ORIGINAL ARTICLE

Real-time saliency detection for greyscale and colour images

Jian-Feng Shi[1] · **Steve Ulrich**[1] · **Stéphane Ruel**[2]

Abstract

Unsupervised salient image generation without the aid of prior assumptions has many applications in computer vision. We present three unique real-time saliency generation algorithms that provide state-of-the-art performance for greyscale and colour images. Our fastest method run under 50 ms per frame on average. Our algorithm introduces a novel weighted histogram of orientation feature to supplement image intensity for monochromatic image manifold ranking. We also provide a method of dimensional reduction for the non-normalized optimal affinity matrix (OAM) using principal components analysis; this novel technique allows faster computation and stabilization of the OAM inversion process. We compare our methods with 18 traditional and recent techniques using three standard and custom datasets including ECSSD, DUT-OMRON and MSRA10K totalling 32,536 images for colour and greyscale variations. The results show our method to be more than $10\times$ faster than the RC and GMR models and having similar or better precision performances.

Keywords Image saliency · Image segmentation · Image features

1 Introduction

Salient image detection is a remarkable skill which consists in focusing attention solely to the most relevant region or objects from an input image. Many disciplines such as cognitive psychology, neuroscience and computer vision studies this evolved ability [13]. Image saliency is one of the main research for the vision community in the last several decades, and it has produced numerous models to either mimic the biological functionality in human fixation or designing for application-driven scenarios. There are many applications in using image saliency, such as scan path prediction [7], image resizing [8], action recognition [9], image compression [21], interest region proposal [22], visual tracking [71],

✉ Jian-Feng Shi
jianfeng.shi@carleton.ca

Steve Ulrich
steve.ulrich@carleton.ca

Stéphane Ruel
sruel@neptec.com

[1] Mechanical and Aerospace Engineering Department, Carleton University, Ottawa, Canada

[2] MDA Neptec Design Group Ltd., Ottawa, Canada

and video [70] and image segmentation [38,87]. Specifically, we aim to use image saliency to generate spacecraft foreground segmentation for pose estimation in the guidance, navigation and controls (GNC) function during spacecraft rendezvous and docking [60]. Most on-orbit imagery consists of a single spacecraft object against a black background. In these scenarios, a simple Otsu thresholding [54] is sufficient in extracting the foreground spacecraft. However, there are scenarios where extreme illuminations from sunlight reflection, hard shadowing, camera lens flare, and Earth background or occulting space objects that are challenging to resolve. Due to limited data bandwidth and highly constrained computational resources, space images are often monochromatic and low resolution [28] with high risks of loss of signal from limited communication coverage. Since the saliency detection function is only a small portion of an overall GNC system, fast algorithms are highly desirable. Furthermore, real on-orbit spacecraft images are costly to produce, which restricts models that require numerous training data. To this end, this investigation focuses on developing salient image detection that is unsupervised, can be used for both colour and greyscale images, independent of temporal information and can be computed in real-time.

We add to the body of knowledge in image saliency by making the following contributions:

1. We provide a principal component analysis (PCA) approximation of the optimal affinity matrix (OAM) inversion which removes limitations on large matrix inversion, reduces computation time, and stabilizes the OAM matrix inversion process.
2. We develop a novel weighted gradient orientation histogram distance to increase greyscale image distinctiveness. Our orientation histogram feature is faster and more precise than conventional image features.
3. We provide three novel image saliency methods for real-time operation. Our speed optimized version is $12\times$ faster than the benchmark method with comparable precision.

The organization of this investigation is as follows, Sect. 2 provides a detailed overview of related saliency models. Section 3 provides our proposed approach. Section 4 provides methods in speed and foreground/background distinction improvements to the traditional graph manifold ranking model. Section 5 provides a novel technique in handling monochromatic images based on gradient orientation and texture. Section 6 provides three enhancements to optimize mean precision, maximum precision, and speed. Section 7 provides details on experimental datasets and metrics used by this investigation. Section 8 provides precision and timing results. Section 9 concludes this investigation.

2 Related work

Salient feature detection can be categorized as top-down and bottom-up approaches. The top-down approach often uses scene understanding [82], having memories of the past feature structures, either globally [12,16] or locally [11,69,75, 83], while performing supervised learning with class labels. The top-down approach resembles functions in the higher regions of our brain forming more complicated and expensive classification to comprehend the global coupling [20]. Often, the top-down approach is combined with bottom-up techniques for generating the saliency map [39,40,44,45,77]. The bottom-up approach is image driven and pre-attentive. It originates from uniqueness, rarity, irregularity, and surprise. These qualities are closely related to stimulations to our visual system. Indeed, the non-salient image signals are pre-filtered before the most useful information is used for higher processing [40].

Due to recent breakthroughs in convolutional neural networks (CNN), many CNN saliency detection models have been introduced [16,29,30,38,46,47,50,61,86]. However, because our application has limited training input, it rules out network-based approach that requires large amounts of labelled training data. Driven by the unsupervised real-time requirement, we choose the bottom-up approach as the starting point of our saliency generation method. For a more comprehensive list of saliency models, a survey is published by Borji [13].[1]

Seminal theories on image saliency are traced to Koch [34], where founding ideas of *proximity* and *similarity* led to the early Itti [31] model. This model proposed an attention system based on centre-surround, image features, and a so-called the "Winner-Takes-All" neural network. This methodology can be categorized as one of the biologically inspired ways of predicting eye fixation. Other eye fixation predictions include spectral residual (SR) [27], phase spectrum of quaternion Fourier transform (QFT) [26], and spatiotemporal attention [85]. In SR [27], the image is separated into salient *innovation* and non-salient *prior knowledge*, and the *innovation* can be computed using the residuals after removing the log spectrum. QFT extended this idea by using a quaternion made from colour, intensity and motion features between frames [26]. Zhai and Shah proposed the temporal attention model [85] where the scale invariant feature transform (SIFT) [48] operator is used to find frame-to-frame correspondence and motion contrast in identifying attention to moving objects.

On the other hand, methods based on salient object detection of global or local regions are computed through systematic raster scans [2,56,88]. Saliency maps may be generated by global frequency filtering [3,90], geodesic [73], or local centre-surround [1]. Other methods compares local resemblance [52,58], directly using colour contrast [18,20] or colour similarity [1,80]. Non-colour-based models were also explored, mainly through the use of image texture [69,78], image features [15] for still images, and optical flow [70] for motion sequences. In an example of the greyscale image saliency detection, Jung et al. [33] proposed the use of orientation from image gradients to form a histogram that is later mean-shift filtered. The saliency map is computed from the gradient magnitude and the cosine distance between the local directional residual from gradient orientation and the nearest dominating angle. However, using orientation features alone will only produce blurry attention regions with low precision [33].

Many older methods in saliency detection are pixel based [1,49,56]. Although results can precisely define foreground boundaries, our experience showed that pixel-based approaches are too computationally intensive for real-time implementation. Recent developments in image superpixels [4,43,65] has been adopted by several modern work in saliency detection [37,41,42,63,64,72,80,81,83,84]. In particular, the simple linear iterative clustering (SLIC) superpixel is the defacto choice for graph manifold-based methods [37,42,55,74,80]. The benefits of using superpixels over image pixels are the ability to operate pixel clusters with

---

similar properties and therefore save computation effort. However, there is a trade-off between the time saved from using the superpixel and the time needed to compute them. In Sect. 4.1 we evaluate three superpixel methods for our proposed saliency technique.

With so many methods in saliency generation, some general trends may be observed: the use of centre-surround either by foreground localization or background seeding from border nodes; the segmentation of local regions and similarity comparisons of these regions; and the combination of multiple response maps to enhance the most salient features. For example, a recent publication by Li et al. [41] contains many of the typical components of a modern bottom-up saliency model, such as a Gaussian mixture model decomposition with hash-based clustering combined with SLIC-based centre prior uniqueness to produce the foreground saliency map. Li et al. [41] then combined this foreground map, border pixel seeded background map, and a smoothness difference, in a manifold ranking optimization scheme for the final saliency generation.

While the bottom-up saliency detection methods are numerous, our interest is on those that are real-time capable. Furthermore, the attention models such as those by Hou and Zhang [27], Itti et al. [31], and Seo and Milanfar [58] compute the locations for where to direct visual attention but do not provide precise regional boundaries of the foreground object. Our approach is to invoke methods that have sufficient precision in generating saliency maps of the desired foreground and is fast enough to be implemented in real-time. Three of the latest methods that stands out in this regard is regional contrast (RC) [20], minimum barrier distance (MBD) [88] and graph manifold ranking (GMR) [80,89]. RC was proposed by Cheng et al. [20] to extract saliency from local regions as the weighted sum of colour contrast. The RC method first uses a graph-based [24] image segmentation to generate the working regions then creates a colour histogram and applies a histogram smoothing technique to reduce noise. Cheng [20] also uses RC to supplement *SaliencyCut*, a *GrabCut* [57] based segmentation in producing high-quality foreground masks. Our testing shows that RC is highly efficient to compute, with the most expensive module being the initial region segmentation. Zhang et al. [88] proposed a fast raster-scanning algorithm to approximate the MBD transform [62]. A central idea in the MBD is the *Image Boundary Connectivity Cue* which assumes that background regions are connected to the image border, a variation of centre-surround [31]. Although more precise than its geodesic counterpart, MBD by itself lacks the desired accuracy when compared to RC. An extended version, MB+, was proposed by adding *Image Boundary Contrast* (IBC) map using border pixels as colour contrast seeds in the whitened colour space. Unfortunately, the proposed IBC Map is computed by using pixels, and our experience shows that this can be a computationally-expensive addition to the MBD. Finally, Yang et al. [80] proposed GMR saliency using document ranking. With this approach, the GMR solves an optimization problem by defining a graph-based ranking cost function. The minimum solution is an OAM based on weights of colour distances in *CIE Lab* colour space. GMR is elegant and fast; however, it requires a matrix inverse that can be time-consuming and unstable when there is a large number of superpixels. To this end, one of our contributions is to enhance the inverse process by using PCA approximations.

3 Method overview

We begin by comparing the timing of traditional and state-of-the-art methods available in the literature. Models are discarded if they cannot be practically implemented in real-time even if they produce high precision saliency maps. Out of the remaining saliency model candidates, we found GMR, MBD and RC to have the most potential for highly precise real-time operations. Coarse-resolution input images can reduce operation timing; however, coarse resolution images can also significantly degrade the precision of the GMR, MBD and RC saliency models. The novelty in our methods consists in using key aspects of state-of-the-art solutions to achieve much faster operation time while maintaining the same precision to high-resolution input images. We evolve the GMR method to meet our timing requirement of 50 ms per frame while still maintaining its precision. Our performance space is bounded by the mean F-measure, the maximum area under the curve (AUC), and the minimum computation time. Instead of developing one approach that dilutes the conflicting requirements, we developed three algorithms that maximize each performance direction. Specifically, we herein call these models precision (*prc*), fast (*fst*), and fast maximum-precision (*fmx*). We achieved our timing requirement with the *fst* and *fmx* models. Our *fmx* model has top precision versus recall curves while our *fst* model is the fastest design with competitive precision. We also achieve the highest mean F-measure overall by using our *prc* model. While the *prc* model did not meet our timing requirement, it is still $3\times$ faster than RC and GMR methods. All developed methods use superpixel ranking, where the *prc* model includes a novel weighted orientation histogram feature that is the fastest compared to other monochrome image descriptors. This improves the saliency generation of the monochromatic image. Both *fmx* and *fst* models have inherited elements from the MBD and RC techniques. Details of each model are provided in Sect. 6.

4 Speed and distinction improvements

The highest computation cost inherent to GMR [80] comes from the image size and the number of graph nodes. Our

approach aims to reduce the number of nodes, add stability to the GMR OAM computation, while increasing the distinctiveness between foreground and background ranking scores, especially when the colour information is not available.
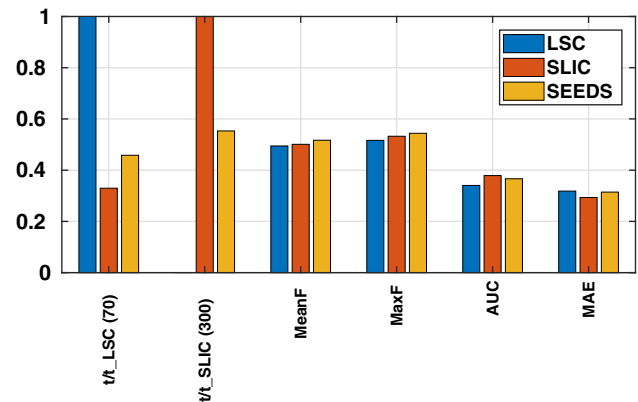
4.1 Superpixels

Superpixels are used to identify locations and boundaries of graph nodes. They reduce the pixel-wise operations by several hundred times and are essential in the design of a real-time algorithm for saliency detection. Three popular superpixel techniques are linear spectral clustering (LSC) [43], simple linear iterative clustering (SLIC) [4], and superpixels extracted via energy-driven sampling (SEEDS) [65]. An example superpixelation of an International Space Station (ISS) infrared image is shown in Fig. 1 for all three methods. LSC uses normalized cuts with colour similarity and space proximity [43]. Unfortunately, LSC contains undesirable local discontinuities such as those shown in Fig. 1a. The LSC computation time was 7.941 ± 0.869 ms for the ISS test image with roughly 70 superpixels.[2] SLIC is a graph-based algorithm that treats each pixel as a graph node [4]. Similar to LSC, the number of pixels can be consistently specified and the pixels are similar in size. SLIC takes longer time to compute than both LSC and SEEDS but it generates the most stable and consistent superpixel map. For the ISS test image, the SLIC method timing was 11.610 ± 0.071 ms. To our knowledge, all GMR-related saliency models uses SLIC to superpixelate the input image. Unlike SLIC which grows superpixels by clustering around centres, SEEDS starts from a grid partitioning and iteratively refines the superpixel boundaries [65]. Our tests show that SEEDS is the fastest superpixel method at 6.078 ± 0.028 ms for the ISS image. However, SEEDS superpixels are not consistent in size and the shape and quantity cannot be directly controlled. Figure 2 compares timing and precision performance for these various superpixel methods. LSC was the slowest and least precise overall. SLIC and SEEDS have roughly the same precision performance. SLIC is faster when the number of superpixels is low, but performs slower when the number of superpixels is higher. SLIC is used in the *prc* and the *fst* model, while SEEDS is used in the *fmx* model.

4.2 Graph manifold ranking

Graph manifold ranking is a rating algorithm that spreads the seeding query scores to neighbouring nodes via the weighted network [91]. GMR has been widely adopted for document [67] and image retrievals [76]. The standard GMR framework is provided as follows: given an image with N number of superpixels, each superpixel is considered as

(a) LSC   **(b)** SLIC   **(c)** SEEDS
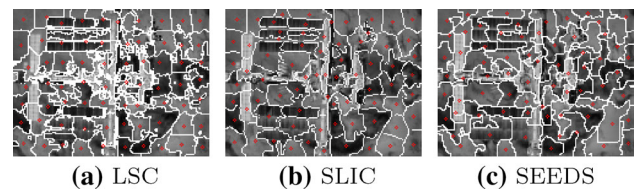
Fig. 1 Superpixel breakdown of an 160×120 infrared ISS image. Subfigures **a** to **c** shows LSC, SLIC, and SEEDS methods, respectively. Red circles indicate locations of the superpixel spatial centroid

Fig. 2 Superpixel performance comparisons. In the first column, images using 70 superpixels were computed using LSC, SLIC and SEEDS; the timing was scaled by LSC which was the slowest algorithm of the three tested methods. In the second column, images using 300 superpixels were computed using SLIC and SEEDS, and the timing was scaled by the SLIC results. The LSC method was eliminated for the 300 superpixels test since it was a significantly slower algorithm than SLIC and SEEDS as indicated by the 70 superpixel result. Columns 3 to 6 provides Mean F-measure, Max F-measure, AUC, and MAE for the three superpixel methods

regions with some given feature vector $\mathbf{h}^{(i)} \in \mathbb{R}^m$. The set of feature vectors for all regions is $\bar{\mathbf{V}} = \{\mathbf{h}^{(1)}, \ldots, \mathbf{h}^{(N)}\}^{m \times N}$. We select a subset of $\bar{\mathbf{V}}$ as seeds and rank the rest of the set base on their relevance to the seeding queries. Let us define $\mathbf{f} : \bar{\mathbf{V}} \rightarrow \mathbb{R}^N$ denote the ranking function for each region in N, such that $\mathbf{f} = \left[f^{(1)}, \ldots, f^{(N)} \right]^{\mathrm{T}}$. Let define $\mathbf{y} = \left[y^{(1)}, \ldots, y^{(N)} \right]^{\mathrm{T}}$ as an indication vector where 1 is to perform a query on $\mathbf{h}^{(i)}$ and 0 otherwise. Then, define a graph $G = \left(\bar{\mathbf{V}}, \bar{\mathbf{E}} \right)$ over the image regions, where the nodes $\bar{\mathbf{V}}$ are region features and $\bar{\mathbf{E}} = \mathcal{E}(\mathbf{x})$ are the edges which are weighted by the *affinity matrix* $\mathbf{W} = [w^{(ij)}]$ where $\mathbf{W} \in \mathbb{R}^{N \times N}$. The *degree matrix* $\mathbf{D} = \mathrm{diag}\{d^{(11)}, \ldots, d^{(NN)}\}$ is computed by $d^{(ii)} = \sum_{j \in N} w^{(ij)}$. The optimal ranking is computed by using the optimization cost function given by Yang et al. [80]. For any given superpixel, its adjacent neighbours and the closed-loop border boundaries are used

---

[2] Discounting the discontinuous subpixels.

for the feature distance differencing. Setting the derivative of the cost function to zero, the resulting ranking function is

$$\mathbf{f}^* = \bar{\mathbf{A}}\mathbf{y} = \left(\mathbf{1} - \alpha\mathbf{D}^{-\frac{1}{2}}\mathbf{W}\mathbf{D}^{-\frac{1}{2}}\right)^{-1}\mathbf{y}, \tag{1}$$

where $\mathbf{1}$ is the identity matrix, $\bar{\mathbf{A}}$ is the normalized OAM, and α is optimally tuned to 0.99. Yang et al. [80] proposed to use the unnormalized OAM $(\mathbf{D} - \alpha\mathbf{W})^{-1}$ for better performance. For background seeding queries, we take the normalized complementary vector for the foreground ranking score

$$\bar{\mathbf{f}}^* = \mathbf{1} - \frac{\mathbf{f}^*}{\underset{f^*}{\operatorname{argmax}} f^{*(i)}}, \tag{2}$$

where $f^{*(i)}$ is the ranking score at the ith node. This ranking score is used to label individual superpixels in the final saliency map. The weighting is computed as

$$w^{(ij)} = \exp\left(\frac{\check{D} - D^{(ij)}}{\delta\left(\hat{D} - \check{D}\right)}\right), \tag{3}$$

where $i, j \in \bar{\mathbf{V}}$, δ is optimally tuned to 0.1. We use the L1-norm instead of the L2-norm for $D^{(ij)}$ [52],

$$D^{(ij)} = \sum_{k=1}^{m} |h_k^{(i)} - h_k^{(j)}|, \tag{4}$$

where the maximum and minimum feature vector difference is $\hat{D} = \operatorname{argmax}_{i,j\in\bar{\mathbf{v}}} D^{(ij)}$, and $\check{D} = \operatorname{argmin}_{i,j\in\bar{\mathbf{v}}} D^{(ij)}$. While the change to L1-norm did not result in increased precision, it is slightly faster than the L2-norm implementation. We also compared the L1 and L2-norm in whitened feature space. Results show the difference in mean ranking scores for the foreground and background computed using the ground truth masks increased significantly; however, the variance of these ranking scores are still not concentrated enough to translate into higher overall precision performance.

The background seeding taken from each of the outer border superpixels is ranked by Eqs. (1) and (2) to form the foreground saliency maps. The foreground maps based on the four borders are piece-wise multiplied $\bar{\mathbf{f}}_f^* = \bar{\mathbf{f}}_t^* \circ \bar{\mathbf{f}}_b^* \circ \bar{\mathbf{f}}_l^* \circ \bar{\mathbf{f}}_r^*$, where the subscripts f, t, b, l, r denotes foreground, top, bottom, left, and right respectively. A mean value binary threshold is applied to $\bar{\mathbf{f}}_f^*$. The resulting normalized ranking scores are fed into Eq. (1) again for the final foreground saliency map. The ranking scores are transformed back to the

saliency image space by $S_f^{(i)} = \bar{f}^{*(i)}$ for $i = \{1\ldots N\}$ where i is the superpixel label index.

4.3 Optimal affinity matrix from PCA inversion

The dimension of the *learnt optimal affinity matrix* $\mathbf{A}$ is $N \times N$. For a low number of superpixels (i.e.$N \leq 200$), the inversion can be computed quickly. However, as the number of superpixels increases, $\mathbf{A}$ takes much longer to compute and has more chances of becoming unstable. In the general application of GMR for document and image retrieval, attempts have been made to reduce the dimension of $\mathbf{A}$ by efficient computations via sparse matrices [25] and by designing the weighting matrix as a separable symmetrical matrix such that $\mathbf{W} = \mathbf{Z}^T\mathbf{Z}$ [76]. In our GMR framework, however, the weighting matrix cannot be decomposed into a single $\mathbf{Z}$ matrix. We derive a new form for representing the non-normalized OAM using PCA approximation for matrix reduction.

First, we decompose the weighting matrix by *Singular Value Decomposition*,

$$\mathbf{W} = \mathbf{U}\Sigma\mathbf{V}^T = \mathbf{PQ}, \tag{5}$$

where $\mathbf{U} = \mathbf{P}$ and contains the columns of eigenvectors of $\mathbf{W}\mathbf{W}^T$. The diagonal of Σ contains the square root of the eigenvalues $\sqrt{\lambda^{(i)}}$, $i = \{1\ldots N\}$ of $\mathbf{W}\mathbf{W}^T$ descending from large to small. Let us define $\mathbf{E} = \mathbf{D}^{-\frac{1}{2}}\mathbf{P}$ and $\mathbf{F} = \mathbf{Q}\mathbf{D}^{-\frac{1}{2}}$, then the normalized OAM can be written as $\bar{\mathbf{A}} = (\mathbf{1} - \alpha\mathbf{EF})^{-1}$.

Lemma 1 *Given the matrix* $\bar{\mathbf{A}} \in \mathbb{R}^{N\times N}$, $\mathbf{E} \in \mathbb{R}^{N\times d}$, $\mathbf{F} \in \mathbb{R}^{d\times N}$, *and* $\bar{\mathbf{A}} = (\mathbf{1} - \alpha\mathbf{EF})^{-1}$ *such that the inverse exists, then an alternative form for* $\bar{\mathbf{A}}$ *can be written as* $\bar{\mathbf{A}} = \mathbf{1} - \mathbf{E}(\mathbf{FE} - \frac{1}{\alpha}\mathbf{1}_d)^{-1}\mathbf{F}$, *where* $\mathbf{1}_d \in \mathbb{R}^{d\times d}$ *is the identity matrix and* $d \leq N$.

The proof of Lemma 1 is provided in Appendix A. We take an additional step to develop the equivalent inverse equation for the non-normalized OAM where

$$\begin{aligned} \mathbf{A} &= \mathbf{D}^{-\frac{1}{2}}\bar{\mathbf{A}}\mathbf{D}^{-\frac{1}{2}} \\ &= \left(\mathbf{D}^{\frac{1}{2}}\mathbf{D}^{\frac{1}{2}} - \alpha\mathbf{D}^{\frac{1}{2}}\mathbf{D}^{-\frac{1}{2}}\mathbf{W}\mathbf{D}^{-\frac{1}{2}}\mathbf{D}^{\frac{1}{2}}\right)^{-1} \\ &= (\mathbf{D} - \alpha\mathbf{PQ})^{-1}. \end{aligned} \tag{6}$$

The inverse form can be developed as the following,

Corollary 1 *Given the matrix* $\mathbf{A} \in \mathbb{R}^{N\times N}$, $\mathbf{P} \in \mathbb{R}^{N\times d}$, $\mathbf{Q} \in \mathbb{R}^{d\times N}$, *and* $\mathbf{A} = (\mathbf{D} - \alpha\mathbf{PQ})^{-1}$ *such that the inverse exists, then an alternative form for* $\mathbf{A}$ *can be written as* $\mathbf{A} = \mathbf{D}^{-1}(\mathbf{1} - \mathbf{P}(\mathbf{QD}^{-1}\mathbf{P} - \frac{1}{\alpha}\mathbf{1}_d)^{-1}\mathbf{QD}^{-1})$, *where* $\mathbf{1}_d \in \mathbb{R}^{d\times d}$ *is the identity matrix and* $d \leq N$.

The proof of Corollary 1 is also provided in Appendix A. Using the results of Corollary 1, the OAM may be approximated by taking the most relevant eigen components effectively reducing the inverse dimension from N to a lower value d. We denote the approximate OAM, degree matrix and decomposition of the weighting matrix using the tilde notation as $\tilde{\mathbf{A}}$, $\tilde{\mathbf{D}}$, $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{Q}}$. We can adaptively compute the number of most relevant eigen components to maintain by setting some threshold to the sum of the full eigen components by the following,

$$\frac{\sum_{i=1}^{d} |\sqrt{\lambda^{(i)}}|}{\sum_{i=1}^{N} |\sqrt{\lambda^{(i)}}|} \geq Thr, \tag{7}$$

where 1 to d is the reduced number of components in maximum descending order. Values for $\sqrt{\lambda^{(i)}}$ may be extracted from Σ of Eq. (5). PCA ensures an optimally minimized OAM approximation; we may manually set the Thr tolerance based on timing and precision requirements. Figure 3 provides results for the various threshold levels varied from 0.6 to 1. We observe that when Thr is set to 1, the reduced method is slower due to the additional steps in the modified form. Time-savings increase with reduced thresholds and increasing number of superpixels. By contrast, the error in the OAM increases with threshold reduction. Interestingly, a minimum point is observed at 168 pixels resulting from the balance between the omitted components and the retained ones. To take a broader view of the OAM reduction effectiveness, we define the distinctiveness as the difference between foreground and background ranking scores using the ground truth mask's intersection with the final $\mathbf{f}^*$. The distinctiveness from both Fig. 3c, d shows a steady reduction in GMR performance as OAM reduction increases for colour and greyscale images, respectively.

4.4 High-frequency response attention driven seeding

The final GMR ranking computed by Eq. (1) is a function of the OAM $\mathbf{A}$ and the estimated nodes of the foreground $\mathbf{y}$ as queries. We replaced the estimated foreground query with the ground truth foreground nodes and found significant improvements in the overall precision. This indicates GMR precision is sensitive to the quality of the seeds used in the ranking optimization. We developed a high-frequency response attention driven seeding scheme to improve background estimation from the surrounding borders. Using the border nodes for background seeding has been used by many models [41,42,68,80,88]. In this work, we remove the border seeds from the background query if there is evidence these are foreground nodes and replace them with the inner-layer neighbour background estimate. The target object in

the spacecraft application typically contains strong artificial lines and edges compared to the softly blended background scene. Therefore, we extract these high-frequency lines by applying a 3×3 Laplacian filter, $\mathcal{L}(\mathbf{x})$, on the greyscale input image $\mathbf{I}$ and blurred by a square box filter, $\mathcal{B}(\mathbf{x})$, of the size K_B, to remove noise. We focus the foreground region from eye fixation attention cues computed by SR. Our tests show SR to be faster and produce better results than QFT. The attention cue is combined with the high-frequency region by using a Gaussian distribution map, $\mathcal{G}(\mathbf{x})$, centred at the SR moment centroid, $\mathcal{MC}(\mathcal{SR}(\mathbf{I}))$. We then take the intersection between the estimated foreground responses with the border seeds to replace any lost nodes with the estimated background node neighbours. The seeding process applied to the ISS monochrome image is depicted in Fig. 4.
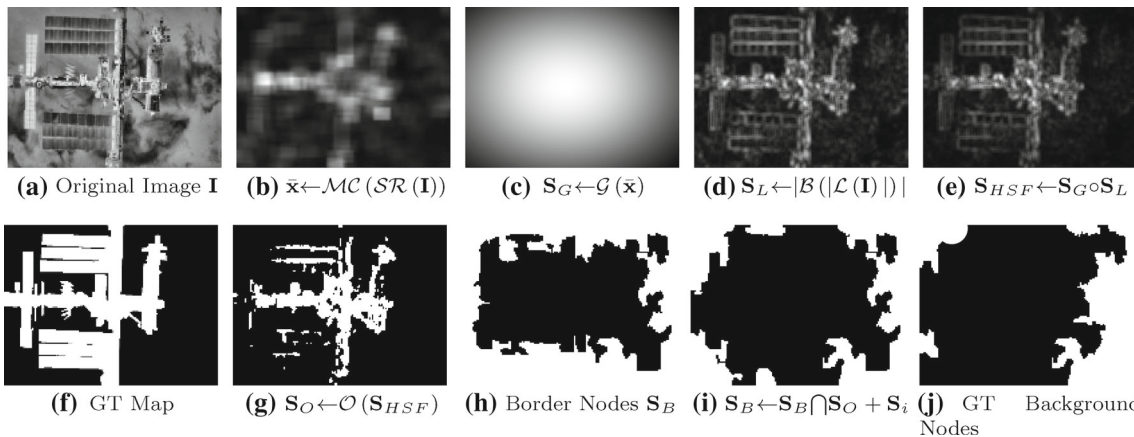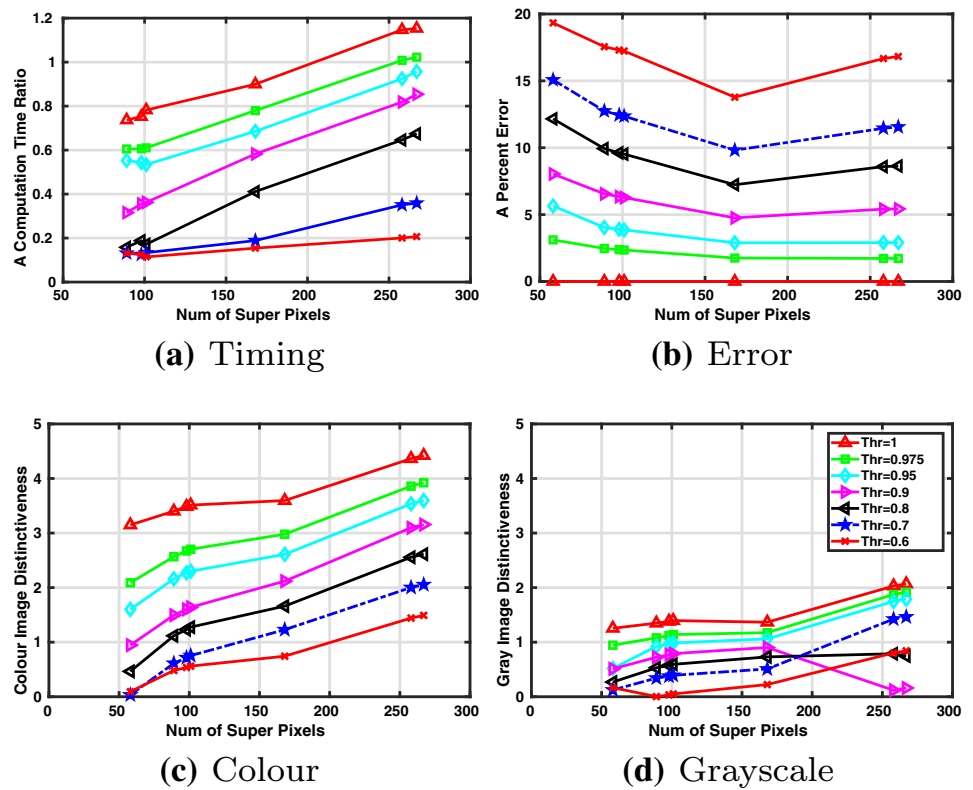
5 Monochromatic features

Colour features are essential in distinguishing objects from its background; we use colour features in the *CIE Lab* colour space similar to other methods [20,80]. Our testing confirms higher performance using *Lab* colour space than *RGB*. For efficiency, we compute the colour space conversion from *RGB* to *Lab*, $\mathbf{I}_{Lab} \leftarrow \mathcal{LAB}(\mathbf{I})$, only once for SLIC superpixelation and in our modified GMR calculation. Unfortunately, in spacecraft GNC applications, data rate and memory restrictions require the use of low-resolution monochromatic images, and only greyscale images are available from infrared imagers. To this end, we herein propose an alternative real-time solution using image orientation and texture instead of colour.

5.1 Weighted histogram of orientation

Local image orientation also provides information between various objects in the scene. We are inspired by the approach of Jung et al. [33] to use an orientation histogram from local image gradients. Instead of finding the orientation differences from the dominant direction, we use the orientation histogram directly as the image feature. We built our orientation histogram weighted by the gradient magnitude; we call this approach the weighted histogram of orientation (WHO) feature since the histogram encodes both the local direction and strength of the image gradient. Let an 8-bit greyscale local patch be denoted by $\mathbf{I}$ and the patch centre is located at the superpixel spatial centroid. The first-order image intensity derivative for an individual pixel in the local patch is $I_x^{(ij)} = \partial I^{(ij)}/\partial x$ and $I_y^{(ij)} = \partial I^{(ij)}/\partial y$, where i and j denotes the row and column index of the local patch. We compute the gradient image by convolving $\mathbf{I}$ with the 3×3 Sobel kernel. From the patch gradients, the orientation and

Fig. 3 OAM reduction results. Each data point is the average of all 32,536 image calculations. Threshold level corresponds to Eq. (7). Figure **a** shows the timing ratio between reduced OAM over non-reduced for a single OAM calculation. Figure **b** shows the average-relative-percent-error in the reduced OAM. Subfigure **c** and **d** shows the distinctive difference between the average foreground and background ranking score for all colour and greyscale images, respectively

(a) Timing

(b) Error

(c) Colour

(d) Grayscale

Thr=1
Thr=0.975
Thr=0.95
Thr=0.9
Thr=0.8
Thr=0.7
Thr=0.6

(a) Original Image $\mathbf{I}$ **(b)** $\bar{\mathbf{x}} \leftarrow \mathcal{MC}(\mathcal{SR}(\mathbf{I}))$ **(c)** $\mathbf{S}_G \leftarrow \mathcal{G}(\bar{\mathbf{x}})$ **(d)** $\mathbf{S}_L \leftarrow |\mathcal{B}(|\mathcal{L}(\mathbf{I})|)|$ **(e)** $\mathbf{S}_{HSF} \leftarrow \mathbf{S}_G \circ \mathbf{S}_L$

(f) GT Map   **(g)** $\mathbf{S}_O \leftarrow \mathcal{O}(\mathbf{S}_{HSF})$ **(h)** Border Nodes $\mathbf{S}_B$ **(i)** $\mathbf{S}_B \leftarrow \mathbf{S}_B \bigcap \mathbf{S}_O + \mathbf{S}_i$ **(j)** GT Background Nodes

Fig. 4 Algorithm sequence for border seeds combined with the foreground estimate. The input image $\mathbf{I}$ in **a** is passed into SR in **b**, where the SR moment centroid is used to centre the Gaussian centredness map $\mathbf{S}_G$ in **c**. A high-frequency response $\mathbf{S}_L$ is computed by the 3×3 Laplacian and the $K_B \times K_B$ box filter in **d**. The final foreground estimate based on image frequency is $\mathbf{S}_{HSF}$ in **e** which can be compared with the ground truth masks in **f**. $\mathbf{S}_{HSF}$ is threshold using the OTSU method in **g**, and then combined with the border nodes $\mathbf{S}_B$ from **h**, for a better estimation of the background border seeds in **i**, where $\mathbf{S}_i$ are the replacement background nodes. $\mathbf{S}_B$ can be compared with the ground truth border seeds in **j**

magnitude may be computed as $\theta^{(ij)} = \arctan\left(I_y^{(ij)}/I_x^{(ij)}\right)$ and $\xi^{(ij)} = \|[I_x^{(ij)}, I_y^{(ij)}]\|_2$. We use the arctan function that limits the angle range between $[-\pi, \pi]$; this may result in two points under the same orientation line image to have two opposite direction angles. To remove these opposite angles that could hinder the distance comparison, we added π to all negative angles, effectively changing the orientation limits between $[0, \pi]$. We then collected the orientation of every pixel in the local patch by adding $\xi^{(ij)}$ to the histogram bin. For each local patch, we normalize the WHO histogram between $[0, 1]$; this allows better intra-node comparisons and avoids possible extreme values from border nodes to over-

whelm the histogram magnitude. For colour images, we keep WHO features for each image channel.

For each graph node, we have features from the *Lab* colour vector and the WHO feature vector having the dimensions of the specified orientation bins. Both *a* and *b* channels are by definition zero when the image is monochromatic. The feature difference as noted by Eq. (4) is used for the colour vector, while the similarity distance of the WHO feature is compared using the *Bhattacharyya* distance between histograms. Since the histogram distance and the intensity distance may have very different magnitudes, we use the following equation to bring the two distances into the same range,

$$D = D_c + \bar{D}_c \left(1 + D_w \right) , \tag{8}$$

where D_c is the L1-norm of colour intensity as given by Eq. (4), $\bar{D}_c$ is the average of all D_c computed recursively, and D_w is the intra-node histogram distance. Equation (8) allows the colour and orientation histogram to be independent while keeping both portions to be the same range as eachother. Equation (4) also allows each term to be zero locally.

We compute a range of local patch size and the number of bins for the WHO feature, and the results are provided in Fig. 5. We observe a 17% timing increase after introducing the WHO feature with the smallest aperture of 3×3 and bin width of $30°$. The timing increased 46% when the patch aperture is 21×21 with a bin width of $5°$. However, the increased computation time did not translate into better precision performance. In fact, the performance is lower with the highest aperture and finest bin width. The optimal aperture bin width is 5×5 and $20°$ respectively. Figure 5b shows that using orientation in addition to the image intensity increased the mean *F*-measure for colour and greyscale images. The *F*-measure is computed using definitions provided in Sect. 7.3. Overall, there is a consistent increase in all performance metrics for greyscale images, but for colour images, the performance tends to reduce in maximum *F*-measure, AUC and MAE. This result is expected since the greyscale image lacks channel variation in the image intensity which is supplemented by the orientation histogram. The colour image already has good object distinctiveness, and its resolution could be diluted by the less precise orientation map. Based on these results, we implement the WHO feature only on greyscale images for optimal time and precision performance.

5.2 Image texture from feature descriptors

Typical texture-based approach uses Gabor [51], Laplacian of Gaussian and Gaussian filters to extract *textons* [32] from the local image. Our testing of the LM, S and MR8 filter banks [66] shows this approach to be too computationally

**(a)** Time   **(b)** WHO Parameters
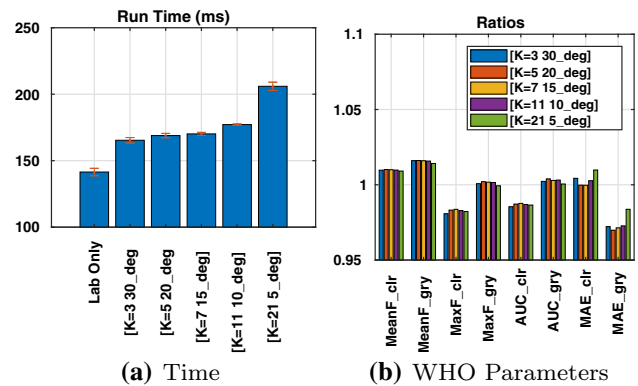
Fig. 5 WHO parameter sensitivity study. Figure **a** provides timing for *Lab* colour-space-only and various $K \times K$ patch size and histogram bin delta angles. Error bars represent one standard deviation of the measured timing. Figure **b** provides ratio of the WHO parameter variation over the *Lab* colour-space-only metrics. The metrics are Mean *F*-measure, Max *F*-measure, AUC, and MAE for colour and greyscale images. Each data point is the average of all 32,536 images

intensive for real-time application. We turned our focus to other image features for their description of the image texture from local difference in illumination. Image features, such as the SIFT descriptors [48], are designed for keypoint matching and SIFT has the property of indicating local intensity gradients. Two neighbouring nodes of the same foreground object under the same illumination should have closer local intensity gradient than with far away nodes of the background scene, and therefore in principle, image descriptors can also be used to describe image texture. During preliminary testing, we have discovered evidence of SIFT and fast retina keypoint (FREAK) [5] image descriptors to be distinctive in the classification of the foreground and background regions. We formally evaluate the effectiveness of six image features using saliency metrics: accelerated-KAZE (AKAZE) [6], FREAK [5], binary robust invariant scalable keypoints (BRISK) [35], binary robust independent elementary features (BRIEF) [14], speed-up robust feature (SURF) [10], and SIFT [48]. The descriptor distance can be computed as $D_f^{(ij)} = |\hat{\mathbf{h}}^{(i)} \cdot \hat{\mathbf{h}}^{(j)}|$ where $\hat{\mathbf{h}}$ is the L2 normalized unit feature vector. We add this feature distance to the colour space L1-norm distance the same way as in Eq. (8) by replacing D_w with D_f.

Results of the various image descriptor ratios over the colour-only baseline compared to the WHO feature ratio in timing and precision are provided in Fig. 6. Figure 6b shows that all descriptors caused increases in the mean *F*-measure for colour and greyscale images, and greyscale images outperformed colour images for all metrics; this confirms the premise that descriptors can be used to add texture feature distinctiveness between nodes. When ordered by publication year, an increase in performance with newer feature descriptors can generally be observed; this again can be expected

(a) Time   **(b)** Image Features
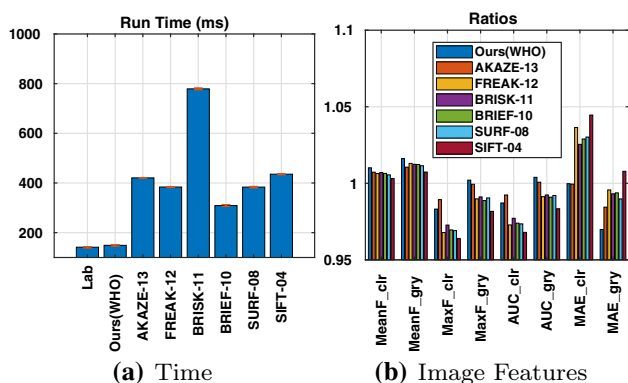
Fig. 6 Image feature sensitivity study. Figure **a** provides timing comparisons for various image feature in addition to the *Lab* colour space. Figure **b** provides ratio of the various image features over the *Lab* performance. The same metrics and number of images are used as per descriptions in Fig. 5. The image feature cases are FREAK [5], AKAZE [6], SURF [10], BRIEF [14], BRISK [35], and SIFT [48]
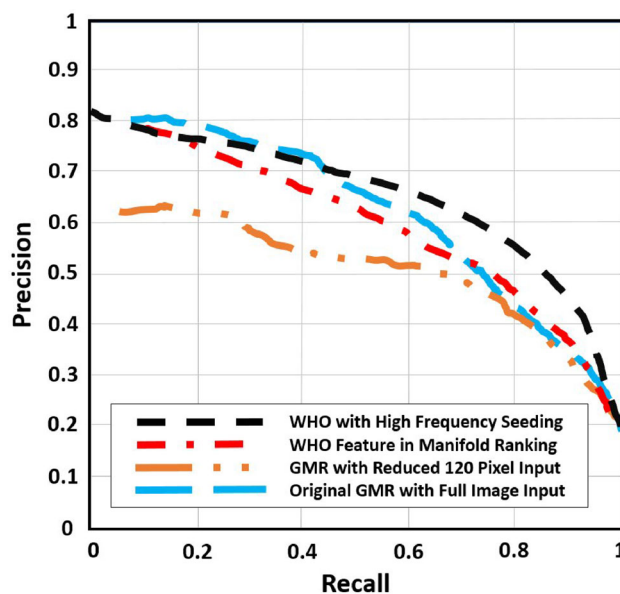

Fig. 7 Precision versus recall for the SatSeg dataset. Original GMR method uses a full-size image as input. The GMR reduced curve uses 120-row pixel input images. The WHO feature curve uses 120-row pixel with WHO feature for node distinction. The enhanced WHO curve uses WHO features and High-frequency response attention-driven seeding as described in Sect. 4.4

as the image descriptors are improved based on similar precision versus recall approach for matching nearest keypoint features. Finally, the WHO feature out-performed all image descriptors with the exception of colour image maximum F-measure, AUC and MAE against AKAZE. More importantly, the WHO feature is $3\times$ to $5\times$ faster than the conventional image descriptors as shown in Fig. 6a. This result alone excludes the use of conventional image descriptors since they would violate our timing requirement. These results support the selection of the proposed WHO features for monochromatic image node representation, in addition to the image intensity.

Figure 7 shows the contribution of the WHO feature and high-frequency response attention-driven seeding in improving the saliency detection quality. The saliency detection precision is reduced by 25% when the input image is reduced to 120-rows, when the WHO feature is used to distinguish the superpixels the precision improved to original levels and when the high-frequency response-driven seeding is applied on top of the WHO features, the saliency precision exceeded the baseline GMR method especially when the recall rate is between 0.6 and 0.8. This demonstrates the effectiveness of the improved seeding and non-colour dependent node features.

6 Proposed saliency models

6.1 Precision model

In the precision model (*prc*), we follow the GMR process described in Sect. 4.2 with the following modifications. First, we resize the input image to 160×120 and set the number of SLIC pixels to approximately 200. The main factor in reduc-

ing computation time is by reducing the size of the input image; however, the image reduction will also degrade algorithm precision when using the original saliency methods. Our proposed saliency method can operate coarse resolution images while maintaining similar precision to the state-of-the-art methods. Computation speed is extremely sensitive to the number of pixels; a small increase in the number of superpixels may not change the precision performance, but it will lead to a significant increase in computation time. We include the WHO features detailed in Sect. 5.1 for greyscale images when computing the Laplacian weighting matrix. Section 5.2 shows the WHO feature is faster than all other tested image descriptors; however, the *Bhattacharyya* distance calculation is still ten times slower than the *Lab* L1-norm and can only be implemented in the *prc* model due to our timing requirement. The weighting matrix is computed by an exhaustive neighbour search through the adjacent matrix for every node. Rather than computing the feature difference during the adjacent node search, we only mark the weighting matrix. The feature difference is then computed in all the marked weights, so there are no duplications in the difference calculation. This enhancement reduced the number of feature difference calculation from $\mathcal{O}\left(n^2\right)$ to $\mathcal{O}\left(n\right)$ where n is the number of adjacent intra-connections. Our timing test shows that this improvement reduced the overall computation time by two-third. Next, we implement PCA OAM reduction with a pixel limit of 150. If the full OAM cannot be inverted, we successively reduce the threshold tolerance by 95% for 5 iter-

ations. In practice, the approximation of the OAM inversion is no longer corrupted after one reduction iteration. Finally, we apply contrast enhancement by a sigmoid function as described in Zhang et al. [88],

$$I^{(ij)} \leftarrow \frac{1}{1 + e^{-b(I^{(ij)} - 0.5)}}, \tag{9}$$

where b is optimally tuned to 10. The sigmoid contrast enhancement reduced the maximum F-measure, but significantly improved the mean F-measure and MAE.

6.2 Speed optimized model

Based on the findings from Sect. 4.4, we propose a speed optimized model (*fst*) that does not use the traditional border nodes as background seeding. Instead, we modify GMR, $\overline{\mathcal{GMR}}(\mathbf{x})$, to use an estimate of the foreground, $\mathbf{S}_{fg}$, and background seeds, $255 - \mathbf{S}_{fg}$, by combining principals of high frequency response, Gaussian centredness, MBD and modified RC. Further, we modify RC, $\overline{\mathcal{RC}}(\mathbf{I}_{Lab})$, by removing the graph segmentation calculation which is the most computationally expensive task in the RC pipeline that is worth 27% of the entire process. Instead, we use the already computed superpixel map and compute the superpixel's spatial centroid. For the ISS test image with 70 superpixels, this reduced the region segmentation calculation from 45 to 12 and 6 ms for SLIC and SEEDS, respectively. We propose a centredness map [52,88] using a Gaussian distribution centred on the spatial moment centroid of the SR map. The Gaussian distribution has the following adaptive scales,

$$\begin{bmatrix} \sigma_x \\ \sigma_y \end{bmatrix} = 4 \left(\begin{bmatrix} C \\ R \end{bmatrix} - 1 \right), \tag{10}$$

where R and C are the numbers of rows and columns of the input image. Let define $\mathcal{MT}(\mathbf{x})$ as the mean binary threshold of an input map. The *fst* algorithm is provided in Algorithm 1.

6.3 Fast maximum precision model

The fast maximum precision model (*fmx*) is optimized to provide maximum precision versus recall curve and AUC. Let define the function that extracts the top, bottom, left, and right border nodes as $\mathcal{BDR}(\mathbf{x})$. The algorithm for *fmx* is provided in Algorithm 2.

Algorithm 1 The *fst* Algorithm

1: **procedure** FST(**I**)
2:     **if** Image larger than 120 rows **then**
3:        Resize image to 120 rows
4:     **end if**
5:     **if** Border frames exist **then**
6:        Remove border frames
7:     **end if**
8:     Check for greyscale image
9:     $\mathbf{I}_{Lab} \leftarrow \mathcal{LAB}(\mathbf{I})$
10:     $\mathbf{S}_S \leftarrow \mathcal{SLIC}(\mathbf{I}_{Lab}, N = 97)$
11:     $\bar{\mathbf{E}} \leftarrow \mathcal{E}(\mathbf{S}_S)$
12:     $\mathbf{S}_{SR} \leftarrow \mathcal{N}(\mathcal{SR}(\mathbf{I}, \sigma = 2))$
13:     $\mathbf{S}_G \leftarrow \mathcal{G}(\mathcal{MC}(\mathbf{S}_{SR}))$
14:     $\mathbf{S}_L \leftarrow \mathcal{N}(|\mathcal{B}(|\mathcal{L}(\mathbf{I})|, K_B = 5)|)$
15:     $\mathbf{S}_{HSF} \leftarrow \mathcal{N}(\mathbf{S}_{SR} + \mathbf{S}_L)$
16:     $\mathbf{S}_{HSF} \leftarrow \mathbf{S}_{HSF} + Mean(\mathbf{S}_{HSF})$
17:     $\mathbf{S}_{MBD} \leftarrow \mathcal{N}(\mathcal{MBD}(\mathbf{I}_{Lab}))$
18:     $\mathbf{S}_{RC} \leftarrow \overline{\mathcal{RC}}(\mathbf{I}_{Lab})$
19:     $\mathbf{S}_{fg} \leftarrow \mathcal{O}(\mathbf{S}_L) \bigcup \mathcal{MT}(\mathbf{S}_G \circ \mathbf{S}_{HSF} \circ \mathbf{S}_{MBD} \circ \mathbf{S}_{RC})$
20:     $\mathbf{S}_{fg} \leftarrow \mathcal{N}(\overline{\mathcal{GMR}}(\mathbf{S}_{fg}) \circ (255 - \overline{\mathcal{GMR}}(255 - \mathbf{S}_{fg})))$
21:     Compute contrast enhancement per Eq. (9)
22:     **if** Resized **then**
23:        Resize saliency map to original size
24:     **end if**
25: **end procedure**

Algorithm 2 The *fmx* Algorithm

1: **procedure** FMX(**I**)
2:     **if** Image larger than 120 rows **then**
3:        Resize image to 120 rows
4:     **end if**
5:     **if** Border frames exist **then**
6:        Remove border frames
7:     **end if**
8:     Check for greyscale image
9:     $\mathbf{S}_G \leftarrow \mathcal{G}(\mathcal{MC}(\mathcal{SR}(\mathbf{I}, \sigma = 8)))$
10:     $\mathbf{S}_L \leftarrow |\mathcal{B}(|\mathcal{L}(\mathbf{I})|, K_B = 3)|$
11:     $\mathbf{S}_{HSF} \leftarrow \mathbf{S}_G \circ \mathbf{S}_L$
12:     $\mathbf{I}_{Lab} \leftarrow \mathcal{LAB}(\mathbf{I})$
13:     $\mathbf{S}_S \leftarrow \mathcal{SEEDS}(\mathbf{I}, N = 225)$
14:     $\bar{\mathbf{E}} \leftarrow \mathcal{E}(\mathbf{S}_S)$
15:     $\mathbf{S}_O \leftarrow \mathcal{O}(\mathbf{S}_{HSF})$
16:     **for** $i = top, bottom, left, right$ **do**
17:        $\mathbf{S}_B^{(i)} \leftarrow \mathcal{BDR}(\mathbf{S}_S, border = i) \bigcap \mathbf{S}_O + \mathbf{S}_i$
18:     **end for**
19:     $\mathbf{S}_{GMR} \leftarrow \overline{\mathcal{GMR}}(\mathbf{S}_B^t, \mathbf{S}_B^b, \mathbf{S}_B^l, \mathbf{S}_B^r)$
20:     $\mathbf{S}_{MBD} \leftarrow \mathcal{MBD}(\mathbf{I}_{Lab})$
21:     $\mathbf{S}_{RC} \leftarrow \overline{\mathcal{RC}}(\mathbf{I}_{Lab})$
22:     $\mathbf{S}_{fg} \leftarrow \mathbf{S}_G \circ \mathbf{S}_{GMR} \circ \mathbf{S}_{MBD} \circ \mathbf{S}_{RC}$
23:     **if** Resized **then**
24:        Resize saliency map to original size
25:     **end if**
26: **end procedure**

---

7 Experimental data and metrics

7.1 Datasets

While our design focuses on spacecraft applications, our developed models can be used for arbitrary images. We

Table 1 List of benchmark models used for comparison

Code	References	Description
OTSU	[54]	OTSU thresholding
WS	[53]	Watershed
GCUT	[57]	Grabcut
LC	[85] [20]*	Colour contrast
SR	[27]	Spectral residual
QFT	[26]	Phase spectrum quaternion Fourier transform
AC	[2] [13]*	Local contrast raster scan
FT	[3]	Frequency tuned
MSS	[1]	Maximum symmetric surround
DRLSE	[36]	Distance regularized level-set evolution
HC	[17]	Histogram-based contrast
GMR	[80]	Graph manifold ranking
GC	[18]	Global cues
BING	[19]	BInarized normal gradients
GD	[88]	Geodesic
MBD	[88]	Minimum barrier distance
MB+	[88]	MBD extended
RC	[20]	Regional contrast

Starred references (*) is where source code or executable was downloaded from

use ECSSD [59], DUT-OMRON [80], and MSRA10K [45] datasets to evaluate the benchmark saliency models. We generate a satellite segmentation (SatSeg) dataset composed of colour and greyscale spacecraft images captured by visible and infrared cameras.

Ground truth data was produced manually as foreground (255) and background (0) masks. All dataset can be downloaded from our project website.[3]

7.2 Model comparisons

We compared our models with 18 traditional and state-of-the-art saliency detection and segmentation models provided in Table 1. These models were selected based on their real-time and near-real-time running performance. For ECSSD, DUT-OMRON and MSRA10K datasets, comparisons were made against LC [85], SR [27], QFT [26], AC [2], FT [3], MSS [1], HC [17], GMR [80], GC [18], GD [88], MBD [88], MB+ [88] and RC [20]. For the SatSeg tests, we added OTSU [54], WS [53], GCUT [57], DRLSE [36] and BING [19] to the previous list. This allowed running-time comparisons between saliency generation, direct thresholding, segmentation and objectiveness methods. WS, GCUT, DRLSE are segmentation methods and can only produce a single foreground-to-background binary mask. Ground truth masks was used as foreground seeding for WS, a border box was used as background seeding for GCUT, and the DRLSE mask

is the inner region of the level-set function. OTSU mask was generated by directly applying Otsu thresholding on the original image. The BING [19] technique is a fast objectness measure. We can normalize all of its bounding-boxes on a single image to generate a saliency map. While the latter method is not typically considered as saliency detection, we may use it to evaluate the object segmentation problem.

7.3 Evaluation metrics

We apply receiver operating characteristics (ROC) [23] as the evaluation metrics for this investigation. For a computed saliency map, $\mathbf{S}$, it is converted to an 8-bit binary mask image, $\mathbf{M}$, by applying a constant threshold. The precision and recall of a single image can be computed from the ground truth by performing the following piece-wise operation to the saliency maps,

$$\text{Prec} = \frac{\text{TP}}{\text{EP}} = \frac{\sum_{i=1}^{R}\sum_{j=1}^{C}|M^{(ij)}\cap G^{(ij)}|}{\sum_{i=1}^{R}\sum_{i=1}^{C}|M^{(ij)}|}, \tag{11}$$

$$\text{Rcal} = \frac{\text{TP}}{\text{AP}} = \frac{\sum_{i=1}^{R}\sum_{j=1}^{C}|M^{(ij)}\cap G^{(ij)}|}{\sum_{i=1}^{R}\sum_{i=1}^{C}|G^{(ij)}|}, \tag{12}$$

where TP, EP, AP are true positive, estimated positive, and actual positive respectively, M_{ij} and G_{ij} are the individual pixels in the saliency and ground truth maps $\mathbf{M}$ and $\mathbf{G}$ respectively. The variables R and C are the rows and columns of the saliency and ground truth maps, respectively. A threshold

[3] http://ai-automata.ca/research/hisafe.html.

ranging from 0 to 255 is used to control recall. We compute the standard units of measure for saliency evaluations including average and maximum F-measure [3] and AUC from the ROC curve. The F-measure is defined as

$$F_\beta = \frac{(1 + \beta^2)\, \text{Prec} \times \text{Rcal}}{\beta^2 \times \text{Prec} + \text{Rcal}}, \tag{13}$$

where β^2 is set to 0.3 [3]. The ROC curve is generated using the following definitions of true positive rate (TPR), where TPR = Rcal, and false positive rate (FPR), where

$$\text{FPR} = \frac{\text{EP} - \text{TP}}{\text{Total} - \text{AP}} = \frac{\sum_{i=1}^{R} \sum_{j=1}^{C} |M^{(ij)} \cap (255 - G^{(ij)})|}{\sum_{i=1}^{R} \sum_{i=1}^{C} |255 - G^{(ij)}|}. \tag{14}$$

We use the mean absolute error (MAE) [13] to take into account continuous saliency map variations. Both the saliency map, $\bar{\mathbf{S}}$, and the ground truth map, $\bar{\mathbf{G}}$, are normalized between the range of [0, 1]. The MAE is defined as

$$MAE = \frac{1}{RC} \sum_{i=1}^{R} \sum_{j=1}^{C} |\bar{S}^{(ij)} - \bar{G}^{(ij)}|. \tag{15}$$

where $\bar{S}^{(ij)}$ and $\bar{G}^{(ij)}$ are the (i, j) elements of $\bar{\mathbf{S}}$ and $\bar{\mathbf{G}}$ respectively.

8 Results and discussions

8.1 Precision performance

We compared our models with the benchmark models using three standard datasets ECSSD [59,79], DUT-OMRON [80], MSRA10K [20,45] and SatSeg over a total of 32,536 test images including colour and greyscale variations. Figures 8 and 9 provides selected qualitative results for the standard and SatSeg datasets respectively. Due to space limitations, we only present the highest quality benchmark models for comparison plus the SR model as a reference. Even though our primary focus is to reduce runtime, our models have exceeded the precision performance of most benchmark models in several areas. Notably, the *Monk* image in Fig. 8a and the *Rose* image in Fig. 8b our *prc* and *fmx* models out-performs all benchmark techniques. In the *ISS Blue Earth* (5th row), *ISS Cloud Earth* (6th row), *ISS Solar Panel* (7th row), *Orbital Express over Earth* (9th row), and *Space Shuttle* (11th row) images, almost all benchmarking models cannot distinguish the Earth background and the foreground vehicle. In contrast,

our *fmx* and *fst* methods almost entirely removed the Earth background. Non-distinctiveness is worsened when only the monochromatic image is available. An example of this is the *Radarsat Model* (10th row); in this case, our *fmx* and *fst* method provide the best match to the ground truth.

Quantitative results for all datasets are shown in Figs. 10 and 11 for colour and greyscale images, respectively. For all datasets, benchmark models that rely on colour perform worse when only the greyscale intensity is available. Our *prc* model has the highest mean F-measure and lowest MAE in all standard datasets and is only eclipsed by MB+ in SatSeg for F-measure, and MB+ and RC for MAE. Our *fmx* model consistently matches the top performers in all datasets in precision versus recall and AUC. Our *prc* model outperforms all models in the greyscale ECSSD dataset measured by the precision versus recall curve, as demonstrated in Fig. 11a. In the greyscale SatSeg dataset in Fig. 11, our *fmx* model has a higher precision versus recall performance than all other models. Our *fst* model matches the top RC and MB+ performance and outperforms all other methods in the mean F-measure. OTSU, DRLSE, and BING have much higher MAE than other methods. This is because OTSU and DRLSE are intensity threshold and segmentation methods, while BING is an objectiveness bounding box technique and was not designed to be a salient image generator. The general saliency performance trend increases with increasing publication year. While our models did not achieve a considerable increase in precision performance, we meet or exceed the benchmark precision and significantly exceeded the benchmark model run speeds.

8.2 Timing analysis

A speed test was performed using the average run time for computing the SatSet dataset. Again, in our specific application, the timing requirement is less than 50 ms. Figure 12 provides the timing for the various models. Majority of the models are implemented in C++ on the 64L platform.[4] WS, GCUT, OTSU, SR and BING are class modules from OpenCV 3.3. AC, GC, and MB+ are windows compiled executables provided by the various authors that can only be run from the 64W platform.[5] The source code for DRLSE was supplied by [36] in MATLAB and was executed on the 32W platform.[6] Figure 12 shows that the slowest model is DRLSE, it has an average run time of approximately 2.5 minutes per image. While the platform and the coding language skewed the DRLSE timing higher, this model requires time-consuming evolutions of the level-set function and is deemed to be too costly for real-time. In the original MB paper, the
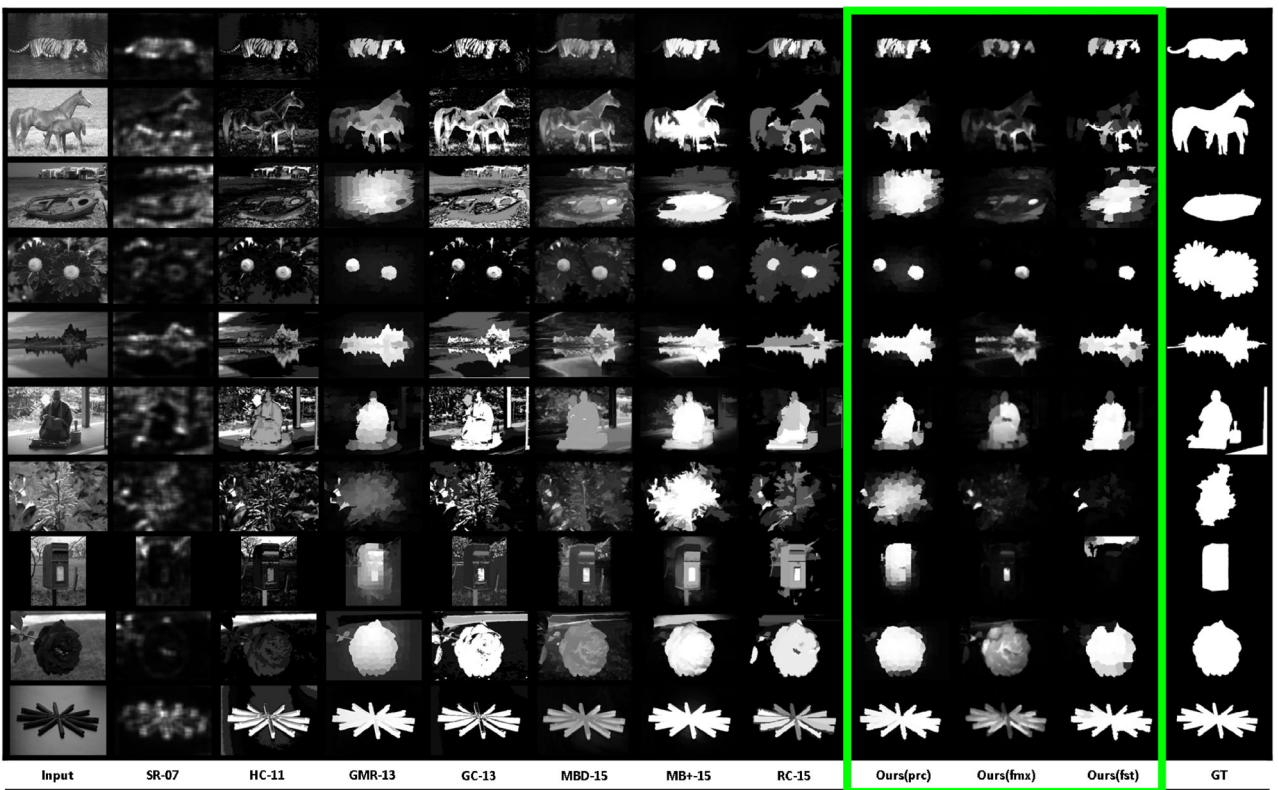
---

[4] AMD FX-8350 8-4.0GHz.

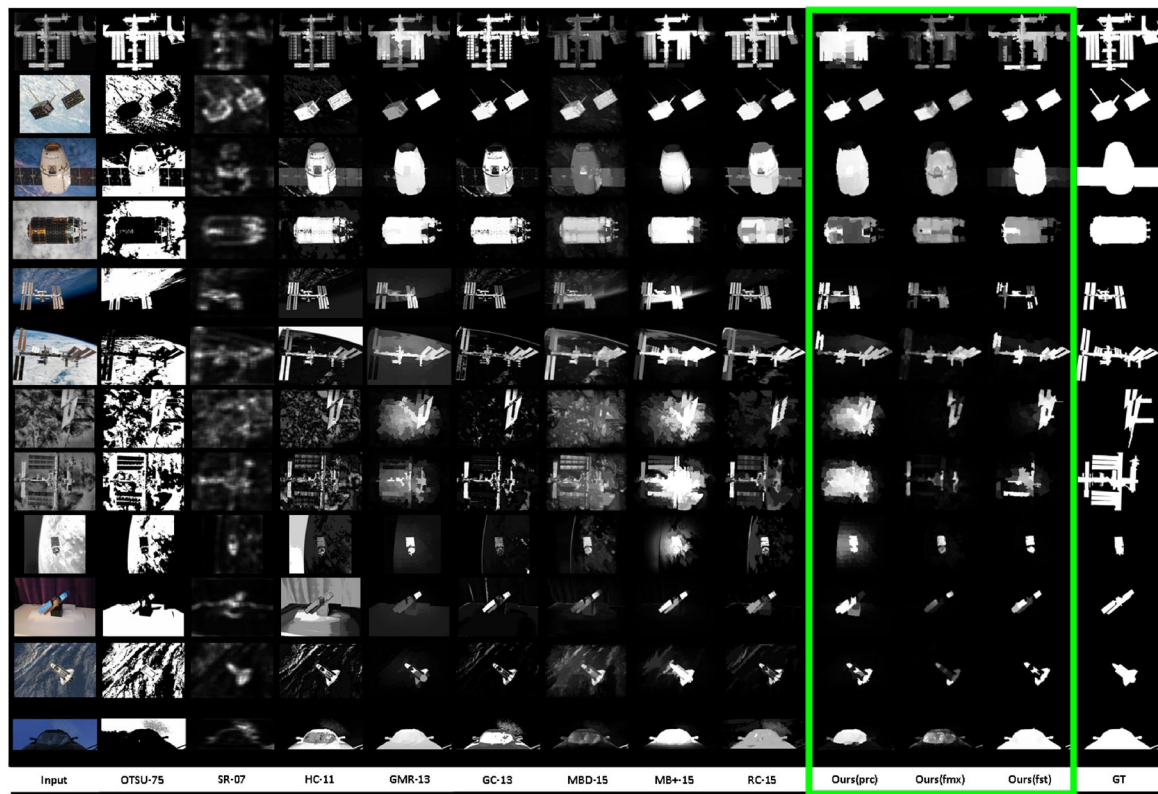[5] AMD A4-5000-APU 4-1.5GHz.

[6] Intel 2-Quad-Q6600 4-2.4GHz.
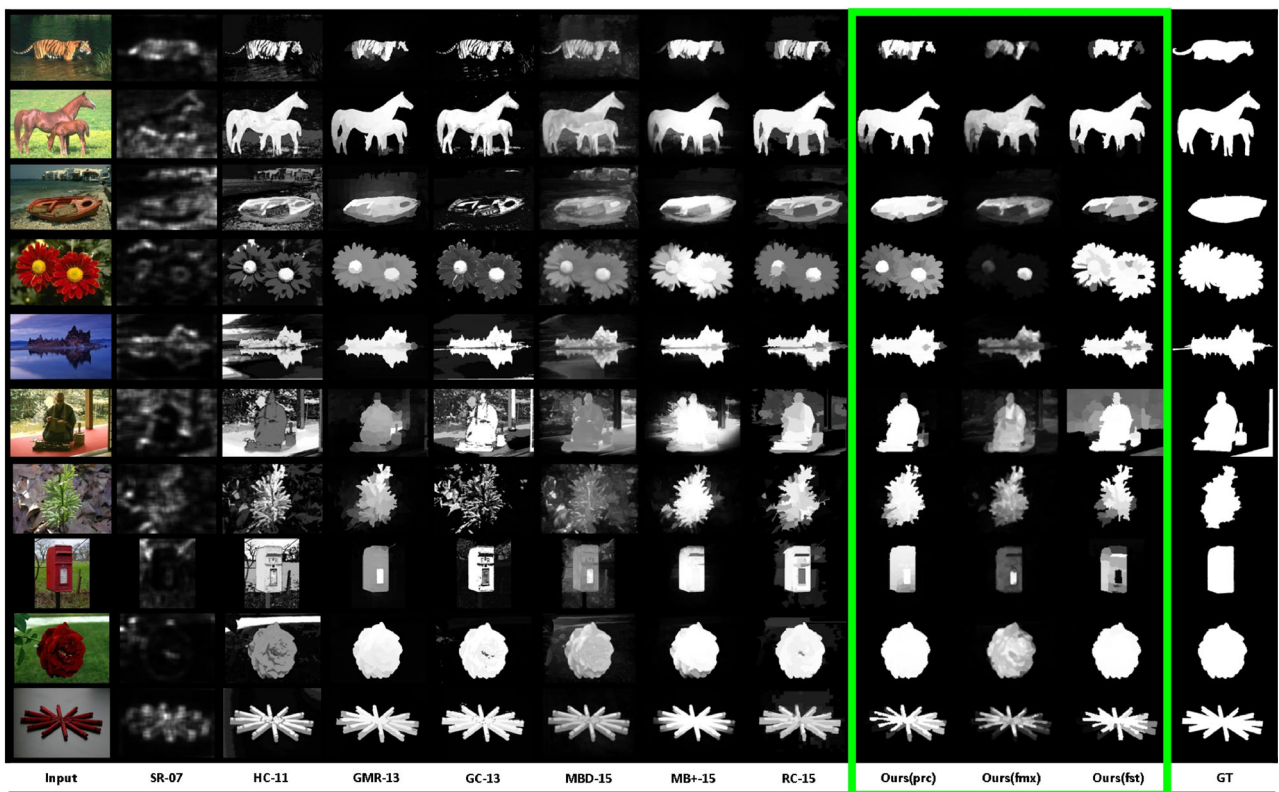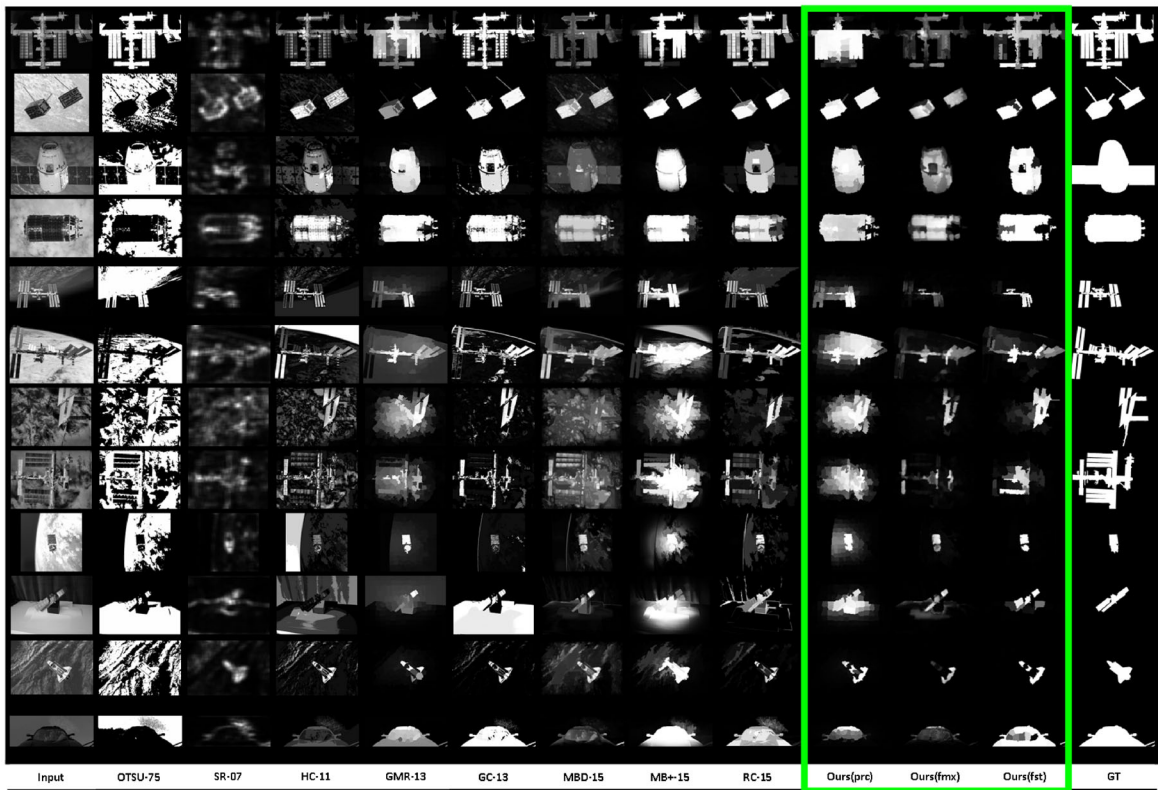
(a) Colour

(b) Grayscale

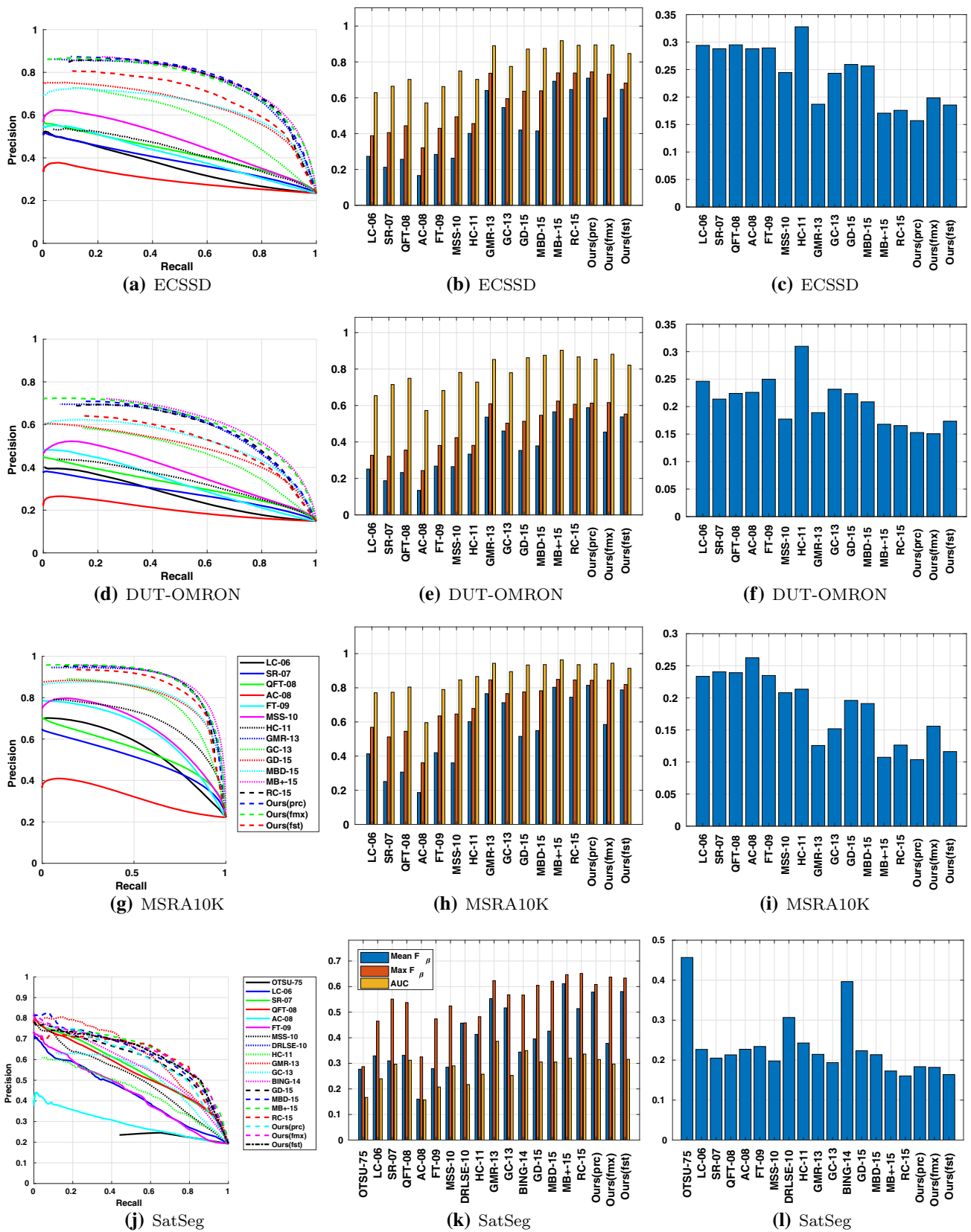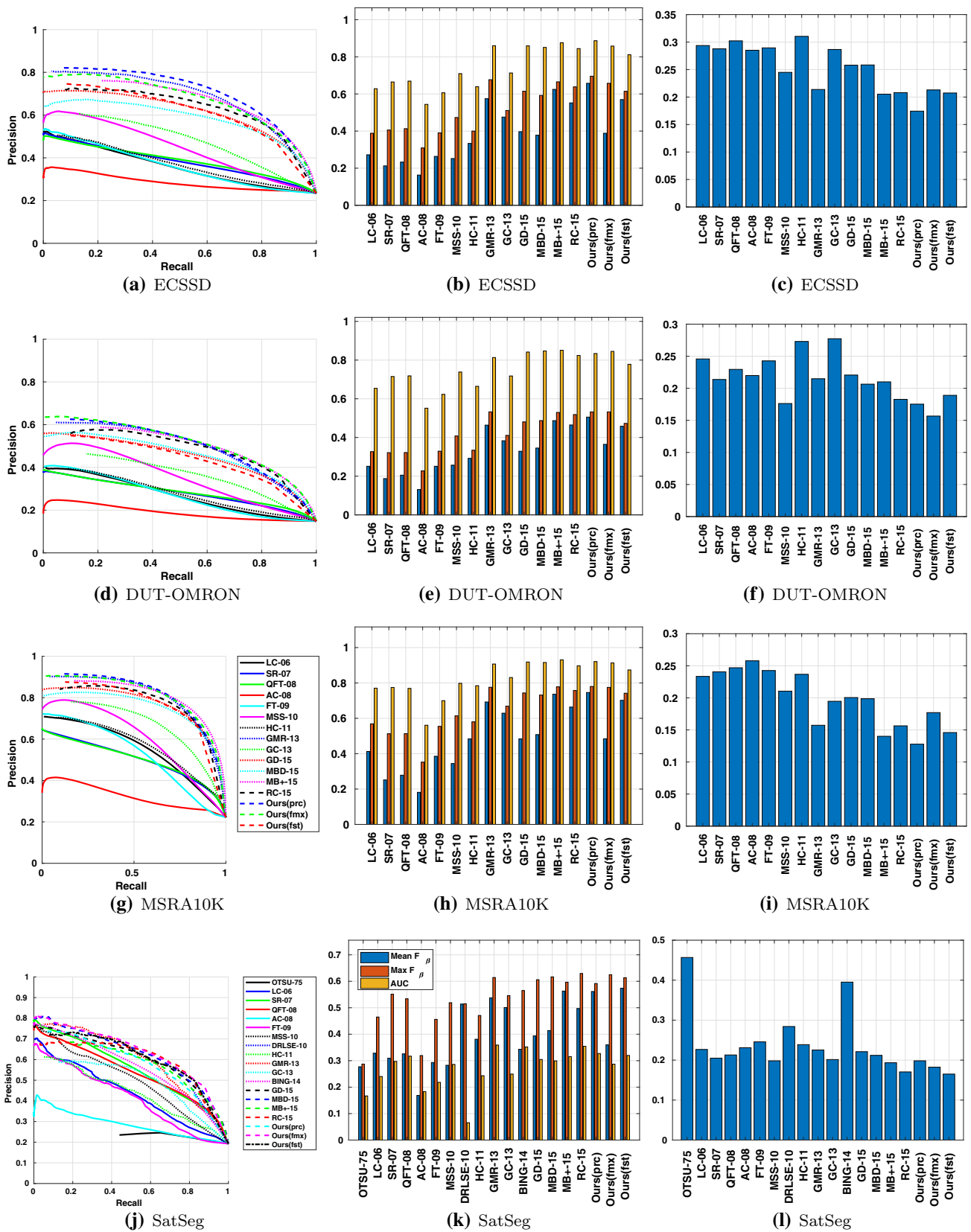**Fig. 8** Standard dataset salient image comparisons

(a) Colour

(b) Grayscale

Fig. 9 SatSeg dataset salient image comparisons

Fig. 10 Saliency detection results for colour images. Column 1: ECSSD, DUT-OMRON, MSRA10K shares the same legend in **g**. Column 2: all figures shares the same legend in **k**

Fig. 11 Saliency detection results for greyscale images. Column 1: ECSSD, DUT-OMRON, MSRA10K shares the same legend in **g**. Column 2: all figures shares the same legend in **k**
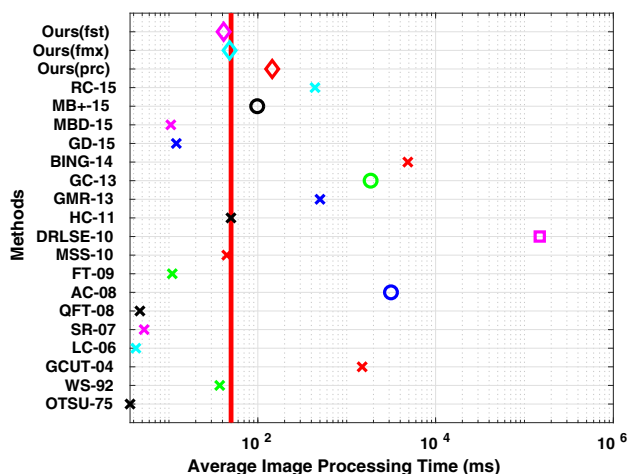
Fig. 12 Saliency model timing comparisons. Vertical red line indicates the required real-time performance. The × markers indicate runs computed using the 64L platform with C++; the ∘ markers indicate runs calculated using the 64W platform with windows compiled executables, the □ markers indicate runs computed using 32W platform with MATLAB, the ◇ markers indicate our models ran on the 64L platform with C++. Refer to Sect. 8.2 for details on the computing platforms

authors reported speed of 80 frames per second. The absolute speed of an algorithm is purely based on the processing hardware and the resolution of the test images. MB+ runs slower than 80 frames per second when using our processing hardware and test images; however, it still has the fastest timing compared to the other Windows executables. It doubles the requirement of 50 ms running at 98.5 ms per image. It is possible this method could be even faster running on the 64L platform; however, our C++ implementation of the MB+ extensions by [88] did not achieve the specified performance. Both RC and GMR exceeded the 50 ms target with 440 and 500 ms per image respectively. Our *fmx* and *fst* models are more than $10\times$ faster at 48.122 ± 0.186 and 41.242 ± 0.097 ms respectively, whereas our *prc* model is still $3\times$ faster at 145.09 ± 1.197 ms. OTSU, LC, SR, QFT are on the order of 3 to 5 ms but fails to achieve the same precision as our model. Overall, our *fst* and *fmx* methods developed in this work achieved the best precision when compared to models with similar speeds.

9 Conclusion

In this work, we presented three high-performance approaches for real-time saliency detection in greyscale images. We showed our approach has better performance than the original GMR and has the same precision performance as MB+ and RC. In developing our saliency generator, we also created the WHO feature histogram and the non-normalized OAM reduction using PCA. We showed speed and precision improvements by using these techniques. Both the WHO fea-

ture and PCA OAM inverse reduction may be used beyond salient image generation. Future work includes advancing greyscale image saliency generation by incorporating generically trained CNN filters.

A Proofs

The proof of Lemma 1 is as follows,

Proof If the alternative form of $\bar{\mathbf{A}}$ from Lemma 1 is equivalent, then when multiplied by the inverse it will result in the identity.

$$(1 - \alpha\mathbf{EF})\left(1 - \mathbf{E}\left(\mathbf{FE} - \frac{1}{\alpha}\mathbf{1}_d\right)^{-1}\mathbf{F}\right)$$

$$= 1 - \alpha\mathbf{EF} - \mathbf{E}\left(\mathbf{FE} - \frac{1}{\alpha}\mathbf{1}_d\right)^{-1}\mathbf{F} + \alpha\mathbf{EFE}\left(\mathbf{FE} - \frac{1}{\alpha}\mathbf{1}_d\right)^{-1}\mathbf{F}$$

$$= 1 - \alpha\mathbf{EF} + \alpha\mathbf{E}\left(-\frac{1}{\alpha}\mathbf{1}_d + \mathbf{FE}\right)\left(\mathbf{FE} - \frac{1}{\alpha}\mathbf{1}_d\right)^{-1}\mathbf{F}$$

$$= 1 \qquad\qquad\square$$

The proof of Corollary 1 is as follows,

Proof Substituting Lemma 1 into Eq. (6), and using the previous definitions for $\mathbf{E}$ and $\mathbf{F}$,

$$\mathbf{A} = (\mathbf{D} - \alpha\mathbf{W})^{-1}$$

$$= \mathbf{D}^{-\frac{1}{2}}\left(1 - \mathbf{E}\left(\mathbf{FE} - \frac{1}{\alpha}\mathbf{1}_d\right)^{-1}\mathbf{F}\right)\mathbf{D}^{-\frac{1}{2}}$$

$$= \mathbf{D}^{-1}\left(1 - \mathbf{P}\left(\mathbf{QD}^{-1}\mathbf{P} - \frac{1}{\alpha}\mathbf{1}_d\right)^{-1}\mathbf{QD}^{-1}\right) \qquad\square$$

References

1. Achanta, R., Süsstrunk, S.: Saliency detection using maximum symmetric surround. In: 17th IEEE International Conference on Image Processing, Hong Kong, China (2010)
2. Achanta, R., Estrada, F., Wils, P., Süsstrunk, S.: Salient region detection and segmentation. In: IEEE International Conference on Computer Vision (2008)
3. Achanta, R., Hemami, S., Estrada, F., Süsstrunk, S.: Frequency-tuned salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1597–1604 (2009)
4. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2274–2281 (2012)
5. Alahi, A., Ortiz, R., Vandergheynst, P.: Freak: Fast retina keypoint. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 510–517 (2012)
6. Alcantarilla, P., Nuevo, J., Bartoli, A.: Fast explicit diffusion for accelerated features in nonlinear scale spaces. In: British Machine Vision Conference, Bristol, UK (2013)

7. Assens, M., Giro-i Nieto, X., McGuinness, K., Ne, O.: Saltinet: scan-path prediction on 360 degree images using saliency volumes. In: IEEE International Conference on Computer Vision Workshop (2017)

8. Avidan, S., Shamir, A.: Seam carving for content-aware image resizing. ACM Trans. Graph. **26**(3), 10–19 (2007)

9. Aye, H., Zaw, S.: Salient object based action recognition using histogram of changing edge orientation. In: IEEE International Conference on Software Engineering Research, Management and Applications (2017)

10. Bay, H., Ess, A., Tuytelaars, T., Gool, L.: Speeded-up robust features. Comput. Vis. Image Underst. **110**(3), 346–359 (2008)

11. Borenstein, E., Malik, J.: Shape guided object segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 969–976 (2006)

12. Borji, A.: Boosting bottom-up and top-down visual features for saliency estimation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 438–445 (2012)

13. Borji, A., Cheng, M., Jiang, H., Li, J.: Salient object detection: a benchmark. IEEE Trans. Image Process. **24**(12), 5706–5723 (2015)

14. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: Brief: binary robust independent elementary features. In: European Conference on Computer Vision, Heraklion, Crete (2010)

15. Chan, K.: Saliency/non-saliency segregation in video sequences using perception-based local ternary pattern features. In: IEEE International Conference on Machine Vision Applications (2017)

16. Chen, T., Lin, L., Liu, L., Luo, X., Li, X.: Disc: deep image saliency computing via progressive representation learning. IEEE Trans. Neural Netw. Learn. Syst. **27**(6), 1135–1149 (2016)

17. Cheng, M., Zhang, G., Mitra, N., Huang, X., Hu, S.: Global contrast based salient region detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 409–416 (2011)

18. Cheng, M., Warrell, J., Lin, W., Zheng, S., Vineet, V., Crook, N.: Efficient salient region detection with soft image abstraction. In: IEEE International Conference on Computer Vision, pp. 1529–1536 (2013)

19. Cheng, M., Zhang, Z., Lin, W., Torr, P.: Bing: binarized normed gradients for objectness estimates at 300 fps. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3286–3293 (2014)

20. Cheng, M., Mitra, N., Huang, X., Torr, P., Hu, S.: Global contrast based salient region detection. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 569–582 (2015)

21. Christopoulos, C., Skodras, A., Ebrahimi, T.: The JPEG2000 still image coding system: an overview. IEEE Trans. Consum. Electron. **46**(4), 1103–1127 (2002)

22. Fattal, A.K., Karg, M., Scharfenberger, C., Adamy, J.: Saliency-guided region proposal network for CNN based object detection. In: IEEE International Conference on Intelligent Transportation System (2017)

23. Fawcett, T.: An introduction to ROC analysis. Pattern Recognit. Lett. **27**, 861–874 (2006)

24. Felzenszwalb, P., Huttenlocher, D.: Efficient graph-based image segmentation. Int. J. Comput. Vis. **59**(2), 167–181 (2004)

25. Fujiwara, Y., Irie, G., Kuroyama, S., Onizuka, M.: Scaling manifold ranking based image retrieval. In: Proceedings of the VLDB Endowment, vol. 8 (2014)

26. Guo, C., Ma, Q., Zhang, L.: Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform. In: IEEE Conference on Computer Vision and Pattern Recognition (2008)

27. Hou, X., Zhang, L.: Saliency detection: a spectral residual approach. In: IEEE Conference on Computer Vision and Pattern Recognition (2007)

28. Howard, R., Heaton, A., Pinson, R., Carrington, C.: Orbital express advanced video guidance sensor. In: IEEE Aerospace Conference, Big Sky, MT (2008)

29. Hu, P., Shuai, B., Liu, J., Wang, G.: Deep levelsets for salient object detection. In: IEEE Conference on Computer Vision and Pattern Recognition (2017)

30. Huang, K., Gao, S.: Image saliency detection via multi-scale iterative CNN. Vis. Comput. J. (2019). https://doi.org/10.1007/s00371-019-01734-2

31. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention, for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **20**(11), 1254–1259 (1998)

32. Julesz, B.: Textons, the elements of texture perception, and their interactions. Nature **290**(5802), 91–97 (1981)

33. Jung, C., Kim, W., Yoo, S., Kim, C.: A novel monochromatic cue for detecting regions of visual interest. J. Image Vis. Comput. **32**, 405–413 (2014)

34. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. Hum. Neurobiol. **4**, 219–227 (1985)

35. Leutenegger, S., Chli, M., Siegwart, R.: Brisk: binary robust invariant scalable keypoints. In: IEEE International Conference on Computer Vision, pp. 2548–2555 (2011)

36. Li, C., Xu, C., Gui, C., Fox, M.: Distance regularized level set evolution and its application to image segmentation. IEEE Trans. Image Process. **19**(12), 3243–3254 (2010)

37. Li, C., Zhang, B., Zhang, S., Sheng, H.: Saliency detection with relative location measure in light field image. In: IEEE International Conference on Image, Vision and Computing, Chengdu, China (2017)

38. Li, G., Xie, Y., Lin, L., Yu, Y.: Instance-level salient object segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 247–256 (2017)

39. Li, J., Tian, Y., Huang, T., Gao, W.: Probabilistic multi-task learning for visual saliency estimation in video. Int. J. Comput. Vis. **90**(2), 150–165 (2010)

40. Li, J., Tian, Y., Chen, X., Huang, T.: Measuring visual surprise jointly from intrinsic and extrinsic contexts for image saliency estimation. Int. J. Comput. Vis. **120**(1), 44–60 (2016)

41. Li, L., Zhou, F., Zheng, Y., Bai, X.: Saliency detection based on foreground appearance and background-prior. Neurocomputing **301**, 46–61 (2018)

42. Li, S., Zeng, C., Fu, Y., Liu, S.: Optimizing multi-graph learning based salient object detection. Signal Process. Image Commun. **55**, 93–105 (2017)

43. Li, Z., Chen, J.: Superpixel segmentation using linear spectral clustering. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1356–1363 (2015)

44. Liu, T., Sun, J., Zheng, N., Tang, X., Shum, H.: Learning to detect a salient object. In: IEEE Conference on Computer Vision and Pattern Recognition (2007)

45. Liu, T., Yuan, Z., Sun, J., Wang, J., Zheng, N., Tang, X., Shum, H.: Learning to detect a salient object. IEEE Trans. Pattern Anal. Mach. Intell. **33**(2), 353–367 (2011)

46. Liu, Z., Tang, J., Zhao, P.: Salient object detection via hybrid upsampling and hybrid loss computing. Vis. Comput. J. (2019). https://doi.org/10.1007/s00371-019-01659-w

47. Liu, Z., Xiang, Q., Tang, J., Wang, Y., Zhao, P.: Robust salient object detection for RGB images. Vis. Comput. J. (2019). https://doi.org/10.1007/s00371-019-01778-4

48. Lowe, D.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)

49. Lu, Y., Zhou, K., Wu, X., Gong, P.: A novel multi-graph framework for salient object detection. Vis. Comput. J. **35**(11), 1683–1699 (2019)

50. Luo, Z., Mishra, A., Achkar, A., Eichel, S., Li, J., Jodoin, P.: Non-local deep features for salient object detection. In: IEEE International Conference on Computer Vision (2017)

51. Malik, J., Perona, P.: Preattentive texture discrimination with early vision mechanisms. J. Opt. Soc. Am. A **7**(5), 923–932 (1990)

52. Margolin, R., Tal, A., Zelnik-Manor, L.: What makes a patch distinct? In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1139–1146 (2013)

53. Meyer, F.: Color image segmentation. In: IET International Conference on Image Processing and Its Applications (1992)

54. Otsu, N.: A threshold selection method from gray-level histograms. Automatica **11**, 23–27 (1975)

55. Qi, W., Cheng, M., Borji, A., Lu, H., Bai, L.: Saliencyrank: two-stage manifold ranking for salient object detection. Comput. Vis. Media **1**(4), 309–320 (2015)

56. Rahtu, E., Kannala, J., Salo, M., Heikkilä, J.: Segmenting salient objects from images and videos. In: European Conference on Computer Vision, pp. 366–379 (2010)

57. Rother, C., Kolmogorov, V., Blake, A.: "grabcut"-interactive foreground extraction using iterated graph cuts. ACM Trans. Graph. **23**(3), 309–314 (2004)

58. Seo, H., Milanfar, P.: Static and space-time visual saliency detection by self-resemblance. J. Vis. **9**(12), 15 (2009)

59. Shi, J., Yan, Q., Xu, L., Jia, J.: Hierarchical image saliency detection on extended CSSD. IEEE Trans. Pattern Anal. Mach. Intell. **38**(4), 717–729 (2016)

60. Shi, J., Ulrich, S., Ruel, S.: Regional method for monocular infrared image spacecraft pose estimation. In: Proceedings of the AIAA Space Conference and Exhibit, Orlando, FL (2018)

61. Shigematsu, R., Feng, D., You, S., Barnes, N.: Learning RGB-D salient object detection using background enclosure, depth contrast, and top-down features. In: IEEE International Conference on Computer Vision (2017)

62. Strand, R., Ciesielski, K., Malmberg, F., Saha, P.: The minimum barrier distance. Comput. Vis. Image Underst. **117**(4), 429–437 (2013)

63. Tan, Z., Wan, L., Feng, W., Pun, C.: Image co-saliency detection by propagating superpixel affinities. In: IEEE International Conference on Acoustics, Speech and Signal Processing (2013)

64. Tian, Z., Zheng, N., Xue, J., Lan, X., Li, C., Zhou, G.: Video object segmentation with shape cue based on spatiotemporal superpixel neighbourhood. IET Comput. Vis. **8**(1), 16–25 (2014)

65. Van den Bergh, M., Boix, X., Roig, G., Van Gool, L.: Seeds: superpixels extracted via energy-driven sampling. Int. J. Comput. Vis. **111**(3), 298–314 (2015)

66. Varma, M., Zisserman, A.: A statistical approach to texture classification from single images. Int. J. Comput. Vis. **62**(1), 61–81 (2005)

67. Wan, X., Yang, J., Xiao, J.: Manifold-ranking based topic-focused multi-document summarization. In: Proceedings of International Joint Conference on Artificial Intelligence, pp. 2903–2908 (2007)

68. Wang, J., Lu, H., Li, X., Tong, N., Liu, W.: Saliency detection via background and foreground seed selection. Neurocomputing **152**, 359–368 (2015)

69. Wang, J., Jiang, H., Yuan, Z., Cheng, M., Hu, X., Zheng, N.: Salient object detection: a discriminative regional feature integration approach. Int. J. Comput. Vis. **123**(2), 251–268 (2017)

70. Wang, W., Shen, J., Yang, R., Porikli, F.: Saliency-aware video object segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **40**(1), 20–33 (2018)

71. Wang, Y., Wei, X., Ding, L., Tang, X., Zhang, H.: A robust visual tracking method via local feature extraction and saliency detection. Vis. Comput. J. (2019). https://doi.org/10.1007/s00371-019-01646-1

72. Wang, Z., Wu, X.: Salient object detection using biogeography-based optimization to combine features. Appl. Intell. **45**(1), 1–17 (2016)

73. Wei, Y., Wen, F., Zhu, W., Sun, J.: Geodesic saliency using background priors. In: European Conference on Computer Vision, pp. 29–42 (2012)

74. Wu, X., Lin, X., Jiang, L., Zhao, D.: An improved manifold ranking based method for saliency detection. In: IEEE International Conference on Systems and Informatics (2017)

75. Xia, C., Li, J., Chen, X., Zheng, A., Zhang, Y.: What is and what is not a salient object, learning salient object detector by ensembling linear exemplar regressors. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 4321–4329 (2017)

76. Xu, B., Bu, J., Chen, C., Cai, D., He, X., Liu, W., Luo, J.: Efficient manifold ranking for image retrieval. In: International ACM Conference on Research and Development in Information Retrieval (2011)

77. Xu, Y., Li, J., Chen, J., Shen, G., Gao, Y.: A novel approach for visual saliency detection and segmentation based on objectness and top-down attention. In: IEEE International Conference on Image, Vision and Computing, pp. 4321–4329 (2017)

78. Yacoob, Y., Davis, L.: Segmentation using meta-texture saliency. In: IEEE International Conference on Computer Vision (2007)

79. Yan, Q., Xu, L., Shi, J., Jia, J.: Hierarchical saliency detection. In: IEEE Conference on Computer Vision and Pattern Recognition (2013)

80. Yang, C., Zhang, L., Lu, H., Ruan, X., Yang, M.: Saliency detection via graph-based manifold ranking. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3166–3173 (2013)

81. Yang, C., Pu, J., Dong, Y., Xie, G., Si, Y., Liu, Z.: Scene classification-oriented saliency detection via the modularized prescription. Vis. Comput. J. **35**(4), 473–488 (2019)

82. Yang, J., Yang, M.: Top-down visual saliency via joint CRF and dictionary learning. IEEE Trans. Pattern Anal. Mach. Intell. **39**(3), 576–588 (2017)

83. Ye, L., Liu, Z., Zhou, X., Shen, L., Zhang, J.: Saliency detection via similar image retrieval. IEEE Signal Process. Lett. **23**(6), 838–842 (2016)

84. Ye, R., Chen, Z.: Universal enhancement of salient object detection. In: IEEE International Conference on Multimedia and Expo (2017)

85. Zhai, Y., Shah, M.: Visual attention detection in video sequences using spatiotemporal cues. In: ACM International Conference on Multimedia (2006)

86. Zhang, D., Han, J., Zhang, Y.: Supervision by fusion: Towards unsupervised learning of deep salient object detector. In: IEEE International Conference on Computer Vision (2017)

87. Zhang, J., Shen, Y.: Spectral segmentation via minimum barrier distance. Multimed. Tools Appl. **76**(24), 25713–25729 (2017)

88. Zhang, J., Sclaroff, S., Lin, Z., Shen, X., Price, B., Měch, R.: Minimum barrier salient object detection at 80 fps. In: IEEE International Conference on Computer Vision (2015)

89. Zhang, L., Yang, C., Lu, H., Ruan, X., Yang, M.: Ranking saliency. IEEE Trans. Pattern Anal. Mach. Intell. **39**(9), 1892–1904 (2017)

90. Zhao, J., Gao, X., Chen, Y., Feng, H.: Optical imaging system-based real-time image saliency extraction method. Opt. Eng. **54**(4), 43101-1–43101-8 (2015)

91. Zhou, D., Weston, J., Gretton, A.: Ranking on data manifolds. In: Conference on Neural Information Processing Systems (2004)

Jian-Feng Shi has over 20 years of experience in spacecraft and robotics GNC. He has lead various flight mission analyses at MDA including ISS free-flyer robotic capture, dynamic simulator designs and satellite servicing mission planning. He received his B.Eng. (2001), M.Sc. (2004) in Aerospace Engineering from Carleton University and the University of Toronto Institute for Aerospace Studies, respectively. He received his PhD (2019) in Aerospace Engineering from Carleton University. He is the vice-chair of the AIAA Space Automation and Robotics Technical Committee, and he is currently a Senior GNC Engineer at MDA specialising in machine vision, machine learning, spacecraft GNC and robotics.
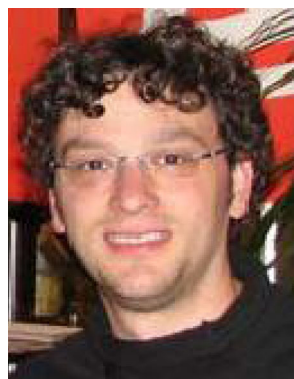
Stéphane Ruel is the Vice President of Development at MDA Neptec Design Group Ltd. in Ottawa Canada. He leads the development of novel 3D computer vision algorithms and sensor software for space applications. Previously, he served as an operations engineer on several Space Vision System Space Shuttle missions. He has a B.Sc. in Computer Engineering and M.Sc. in Aerospace Engineering from Laval University. He is also an alumnus of the International Space University SSP in 2004.

Steve Ulrich is an Associate Professor at the Department of Mechanical and Aerospace Engineering at Carleton University. He received his B.Eng. (2004) and M.Sc. (2006) in Electrical Engineering from the Université de Sherbrooke and his PhD in Aerospace Engineering (2012) from Carleton University. From 2006 to 2008, he was a Spacecraft GNC Research Engineer with NGC Aerospace Ltd, where he led the development of the perturbation estimation system for the ESA PROBA-2 spacecraft.