Unsupervised Saliency Detection in 3D video based on Multi-scale Segmentation and Refinement

Ping Zhang*, Pengyu Yan, Jiang Wu

***Abstract*—In this paper, we propose an unsupervised saliency objects extraction method for 3D video. The proposed framework consists of three main stages: (i) The input video frame is segmented into non-overlapping superpixels by combining both appearance and depth information at the input. Depth information can be used to improve the accuracy of segmentation in complex regions, *i.e.*, foreground object with similar appearance as background. A multi-scale segmentation scheme is also deployed by using different segmentation parameters to extract varying shapes of the foregrounds in each frame. (ii) The initial saliency score of each segmented superpixel in each scale is calculated via global contrast which is defined by appearance, depth, and motion cues from two consecutive frames. (iii) The initial saliency scores in each scale are refined by smoothing over a graph built by the spatial neighboring of all the superpixels in the frame. The final result is generated by fusing the saliency maps in all scales. The experiments on two widely-used datasets illustrate that our method outperforms state-of-the-art algorithms accuracy, robustness, and reliability.**

***Index Terms*—Saliency detection, Segmentation, Appearance, Motion, Depth**

I. Introduction

Human visual attention is most significant in human visual system, as it ensures us to detect the most important object from the very complex scenery. This object is called the salient object and the corresponding area in the picture is saliency area.

Center-surround visual attention in [1] used three elementary features, color, luminance and direction, to calculate the contrast between the center and surround areas. By knowing the mechanism of the human visual perception, Itti [1] defined “saliency” as the area which is much different from the surrounding. This definition was broadly accepted by the research community. In [2], [3], the center-surround algorithm was upgraded using the KL distance of two feature histograms. Later, instead of using the part contrast, global contrast algorithm was designed to segment the saliency area [4]. Considering the feature dispersal effect, a “saliency filter” [5] was designed to calculate both the contrast and dispersal degree of color to determine the saliency area. Apart from the contrast algorithm, graphical model is used in saliency detection starting in 2006. Harel et. al. [6] modified the method of Itti [1] and proposed a graphical model to represent the similarity between each part within the picture. Random Walk (RW) algorithm was adopted to give the grade of each area. In 2013, Yang et. [7] used Manifold Ranking algorithm to distinguish the foreground and background areas.

In 3D scenery, depth information is naturally incorporated in the saliency detection model. Zhang et. [8] used appearance, depth, motion, illumination and direction to calculate bottom-up saliency in a picture and the closer the object from viewer, the more salient the object. In [9], depth is used to weigh the 2D saliency result. Based on that, Niu [10] constructed a depth weight curve by presuming that the comfort zone and popping out object are more salient than any other region or object. In [11], [12], instead using depth as a weight bias, the authors considered it as another feature and calculated the depth-only saliency map via global contrast. Finally, they combined the 2D saliency map with the depth saliency map and obtained 3D saliency result.

Later, with the growing focus on video, motion and depth information is naturally utilized in saliency detection. In 2010, Zhang proposed the 3DV [8] method which fused the depth feature map with appearance and motion saliency map. In 2015, Lino [13] designed a saliency detection method which spatial, motion and depth saliency maps is obtained based on Itti’s model, block matching and Difference of Gaussian (DoG) filter, separately. These results are fused to yield the final saliency map.

Although there are many proposed methods, their performance is limited in preserving clear edges of the foreground object. The performance is not sufficiently robust for a wide range of image and 3D video.

In order to overcome the aforementioned challenges, we proposed a multi-scale architecture and a refining method based on graphical model to make the detection method general and robust. In our paper, we upgraded Simple Linear Iterative Cluster (SLIC) algorithm [14] by adding the depth information as one feature while clustering the pixels into superpixels. As a result, two regions which share a similar appearance but belong to different objects can be separated and then the edge of the objects can be preserved well. Multi-scale architecture is generated by setting different number of superpixels in each scale and then fusing the final result of each scale together. This architecture enables robustness and generality by combining the merit under each scale and fixing the problem of setting optimized superpixel number due to the continuously changed inputs. After motion feature was calculated by TV-L_1 optical flow [15], [16], the initial saliency map are obtained via global contrast [11] with appearance, depth, and motion information. Finally, the refinement based on graphical model [17] is utilized to refine the initial saliency

Ping Zhang, associate professor with the University of Electronic Science and Technology of China.

Pengyu Yan and Jiang Wu are students in University of Electronic Science and Technology of China

*:Respondence author: pingzh@uestc.edu.cn

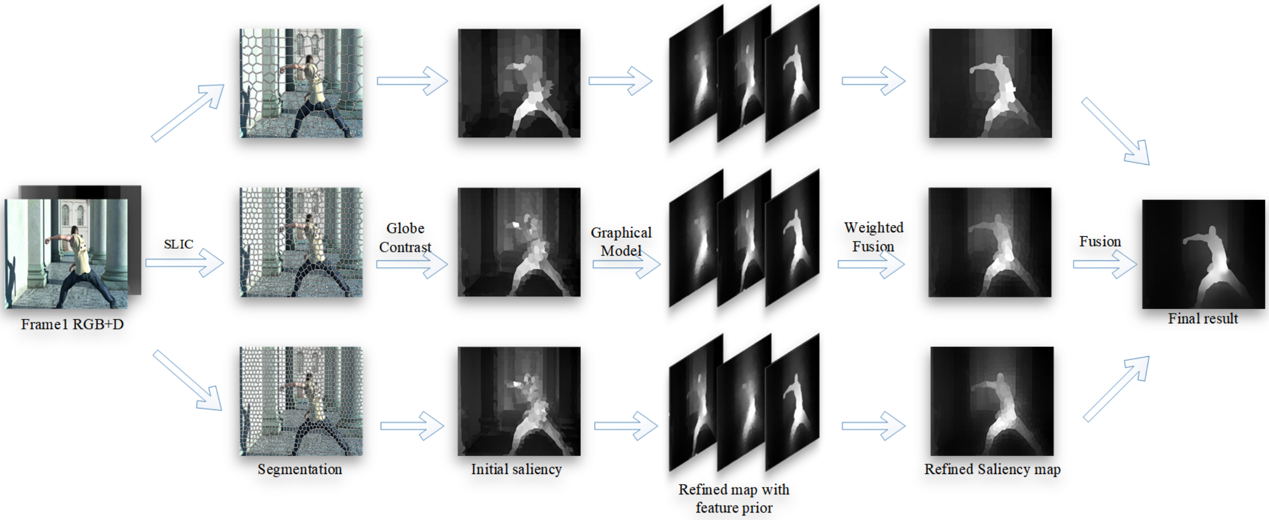


Fig. 1: Flowchart of the proposed method

maps, which makes the saliency area smooth in interior and regular in shape.

Overall, the main contributions of our method are summarized as follows:

- In order to segment the input with a well-preserved edge and improve the performance, we upgrade the SLIC algorithm by adding the depth information as one term in K-mean clustering algorithm. Also the weight between the distance, appearance and depth are redesigned and set by the experiment.
- Since the inputs varied significantly when processing the video sequence, we generated a multi scale architecture based on modified SLIC algorithm to make the model generic and robust. One input matches three output with different number of superpixels. From the first to third scale, we use 200, 600, 1000 superpixels to segment the input. The letter mechanism is all based on the level of minor scale. The result of each scale would be fused in the last step.
- After the initial saliency calculation, we adopt the graphical model to refine our saliency map. By calculating transition probability among center-surrounded superpixel via appearance, depth and motion, we fine tune the saliency score of each superpixel. Finally, the interior of the saliency area would be smoothed and the shape and edge of the object would be preserved well.

The rest of this paper is organized as follows. Section II described the proposal in detail including multi scale based on upgraded SLIC, initial saliency calculation via global contrast and refinement by graphical model. Then in section III, exclusive experiment and comparisons are discussed. In the last section IV, a conclusion is made and the future work is anticipated.

II. Proposed Method

Fig. 1 shows the framework of the proposed method which consists of three parts: (i) Based on upgraded SLIC algorithm, each input image (frame) with appearance and depth is segmented to three images with different superpixel number, as three scales. (ii) The score of each superpixel in each image under each scale is computed using global contrast which is defined by appearance, depth, and motion cues from two consecutive frames. Consequently, the initial saliency map in each scale is obtained. (iii) In order to refine the initial saliency map, graphical model is adopted to smooth and regulate the saliency area by appearance, motion and depth prior. Then the final result is obtained by fusing three refined saliency map in all scales. A guided filter is also adopted to process the final result. The proposed method is unsupervised since ground truth is only used for testing. Detailed description of each part is presented in the following subsections.

A. *Segmentation with multi-scale*

1) Upgraded SLIC Model: In this part, our aim is to decompose the input image into non-overlapping segments and preserve accurate object's shape and edge. While the traditional K-means algorithm in SLIC used appearance, *i.e.*, color, and distance to cluster pixels, we further add depth information with appropriate weight, in the proposed method. We set the depth feature distance as $D_d = \sqrt{(d_j - d_i)^2}$, where d_j and d_i represent the j^{th} and i^{th} superpixels' depth value, respectively. By adding the weight coefficient of depth feature distance as β, the whole feature distance evaluation coefficient becomes:

$$D' = (1-\beta)D_c + \beta D_d \tag{1}$$

$$D = \sqrt{(\frac{D'}{m})^2 + (\frac{D_S}{S})^2} \tag{2}$$

Here D' is the weighted mean of color and depth feature distance, noted as D_c and D_d, respectively. D_S is the space

Euclidean distance calculated by the coordinate of each pixel. m is a parameter related to the largest color distance in one image, with typical range of $[1, 40]$ and S is the expected number of the superpixel. Throughout the experiment, β and m are set to 0.5 and 15, as they give the best segmentation performance.

By computing the color, depth and spatial distance, the algorithm clusters pixels with smallest distance. The algorithm yields excellent segmentation result with 10 iterations.

2) *Multi-scale Architecture:*

The number of superpixels is a parameter in image segmentation. Small number of superpixels reduces the algorithm's ability to preserve fine edges of an object, while large number of superpixels misses important corners. Furthermore, the number of optimal superpixels depends on the input images.

In order to increase the generality and robustness of the proposed algorithm, a Multi-scale approach is integrated with the SLIC algorithm. The first part of Fig. 1 shows the modified SLIC algorithm's output for a image with appearance and depth information and segmented to 200, 600 and 1000 superpixels. The feature value in each superpixel is the mean of all pixel's feature value in the superpixel area.

The rest part of the whole model is calculated upon superpixel level and within each scale. At the last step, the result of all scales would be finally fused together and we would introduce this fusion way in section II-D.

B. *Initial saliency*

According to visual perceptual research [1], image contrast is the most important factor in visual saliency. In a static image, visual attention centers on the objects which have a high contrast comparing to their surrounding. In 3D videos, the object with hign-contrast appearance, depth and motion would cause the visual attention. In this part, two steps are proposed.

First, we use CIE-Lab color space to represent the appearance of input image, since LAB color correlates better to perceptual of human eyes. Motion is extracted by the TV-L_1 Optical Flow [15], [16] and the two parameters of motion, v_X and v_y, can be obtained by solving the equation $I(x, y, t) = I(x + \mathrm{d}x, y + \mathrm{d}y, t + \mathrm{d}t)$ with the constraint of by $I_x v_x + I_y v_y + I_t = 0$. I is the luminance.

Second, we use global contrast with color, depth, motion and spatial distance to generate the initial saliency maps. Since the images have been segmented by SLIC, feature contrast can be calculated on the superpixel level. We use arithmetic mean of color, motion and depth of all pixels in each superpixel area to represent the feature value of whole superpixel. The spatial distance is calculated among the barycenters of superpixels.

Similar to the distance calculation in SLIC in II-A, the feature distance of color, denoted as d_C and motion, denoted as d_M can be calculated by Euclidean distance:

$$d_C(R_j, R_i) = \sqrt{(L_{R_j} - L_{R_i})^2 + (a_{R_j} - a_{R_i})^2 + (b_{R_j} - b_{R_i})^2} \quad (3)$$

$$d_M(R_j, R_i) = \sqrt{(v_{x_{R_j}} - v_{x_{R_i}})^2 + (v_{y_{R_j}} - v_{y_{R_i}})^2} \quad (4)$$

where R_i denotes the i^{th} superpixel, and L, a, b are the color values under CIE-Lab.

We set the gray difference between two superpixels R_j, R_i as the depth feature distance d_D, and D_{R_i} represents the depth value of the R_i^{th} superpixels:

$$d_D(R_j, R_i) = \sqrt{(D_{R_j} - D_{R_i})^2} \quad (5)$$

After the feature distance calculation, we use weighted mean of these four feature distances, namely color, motion, depth and spatial distance, to represent the saliency score of each superpixel.

According to the visual mechanism of human which the closer the two superpixels are, the more influence they had for each other, we designed the weight coefficient based on spatial distance as follows:

$$\omega(R_j, R_i) = \exp(-\frac{d_S(R_j, R_i)}{\sigma}) \quad (6)$$

d_S represents the spatial distance among the barycenter of the superpixels. When two superpixels are closer, the value of ω is larger. The value of the parameter σ is 0.6.

We can obtain the score of each superpixel under each feature:

$$\begin{aligned} S_F(R_i) &= \sum_{R_j \in \Omega} s_F(R_j, R_i) \\ &= \sum_{R_j \in \Omega} \omega(R_j, R_i) \cdot d_F(R_j, R_i) \end{aligned} \quad (7)$$

where the notation "F" stands for either "C","M" or "D", *i.e.*, color, motion, and depth, correspondingly. s_F is "F" feature distance between two superpixels, which is weighted by spatial distance and $S_F(R_i)$ is the saliency score of R_i^{th} superpixel under "F" feature. Ω is the set of all the superpixels in one frame.

By normalizing all superpixels' score, the "F" feature saliency map is obtained. Since the quality of each feature saliency map is different, a weighted fusion method is conducted to fuse color feature saliency map, motion feature saliency map and depth feature saliency map to obtain the initial saliency map.

According to perceptual research [1], the more saliency area aggregated, the higher the quality of the saliency map is. Thus we design a way to calculate the aggregation degree:

$$\mu_F = \frac{\sum\limits_{R_i \in \Omega} \sqrt{(y_{R_i} - \overline{p_{y_F}})^2 + (x_{R_i} - \overline{p_{x_F}})^2} \cdot S_F(R_i)}{\sum\limits_{R_i \in \Omega} S_F(R_i)} \quad (8)$$

Here, "F" means this equation is used to calculate value of "F" feature saliency map. The (x_{R_i}, y_{R_i}) is the coordinate barycenter of the R_i^{th} superpixel, $(\overline{p_{x_F}}, \overline{p_{y_F}})$ is the barycenter of the saliency area calculated by saliency scores of all superpixels, and μ_F is the aggregation degree of each feature saliency map. $S_F(R_i)$ is the "F" feature saliency score of R_i^{th} superpixel.

Based on saliency aggregation degree, the initial saliency maps can be computed as follows:

$$S = \frac{\sum_{F \in \{C,M,D\}} 1/\mu_F \times S_F}{3} \quad (9)$$

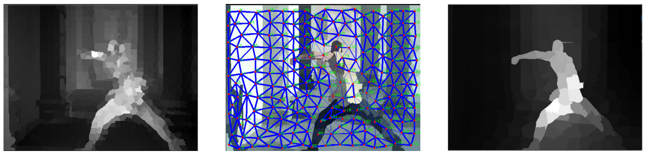
C. Saliency refinement based on Graphical Model

Initial saliency map obtained from global contrast has many artifacts at object's edge, as seen in Fig. 2(a). To regularize the saliency map and maintain the object's shape, a refinement method based on graphical model is exploited. A probability transition function [17] can reduce the great saliency value diversity between two near superpixels inside the saliency region and finally make the saliency area smoother and the edge between salient and none-salient area clearer.

First, upon initial saliency maps, we separately use motion, color and depth features to calculate the transition probability $P_{R_j,R_i,F}$:

$$P_{R_j,R_i,F} = \frac{W_{R_i,R_j,F}}{\sum_{R_j\in\Omega_{R_i}} W_{R_i,F}} \qquad (10)$$

where $F \in \{C, M, D\}$, and $W_{R_i,R_j,F} = \exp(-\frac{d_F(R_j,R_i)}{\sigma})$. We can see that the more different the two superpixels feature is, the smaller the transition probability would be.



(a) Initial Saliency (b) Graphical Model (c) Optimized Saliency

Fig. 2: Optimization based on Graphical Model

Based on the transition probability, we draw the lines among the superpixels' barycenters and use the thickness of the lines to represent the value of transition probability. If the feature contrast of these two superpixels are small, the line would be thick, vice versa (Fig. 2(b)). In Fig. 2(b), the rough outline of the objects could be presented by the thin lines, which reflect a big difference between the objects and the background. It is also the places where the line just cross the objects' edge.

Second, we use the transition probability as the weight to process the initial saliency maps:

$$U_F = \sum_{R_j,R_i\in\Omega} P_{R_j,R_i,F}\cdot s(R_j,R_i) \qquad (11)$$

Here, $s(R_j, R_i)$ means the fused feature distance between two superpixels implied in the initial saliency maps and it can be calculated from the eq. (7) and eq. (8) by $s(R_i, R_j) = \frac{1}{3}\sum_{F\in C,M,D}\frac{1}{\mu_F}s_F(R_j, R_i)$.

From this equation, we can gain the refined saliency map by each feature prior, for example U_M represents the refined saliency map by motion prior.

Third, after the normalization, the three refined saliency maps by three features priors would be fused by a simple way:

$$U = \frac{1}{3}\sum\nolimits_{F\in\{C,M,D\}} U_F \qquad (12)$$

The proposed fusion method yields refined saliency map with well-preserved object's shape and edge, as shown in Fig. 2(c).

Now, the refined saliency map under one scale is obtained. Similarly, refined saliency maps in the other two scales are calculated as the same way.

D. Fusion of Multi Scale

After the section II-C, there are three refined saliency maps under three scales. We name each saliency map by U_n, *e.g.*, U_1 means the saliency map in the first scale.

$$U_A = \frac{1}{3}\sum\nolimits_{n=1}^{3} U_n + \prod\nolimits_{n=1}^{3} U_n \qquad (13)$$

Considering the slight difference among saliency maps of all scales, we designed a fusion way by combining an average sum and multiplication. In the sum part, we average it to make the saliency value of each superpixel still between $[0, 1]$. In the multiplication part, we multiple saliency maps together and then the common saliency areas in all saliency maps would be enlarged, and the none-common saliency area or none saliency area in all saliency maps would be greatly minified (because it is the multiply between 0 and 1). In multiplication part, the saliency value of each superpixel is among $[0, 1]$ as well. Then we add these two part together to improve the performance of the final refused result. The normalization: $U(R_i) = [U_A(R_i) - min(U_A)]/[max(U_A) - min(U_A)]$ is also used and U represent the final saliency result of the whole proposal.

III. Experiments and Evaluation

A. Datasets and settings

Two public datasets, 3D sequence supplied by 3D-HEVC [18] and NAMA3DS1 [19], are used to evaluate the proposed method. 3D-HEVC provided totally 8 groups of 3D sequences with both appearance and depth, and the resolution is 1280×720 by 'yuv' form. NAMA3DS1 dataset includes ten 3D full HD stereoscopic sequences with 25 frames per seconds and contains the depth images generated by disparity estimation algorithm. Since each dataset contains plenty of depth information and enough sequence to calculate motion, we choose ten groups of sequence in each datasets to test the proposed method. The ground truth is made manually by Adobe Photoshop.

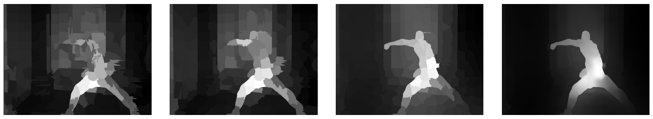


Fig. 3: the images from left to right are:initial saliency map without depth, initial saliency map with depth, refining based on graphical model, result of multi-scale approach

B. Evaluation Method

In this section, two steps are adapted to validate our experiment. The first step is to vindicate that every part of our proposal is rational and necessary. The second step is to compare our method with other state-of-the-art methods.

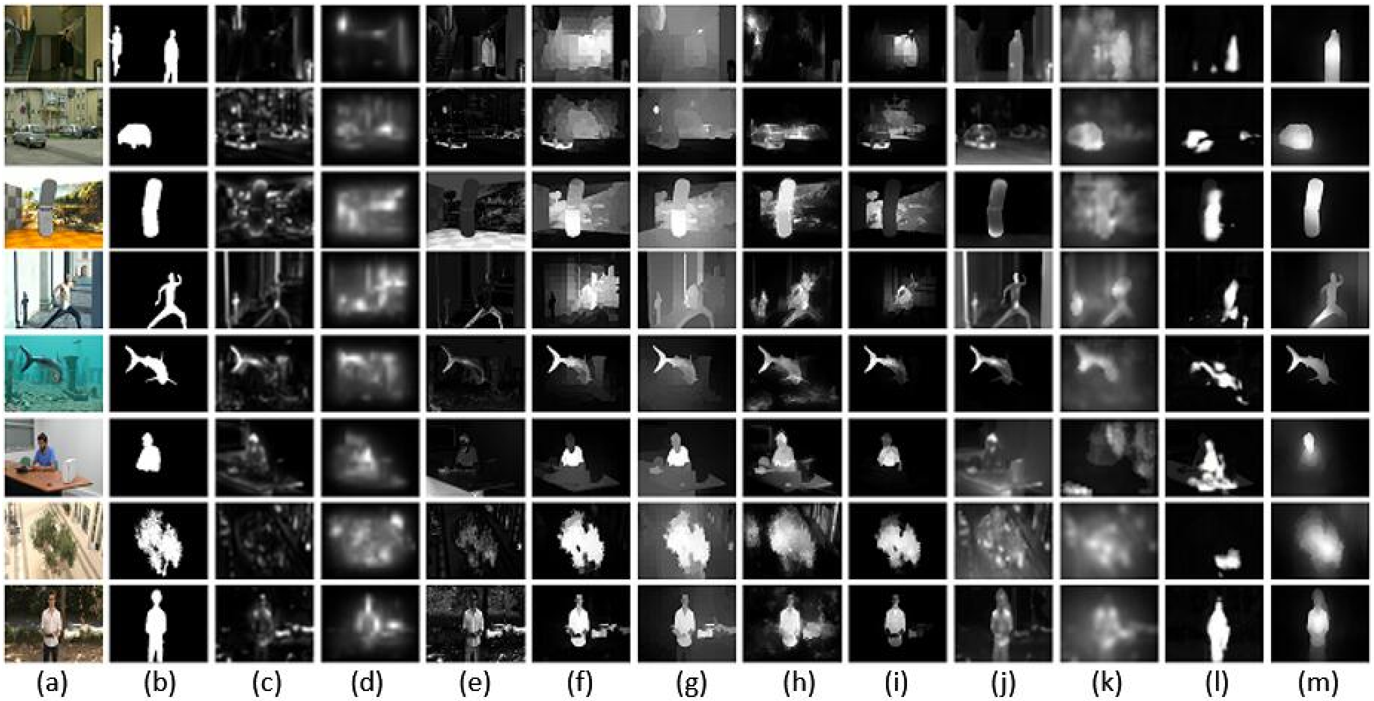
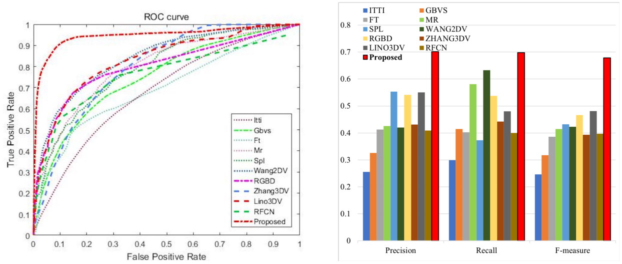


Fig. 4: Example of saliency results of differen methods. (a) Original frame. (b) Ground Truth. (c)-(l) Results of ITTI [1], GBVS [17], FT [20], MR [7], SPL [21], WANG2DV [22], RGBD [12], ZHANG3DV [8], LINO3DV [13], RFCN [23]. (m) Our proposal



(a) ROC Curves

(b) Evaluate Chart

Fig. 5: Evaluation

1) Vindication of Proposal:

From Fig. 3, by comparing the result of each steps of our approach, we can observe the importance of each step in our proposal. As observed in the first two sub-images in Fig. 3, adding depth could help to improve the performance of the segmentation and initial saliency maps. After refinement, the quality of the saliency maps improved greatly, which preserved the object edge and regularized the object shape. The multi-scale approach further improves the saliency result in the last subfigure as shown in Fig. 3.

TABLE I: Performance of two Fusion Ways

Fusion Way	Precision	Recall	F-measure
Only Average Sum	0.6717	0.7319	0.6683
Average Sum + Multiplication	**0.7020**	**0.6977**	**0.6783**

Meanwhile, as shown in Table I, the performance of the second fusion way is better than the first one which indicate that our proposed fusion method in Section II-D is reasonable and well-performed.

2) Comparison with other Proposal:

In the next experiments, we compare our proposal with other ten state-of-the-art salient detection methods: ITTI [1], GBVS [17], FT [20], MR [7], SPL [21], WANG2DV [22], RGBD [12], ZHANG3DV [8] and LINO3DV [13], also, a deep learning way in 2016, RFCN [23], is compared as well. In each proposal, the parameters are set as default.

TABLE II: Measurement

Models	Precision	Recall	F-measure
ITTI	0.2549	0.2992	0.2464
GBVS	0.3255	0.4143	0.3171
FT	0.4123	0.4023	0.3859
MR	0.4255	0.5811	0.4142
SPL	0.5536	0.3726	0.4322
WANG2DV	0.4197	0.6325	0.4231
RGBD	0.5416	0.5378	0.4661
ZHANG3DV	0.4308	0.4422	0.3931
LINO3DV	0.5507	0.4804	0.4813
RFCN	0.4087	0.3993	0.3968
Proposed	**0.7020**	**0.6977**	**0.6783**

Results of seven sets are selected (shown in Fig. 4). The first four curves are datasets from 3D-HEVC and the last three are selected from NAMA3DS1. Each datasets are sequences with the depth information and we pick up several exampled frame for illustration. From the results, we can see that our proposal performs well at extracting the saliency object with a well-preserved edge and shape. In the second line, the other proposal failed to extract a whole clear car, while in our method, the car maintains consistent shape and edge and the background is restrained greatly. In complex scenario, the methods in (d), (f) and (i) yield poor results with unrecognizable objects, whereas our method preserves object's shape and edge. The RFCN method, which shows in the (l) row, presents good performance in specific scenes but becomes poor in others like the second and seventh line. However, our method perform well in all situation, which indicate the robustness and general of our proposal.

In addition, ROC curves, average Recall, F-measure and accuracy have been measured as well. Fig. 5(a) summarizes the excellent performance of our algorithm over other state-of-the-art algorithms. And as shown in fig. 5(b), it is easy to see the great performance of our proposal among other method. Table II again verifies that the proposed algorithm achieves the best performance.

C. Discussion

In our experiments, we also find some limitations of the proposed method. Fig. 6 shows that when there is an another big object with a large contrast with surroundings, our method can not segment the object where the scene also includes a foreground salient object. In this scenario, one could argue that the proposed method yields sensible salient map, as the foreground object could also be salient. This example showcases the robustness of the proposed method.

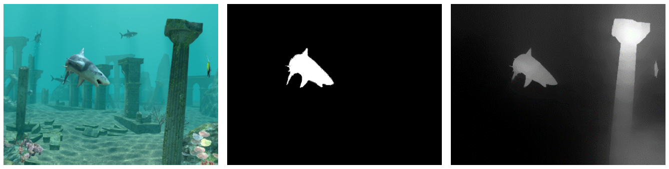

(a) original image (b) ground truth (c) saliency result

Fig. 6: Limitation illustration

IV. Conclusion

In this paper, we proposed an improved SLIC segmentation method and a multiscale architecture. Additionally, refining method based on graphical model is adopted to improve the object's shape and edge in the saliency maps. The experiments on 3D-HEVC and NAMA3DS1 dataset verify the excellent performance of our proposal. These datasets cover a wide variety of objects with background. The saliency maps using the proposed method consist of objects with clear shape and edge, demonstrating the generality and robustness of the proposed algorithm.

Future work includes developing an improved multi-scale algorithm with optimal number of superpixels and the corresponding weights.

References

[1] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, Nov 1998.

[2] D. Gao and N. Vasconcelos, "Bottom-up saliency is a discriminant process," in *2007 IEEE 11th International Conference on Computer Vision*, Oct 2007, pp. 1–6.

[3] D. Gao, V. Mahadevan, and N. Vasconcelos, "On the plausibility of the discriminant center-surround hypothesis for visual saliency," *Journal of Vision*, vol. 8, no. 7, pp. 1–18, 2008.

[4] M. M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S. M. Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, March 2015.

[5] F. Perazzi, P. Krhenbhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 733–740.

[6] B. Schlkopf, J. Platt, and T. Hofmann, *Graph-Based Visual Saliency*. MIT Press, 2007, pp. 545–552. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6287326

[7] C. Yang, L. Zhang, H. Lu, X. Ruan, and M. H. Yang, "Saliency detection via graph-based manifold ranking," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, June 2013, pp. 3166–3173.

[8] Y. Zhang, G. Jiang, M. Yu, and K. Chen, "Stereoscopic visual attention model for 3d video," in *International Conference on Advances in Multimedia Modeling*, 2010, pp. 314–324.

[9] C. Chamaret, S. Godeffroy, P. Lopez, and O. L. Meur, "Adaptive 3d rendering based on region-of-interest," *Proceedings of SPIE - The International Society for Optical Engineering*, vol. 7524, 2010.

[10] Y. Niu, Y. Geng, X. Li, and F. Liu, "Leveraging stereopsis for saliency analysis," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, June 2012, pp. 454–461.

[11] Y. Fang, J. Wang, M. Narwaria, P. L. Callet, and W. Lin, "Saliency detection for stereoscopic images," *IEEE Transactions on Image Processing*, vol. 23, no. 6, pp. 2625–2636, June 2014.

[12] H. Peng, B. Li, W. Xiong, W. Hu, and R. Ji, "Rgbd salient object detection: A benchmark and algorithms," *National Laboratory of Pattern Recognition*, vol. 8691, pp. 92–109, 2014.

[13] L. Ferreira, L. A. da Silva Cruz, and P. Assuncao, "A method to compute saliency regions in 3d video based on fusion of feature maps," in *2015 IEEE International Conference on Multimedia and Expo (ICME)*, June 2015, pp. 1–6.

[14] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[15] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime tv-l 1 optical flow," *Pattern Recognition*, pp. 214–223, 2007.

[16] J. S. Pérez, E. Meinhardt-Llopis, and G. Facciolo, "Tv-l1 optical flow estimation," *Image Processing On Line*, vol. 2013, pp. 137–150, 2013.

[17] B. Scholkopf, J. Platt, and T. Hofmann, *Graph-Based Visual Saliency*. MIT Press, 2007, pp. 545–552. [Online]. Available: http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6287326

[18] K. Mller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, F. H. Rhee, G. Tech, M. Winken, and T. Wiegand, "3d high-efficiency video coding for multi-view video and depth data," *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3366–3378, Sept 2013.

[19] M. Urvoy, M. Barkowsky, R. Cousseau, Y. Koudota, V. Ricorde, P. L. Callet, J. Gutirrez, and N. Garca, "Nama3ds1-cospad1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3d stereoscopic sequences," in *2012 Fourth International Workshop on Quality of Multimedia Experience*, July 2012, pp. 109–114.

[20] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 1597–1604.

[21] C. Yang, L. Zhang, and H. Lu, "Graph-regularized saliency detection with convex-hull-based center prior," *IEEE Signal Processing Letters*, vol. 20, no. 7, pp. 637–640, July 2013.

[22] W. Wang, J. Shen, and F. Porikli, "Saliency-aware geodesic video object segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition*, June 2015, pp. 3395–3402.

[23] L. Wang, L. Wang, H. Lu, P. Zhang, and X. Ruan, "Saliency detection with recurrent fully convolutional networks," in *European Conference on Computer Vision*. Springer, 2016, pp. 825–841.

PLACE
PHOTO
HERE

Michael Shell Biography text here.

John Doe Biography text here.

Jane Doe Biography text here.