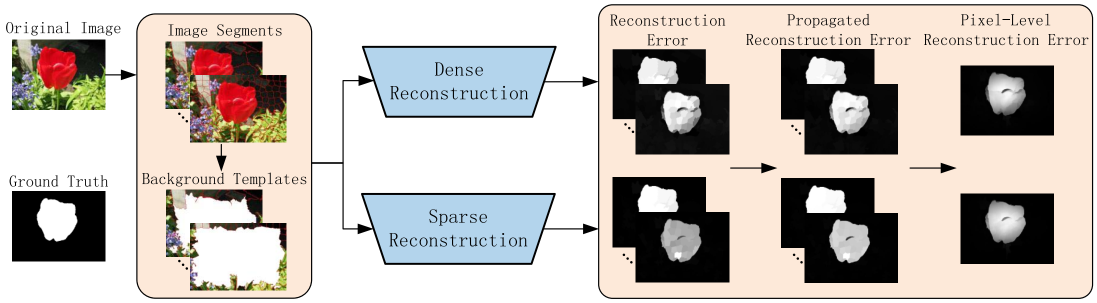


Fig. 2. Main steps of dense and sparse reconstruction errors.

a dense appearance model renders more expressive and generic descriptions of background templates, whereas a sparse appearance model generates unique and compact representations. It is well known that dense appearance models are more sensitive to noise. For cluttered scenes, dense appearance models may be less effective in measuring salient objects via reconstruction errors. On the other hand, solutions (i.e., coefficients) by sparse representation are less stable (e.g., similar regions may have different sparse coefficients), which may lead to discontinuous saliency detection results. In this work, we use both representations to model regions and measure saliency based on reconstruction errors.

The dense and sparse reconstruction errors are obtained as shown in Figure 2. First, we extract the image boundary segments as the background templates for saliency detection. Second, we reconstruct all the image regions based on the background templates and normalize the reconstruction errors to the range of $[0, 1]$. Third, a propagation mechanism is proposed to exploit local contexts obtained from K-means clustering. Forth, pixel-level reconstruction error is computed by integrating multi-scale reconstruction errors.

A. Background Templates

To better capture structural information, we first generate superpixels using the simple linear iterative clustering (SLIC) algorithm [31] to segment an input image into multiple uniform and compact regions (i.e., segments). We use the mean Lab and RGB color features and coordinates of pixels to describe each segment by $\mathbf{x} = \{L, a, b, R, G, B, x, y\}^\top$. The entire image is then represented as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N] \in \mathbb{R}^{D\times N}$, where N is the number of segments and D is the feature dimension. Motivated by the representation ability of image boundary, we extract the D-dimensional feature of each boundary segment as $\mathbf{b}$ and construct the background template set as $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_M]$, where M is the number of image boundary segments Figure 2 shows some background templates extracted at different scales. Given the background templates, we compute two reconstruction errors by dense and sparse representation for each image region, respectively.

B. Dense Reconstruction Error

A segment with larger reconstruction error based on the background templates is more likely to be the foreground. Based on this concern, the reconstruction error of each region is computed based on the dense appearance model generated from the background templates $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_M]$, $\mathbf{B} \in \mathbb{R}^{D\times M}$ using Principal Component Analysis (PCA).

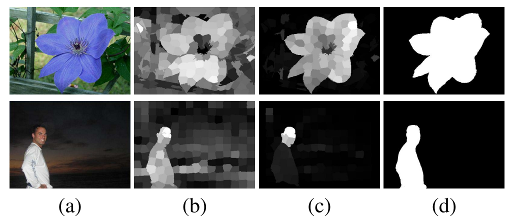


Fig. 3. Saliency maps based on dense and sparse reconstruction errors. Brighter pixels indicate higher saliency values. (a) Original images. (b) Saliency maps from dense reconstruction. (c) Saliency maps from sparse reconstruction. (d) Ground truth.

The eigenvectors from the normalized covariance matrix of $\mathbf{B}$, $\mathbf{U_B} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{D'}]$, corresponding to the largest D' eigenvalues, are computed to form the PCA bases of the background templates. With the PCA bases $\mathbf{U_B}$, we compute the reconstruction coefficient of segment i $(i \in [1, N])$.

$$\boldsymbol{\beta}_i = \mathbf{U_B}^\top(\mathbf{x}_i - \bar{\mathbf{x}}), \tag{1}$$

and the dense reconstruction error of segment i is

$$\varepsilon_i^d = \|\mathbf{x}_i - (\mathbf{U_B}\boldsymbol{\beta}_i + \bar{\mathbf{x}})\|_2^2, \tag{2}$$

where $\bar{\mathbf{x}}$ is the mean feature of $\mathbf{X}$. The saliency measure is proportional to the normalized reconstruction error (within the range of $[0, 1]$).

Figure 3(b) shows some saliency detection results via dense reconstruction. Dense representations model data points with a multivariate Gaussian distribution in the feature space, and thus it may be difficult to capture multiple scattered patterns especially when the number of examples is limited. Therefore, when image saliency detection encounters complicated background, it may be difficult for the dense appearance model to train a set of background bases which could extract complete background information, thus leading to background noise in saliency map. The first row of Figure 3 shows an example where some background regions have large dense reconstruction errors (i.e., inaccurate saliency measure).

C. Sparse Reconstruction Error

Motivated by the demonstrated success of sparsity-based classifiers for computer vision tasks [32], [33], we make an assumption that the background can be better represented than the foreground by a linear combination of the background templates. We use the set of background templates $\mathbf{B}$ as the bases for sparse representation, and encode the image segment i by

$$\boldsymbol{\alpha}_i^* = \underset{\boldsymbol{\alpha}_i}{\operatorname{argmin}} \|\mathbf{x}_i - \mathbf{B}\boldsymbol{\alpha}_i\|_2^2 + \lambda\|\boldsymbol{\alpha}_i\|_1, \tag{3}$$

and the sparse reconstruction error is

$$\varepsilon_i^s = \left\|\mathbf{x}_i - \mathbf{B}\boldsymbol{\alpha}_i^*\right\|_2^2. \tag{4}$$

Since all the background templates are regarded as the basis functions, sparse reconstruction error can better suppress the background compared with dense reconstruction error especially in cluttered images, as shown in the middle row of Figure 3.

Nevertheless, there are some drawbacks in measuring saliency with sparse reconstruction errors. If some foreground segments are collected into the background templates (e.g., when objects appear at the image boundaries), their saliency measures are close to 0 due to low sparse reconstruction errors. In addition, the saliency measures for the other regions are less accurate due to inaccurate inclusion of foreground segments as part of sparse basis functions. On the other hand, the dense appearance model is not affected by this problem. When foreground segments are mistakenly included in the background templates, the extracted principle components from the dense appearance model may be less effective in describing these foreground regions. As shown in the second row of Figure 3, when some foreground segments at the image boundary (e.g., torso and arm) are not detected via sparse reconstruction, these regions are still be detected by the dense counterpart.

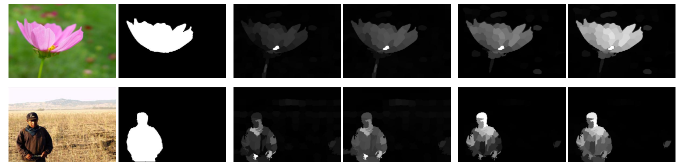
We note sparse reconstruction error is more robust to deal with complicated background, while dense reconstruction error is more accurate to handle the object segments at image boundaries. Therefore, dense and sparse reconstruction errors are complementary in measuring saliency.

D. Context-Based Reconstruction Error Propagation

Considering that even the best segmentation algorithms can not avoid to separate an image region into multiple smaller homogeneous ones. Thus two segments sharing similar features in the feature space may share different reconstruction errors, which results in discontinuous saliency maps. On the other hand, even though salient objects do not touch the image boundaries in most cases, some background templates may be part of foreground in fact. In this case, the reconstruction errors may not precisely represent the contrast with the true background and consequently results in mistakes in saliency detection.

To overcome the above two problems, we propose a context-based error propagation method to smooth the reconstruction errors generated by dense and sparse appearance models. Both dense and sparse reconstruction errors of segment i (i.e., ε_i^d and ε_i^s) are denoted by ε_i for conciseness.

We first apply the K-means algorithm to cluster N image segments into K clusters via their D-dimensional features and



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

Fig. 4. Saliency maps with the context-based error propagation. (a) and (b) are original images and ground truth. (c) and (d) are original and propagated dense reconstruction errors. (e) and (f) are original and propagated sparse reconstruction errors.

initialize the propagated reconstruction error of segment i as $\tilde{\varepsilon}_i = \varepsilon_i$. All the segments are sorted in descending order by their reconstruction errors and considered as multiple hypotheses. They are processed sequentially by propagating the reconstruction errors in each cluster. The propagated reconstruction error of segment i belonging to cluster k $(k = 1, 2, \ldots, K)$, is modified by considering its appearance-based context consisting of the other segments in cluster k as follows:

$$\tilde{\varepsilon}_i = \tau \sum_{j=1}^{N_c} \mathrm{w}_{ik_j} \tilde{\varepsilon}_{k_j} + (1-\tau)\,\varepsilon_i, \tag{5}$$

$$\mathrm{w}_{ik_j} = \frac{\exp\left(-\frac{\left\|\mathbf{x}_i - \mathbf{x}_{k_j}\right\|^2}{2\sigma_{\mathbf{x}}^2}\right)\left(1 - \delta\left(k_j - i\right)\right)}{\sum_{j=1}^{N_c} \exp\left(-\frac{\left\|\mathbf{x}_i - \mathbf{x}_{k_j}\right\|^2}{2\sigma_{\mathbf{x}}^2}\right)}, \tag{6}$$

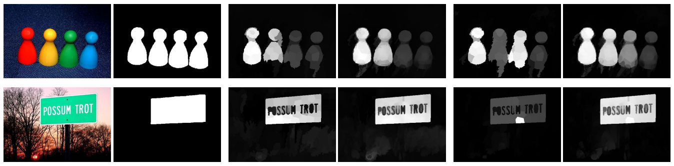
where $\{k_1, k_2, \ldots, k_{N_c}\}$ denote the N_c segment labels in cluster k and τ is a weight parameter. The first term on the righthand side of Eq. 5 is the weighted averaging reconstruction error of the other segments in the same cluster, and the second term is the initial dense or sparse reconstruction error. That is, for segment i, by considering all the other segments belonging to the same cluster k (i.e., the appearance-based local context), the reconstruction error can be better estimated. The weight of each segment context is defined by its normalized similarity with segment i in Eq. 6, where $\sigma_{\mathbf{x}}^2$ is the sum of the variance in each feature dimension of $\mathbf{X}$ and $\delta(\cdot)$ is the indicator function.

Figure 4 shows two examples where the context-based propagation mechanism smooths the reconstruction errors in a cluster, thereby uniformly highlighting the image objects. The bottom row of Figure 4 presents one case that several segments of the object (e.g., torso) are mistakenly included in the background templates, and therefore they are not correctly identified by the dense and sparse appearance models. Nevertheless, the reconstruction errors of these segments are modified by taking the contributions of their contexts into consideration using Eq. 5.

E. Pixel-Level Reconstruction Error

For a full-resolution saliency map, we assign saliency to each pixel by integrating results from multi-scale reconstruction errors.

To handle the scale problem, we generate superpixels at N_s different scales. We compute and propagate both dense



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

Fig. 5. Saliency maps with the multi-scale integration of propagated reconstruction errors. (a) and (b) are original images and ground truth. (c) and (d) are propagated dense reconstruction errors without and with integration. (e) and (f) are propagated sparse reconstruction errors without and with integration.

and sparse reconstruction errors for each scale. We integrate multi-scale reconstruction errors and compute the pixel-level reconstruction error by

$$\overline{E}(z) = \frac{\sum_{s=1}^{N_s} \omega_{zn^{(s)}} \tilde{\varepsilon}_{n^{(s)}}}{\sum_{s=1}^{N_s} \omega_{zn^{(s)}}}, \quad \omega_{zn^{(s)}} = \frac{1}{\left\|\mathbf{d}_z - \mathbf{x}_{n^{(s)}}\right\|_2}, \tag{7}$$

where $\mathbf{d}_z$ is a D-dimensional feature of pixel z and $n^{(s)}$ denotes the label of the segment containing pixel z at scale s. Similarly to [14], we utilize the similarity between pixel z and its corresponding segment $n^{(s)}$ as the weight to average the multi-scale reconstruction errors.

Figure 5 shows some examples where objects are more precisely identified by the reconstruction errors with multi-scale integration, which suggests the effectiveness of using multi-scale integration mechanism to measure saliency.

IV. Compactness Weighted Reconstruction Error

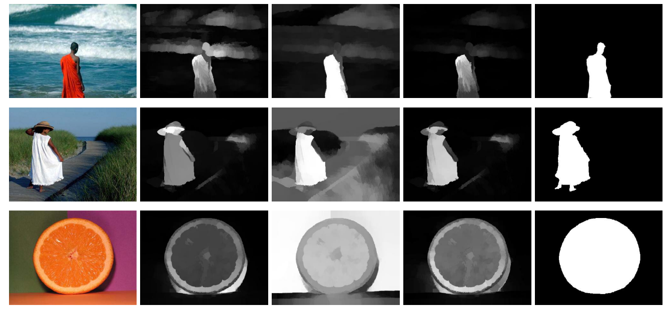
Considering that salient regions generally group compactly in the spatial domain, while background ones always distribute over the entire image with higher spatial variance, we conclude that region compactness in the image spatial domain is vital to saliency detection.

The dense and sparse reconstruction errors introduced in Section III imply the feature distance of an image region to the background templates in the color space, which exhibits promising performance in suppressing the background noise. However, spatially adjacent regions may still share largely different reconstruction errors without considering the color compactness or distribution in spatial domain. Therefore we propose a compactness weighted reconstruction error to measure saliency by taking the color distribution into consideration in order to smooth the object saliency.

A. Region Compactness

Based on the observation that salient regions tend to distribute compactly, we define the region compactness by its inverse spatial variance, i.e. the smaller the spatial variance is, the larger the compactness is, thus the more salient the region will be.

We represent the image segment i as $\mathbf{x}_i = [\mathbf{f}_i; \mathbf{p}_i]$ where $\mathbf{f} = \{L, a, b, R, G, B\}^\top$ and $\mathbf{p} = \{x, y\}^\top$ denote the color



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

Fig. 6. Saliency maps based on compactness weighted reconstruction error. (a) Original images. (b) Saliency maps from sparse reconstruction. (c) Saliency maps generated by region compactness. (d) Saliency maps based on compactness weighted reconstruction error. (e) Ground truth.

feature and position information of each segment respectively. Then the region compactness of segment i can be defined as

$$c_i = 1 - \left|\left| \sum_{j=1}^{N} \pi_{ij} \mathbf{p}_j^{\,2} - \left(\sum_{j=1}^{N} \pi_{ij} \mathbf{p}_j\right)^2 \right|\right|_1, \tag{8}$$

where

$$\pi_{ij} = \frac{\exp\left(-\frac{\|\mathbf{f}_i - \mathbf{f}_j\|^2}{2\sigma_{\mathbf{f}}^2}\right)}{\sum_{j=1}^{N} \exp\left(-\frac{\|\mathbf{f}_i - \mathbf{f}_j\|^2}{2\sigma_{\mathbf{f}}^2}\right)}. \tag{9}$$

The squared position is computed as $\mathbf{p}^2 = \{x^2, y^2\}^\top$ in Eq. 8. The region position is weighted by the normalized color similarity in Eq. 9 similarly to [19], where $\sigma_{\mathbf{f}}^2$ is the sum of the variance in each feature dimension similarly to $\sigma_{\mathbf{x}}^2$ in Eq. 6. The color weighted region position minus the mean weighted region position effectively describes the spatial distribution of region i, and L1-norm combines the horizontal and vertical variances together. As summarized in Eq. 8, we inverse the description of the spatial distribution to calculate region compactness over the entire image.

Figure 6(c) shows some saliency maps generated by the region compactness. Compared to the sparse reconstruction error (Figure 6(b)), the region compactness could highlight the salient object more uniformly due to the spatially compact distribution of object color. However, the region compactness is more sensitive to background noise. The bottom row of Figure 6 shows a failure case where the object is of large size and the background color is uniformly distributed, which consequently leads to false object detection. Therefore we propose a compactness weighted reconstruction error in order to further enhance the contrast between salient object and background.

B. Compactness Weighted Reconstruction Error

Without taking region compactness into consideration, the reconstruction error of salient object is not grouped as well as the color feature. Therefore we weigh the reconstruction error introduced in Section III by the region compactness.

We first propagate the initial region compactness (i.e., c_i for each segment in Eq. 8) and integrate the multi-scale results to obtain pixel-level compactness as $\overline{C}(z)$ for pixel z, similarly to the reconstruction error $\overline{E}(z)$ in Eq. 7. Then we calculate the compactness weighted reconstruction error for pixel z as

$$E(z) = w^C(z) * \overline{E}(z), \tag{10}$$

where function $w^C(\cdot)$ can be any positive weight function (e.g., exponential, log and sigmoid function) of region compactness and we simply define it as $w^C(z) = \overline{C}(z)$ in this work.

Figure 6 shows three examples where the compactness weighted sparse reconstruction error performs better than the non-weighted one. As shown on the top and middle row, the salient object uniformly pops out due to the region compactness which presents larger contrast between the foreground and background than the reconstruction error. However, we note that region compactness may be more sensitive to the background noise than the reconstruction error (see the bottom two rows of Figure 6). By taking spatial distribution into account, the compactness weighted reconstruction error could reduce the negative impact of region compactness and highlight the foreground as well as suppress the background. The weighted reconstruction error detects the salient regions more accurately and uniformly, even correcting the false detection by the region compactness in the bottom image. In summary, as an important factor for saliency detection, the region compactness largely helps to locate the object by weighing the reconstruction error.

C. Saliency Assignment Refined by Object-Biased Gaussian

Borji *et al.* show that there is a center bias in some saliency detection datasets [26]. Recently center prior has been used in [14], [15], [18], [19], and [34] and usually formulated as a Gaussian model,

$$G(z) = \exp\left[-\left(\frac{(x_z-\mu_x)^2}{2\sigma_x{}^2} + \frac{(y_z-\mu_y)^2}{2\sigma_y{}^2}\right)\right], \tag{11}$$

where $\mu_x = x_c$ and $\mu_y = y_c$ denote the coordinates of the image center and x_z and y_z are the coordinates of pixel z. Since salient objects do not always appear at the image center as Figure 7 shows, the center-biased Gaussian model is not effective and may include background pixels or miss the foreground regions. We use an object-biased Gaussian model G_o with $\mu_x = x_o$ and $\mu_y = y_o$, where x_o and y_o denote the object center derived from the pixel error in Eq. 7:

$$x_o = \sum_i \frac{E(i)}{\sum_j E(j)} x_i, \quad y_o = \sum_i \frac{E(i)}{\sum_j E(j)} y_i. \tag{12}$$

We set $\sigma_x = 0.25 \times H$ and $\sigma_y = 0.25 \times W$, where W and H respectively denote the width and height of an image. With the object-biased Gaussian model, the saliency of pixel z is computed by

$$S(z) = G_o(z) * E(z). \tag{13}$$

Figure 7 shows an example when the object does not locate at the image center. Comparing the two refined maps of the saliency via dense or sparse reconstruction in the bottom row, the proposed object-biased Gaussian model renders more accurate object center, and therefore better refines the saliency detection results.

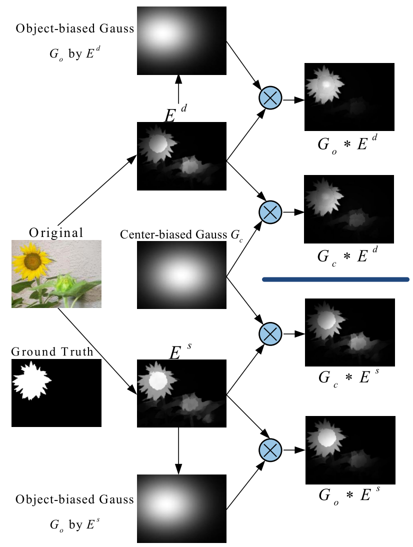


Fig. 7. Comparison of center-biased (G_c) and object-biased (G_o) Gaussian refinement. E^d and E^s are the multi-scale integrated dense and sparse reconstruction error maps, respectively.

V. Bayesian Integration of Saliency Maps

As mentioned in Section III, the saliency measures by dense and sparse reconstruction errors are complementary to each other. To integrate both the saliency measures, we propose an integration method by Bayesian inference, different from the conventional integration strategy simply by weighted averaging saliency maps in [25] and [26].

A. Bayes Formula

Recently, the Bayes formula has been used to measure saliency by the posterior probability in [29] and [35]:

$$p(F|H(z)) = \frac{p(F)p(H(z)|F)}{p(F)p(H(z)|F) + (1-p(F))p(H(z)|B)}, \tag{14}$$

where the prior probability $p(F)$ is a uniform [29] or a saliency map [35] and $H(z)$ is a feature vector of pixel z. The observation likelihood probabilities are computed as:

$$p(H(z)|F) = \prod_{r\in\{L,a,b\}} \frac{N_{b_F(r(z))}}{N_F},$$
$$p(H(z)|B) = \prod_{r\in\{L,a,b\}} \frac{N_{b_B(r(z))}}{N_B}, \tag{15}$$

where N_F denotes the number of pixels in the foreground and $N_{b_F(r(z))} (r \in \{L, a, b\})$ is the number of pixels whose color features fall into the foreground bin $b_F(r(z))$ which contains feature $r(z)$, while the color distribution histogram of the background is denoted likewise by N_B and $N_{b_B(r(z))}$. However, the noise in color space may be introduced again

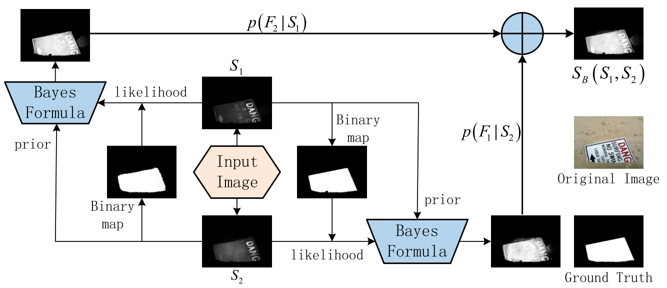


Fig. 8. Bayesian integration of saliency maps. The two saliency measures via dense and sparse reconstruction are respectively denoted by S_1 and S_2.

though it has been removed by the prior probability, and results in inaccurate posterior probability, which makes the posterior even worse than the prior in some cases.

Considering this problem, we take one saliency map in this work as the prior and use the other one instead of Lab color information to compute the likelihoods, which integrates more diverse information from different saliency maps.

B. Bayesian Integration Formula

Given two saliency maps S_1 and S_2 (i.e., from dense and sparse reconstruction), we treat one of them as the prior $S_i (i = \{1, 2\})$ and use the other one $S_j (j \neq i, j = \{1, 2\})$ to compute the likelihood, as shown in Figure 8. First, we threshold the map S_i by its mean saliency value and obtain its foreground and background regions described by F_i and B_i, respectively. In each region, we compute the likelihoods by comparing S_j and S_i in terms of the foreground and background bins at pixel z:

$$p(S_j(z)|F_i) = \frac{N_{b_{F_i}(S_j(z))}}{N_{F_i}}, \; p(S_j(z)|B_i) = \frac{N_{b_{B_i}(S_j(z))}}{N_{B_i}}. \quad (16)$$

Consequently the posterior probability is computed with S_i as the prior by

$$p(F_i|S_j(z)) = \frac{S_i(z)p(S_j(z)|F_i)}{S_i(z)p(S_j(z)|F_i) + (1 - S_i(z))p(S_j(z)|B_i)}. \quad (17)$$

Similarly, the posterior saliency with S_j as the prior is computed. We use these two posterior probabilities to compute an integrated saliency map, $S_B(S_1(z), S_2(z))$, based on Bayesian integration:

$$S_B(S_1(z), S_2(z)) = p(F_1|S_2(z)) + p(F_2|S_1(z)). \quad (18)$$

The proposed Bayesian integration of saliency maps is illustrated in Figure 8. It should be noted that Bayesian integration enforces these two maps to serve as the prior and cooperate with each other in an effective manner, which uniformly highlights salient objects in an image. The proposed saliency detection algorithm via dense and sparse reconstruction is summarized in Algorithm 1.

Algorithm 1 Saliency Detection via Dense and Sparse Reconstruction

Given: An input image $\{\mathbf{X}_s\}_{s=1}^{N_s}$ obtained from SLIC segmentation at N_s different scales, where $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{N}$ for one scale.
For $s = 1, 2, ..., N_s$ **do**
  Obtain the background template set $\mathbf{B}$.
  For $i = 1, 2, ..., N$ **do**
    Calculate the dense reconstruction error ε_i^d by Eq. 1-2.
    Calculate the sparse reconstruction error ε_i^s by Eq. 3-4.
    Calculate the region compactness c_i by Eq. 8-9.
    Calculate the propagated reconstruction errors $\{\tilde{\varepsilon}_i^d, \tilde{\varepsilon}_i^s\}$ and region compactness $\tilde{c}_i$ by Eq. 5-6.
  End
End
Integrate the multi-scale results and obtain the pixel-level reconstruction errors $\{\overline{E}^d(z), \overline{E}^s(z)\}$ and compactness $\overline{C}(z)$ by Eq. 7.
Obtain the weighted reconstruction errors $\{E^d(z), E^s(z)\}$ by Eq. 10.
Calculate the dense and sparse saliency S^d and S^s by Eq. 11-13.
Obtain the Bayesian integrated saliency $S_B(S^d, S^s)$ by Eq. 16-18.
Output: The final saliency map $S_B(S^d, S^s)$.

VI. Experiments

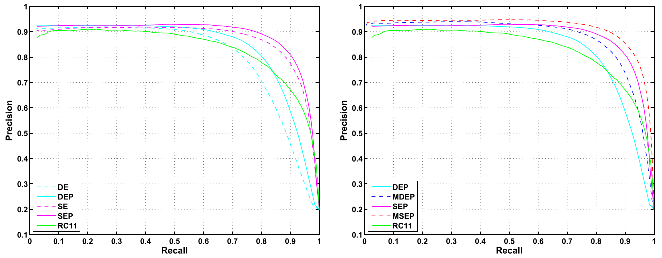
We evaluate the proposed algorithm with twenty-four state-of-the-art algorithms including IT [11], MZ [36], LC [37], GB [38], SR [39], AC [40], FT [16], CA [13], RA [29], RC [17], CB [14], SVO [41], DW [18], SDS [42], SF [19], LR [34], GS [25], XL [35], SIA [43], HS [44], wCO [45], HCT [46], MKB [47], and DSR1 [48] on three benchmark data sets: ASD, MSRA and SOD.

A. Data Sets

The MSRA database [49] contains 5000 images. The ASD database [16] includes 1000 images selected from the MSRA database. Most images in the MSRA and ASD databases have only one salient object and there are usually strong contrast between objects and backgrounds. In addition, we evaluate the proposed algorithm on the SOD database. The SOD database [50] is based on the Berkeley segmentation dataset which is more challenging than the other databases with multiple objects of different sizes and locations in more complicated backgrounds. All images in these dataset correspond to manually labeled ground truth. In order to evaluate the effectiveness of saliency detection models, we employ the common used Precision-Recall (PR) curve and F-measure.

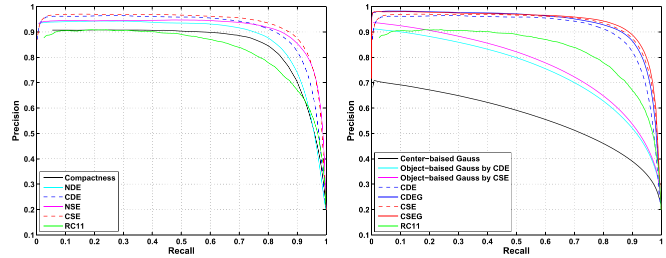
B. Parameter Setting

The two main parameters of our method are the number of clusters K and the weight factor τ in Eq. 5. We set $K = 8$ and $\tau = 0.5$ in all experiments. The parameter λ of Eq. 3, is empirically set to 0.01. For dense reconstruction, we use the eigenvectors corresponding to the biggest eigenvalues which retain 95% of the energy. For background template update, we empirically set the maximal iteration number to 3 in our experiment. For multi-scale reconstruction errors, we generate superpixels at eight different scales respectively with 50 to 400 superpixels. The developed MATLAB code will be made available to the public.



(a) (b)

Fig. 9. Evaluation of saliency via reconstruction error. (a) Based on the context-based propagation. (b) Based on the multi-scale reconstruction error integration. DE: dense reconstruction error; DEP: propagated DE; MDEP: multi-scale integrated DEP; SE: sparse reconstruction error; SEP: propagated SE; MSEP: multi-scale integrated SEP; RC11: baseline method [17].



(a) (b)

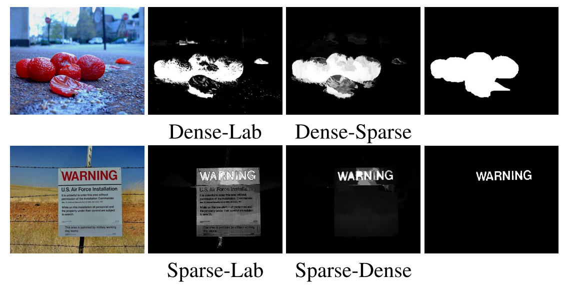
Fig. 10. Evaluation of saliency via compactness weighted reconstruction error. (a) Based on the compactness weighted reconstruction error. (b) Based on the object-biased Gaussian refinement. Compactness: pixel-level compactness; CDE and CSE: compactness weighted dense and sparse reconstruction error; NDE and NSE: non-weighted dense and sparse reconstruction error; CDEG and CSEG: Gaussian refined CDE and CSE; RC11: baseline model [17].

C. Evaluation of Reconstruction Error

We evaluate the proposed dense and sparse reconstruction errors as well as the compactness weighted ones on the ASD database.

1) Reconstruction Error: We evaluate the contribution of the context-based propagation and multi-scale reconstruction error integration in Figure 9. The approach in [17] (referred as RC11) is also presented as a baseline model for comparisons. Figure 9(a) shows that the sparse reconstruction error based on background templates achieves better accuracy in detecting salient objects than RC11 [17], while the dense one is comparable with it. This is due to the strong capacity of dense and sparse reconstruction techniques to model background appearance characters. The context-based reconstruction error propagation method uses segment contexts through K-means clustering to smooth the reconstruction errors and minimize the detection mistakes introduced by the object segments in background templates with improved performance (Figure 9(a)). The reconstruction error of a pixel is assigned by integrating the multi-scale reconstruction errors, which helps generate more accurate and uniform saliency maps. Figure 9(b) shows the improved performance due to the integration of multi-scale reconstruction errors.

2) Compactness Weighted Reconstruction Error: To evaluate the compactness weighted reconstruction error, we calculate pixel-level compactness by the context-based propagation and multi-scale integration and directly utilize it to describe image saliency. We quantitatively compare the performance of the non-weighted and weighted reconstruction errors to figure out the contribution of region compactness to saliency detection. We can see from Figure 10(a) that the pixel-level compactness shows higher precision than RC11 [17], which demonstrates that color compactness is of as much importance as color uniqueness that used in RC11 [17]. Owing to the incorporated compactness factor, the weighted reconstruction error achieves better performance in detecting saliency than the non-weighted one as shown in Figure 10(a). In addition, we also evaluate the object-biased Gaussian refinement for the weighted reconstruction error. Figure 10(b) shows that the object-biased Gaussian model further refines the results and performs better than the center-biased one.



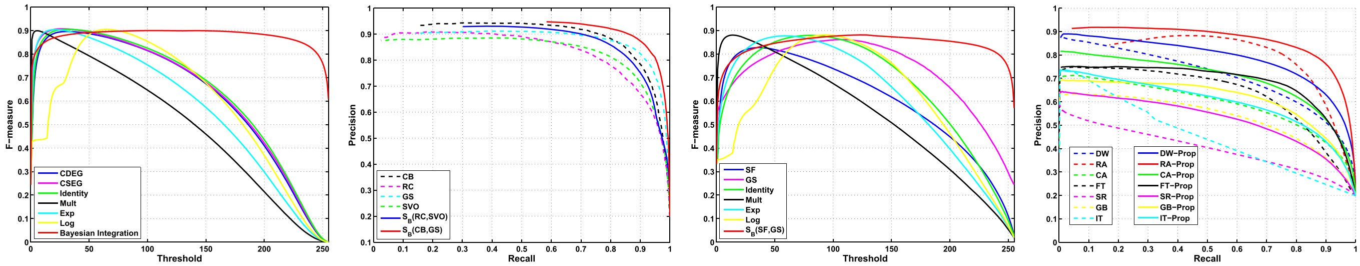
(a) (b) (c) (d)

Fig. 11. Saliency maps based on different Bayes posterior probabilities. (a) and (d) are original image and ground truth. (b) and (c) are saliency maps from Bayes posterior probability. We treat the saliency map by dense (or sparse) reconstruction as the prior, and use the other saliency map by sparse (or dense) reconstruction and Lab color to compute the likelihood, denoted by Dense-Sparse (or Sparse-Dense) and Dense-Lab (or Sparse-Lab), respectively.

D. Evaluation of Bayesian Integration

We also evaluate the proposed Bayesian integration method for combining saliency maps.

1) Bayesian Integration of the Dense and Sparse Saliency Maps: In Section V, we discuss that the posterior probability can be more accurate with likelihood computed by a saliency map rather than the CIELab color space on the condition of the same prior in the Bayes formula. We present experimental results in which we treat the saliency map by dense (or sparse) reconstruction as the prior, and use the other saliency map by sparse (or dense) reconstruction and Lab color to compute the likelihood probability, denoted respectively by Dense-Sparse (or Sparse-Dense) and Dense-Lab (or Sparse-Lab) in Figure 11. With the saliency generated by dense (or sparse) reconstruction as the prior, the result with the likelihood based on sparse (or dense) reconstruction (Figure 11(c)) is more accurate than that with the CIELab color space (Figure 11(b)), which suggests that the Bayes posterior probability with likelihood computed by a saliency map can achieve higher recall (see the top row) and precision (see the bottom row) than that computed by color information, due to the less noise in the saliency map than CIELab color.



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

Fig. 12. (a) F-measure curves of the proposed Bayesian integrated saliency and four other integrated ones of CDEG and CSEG. (b) Precision-recall curves of Bayesian integrated saliency of four state-of-the-art methods. (c) F-measure curves of the proposed Bayesian integrated saliency and four other integrated ones of SF [19] and GS [25]. (d) Further improvement of seven state-of-the-art methods by the proposed saliency propagation, where dashed lines are precision-recall curves of the original state-of-the-art methods, while solid ones are the variants (i.e., the propagated results) denoted by the original names suffixed with *-Prop*.

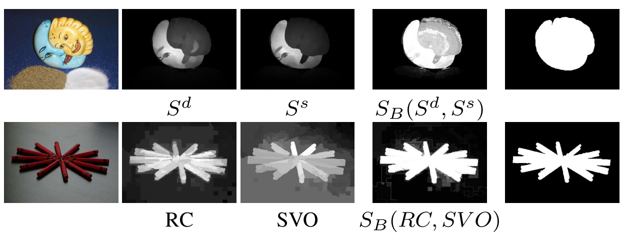


Fig. 13. Saliency maps based on the proposed Bayesian integration. (a) Original images. (d) The Bayesian integrated saliency map of (b) and (c). (e) Ground truth. S^d and S^s are saliency maps via dense and sparse reconstruction, respectively. RC [17], SVO [41], CB [14] and GS [25] denote four state-of-the-art saliency maps.

In addition, we also present the F-measure curve depicted by the mean F-measure at each threshold from 0 to 255 in Figure 12(a). We evaluate the performance of Bayesian integrated saliency map S_B by comparing it with the integration strategies formulated in [26]:

$$S_c = \frac{1}{Z} \sum_i Q(S_i) \quad \text{or} \quad S_c = \frac{1}{Z} \prod_i Q(S_i), \tag{19}$$

where Z is the partition function. In Figure 12(a), we denote the linear summation S_c with $Q(x) = \{x, \exp(x), -1/\log(x)\}$ respectively by *Identity*, *Exp* and *Log*, while denote the accumulation S_c with $Q(x) = x$ by *Mult*. Figure 12(a) shows that the F-measure of the proposed Bayesian integrated saliency map is higher than the other methods at most thresholds, which demonstrates the effectiveness of Bayesian integration.

2) Bayesian Integration of State-of-the-Art Saliency Maps: To further validate the effectiveness of the proposed Bayesian integration mechanism, we implement it on the state-of-the-art methods and evaluate the performance of the integrated saliency maps. We employ Bayesian integration to combine several best salient object detection models reported by [26], including CB [14], SVO [41], RC [17], SF [19] and GS [25]. Figure 12(b) shows that the Bayesian integrated saliency results $S_B(RC, SVO)$ and $S_B(CB, GS)$ achieve better precision-recall curves than either individual saliency. The Bayesian integrated saliency maps (Figure 13(d)) have comparable capability to suppress the background with both the two individual saliency maps (Figure 13(b) and (c)), and simultaneously highlight the salient object more uniformly than them. Due to the uniformly highlighted salient object, the recall value of the integrated saliency map is largely improved, which can be figured out from the quantitative comparisons (see Figure 12(b)) where the minimum recall value of the Bayesian integrated saliency is much higher than others.

We also implement the Bayesian integration on SF [19] and GS [25], and compare the integrated result $S_B(SF, GS)$ with those obtained by other four conventional integration formulas (Eq. 19). Figure 12(c) presents the F-measure curves of the integrated results of SF and GS, including *Identity*, *Mult*, *Exp*, *Log*, and $S_B(SF, GS)$. As shown in Figure 12(c), the Bayesian integrated result $S_B(SF, GS)$ achieves higher F-measure than other integrated ones at most thresholds, which further demonstrates the effectiveness of Bayesian integration.

E. Comparisons With State-of-the-Art Methods

We present the evaluation results of the proposed algorithm compared with the state-of-the-art saliency detection methods on the ASD database in Figure 14, and the MSRA and SOD databases in Figure 15. The precision-recall curves show the proposed algorithm achieves consistent and favorable performance against the state-of-the-art methods. In the bar graphs, the precision, recall and F-measure of the proposed algorithm are comparable with those of the other methods, especially with higher recall and F-measure value. Figure 16 shows that the proposed model generates more accurate saliency maps with uniformly highlighted foreground and well suppressed background on the ASD, MSRA and SOD databases.

Compared with the latest and best salient object detection models (e.g., HS [44], wCO [45], HCT [46] and MKB [47]), our method achieves better or comparable performance among the three databases as shown in Figure 14 and 15. Due to the robust dense and sparse reconstruction model, our saliency map shows better performance than GS [25] which also exploits image boundaries as background priors. wCO [45] introduces the concept of boundary connectivity which describes the likelihood of a region belonging to background effectively. Figure 14 and 15 demonstrate that our background templates based reconstruction error acquire similar performance with this approach. HCT [46] combine feature

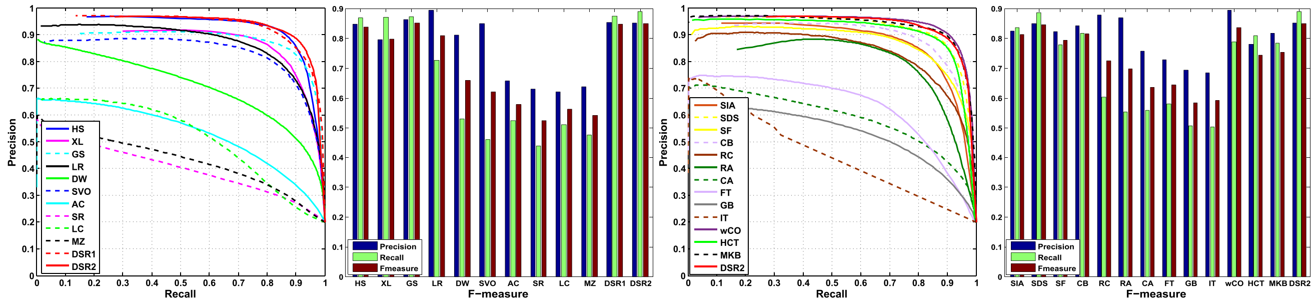
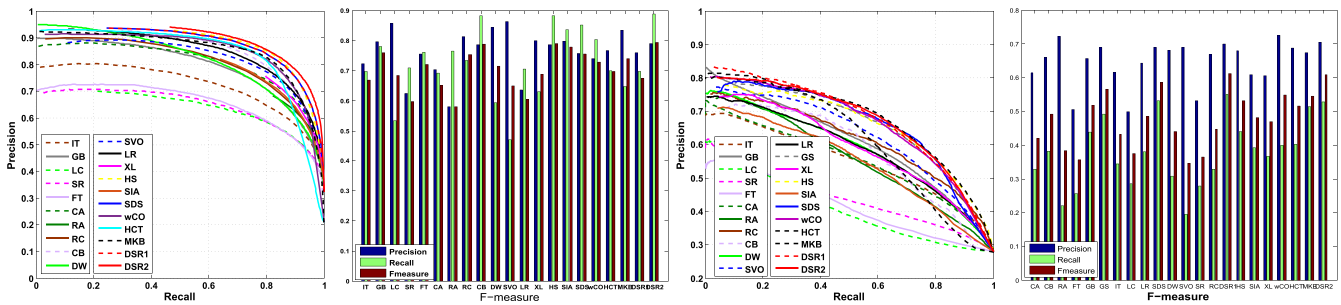
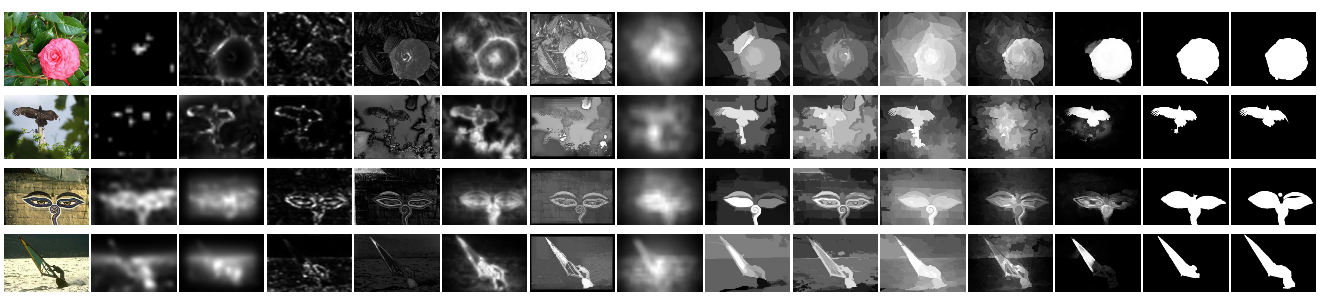


Fig. 14. Performance of the proposed method compared with twenty-four state-of-the-art methods on the ASD database.



(a) (b)

Fig. 15. Performance of the proposed algorithm compared with other state-of-the-art methods on the MSRA and SOD databases, respectively. (a) MSRA. (b) SOD.



Original IT [11] GB [38] SR [39] FT [16] CA [13] RA [29] DW [18] CB [14] RC [17] SVO [41] LR [34] DSR DSR cut GT

Fig. 16. Comparisons of saliency maps. Top, middle and bottom two rows are images from the ASD, SOD and MSRA data sets, respectively. DSR: the proposed algorithm based on dense and sparse reconstruction. DSR cut: cut map using the generated saliency map. GT: ground truth.

TABLE I
COMPARISON OF AVERAGE RUN TIME (SECONDS PER IMAGE)

Method	SR [39]	GB [38]	DW [18]	LR [34]	CA [13]	RA [29]	CB [14]	SVO [41]	Our
Time(s)	0.016	0.460	0.152	20.725	26.878	9.244	1.385	50.526	5.010
Code Type	Matlab	Matlab & C	Matlab	Matlab	Matlab	Matlab	Matlab & C	Matlab	Matlab

vectors in high-dimensional color space linearly to distinct the salient object and background of the input image. However, this algorithm is sensitive to the initial color seed and the high-dimensional color transformation does not fully accord with human visual perception. Therefore, the robustness of this framework is undesirable. MKB [47] learns a multi-kernel boost Support Vector Machine (SVM) classifier within a single image. Because of the lack of samples, MKB [47] performs less robust than our proposed algorithm. The proposed algorithm DSR2 in this paper also performs better than the published conference version DSR1 [48] on the ASD and MSRA databases, and achieves comparable experimental results on the challenging SOD database.

Run Time: The average run time of the proposed method and currently top-performance methods on the SOD database are presented in Table I based on a machine with Intel(R) Core(TM) i7-3770 3.4GHz CPU and 32GB RAM. Based on the current implementation without code optimization, the proposed algorithm takes about 5 seconds to process an image (where the most time-consuming part is the multi-scale superpixel segmentation), which costs less time than the state-of-the-art salient object detection models (e.g., RA [29], CA [13], LR [34] and SVO [41]).

F. Further Improvement of State-of-the-Art Saliency Maps

As discussed in Section VI-C, the propagated reconstruction error is more accurate than the non-propagated one (see Figure 9(a)), which validates the effectiveness of the context-based error propagation in this work. Intuitively, the proposed saliency propagation mechanism (Section III-D) may also further improve the performance of other state-of-the-art methods by smoothing saliency among image contexts.

We implement the context-based propagation on seven state-of-the-art models including DW [18], RA [29], CA [13], FT [16], SR [39], GB [38] and IT [11]. In detail, we first calculate the mean pixel saliency of each SLIC superpixel as the initial saliency for the corresponding segment. Then we propagate the saliency of each image segment by taking the contextual information into consideration using Eq. 5-6. The mean propagated saliency value for each pixel is finally obtained by the multi-scale saliency integration from Eq. 7. We evaluate the propagated saliency map of each method and compare it with the original result in Figure 12(d). The propagated saliency results (solid lines) achieve better performance of precision-recall curves than the original non-propagated ones (dashed lines), which attributes to the significant contribution of image contexts in the propagation mechanism (see Eq. 5).

VII. Conclusions

In this paper, we present a saliency detection algorithm via dense and sparse reconstruction based on the background templates. Considering the prominent contribution of color compactness for saliency detection, we propose a compactness weighted reconstruction error to better measure saliency. A context-based propagation mechanism is designed to propagate the reconstruction errors through local context obtained by K-means clustering. The pixel-level saliency is finally computed by an integration of multi-scale reconstruction errors followed by an object-biased Gaussian refinement. To combine the two saliency maps via dense and sparse reconstruction, we introduce a Bayesian integration method which performs better than the conventional integration strategy. Experimental results show the performance improvement of the proposed method compared to twenty-four state-of-the-art models. Our saliency map can well suppress the background while uniformly highlight the foreground objects.

References

[1] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Rev. Neurosci.*, vol. 2, no. 3, pp. 194–201, 2001.

[2] J. M. Wolfe and T. S. Horowitz, "What attributes guide the deployment of visual attention and how do they do it?" *Nature Rev. Neurosci.*, vol. 5, no. 6, pp. 495–501, 2004.

[3] J. Han, K. N. Ngan, M. Li, and H.-J. Zhang, "Unsupervised extraction of visual attention objects in color images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 1, pp. 141–145, Jan. 2006.

[4] C. Siagian and L. Itti, "Rapid biologically-inspired scene classification using features shared with visual attention," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 300–312, Feb. 2007.

[5] C. M. Privitera and L. W. Stark, "Algorithms for defining visual regions-of-interest: Comparison with eye fixations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 9, pp. 970–982, Sep. 2000.

[6] D. Gao, S. Han, and N. Vasconcelos, "Discriminant saliency, the detection of suspicious coincidences, and applications to visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 6, pp. 989–1005, Jun. 2009.

[7] U. Rutishauser, D. Walther, C. Koch, and P. Perona, "Is bottom-up attention useful for object recognition?" in *Proc. CVPR*, 2004, pp. 37–44.

[8] L. Marchesotti, C. Cifarelli, and G. Csurka, "A framework for visual saliency detection with applications to image thumbnailing," in *Proc. ICCV*, 2009, pp. 2232–2239.

[9] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 185–198, Jan. 2010.

[10] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Trans. Image Process.*, vol. 13, no. 10, pp. 1304–1318, Oct. 2004.

[11] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[12] D. A. Klein and S. Frintrop, "Center-surround divergence of feature statistics for salient object detection," in *Proc. ICCV*, 2011, pp. 2214–2219.

[13] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *Proc. CVPR*, 2010, pp. 2376–2383.

[14] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li, "Automatic salient object segmentation based on context and shape prior," in *Proc. BMVC*, 2011, pp. 1–12.

[15] A. Borji and L. Itti, "Exploiting local and global patch rarities for saliency detection," in *Proc. CVPR*, 2012, pp. 478–485.

[16] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. CVPR*, 2009, pp. 1597–1604.

[17] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Proc. CVPR*, 2011, pp. 409–416.

[18] L. Duan, C. Wu, J. Miao, L. Qing, and Y. Fu, "Visual saliency detection by spatially weighted dissimilarity," in *Proc. CVPR*, 2011, pp. 473–480.

[19] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. CVPR*, 2012, pp. 733–740.

[20] S. Lu, V. Mahadevan, and N. Vasconcelos, "Learning optimal seeds for diffusion-based salient object detection," in *Proc. CVPR*, 2014, pp. 2790–2797.

[21] R. Liu, J. Cao, Z. Lin, and S. Shan, "Adaptive partial differential equation learning for visual saliency detection," in *Proc. CVPR*, 2014, pp. 3866–3873.

[22] R. Zhao, W. Ouyang, H. Li, and X. Wang, "Saliency detection by multi-context deep learning," in *Proc. CVPR*, 2015, pp. 1265–1274.

[23] G. Li and Y. Yu, "Visual saliency based on multiscale deep features," in *Proc. CVPR*, 2015, pp. 5455–5463.

[24] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, "SUN: A Bayesian framework for saliency using natural statistics," *J. Vis.*, vol. 8, no. 7, p. 32, 2008.

[25] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Proc. ECCV*, 2012, pp. 29–42.

[26] A. Borji, D. N. Sihite, and L. Itti, "Salient object detection: A benchmark," in *Proc. ECCV*, 2012, pp. 414–429.

[27] X. Hou and L. Zhang, "Dynamic visual attention: Searching for coding length increments," in *Proc. NIPS*, vol. 21. 2008, pp. 681–688.

[28] W. Wang, Y. Wang, Q. Huang, and W. Gao, "Measuring visual saliency by site entropy rate," in *Proc. CVPR*, 2010, pp. 2368–2375.

[29] E. Rahtu, J. Kannala, M. Salo, and J. Heikkilä, "Segmenting salient objects from images and videos," in *Proc. ECCV*, 2010, pp. 366–379.

[30] Y. Xie and H. Lu, "Visual saliency detection based on Bayesian model," in *Proc. ICIP*, 2011, pp. 645–648.

[31] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels," EPFL, Lausanne, Switzerland, Tech. Rep. 149300, 2010.

[32] W. Zhong, H. Lu, and M.-H. Yang, "Robust object tracking via sparsity-based collaborative model," in *Proc. CVPR*, 2012, pp. 1838–1845.

[33] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[34] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proc. CVPR*, 2012, pp. 853–860.

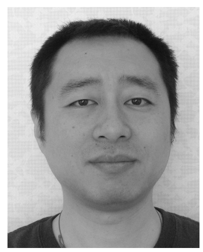
[35] Y. Xie, H. Lu, and M.-H. Yang, "Bayesian saliency via low and mid level cues," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1689–1698, May 2013.
[36] Y.-F. Ma and H.-J. Zhang, "Contrast-based image attention analysis by using fuzzy growing," in *Proc. ACM Multimedia*, 2003, pp. 374–381.
[37] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proc. ACM Multimedia*, 2006, pp. 815–824.
[38] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. NIPS*, 2006, pp. 1–8.
[39] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. CVPR*, 2007, pp. 1–8.
[40] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk, "Salient region detection and segmentation," in *Proc. ICVS*, 2008, pp. 66–75.
[41] K.-Y. Chang, T.-L. Liu, H.-T. Chen, and S.-H. Lai, "Fusing generic objectness and visual saliency for salient object detection," in *Proc. ICCV*, 2011, pp. 914–921.
[42] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. Torr, and S.-M. Hu, "Salient object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 2, no. 3, p. 9, Jun. 2011.
[43] M.-M. Cheng, J. Warrell, W.-Y. Lin, S. Zheng, V. Vineet, and N. Crook, "Efficient salient region detection with soft image abstraction," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1529–1536.
[44] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proc. CVPR*, 2013, pp. 1155–1162.
[45] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proc. CVPR*, 2014, pp. 2814–2821.
[46] J. Kim, D. Han, Y.-W. Tai, and J. Kim, "Salient region detection via high-dimensional color transform," in *Proc. CVPR*, Jun. 2014, pp. 883–890.
[47] N. Tong, H. Lu, and M.-H. Yang, "Salient object detection via bootstrap learning," in *Proc. CVPR*, 2015, pp. 1884–1892.
[48] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *Proc. ICCV*, 2013, pp. 2976–2983.
[49] T. Liu, J. Sun, N.-N. Zheng, X. Tang, and H.-Y. Shum, "Learning to detect a salient object," in *Proc. CVPR*, 2007, pp. 1–8.
[50] V. Movahedi and J. H. Elder, "Design and perceptual validation of performance measures for salient object segmentation," in *Proc. CVPRW*, 2010, pp. 49–56.

Huchuan Lu (SM'12) received the M.Sc. degree in signal and information processing and the Ph.D. degree in system engineering from the Dalian University of Technology (DUT), Dalian, China, in 1998 and 2008, respectively. He joined DUT in 1998, as a Faculty Member, where he is currently a Full Professor with the School of Information and Communication Engineering. His current research interests include computer vision and pattern recognition with a focus on visual tracking, saliency detection, and segmentation. He is a member of the Association for Computing Machinery and an Associate Editor of the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS.

Xiaohui Li received the B.E. degree in electronic information engineering and the M.S. degree in signal and information processing from the Dalian University of Technology, Dalian, China, in 2011 and 2014, respectively. Her research interest is in saliency detection.

Lihe Zhang received the Ph.D. degree in signal and information processing from the Beijing University of Posts and Telecommunications, Beijing, China, in 2004. He is currently an Associate Professor with the School of Information and Communication Engineering, Dalian University of Technology.

His current research interests include computer vision and pattern recognition.



Xiang Ruan received the B.E. degree from Shanghai Jiao Tong University, Shanghai China, in 1997, and the M.E and Ph.D. degrees from Osaka City University, Osaka, Japan, in 2001 and 2004, respectively. He is currently a Research Engineer with OMRON Corporation, Kusatsu, Japan. His current research interests include computer vision, machine learning, and image processing.

Ming-Hsuan Yang (SM'06) received the Ph.D. degree in computer science from the University of Illinois at Urbana–Champaign, Urbana, in 2000. He was a Senior Research Scientist with the Honda Research Institute, working on vision problems related to humanoid robots. He is currently an Assistant Professor with the Department of Electrical Engineering and Computer Science, University of California (UC), Merced. He has co-authored the book *Face Detection and Gesture Recognition for Human–Computer Interaction* (Kluwer, 2001) and edited the special issue on face recognition for *Computer Vision and Image Understanding* in 2003. He is a Senior Member of the Association for Computing Machinery. He was a recipient of the Ray Ozzie fellowship for his research work in 1999. He received the Natural Science Foundation CAREER Award in 2012, the Campus Wide Senate Award for Distinguished Early Career Research at UC in 2011, and the Google Faculty Award in 2009. He edited a special issue on real world face recognition for the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. He serves as an Area Chair for the IEEE International Conference on Computer Vision in 2011, the IEEE Conference on Computer Vision and Pattern Recognition in 2008 and 2009, and the Asian Conference on Computer in 2009, 2010, and 2012. He served as an Associate Editor of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE from 2007 to 2011, and the *Image and Vision Computing*.