Image and Vision Computing 61 (2017) 82–97

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Saliency and KAZE features assisted object segmentation ☆

Prerana Mukherjee*, Brejesh Lall

Department of Electrical Engineering, Indian Institute of Technology, Delhi, India

ARTICLE INFO

Article history:
Received 1 June 2016
Received in revised form 7 November 2016
Accepted 18 February 2017
Available online 6 March 2017

Keywords:
Saliency
KAZE
Unsupervised segmentation
Occlusion boundaries

ABSTRACT

In this paper, we propose an unsupervised salient object segmentation approach using saliency and object features. In the proposed method, we utilize occlusion boundaries to construct a region-prior map which is then enhanced using object properties. To reject the non-salient regions, a region rejection strategy is employed based on the amount of detail (saliency information) and density of KAZE keypoints contained in them. Using the region rejection scheme, we obtain a threshold for binarizing the saliency map. The binarized saliency map is used to form a salient superpixel cluster. Finally, an iterative grabcut segmentation is applied with salient texture keypoints (SIFT keypoints on the Gabor convolved texture map) supplemented with salient KAZE keypoints (keypoints inside saliency cluster) as the foreground seeds and the binarized saliency map (obtained using the region rejection strategy) as a probably foreground region. We perform experiments on several datasets and show that the proposed segmentation framework outperforms the state of the art unsupervised salient object segmentation approaches on various performance metrics.

1. Introduction

Salient object detection and segmentation methods are very important research areas in computer vision because of their widespread applicability in various domains [1–7]. It is quite trivial for human beings to distinguish the salient object from any kind of complex background, or amongst a number of salient objects whereas for a machine it is a very challenging task to identify such objects. The problem of being able to identify and segregate the visually important objects i.e. *salient objects* is defined as salient object segmentation. Numerous techniques have been proposed in literature for solving various aspects of object detection and image segmentation [8–11]. However, solutions which achieve human level accuracy and intelligence are still far from being realized. This can firstly be attributed to the fact that the focus of most of the saliency techniques is either on single salient object segmentation or requirement of user interaction to start with and improve upon successive feedback. Secondly, various attentional models identify different regions as salient and incorrectly capture irrelevant background details. Therefore, achieving a near perfect segmentation using saliency alone is very challenging. The problem is compounded by the fact that in most cases the image is never composed of a single salient object. Finding the object in a highly cluttered environment requires immensely intelligent processing which can be supported by less computationally intensive saliency models. Also in other cases, the object of interest is heavily occluded which makes its segmentation even more challenging. These scenarios motivate the necessity of a segmentation scheme which is able to effectively characterize the object and segment it against complex background. The need for extensive training can be overcome by designing an unsupervised scheme. For the scenarios where training is prohibitive, the proposed unsupervised segmentation is an attractive option because it achieves performance close to deep learning methods. Unsupervised segmentation requires no prior knowledge about the model to which the object should belong instead it clusters the similar tokens or features in the image on the basis of their similarity. Most automatic unsupervised salient object segmentation algorithms make use of the bottom-up saliency computational models [1,12,13] while a few of them rely on the variability in shape or color/texture [14,15].

The proposed saliency based segmentation technique is augmented with cues that *define* the characteristics of an object. In this regard, region/object proposals have found success in recent times [10,16-18]. Alexe et al. [3], stated that an object can be characterized by a well defined boundary, distinctive appearance and region uniqueness. Preserving the edge information (giving the well defined boundary of an object) is crucial. KAZE features [19] retain the boundary information of the objects which is an inherent property of objectness [3]. It uses anisotropic diffusion filtering which is based on non-linear scale space. KAZE detector thus helps in retaining the edge information as opposed to other feature

☆ This paper has been recommended for acceptance by Ming-Hsuan Yang.

* Corresponding author.
E-mail addresses: eez138300@ee.iitd.ac.in (P. Mukherjee), brejesh@ee.iitd.ac.in (B. Lall).

http://dx.doi.org/10.1016/j.imavis.2017.02.008
0262-8856/

detectors such as Scale Invariant Feature Transform (SIFT) [20] and its variants [21,22], which rely on the linear scale space construction. The proposed technique is motivated by a previous work [23] where we showed that objects can be characterized by using a combination of SIFT and KAZE features. It was found that KAZE features are more responsive to the object boundaries, while SIFT keypoints are mostly localized in the salient regions. The combination of SIFT and KAZE was found to be a good mixture of features as they capture both the saliency and boundary properties (key elements in object characterization). Such features constitute just one attribute of an *object* while as mentioned earlier, it constitutes a well defined boundary and unique region with respect to the background. Uniqueness is incorporated by considering the salient features of an object which includes color, intensity, motion, blur, etc. Most of the bottom-up visual attention models address these low-level features for the construction of the saliency maps. If they are supplemented with a prior macro segment regions (based on edge, surface and depth cues) and knowledge of saliency, one can achieve improved segmentation results. In this paper, we provide an end-to-end architecture for unsupervised salient object segmentation by augmenting the saliency map with an edge-aware region prior map. We also propose a novel technique for selecting an appropriate threshold for binarization of the saliency map. It involves utilizing an enhanced weighted saliency map and binarizing it at different grayscale threshold levels. To this end, we make use of salient keypoints to provide the foreground seeds for leveraging the strength in object representation. The key contributions can be summarized as:

1. To the best of our knowledge, this is the first work to exploit the combination of saliency and KAZE features for effective object segmentation. The improvement is achieved by constructing a region prior map which provides a macro level of segmentation in which the regions with high saliency and high density of salient KAZE keypoints are chosen.
2. We show that KAZE keypoints are most suited for characterization of boundaryness. To validate this fact, we also provide an exhaustive empirical evaluation of the effectiveness of KAZE keypoints as compared to other corner and edge based feature detectors.
3. We propose a novel saliency cluster approach to obtain salient keypoints. We also show that the objectness level information is enhanced with the help of these salient keypoints as foreground seeds. The salient KAZE keypoints are supplemented with texture keypoints obtained by SIFT on the Gabor convolved texture map, further strengthening our unsupervised segmentation. These contributions result in improvement in segmentation performance and outperforming state of the art unsupervised segmentation schemes while closing the gap with deep learning based methods.

Overall structure of the paper is as follows. Background about the related works is provided in Section 2. The proposed methodology is explained in Section 3. Results and discussion are given in Section 4 followed by conclusion in Section 5.

2. Related work

In this section, we briefly describe the classical methods which utilize saliency and regional features for salient object segmentation.

2.1. Saliency-based methods

In Ref. [24], the authors focus on two relevant problems in saliency detection: salient object detection and fixation prediction. The saliency detection and fixation algorithms are merged to generate a saliency map. They address the conflict that arises in choosing the salient object identified by the fixation models due to various design biases involved in the construction of the saliency models. A pool of object candidates obtained by Ref. [9] is ranked based on the density of fixation points over them which would essentially result in best salient candidates. Hence, the authors establish a conjunction between human fixations and the saliency models. Another relevant paper in this context [25], entails an automatic object segmentation method using probabilistic edge map with the help of static and motion cues. The edge map is converted to a polar form assuming the fixation points as the poles. The segmentation results are iteratively improved by changing the probabilities in the edge map. As the image is not composed of a single salient object and different objects appear as salient to different people, the authors rely on the fixation points for correctly segmenting the object. Cheng et al. [1] propose a bottom-up data driven saliency detection approach in which global contrast is incorporated to seamlessly highlight the contiguous salient regions in the image. In Ref. [26], the authors integrate context and shape prior for salient object segmentation. Instead of just accounting for the regional saliency alone, the contextual knowledge about the region is also considered. The object prior is calculated by using a Pb detector [27] to accentuate the edges around the object boundaries by gap filling. The final segmentation is obtained by applying an energy minimization framework. Authors in Ref. [28] propose a saliency detection framework using graph based manifold ranking. The image is partitioned in a graph and similarity is computed with respect to the foreground and background cues. Each node on the boundary prior is considered as background labeled query. The saliency map is computed based on the relevance scores given to the query labels (foreground/background). The binary segmentation of the foreground nodes gives the salient queries. In Ref. [29], Liu et al. propose a supervised salient object detection approach. The authors employ local, global and regional saliency features to identify the salient object. Conditional Random Field learning is incorporated for the segmentation. In Ref. [30], an exhaustive empirical comparative analysis between various state of the art saliency detection methods and fixation models has been provided. The models used here have been tested against 7 datasets to show an effective evaluation. In Ref. [31], the authors utilize a graph based salient object ranking method and associate a relationship of the salient regions with the background features.

We propose an unsupervised saliency based segmentation approach in this paper similar to Refs [1,26] and achieve superior performance to these approaches. Refs. [24,29] employ a learning based framework. The method used in Ref. [24] is heavily dependent on the fixation models and on the choice of the segmentation technique to assign saliency score to the salient segments. We exploit the strength of low-level features like SIFT and KAZE with saliency for getting efficient salient object segmentation. Similar to the analogy in Ref. [31] that background features can assist in improving the segmentation accuracy we incorporate the boundary features of the salient objects to retain the distinctiveness of the salient objects.

2.2. Object proposal-based methods

Since last few decades, the research concentrated on generating good object hypotheses has got a lot of thrust. Object proposals are either rectangular bounding boxes enclosing the objects or the regions having higher probability of finding an object. The object proposals aid in segmentation, recognition and detection tasks. In Ref. [32], a highly effective and robust technique for generating and merging object hypotheses is discussed. The effectiveness can be corroborated by the fact that it uses a greedy, hierarchical grouping algorithm as in Ref. [17] using stronger region features than Refs. [10,16] for capturing object proposals. Random Forest based learning is utilized at each stage of the hierarchy to merge stable regions. Authors in Ref. [18]

pioneer the object localization task by introducing objectness measure which steers the localization task towards objects based on the saliency cues. Objectness measure gives a set of bounding boxes which cover the object regions better as compared to contemporary techniques which had maximum coverage on the background regions. It helps in drastically reducing the number of false positive windows returned by most of the class-specific object detectors or blob detectors. In the last few years, despite of having significant improvements in the algorithms generating efficient bounding boxes [33–35], objectness measure stands out as a classic method for obtaining object proposals. Another paper [17] has gained immense popularity in this domain which characteristically embed various sampling strategies for the bounding boxes to generate good object locations.

Apart from bounding boxes, region proposals also serve as good object location candidates. The high level performance of region proposals can be attributed to two reasons. Firstly, pixel-level classification (candidate regions) is better than window classification as it uses additional some other cues for achieving the goal. Secondly, the brute force nature of window classification is computationally expensive. Authors in Refs. [10,16] have done seminal work by generating strong region proposals. They rely on regional features like color similarity, texture similarity, edge density, size, etc. which drive the object localization task.

This paper utilizes a region prior map to select the salient macro segments. We do not consider the bounding box based approaches [17,33-35] for selection of regions. Authors in Ref. [16] generate an overcomplete set of segments taking various permutations of seed values. Our method incorporates the approach in Ref. [36] to generate initial segments and create independent segments.

3. Proposed methodology

In this section, we give a detailed overview of the proposed segmentation scheme. The workflow of the proposed method is shown in Fig. 1. In the following subsections, we describe the components of the proposed method in detail.

3.1. Macro segmentation—region prior map based on occlusion boundaries

Occlusion edges help to separate the cluttered regions (one region occluding other) in the scene. As opposed to occlusion edges, non-occlusion edges are due to shadow casts, reflectance or material/surface changes. Thus, occlusion information has great potential in segregating objects from the background. We have utilized the regions obtained by occlusion boundaries [36] to form a region prior map that provides a macro-level segmentation. The inherent assumption in Ref. [36] is that the image is a 2D projection where the objects get occluded by other objects of a otherwise spatially

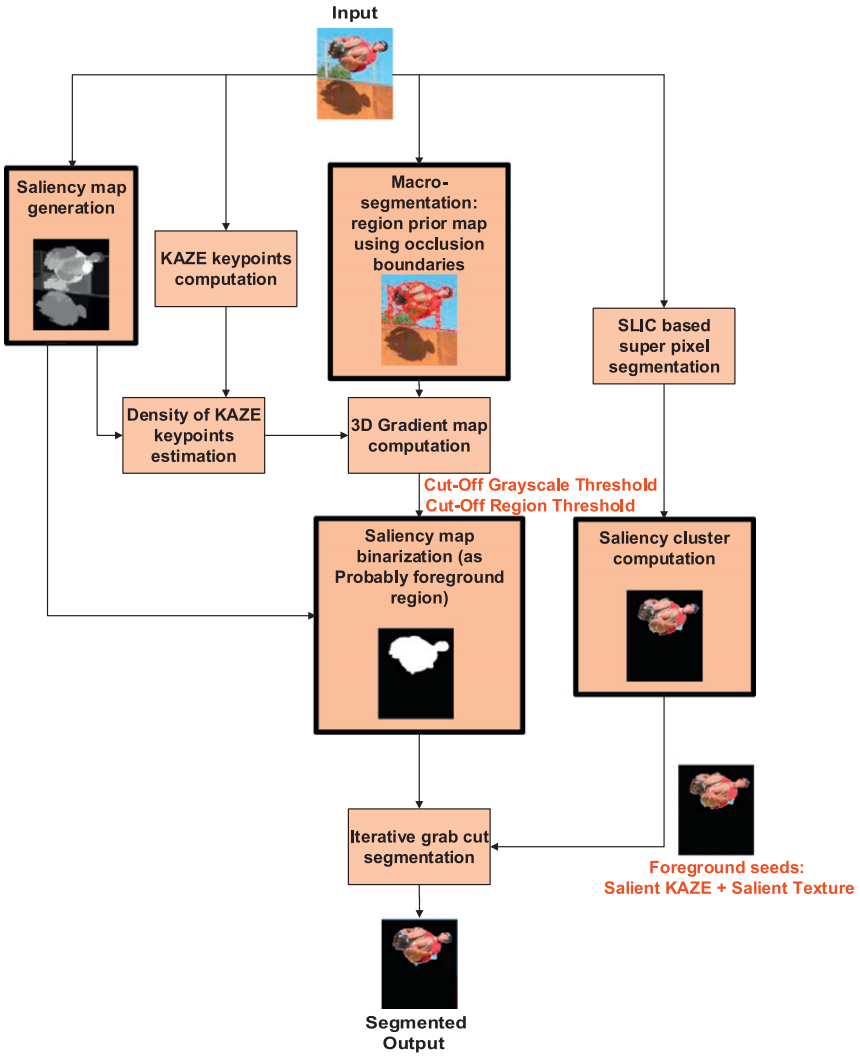

Fig. 1. Workflow of the proposed method.

separated 3D space. It recovers the boundaries and depth ordering of salient objects thus providing a sense of figure/ground separation. Confidence assignment is computed for the boundaries as well as regions. The boundary estimation uses color, surface, depth and Gestalt cues. The initial over-segmentation is constrained by the soft probabilistic boundary map which helps in the removal of weak boundaries resulting in larger regions. The result obtained after this preprocessing is illustrated in Fig. 2. We utilize these regions to select the final candidate regions for the binarized saliency map as explained in the next subsection.

3.2. Salient macro segment selection

A region rejection strategy (Algorithm 1) based on occlusion maps is used to remove the non-salient macro segments from wrongly being classified as salient. For this, we utilize an enhanced weighted saliency map (Section 3.2.1) and binarize it at different grayscale threshold levels to select an appropriate adaptive threshold for binarizing the saliency map. We select only those region proposals which have high percentage of saliency and high density of salient KAZE keypoints. The density of keypoints indicating the number of responses per unit area and the discriminative power of the intensity patches (characterized by high salient regions) captures high distinctiveness of the regions [37]. KAZE features are computed based on non-linear scale space construction (Eqs. (1)–(2)) which otherwise helps in retaining the edge and boundary information which gets lost in linear filtering based features. Non-linear filtering is based on a diffusion coefficient which computes the concentration gradient of gray-values in the image. If the diffusion tensor is constant over the whole image it is called homogeneous/isotropic blurring (as in the case of linear filters) and if it is space-dependent then it is termed as inhomogeneous/anisotropic blurring (non-linear filters) (refer Section A.1.2). The non-linear diffusion equation is given as:

$$\frac{\partial U}{\partial t} = \nabla . \left\{ \left(c\left(x,y,t\right) . \nabla \left(U\right) \right) \right\}, \tag{1}$$

Fig. 2. Region prior map using occlusion edges.

where $\{\nabla .\}$ is the divergence operator, $\nabla(U)$ is the gradient of the original image U, c is the conductivity function and t is the scale parameter. The conductivity function c, is represented as a gradient (Eq. (2)) which helps in retaining edges while smoothening the non-edge regions. Conductivity function c, is given as:

$$c\left(x,y,t\right) = g\left(\left| \nabla U_{\sigma}\left(x,y,t\right) \right| \right), \tag{2}$$

where ∇U_{σ} is the gradient of a Gaussian smoothed original image U and σ is representative of the amount of blur. Other variants of the conductivity functions chosen in Ref. [38], promote high contrast, wider regions or smoothening on both sides of the edges.

Algorithm 1. Algorithm for region rejection.

1: **procedure** REGION–REJECT
2: Compute binary saliency map $S_B^{G_i}$ for saliency map (*salmap*) at different grayscale threshold levels, G_i, $i\epsilon\{1...L\}$ levels between [0-255].
3: Obtain set of primitive regions denoted as, $\{R_k\}_{k=1}^{N}$ using occlusion boundaries, where N, denotes total number of regions. Evaluate percentage of saliency in each primitive region R_k. If the saliency percentage in $S_B^{G_i}$ falls below region threshold (percentage of salient pixels) T_j, $j\epsilon[0.1, 1]$, then the region R_k is rejected.
4: The number of salient KAZE keypoints retained in each saliency map, $S_B^{G_i}$ is noted as a function of region threshold T_j and grayscale threshold G_i, $Salient_{KAZE}(G_i, T_j)$.
5: for each keypoint $KP_{(p)}$
6: $Salient_{KAZE} = \{[Salient_{KAZE}\, KP_{(p)}] \mid KP_{(p)} : S_B^{G_i}\{KP_{(p)}\} = 1\}$
7: end for
8: Compute the gradient of the $Salient_{KAZE}(G_i, T_j)$. Evaluate global maxima of the 3D gradient map and take the corresponding projections on the threshold's coordinate axes to obtain the desired cutoff region threshold and grayscale threshold.
9: Also, the local maxima around the global maxima are used to decide the cutoff thresholds.
10: **end procedure**

KAZE keypoints are highly localized around the object boundaries, thus making them a suitable choice for segmentation (an empirical evaluation of the characterization of the boundaryness measure using KAZE features as compared to other edge-based feature detectors is provided in Appendix A. The analysis clearly demonstrates the superiority of KAZE features for boundaryness). We use this characteristic property to obtain a threshold for binarizing the saliency map. The threshold therefore does not need to be set a priori. We use the fact that a sharp decline (large gradient) in the number of KAZE keypoints as a function of grayscale threshold G_i and region threshold T_j is a good indicator of the operating point for binarization. Using this value of cut-off threshold we obtain all the salient primitive regions in the binarized *salmap* containing high density of KAZE keypoints. The effectiveness of the computed threshold is demonstrated by an empirical comparative analysis with the thresholding techniques used in state of the art salient object detection methods (Section 3.3, Fig. 6).

3.2.1. Weighted saliency map

We utilize the Principal Component Analysis (PCA) based weighted combination of two state-of-the-art saliency map estimation methods: Discriminative Regional Feature Integration (DRFI [39]) and Hierarchical Saliency (HS [40]) approaches. The choice behind these two saliency maps lies in the fact that DRFI selects distinctive regional features using regression scheme based on the contrast and backgroundness details at multiple levels of segmentation. HS is able to highlight the salient objects at multiple granularity levels. In the PCA based method, we obtain a column vector from each saliency map which is used to calculate a covariance matrix. We normalize

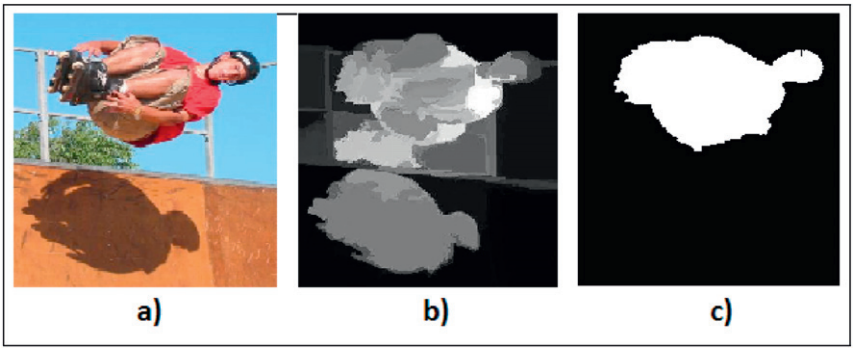

Fig. 3. (a) Original image, (b) weighted fused saliency map, (c) final binarized saliency map based on region rejection strategy.

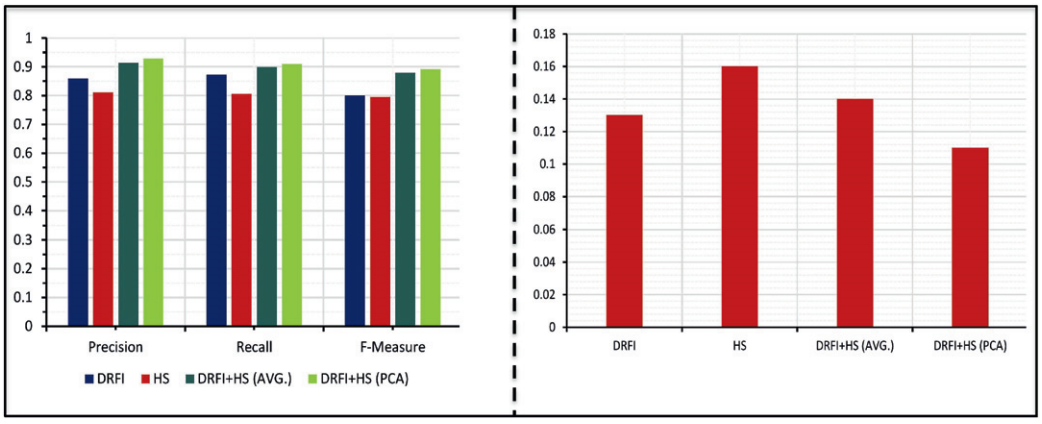

Fig. 4. Performance evaluation of the fusion strategies.

the column vector corresponding to the largest eigen value by dividing with the mean of eigen vectors. Normalized eigen vector values act as the weights. The linear combination of the respective saliency maps with these weights gives the weighted fused saliency map (Fig. 3 (b)). To motivate the use of PCA based weights we compare it with the averaged fusion of the saliency maps and report the performance boost achieved by PCA based weights over simple averaging results. We report the average precision, recall, F-Measure and MAE (mean absolute error) rates on MSRA-1000 dataset using these fusion strategies in Fig. 4. PCA based combination achieves 92.83% precision and 90.99% recall while averaging gives 91.28% precision and 89.94% recall. As compared to the methods before fusion, F-Measure increases from 0.8002 (DRFI) and 0.7954 (HS) to 0.8912 (DRFI + HS-PCA) and there is an increase in precision rate by 7% over DRFI and 10% over HS. The reason for the performance gain is due to the fact that extracting the principal components results in minimization of redundancy and the selection of self adaptive weighted coefficients reduces the MAE (mean absolute error) by a margin of 3% as compared to direct averaging methods. MAE is given as the average of pixelwise absolute difference between the binary ground truth and the computed saliency map [41].

During experimentation, we found that utilizing these saliency maps alone as the initial input to any segmentation method is not able to clearly segment the objects and retain the object boundaries (Section 4.2). It is observed that although the use of stronger saliency maps is able to detect the salient objects pretty well but their results obtained on graph based segmentation methods [1] (SaliencyCut: an improvised version of iterative GrabCut for segmenting salient objects) are not good. Thus, we show that our proposed segmentation method has a clear advantage over saliency map only being used as initial masks for segmentation.

3.3. Modified iterative grab-cut segmentation

In addition to binarized saliency map as a mask for segmentation we provide object features as foreground seeds. First, we perform super-pixel segmentation to obtain a set of superpixels.

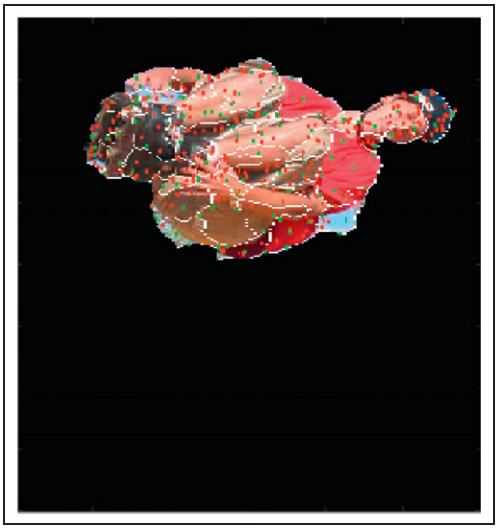

Fig. 5. Salient KAZE and TEXTURE keypoints inside the saliency cluster. Red denotes the KAZE keypoints and green denotes SIFT keypoints.

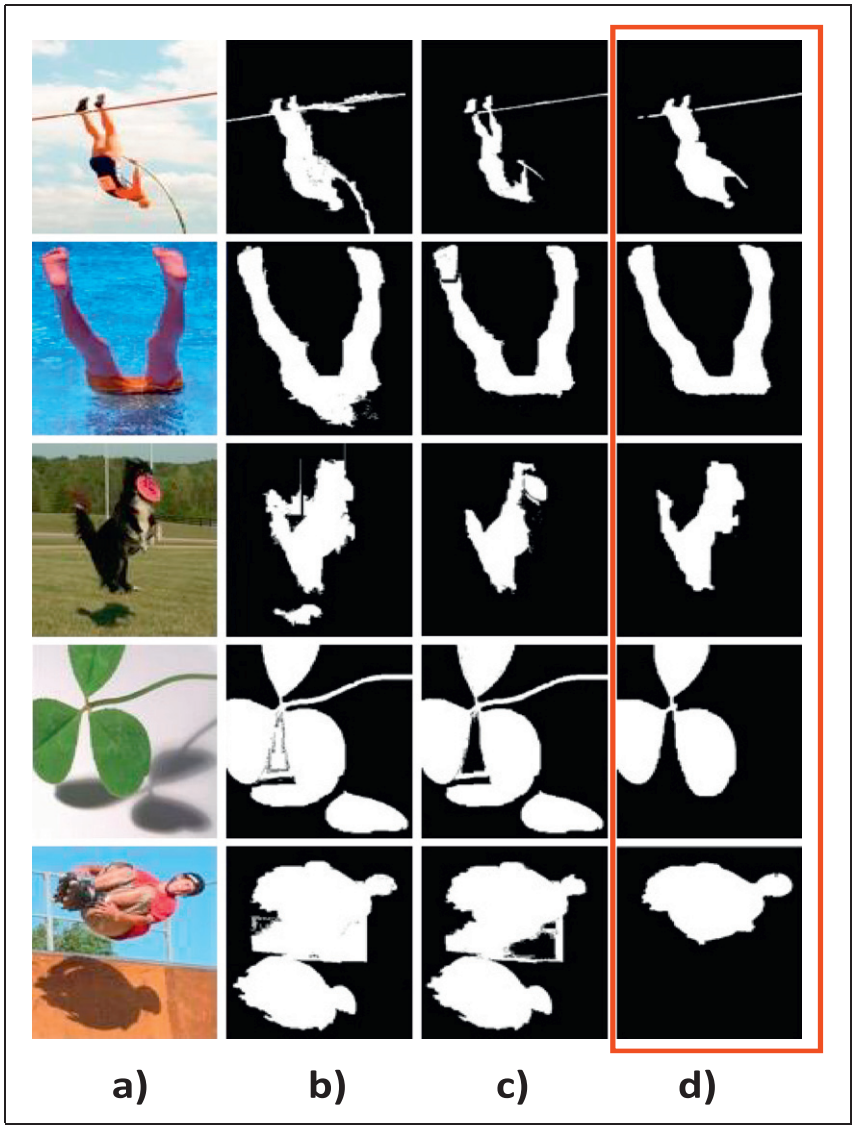

Fig. 6. (a) Original image, (b) thresholding using mean [43], (c) thresholding using Otsu's method [45], (d) binarization by our proposed method using 3D gradient plot.

Using, the region rejection strategy as explained in Section 3.2 we obtain the final binarized saliency map. We then form a group of salient superpixel cluster, saliency cluster $\{SC_i\}_{i=1}^{K}$, containing all the superpixels which are *highly* salient (more than 50% salient) using the binarized saliency map. We then calculate the SIFT keypoints on the original image. An orientation histogram is formed using the orientations of the SIFT keypoints. Using the bin centers as the orientations, the image is convolved with Gabor filter giving a texture map. Next, the SIFT keypoints are computed on the texture map and only those keypoints which lie in the saliency cluster SC_i are retained and termed as the salient texture keypoints $Salient_{Texture}$. Similarly, the final salient KAZE keypoints $Salient_{KAZE}$ retained are the ones which lie inside the saliency cluster. Salient keypoints refer to the keypoints with high saliency score (lying in the saliency cluster). These salient keypoints (texture and KAZE) are used as foreground seeds in the iterative grab-cut segmentation [42]. The binarized saliency map, $S_B^{G_i}$ is used as the probable foreground region prior for segmentation. The inclusion of these keypoints enhances the objectness measure compatibility of the proposed features thus, resulting in robust salient object segmentation (Fig. 5). The algorithm for the selection of the foreground seeds in iterative grabcut segmentation scheme has been detailed in Algorithm 2.

Algorithm 2. Modified iterative grab-cut segmentation.

1: **procedure** ITERATIVE −GRAB-CUT
2: Obtain set of superpixels on original image. Form saliency cluster $\{SC_i\}_{i=1}^{K}$, K: total number of salient superpixels, $\left\{i \epsilon SC_{i=1}^{K} | \frac{\#salient}{\#Total} > 0.5\right\}$,
where $\#salient$:total number of salient pixels in superpixel i, $\#total$:total number of pixels in superpixel i.
3: Foreground seeds for grabcut segmentation:
Compute SIFT keypoints on original image.
Form orientation histogram using orientations of computed SIFT keypoints.
Convolve Gabor filter using cluster centers of the bins of orientation histogram. $G_{\lambda,\theta,\psi,\sigma,\gamma}(x,y) = e^{\frac{-u'^2+\gamma^2 v'^2}{2\sigma^2}} e^{i(\frac{2\pi u'}{\lambda}+\psi)}$,
where λ: wavelength of sinusoidal function, θ: orientation of the function, ψ: phase offset, σ: standard deviation of the Gaussian function, γ: spatial offset ratio, $u' = x\cos\theta + y\sin\theta$, $v' = -x\cos\theta + y\cos\theta$.
4: $Salient_{Texture} = \left\{SIFT_{KP} | SIFT_{KP} \epsilon \{SC_i\}_{i=1}^{K}\right\}$.
5: $Salient_{KAZE} = \left\{KAZE_{KP} | KAZE_{KP} \epsilon \{SC_i\}_{i=1}^{K}\right\}$.
6: Input for grab cut segmentation:
$ProbablyFG \leftarrow S_B^{G_i} : binarized \quad saliency \quad map$
$FGseeds \leftarrow [Salient_{Texture} \quad Salient_{KAZE}]$,
where FG: foreground
7: **end procedure**

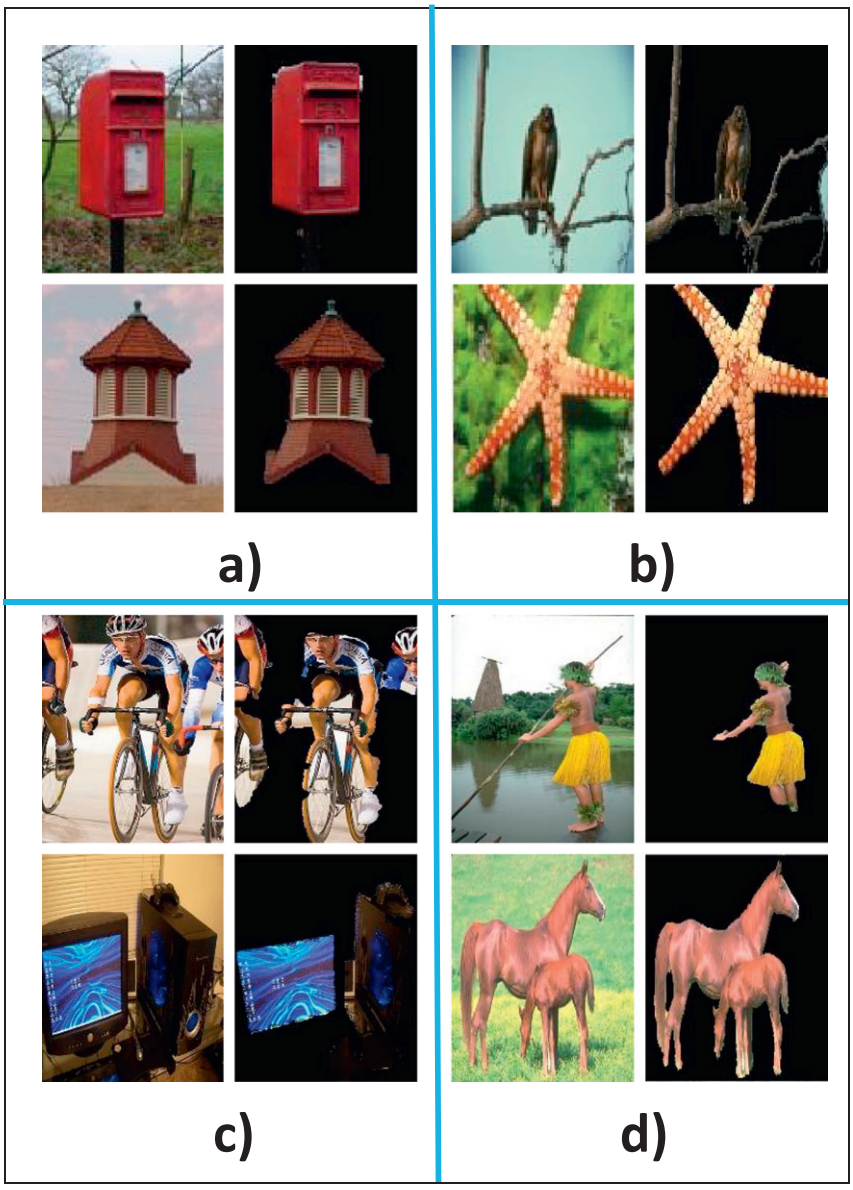

Fig. 7. Segmentation results obtained by the proposed method: single salient object per image in (a) MSRA1000 dataset, (b) SOD dataset; multiple salient objects per image in (c) PASCAL-1500 dataset, (d) ECSSD dataset.

Most of the saliency detection methods assume a fixed threshold for binarizing the saliency map [1,43]. They assume the threshold as a value between [0 and 255] where the maps give good precision and recall rates. These approaches however require the ground truth to be known beforehand. This choice of threshold introduces a dataset bias. In adaptive thresholding, authors utilize some statistical measure like mean [43], variance [44] (as used in Otsu's method) for the saliency map binarization. This may not work in the case of highly cluttered and occluded cases. In our proposed method, we have removed such dataset biases by making threshold as a function of saliency as well as density of salient object features. As illustrated in Fig. 6 (for MSRA-1000 dataset) mean and Otsu's threshold fail to clearly binarize the salient object (give some redundant background as well). Another strong aspect in the proposed segmentation scheme is that we have also taken salient texture keypoints along with salient boundary keypoints resulting in a well defined boundary (given by KAZE) and distinctive appearance (combined by texture and saliency in our case) which are the characteristic features of an object. Since, this method makes use of the occlusion reasoning for the region prior map, every object is thoroughly analyzed even if it is partially visible and is able to label it as a single salient object. Prior methods found this challenging as they couldn't recognize object with multiple color and positioning as a single entity because of the discontinuity/non-connectivity of the object.

4. Experimental results and analysis

4.1. Experimental setup

The experiments were performed on a 32 GB RAM machine with Xeon 1650 processor and 1 GB NVIDIA Graphics Card. Matlab 2015b was used as the programming platform. The evaluation of

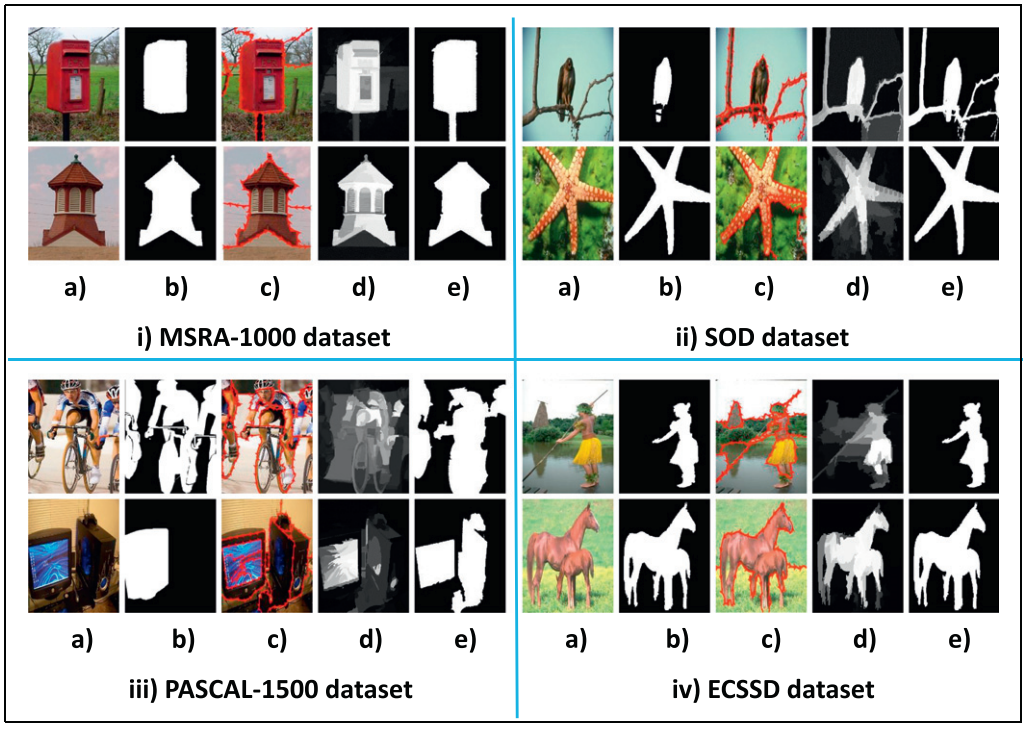

Fig. 8. Corresponding intermediate results for i) MSRA-1000, ii) SOD, iii) PASCAL-1500 and iv) ECSSD datasets: (a) original images, (b) ground truth, (c) region prior map obtained from the occlusion boundaries, (d) weighted saliency map, (e) final binarized saliency map obtained from the region rejection strategy in Algorithm 1. This binarized saliency map is given as a probably foreground region to the iterative grabcut algorithm.

the proposed methodology was performed on the publicly available saliency datasets: MSRA1000 [43], SOD [46], PASCAL-1500 [47] and ECSSD [40]. MSRA1000 dataset predominantly contains single salient object per image. SOD dataset contains 300 images from the Berkeley segmentation dataset. SOD and PASCAL-1500 are quite challenging saliency datasets. PASCAL-1500 dataset contains a subset of selected images from PASCAL VOC segmentation challenge dataset [48]. It contains multiple salient objects in the images appearing at various scales and locations in cluttered background which makes it challenging. ECSSD dataset is another recently introduced challenging dataset. It contains structurally complex images which are semantically meaningful. In the following subsections, we discuss the results and provide a comparative analysis with the state of the art methods.

4.2. Segmentation results

Since, we have proposed an unsupervised saliency and object keypoint driven segmentation scheme, we have done comparative analysis with saliency based segmentation methods and automatic segmentation schemes. Fig. 7 shows the segmentation results on few images from the tested datasets using the proposed approach. To put our work in perspective, we show that the proposed technique gives segmentation results close to deep learning based approaches.

The intermediate results consisting the region prior map, weighted saliency map, final binarized saliency map are shown in Fig. 8. This region map is used to obtain a 3D gradient map (Algorithm 1). The cut-off threshold values obtained using Algorithm 1 are used to get the binarized saliency map. So, rather than choosing a threshold based on fixed or adaptive threshold, this is a better measure for computing threshold as the object boundaries are preserved which ordinarily get lost when mean or variance values are used for thresholding. The number of region thresholds is chosen in the range [0.1–1] with a step size of 0.1. The number of gray level thresholds is set between [0 and 255] with 25 intermediate levels. In our experiments we found that these values provide the best results. The binarized saliency map obtained by the region rejection strategy is used as the probable foreground region in the iterative grabcut segmentation scheme. We use SLIC superpixel segmentation scheme [49] to obtain the salient keypoints (Section 3.3). The initial number of superpixels was set to 200. The superpixels containing more than 50% salient pixels were chosen which forms the saliency cluster. We use the KAZE features [19] implementation and OpenCV 3.1.0 [50] for setup of KAZE. For calculating SIFT features we use VLFeat [51] implementation. The number of bins for the orientation histogram in the iterative grabcut segmentation (Algorithm 2) is set to 10. The parameter values chosen for the Gabor filter are λ: 8, θ: 0, ψ: 0, γ: 0.5, σ: 1. The results of the SLIC based segmentation giving the saliency cluster and salient keypoints are shown in Fig. 9.

Fig. 10 shows the output of the proposed segmentation scheme as compared to other approaches. It can be observed that the shadow of the skater's image and the green foliage (MSRA-1000 dataset) forms the part of the final segmented image with most of the methods [1,26,39,40,43,44,52-54]. While it is not the case with the proposed method. The reason is that the methods in Refs. [1,26,53,54] detect the shadow as a salient image which is not a part in the saliency ground truth. Similar observation can be made for the green foliage (skater's image) which is labeled as salient by Refs. [1,26,39,40,44,52-54]. Similarly, most methods fail to segment out all the cyclists correctly (PASCAL-1500). Congruent analogies can be drawn from other images. For a quantitative evaluation comparing prior works, we use the following standard objective measures, Probabilistic Rand Index (PRI), Variation of Information

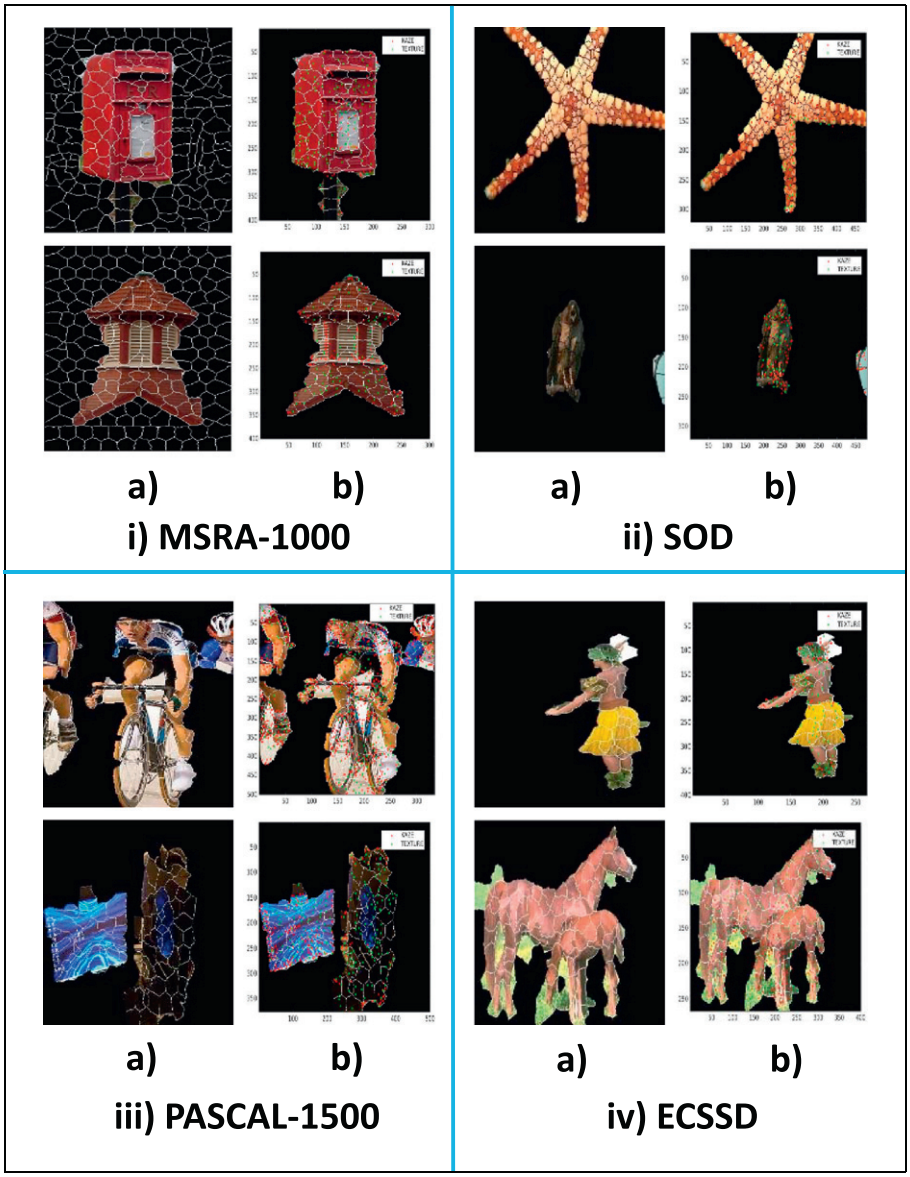

Fig. 9. (a) and (c): Saliency cluster based on regions obtained by SLIC based segmentation on MSRA-1000 and SOD datasets respectively; (b) and (d): salient keypoints in the saliency cluster on PASCAL-1500 and ECSSD datasets respectively (KAZE keypoints: red and TEXTURE SIFT: green) given as the foreground seeds.

(VoI), Global Consistency Error (GCE) and Boundary Displacement Error (BDE) [55]. The range of PRI and GCE is [0, 1] and range of VoI and BDE is [0, ∞]. We report the average value of these measures on these datasets. Higher value of PRI and lower values of GCE, VoI and BDE indicate good performance in segmentation. We utilize the saliency maps obtained by FT [43], SEG [52], RC [1], DRFI [39], HS [40], PISA [54], MDF [56], DeepSaliency [57] as the initial masks to iterative grabcut for salient objects [1] (SaliencyCut) and report their performance evaluation with our method. In order to generate the saliency maps by the deep learning methods, the publicly available Matlab code [56,57] was run on a system with NVIDIA GeForce GTX Titan X GPU, 3072 cores and Intel Xeon 3.5 GHz processor on Ubuntu 14.04, Matlab 2015b. The publicly available code for SaliencyCut [1] was used and run on Visual Studio 2012 integrated with OpenCV 3.1.0. SaliencyCut uses the computed saliency maps and aids in automated salient object segmentation and doesn't require any manual annotation to be provided. Iterative runs of GrabCut [42] with adaptive fitting reduce the noise which might get incorporated in the saliency map computation thus resulting in a better and robust segmentation. Our method was able to outperform the prior methods on MSRA1000 dataset (Table 1) in terms of VoI. The second best method [57] lags by a margin of 0.0153 (VoI value). Deep learning methods outperform for the other three indexes PRI, GCE and BDE values (PRI: higher by a margin of 0.0154, GCE: higher by a margin of 0.0175 and BDE: higher by 0.0773). In the case of SOD dataset (Table 2), the proposed method lags behind the DeepSaliency method [57] (PRI: less by 0.0642, GCE: less by 0.0011 and BDE: less by 0.4958) but for VoI index proposed method performs much better as compared to other methods. For Pascal-1500 dataset (Table 3) the proposed method is able to outperform in terms of VoI and GCE by a margin of 0.0044 and 0.0169 respectively with respect to the second best. In the case of ECSSD (Table 4), the proposed method is able to outperform the other comparable methods in terms of low VoI, GCE and BDE. Low values of BDE suggest that the contours

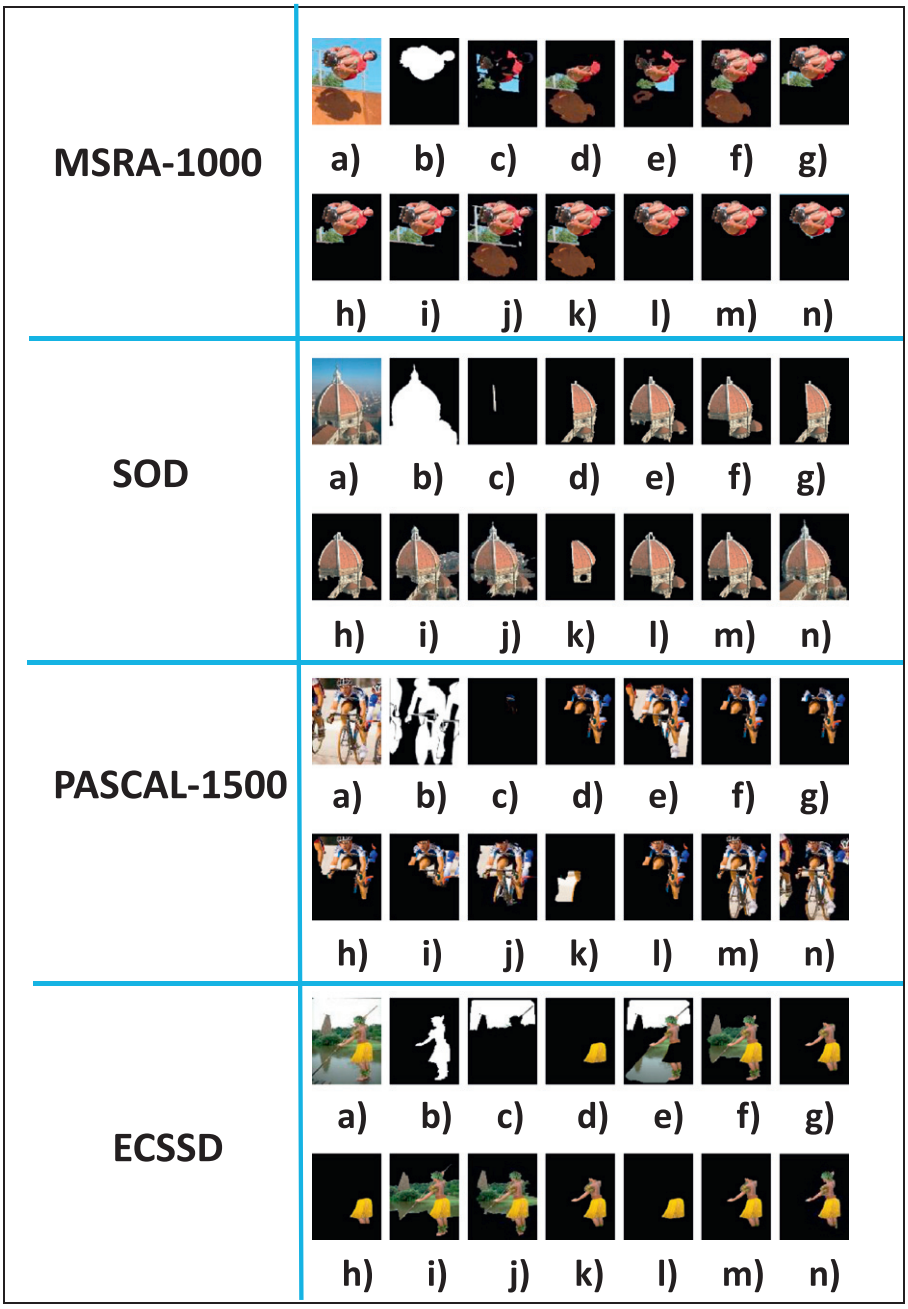

Fig. 10. (a) Original image, (b) ground truth, (c) frequency tuned + SaliencyCut [43], (d) context and shape prior based segmentation [26], (e) SEG + SaliencyCut [52], (f) RC + SaliencyCut [1], (g) DRFI + SaliencyCut [39], (h) HS + SaliencyCut [40], (i) salient Harris + star shape prior [44], (j) QCUT [53], (k) PISA + SaliencyCut [54], (l) MDF + SaliencyCut [56], (m) DeepSaliency + SaliencyCut [57], (n) proposed approach.

are preserved better. It is observed that the proposed method (combination of DRFI + HS(PCA) along with salient keypoints) is able to outperform the unsupervised salient object segmentation techniques considerably while also closing the gap with CNN-based techniques [56,57], reducing the computational overhead of training with a huge corpus of image data with pixel-level segmentation annotations. It significantly outperforms the DRFI and HS saliency schemes when used individually as priors for segmentation. The results indicate that the proposed scheme is consistently able to give good results on the datasets having large variation in terms of number of salient objects, size of salient object and contrast between them. Furthermore, in more challenging datasets like PASCAL-1500 and ECSSD it is able to characterize the salient objects and also retain the object boundaries. The average execution time of the proposed unsupervised segmentation framework on MSRA-1000 dataset (400 × 300 pixels) including the overhead of saliency computation, region prior map construction, low-level features computation (foreground seeds) is provided in Table 5.

5. Conclusion

We have proposed an automatic unsupervised salient object segmentation approach using saliency and object features. The prior

Table 1
MSRA1000: performance analysis.

Method	PRI	VoI	GCE	BDE
FT + SaliencyCut [43]	0.8704	0.5483	0.0897	16.3453
CB [26]	0.9334	0.2912	0.0433	7.3414
SEG + SaliencyCut [52]	0.8813	0.4789	0.0722	13.4110
RC + SaliencyCut [1]	0.9386	0.2836	0.0434	6.7060
DFRI + SaliencyCut [39]	0.9346	0.2348	0.0492	4.9232
HS + SaliencyCut [40]	0.9105	0.2592	0.0432	4.9284
Salient Harris + star shape prior [44]	0.9396	5.4231	0.8256	0.3452
Saliency + random walk [58]	0.9276	8.5423	2.2763	0.2934
Active contour [59]	0.7982	0.6221	0.1052	14.2343
QCUT [53]	0.9023	0.2843	0.8772	4.3293
PISA + SaliencyCut [54]	0.9345	0.2668	0.0528	4.2384
MDF + SaliencyCut [56]	0.9584	0.1837	0.0348	0.2532
DeepSaliency + SaliencyCut [57]	**0.9606**	0.1429	**0.0297**	**0.2040**
Proposed method	0.9452	**0.1276**	0.0472	0.2813

The value in bold indicates the highest PRI value and lowest values of VoI, GCE and BDE for Tables 1–4.

Table 2
SOD: performance analysis.

Method	PRI	VoI	GCE	BDE
FT + SaliencyCut [43]	0.8342	0.6777	0.7234	15.1412
CB [26]	0.8763	0.0235	0.0833	7.7664
SEG + SaliencyCut [52]	0.8412	0.5437	0.0857	14.5728
RC + SaliencyCut [1]	0.9137	0.3125	0.0683	8.0542
DFRI + SaliencyCut [39]	0.8724	0.5278	0.2339	8.2343
HS + SaliencyCut [40]	0.8612	0.5582	0.2623	7.2345
Salient Harris + star shape prior [44]	0.8772	9.4231	0.1296	7.5673
Saliency + random walk [58]	0.8523	8.2323	3.7225	8.3884
Active contour [59]	0.7198	0.9178	0.1447	31.9205
QCUT [53]	0.8623	0.6792	0.8232	7.1298
PISA + SaliencyCut [54]	0.9237	0.5239	0.3776	7.2387
MDF + SaliencyCut [56]	0.9387	0.0287	0.0602	6.7773
DeepSaliency + SaliencyCut [57]	**0.9476**	0.0176	**0.0523**	**6.2887**
Proposed method	0.8834	**0.0133**	0.0534	6.7845

The value in bold indicates the highest PRI value and lowest values of VoI, GCE and BDE for Tables 1–4.

Table 3
PASCAL-1500: performance analysis.

Method	PRI	VoI	GCE	BDE
FT + SaliencyCut [43]	0.6927	7.2998	0.7234	15.1412
CB [26]	0.7623	0.3476	0.9827	10.5839
SEG + SaliencyCut [52]	0.7458	0.9376	0.5482	11.3872
RC + SaliencyCut [1]	0.7876	0.3387	0.9878	6.7284
DFRI + SaliencyCut [39]	0.8019	0.3498	0.7882	4.4599
HS + SaliencyCut [40]	0.7965	0.4534	0.8712	4.0923
Salient Harris + star shape prior [44]	0.7989	0.2378	0.1998	4.7656
Saliency + random walk [58]	0.8176	0.5423	2.2265	5.3434
Active contour [59]	0.6892	6.7223	5.6723	26.7254
QCUT [53]	0.8012	0.3454	2.3776	5.0234
PISA + SaliencyCut [54]	0.8097	0.4512	0.3487	4.3498
MDF + SaliencyCut [56]	0.8432	0.4387	0.0498	3.2934
DeepSaliency + SaliencyCut [57]	**0.8472**	0.0400	0.0401	**3.0972**
Proposed method	0.8364	**0.0356**	**0.0232**	3.2465

The value in bold indicates the highest PRI value and lowest values of VoI, GCE and BDE for Tables 1–4.

region map is obtained using occlusion boundaries. A region rejection strategy is presented to reject the less salient regions. An iterative grabcut is then provided with salient boundary (KAZE) and salient texture (SIFT) keypoints as the foreground seeds. The proposed method is shown to outperform the unsupervised state of the art techniques while closing the gap with state of the art CNN-based approaches with exhaustive experiments on numerous datasets. Additionally, the efficacy of KAZE features over other edge based and corner features for the characterization of boundaryness is demonstrated.

Table 4
ECSSD: performance analysis.

Method	PRI	VoI	GCE	BDE
FT + SaliencyCut [43]	0.7342	0.9882	0.7824	19.2991
CB [26]	0.7823	0.2332	0.1833	7.4552
SEG + SaliencyCut [52]	0.7934	0.2454	0.1857	17.1224
RC + SaliencyCut [1]	0.8002	0.3235	0.0883	9.0568
DFRI + SaliencyCut [39]	0.8343	0.1229	0.0816	8.0128
HS + SaliencyCut [40]	0.8354	0.2893	0.1993	8.0228
Salient Harris + star shape prior [44]	0.8232	7.2321	0.1221	6.5672
Saliency + random walk [58]	0.8076	9.3776	5.1213	5.2974
Active contour [59]	0.6523	0.8233	0.1776	25.9105
QCUT [53]	0.8382	0.7283	0.1882	8.7823
PISA + SaliencyCut [54]	0.8513	0.0556	0.0968	6.0002
MDF + SaliencyCut [56]	0.8847	0.0423	0.0876	5.9217
DeepSaliency + SaliencyCut [57]	**0.8876**	0.0413	0.0822	5.2242
Proposed method	0.8737	**0.0376**	**0.0789**	**4.7845**

The value in bold indicates the highest PRI value and lowest values of VoI, GCE and BDE for Tables 1–4.

Table 5
Average execution time of the proposed method on MSRA-1000 dataset.

Steps	Region prior map computation using occlusion boundaries	Saliency map computation	Threshold obtained by region rejection strategy	Steps foreground seeds (SIFT on Gabor texture map + KAZE)	Iterative grab cut
Average cost (in s)	10.23	12.11	11.04	5.36	1.03

Total cost (in s): 39.77

Appendix A. Characterization of boundaryness

Local features give efficient image representation and accentuates the details in the image. They are generally concentrated on the local areas in the image including boundary segments, curvature and corner points. They give keypoints which are either corners or blobs. The keypoints thus obtained rarely get localized near the edges. In case we need to obtain the corner points on edges generally the corner points near the edges (given by Canny detector) are selected. The strong edges are mostly located on the object boundaries. Edges change their appearance as they change their locations when the image is scaled over multiple scales [60]. In Ref. [60], authors introduce edge based features (EDGELAP KP) which focuses on selecting keypoints along edges. The edge appearance is captured by building a scale space representation. Laplacian responses are calculated for several scales around the edge points to select the extremum points. This EdgeLap detector essentially uses a Gaussian scale space due to which the edges lose their appearance due to blurring. Other edge specific interest point detectors include edge-foci interest points (EFI KP) [61] and edge-based region keypoints (EBR KP) [62]. We also compare KAZE keypoints with Context Aware Keypoint Extractor keypoints (CAKE KP) which give keypoints corresponding to local structures in the image based on the image context. KAZE features [19] build up a non-linear scale space which is robust to noise and increases localization accuracy. It maintains the object boundaries. We have compared the robustness of KAZE keypoints against various edge based keypoint detector and corner detectors to show the effectiveness of KAZE particularly for maintaining boundaryness. In the first set of experiments we see the distribution of KAZE keypoints around the gradient responses at various scales. In the second set of experiments we compare the effectiveness of KAZE keypoint responses with other popular edge based detectors around the boundary regions. In order to explain the effectiveness of KAZE against other techniques, we provide a comparison of linear scale

space and non-linear scale space construction followed by an empirical evaluation strengthening the theoretical justification to bring out their effectiveness in boundary representation.

A.1. Scale space representation

A.1.1. Linear scale space

Given the initial image $f(x,y)$, first we treat it with zero order scale space (i.e. Gaussian scale space). In such a representation, significant image structures coexist in the smoothened regions which have high contrast with respect to their background and are highly distinct from their surroundings. Such structures are called *blobs* [63]. Detection of such image structures and relation between them at different scales along with scales at which these arise result in salient features being segmented out at coarser scale.

The following scale space representation is obtained for the given image, $L : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$ such that:

$$L(\cdot; t) = K(\cdot; t) * f(\cdot), \tag{A.1}$$

where, t is scale parameter and K denotes kernel of successively increasing width, and

$$L(\cdot; 0) = f(\cdot). \tag{A.2}$$

In terms of differential equation, L can be described by the *diffusion equation*:

$$\partial_t L = \frac{1}{2}\nabla^2 L, \tag{A.3}$$

where ∇ denotes the gradient operator. *Causality* property ensures that new level surfaces do not spur out when the scale parameter is increased. *Homogeneity and Isotropy* property essentially dictates the treatment of all spatial points and scale levels in a similar manner. *Linearity and Shift Invariance* properties are ensured in the scale space construction [64]. It can be shown that the kernel $K(\cdot; t)$ assumes the form of Gaussian kernel [65]. Thus, Eq. (A. 1) can be rewritten with a Gaussian kernel $G : \mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ as:

$$G(x,y;t) = \frac{1}{2\pi t} e^{-(x^2+y^2)/2t}. \tag{A.4}$$

Assuming $t = \sigma^2$, the heat diffusion Eq. (A. 3) becomes:

$$\begin{aligned} &\frac{\partial L}{\partial \sigma^2} = \frac{1}{2}\nabla^2 L \\ \Rightarrow\ &\frac{\partial L}{\partial \sigma} = \sigma\nabla^2 L [\because \partial\sigma^2 = 2\sigma\partial\sigma], \end{aligned} \tag{A.5}$$

which result to (using Eq. (A. 1)):

$$\begin{aligned} &\frac{\partial (f * G)}{\partial \sigma} = \sigma\nabla^2 (f * G) \\ \Rightarrow\ &f * \frac{\partial G}{\partial \sigma} = \sigma f * \nabla^2 G \\ \Rightarrow\ &\frac{\partial G}{\partial \sigma} = \sigma\nabla^2 G. \end{aligned} \tag{A.6}$$

The interest points are the local maxima in the scale space of Laplacian of Gaussian (LoG). The Laplacian pyramid is obtained using the Difference-of-Gaussian (DoG) function which acts as a band pass filter. DoG function is given as:

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y). \tag{A.7}$$

This equation represents difference of two nearby scales separated by a constant multiplicative factor k. The scale normalized Laplacian of Gaussian is $\sigma\nabla^2 G$. Drawing analogy from the heat equation: $\sigma\nabla^2 G = \frac{\partial G}{\partial \sigma}$ and using the approximation, $\frac{\partial G}{\partial \sigma} \approx \frac{G(x,y,k\sigma) - G(x,y,\sigma)}{(k\sigma - \sigma)}$. We deduce that:

$$G(x,y,k\sigma) - G(x,y,\sigma) \approx (k-1)\sigma^2\nabla^2 G \approx (k-1)\sigma(LoG). \tag{A.8}$$

The factor $(k-1)$ in the equation is a constant over all scales and therefore does not influence extrema location. The next task is to detect the accurate keypoints in the image that is done by comparing the keypoints with their neighbors. Thus, the point of extrema in DoG as selected in Ref. [20] will have a corresponding set with addition of scale space points to incorporate qualitative information about the image. As we move from finer to coarser scales, linear diffusion filtering dislocates the image structures. So, we need some localization method for these structures. This can be achieved by using the Taylor expansion of the scale space function or DoG (Difference of Gaussian), $D(x,y,\sigma)$, shifted so that origin is at the stable keypoint.

A.1.2. Non-linear scale space

Linear diffusion does not preserve the edge information and smoothens the entire image (I) uniformly (Eq. (A. 4)). Higher 'σ' value blurs over a wider radius. This also results in a larger kernel matrix to capture most of the function's energy. The aim is to capture the maximum area of the object with minimal overlap with other objects. Non-linear diffusion on the other hand reduces noise and preserves the contours/boundaries in the images. The diffusion coefficient is adaptive to the image data and remains negligible in the case of object boundaries. Diffusion is a physical process which brings equilibrium in the concentration differences for a system. Using Fick's law [66]:

$$j = -D.\nabla u, \tag{A.9}$$

where j is the flux, ∇u is the concentration gradient and D is the tensor value. Since, diffusion involves neither creation nor destruction of mass, using the equation of continuity (conservation of mass) one obtains:

$$\partial_t u = -(\nabla . j), \tag{A.10}$$

where $\nabla . j$ denotes the divergence of the flux. From Eq. (A. 9) we get:

$$\partial_t u = -\left(\nabla .(-D.\nabla u)\right). \tag{A.11}$$

Replacing tensor value 'D' by scalar diffusivity 'd' results in:

$$\partial_t u = \left(\nabla .(d\nabla u)\right). \tag{A.12}$$

The scalar diffusion constant 'd' can be replaced by a scalar valued function '$g|\nabla u|$' which is the gradient of the gray levels in the image [38]. Hence we have:

$$\frac{\partial u}{\partial t} = \nabla .\left(g(|\nabla u|)\nabla u\right), \tag{A.13}$$

where $g(|\nabla u|)$ is given as,

$$g(|\nabla u|) = C(x,y,\sigma) \tag{A.14}$$

where C is the conductivity equation and σ gives the amount of blur. In non-linear diffusion equation [38], Perona and Malik used $C(x,y,t)$ as a function of the gradient magnitude (isotropic diffusion). The

value of C chosen in this way reduces the diffusion at the location of edges, encouraging smoothing within a region instead of smoothing across boundaries. Thus, C is chosen as:

$$C(x,y,t) = g(|\nabla L_\sigma(x,y,t)|), \tag{A.15}$$

where ∇L_σ (luminance function) is the gradient of a Gaussian smoothed original image L. Instead, one can choose $C(x,y,t)$ (similar to 'g' as proposed in Ref. [67]) such that:

$$g = \begin{cases} \frac{1}{2} & ,|\nabla L_\sigma|^2 = 0 \\ 1 - exp\left(-\frac{3.315}{(|\nabla L_\sigma|/k)^8}\right) & ,|\nabla L_\sigma|^2 > 0. \end{cases} \tag{A.16}$$

Thus, we have two different diffusion equation given by:

$$\partial_t L = div\left[\frac{1}{2} \cdot \nabla L\right] \tag{A.17}$$

corresponding to the points in the scale space, where $|\nabla L_\sigma|^2 = 0$, and

$$\partial_t L = div\left[\left\{1 - exp\left(-\frac{3.315}{(|\nabla L_\sigma|/k)^8}\right)\right\} \cdot \nabla L\right] \tag{A.18}$$

corresponding to the points in the scale space, where $|\nabla L_\sigma|^2 > 0$. Here, *div* denotes the divergence operator. We show the distribution of KAZE keypoints along the gradient images of non-linear scale space (KAZE) in Fig. A.11. The plot of KAZE keypoints in the gradient images of the evolution images is shown in Fig. A.12. In Fig. A.13 we have shown a failure case of KAZE detector. In the case of objects having smoothened boundary it fails to detect any keypoints at different levels of the scale space.

A.2. Comparison of KAZE keypoints along the boundary of the salient objects with other interest point detectors

In order to characterize the effectiveness of KAZE for boundaryness we compare the distribution of the keypoints along the ground-truth boundary annotations Fig. A.14. The experimentation

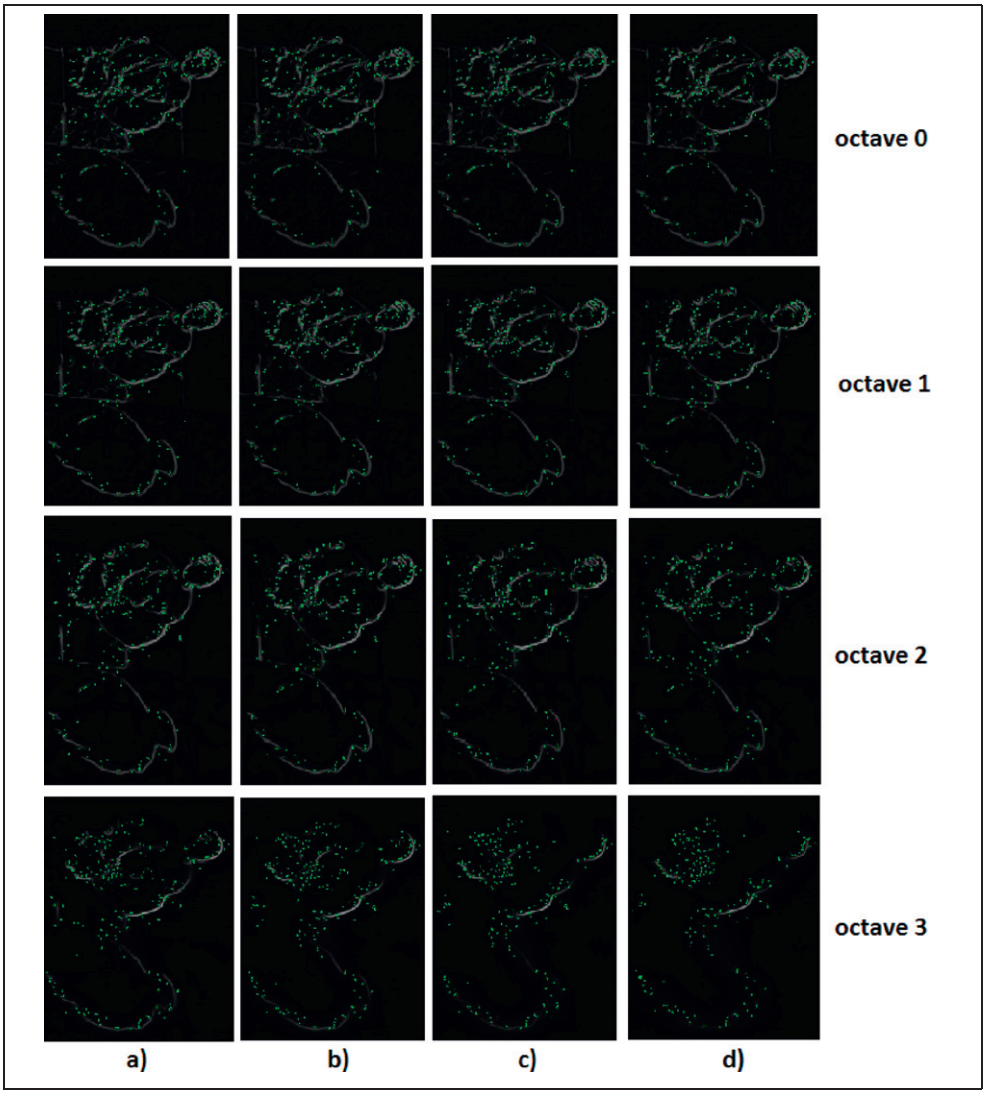

Fig. A.11. Total number of keypoints detected are 577 in the original image (Fig. 3 (a)). Majority of KAZE keypoints contained a gradient value >0 tracing the object contours. The number of octaves in the KAZE implementation is 4 and the number of evolutions in each octave is 4. Gradient images of non-linear scale space a)–d) evolution 1–4 of respective octaves.

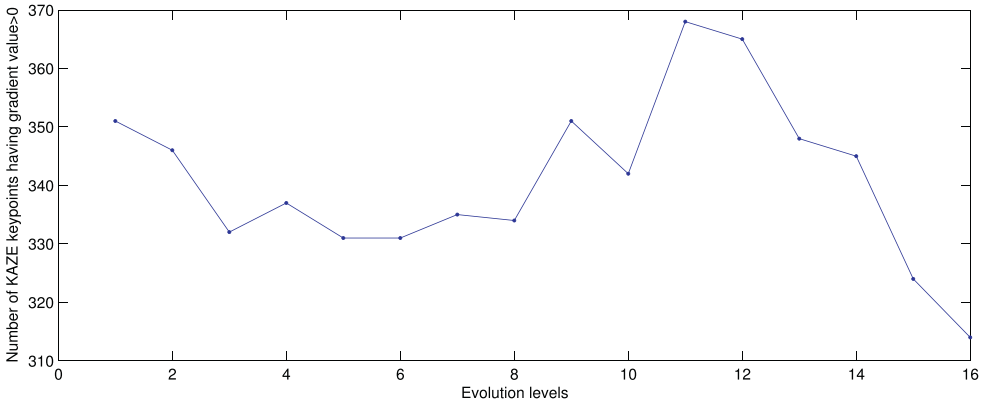

Fig. A.12. Number of KAZE keypoints with gradient value>0 in the gradient image of the evolution images.

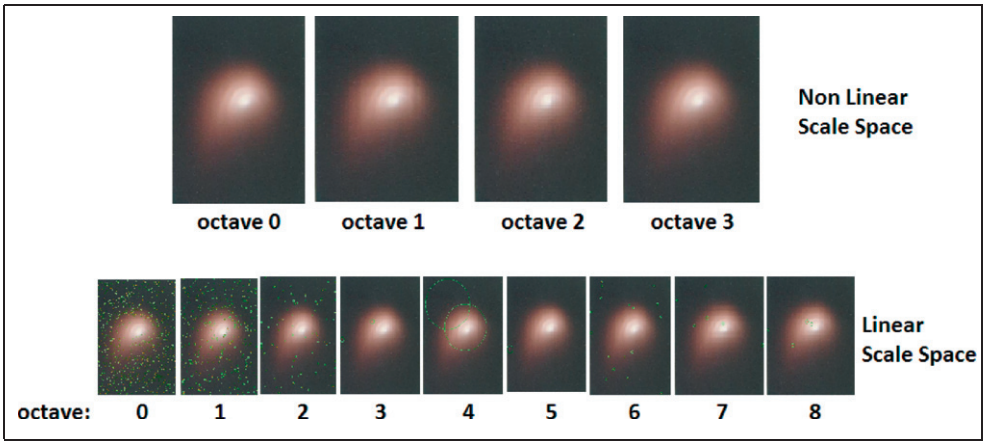

Fig. A.13. Failure case of KAZE keypoint detection: in the first row the keypoints detected at different octaves using KAZE (non-linear scale space). KAZE fails to detect any keypoints in the image as it does not find any well defined boundaries of the object. In the second row the keypoints detected at different octaves using SIFT (linear scale space) have been shown. Since SIFT is based on linear scale space it gives keypoints over the regions.

was performed on the MSRA-1000 dataset [29]. We calculate the ground-truth boundary annotations using Canny detector [68] on the ground-truth images. We calculate the density of keypoints for each salient object in the band of region along the boundary. The band of region was chosen as $\alpha = \pm 20$ pixels along the boundary. We observed that when α was increased the band of region started covering the spurious keypoints in the background regions. The density of keypoints is given as:

$$Avg(KP \quad density) = \sum_{i=1}^{I} \sum_{o=1}^{S_O} \left[\frac{\#KP \quad (boundary \pm \alpha)}{\#Total \quad KP} \right], \tag{A.19}$$

where S_O is the total number of salient objects in the image, KP denotes the keypoint detected in the image and I indicates the total number of images. The distribution of KAZE keypoints is compared with FAST corner points [69], Harris corner points [70], edge-foci interest points (EFI KP) [61], EDGELAP detector [60], edge-based region keypoints (EBR KP) [62] and Context Aware Keypoints Extraction (CAKE KP) [71] in Fig. A.14. We observe an increasing trend in the curve for KAZE features indicating a continuous increase in the average density of keypoints near the boundaries. For smaller values of N, the reduction in the average density of KAZE keypoints is less relevant. Curves corresponding edge-foci interest points and edge-based region keypoints are more flat. Harris keypoints curve saturates i.e. density remains almost the same even near edges (particularly object boundaries in this case). EDGELAP, CAKE and FAST keypoints show good number of keypoints near boundaries but lag behind KAZE keypoints.

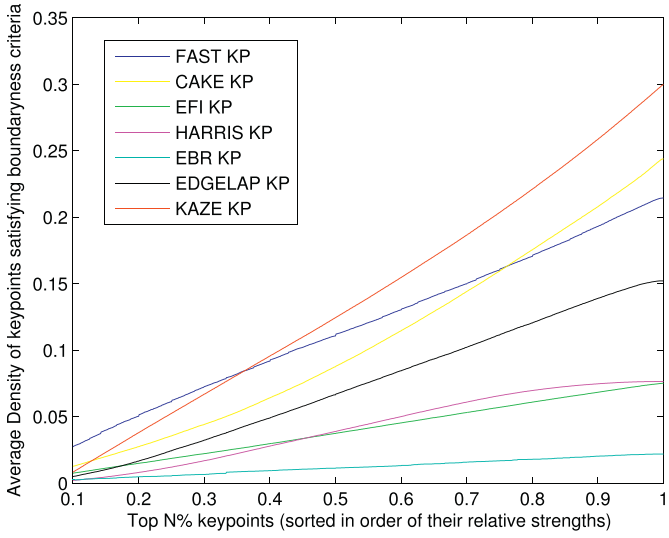

Fig. A.14. Average density of keypoints satisfying the boundaryness criteria vs top N% keypoints.

References

[1] M. Cheng, N.J. Mitra, X. Huang, P.H. Torr, S. Hu, Global contrast based salient region detection, IEEE Trans. Pattern Anal. Mach. Intell. 37 (3) (2015) 569–582.
[2] B.C. Russell, A. Torralba, K.P. Murphy, W.T. Freeman, LabelMe: a database and web-based tool for image annotation, Int. J. Comput. Vis. 77 (1-3) (2008) 157–173.
[3] B. Alexe, T. Deselaers, V. Ferrari, What is an object? Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE. 2010, pp. 73–80.
[4] C. Guo, L. Zhang, A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression, IEEE Trans. Image Process. 19 (1) (2010) 185–198.
[5] J. Ghosh, Y.J. Lee, K. Grauman, Discovering important people and objects for egocentric video summarization, 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE. 2012, pp. 1346–1353.
[6] H. Hadizadeh, I.V. Bajic, Saliency-aware video compression, IEEE Trans. Image Process. 23 (1) (2014) 19–33.
[7] G. Sharma, F. Jurie, C. Schmid, Discriminative spatial saliency for image classification, Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE. 2012, pp. 3506–3513.
[8] V. Gulshan, C. Rother, A. Criminisi, A. Blake, A. Zisserman, Geodesic star convexity for interactive image segmentation, Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE. 2010, pp. 3129–3136.
[9] J. Carreira, C. Sminchisescu, Constrained parametric min-cuts for automatic object segmentation, Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE. 2010, pp. 3241–3248.
[10] I. Endres, D. Hoiem, Category-independent object proposals with diverse ranking, IEEE Trans. Pattern Anal. Mach. Intell. 36 (2) (2014) 222–234.
[11] B. Hariharan, P. Arbeláez, R. Girshick, J. Malik, Simultaneous detection and segmentation, European Conference on Computer Vision, Springer. 2014, pp. 297–312.
[12] X. Cao, F. Wang, B. Zhang, H. Fu, C. Li, Unsupervised pixel-level video foreground object segmentation via shortest path algorithm, Neurocomputing 172 (2016) 235–243.
[13] M.-M. Cheng, N.J. Mitra, X. Huang, S.-M. Hu, Salientshape: group saliency in image collections, Vis. Comput. 30 (4) (2014) 443–453.
[14] G. Liu, Z. Lin, X. Tang, Y. Yu, Unsupervised object segmentation with a hybrid graph model (HGM), IEEE Trans. Pattern Anal. Mach. Intell. 32 (5) (2010) 910–924.
[15] J. Winn, N. Jojic, Locus: learning object classes with unsupervised segmentation, Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, 1, IEEE. 2005, pp. 756–763.
[16] J. Carreira, C. Sminchisescu, CPMC: automatic object segmentation using constrained parametric min-cuts, IEEE Trans. Pattern Anal. Mach. Intell. 34 (7) (2012) 1312–1328.
[17] J.R. Uijlings, K.E. van de Sande, T. Gevers, A.W. Smeulders, Selective search for object recognition, Int. J. Comput. Vis. 104 (2) (2013) 154–171.
[18] B. Alexe, T. Deselaers, V. Ferrari, Measuring the objectness of image windows, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2189–2202.
[19] P.F. Alcantarilla, A. Bartoli, A.J. Davison, KAZE features, Computer Vision—ECCV 2012 Springer. 2012, pp. 214–227.
[20] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[21] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Trans. Pattern Anal. Mach. Intell. 27 (10) (2005) 1615–1630.
[22] K.E. Van de Sande, T. Gevers, C.G. Snoek, A comparison of color features for visual concept classification, Proceedings of the 2008 International Conference on Content-Based Image and Video Retrieval, ACM. 2008, pp. 141–150.
[23] S. Srivastava, P. Mukherjee, B. Lall, Characterizing objects with SIKA features for multiclass classification, Appl. Soft Comput. (2015)
[24] Y. Li, X. Hou, C. Koch, J.M. Rehg, A.L. Yuille, The secrets of salient object segmentation, Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, IEEE. 2014, pp. 280–287.
[25] A.K. Mishra, Y. Aloimonos, L.-F. Cheong, A. Kassim, et al. Active visual segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 34 (4) (2012) 639–653.
[26] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, S. Li, Automatic salient object segmentation based on context and shape prior., BMVC, 6, 2011. pp. 9.
[27] D.R. Martin, C.C. Fowlkes, J. Malik, Learning to detect natural image boundaries using local brightness, color, and texture cues, IEEE Trans. Pattern Anal. Mach. Intell. 26 (5) (2004) 530–549.
[28] C. Yang, L. Zhang, H. Lu, X. Ruan, M.-H. Yang, Saliency detection via graph-based manifold ranking, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013. pp. 3166–3173.
[29] T. Liu, Z. Yuan, J. Sun, J. Wang, N. Zheng, X. Tang, H.-Y. Shum, Learning to detect a salient object, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2) (2011) 353–367.
[30] A. Borji, M.-M. Cheng, H. Jiang, J. Li, Salient object detection: a benchmark, IEEE Trans. Image Process. 24 (12) (2015) 5706–5722.
[31] W. Qi, M.-M. Cheng, A. Borji, H. Lu, L.-F. Bai, SaliencyRank: two-stage manifold ranking for salient object detection, Comput. Vis. Media 1 (4) (2015) 309–320.
[32] V. Yanulevskaya, J. Uijlings, N. Sebe, Learning to group objects, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014. pp. 3134–3141.
[33] M.-M. Cheng, Z. Zhang, W.-Y. Lin, P. Torr, BING: binarized normed gradients for objectness estimation at 300 fps, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014. pp. 3286–3293.
[34] S. Manen, M. Guillaumin, L. Gool, Prime object proposals with randomized Prim's algorithm, Proceedings of the IEEE International Conference on Computer Vision, 2013. pp. 2536–2543.
[35] C.L. Zitnick, P. Dollár, Edge boxes: locating object proposals from edges, Computer Vision—ECCV 2014 Springer. 2014, pp. 391–405.
[36] D. Hoiem, A.N. Stein, A.A. Efros, M. Hebert, Recovering occlusion boundaries from a single image, 2007 IEEE 11th International Conference on Computer Vision, IEEE. 2007, pp. 1–8.
[37] J. Aldana-Iuit, D. Mishkin, O. Chum, J. Matas, In the Saddle: Chasing Fast and Repeatable Features, Dec. 2016.
[38] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Trans. Pattern Anal. Mach. Intell. 12 (7) (1990) 629–639.
[39] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, S. Li, Salient object detection: a discriminative regional feature integration approach, Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE. 2013, pp. 2083–2090.
[40] Q. Yan, L. Xu, J. Shi, J. Jia, Hierarchical saliency detection, Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE. 2013, pp. 1155–1162.
[41] F. Perazzi, P. Krähenbühl, Y. Pritch, A. Hornung, Saliency filters: contrast based filtering for salient region detection, Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE. 2012, pp. 733–740.
[42] C. Rother, V. Kolmogorov, A. Blake, GrabCut: interactive foreground extraction using iterated graph cuts, ACM Transactions on Graphics (TOG), 23, ACM. 2004, pp. 309–314.
[43] R. Achanta, S. Hemami, F. Estrada, S. Susstrunk, Frequency-tuned salient region detection, Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, IEEE. 2009, pp. 1597–1604.
[44] X. Liao, H. Xu, Y. Zhou, K. Li, W. Tao, Q. Guo, L. Liu, Automatic image segmentation using salient key point extraction and star shape prior, Signal Process. 105 (2014) 122–136.
[45] N. Otsu, A threshold selection method from gray-level histograms, Automatica 11 (285-296) (1975) 23–27.
[46] V. Movahedi, J.H. Elder, Design and perceptual validation of performance measures for salient object segmentation, Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on, IEEE. 2010, pp. 49–56.
[47] W. Zou, K. Kpalma, Z. Liu, J. Ronsin, Segmentation driven low-rank matrix recovery for saliency detection, 24th British Machine Vision Conference (BMVC), 2013. pp. 1–13.
[48] M. Everingham, L. Van Gool, C.K. Williams, J. Winn, A. Zisserman, The Pascal visual object classes (VOC) challenge, Int. J. Comput. Vis. 88 (2) (2010) 303–338.
[49] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Susstrunk, SLIC superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2274–2282.
[50] G. Bradski, The OpenCV Library (2000), Dr. Dobbs J. Softw. Tools (2000).
[51] A. Vedaldi, B. Fulkerson, VLFeat: An Open and Portable Library of Computer Vision Algorithms, 2008, http://www.vlfeat.org/.
[52] E. Rahtu, J. Kannala, M. Salo, J. Heikkilä, Segmenting salient objects from images and videos, Computer Vision—ECCV 2010 Springer. 2010, pp. 366–379.
[53] Ç. Aytekin, E.C. Ozan, S. Kiranyaz, M. Gabbouj, Visual saliency by extended quantum cuts, Image Processing (ICIP), 2015 IEEE International Conference on, IEEE. 2015, pp. 1692–1696.
[54] K. Wang, L. Lin, J. Lu, C. Li, K. Shi, Plsa: pixelwise image saliency by aggregating complementary appearance contrast measures with edge-preserving coherence, IEEE Trans. Image Process. 24 (10) (2015) 3019–3033.
[55] R. Unnikrishnan, C. Pantofaru, M. Hebert, Toward objective evaluation of image segmentation algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 29 (6) (2007) 929–944.
[56] G. Li, Y. Yu, Visual saliency based on multiscale deep features, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015. pp. 5455–5463.
[57] X. Li, L. Zhao, L. Wei, M.-H. Yang, F. Wu, Y. Zhuang, H. Ling, J. Wang, DeepSaliency: multi-task deep neural network model for salient object detection, IEEE Trans. Image Process. 25 (8) (2016) 3919–3930.
[58] C. Qin, G. Zhang, Y. Zhou, W. Tao, Z. Cao, Integration of the saliency-based seed extraction and random walks for image segmentation, Neurocomputing 129 (2014) 378–391.
[59] N. Houhou, J.-P. Thiran, X. Bresson, Fast texture segmentation based on semi-local region descriptor and active contour, Numer. Math. Theory, Methods Appl. 2 (EPFL-ARTICLE-140431) (2009) 445–468.
[60] K. Mikolajczyk, A. Zisserman, C. Schmid, Shape recognition with edge-based features, British Machine Vision Conference (BMVC'03), 2, The British Machine Vision Association. 2003, pp. 779–788.
[61] C.L. Zitnick, K. Ramnath, Edge foci interest points, Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE. 2011, pp. 359–366.
[62] T. Tuytelaars, L. Van Gool, Matching widely separated views based on affine invariant regions, Int. J. Comput. Vis. 59 (1) (2004) 61–85.
[63] T. Lindeberg, Scale-Space Theory in Computer Vision, 256. Springer Science & Business Media. 2013.
[64] J.J. Koenderink, The structure of images, Biol. Cybern. 50 (5) (1984) 363–370.
[65] L. Florack, B.M. ter Haar Romeny, J.J. Koenderink, M.A. Viergever, Linear scale-space, J. Math. Imaging Vision 4 (4) (1994) 325–351.
[66] W. Wakeham, Fick's Law of Diffusion, 1980.
[67] J. Weickert, Efficient image segmentation using partial differential equations and morphology, Pattern Recognit. 34 (9) (2001) 1813–1824.

[68] J. Canny, A computational approach to edge detection, IEEE Trans. Pattern Anal. Mach. Intell. (6) (1986) 679–698.
[69] E. Rosten, T. Drummond, Machine learning for high-speed corner detection, Computer Vision—ECCV 2006 Springer. 2006, pp. 430–443.
[70] C. Harris, M. Stephens, A combined corner and edge detector, Alvey Vision Conference, 15, Citeseer. 1988, pp. 50.
[71] P. Martins, P.D. Carvalho, C. Gatta, Context aware keypoint extraction for robust image representation., BMVC, 2012. pp. 1–12.