300-FPS Salient Object Detection via Minimum Directional Contrast

Xiaoming Huang and Yu-Jin Zhang, *Senior Member, IEEE*

***Abstract*— Global contrast considers the color difference between a target region or pixel and the rest of the image. It is frequently used to measure the saliency of the region or pixel. In previous global contrast-based methods, saliency is usually measured by the sum of contrast from the entire image. We find that the spatial distribution of contrast is one important cue of saliency that is neglected by previous works. Foreground pixel usually has high contrast from all directions, since it is surrounded by the background. Background pixel often shows low contrast in at least one direction, as it has to connect to the background. Motivated by this intuition, we first compute directional contrast from different directions for each pixel, and propose minimum directional contrast (MDC) as raw saliency metric. Then an O(1) computation of MDC using integral image is proposed. It takes only 1.5 ms for an input image of the QVGA resolution. In saliency post-processing, we use marker-based watershed algorithm to estimate each pixel as foreground or background, followed by one linear function to highlight or suppress its saliency. Performance evaluation is carried on four public data sets. The proposed method significantly outperforms other global contrast-based methods, and achieves comparable or better performance than the state-of-the-art methods. The proposed method runs at 300 FPS and shows six times improvement in runtime over the state-of-the-art methods.**

***Index Terms*— Saliency detection, minimum directional contrast, directional contrast, distribution, saliency enhancement.**

I. Introduction

SALIENCY detection aims to simulate the human visual system for detecting pixels or regions that are most attractive. Saliency detection can be regarded as one key preprocessing step in many applications such as image classification [1], [2], object detection and recognition [3], [4], image compression [5], and image segmentation [6], [7]. Due to the increasing need of mobile applications and the growing scale of datasets, a desirable salient object detection method should not only provides high quality saliency maps, but also has high computational efficiency. In this paper, we address both the accuracy and efficiency for salient object detection.

Manuscript received August 26, 2016; revised March 23, 2017 and April 28, 2017; accepted May 15, 2017. Date of publication June 1, 2017; date of current version June 23, 2017. This work was supported by the National Nature Science Foundation under Grant 61171118, Grant 61673234, and Grant U1636124. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Guoliang Fan. *(Corresponding author: Xiaoming Huang.)*

The authors are with the Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China, and also with the Department of Electronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: huangxm14@mails.tsinghua.edu.cn; zhang-yj@tsinghua.edu.cn).

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.2017.2710636

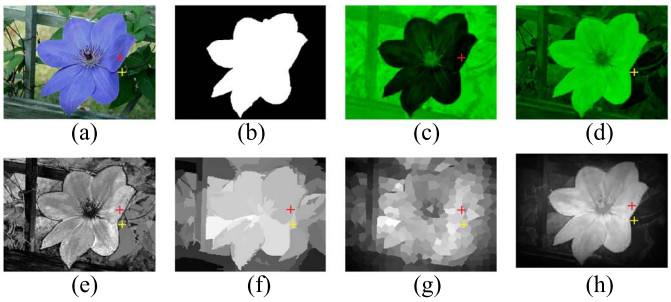

Fig. 1. Motivation of proposed method. (a) Input image with two target pixels at red and yellow crossing. (b) Saliency ground truth. (c)(d) Contrast from entire image to two target pixels differ greatly in spatial distribution. (e)(f)(g) Raw saliency of HC[18], RC [18] and SF [20]. (h) Our raw saliency.

Motivated by increasing application demand, a number of algorithms [8]–[25] have been proposed for saliency detection. Recently, bottom-up methods have made significant progress. These methods adopt low-level visual cues such as color, intensity, texture, etc. One representative research field is based on the global contrast which consider difference between image regions and entire image. The definition of contrast can leverage different types of image features, such as color, edges, gradients, etc. In this paper, we use color contrast which is widely researched in recent years.

Many global contrast based saliency detection methods [17], [18], [20] have been proposed in recent years. In these previous global contrast based methods, saliency is usually measured by the sum of contrast from the entire image [18], [20], or defined as contrast with average image color [17]. Spatial distribution of contrast is neglected by previous works. We present an example to demonstrate that the spatial distribution of contrast can play a critical role for distinguishing salient objects:

Fig. 1. (a) is one source image with two marked target pixels (one foreground pixel at the red crossing and one background pixel at the yellow crossing). The contrast from the entire image to these two target pixels using Euclidean distance in color space are shown in Fig. 1. (c) and Fig. 1. (d), more green indicates higher contrast. The sum of contrast in Fig. 1. (c) and that in Fig. 1. (d) both have high value (about 10% difference), but the spatial distributions of contrast differ greatly. If the target pixel is regarded as the center of view, foreground pixel has high contrast from almost all directions in Fig. 1. (c), while background pixel shows low contrast from some directions (bottom right) and high contrast from other directions in Fig. 1. (d). If saliency is simply defined as the sum of contrast from the entire image (HC [18],

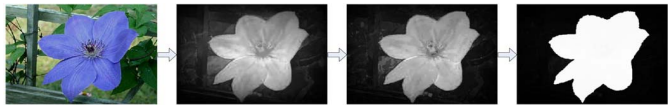

Fig. 2. Main flowchart of our work. (a) Input image. (b) Raw saliency. (c) Smoothing. (d) Enhancement.

RC [18], SF [20]), these two target pixels both show high saliency as shown in Fig. 1. (e)-(g). In order to distinguish foreground pixels from background pixels, we consider the spatial distribution of contrast.

For each pixel, we compute *directional contrast* from different directions, and propose *minimum directional contrast* (*MDC*) as the raw saliency metric. Since a foreground pixel is usually surrounded by the background, it often has high contrast from all directions, and thus shows high *MDC*. On the contrary, the *MDC* of a background pixel usually is small, as it has to connect to the background through at least one direction. Foreground pixels can be distinguished from background pixels in the proposed work. Fig. 1. (h) shows the proposed raw saliency map. The foreground pixel (red crossing) has a significantly higher saliency value than the background pixel (yellow cross). This result resembles closely to the ground-truth shown in Fig. 1.(b). More examples can be found in Fig. 7.

Fig. 2 shows the pipeline of our proposed method. Given an input image as in Fig. 2. (a), we first compute its pixel-wise *MDC* as the raw saliency map shown in Fig. 2. (b). This map is further smoothed and enhanced as in Fig. 2. (c)(d). For *MDC*, we propose a highly efficient implementation using integral image. Our implementation has a time complexity of $O(1)$ per pixel and only requires 1.5ms for an input image of QVGA resolution. In saliency smoothing, we adopt *boundary connectivity* prior [37]. In saliency enhancement, each pixel is initially marked as *foreground*, *background*, or *unknown* according to its smoothed saliency value. Then a fast marker-based watershed algorithm [39] will label each pixel as *foreground*, *background* or *watershed*. The saliency of *foreground* pixel will be raised to highlight the object, and saliency of *background* pixel will be suppressed.

Performance evaluation on four public datasets shows that the proposed method significantly outperforms other global contrast based method, and achieves comparable or better performance than the state-of-the-art methods. Due to fast raw saliency computation and enhancement, the proposed method runs at 300 FPS. In state-of-the-art methods, the highest speed performance is about 50 FPS (MB+ [40]), our method shows 6 times speed improvement.

The main contributions of this work include:

1) We propose a novel saliency metric *MDC* that leverages minimum contrast from all directions. *MDC* significantly outperforms previous global contrast based methods.
2) We present a highly efficient $O(1)$ computation of pixel-wise *MDC* using integral image. Our implementation only requires 1.5ms for an input image of QVGA resolution.
3) We propose a simple and effective saliency enhancement step. Our method generates a binary mask of salient objects to highlight foreground and suppress background.

The remainder of this paper is organized as follows. Section II contains a review of related saliency detection methods. Proposed saliency detection method and experimental results are described in Section III and Section IV, respectively. Finally, Section V presents the conclusion and future work.

II. Related Work

Saliency detection methods can be divided into two categories: bottom-up (stimulus-driven) methods [8]–[23] and top-down (goal-driven) methods [24], [25]. According to the type of visual cue, bottom-up methods can be based on contrast, compactness, background, etc. The main advantage of such methods is fast and need not specific prior knowledge. Top-down methods learn high-level knowledge from training examples with manually labeled ground truth. Some priors including *center*, *boundary*, *connectivity*, *objectness* are widely adopted in saliency detection. A comprehensive survey of saliency detection can be found in [26] and [27], and a quantitative comparison of different methods was provided in [28].

Many contrast based saliency detection methods have been proposed in recent years. These methods can be roughly divided into local methods [14], [16], [30] and global methods [17], [18], [20]. Local contrast based methods mainly investigate the rarity of the pixel or region with respect to small local neighborhoods, whereas global contrast based methods consider contrast relationships over the entire image. In general, global contrast based methods achieve better performance than local methods. We focus on global contrast based method.

Achanta *et al*. [17] present a frequency tuned algorithm that defines pixel saliency as its color difference from the average image color. This simple but very fast approach usually failed in complex natural images.

A pixel wise global contrast method HC is proposed in [18], the saliency is defined as its color distance to all other pixels. In order to improve speed performance, author quantizes each channel into 12 levels, choose more frequently used colors to generate a color histogram and represent the entire image, saliency can be efficiently calculated on this histogram. The pixels with the same color have the same saliency in this algorithm, spatial relationship is ignored.

A region based saliency algorithm RC is also introduced in [18]. The input image is first segmented into M regions r_i ($i = 1 .. M$), saliency value of the region r_i can be measured as the sum of contrast from all other regions with spatial weight:

$$S_i^{RC} = \sum_{j=1}^{M} e^{-\frac{D(r_i, r_j)}{\delta^2}} w(r_j) D_r(r_i, r_j) \tag{1}$$

where $D(r_i, r_j)$ and $D_r(r_i, r_j)$ are the spatial and color distance between two regions,δ controls the strength of spatial distance weight, $w(r_j)$ is the number of pixels in the region r_j. The main contribution of this work is considering spatial weighted global contrast from the entire image.

Perazzi *et al.* [20] adopt global contrast and spatial distribution of color as saliency metric. In this paper, global

contrast also named as uniqueness is defined as Euclidean color distance to entire image with Gaussian spatial weight:

$$S_i^{SF} = \sum_{j=1}^{M} ||c_i - c_j||^2 w_{ij} = c_i^2 \sum_{j=1}^{M} w_{ij} - 2c_i \sum_{j=1}^{M} c_j w_{ij} + \sum_{j=1}^{M} c_j^2 w_{ij} \quad (2)$$

where c_i is the average color of the region or pixel r_i. Since w_{ij} is a Gaussian spatial weight function, using Gaussian blurring kernel on c_i and c_i^2, the computational complexity for each pixel can be reduced from $O(M)$ to $O(1)$. Nevertheless, Gaussian blurring still needs high time expenditure.

In above global contrast based work, saliency is measured by the sum of contrast from the entire image, some background regions also show high saliency due to large contrast contribution from the object. An example is shown in Fig. 1. (e)(f)(g), the background pixel at yellow crossing and the foreground pixel at red crossing demonstrate comparable saliency. In this paper, we consider spatial distribution of contrast, and adopt *MDC* as saliency metric, contrast contribution from object to background can be suppressed.

Besides, RC [18] adopt graph-based segmentation [34] which needs 60ms for each image, expensive time cost for region segmentation limits speed performance.

The *boundary* and *connectivity* priors [21] are also shown to be effective in salient object detection. These priors assume that background regions are usually connected to the image boundary. Geodesic distance [21] and Minimum Barrier Distance (MBD) [46] are widely used distance transform to measure regions *connectivity* to the image boundary. In [40], authors propose one approximate MBD implementation with raster scan. Three passes scan on each color channel also limit the speed performance. Another approximation method on minimum spanning tree (MST) in presented in [45], additional time cost to build tree leads to worse speed performance than [40]. Both these two algorithms show that MBD is more robust to noise and blur than the geodesic distance.

Recently, deep learning has achieved great successes in many computer vision tasks. Some researchers have already applied deep neural networks to saliency detection. Wang *et al.* [47] use a CNN to predict saliency for each pixel in local context, then refine the saliency on object proposal over the global view. Zhao *et al.* [48] consider global and local context simultaneously in a multi-context CNN, then combine them to predict saliency. Deep learning based works both achieve better performance with very low speed.

III. Salient Object Detection Method

In this section, we present an efficient salient object detection method. We first propose one raw saliency metric *MDC* which considers spatial distribution of contrast. Next, an $O(1)$ implementation of *MDC* is proposed. In post processing, saliency smoothing and enhancement will be introduced.

A. Minimum Directional Contrast (MDC)

To measure the saliency of a region or pixel, contrast is the most frequently used feature. Global contrast is widely studied in salient object detection which considers the color difference between the target region or pixel and the entire image.

Global contrast can be calculated at pixel or region level. Region level method needs higher time cost for image segmentation. In RC [18], graph-based segmentation [34] needs about 60ms for each image on MSRA-10K dataset [18] [33]. In SLIC superpixel segmentation based method SF [20], segmentation needs about 110ms. In order to achieve higher speed performance, we adopt pixel level saliency detection which needs not time cost for region segmentation.

As discussed in introduction part, in previous global contrast based methods, saliency is simply measured by the sum of contrast from the entire image [18], [20], or defined as contrast with the average image color [17]. Spatial distribution of contrast is neglected.

In this paper, we further detail analyze contrast from different spatial directions. If the target pixel i is regarded as the center of view, the entire image can be divided into several regions based on their location w.r.t. pixel i, i.e., top left (*TL*), top right (*TR*), bottom left (*BL*), bottom right (*BR*). *Directional contrast* (*DC*) from each region Ω can be calculated as:

$$DC_{i,\Omega} = \sqrt{\sum_{j \in \Omega} \sum_{ch=1}^{K} (I_{i,ch} - I_{j,ch})^2} \quad (3)$$

where I denotes one input image with K color channels in CIE-Lab color space. Fig. 3 (a) is one input image, two target pixels are shown in Fig. 3 (b), one foreground pixel at the red crossing in the top row, and one background pixel at the yellow crossing in the bottom row. The entire image is simply divided into four regions by red or yellow line. *DC* result of two target pixels is shown in Fig. 3 (c).

From *DC* result in Fig. 3 (c), we can find the distribution of *DC* differ greatly between foreground and background pixel. In the top row, foreground pixel shows high *DC* in almost all directions, *minimum directional contrast* (*MDC*) still has high value. In the bottom row, background pixel demonstrates very low *DC* in direction bottom right and high *DC* in other directions, *MDC* will show very low value. More generally, since foreground pixel is usually surrounded by the background, it often has high contrast from all directions, *MDC* shows high value. On the contrary, the *MDC* of a background pixel is usually small, as it has to connect to the background through one of the directions. This suggests that we can define *MDC* which means the minimum contrast from all directions as the raw saliency metric:

$$S(i) = \min_{\Omega \in TL,TR,BL,BR} DC_{i,\Omega} = \sqrt{\min_{\Omega \in TL,TR,BL,BR} (\sum_{j \in \Omega} \sum_{ch=1}^{K} (I_{i,ch} - I_{j,ch})^2} \quad (4)$$

The *MDC* of two target pixels are shown in Fig. 3 (c), the foreground pixel at red crossing shows obviously higher *MDC* than the background pixel at yellow crossing. The *MDC* based raw saliency of all pixels are shown in Fig. 3 (d).

In previous global contrast based methods (HC [18], RC [18], SF [20]), saliency is simply defined as the sum of

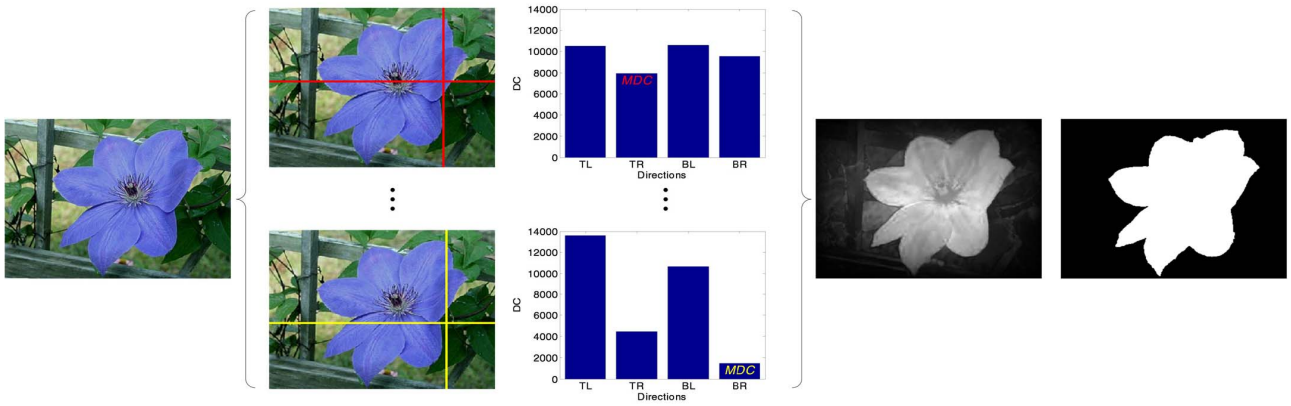

Fig. 3. Raw saliency computation. The input image (a) is simply divided into four directions as shown in (b) for each pixel (top row is one foreground pixel at red crossing, and bottom row is one background pixel at yellow crossing). DC from four directions and MDC is shown in (c). MDC of all pixels is shown in (d). (a) Input image. (b) Four direction. (c) *DC* & *MDC*. (d) Raw saliency. (e) Ground truth.

contrast from entire image, the background pixel at yellow crossing also has high contrast from object in some directions (*TL*, *BL*), sum operation can lead to high saliency which is shown in Fig. 1. (e)-(g). Our method adopts minimum operation obviously can suppress contrast contribution from object to background.

Raw saliency evaluation between proposed *MDC* metric and other global contrast based methods is presented at Sec.IV-D, including quality comparison in Fig. 7, and quantitative evaluation in Fig. 8. Our method significantly outperforms previous global contrast methods.

B. O(1) MDC Computation

In order to compute *MDC* of each pixel, we need to compute contrast from four directions with the equation $\sum_{j\in\Omega}\sum_{ch=1}^{K}(I_{i,ch}-I_{j,ch})^2$. If we naively compute this equation, computation complexity for each pixel is $O(N)$ where N is pixel number of entire image, time cost is too expensive even for medium sized images. We show that computational complexity can be optimized to $O(1)$ as following:

$$\sum_{j\in\Omega}\sum_{ch=1}^{K}(I_{i,ch}-I_{j,ch})^2 = \sum_{j\in\Omega}\sum_{ch=1}^{K}I_{j,ch}^2 - 2\sum_{ch=1}^{K}(\sum_{j\in\Omega}I_{j,ch})I_{i,ch} + |\Omega|\sum_{ch=1}^{K}I_{i,ch}^2 \quad (5)$$

where $|\Omega|$ means the number of pixels in direction Ω. In (5), $\sum_{j\in\Omega}\sum_{ch=1}^{K}I_{j,ch}^2$ and $\sum_{j\in\Omega}I_{j,ch}$ can be calculated fast by integral image [38] with $O(1)$ computation complexity per pixel.

Due to $O(1)$ implementation, pixel-wise *MDC* only needs 1.5ms for an input image of QVGA resolution. The hardware platform for time cost analysis is introduced in Sec.IV-F.

C. Saliency Smoothing

Our fast saliency smoothing method with *boundary connectivity* [37] prior is described in this section.

Similar to color reduction introduced in RC [18], we first quantize each color channel to L different values, thus the number of colors reduced from 256^3 to L^3. The saliency of pixels with same quantized color will be smoothed.

In [37], authors hold that background regions are much more connected to image boundaries than foreground regions, and propose *boundary connectivity* to measure background. In this work, we calculate this background measure at quantized color level. Let R_q denotes pixels with same quantized color q, we calculate the *boundary connectivity* of R_q as following:

$$BndCon(R_q) = \frac{\sum_{j\in R_q}\delta(j\ is\ boundary)}{\sqrt{|R_q|}} \quad (6)$$

where δ () function is 1 for pixel on the image boundary and 0 otherwise, $|R_q|$ means the number of pixels. Then *boundary connectivity* weighted average saliency of R_q is defined as:

$$S_{aver}(R_q) = \underset{j\in R_q}{average}\, S(j)\cdot e^{-w\cdot BndCon(R_q)} \quad (7)$$

where w is the parameter to control *boundary connectivity* weight. Final saliency is a fusion of the average saliency and raw saliency as following:

$$S_{smooth}(i) = 0.5\cdot S(i) + 0.5\cdot S_{aver}(R_q) \quad (8)$$

where pixel i has quantized color q. Saliency smoothing examples are shown in Fig. 4 (c), a little artifacts introduced by the quantization will be suppressed in saliency enhancement.

D. Saliency Enhancement

In order to increase the contrast between foreground and background regions, we propose one simple yet effective watershed based saliency enhancement method. We firstly use OTSU [41] method on smoothed saliency S_{smooth}, and get binarization threshold T. Then some pixels are marked as *foreground* or *background* using reliable conditions, and others marked as *unknown* as following:

$$M(i) = \begin{cases} foreground & S_{smooth}(i) > (1+\theta)T \\ background & S_{smooth}(i) < (1-\theta)T \\ unknown & others \end{cases} \quad (9)$$

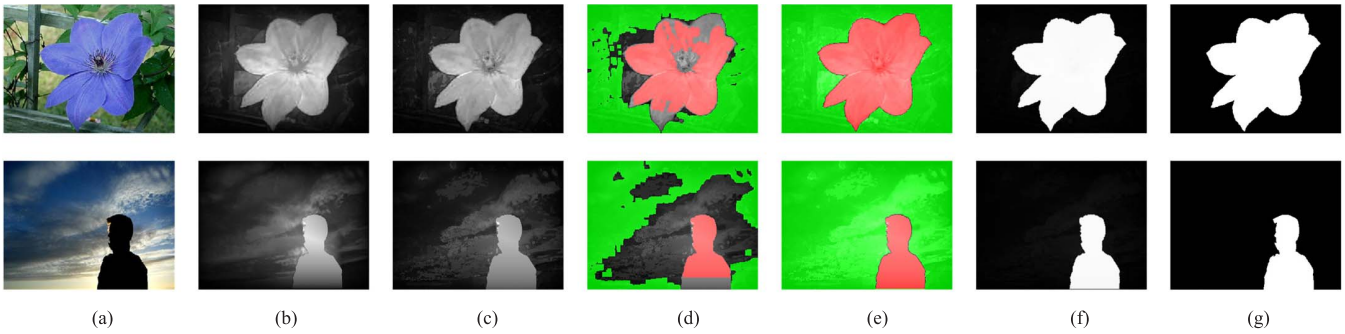

Fig. 4. Saliency post-processing. Raw and smoothed saliency are shown in (b) (c). Each pixel is initialy marked as *foreground, background* or *unknown* (shown with red, green or no mask) in (d), then apply marker-based watershed algorithm [39] to estimate *unknown* pixel shown in (e), final enhanced saliency is shown in (f). (a) Input image. (b) Raw saliency. (c) Smoothed saliency. (d) Initial marker. (e) Final marker. (f) Enhanced saliency. (g) Ground truth.

where θ is a parameter to control initial mark region. In order to remove small noises in the initial marker, we use a morphological opening filter with adaptive kernel size:

$$k = 0.05\sqrt{F} \tag{10}$$

where F is the number of initial *foreground* marker. Then we apply marker-based watershed algorithm [39], all pixels will be labeled as *foreground*, *background* or *watershed*. The saliency of each pixel is assigned as following:

$$S_{enhance}(i) = \begin{cases} 1-\alpha(1-S_{smooth}(i)) & i\ is\ foreground \\ \beta S_{smooth}(i) & i\ is\ background \\ S_{smooth}(i) & i\ is\ watershed \end{cases} \tag{11}$$

where α and β is parameter ($\in [0, 1]$) to control enhancement, the saliency of foreground and background pixel will be mapped to $[1-\alpha, 1]$ and $[0, \beta]$ respectively. Lower value of α and β means foreground pixel will be assigned higher saliency, background pixel will be assigned lower saliency. In Fig. 4 we show some saliency enhancement examples. Fig. 4 (a) and Fig. 4 (c) are source image and smoothed saliency, initial mark using conditions (9)(10) is shown in Fig. 4 (d), *foreground*, *background* and *unknown* mark are shown with red, green and no mask respectively. Fig. 4. (e) is the result of marker-based watershed algorithm [39]. Final enhanced saliency using (11) is shown in Fig. 4. (f).

IV. Experimental Results

In this section, we present an evaluation of the proposed method against several methods on four public datasets. Parameter setting, ablation study, performance and speed evaluation will be discussed.

Our source code and one demo program is provided at https://github.com/huangxm14-thu/SaliencyMDC.

A. Datasets

We evaluate the proposed method on four public datasets. The first MSRA-10K dataset [18], [33] consists of 10000 images, each of which has an unambiguous salient object. The second ECSSD [10] contains more salient objects under complex scenes and some images come from the challenging Berkeley-300 dataset. The third DUTOMRON dataset [23] contains 5166 challenging images. The final PASCAL-S dataset [43] constructed on the validation set of recent PASCAL VOC segmentation challenge. It contains 850 natural images with multiple complex objects and cluttered backgrounds.

MB+ [40] is one recent work with highest speed performance in the state-of-the-arts. In this method, all input images are first resized so that the maximum dimension is 300 pixels. We adopt same resize operation for fair comparison.

B. Evaluation Metrics

We evaluate the performance by precision-recall (PR) curve, F-Measure and MAE (mean absolute error).

1). Precision-recall (PR) curve. For a saliency map S, we can convert it to a binary mask M with threshold which changes from 0 to 255 and compute *Precision* and *Recall* by comparing M with ground-truth G:

$$Precision = \frac{|M \cap G|}{|M|}, \quad Recall = \frac{|M \cap G|}{|G|} \tag{12}$$

where $|\cdot|$ means the number of non-zero entries in the mask.

2). F-measure is the overall performance measurement computed by the weighted harmonic of precision and recall at adaptive threshold which is defined as twice the mean saliency:

$$F_\beta = \frac{(1+\beta^2)\,\mathrm{Pr}\,ecision \times \mathrm{R}ecall}{\beta^2\,\mathrm{Pr}\,ecision + \mathrm{R}ecall} \tag{13}$$

where β^2 is set to 0.3 to raise more importance on precision.

3). Mean absolute error (MAE) between the saliency map S and the ground-truth G can be calculated as:

$$MAE = \frac{1}{|I|}\sum_{i \in I} |S(i) - G(i)| \tag{14}$$

C. Parameter Setting

We carried out a series of experiments to investigate the influence of parameters on the saliency detection. The experiments are carried on MSRA-10K dataset.

In this work, saliency metric *MDC* is defined as minimum contrast from all directions, the number of directions is one important parameter. Taking into account the computational complexity, we mainly compare *MDC* calculation in four and eight directions. With the increase of direction number,

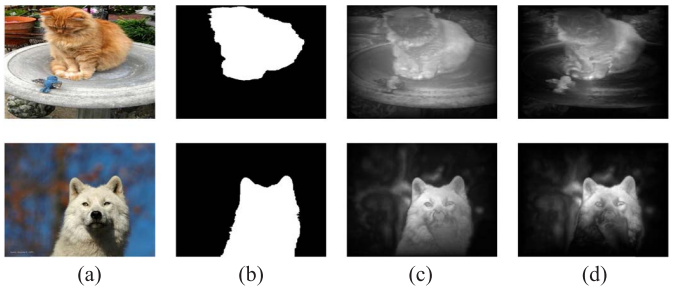

Fig. 5. *MDC* computation in different directions. Both background and foreground near to boundary are easier to show lower *MDC* in 8 directions. (a) input image. (b) ground truth. (c) 4 directions. (d) 8 directions.

background region is easier to get lower *MDC*, while the object near to image boundary also shows lower *MDC*. Two examples are shown in Fig. 5. In top row, background shows lower *MDC* in eight directions. In the bottom row, the dog near to image boundary shows lower *MDC* in eight directions. Quantitative comparison of raw saliency on MSRA-10K dataset is presented in top row of Fig. 6. Computation in eight directions demonstrates better MAE but shows worse PR curve and F-Measure. What's more, calculation in eight directions needs twice time cost. Considering accuracy and time factor, *MDC* calculation from four directions in adopted in our work.

There are five parameters in saliency post-processing: L which decides color reduction for saliency smoothing in Sec.III-C, ω in (7) which control *boundary connectivity* weight in saliency smoothing, θ in (9) which determine initial mark region in saliency enhancement, α and β in (11) which control saliency enhancement.

The second row of Fig. 6 is evaluation result of parameter L. Our results are insensitive to this parameter when L is between 16 to 32, we select 24 as the default value.

The third row of Fig. 6 is evaluation result of parameter ω. Higher ω leads to better performance than $\omega = 0$, this result proves that *boundary connectivity* weight is helpful in saliency smoothing. When ω is greater than 3.0, the performance has no evident improvement. We select 3.0 as the optional value.

The fourth row of Fig. 6 is evaluation result of parameter θ. This parameter will initially mark each pixel as *foreground*, *background* or *unknown* with (9). Too small or large parameter will result in too small or large *unknown* mark. $\theta = 0.6$ shows the best performance.

The bottom two rows of Fig. 6. are evaluation result of parameter α and β. With (11) of paper, lower value of α means foreground pixel will be assigned higher saliency, and lower value of β indicates background pixel will be assigned lower saliency. In general, lower α and β results in lower MAE and higher F-Measure. But too lower α(less than 0.5) and β (less than 0.2) will leads to F-Measure decline slightly.

According to above discussion, we set the optional parameters as: $L = 24$, $\omega = 3.0$, $\theta = 0.6$, $\alpha = 0.05$, $\beta = 0.2$.

D. Ablation Study

We evaluate the effectiveness of proposed contrast method and post-processing in this section.

1) Contrast Method Comparison: The most important contribution of this paper is we propose *MDC* as saliency metric.

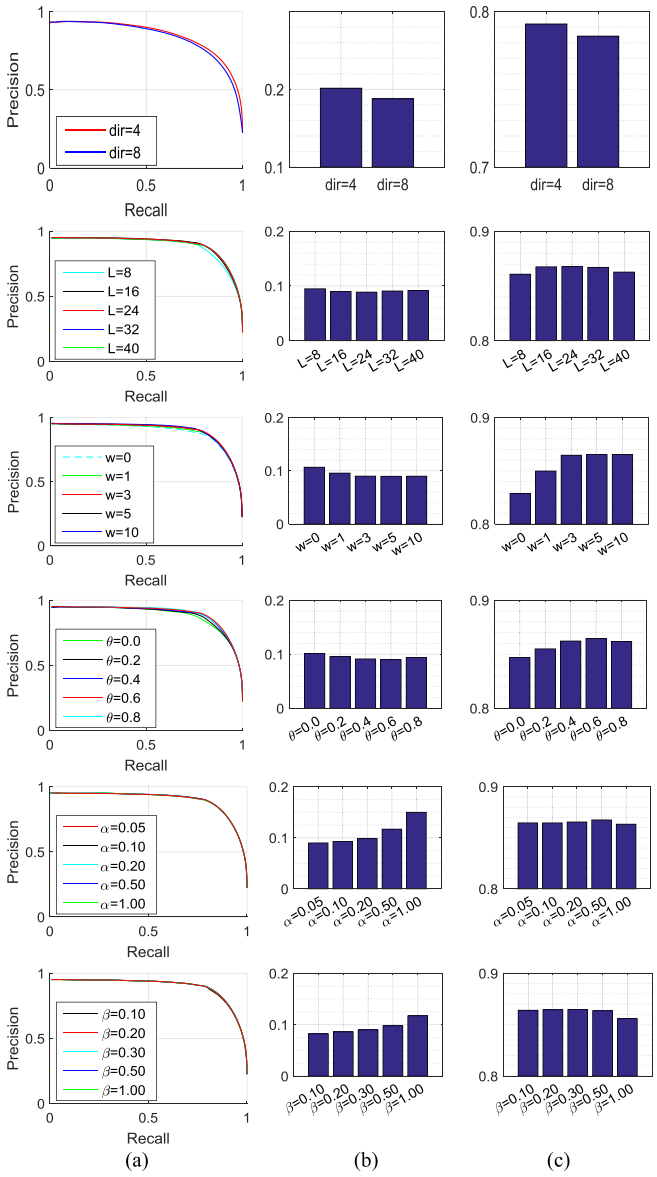

Fig. 6. Parameter analysis. Top row is parameter evaluation of *MDC*, other rows are parameter comparison of saliency post-processing. (a) PR curve. (b) MAE. (c) F-Measure.

The main difference between our method and previous global contrast methods (FT [17], HC [18], SF [20], RC [18]) is considering the distribution of contrast. We take an evaluation between proposed contrast method and other global contrast based methods on four public datasets. In this evaluation, post-processing is turn off in proposed method, HC [18] and RC [18] (denoted as Ours*, HC* and RC*), only global contrast is adopted in SF [20] (denoted as SF*). In one word, raw saliency computed by contrast method is taken for comparison.

Some examples for quality comparison are shown in Fig. 7. Top two rows are examples with simple scene background, proposed method can detect the object with clear contour and suppress saliency of background, RC* [18] and SF* [20] show fuzzy contour, HC* [18] and FT [17] provide high saliency on some background region. Bottom three rows are

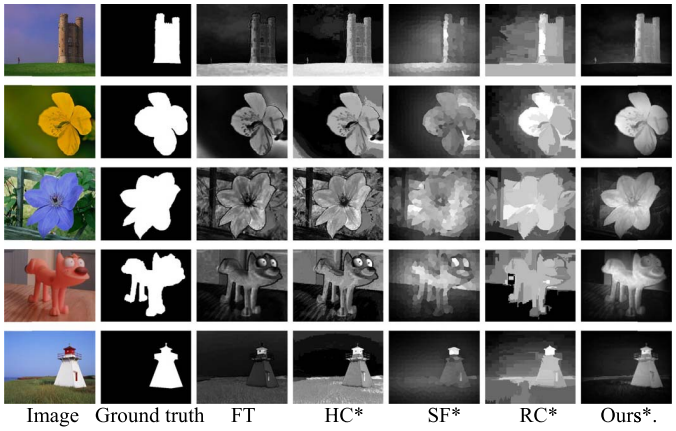

Fig. 7. Contrast method (raw saliency) quality comparison. Our method significantly perform other global contrast based methods.

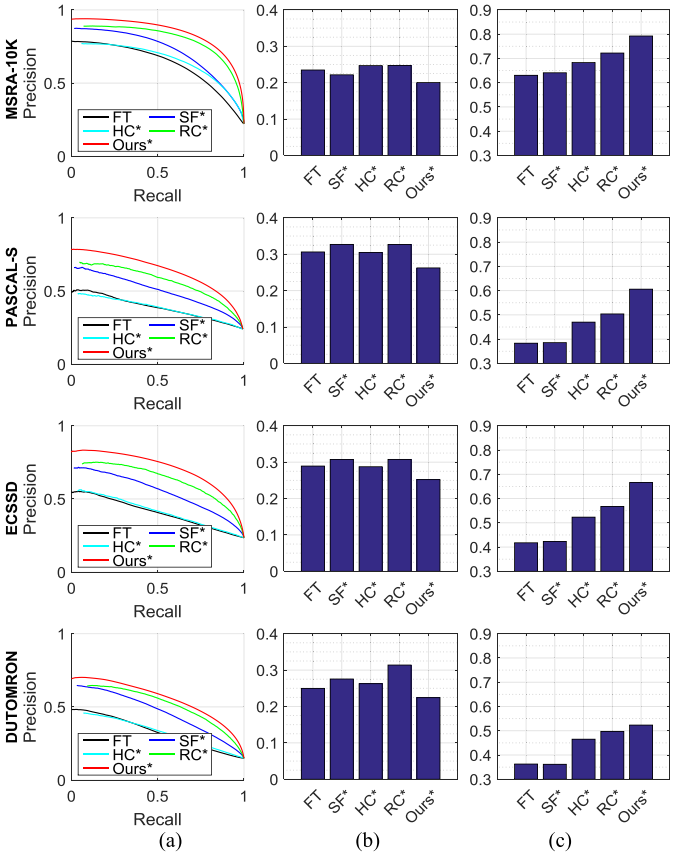

Fig. 8. Contrast method (raw saliency) quantitative evaluation between global contrast based methods. (a) PR curve. (b) MAE. (c) F-Measure.

examples with complex scene background, proposed method also significantly perform other methods.

Quantitative comparison (PR curve, F-measure and MAE measure) on four public datasets is shown in Fig. 8. Our method significantly outperforms other global contrast based methods on all datasets and every evaluation metric. These experimental results prove that *MDC* is a better metric of saliency.

2) Post-Processing Analysis: In saliency post-processing, we adopt saliency smoothing and one simple saliency enhancement method. We carried out some experiments to analyze the

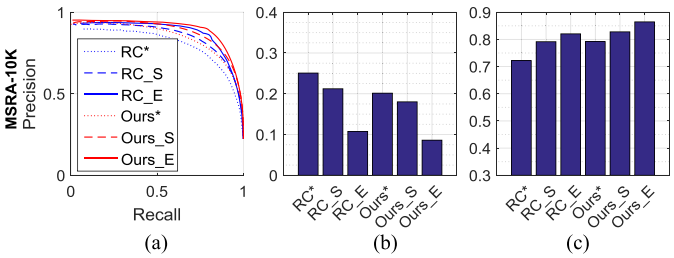

Fig. 9. Post-processing analysis. Both saliency smoothing and enhancement can improve performance on raw saliency of RC and our work. (a) PR curve. (b) MAE. (c) F-Measure.

effectiveness of post-processing step. In order to investigate the influence of post-processing on other saliency detection methods, we also carry the same refine step on RC [18].

Quantitative evaluation of raw saliency, smoothed saliency and enhanced saliency (suffix denoted as *, _S and _E) on MSRA-10K dataset is shown in Fig. 9. We can draw three conclusions from the figure:

1) Both saliency smoothing and enhancement can improve performance on PR curve, MAE and F-Measure. Especially saliency enhancement can provide obvious decline on MAE.
2) Our post-processing step also can achieve performance improvement on RC [18].
3) The same post-processing is carried on raw saliency of our work and RC [18], but our refined result shows better performance than refined result of RC [18]. This phenomenon illustrates raw saliency computation is very important in saliency detection.

E. Performance Evaluation

This evaluation is carried on final saliency result, some methods adopt post-processing or other saliency metric except contrast. The proposed algorithm is firstly compared with global contrast based methods: FT [17], HC [18], SF [20], RC [18]. We also compare with some state-of-the-arts or recent work: MST [45], MAP [49], DBCL [50], MC [42], MB+ [40], MR [23], RBD [37] and DSR [11]. Besides, two deep learning based method LEGS [47] and MCDL [48] are also compared. To evaluate these methods, we run their implementations based on the available codes or software.

Some images from datasets are used as training data in deep learning based approaches while traditional methods usually just based on handcrafted features. Therefore, the experimental comparisons are divided into traditional methods and deep learning based methods.

1) Compared With Traditional Methods: Fig. 10 shows a few saliency maps of the evaluated methods. Compared with global contrast based method FT [17], HC [18], SF [20] and RC [18], our model can tackle even more complicated scenarios, examples are presented at first and second row. The second advantage of our method is providing highlighted object while other methods failed. In the third row, one green fruit is located at input image, RC [18], MST [45] and MB+ [40] shows fuzzy object contour, while we demonstrate highlighted object. The fourth row is one example of the

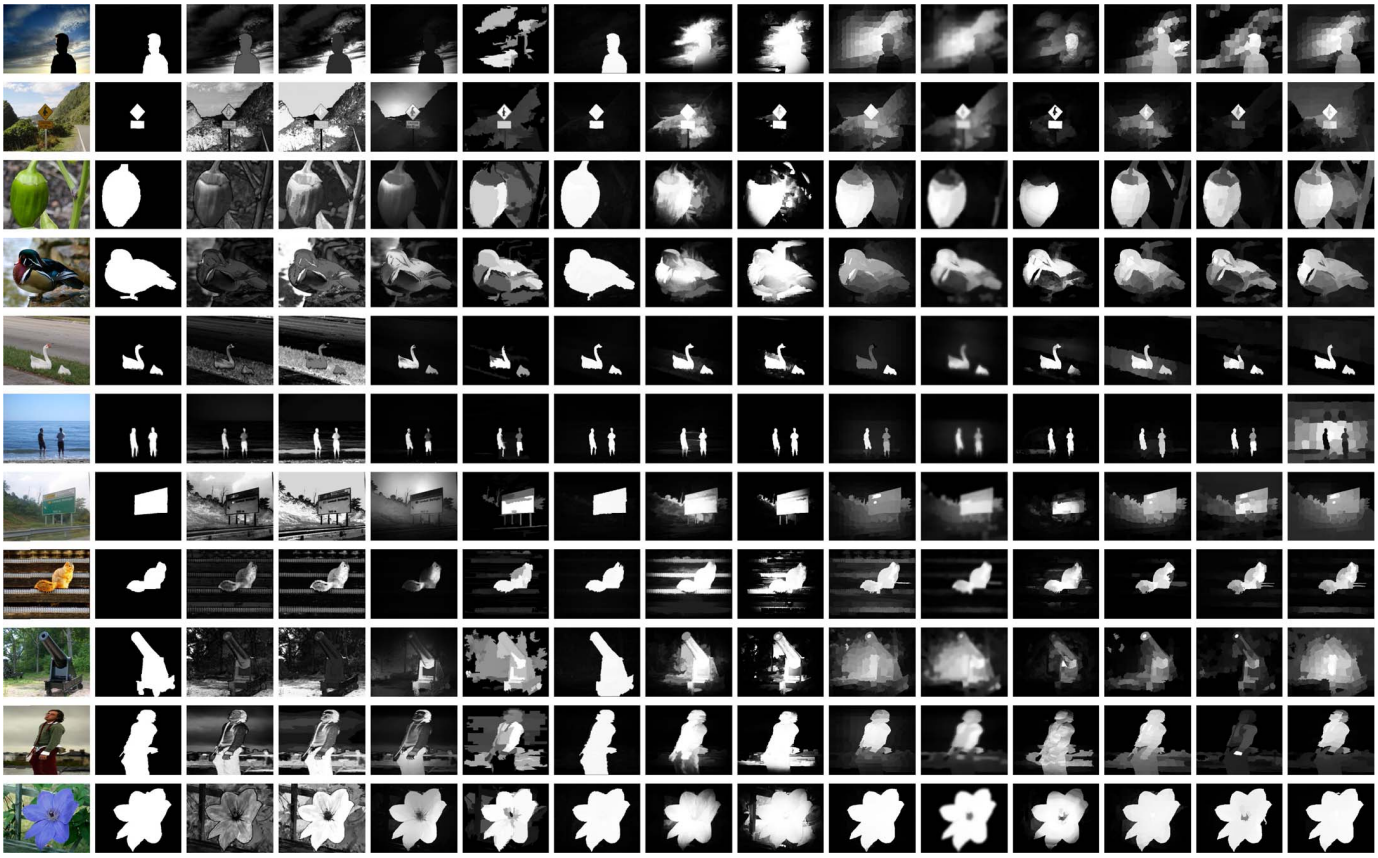

Fig. 10. Final saliency quality comparison.

object with different color, our method also shows highlighted object. Our method also shows promised saliency result on the image with multi objects, examples are shown in the fifth and sixth row.

Fig. 11 shows quantitative comparison between the proposed method and some previous methods. Compared with global contrast based methods FT [17], HC [18], SF [20] and RC [18], we show significantly better performance on all datasets and every evaluation metrics. Compared with MB+ [40] which is a recent top-performing method, we show comparable PR curve but obviously better MAE, we also demonstrate comparable F-measure on PASCAL-S dataset and better F-measure on other datasets. We demonstrate better or comparable performance than RBD [37] and DSR [11] except a little worse MAE on DUTOMRON dataset. Our method shows comparable F-Measure and better PR curve and MAE than MC [42]. We outperform other state-of-the-arts such as MR [23], MST [45] and MAP [49] on all datasets and every evaluation metrics.

2) Compared With Deep Learning Based Methods: Since some images of MSRA-10K and PASCAL-S are selected as train data by LEGS [47] and MCDL [48], we only evaluate on challenging dataset ECSSD and DUTOMRON.

Quantitative comparison is shown in Fig. 12. Deep learning based methods show significantly better performance. Our perspective is that computational efficiency must be an important consideration since saliency detection is typically a preprocessing step. In Sec.IV-F, speed performance shows that our proposed method runs at about 300 FPS on CPU, while LEGS [47] and MCDL [48] only achieves 0.03 and 0.04 FPS on the same CPU platform. That is to say, our work shows about 10000 times faster. The trade-off between accuracy and speed is also shown in Fig. 14.

We further analyze performance gap between our work and deep learning based methods. Deep learning based work can learn or adopt multi-level feature: low-level feature such as color, edge and contrast, high level feature such as object concept. We only consider contrast as saliency metric. In LEGS [47], author firstly uses a deep neural network to learn low-level feature (including color, contrast, texture, etc.) for initial saliency estimation, then refine saliency by exploring the high level object concepts. To this end, they utilize the geodesic object proposal (GOP) [29] method to extract a set of object candidates, then calculate confidence of each object with initial saliency. The refined saliency is generated by the top K ($= 20$) confident objects. High level object proposals is integrated with low level initial saliency. Similar to this work, we use smoothed *MDC* as initial saliency, then integrate with object proposal to refine saliency, finally adopt proposed saliency enhancement. Quantitative evaluation of our modified method (denoted as Ours+) is also shown in Fig. 12. Two examples are shown in Fig. 13. Our method

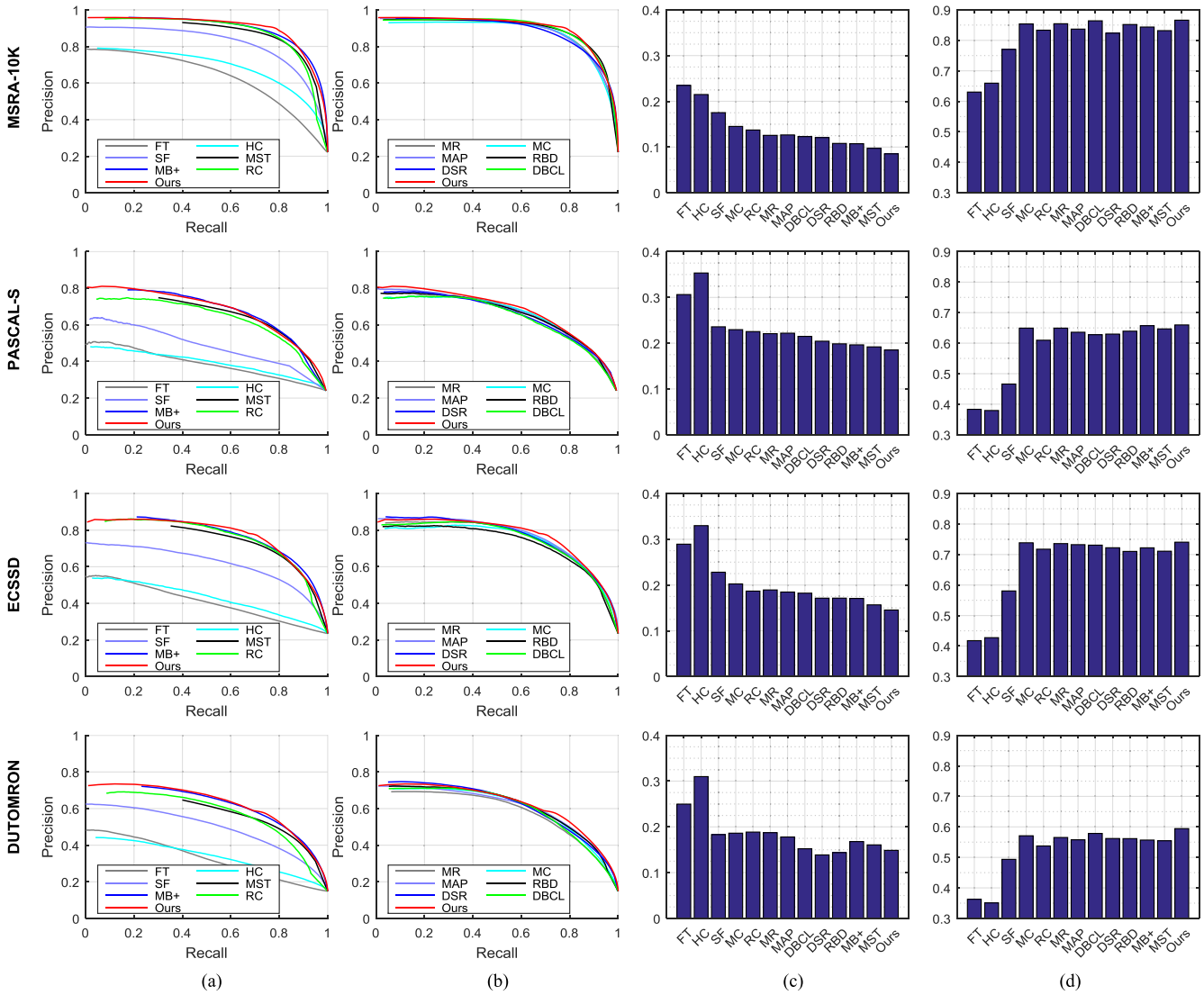

Fig. 11. Final saliency quantitative comparison. (a) PR curve. (b) PR curve. (c) MAE. (d) F-Measure.

integrated with object proposal achieves significant performance improvement. Note that we only use contrast as low level feature, and use unsupervised GOP [29] to generate object candidates. As a result, our modified method still shows a little worse performance than LEGS [47] and MCDL [48] which is supervised method.

Integration with object proposal can greatly improve performance, but object proposal generation needs about average 0.5s on four datasets. Our modified method only achieves 2 FPS speed performance. Nevertheless, we still show obvious speed advantage than LEGS [47] and MCDL [48] (0.03 and 0.04 FPS). We will focus on how to fast integrate high level object feature in future work.

F. Speed Comparison

The speed performance test is carried on a 64-bit PC with Intel Core i5-4590 CPU @ 3.3 GHz and 8GB RAM. Average speed is computed on all images of four public datasets. We do not count file I/O time, and do not allow processing multiple images in parallel. Our work, MB+ [40], RC [18], MST [45], FT [17], HC [18] are implemented in C code. Other methods use Matlab code but the most time-consuming superpixel segmentation or deep learning is implemented in C code.

In order to comprehensively compare performance and speed, we illustrate the trade-off between MAE, F-Measure and speed in Fig. 14. Both performance and speed are average result on four datasets. The speed evaluation results of all methods are also shown in Table I.

Our proposed method runs at about 300 FPS. Compared with MB+ [40] which have highest speed performance (52 FPS) in the state-of-the-arts, we show 6 times speed improvement. Compared with RC [18] which is the state-of-the-art also based on global contrast, we perform 25 times speed advantage. FT [17] and HC [18] show high speed (148 and 168 FPS) but with significantly worse performance. Deep learning based work LEGS [47] and MCDL [48] achieves better performance with very low speed.

Given an input image of QVGA resolution, raw saliency computation via *MDC* needs about 1.5ms, saliency smoothing

TABLE I
SPEED EVALUATION RESULTS

Methods		Code Type	fps
state-of-the-arts	**Ours**	**C**	**305**
	MB+ [40]	C	52
	MST [45]	C	40
	RC [18]	C	12
	MC [42]	C, Matlab	8
	MR [23]	C, Matlab	7
	MAP [49]	C, Matlab	7
	RBD [37]	C, Matlab	6
	DBCL [50]	C, Matlab	0.9
	DSR [11]	C, Matlab	0.3
	LEGS [47]	C, Matlab	0.03
	MCDL [48]	C, Matlab	0.04
other works	SF [20]	C, Matlab	7
	FT [17]	C	148
	HC [18]	C	168

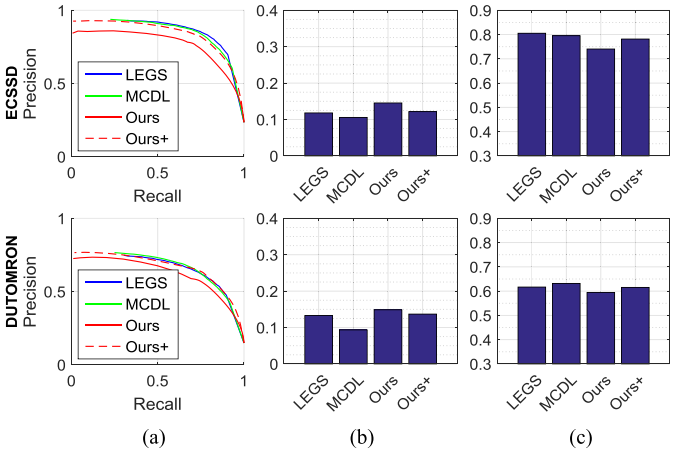

Fig. 12. Quantitative comparison with deep learning based method. Our work shows worse performance than deep learning based algorithm. If integrated with high level object proposal, our performance can be significantly improved. (a) PR curve. (b) MAE. (c) F-Measure.

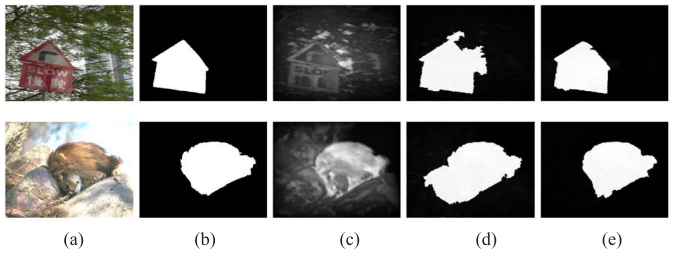

Fig. 13. Quantity comparison of our work integrated with objectness or not. (a) Input image. (b) Ground truth. (c) MDC. (d) Ours. (e) Ours+

and saliency enhancement needs 0.4ms and 0.9ms respectively, other steps need 0.5ms. We achieve significantly higher speed than the state-of-the-art mainly due to following factors:

(1) We calculate saliency at pixel level instead of region level which needs time cost for segmentation. In order to get region segmentation, RC [18] adopt graph-based segmentation [34] which needs 60ms for each image, SLIC superpixel

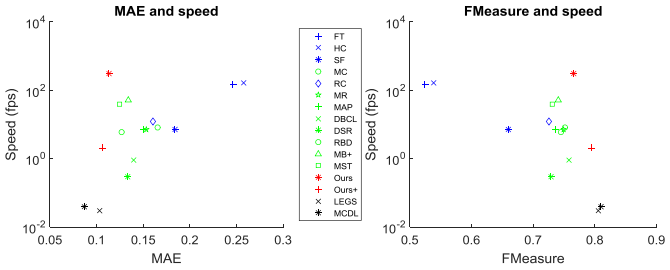

Fig. 14. Trade-off between accuracy and speed. Our method (in red) shows better trade-off than other global contrast methods (in blue), state-of-the-arts traditional methods (in green) and deep learning based work (in black).

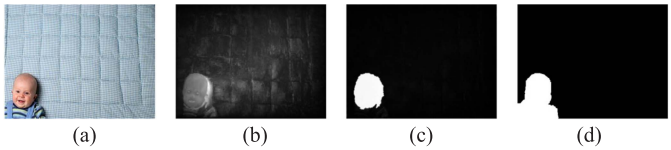

Fig. 15. Fail case when salient object locates at image corners. (a) Input image. (b) Raw saliency. (c) Final saliency. (d) Ground truth.

based methods (such as RBD [37], MAP [49], MR [23]) need about 110ms for segmentation.

(2) In raw saliency detection step, we propose efficient $O(1)$ implementation of *MDC* with integral image. Raw saliency computation needs raster scan image once and consumes 1.5ms for an input image of QVGA resolution. In MB+ [40] which is also pixel level method, raster scan needs three passes for each color channel, that is to say, raster scan totally needs nine times.

(3) In saliency post-processing, we use a simple but effective marker-based watershed algorithm to enhance saliency.

G. Limitation

MDC will fail when the salient objects are just located at image corners. In Fig. 15, one baby is located at the bottom left corner of the input image which is shown in Fig. 15. (a), *MDC* of the baby is contrast from bottom left direction, head part which near the corner still shows high *MDC*, while other part which touches the corner has very low *MDC*. Final saliency in Fig. 15. (c) only highlights head of the baby.

Note that if the salient object touches the boundary of the image but doesn't touch the image corner, *MDC* usually can handle this case. One example is shown in the bottom row of Fig. 4, the people in the image touch bottom boundary but still shows high *MDC*. Statistics on MSRA-10K dataset shows that about 4% salient objects touch image boundary, while salient objects locate at image corner less than 1%. In this respect, corner-sensitive *MDC* method is robust than boundary-sensitive algorithms such as MB+ [40].

V. CONCLUSIONS

We find that the spatial distribution of contrast is a very important cue of saliency that is neglected by previous works. Then *MDC* based saliency metric and $O(1)$ implementation with integral image are proposed. In saliency post-processing,

a simple yet effective watershed based saliency enhancement algorithm is also proposed.

Experimental results on four datasets demonstrate the proposed method significantly outperforms other global contrast based methods, and achieves comparable or better performance than the state-of-the-art methods. The proposed method runs at 300 FPS. In the state-of-the-art methods, MB+ [40] shows the highest speed at 52 FPS, our method shows 6 times speed improvement.

Given an input image of QVGA resolution, our raw saliency computation only needs about 1.5ms, while saliency post-processing needs 1.3ms (0.4ms for smoothing and 0.9ms for enhancement). We will focus on more effective and fast post-processing method in future work to achieve higher speed.

REFERENCES

[1] 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.

[2] G. Sharma, F. Jurie, and C. Schmid, "Discriminative spatial saliency for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 3506–3513.

[3] D. Walther, U. Rutishauser, C. Koch, and P. Perona, "Selective visual attention enables learning and recognition of multiple objects in cluttered scenes," *Comput. Vis. Image Understand.*, vol. 100, nos. 1–2, pp. 41–63, Oct. 2005.

[4] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2189–2202, Nov. 2012.

[5] 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.

[6] L. Wang, J. Xue, N. Zheng, and G. Hua, "Automatic salient object extraction with contextual cue," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 105–112.

[7] C. Jung and C. Kim, "A unified spectral-domain approach for saliency detection and its application to automatic object segmentation," *IEEE Trans. Image Process.*, vol. 21, no. 3, pp. 1272–1283, Mar. 2012.

[8] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 853–860.

[9] P. Jiang, H. Ling, J. Yu, and J. Peng, "Salient region detection by UFO: Uniqueness, focusness and objectness," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 1976–1983.

[10] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 1155–1162.

[11] X. Li, H. Lu, L. Zhang, X. Ruan, and M.-H. Yang, "Saliency detection via dense and sparse reconstruction," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 2976–2983.

[12] R. Liu, J. Cao, Z. Lin, and S. Shan, "Adaptive partial differential equation learning for visual saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 3866–3873.

[13] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu, "Saliency detection on light field," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 2806–2813.

[14] 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.

[15] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2006, pp. 545–552.

[16] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.

[17] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 1597–1604.

[18] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, Mar. 2015.

[19] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li, "Automatic salient object segmentation based on context and shape prior," in *Proc. Brit. Mach. Vis. Conf.*, Aug. 2011, p. 9.

[20] F. Perazzi, P. Krähenbühl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 733–740.

[21] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Proc. 12th Eur. Conf. Comput. Vis. (ECCV)*, 2012, pp. 29–42.

[22] 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.

[23] C. Yang, L. Zhang, H. Lu, X. Ruan, and M.-H. Yang, "Saliency detection via graph-based manifold ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 3166–3173.

[24] C. Kanan, M. H. Tong, L. Zhang, and G. W. Cottrell, "SUN: Top-down saliency using natural statistics," *Vis. Cognit.*, vol. 17, nos. 6–7, pp. 979–1003, 2009.

[25] J. Yang and M.-H. Yang, "Top-down visual saliency via joint CRF and dictionary learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2296–2303.

[26] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 185–207, Jan. 2013.

[27] A. Borji, M.-M. Cheng, H. Jiang, and J. Li. (Nov. 2014). "Salient object detection: A survey." [Online]. Available: https://arxiv.org/abs/1411.5878

[28] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, "Salient object detection: A benchmark," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5706–5722, Dec. 2015.

[29] P. Krähenbühl and V. Koltun, "Geodesic object proposals," in *Proc. 14th Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 725–739.

[30] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 2376–2383.

[31] T. Liu *et al.*, "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 353–367, Feb. 2011.

[32] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proc. 14th Annu. ACM Int. Conf. Multimedia*, Oct . 2006, pp. 815–824.

[33] 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. (ICCV)*, Dec. 2013, pp. 1529–1536.

[34] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.

[35] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[36] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[37] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 2814–2821.

[38] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2001, pp. 511–518.

[39] F. Meyer, "Color image segmentation," in *Proc. Int. Conf. Image Process. Appl.*, Apr. 1992, pp. 303–306.

[40] J. Zhang, S. Sclaroff, Z. Lin, X. Shen, B. Price, and R. Mech, "Minimum barrier salient object detection at 80 FPS," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1404–1412.

[41] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.

[42] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, "Saliency detection via absorbing Markov chain," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1665–1672.

[43] Y. Li, X. Hou, C. Koch, J. M. Rehg, and A. L. Yuille, "The secrets of salient object segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, May 2014, pp. 280–287.
[44] J. Kim, D. Han, Y.-W. Tai, and J. Kim, "Salient region detection via high-dimensional color transform," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 883–890.
[45] W.-C. Tu, S. He, Q. Yang, and S.-Y. Chien, "Real-time salient object detection with a minimum spanning tree," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2334–2342.
[46] R. Strand, K. C. Ciesielski, F. Malmberg, and P. K. Saha, "The minimum barrier distance," *Comput. Vis. Image Understand.*, vol. 117, no. 4, pp. 429–437, Apr. 2013.
[47] L. Wang, H. Lu, X. Ruan, and M.-H. Yang, "Deep networks for saliency detection via local estimation and global search," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3183–3192.
[48] R. Zhao, W. Ouyang, H. Li, and X. Wang, "Saliency detection by multi-context deep learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1265–1274.
[49] J. Sun, H. Lu, and X. Liu, "Saliency region detection based on Markov absorption probabilities," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1639–1649, May 2015.
[50] L. Zhou, Z. Yang, Q. Yuan, Z. Zhou, and D. Hu, "Salient region detection via integrating diffusion-based compactness and local contrast," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3308–3320, Nov. 2015.

Xiaoming Huang received the bachelor's degree from Lanzhou University and the master's degree from Peking University. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering, Tsinghua University, Beijing, China. His research interests include image processing, computer vision, and machine learning.

Yu-Jin Zhang (SM'99) received the Ph.D. degree in applied science from the University of Liege, Belgium, in 1989. He was a Post-Doctoral Fellow and a Research Fellow with the Delft University of Technology, The Netherlands, from 1989 to 1993. He joined Tsinghua University, Beijing, China, where he has been a Professor since 1997. He has authored 38 books and 500 papers in image engineering, include image processing, image analysis, and image understanding. He is a fellow of the SPIE for achievements in image engineering. He is a Program Chair of the ICIP2017.