Adaptive Integration of Depth and Color for Objectness estimation

Xiangyang Xu, Ling Ge, Tongwei Ren and Gangshan Wu

State Key Laboratory for Novel Software Technology, Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing University, China

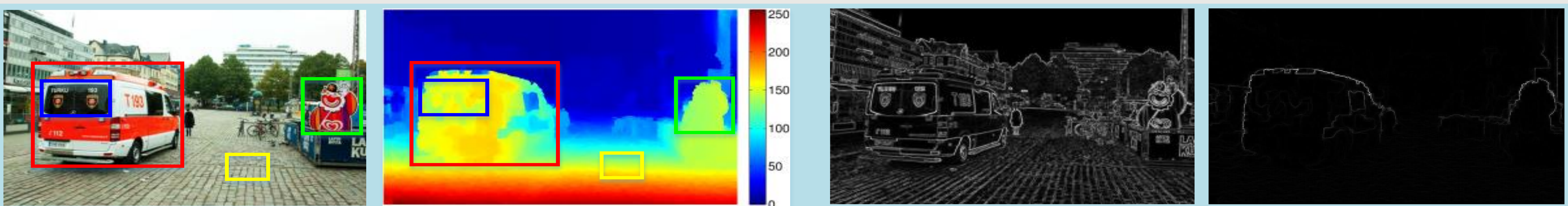
Introduction

The goal of objectness estimation is to predict a moderate number of proposals of all possible objects in a given image with high efficiency. Most existing works solve this problem solely in conventional 2D color images. In this paper, we demonstrate that the depth information could benefit the estimation as a complementary cue to color information.

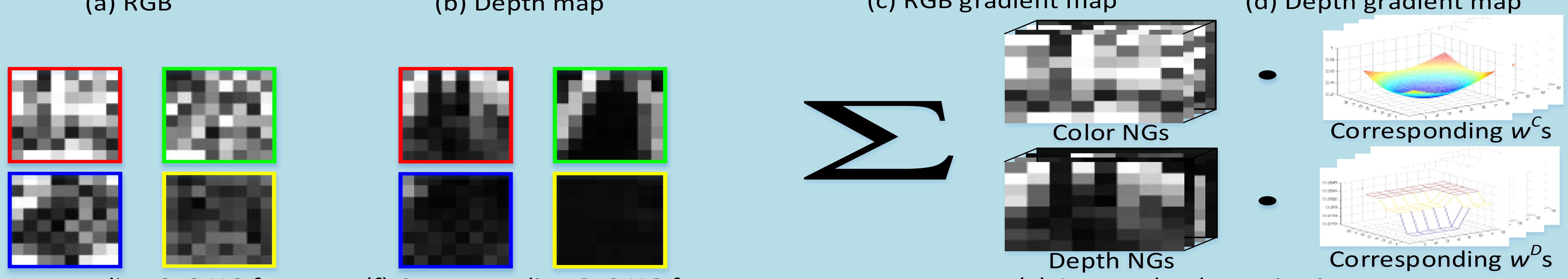
After detailed analysis of depth characteristics, we present an adaptively integrated description for generic objects, which could take full advantages of both depth and color. With the proposed objectness description, the ambiguous area, especially the highly textured regions in original color maps, can be effectively discriminated. Meanwhile, the object boundary areas could be further emphasized, which leads to a more powerful objectness description.

To evaluate the performance of the proposed approach, we conduct the experiments on two challenging datasets. The experimental results show that our proposed objectness description is more powerful and effective than state-of-the-art alternatives.

Proposed approach



(a) RGB (b) Depth map (c) RGB gradient map (d) Depth gradient map



(e) Corresponding 8×8 NG feature (f) Corresponding 8×8 NG feature (g) Approach schematization

Cheng et al. argued that the objects share strong correlation in the small normed gradient (NG) space, e.g., 8 × 8, as the "bounded" boxes shown in the Figure (e) above. Nevertheless, the "blue" van window (Figure (a)), a part of the van, is a suspicious false positive object in color space. Therefore, we reformulate the object window gradient with the depth and color weight maps. More specifically, for depth gradient map, we formulate w^D using the Bayes' rule:

$$w^D = P_h(p \in O|D) = \frac{P_h(p \in O, D)}{P_h(D)} = \frac{P_h(D|p \in O)P_h(p \in O)}{P_h(D)}.$$

For the color gradient, the inner parts are possibly distractive and should be suppressed according to the depth prior:

$$w^C = 1 - \delta \cdot G(x, y, \sigma_x, \sigma_y) = 1 - \delta \cdot A \cdot \exp(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}).$$

Due to the adaptive integration, our proposed objectness description (as Figure (g) depicts) has several advantages. First of all, when depth has strong intensity, w^D will contribute more and the inner parts of color normed gradient map (as "blue" bounding box in Figure (c)) are successfully suppressed. Meanwhile, object bounding borders can be emphasized by corresponding depth and color normed gradient map. What's more, as the distance between object and viewer increases, w^D adaptively decays and the color normed gradient will dominate the proposal prediction at faraway places.

Performance Evaluation

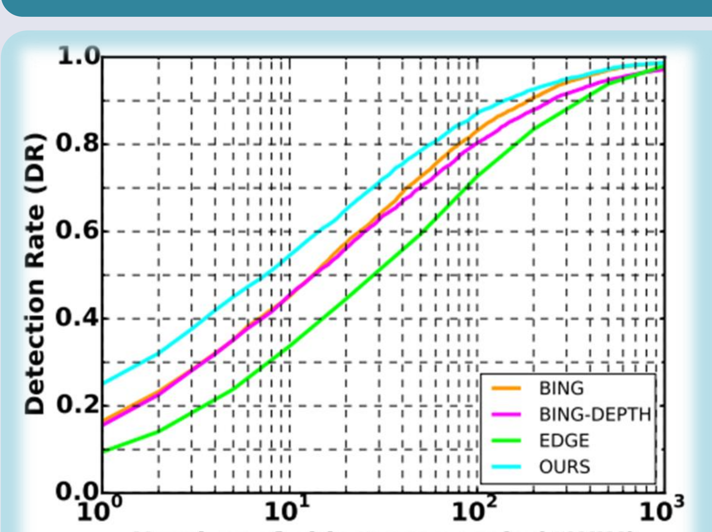


Fig. 2. Comparison of various approaches in stereo objectness dataset.

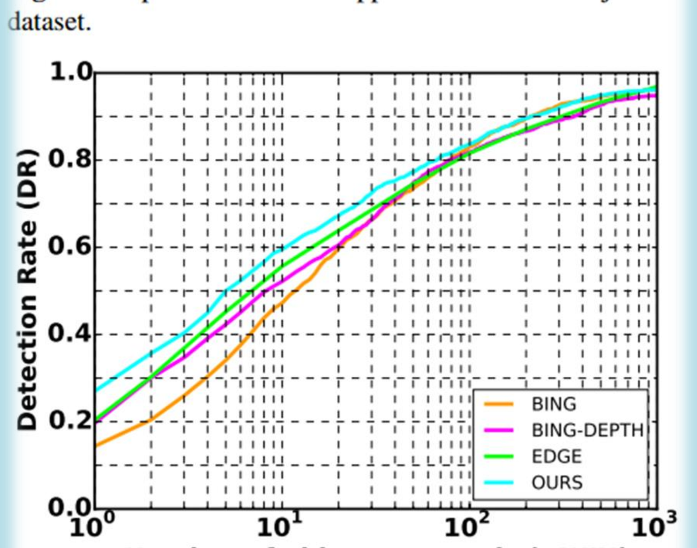


Fig. 3. Comparison of various approaches in RGBD dataset.

Fig. 4. Exemplars of some positive proposals closest to the groundtruth in the test images. The top two rows are from our stereo objectness dataset and bottom two are from RGBD dataset.

Some References

[1] Ming-Ming Cheng, et al., "BING: Binarized normed gradients for objectness estimation at 300fps," in CVPR, 2014.

[2] C Lawrence Zitnick, et al., "Edge boxes: Locating object proposals from edges," in ECCV, 2014.

[3] Houwen Peng, et al., "Rgbd salient object detection: A benchmark and algorithms," in ECCV, 2014.

[4] Sam Hare, et al., "Efficient online structured output learning for keypoint-based object tracking," in CVPR, 2012.

[5] Bogdan Alexe, et al., "Measuring the objectness of image windows," TPAMI, 2012.

Conclusion

By adaptively integrating depth and color cues, we propose a generic object description approach for objectness estimation. Based on the depth priors, object inner distractive regions can be effectively suppressed. Meanwhile, the object boundaries can be emphasized by the complementarily informative parts in depth and color gradient map. On the contrary, as the distance between object and viewer increases, the effect of the depth cue adaptively decays and the color will dominate the proposal prediction at places far away. Experimental results on two challenge datasets show the superiority of the proposed approach.

Research Team

Our research team, Multimedia Computing Group (MCG), belongs to the State Key Laboratory for Novel Software Technology, Nanjing University, China, and the leader is Professor Gangshan Wu. For more information, please contact Prof. Gangshan Wu (gswu@nju.edu.cn).