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

David Feng, Nick Barnes, Shaodi You
NICTA; RSE, Australian National University
firstname.lastname@nicta.com.au

Chris McCarthy
Swinburne University of Technology
cdmccarthy@swin.edu.au

1. 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.

2. 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:

- GP-D, the global depth contrast term from a state-of-the-art RGBD salient object detection system [3]:
$$S_{GP}\left(r_i\right) = \sum_{j \neq i} A\left(r_j\right) C_{GP}\left(r_i, r_j\right), \tag{1}$$
where i, j are region indices, $A\left(r_i\right)$ gives the area of region r_i, and
$$C_{GP}\left(r_i, r_j\right) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{2\sigma_{\mathbf{x}}^2}\right)\|\mathbf{d}_i - \mathbf{d}_j\|_2, \tag{2}$$
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.

- GP-SD, a signed depth version of GP-D which excludes patches with a lower depth than r_i from the contrast computation.

- LMH-D, a multi-contextual depth contrast term from a state-of-the-art RGBD salient object detection system [2]:

$$S_{LMH}\left(P\right) = \prod_{k \in \{L,G,B\}} C_{LMH}\left(P, \Psi^k\right), \tag{3}$$

  where Ψ^L is the local context which consists of the 32 closest patches to P, Ψ^G is the global context which is the set of all patches, Ψ^B is the pseudo-background context which consists of the 36 patches closest to the image corners, and C_{LMH} is the contrast function given by:

$$C_{LMH}\left(P, \Psi^k\right) = -\log\left(\frac{1}{n_k}\sum_{j=1}^{n_k} \alpha_j^k \exp\left(-\frac{\|d - d_j^k\|^2}{2\left(\sigma^k\right)^2}\right)\right), \tag{4}$$

  where n_k is the number of patches in context Ψ^k, α_j^k is the normalised area of context patch j, d_j^k is the depth of context patch j, d is the depth of P, and σ^k is the bandwidth of context Ψ^k.

- LMH-SD, a signed depth version of LMH-D which excludes patches with lower depth than P from the contrast computation.

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

We use a custom implementation of GP-D [3], GP-SD, LMH-D [2] and LMH-SD, while code for ACSD [1] was obtained from the author's website.

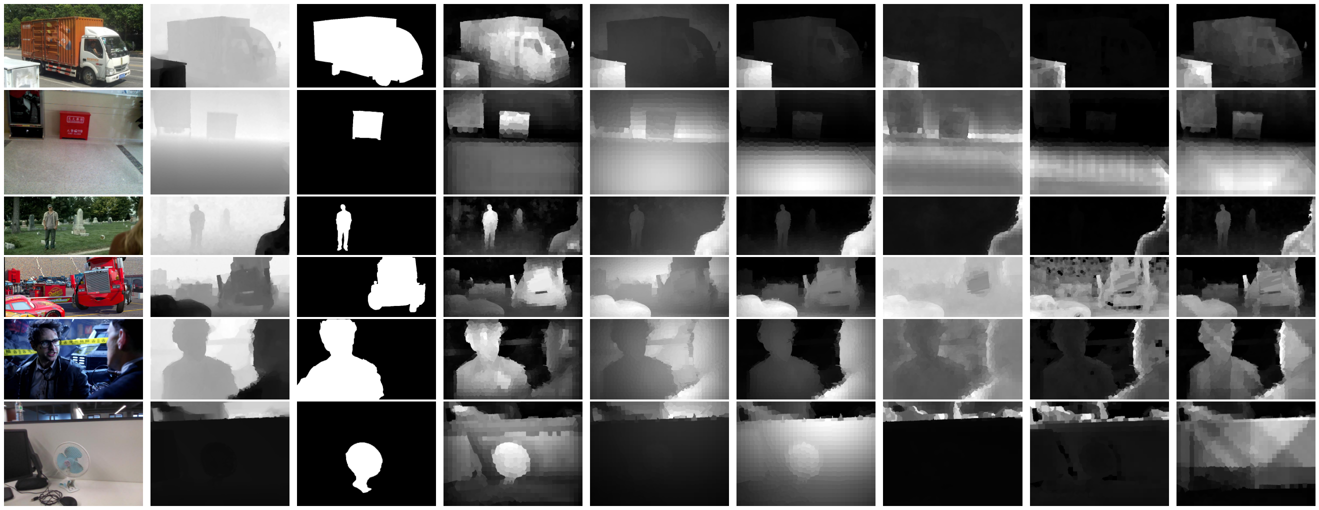
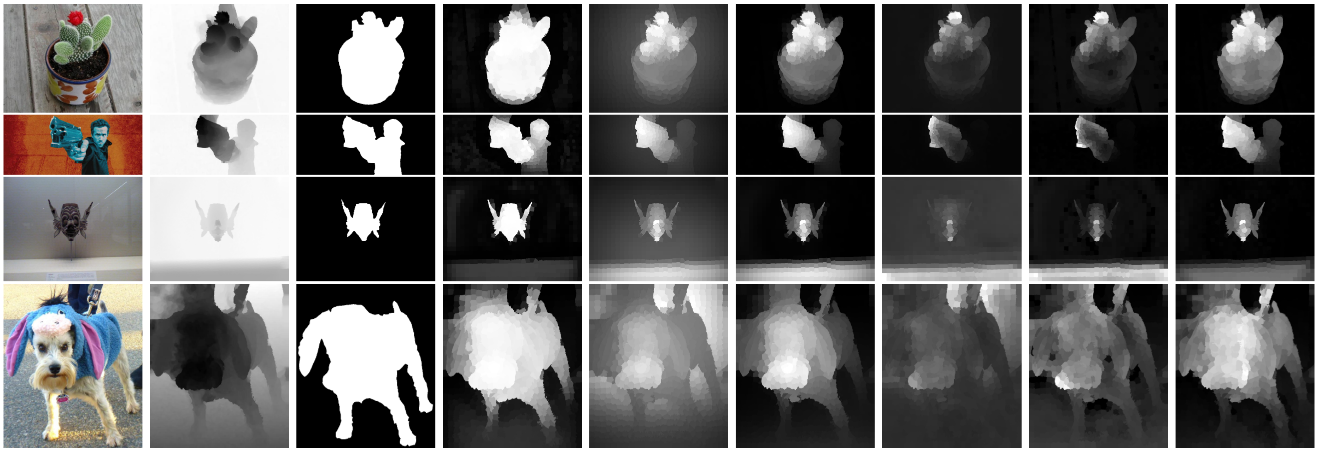


Figure 1. Examples of performance on images where the foreground has low depth contrast compared to the background using the raw depth saliency features of LBE, state-of-the-art methods ACSD[1], DC[3], and SDC[3] which is a signed variant of depth contrast. The depth contrast based methods perform poorly,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

2.1. 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 1 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.

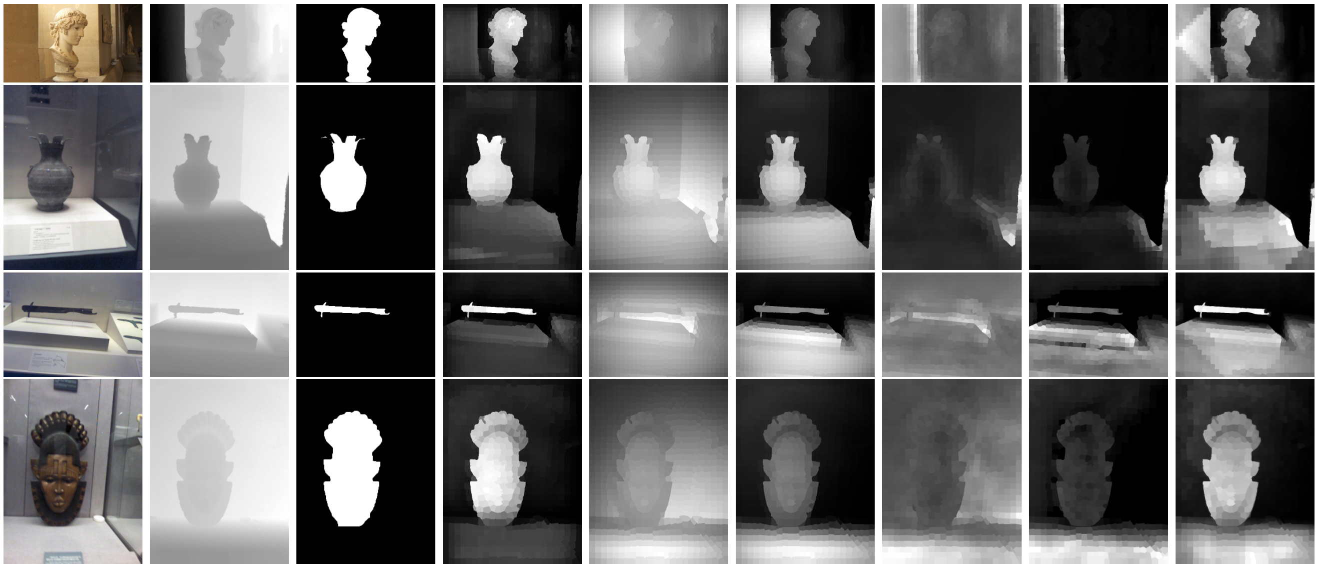


(a) Colour (b) Depth (c) Groundtruth (d) LBE (Ours) (e) GP-D [3] (f) GP-SD (g) LMH-D [2] (h) LMH-SD (i) ACSD [1]

Figure 2. 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.

2.2. 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 2. 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.

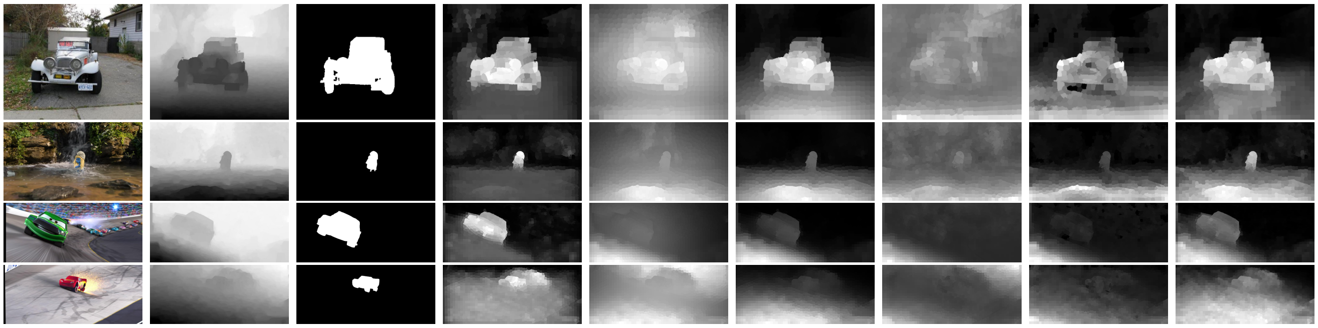


(a) Colour (b) Depth (c) Groundtruth (d) LBE (Ours) (e) GP-D [3] (f) GP-SD (g) LMH-D [2] (h) LMH-SD (i) ACSD [1]

Figure 3. 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.

2.3. 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 3. 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.

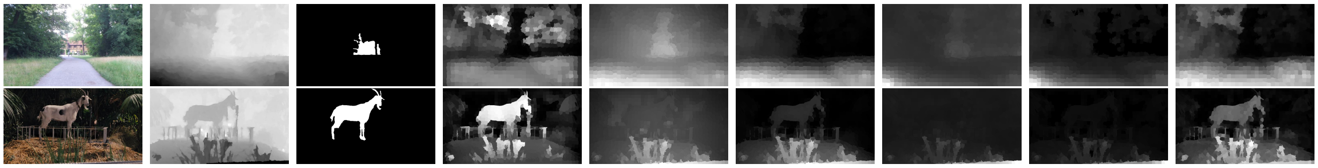


(a) Colour (b) Depth (c) Groundtruth (d) LBE (Ours) (e) GP-D [3] (f) GP-SD (g) LMH-D [2] (h) LMH-SD (i) ACSD [1]

Figure 4. 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.

2.4. 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 4. 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.



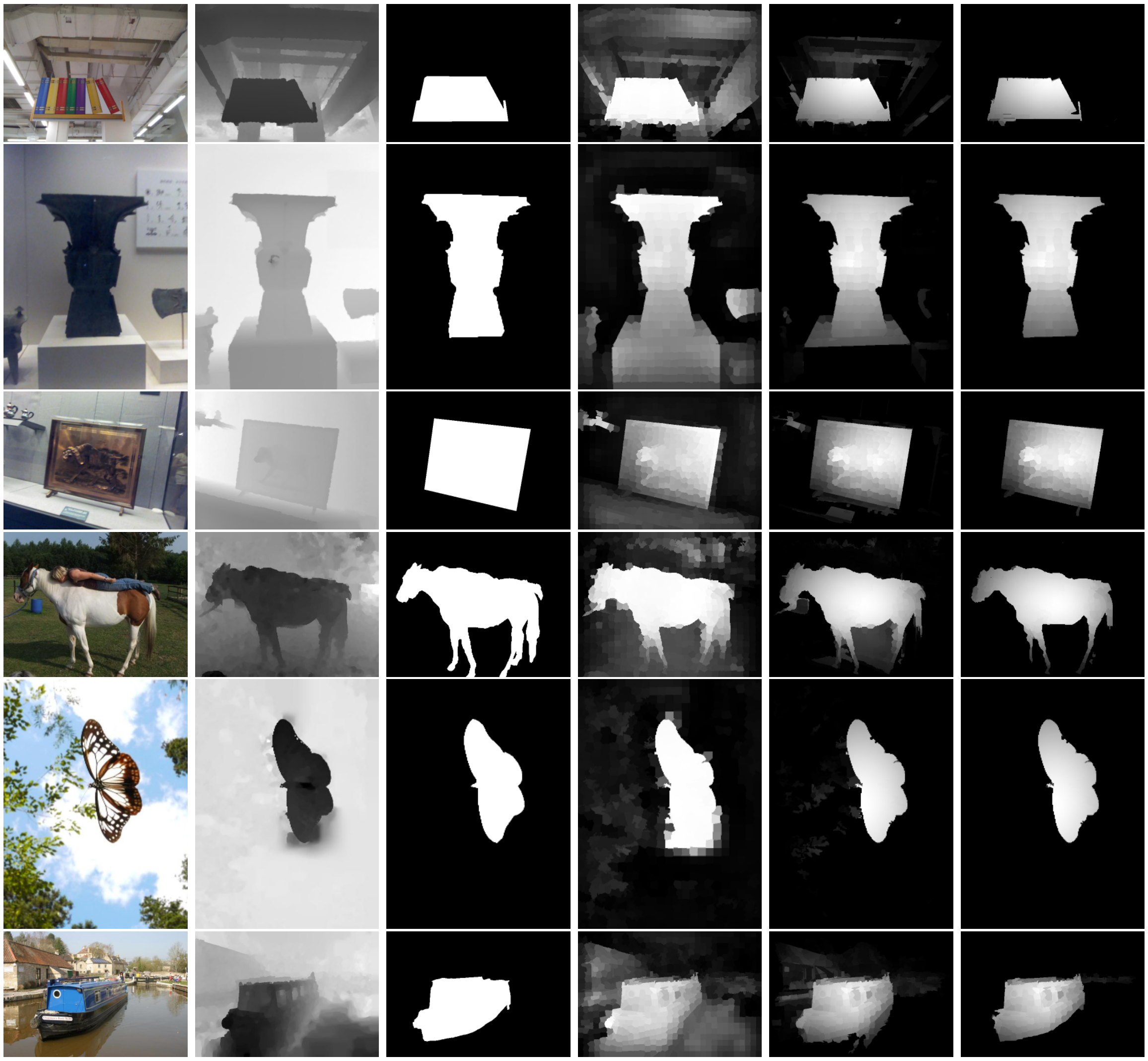
(a) Colour (b) Depth (c) Groundtruth (d) LBE (Ours) (e) GP-D [3] (f) GP-SD (g) LMH-D [2] (h) LMH-SD (i) ACSD [1]

Figure 5. 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.

2.5. 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 5. 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 5, 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 6. 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.

3. 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 6 and 7 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.

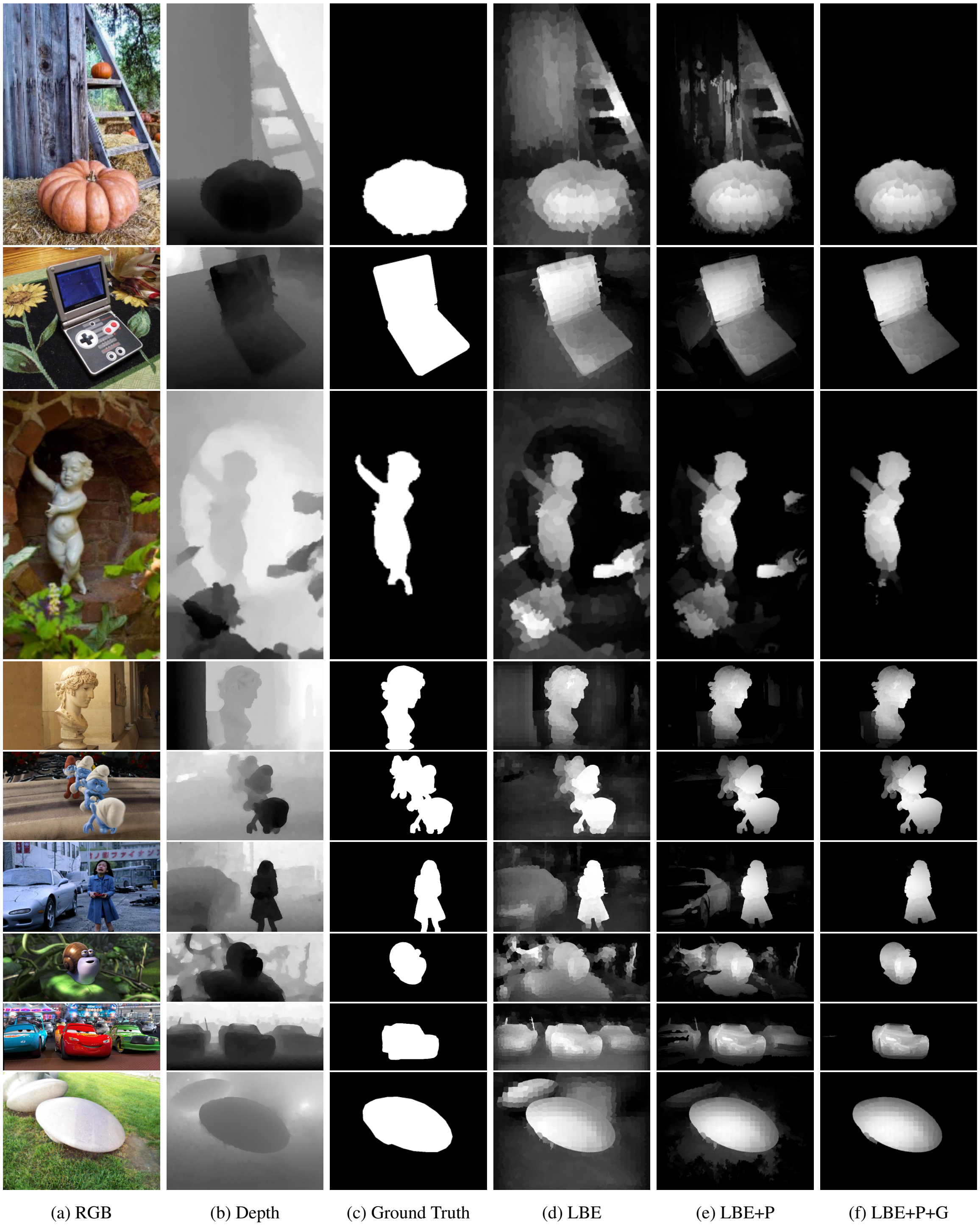


Figure 7. Output of the different stages of our salient object detection system. LBE denotes our proposed depth feature, LBE+P shows the result of depth, spatial, and background prior application, and LBE+P+G illustrates the final output of our salient object detection system after applying Grabcut refinement.

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, 7

[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, 2, 3, 4, 5, 6, 7

[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, 7