Local Background Enclosure for RGB-D Salient Object Detection - Supplementary Results

David Feng, Nick Barnes, Shaodi You, Chris McCarthy

7. Introduction

The purpose of this supplementary material is to examine in detail the contributions of our proposed Local Background Enclosure (LBE) feature. A comparison of LBE with the contrast based depth features used in state-of-the-art salient object detection systems is presented. The LBE feature is compared with the raw depth features ACSD [1], DC [3] and a signed version of DC denoted SDC on the RGBD1000 [2] and NJUDS2000 [1] datasets. We then qualitatively assess the use of the prior application and Grabcut refinement stages of our saliency system to enhance the LBE saliency map.

Please find attached the outputs of the tested salient object detection systems on each dataset. The saliency maps for ACSD [1], LMH [2], and GP [3] were obtained by running code from the author's websites. Note that the saliency maps have been resized to conserve space.

8. Analysis of Local Background Enclosure Feature

Figure 5 of the manuscript shows a graph illustrating that the LBE feature gives superior performance compared to existing state-of-the-art contrast based depth features. Here, we provide a detailed analysis on the nature of the improvements gained by our depth feature, specifically the ways in which false negatives and false positives produced by contrast based methods are reduced when using LBE. We then examine failure cases for our feature. Note that since we are comparing depth features, no colour information is used at this stage. Additionally, no priors are applied to the generated saliency maps. The contrast based depth features that we compare with are:

- depth contrast (DC), the global depth contrast term from a state-of-the-art RGBD salient object detection system [3]:

$$S_{DC}\left(r_i\right) = \sum_{j \neq i} A\left(r_j\right) C\left(r_i, r_j\right) \tag{1}$$

  where i, j are region indices, $A\left(r_i\right)$ gives the area of region r_i, and

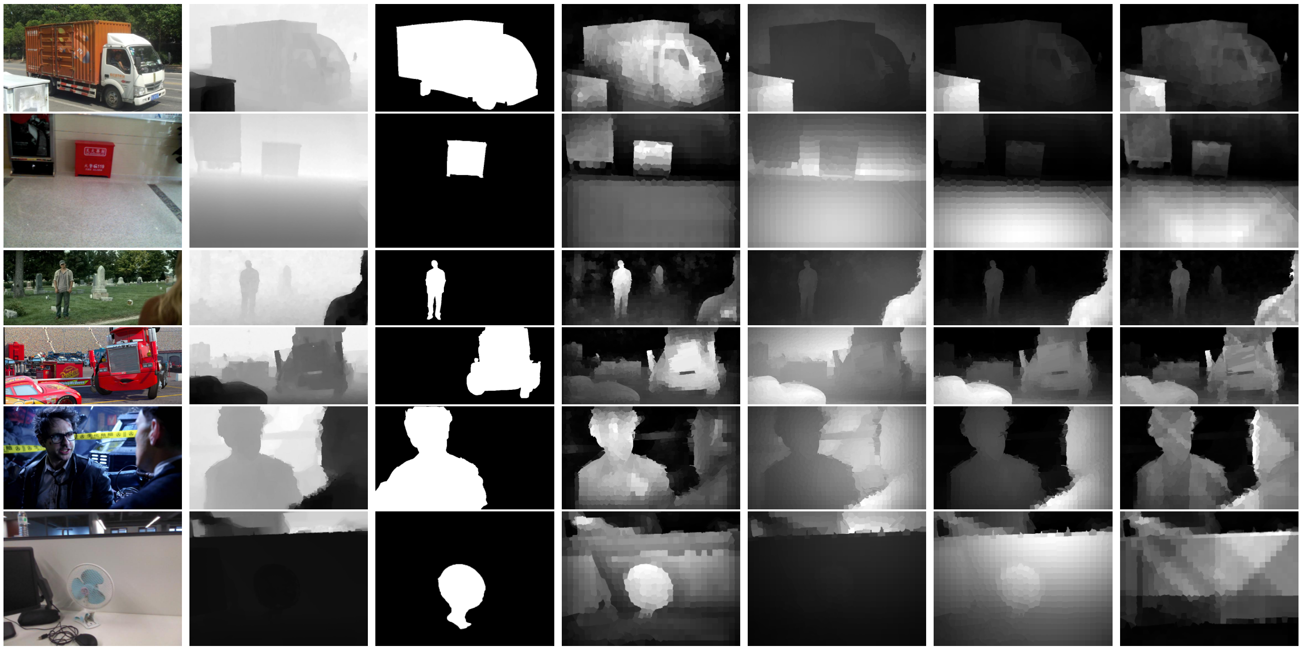
$$C\left(r_i, r_j\right) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{2\sigma_{\mathbf{x}}^2}\right)\|\mathbf{d}_i - \mathbf{d}_j\|_2 \tag{2}$$

  where $\mathbf{d}_i$ denotes the mean depth of region i, $\mathbf{x_i}$ is the centroid of region r_i, and $\sigma_{\mathbf{x}}$ is the standard deviation of the distance between two region centroids.

- signed depth contrast (SDC), which is similar to DC but makes the assumption that objects are closer than the background by ignoring closer regions than the candidate region during contrast computation.

- ACSD [1], a state-of-the-art depth feature.

We use a custom implementation of DC [3] and SDC, while code for ACSD [1] was obtained from the author's website.

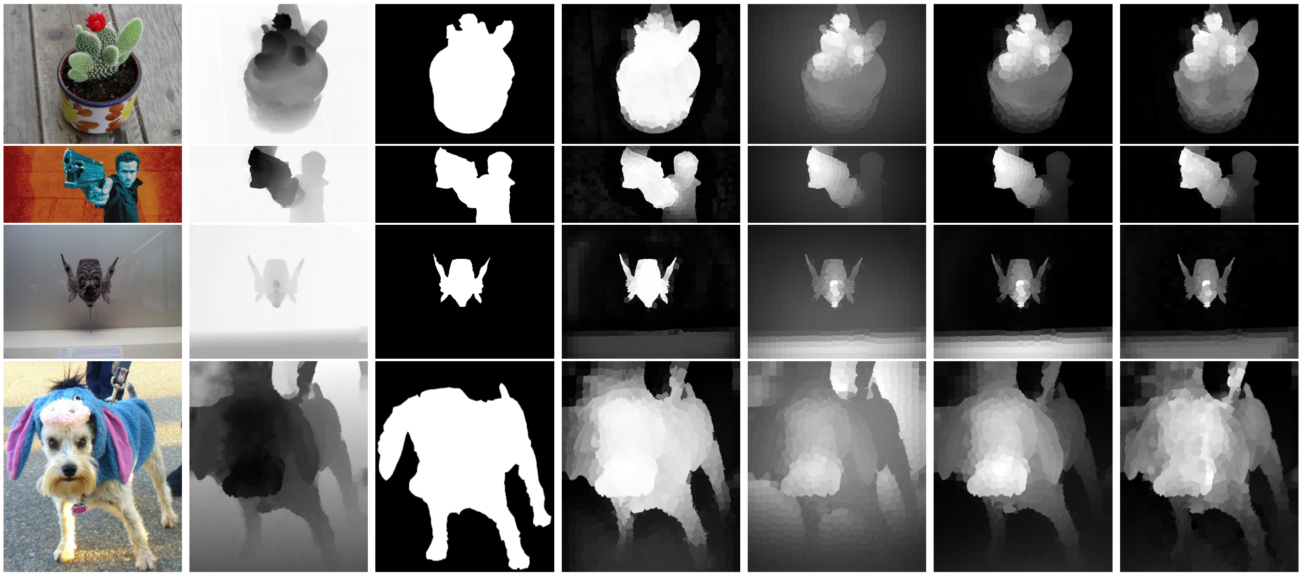


(a) Colour (b) Depth (c) Ground Truth (d) LBE (Ours) (e) DC [3] (f) SDC [3] (g) ACSD [1]

Figure 9. Examples of performance on images where the foreground has low depth contrast compared to the background using the raw depth saliency features of LBE, state-of-the-art methods ACSD[1], DC[3], and SDC[3] which is a signed variant of depth contrast. The depth contrast based methods perform poorly, while our method identifies the salient objects.

8.1. Reducing False Negatives: Low Contrast Foreground

A common pitfall of existing contrast based depth features is sensitivity to depth difference magnitude. These features produce false negatives when the object has lower contrast than the background. Figure 9 shows example scenes where depth contrast based methods incorrectly assign a low saliency score to the salient object because it has relatively low depth contrast. For example in the first row, the white box in the bottom right corner of the image has the greatest depth difference with the surroundings. This object is not salient according to the ground truth, however existing methods identify the box as the salient object. In these cases our depth feature correctly identifies the salient object based on its pop-out structure as measured using local background enclosure.

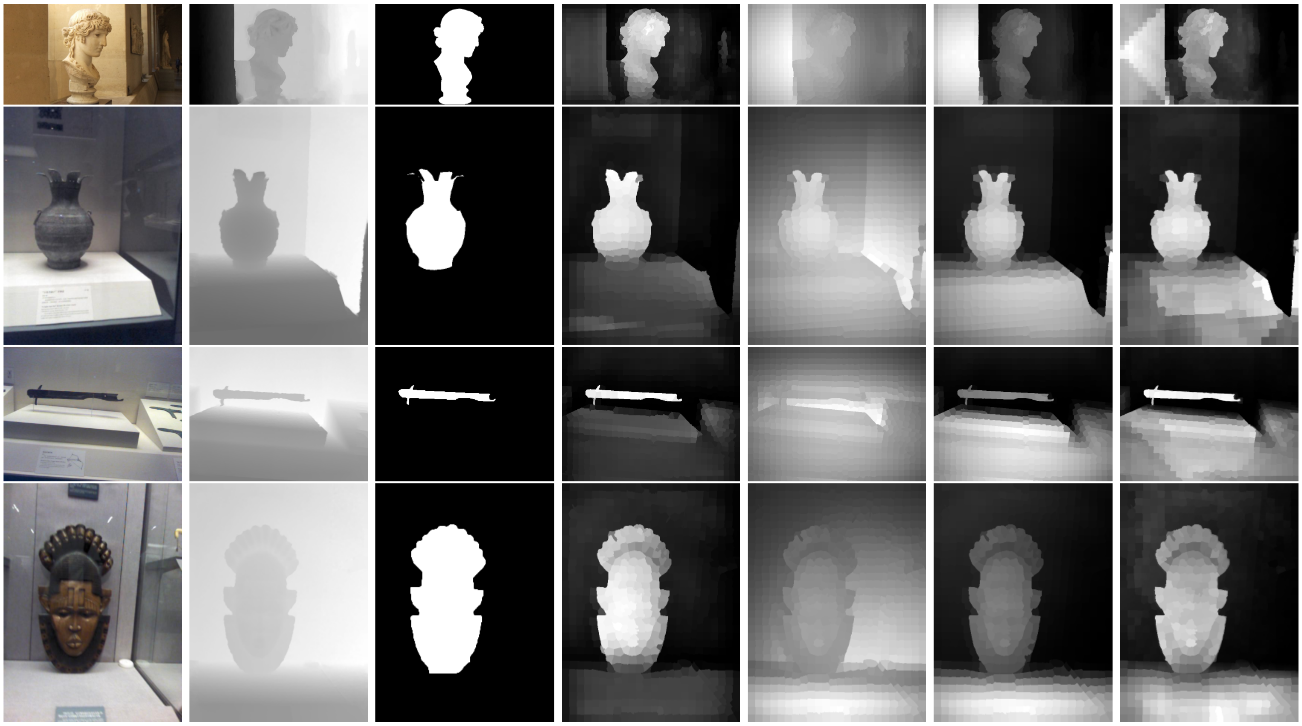


(a) Colour (b) Depth (c) Ground Truth (d) LBE (Ours) (e) DC [3] (f) SDC [3] (g) ACSD [1]

Figure 10. Examples of objects containing different depth contrast values, producing a non-uniform saliency response from the depth contrast based methods ACSD [1], DC [3] and SDC [3] which is a signed version of depth contrast. Our method LBE is able to obtain a uniform saliency response across an object when the object pops out from the surroundings.

8.2. Reducing False Negatives: Objects with Large Depth Range

Salient objects that contain a relatively large range of depth values tend to have a high variation in the saliency response across the object, as shown in Figure 10. For example, in the first row, there is a significant difference between the saliency values of the top of the plant and the pot for contrast based features. For these images and others like them, our feature produces a more uniform response across objects which have a pop-out shape.

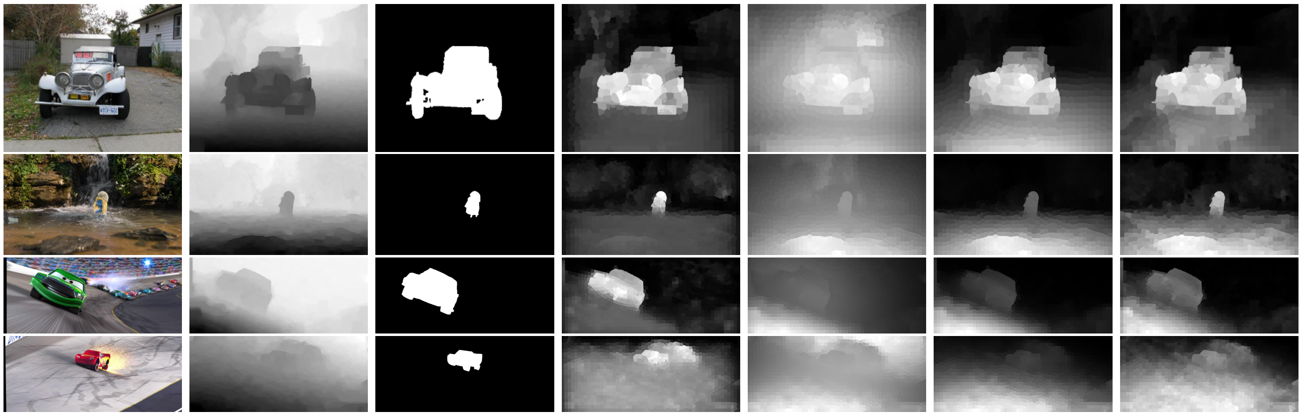


(a) Colour (b) Depth (c) Ground Truth (d) LBE (Ours) (e) DC [3] (f) SDC [3] (g) ACSD [1]

Figure 11. Examples of performance on images where the background exhibits high depth contrast using the raw depth saliency features of LBE, state-of-the-art methods ACSD[1], DC [3], and SDC [3] which is a signed variant of depth contrast. In this type of situation using LBE to measure pop-out structure more robustly identifies foreground regions compared to measuring depth contrast.

8.3. Reducing False Positives: High Contrast Background

Background structure adjacent to a large depth drop-off is a common source of false positives for depth contrast methods, producing high depth contrast values in the background region, as shown in Figure 11. For example, in the first row, the wall on the left is assigned a relatively high saliency by depth contrast based features because it has large depth difference with the adjacent region. Since background structure usually does not have a pop-out shape, our feature is able to suppress these regions.

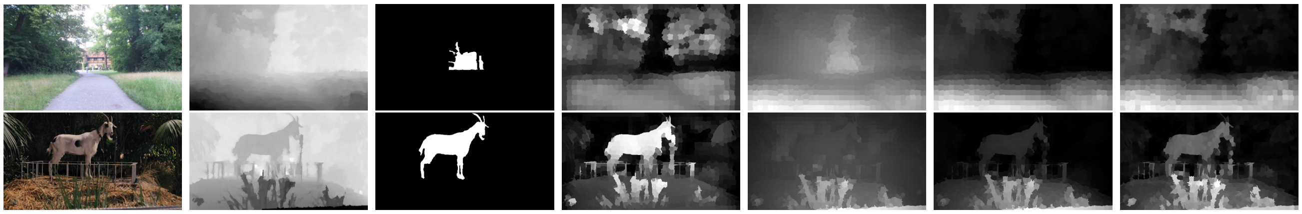


(a) Colour (b) Depth (c) Ground Truth (d) LBE (Ours) (e) DC [3] (f) SDC [3] (g) ACSD [1]

Figure 12. Examples of saliency output on scenes with large angled planar surfaces, which have a high depth contrast, using the raw depth saliency features LBE, ACSD[1], DC [3], and SDC [3] which is a signed variant of depth contrast. Planar surfaces are generally unavoidable, and produce a high saliency response for depth contrast based methods. These types of surfaces do not exhibit pop-out structure however, and are therefore assigned a low saliency score by our depth feature.

8.4. Reducing False Positives: Angled Planar Surfaces

Flat surfaces that are angled towards the camera are one particularly common type of background structure that exhibits depth contrast. These surfaces frequently produce false positives for depth contrast methods, since points along the surface can have a wide range of depth values. Some examples are shown in Figure 12. Our method significantly reduces the false positives caused by this type of structure, since depth difference across the surface is ignored, and since large planar surfaces tend to have a low background enclosure.



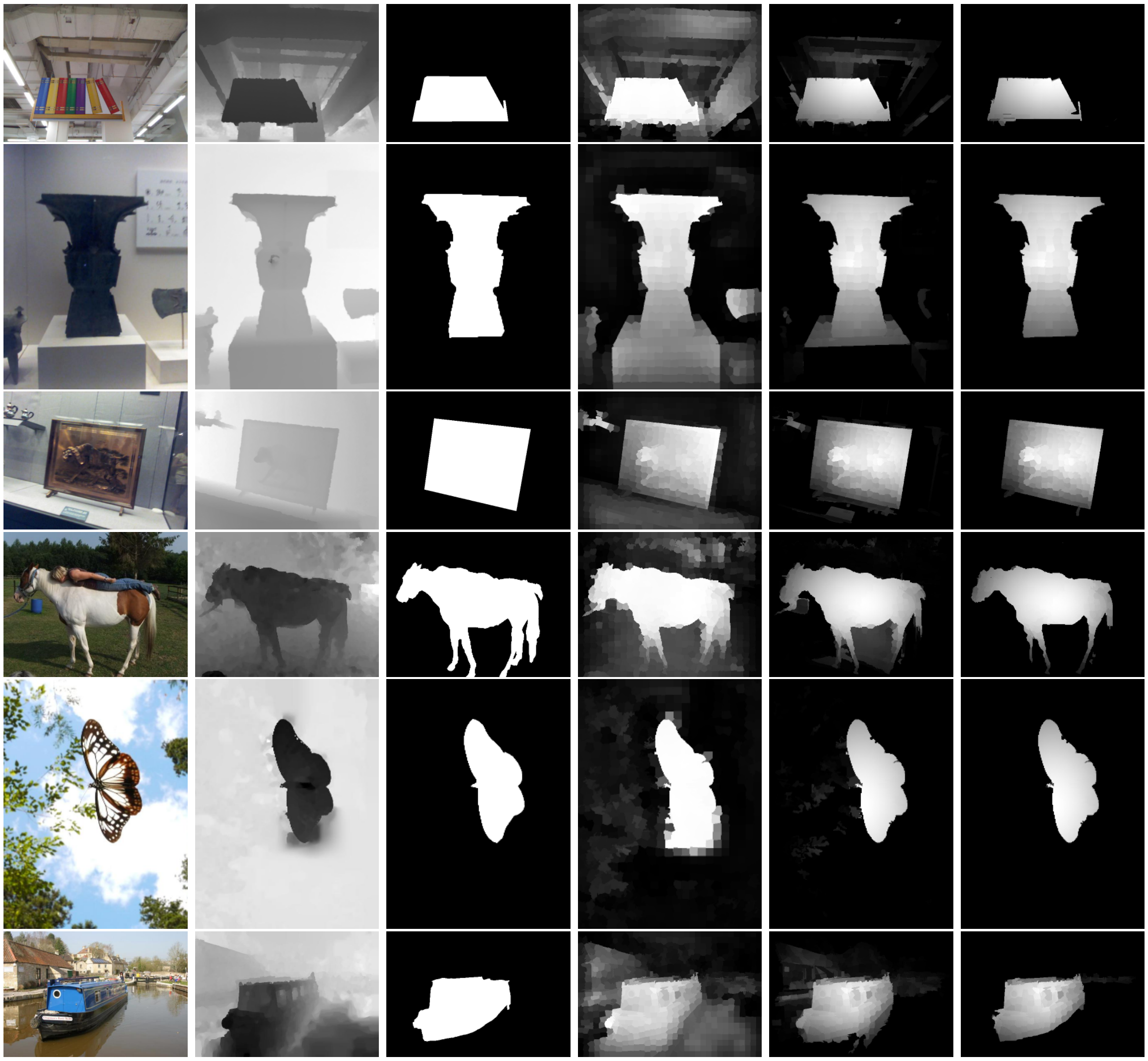
(a) Colour (b) Depth (c) Ground Truth (d) LBE (Ours) (e) DC [3] (f) SDC [3] (g) ACSD [1]

Figure 13. Examples of failure cases, showing saliency output from the raw depth saliency features LBE, ACSD [2], DC [3], and SDC [3], which is a signed version of depth contrast. In the first row, the object is surrounded in all directions by closer surfaces. This is a rare occurrence, as salient objects tend to be in front of their surroundings. The second row shows a situation where a background region has strong pop-out structure. This leads to false positives for all methods, and our method produces the best result in this case.

8.5. Failure Cases

Since our method measures pop-out structure, it does not produce good results when the salient object is surrounded in all directions by background with lower depth. An example is shown in Figure 13. Note that this is a rare occurrence, and the other depth saliency methods with the exception of DC also produce poor results in this case. In these situations, it is questionable whether the object can be considered to be salient. Note that DC produces the best results in this image because it does not assume that salient objects are in front of the background, however this leads to poor performance on the datasets.

Occasionally the background can have some degree of pop-out structure, such as the grass in the second row of Figure 13, leading to false positives from our feature. However the response is generally weaker than for the salient object, and it is a less common occurrence than the background having high depth contrast. Our depth feature still produces the best overall result compared to contrast based depth features, which are also affected by this problem.

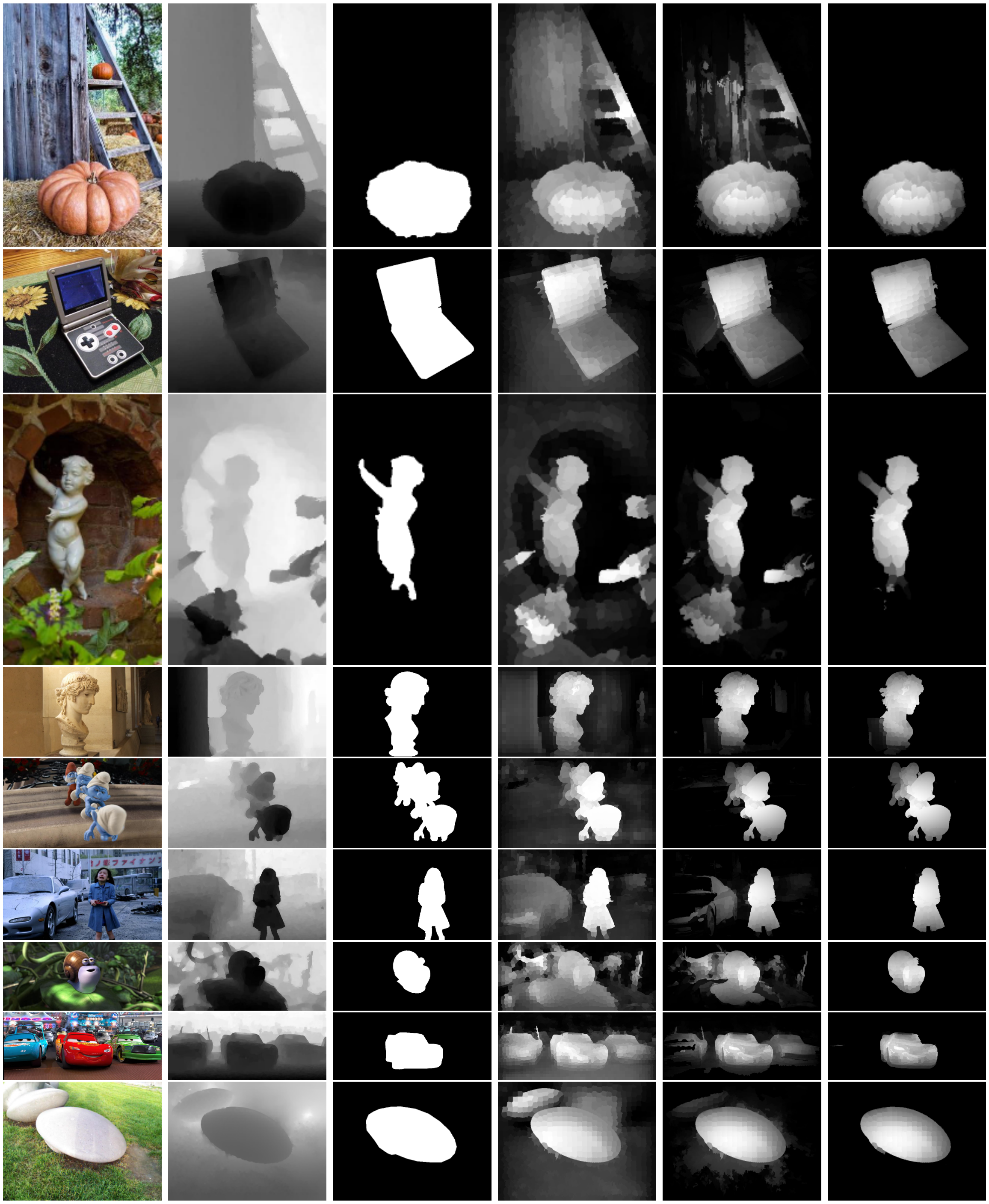


(a) RGB (b) Depth (c) Ground Truth (d) LBE (e) LBE+P (f) LBE+P+G

Figure 14. Output of the different stages of our salient object detection system. LBE denotes our proposed depth feature, LBE+P shows the result of depth, spatial, and background prior application, and LBE+P+G illustrates the final output of our salient object detection system after applying Grabcut refinement.

9. Saliency Detection System: LBE, Priors, and Grabcut Outputs

Figures 6 and 7 of the manuscript show the quantitative contributions of each of the three stages of our saliency detection system. In this section, Figures 14 and 15 will give examples showing the output from each stage of our system. First the LBE feature is applied to the depth image, identifying the salient object and sometimes producing a non-zero response for background regions with pop-out structure. These background regions are trimmed based on depth, spatial position and colour during the prior application stage. The resulting map is further pruned in the Grabcut refinement stage.



(a) RGB (b) Depth (c) Ground Truth (d) LBE (e) LBE+P (f) LBE+P+G

Figure 15. Output of the different stages of our salient object detection system. LBE denotes our proposed depth feature, LBE+P shows the result of depth, spatial, and background prior application, and LBE+P+G illustrates the final output of our salient object detection system after applying Grabcut refinement.

10. Saliency Detection System: Comparison with Other Methods

Figures 6 and 7 of the manuscript show that our saliency system outperforms all other state-of-the-art RGB-D salient object detection systems. In addition to the highest F-score, our method achieves the highest recall on both datasets, reflecting the fact that our depth feature correctly identifies a larger portion of the foreground compared to contrast based methods. From Figure 6b we see that our method has the highest PR curve. Figure 7b shows that our system has high precision up to around 0.65 recall, and is significantly in front in the region of high precision. From the PR curve in Figure 6b, we see that the RGB methods perform significantly worse than most depth-aware methods.

Figure 8 of the manuscript gives a comparison of the saliency maps produced by our method and by existing state-of-the-art methods. This figure is reproduced below in Figure 16 for the reader's convenience. Our method is able to reliably identify the salient object, unlike the contrast based methods which are sensitive to scene arrangement (e.g. fourth row from the bottom). This shows that the selection of an appropriate basic depth feature is crucial, since it is difficult for a system to compensate if the salient object is missed or if too much background is accepted in the early stages of computation. For example, while our prior and Grabcut application stages can refine results, the quality of the final saliency map is highly dependent on the ability of the LBE depth feature to accurately identify salient objects.

References

[1] R. Ju, L. Ge, W. Geng, T. Ren, and G. Wu. Depth saliency based on anisotropic center-surround difference. *ICIP. IEEE*, 2014. 1. 1, 2, 3, 4, 5, 6, 10

[2] H. Peng, B. Li, W. Xiong, W. Hu, and R. Ji. Rgbd salient object detection: a benchmark and algorithms. In *Computer VisionECCV 2014*, pages 92–109. Springer, 2014. 3. 1, 6, 10

[3] J. Ren, X. Gong, L. Yu, W. Zhou, and M. Yang. Exploiting Global Priors for RGB-D Saliency Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 25–32, 2015. 1, 2, 3, 4, 5, 6, 10

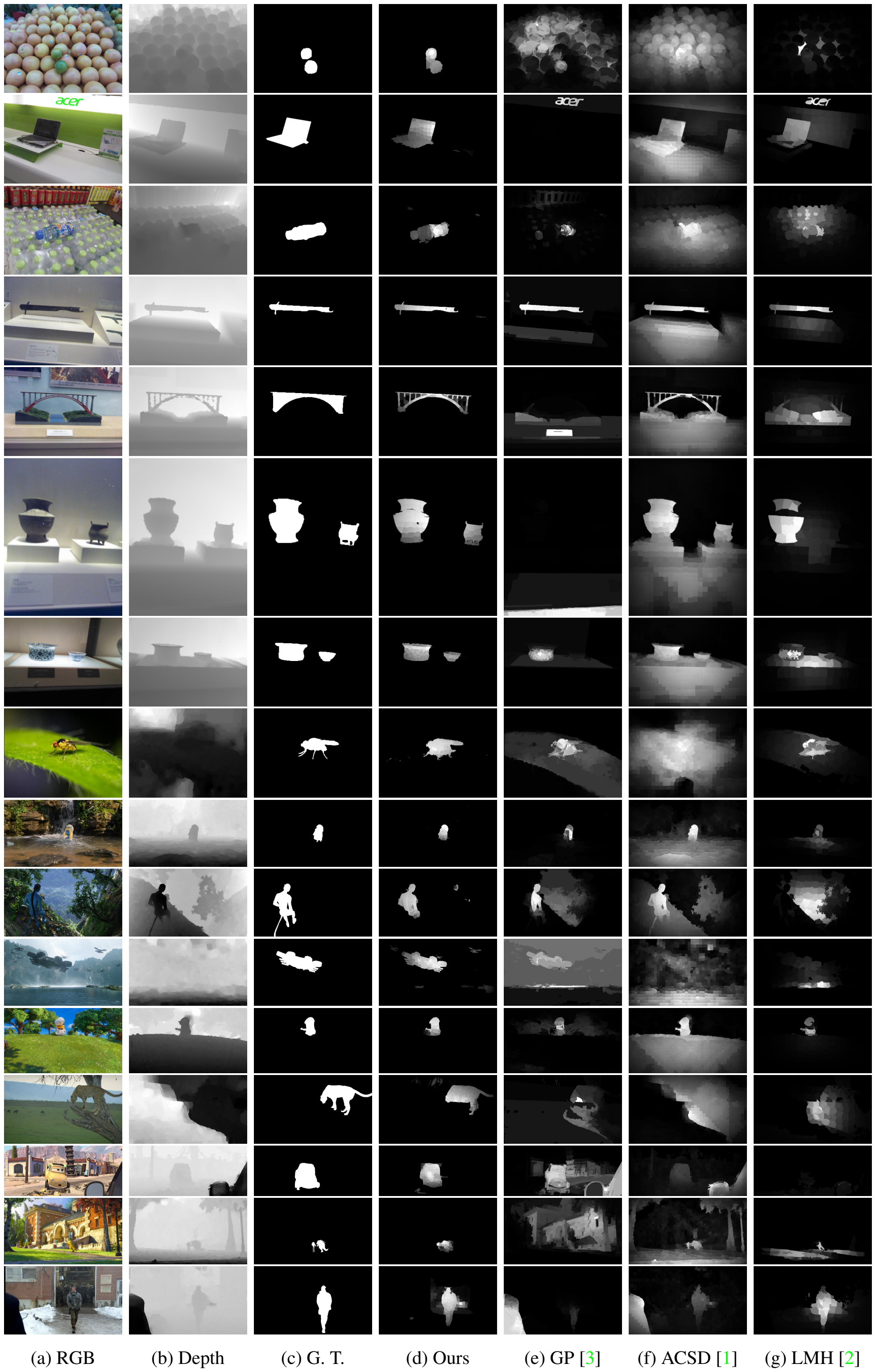


Figure 16. Comparison of output saliency maps produced by our salient object detection system against the output of GP [3], ACSD [1], and LMH [2]. Our robust LBE depth feature allows for a more accurate final saliency map compared to methods using contrast based depth features. Note that G. T. denotes Ground Truth.