IPSJ Transactions on Mathematical Modeling and Its Applications Vol.6 No.2 78–85 (Aug. 2013)

Regular Paper

Salient Object Detection Based on Direct Density-ratio Estimation

Masao Yamanaka[1,2,a) Masakazu Matsugu[2,b) Masashi Sugiyama[3,c)

Received: August 19, 2012, Revised: November 6, 2012,
Accepted: January 29, 2013

Abstract: Detection of salient objects in images has been an active area of research in the computer vision community. However, existing approaches tend to perform poorly in noisy environments because probability density estimation involved in the evaluation of visual saliency is not reliable. Recently, a novel machine learning approach that directly estimates the ratio of probability densities was demonstrated to be a promising alternative to density estimation. In this paper, we propose a salient object detection method based on direct density-ratio estimation, and demonstrate its usefulness in experiments.

Keywords: salient object detection, direct density-ratio estimation, relative density-ratio estimation, Shannon entropy, density estimation

1. Introduction

Detecting salient objects in images has been extensively investigated in many computer vision applications such as general object detection in web images [2] and over image thumbnails [11], and computing a joint focus of attention in human robot interaction [12]. Here, a salient object indicates a region in an image that visually stands out from its surroundings and is likely to attract human attention, as illustrated in **Fig. 1**. A key property that makes an object salient is the visual difference from the background.

Methods of salient object detection can be divided into the top-down approach based on supervised learning [2], [10] and the bottom-up approach based on unsupervised learning [3], [6]. So far, various top-down methods have been proposed, for example, Alexe et al. [2] combined multi-scale saliency, color contrast, edge density, and super-pixels in a Bayesian framework, and Liu et al. [10] combined multi-scale contrast, center-surround histograms, and color spatial-distributions with conditional random fields. However, the performance of the top-down approach depends heavily on the quality and quantity of ground truth data used for supervised learning, and gathering a large number of high-quality training data is costly. Furthermore, adding a new object category is not straightforward and human subjectivity often causes ambiguity.

On the other hand, the bottom-up approach can be easily applied in an on-line fashion with no labeling cost. A seminal work by Itti et al. [6] is based on a feature integration theory in cognitive science [14]. This method identifies salient objects based on conspicuity maps that are generated from the spatial contrast of features such as the luminance value, edge intensity, and gradient orientation. While many computational models have been developed and their applications have been explored based on this structure [3], [4], [16], fusion of feature channels remains somewhat arbitrary. Furthermore, the performance of the bottom-up approach depends on the characteristics of image features—if features are sensitive to environmental and observational noise, lighting conditions and system-specific noise can cause severe performance degradation.

The above approaches are mostly motivated biologically. On the other hand, several recent approaches attempted to model saliency computationally and mathematically. For example, Kadir et al. [7] introduced entropy-based saliency, and Hou and Zhang [5] computed the incremental coding length to measure the perspective entropy gain. However, entropy-based methods tend to identify objects with various structures as salient, which is not always appropriate in practice.

In this paper, we propose a new bottom-up method to detect salient objects in images that is robust against noise incurred by various environmental and observational factors. Our approach is based on the standard structure of cognitive visual attention models [14], where several feature channels are investigated in parallel and the conspicuity maps are fused to a single saliency map. We choose the lightness, edge intensity, and color as feature channels because they are basic features of the human attention system [17].

Our saliency computation consists of two steps: First, we sample low-dimensional features such as intensities and colors in different scales. Then, in the second step, the center-surround con-

[1] Graduate School of Science and Engineering, Tokyo Institute of Technology, Yokohama, Kanagawa 226–8502, Japan
[2] CANON Inc., Ohta, Tokyo 146–8501, Japan
[3] Graduate School of Information Science and Engineering, Tokyo Institute of Technology, Meguro, Tokyo 152–8552, Japan
[a)] yamanaka.masao@canon.co.jp
[b)] matsugu.masakazu@canon.co.jp
[c)] sugi@cs.titech.ac.jp

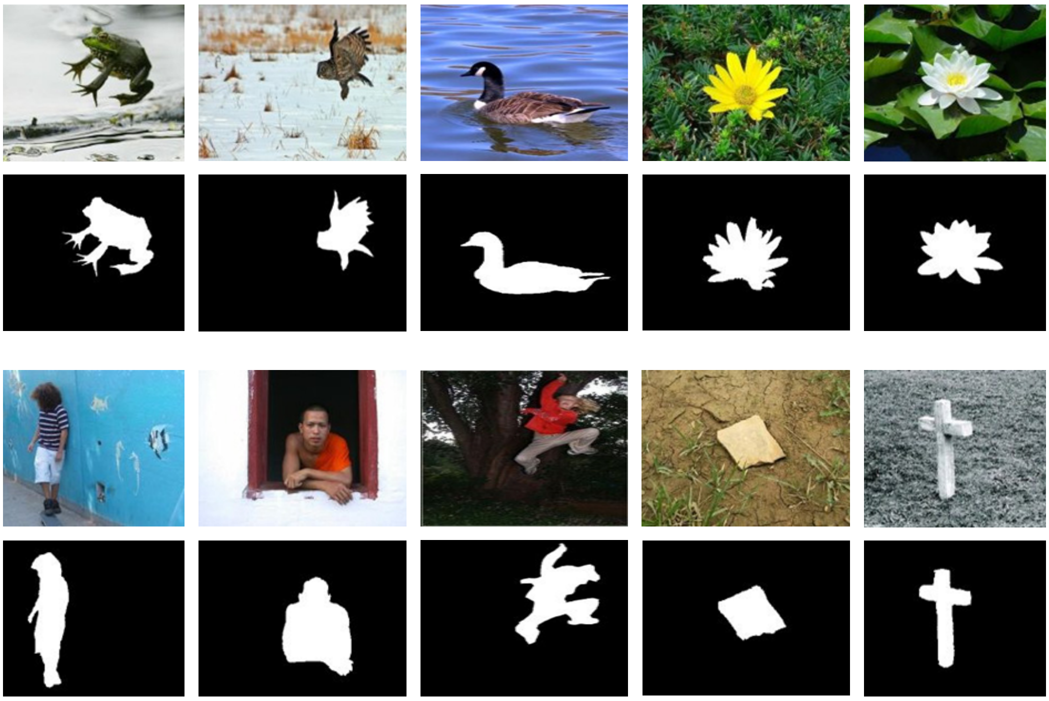


Fig. 1 Examples of salient objects in images [10].

trast is evaluated with a machine learning technique. More specifically, two probability densities of visual feature occurrences are considered for a center region and a surround region, and a divergence between these densities is evaluated by the state-of-the-art machine learning method called *direct density-ratio estimation* [13]: The *ratio* of probability densities is directly estimated without separate estimation of each density. Because density estimation is known to be a hard task [15], avoiding density estimation and directly estimating the density ratio would be more promising.

Through experiments, we demonstrate that the proposed method allows robust computation of visual saliency in various scales and tends to outperform the method that directly estimates probability density for low-dimensional features [7].

2. Problem Formulation

In this section, we formulate our salient object detection problem based on density ratios.

Let $\boldsymbol{x}$ be a low-dimensional feature (e.g., lightness, color, and edge) extracted from an image. Our task is to detect whether there exists a salient object from the low-dimensional feature. A naive approach to this problem would be to first estimate the center and surround probability density functions for low-dimensional features separately, and then to evaluate the difference between center and surround regions by comparing the estimated probability density functions.

However, since non-parametric density estimation is known to be a hard problem [15], this naive approach to salient object detection may not be effective in practice. Instead, directly estimating the *ratio* of probability densities without going through density estimation was shown to be more promising [13]. Motivated by this line of research, we develop a saliency detection algorithm based on the relative density ratio for low-dimensional feature $\boldsymbol{x}$:

$$w(\boldsymbol{x}) = \frac{p_{\mathrm{c}}(\boldsymbol{x})}{\beta p_{\mathrm{c}}(\boldsymbol{x}) + (1-\beta)p_{\mathrm{s}}(\boldsymbol{x})}, \tag{1}$$

where $p_{\mathrm{c}}(\boldsymbol{x})$ and $p_{\mathrm{s}}(\boldsymbol{x})$ are the probability density functions for center and surround low-dimensional features, respectively. β $(0 \le \beta < 1)$ is the relative parameter that controls the "smoothness" of the density ratio; $\beta = 0$ corresponds to the plain density ratio $p_{\mathrm{c}}(\boldsymbol{x})/p_{\mathrm{s}}(\boldsymbol{x})$ and the relative density ratio tends to be smoother as β gets larger. See Ref. [18] for a theoretical background for this relative parameter.

We use the *relative Pearson divergence* from $p_{\mathrm{c}}(\boldsymbol{x})$ to $p_{\mathrm{s}}(\boldsymbol{x})$ [9], [18] as our saliency score S:

$$S = \frac{1}{2}\int (w(\boldsymbol{x}) - 1)^2 \left(\beta p_{\mathrm{c}}(\boldsymbol{x}) + (1-\beta)p_{\mathrm{s}}(\boldsymbol{x})\right) \mathrm{d}\boldsymbol{x},$$

which is always non-negative and zero if and only if $p_{\mathrm{c}} = p_{\mathrm{s}}$. **Figure 2** illustrates the relation between visual saliency in an image and the relative density ratio for low-dimensional features (e.g., color). The visual saliency of the center-surround region in the left-hand side of Fig. 2 is high, yielding the saliency score S to be large. On the other hand, the visual saliency of the center-surround region in the right-hand side of Fig. 2 is low, resulting in a small saliency score S.

In practice, it may be difficult to determine the size of center-surround regions to properly detect salient objects without any prior knowledge. In this paper, we mitigate this difficulty by considering a hierarchy of center-surround regions, as illustrated in **Fig. 3**. This hierarchical structure makes it possible to detect salient objects in different scales from micro to macro levels. We further consider multiple types of low-dimensional features such as lightness, color, and edge, which are linearly combined as

$$\sum_{h=1}^{H} \left(S_h^{\mathrm{l}} + S_h^{\mathrm{c}} + S_h^{\mathrm{e}}\right),$$

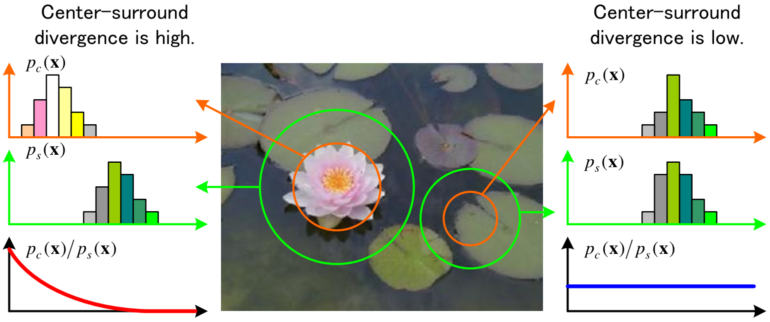


Fig. 2 Relation between visual saliency in an image and probability density ratios for low-dimensional features (e.g., color).

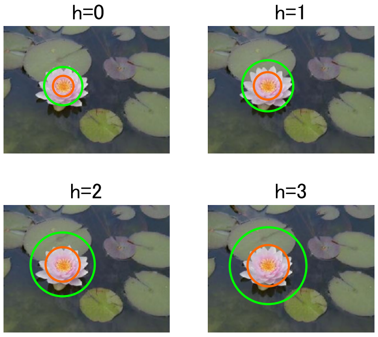


Fig. 3 Center-surround region and down sampling.

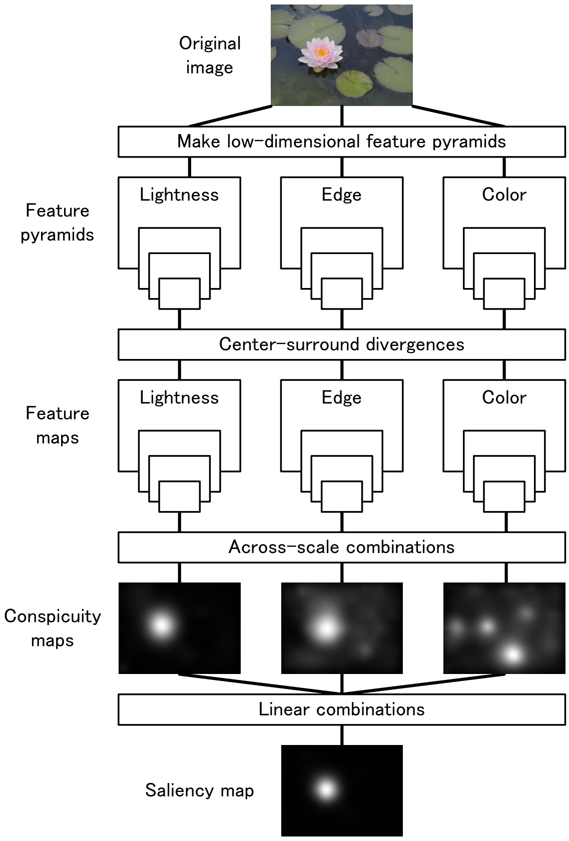


Fig. 4 Schematic overview of our saliency detection system.

where S_h^{l}, S_h^{c}, and S_h^{e} are the scores of lightness, color, and edge in the h-th hierarchy, respectively. Finally, we build a saliency map by calculating the saliency score exhaustively at all position in the image. The above formulation is summarized in **Fig. 4**, which illustrates the schematic overview of our saliency detection system.

The remaining question in the proposed procedure is how to evaluate the relative Pearson divergence S from data, which is discussed in the next section.

3. Direct Approximation of Pearson Divergence

In this section, we show how the relative Pearson divergence is evaluated from data.

3.1 Approximation of Relative Pearson Divergence by Relative Density-ratio Estimation

Suppose we are given a set of N_{c} samples extracted from a center region (see **Fig. 5**) that are drawn independently from a probability distribution P_{c} with density p_{c}:

$$\mathcal{X}_{\mathrm{c}} = \left\{\boldsymbol{x}_i^{\mathrm{c}} \mid \boldsymbol{x}_i^{\mathrm{c}} \in \Re^d\right\}_{i=1}^{N_{\mathrm{c}}} \overset{\mathrm{i.i.d}}{\sim} P_{\mathrm{c}}.$$

We also suppose that another set of N_{s} samples extracted from a surround region (see **Fig. 6**) that are drawn independently from (possibly) another probability distribution P_{s} with density p_{s}:

$$\mathcal{X}_{\mathrm{s}} = \left\{\boldsymbol{x}_j^{\mathrm{s}} \mid \boldsymbol{x}_j^{\mathrm{s}} \in \Re^d\right\}_{j=1}^{N_{\mathrm{s}}} \overset{\mathrm{i.i.d}}{\sim} P_{\mathrm{s}}.$$

From the samples $\mathcal{X}_{\mathrm{c}}$ and $\mathcal{X}_{\mathrm{s}}$, we estimate the relative density ratio defined by Eq. (1). Let $\widehat{w}(\boldsymbol{x})$ be an estimate of the relative density ratio. Then the Pearson divergence can be approximated as

$$\widehat{S} = -\frac{\beta}{2N_{\mathrm{c}}}\sum_{i=1}^{N_{\mathrm{c}}} w(\boldsymbol{x}_i^{\mathrm{c}})^2 - \frac{1-\beta}{2N_{\mathrm{s}}}\sum_{j=1}^{N_{\mathrm{s}}} w(\boldsymbol{x}_j^{\mathrm{s}})^2 + \frac{1}{N_{\mathrm{c}}}\sum_{i=1}^{Nc} w(\boldsymbol{x}_i^{\mathrm{c}}) - \frac{1}{2}.$$

Below, we explain how the relative density ratio can be directly estimated without going through density estimation.

3.2 Unconstrained Least-squares Approach to Relative Density-ratio Estimation

Here, we review a density-ratio estimation method called *relative unconstrained least-squares importance fitting* (RuLSIF) [18].

Let us model the relative density ratio function $w(\boldsymbol{x})$ by the following kernel model:

$$\widetilde{w}(\boldsymbol{x}) = \sum_{i=1}^{N_{\mathrm{c}}} \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i^{\mathrm{c}}) = \boldsymbol{\alpha}'\boldsymbol{k}(\boldsymbol{x}),$$

where

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_{N_c})'$$

are parameters to be learned from data samples, $\bullet'$ denotes the transpose of a matrix or a vector, and

$$\boldsymbol{k}(\boldsymbol{x}) = \left(K(\boldsymbol{x}, \boldsymbol{x}_1^c), K(\boldsymbol{x}, \boldsymbol{x}_2^c), \ldots, K(\boldsymbol{x}, \boldsymbol{x}_{N_c}^c)\right)'$$

are kernel basis functions. A popular choice of the kernel is the Gaussian function:

$$K(\boldsymbol{x}, \boldsymbol{y}) = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{y}\|^2}{2\sigma^2}\right), \qquad (2)$$

where $\sigma > 0$ is the Gaussian width.

We determine the parameter $\boldsymbol{\alpha}$ in the model $\widetilde{w}(\boldsymbol{x})$ so that the following squared-error J_0 is minimized:

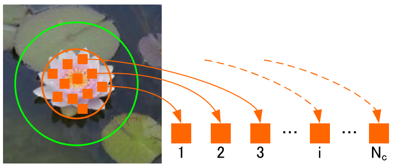


Fig. 5 Low-dimensional feature samples randomly extracted from a center region.

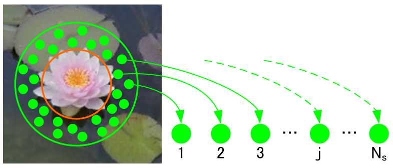


Fig. 6 Low-dimensional feature samples randomly extracted from a surround region.

$$\begin{aligned} J_0 &= \frac{1}{2}\int \left(\widetilde{w}(\boldsymbol{x}) - w(\boldsymbol{x})\right)^2 (\beta p_c(\boldsymbol{x}) + (1-\beta)p_s(\boldsymbol{x}))\,\mathrm{d}\boldsymbol{x} \\ &= \frac{\beta}{2}\int \widetilde{w}(\boldsymbol{x})^2 p_c(\boldsymbol{x})\mathrm{d}\boldsymbol{x} + \frac{1-\beta}{2}\int \widetilde{w}(\boldsymbol{x})^2 p_s(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \\ &\quad - \int \widetilde{w}(\boldsymbol{x}) p_c(\boldsymbol{x})\mathrm{d}\boldsymbol{x} + \mathrm{Const.} \end{aligned}$$

Let us denote the first three terms by J. Since J contains the expectation over unknown densities $p_s(\boldsymbol{x})$ and $p_c(\boldsymbol{x})$, we approximate the expectations by empirical averages. Then we obtain

$$\widehat{J}(\boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{\alpha}'\widehat{\boldsymbol{H}}\boldsymbol{\alpha} - \boldsymbol{\alpha}'\widehat{\boldsymbol{h}},$$

where $\widehat{\boldsymbol{H}}$ is the $N_c \times N_c$ matrix defined by

$$\widehat{\boldsymbol{H}} = \frac{\beta}{N_c}\sum_{i=1}^{N_c} \boldsymbol{k}(\boldsymbol{x}_i^c)\boldsymbol{k}(\boldsymbol{x}_i^c)' + \frac{1-\beta}{N_s}\sum_{j=1}^{N_s} \boldsymbol{k}(\boldsymbol{x}_j^s)\boldsymbol{k}(\boldsymbol{x}_j^s)',$$

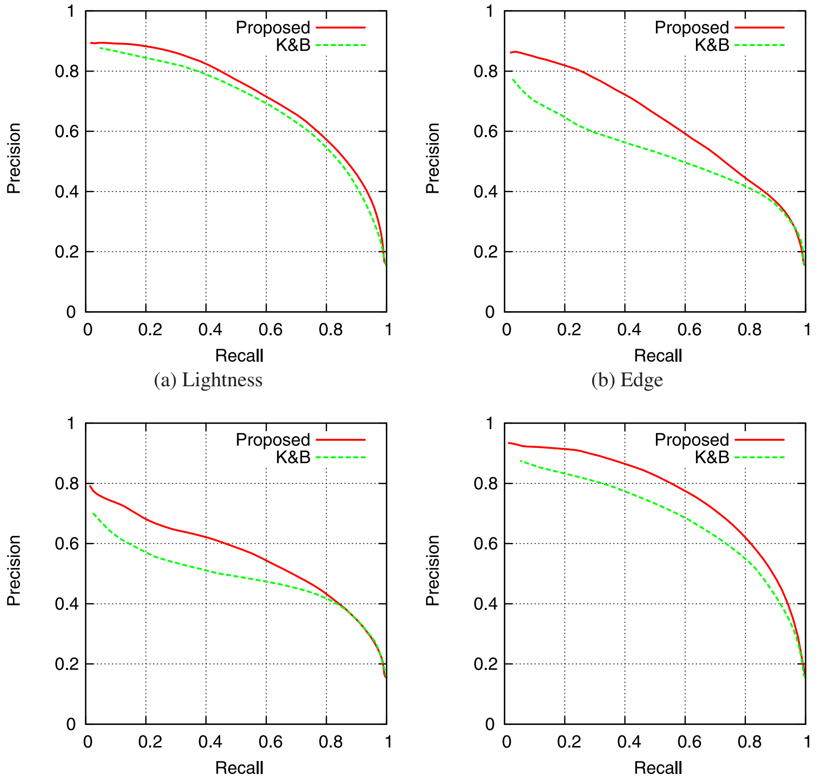
and $\widehat{\boldsymbol{h}}$ is the N_c-dimensional vector defined by

$$\widehat{\boldsymbol{h}} = \frac{1}{N_c}\sum_{i=1}^{N_c} \boldsymbol{k}(\boldsymbol{x}_i^c).$$

By including a regularization term, the RuLSIF optimization problem is formulated as

$$\widehat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}}\left[\frac{1}{2}\boldsymbol{\alpha}'\widehat{\boldsymbol{H}}\boldsymbol{\alpha} - \boldsymbol{\alpha}'\widehat{\boldsymbol{h}} + \frac{\lambda}{2}\boldsymbol{\alpha}'\boldsymbol{\alpha}\right],$$

where $\boldsymbol{\alpha}'\boldsymbol{\alpha}/2$ is a regularizer and λ (≥ 0) is the regularization parameter that controls the strength of regularization. By taking the derivative of the above objective function with respect to the parameter $\boldsymbol{\alpha}$ and equating it to zero, we can analytically obtain the



(a) Lightness

(b) Edge

(c) Color

(d) Combination

Fig. 7 Precision-recall curves.

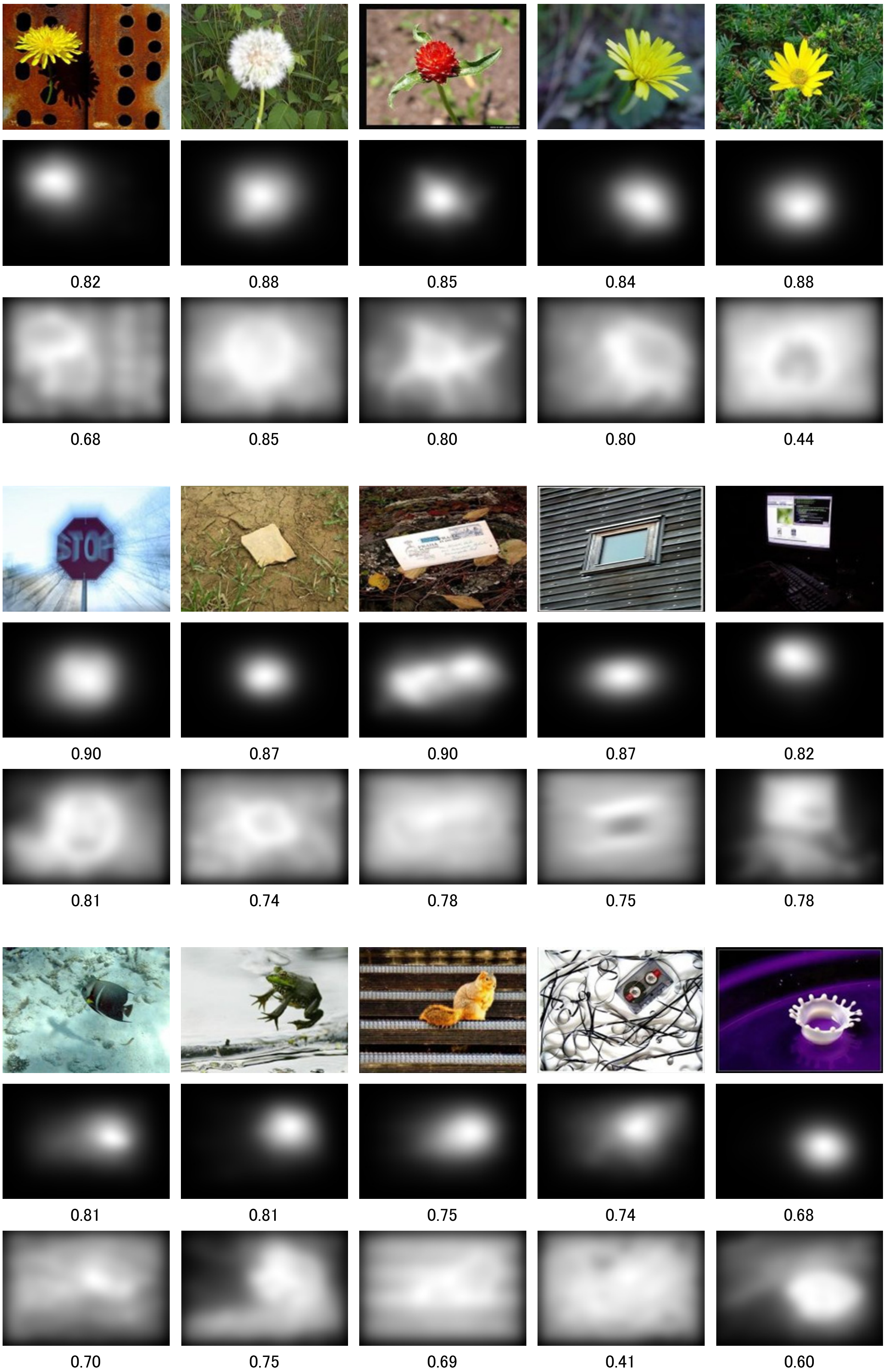


Fig. 8 Experimental results on the MSRA dataset. Top rows: Original images. Middle rows: Saliency maps obtained by the proposed method. Bottom rows: Saliency maps obtained by the K&B method. The number below each image is the maximum F-score.

solution $\widehat{\boldsymbol{\alpha}}$ as

$$\widehat{\boldsymbol{\alpha}} = (\widehat{\boldsymbol{H}} + \lambda \boldsymbol{I})^{-1}\widehat{\boldsymbol{h}},$$

where $\boldsymbol{I}$ denotes the identity matrix. Finally, a density ratio estimator $\widehat{w}(\boldsymbol{x})$ is given by

$$\widehat{w}(\boldsymbol{x}) = \widehat{\boldsymbol{\alpha}}' \boldsymbol{k}(\boldsymbol{x}).$$

Thanks to the simple analytic-form expression, RuLSIF is computationally more efficient than alternative density-ratio estimators which involve non-linear optimization [13].

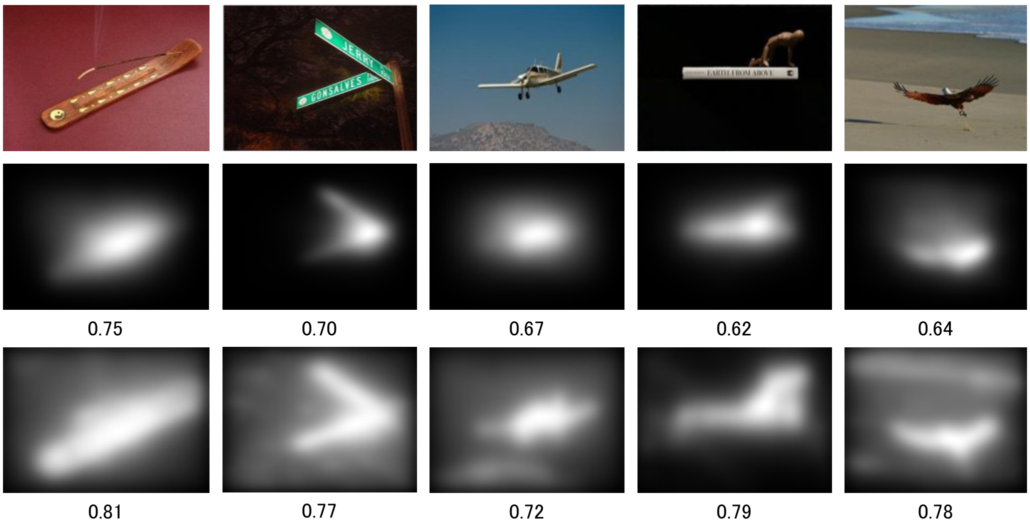


Fig. 9 Examples where the proposed methods do not perform well. The proposed method is not suitable for finding elongated salient regions with the current implementation.

3.3 Model Selection by Cross-validation

The practical performance of RuLSIF depends on the choice of the kernel function (e.g., the kernel width σ in the case of Gaussian kernel Eq. (2)) and the regularization parameter λ. Model selection of RuLSIF is possible based on *cross-validation* with respect to the error criterion J.

More specifically, each of the sample sets $\mathcal{X}_{\mathrm{c}} = \{\boldsymbol{x}_i^{\mathrm{c}}\}_{i=1}^{N_{\mathrm{c}}}$ and $\mathcal{X}_{\mathrm{s}} = \{\boldsymbol{x}_j^{\mathrm{s}}\}_{j=1}^{N_{\mathrm{s}}}$ is divided into L disjoint sets $\{\mathcal{X}_{\mathrm{c}}^l\}_{l=1}^L$ and $\{\mathcal{X}_{\mathrm{s}}^l\}_{l=1}^L$. Then an RuLSIF solution $\widetilde{w}_l(\boldsymbol{x})$ is obtained using $\mathcal{X}_{\mathrm{c}} \backslash \mathcal{X}_{\mathrm{c}}^l$ and $\mathcal{X}_{\mathrm{s}} \backslash \mathcal{X}_{\mathrm{s}}^l$ (i.e., all samples without $\mathcal{X}_{\mathrm{c}}^l$ and $\mathcal{X}_{\mathrm{s}}^l$), and its J-value for the hold-out samples $\mathcal{X}_{\mathrm{c}}^l$ and $\mathcal{X}_{\mathrm{s}}^l$ is computed as

$$\widehat{J}_{\mathrm{CV}}^{\,l} = \frac{\beta}{2|\mathcal{X}_{\mathrm{c}}^l|} \sum_{\boldsymbol{x}_{\mathrm{c}} \in \mathcal{X}_{\mathrm{c}}^l} \widetilde{w}_l(\boldsymbol{x}_{\mathrm{c}})^2 + \frac{1-\beta}{2|\mathcal{X}_{\mathrm{s}}^l|} \sum_{\boldsymbol{x}_{\mathrm{s}} \in \mathcal{X}_{\mathrm{s}}^l} \widetilde{w}_l(\boldsymbol{x}_{\mathrm{s}})^2 - \frac{1}{|\mathcal{X}_{\mathrm{c}}^l|} \sum_{\boldsymbol{x}_{\mathrm{c}} \in \mathcal{X}_{\mathrm{c}}^l} \widetilde{w}_l(\boldsymbol{x}_{\mathrm{c}}),$$

where $|\mathcal{X}|$ denotes the number of elements in the set $\mathcal{X}$. This procedure is repeated for $l = 1, \ldots, L$, and the average of $\widehat{J}_{\mathrm{CV}}^{\,l}$ over all l is computed as

$$\widehat{J}_{\mathrm{CV}} = \frac{1}{L} \sum_{l=1}^{L} \widehat{J}_{\mathrm{CV}}^{\,l}.$$

Finally, the model (the kernel width σ and the regularization parameter λ in the current setup) that minimizes $\widehat{J}_{\mathrm{CV}}^{\,l}$ is chosen as the most suitable one.

4. Experiments

In this section, we experimentally compare the proposed method with the method proposed by Kadir and Brady [7] (K&B) that separately estimates probability densities for low-dimensional features. We use the *MSRA salient object database* [10] for evaluation.

In the K&B method, visual saliency is defined as

$$H(r) \cdot W(r), \tag{3}$$

where $H(r)$ denotes the Shannon entropy in the local region with size r:

$$H(r) = -\int p(I, r) \log p(I, r) \mathrm{d}I.$$

Here, $p(I, r)$ represents the probability density for low-dimensional feature I (e.g., lightness, edge, and color). $W(r)$ in Eq. (3) is the magnitude of low-dimensional feature I defined by

$$W(r) = \frac{r^2}{2r-1} \int \left| \frac{\partial p(I, r)}{\partial r} \right| \mathrm{d}I.$$

Figure 1 shows some images in the database [*1] and ground-truth saliency maps. In the proposed RuLSIF-based method, we fix the parameters at $N_{\mathrm{c}} = 50$, $N_{\mathrm{s}} = 50$, and $\beta = 0.1$. The sizes of center and surround regions are set to 0.2 and 0.3 of the entire image, and the hierarchy of center-surround regions (see Fig. 3) is constructed by decreasing the size by factor $1/\sqrt{2}$ with depth $H = 8$.

The quality of an obtained saliency map is evaluated according to Achanta et al. [1]: A binary map is constructed from an obtained saliency map by varying a threshold on the intensity values in $[0, 255]$. Then each of these 256 maps is compared with the ground-truth binary map and the precision and recall scores are computed.

We compare precision-recall curves for each low-dimensional feature (lightness, edge, and color) and also for the combined feature in **Fig. 7**. The graphs show that our method tends to outperform the K&B method for all low-dimensional feature channels and it more clearly outperforms the K&B method for the combined feature.

Figure 8 shows examples of saliency maps obtained by the proposed approach and the K&B method; below each image, the maximum F-score (i.e., the maximum of the harmonic mean of precision and recall) is described. Overall, the proposed method gives much better results both in visual quality and the F-score.

The processing time necessary for building a conspicuity map of size 300 by 200 [pixels] was about 879 [msec] on a PC with Intel Core2 Duo 2.53 [GHz] and 2.0 [GB] memory. Although this is about 5 times slower than the processing time of the K&B method, our naive implementation may be further improved, e.g.,

[*1] As pointed out in Liu et al. [10], saliency detection for images with large objects is too easy. Here, we choose 200 images from the database that contain small objects of size less than 20% of the image size.

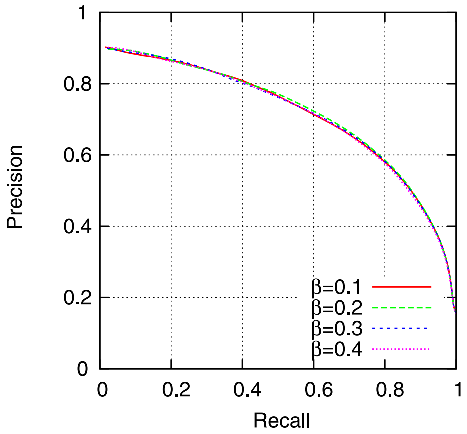


Fig. 10 Precision-recall curves for the lightness channel for several different relative parameters β.

by introducing a coarse-to-fine strategy and parallel computation.

A potential weakness of the proposed method is that if a salient object is highly elongated, its shape cannot be extracted sharply (see **Fig. 9**). This weakness is caused by our search strategy that density ratios between *spherical* regions are estimated. Thus, this weakness may be overcome by considering elongated regions in the saliency search, in exchange for an increase in the computational cost.

Figure 10 depict examples of precision-recall curve of the lightness channel when the relative parameter β in Eq. (1) is changed. The graph shows that the behavior of the proposed algorithm is highly stable with respect to the changes in β, which is a desirable property in practice.

5. Conclusions

We presented a new approach to computing visual saliency based on direct density-ratio estimation. Direct density-ratio estimation is an emerging machine learning technique that allows us to systematically avoid density estimation, which is known to be a hard task. Based on an estimated density ratio, we determined the contrast of the center and the surround feature distributions for lightness, edge, and color channels. Through experiments, we demonstrated that our proposed approach outperforms the K&B method which is based on probability density estimation.

We experimentally found that the proposed method cannot sharply identify a salient object if its shape is elongated, which is due to our search strategy that density ratios between *spherical* regions are estimated. If elongated regions are used for saliency search, this problem can be overcome in principle. Thus, this weakness of the proposed method is not an essential limitation. However, naively employing various elongated regions in saliency search increases the computation cost significantly. Thus, we will develop a computationally efficient way to handle this problem in our future work.

Another important future work is to perform larger-scale and more systematic experiments, including comparison with supervised methods.

References

[1] Achanta, R., Hemami, S., Estrada, F. and Susstrunk, S.: Frequency-tuned salient region detection, *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.1597–1604 (2009).

[2] Alexe, B., Deselaers, T. and Ferrari, V.: What is an object? *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pp.73–80 (2010).

[3] Frintrop, S.: *VOCUS: A Visual Attention System for Object Detection and Goal-directed Search*, Vol.3899 of *Lecture Notes in Artificial Intelligence*, Springer (2006).

[4] Frintrop, S., Rome, E. and Christensen, H.I.: Computational visual attention systems and their cognitive foundation: A survey, *ACM Trans. Applied Perception*, Vol.7, No.1, pp.6:1–6:39 (2010).

[5] Hou, X. and Zhang, L.: Dynamic visual attention: Searching for coding length increments, *Advances in Neural Information Processing Systems 21*, pp.681–688 (2009).

[6] Itti, L., Koch, C. and Niebur, E.: A model of saliency-based visual attention for rapid scene analysis, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.20, No.11, pp.1254–1259 (1998).

[7] Kadir, T., Zisserman, A. and Brady, M.: *An Affine Invariant Salient Region Detector*, Vol.3021 of *Lecture Notes in Computer Science*, Springer (2004).

[8] Kanamori, T., Hido, S. and Sugiyama, M.: A least-squares approach to direct importance estimation, *Journal of Machine Learning Research*, Vol.10, pp.1391–1445 (2009).

[9] Liu, S., Yamada, M. and Sugiyama, M.: Change-point detection in time-series data by relative density-ratio estimation, Technical Report IBISML2011-70, IEICE Technical Report (2011).

[10] Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X. and Shum, H.Y.: Learning to detect a salient object, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.33, No.2, pp.353–367 (2011).

[11] Marchesotti, L., Cifarelli, C. and Csurka, G.: A framework for visual saliency detection with applications to image thumbnailing, *Proc. IEEE International Conference on Computer Vision and Pattern Recognition*, pp.2232–2239 (2009).

[12] Schauerte, B. and Fink, G.A.: Focusing computational visual attention in multi-modal human-robot interaction, *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, pp.6:1–6:8 (2010).

[13] Sugiyama, M., Suzuki, T. and Kanamori, T.: *Density Ratio Estimation in Machine Learning*, Cambridge University Press, Cambridge, UK (2012).

[14] Treisman, A.M. and Gelade, G.: A feature-integration theory of attention, *Cognitive Psychology*, Vol.12, No.1, pp.97–136 (1980).

[15] Vapnik, V.N.: *Statistical Learning Theory*, Wiley, New York, NY, USA (1998).

[16] Walther, D. and Koch, C.: Modeling attention to salient proto-objects, *Neural Networks*, Vol.19, No.9, pp.1395–1407 (2006).

[17] Wolfe, J.M. and Horowitz, T.S.: What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, Vol.5, No.6, pp.495–501 (2004).

[18] Yamada, M., Suzuki, T., Kanamori, T., Hachiya, H. and Sugiyama, M.: Relative density-ratio estimation for robust distribution comparison, *Advances in Neural Information Processing Systems 24*, pp.594–602 (2011).

Masao Yamanaka has worked for corporate R&D Headquaters, Canon Inc. He has been engaged in the development of image information processing technology, such as face recognition in still image, pose estimation in depth image, and anomaly detection of human actions in video so far. His recent research interests include not only a wide range of image information processing techniques but also bioinformatics using advanced machine learning.

Masakazu Matsugu has worked on pattern recognition and neural networks. He received 2002 ICONIP Best Paper Award and 2003 FIT Outstanding Paper Award. He realized a visual inspection system in industrial production system and face identification as well as facial expression functionalities in digital cameras. He holds 110 Japanese patents and 95 US patents. He is a member of INNS.

Masashi Sugiyama received his B.E., M.E., and Ph.D. degrees from Department of Computer Science, Tokyo Institute of Technology, Tokyo, Japan, in 1997, 1999, and 2001. In 2001, he was appointed as a Research Associate in the same institute, and from 2003, he is an Associate Professor. His research interests include theory and application of machine learning and signal/image processing.