An Investigation of Gradient as a Feature Cue for Saliency Detection

Document Version
Author Accepted version

An Investigation of Gradient as a Feature Cue for Saliency Detection

Christopher Cooley
Ulster University
Northland Rd
Derry/Londonderry, BT48 7JL
+44(0) 28 7167 5522
cooley-c3@ulster.ac.uk

Sonya Coleman
Ulster University
Northland Rd
Derry/Londonderry, BT48 7JL
+44(0) 28 7167 5030
sa.coleman@ulster.ac.uk

Bryan Gardiner
Ulster University
Northland Rd
Derry/Londonderry, BT48 7JL
+44(0) 28 7167 5081
b.gardiner@ulster.ac.uk

Bryan Scotney
Ulster University
Cromore Road
Coleraine, BT52 1SA
+44(0) 28 7012 4648
bw.scotney@ulster.ac.uk

ABSTRACT

Salient object detection is a prominent research topic, based on a human's ability to selectively process conspicuous objects/regions within a scene. With many low-level features being adopted into saliency models, gradient is often overlooked. We investigate the effectiveness of gradient as a feature, applying and evaluating multiple image gradient operators. Scale is also addressed via the use of different sizes of convolutional masks and by varying the neighbour region to calculate gradient contrast. Finally, we present and evaluate a single scale saliency model with the respective gradient cue from each operator, for the detection of salient objects. Each model is evaluated on the publicly available MSRA10K salient object dataset.

CCS Concepts

Computing Methodologies→ Artificial Intelligence→ Computer Vision→ Computer Vision Problems→ Object Detection

Keywords

Saliency Detection; Gradient Operators; Gradient Feature.

1. INTRODUCTION

The human eye captures an incredible amount of visual information. Correspondingly, a certain characteristic of the visual system, the selective attention mechanism, acts in a filter-like manner allocating attention to regions of interest for further processing. Motivated by this, saliency detection has been widely adapted for multiple problems within computer vision, such as

object classification [1], image segmentation [2] and video compression [3]. Traditionally, saliency detection was defined as estimating areas within images that humans would fixate on [4], but more recently has broadened to include detecting and segmenting eye-catching objects/regions within a scene [5].

The latter definition is the focus of this paper. Salient object detection models can be divided into two categories: top-down models, which are time-consuming, requiring prior knowledge of the scene, and often depend upon extensive training; and bottom-up approaches which are generally fast and driven by low-level stimulus. Within bottom-up saliency detection, a wide variety of features have been implemented, namely colour, spatial distribution, texture and centre bias. However, one feature cue that is under-utilised is gradient. Gradient information, known for its use in edge detection, is often overlooked as an effective feature cue within saliency algorithms. Nevertheless, object gradient information is often distorted amidst the calculation of other feature cues, but can play a vital role in saliency detection. Salient object detection can be likened to a binary classification problem: does the current pixel form part of the salient object? Therefore, each pixel is either categorised as salient or non-salient. Gradient information is useful in this process as it preserves the contour of the object, and suppresses background noise due to the lack of intensity changes. This paper investigates the use of gradient magnitude as a feature cue in detecting salient objects, comparing and evaluating multiple operators. Gradient contrast is calculated across a neighbouring region. We investigate the optimal neighbourhood region for the feature calculation, by varying the neighbourhood size from 3×3 to 9×9. This paper focuses on the low-level mathematical principles involved in saliency detection rather than shallow or deep learning.

The paper is organised as follows. Section 2 describes related work within salient object detection. Section 3 and Section 4 give an overview of the operators and feature cues forming a saliency model respectively. Section 5 delineates the evaluation and comparison of each implemented feature cue. Conclusions are drawn and future work outlined in Section 6.

2. RELATED WORK

Within saliency, a large number of feature cues have been adopted. Itti et al. [5], proposed what is known to be the first saliency model.

Their approach consisted of colour, orientation and intensity calculated at multiple scales, using centre surround operations. Normalisation and linear summation were subsequently performed to produce the final saliency map. Many algorithms have taken inspiration from, and based their algorithm on, Itti's model. Achanta et al. considered colour and luminance as feature cues, computing saliency using image frequency information [6]. A global contrast saliency algorithm is outlined in [7] where Cheng et al., evaluate the contrast of the current region with the remainder of the image domain.

A number of different methods have considered scale as a means of improving salient object detection. In [8], a multi-scale method is presented, which extracts three image layers from the input, each layer containing different levels of details. Feature cues are computed on each layer, and fused using a graph model to produce a final output saliency map. Within this paper, the main investigation is the effectiveness of gradient as a feature cue for saliency. Anwar et al., describe their method of non-linearly integrating colour and gradient [9]. Saliency is computed using average patch dissimilarity and five different patch sizes for scaling. In [10], the input image was subsampled using a Gaussian pyramid to produce three sub-images. Morphological gradient is implemented in this approach, which is the difference of erosion and dilation operations. The work in [11] employs gradient contrast using the Sobel operator across a $[3 \times 3]$ neighbouring region.

3. GRADIENT OPERATORS

The key research question in this section is, does the specific gradient operator have a direct link to the success of gradient as a feature cue? The baseline operator chosen in this paper for benchmarking gradient is the Sobel [16]; one of the most widely used gradient operators within image processing. Gradient is computed by convolving an image with the x and y derivative masks. In the gradient calculation, the Sobel operator employs 3×3 convolution masks. In previous work, families of multiscale derivative operators [12, 13, 14] were developed, and in this work, we utilise the 5x5 Linear Gaussian [12], 5x5 Bilinear Gaussian [13] and a 5x5 Near-Circular [14], to answer the proposed research question. These operators are based on the finite element framework. The design procedure used is based on the use of a virtual mesh. For any node centred at i, within the virtual mesh, with coordinates (x_i, y_i), a piecewise linear or bilinear basis function is associated with i, depending on the operator employed, having the following properties:

$$(x_j, y_j) = \begin{cases} 1 \; if \; i = j \\ 0 \; if \; i \neq j \end{cases} i = 1,..,N; j = 1,.,N \quad (1)$$

where (x_j, y_j) are the coordinates of node j. Smoothing is also built into each operator using a Gaussian function ξ_i^g around node i defined within a neighbourhood. The Near-Circular operator [14] filter values are presented in Figure 1.

$$X = \begin{bmatrix} 0 & -0.0199 & 0 & 0.0199 & 0 \\ -0.0409 & -0.0340 & 0 & 0.0340 & 0.0409 \\ -0.0398 & -0.0916 & 0 & 0.0916 & 0.0398 \\ -0.0409 & -0.0340 & 0 & 0.0340 & 0.0409 \\ 0 & 0.0199 & 0 & 0.0199 & 0 \end{bmatrix}$$

$$Y = \begin{bmatrix} 0 & 0.0409 & 0.0398 & 0.0409 & 0 \\ 0.0199 & 0.0340 & 0.0916 & 0.0340 & 0.0199 \\ 0 & 0 & 0 & 0 & 0 \\ -0.0199 & -0.0340 & -0.0916 & -0.0340 & -0.0199 \\ 0 & -0.0409 & -0.0398 & -0.0409 & 0 \end{bmatrix}$$

Figure 1. x and y masks of Near-Circular gradient operator

An example of the gradient magnitude output from each operator is presented in Figure 2.
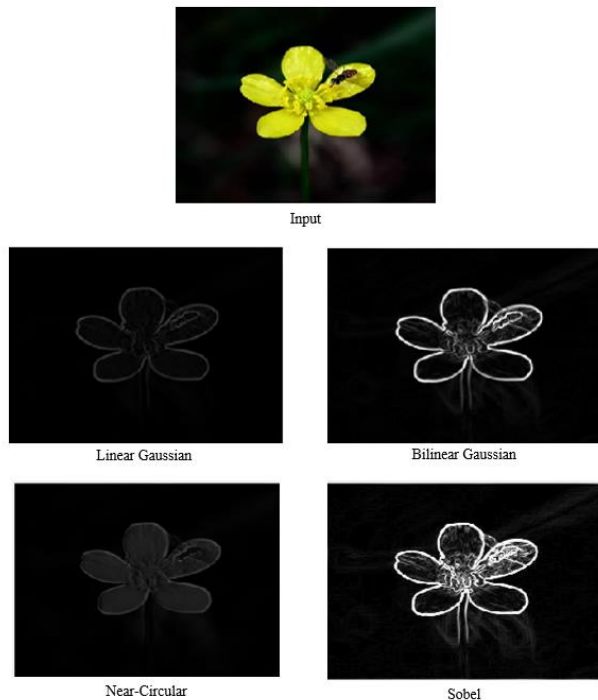

Figure 2. A visual comparison of the gradient magnitude produced by each operator

4. SALIENCY MODEL

Low-level features are an essential part of saliency detection; hence a number of different features derived from image gradient are assessed, namely gradient magnitude and gradient contrast. Finally, the gradient output from each operator is combined with colour and evaluated as part of a single scale saliency model.

4.1 Gradient Magnitude

Image gradient defines the change in intensity within an image. Gradient direction specifies the direction in which the intensity is changing, whereas, the magnitude describes how quickly that intensity is changing. Gradient magnitude (GM) is calculated as the root mean square of the gradient responses as follows:

$$GM = \sqrt{G_y{}^2 + G_x{}^2} \quad (2)$$

Works on saliency [10, 11] have adopted gradient contrast as a feature cue, often calculated from the magnitude, however observation of gradient magnitude has shown this could potentially be incorporated as an independent saliency feature cue. Figure 2 illustrates the visual difference for the gradient magnitude produced by each operator.

4.2 Gradient Contrast

Gradient contrast (GC) can play an integral part in approximating the salient object within a scene. While suppressing non-important information e.g. the background of an image, it also helps to preserve the contours of the salient object. Gradient contrast is computed per pixel over a neighbourhood region. The scale of the

Table 1. Gradient operator and feature cue results

Feature	Linear Gaussian			Bilinear			Near-circular			Sobel		
	Avg. Accuracy	Max. Accuracy	Runtime	Avg. Accuracy	Max. Accuracy	Runtime	Avg. Accuracy	Max. Accuracy	Runtime	Avg. Accuracy	Max. Accuracy	Runtime
CONTRAST												
3x3 Region	80.30%	94.80%	0.0807 secs.	80.40%	94.80%	0.0796 secs.	80.30%	94.20%	0.0824 secs.	80.50%	94.80%	0.0802 secs.
5x5 Region	80.50%	94.70%	0.0823 secs.	80.60%	94.70%	0.0829 secs.	80.60%	95%	0.0840 secs.	80.80%	94.70%	0.0895 secs.
7x7 Region	80.70%	94.80%	0.1080 secs.	80.80%	94.70%	0.0914 secs.	80.70%	95.10%	0.0910 secs.	80.90%	94.70%	0.0983 secs.
9x9 Region	80.80%	94.90%	0.1163 secs	80.90%	94.70%	0.1058 secs.	80.80%	95.30%	0.1411 secs.	80.80%	94.80%	0.1111 secs.
MAGNITUDE	81.90%	96.60%	0.0112 secs.	80.90%	95.50%	0.0110 secs.	83.70%	98.30%	0.0112 secs.	81%	96.60%	0.0106 secs.

neighbourhood region is varied to determine its impact on the feature strength. Gradient contrast (GC) is defined as:

$$GC_{(i,j)} = \sum_{i=1}^{N} \sum_{j=1}^{N} \left\| g_{(i,j)} - g_{(n)} \right\| \qquad (3)$$

where N is the total number of pixels within a neighbourhood and $g_{(i,j)}$ and $g_{(n)}$ are the respective gradient values for pixel i, j, and the sum of gradient values across a small region. The neighbourhood sizes ranged from 3x3 to 9x9 corresponding to [10, 11].

4.3 Feature Fusion

Salient object detection models often consist of two or more feature cues, as one feature often struggles when processing dynamic scenes with varying colours, objects, backgrounds and lighting. When analysing a scene, the human eye is more sensitive to colour, in comparison to other visual features [17], therefore gradient features are combined with colour to complete the saliency model. Feature maps from local colour contrast, calculated per pixel over a defined neighbourhood region, tends to be noisy and mainly highlights the edges of the salient area/object, whereas global contrast computes the contrast of a pixel/region in relation to the colour value of all the remaining pixels/regions within an image. The calculation to compute the colour contrast (CC) of a pixel against the average colour value of the remaining pixels using the Euclidean distance in $L^*a^*b^*$ colour space, is defined as:

$$CC_{(i)} = \sum_{j=1}^{N} \sqrt{L_{(i)} - L_{(j)})^2 + (a_{(i)} - a_{(j)})^2 + (b_{(i)} - b_{(j)})^2} \quad (4)$$

where N is the number of pixels in the image and the LAB colour values at pixels i and j are defined as $L_{(i)}$, $L_{(j)}$, $a_{(i)}$, $a_{(j)}$, $b_{(i)}$ and $b_{(j)}$ respectively. The resultant saliency map FS of the final model can be obtained by normalising the respective feature cues in the range [0, 1], before performing linear fusion:

$$FS_{(i,j)} = \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} (\rho \times CC_{(i,j)}) + (\tau \times GC_{(i,j)}) \qquad (5)$$

where ρ and τ are weight coefficients set to 0.7 and 0.6 respectively. CC and GC are colour contrast and gradient contrast values at pixel i, j. Figure 3 visualises the pipeline of the model, showing outputs for each stage, with feature cues being calculated from the input image. These features are fused, then binarised using a threshold for evaluation.

5. EXPERIMENTAL RESULTS

To assess the effectiveness of each gradient operator and the respective feature cues, we evaluate the output using an accuracy measure to compare the acquired output with the ground-truth image. Computation time is also considered to assess the efficiency of each feature, the operator size and the final saliency approach.

5.1 Evaluation Metrics

For a quantitative evaluation, we selected an algorithmic accuracy metric A, comparing the saliency map FS with its associated ground-truth mask M. Thus, this enables individual features to be evaluated, as well as combined features amounting to saliency models. As with all state-of-the-art saliency approaches, the saliency map FS is binarised using the threshold $[0, 255]$. At each threshold, the segmented saliency map is compared with the ground-truth binary mask and the accuracy recorded using the following equation:

$$A = \frac{(tp + tn)}{(tp + tn + fp + fn)} \qquad (6)$$

where tp, tn, fp and fn are defined as true positives, true negatives, false positives and false negatives respectively.

For performance validation, individual features as well as feature combinations are evaluated using the publicly available MSRA10K salient object dataset [15]. The average accuracy, maximum accuracy and runtime are recorded to quantitatively evaluate each operator and the final saliency model.

5.2 Results

The results for each individual feature, determined using the described gradient operators, are presented in Table 1. Besides this, a visual comparison of the gradient contrast output can be seen in Figure 4. We assessed the two features, outlined in Table 1, across 100 images from the MSRA10K dataset. These images were manually selected to test each feature under varying conditions such as type, position and size of object, background colours and complexity. Each operator scored a comparable average accuracy for gradient contrast. Generally, as the neighbourhood contrast region was increased, so did the average accuracy. Gradient magnitude has separated the operators in terms of their scored accuracies. Bilinear-Gaussian and Sobel achieved averages of 80.9% and 81% respectively, however, the Sobel scored a higher maximum accuracy of 96.6% on par with the Linear-Gaussian operator. The Near-Circular operator recorded the best accuracies with an average of 83.7% and a maximum of 98.3%. With regards to runtime, the operators were able to compute each feature with minute differences. Visually, it can be observed that the model most suited for saliency detection is the Near-Circular (see Figure 4(c)). It successfully suppressed more background noise than any of the other operators, whilst detecting the salient object's edges. Figure 5 shows the final saliency maps of the Sobel and Near-Circular operators. As seen, the Sobel's output produces noisy edges, whereas the Near-Circular produces smooth edges with finer detail. The resultant model, gradient contrast over a neighbourhood of 9x9 combined with colour, achieved an average accuracy of 89.3%,
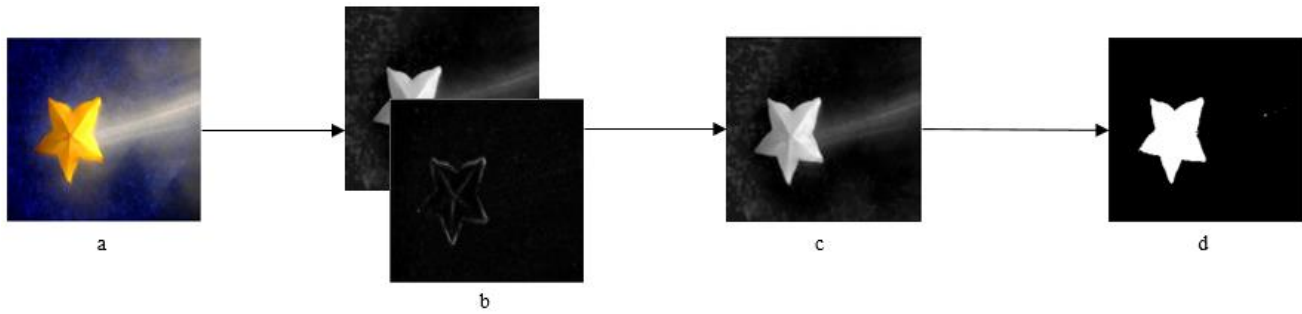
Figure 3. Overview of saliency model. a) is the input image. b) feature cues output maps. c) fused saliency map and d) is the binarised saliency image for comparison with ground truth image.

with the Near-Circular operator. The Sobel's average accuracy was 84.8%, a shortfall of 4.5% compared to the Near-Circular saliency model. Correspondingly, the main gain from the presented results in Table 2 comes from the use of the 5x5 Near-Circular gradient operator. While these results are encouraging, the combined saliency models suffer from runtime deficiencies, due to the incorporation of the global colour contrast feature cue.

Table 2. Gradient operator results combined with colour

Operator	Avg. Acc.	Max Acc.	Runtime
Linear Gaussian	88.60%	99.30%	15.5478 secs.
Bilinear Gaussian	85%	98.90%	16.5775 secs.
Near-Circular	89.30%	99.70%	15.2371 secs.
Sobel	84.80%	99.50%	14.8950 secs.

6. CONCLUSION

This paper investigated gradient information as a feature for use in salient object detection. The outlined features include gradient magnitude and gradient contrast. We implemented and compared multiple gradient operators and found that operators do have an impact on the success of gradient as a feature, with the Near-Circular being best suited for proposing gradient as a saliency feature. After evaluation, we incorporated the Near-Circular derivative operator as part of the proposed model. The presented algorithm combines gradient contrast with colour. When evaluated on the MSRA10K dataset, and compared with the proposed model using a Sobel operator, accuracy increases of 4.5% were achieved using the Near-Circular gradient operator.

Further investigation is required to improve the computational efficiency of the proposed algorithm as highlighted by the obtained runtimes. Other feature cues such as depth, texture and motion will be considered as means of improving the robustness of the proposed approach. Multiple scale feature extraction for saliency is another avenue we aim to explore.

7. REFERENCES

[1] Cooley, C, Coleman, SA, Gardiner, B, and Scotney BW, "Saliency Detection and Object Classification." Irish Machine Vision and Image Processing Conference, 2017.

[2] C. Qin, G. Zhang, Y. Zhou, W. Tao, and Z. Cao, "Integration of the saliency-based seed extraction and random walks for image segmentation," *Neurocomputing*, vol. 129, pp. 378–391, 2014.

[3] C. Qin, G. Zhang, Y. Zhou, W. Tao, and Z. Cao, "Integration of the saliency-based seed extraction and random walks for image segmentation," *Neurocomputing*, vol. 129, pp. 378–391, 2014.

[4] A. Borji and L. Itti, "State-of-the-Art in Visual Attention Modeling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 185–207, 2013.

[5] L. Itti, C. Koch, and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1998, vol. 20, no. 11, pp. 1254–1259.

[6] R. Achanta, S. Hemami, F. Estrada, S. Sabine, and D. L. Epfl, "Frequency-tuned Salient Region Detection," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1597–1604.

[7] M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S. Hu, "Global Contrast Based Salient Region Detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, 2015.

[8] Q. Yan, L. Xu, J. Shi, and J. Jia, "Hierarchical Saliency Detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[9] S. Anwar, Q. Zhaot, and M. F. Manzoor, "Saliency Detection using Parallel Non-Linear Integration of Color and Gradient using Covariances," in *International Conference on the Innovative Computing Technology (INTECH 2014)*, 2014, pp. 197–201.

[10] W. Li, J. Qui, and X. Li, "Visual Saliency Detection based on Gradient Contrast and Color Complexity," in *International Conference on Internet Multimedia Computing and Service*, 2015, pp. 1–5.

[11] Cooley, C, Coleman, SA, Gardiner, B, Scotney, BW, "Salient Obstacle Avoidance for Robotic Systems." Irish Machine Vision and Image Processing Conference, 2018.

[12] S. A. Coleman and B. W. Scotney, "IMAGE FEATURE DETECTION ON CONTENT-BASED MESHES," in IEEE International Conference on Image Processing (ICIP 2002), 2002, pp. 844–847.

[13] B. W. Scotney, S. A. Coleman, M. G. Herron, and S. Engineering, "Device Space Design for Efficient Scale-Space Edge Detection," in International Conference on Computational Science (ICCS 2002), 2002, pp. 1077–1086.

[14] B. W. Scotney, S. A. Coleman, and M. G. Hewon, "A SYSTEMATIC DESIGN PROCEDURE FOR SCALABLE NEAR-CIRCULAR GAUSSIAN OPERATORS," in IEEE International Conference on Image Processing (ICIP 2001), 2001, pp. 844–847.

[15] M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S. Hu, "Global Contrast Based Salient Region Detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 3, pp. 569–582, 2015.

[16] Sobel, I., "Camera Models and Machine Perception", Standford AI Memo, No. 121, May 1970.

[17] S. Wan, P. Jin, and L. Yue, "An Approach for Image Retrieval Based on Visual Saliency," in *2009 International Conference on Image Analysis and Signal Processing*, 2009, pp. 172–175
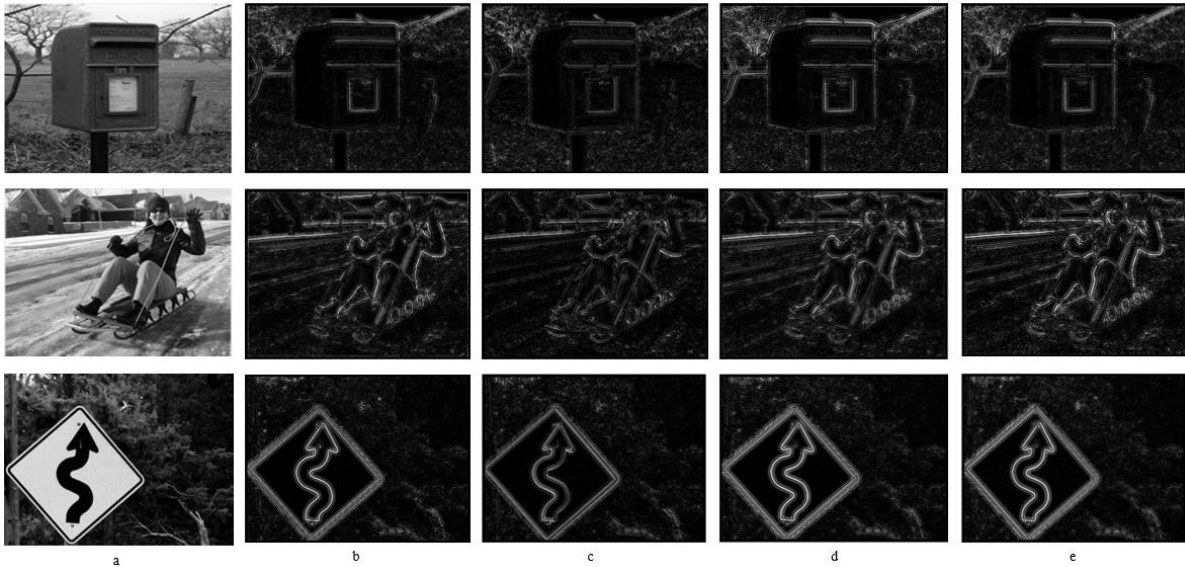
Figure 4. Visual comparison of gradient contrast feature maps from each operator. a) Grayscale image. b) Sobel operator output. c) Near-Circular operator output. d) Bi-linear operator output and e) is the Linear Gaussian operator output.
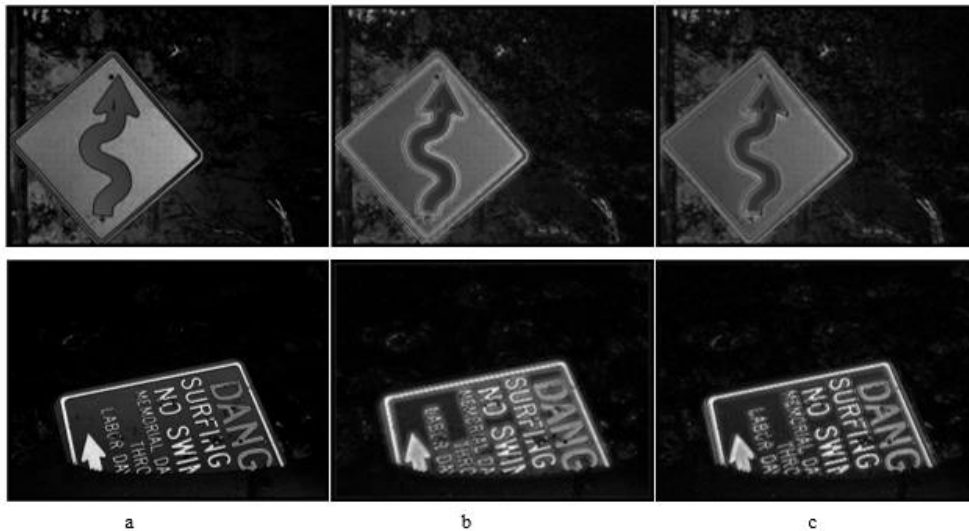

Figure 5. Visual comparison of saliency maps with gradient. a) Colour saliency map. b) Final saliency map with Sobel gradient. c) Final saliency map with Near-Circular gradient.