Abstract

Salient object detection (SOD) in images aims to identify and detect complete regions that represent visually distinct objects. SOD is often used in many tasks like image segmentation, image manipulation, foreground annotation, visual tracking and robotic handling. This thesis investigates a few issues related to the accurate detection of salient objects both in RGB and RGB-D images. It proposes novel methods to detect multiple salient regions, to leverage estimated depth for enhancing detection, to fuse complementary RGB and depth information, and to achieve effective fusion in the absence of data-driven supervision.

The proposed multiple salient region detection technique for RGB images performs locally adaptive center-surround operations on proto-object partitions obtained through color consistency and spatial proximity analysis. The multi-scale center-surround operations are performed by specially designed operator masks that are adapted to the local structure, which yields salient regions with accurate boundaries.

Depth information about image contents can play a significant role in SOD, but RGB images do not contain such data. However, sophisticated techniques similar to feature extractors exist that estimate depth information from RGB images. Even though estimated depth data may not be as accurate as the actual, SOD in RGB images can benefit from them by learning to adapt to them as a feature input along with the RGB data. To this end, a novel estimated depth-based interactive framework is proposed that estimates depth from the input RGB image using a pre-trained model, and fuses hierarchical features of both the estimated depth and the input RGB data to achieve SOD for RGB images.

When both color and depth data are available, RGB-D SOD can be performed where RGB data provide appearance details and depth maps offer vivid spatial structure. This thesis introduces a novel approach for RGB-D SOD that utilizes varied contexts to model different influences while extracting features consistently from object parts and surroundings leading to their saliency determination. In addition, the approach performs an RGB-D fusion unlike before using ordered multiple attention forms, to ensure that only significant components are sequentially allowed in the fused feature representation.

Data-driven deep fusion methods usually require sufficient annotated data for training to achieve satisfactory SOD performance, which in some cases like industrial applications are extremely difficult to obtain. A zero-shot deep RGB-D fusion approach based on the novel concept of fuzzy membership learning is proposed in this thesis, which does not require any ground truth data for training. The learnable parameters of the fuzzy membership learning network are optimized based on a novel membership similarity measure (MSM) to achieve the fusion using a fuzzy inference system.

Multiple quantitative and qualitative result analyses on standard RGB and RGB-D SOD datasets signify that the proposed methods consistently achieve state-of-the-art SOD performance. The utility of all the proposed SOD methods is demonstrated in a robotic workspace, where a miniature industrial robot performs manipulation operations based on SOD. The effectiveness of our methods is shown in terms of localization accuracy, and detection and recognition rates.

Keywords: Structure-aware multiple salient region detection, Consistent feature computation, Attention-based RGB-D fusion, Zero-shot fusion, Fuzzy membership learning