Bottom-up Fusion for Complementary Cues From Multiple Segmentations

No Author Given

No Institute Given

Abstract. Image segmentation is used as a vital prerequisite in several computer vision applications, including object tracking, object recognition and object localization. However, Image segmentation is known to be unstable because it is highly affected by image perturbations such as shadows, shading and highlights. In our work, we combine different cues from multiple segmentations of each image in a bottom up fashion to obtain better segmentations. These resulting segmentations are better in terms of the quality of the segments and the information the segments carry about the objects in the image. Our idea was inspired by the *object class segmentation* problem but can be extended and applied to other applications as well.

1 Introduction

Image segmentation is a computer vision process focusing on partitioning an image into a set of non-overlapping regions. This is an extremely challenging task for real images. The shape variations of the objects provoke several effects related with the illumination such as shadows, shadings and highlights. These effects are one of the main problems that should be solved in order to obtain an efficient segmentation.

Image segmentation algorithms can be divided in several ways, however, all of the existing approaches can first be divided into two main hierarchies: bottom-up approaches and top-down approaches.

Bottom-up segmentation approaches mainly examine the image and try to figure out how to divide it into coherent and meaningful segments. Comprehensive surveys [3] drew the basis for the current classification of bottom-up segmentation techniques. From all of these existing methods, segmentation methods can be divided into four main categories: feature-based, image-based, physics based and hybrid approaches.

Feature based approaches mainly focus on the photometric information of an image represented by its histogram [1]. Image-based approaches exploit the spatial coherence of color in an image [7]. Physics-based methods use physics and psychophysics information to perform the segmentation. Finally, hybrid techniques combine methods of the previous categories.

Top-down approaches, on the other hand, are guided primarily by high-level information and the use of class-specific criteria. The motivation for using these

class-specific criteria, as shown by [2], has two parts. The first is that although recent image segmentation algorithms provide impressive results, they still often fail to capture meaningful and at times crucial parts of the objects in the image. The second is that these methods are analogous to human vision in the sense of indicating high-level, class-based criteria to segment the images in a meaningful manner. Examples of using top-down approaches to perform segmentation are shown in [2, 11].

Image segmentation has been applied to solve a variety of computer vision problems. A robust and efficient segmentation is an important preprocessing step for several computer vision tasks. However, extensive experiments [19] show that when using a single region generated by an image segmentation algorithm, the segmentation quality is highly variable and dependant on image data, the segmentation algorithms and the parameters used to create this segmentation. This was a motivation for the emerging new trend in object recognition that uses the segments generated from multiple segmentation algorithms and tries to merge them efficiently to recognize objects in the scene [16, 14].

In our work, instead of combining the segments generated from multiple segmentation algorithms for object recognition in a top-down fashion, we combine these different segmentations to build a robust and efficient segmentation method that is built in a bottom-up fashion. We combine these segmentations by carefully selecting the segments from each segmentation method and evaluating them using the criteria that we call "goodness measure".

2 Generating Multiple Segmentations

Different bottom-up image segmentation techniques use unique criteria to construct segments. Coherence in the color of the image's segments is a cue that many segmentation techniques rely upon. Another cue is the existence of the edges to separate the segments (or objects) in the image. Also spatial coherence of segments of the same object is an important factor that other segmentation techniques consider. In order to build a more reliable segmentation, we construct different segmentations for each image where each one contains at least one of the properties that we desire in our final segmentation. However, these properties cannot be obtained from a single segmentation method.

We generate a segmentation based on mean-shift [4]. In this case, we guarantee that the segments mainly vary in the color distribution. We also generate another segmentation based on the graph-based method [6] to guarantee a variation along the edges. Moreover, we generate a normalized cut [18] segmentation to guarantee some large segments that are spatially coherent based on efficient clustering techniques. Using these segmentations, we obtained most of the qualities that we desire in our final segmentation.
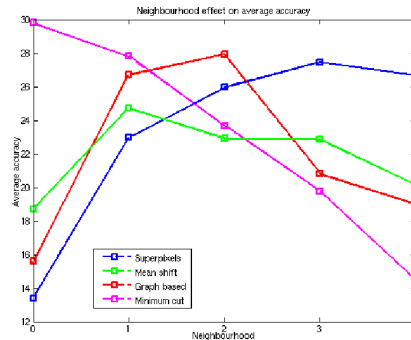
Fig. 1. Effect of increasing neighborhood values on the average accuracies (Best viewed in colors).

3 Bottom-Up Fusion of Multiple Segmentations

In this section we propose the idea of combining multiple segmentations in a bottom-up fashion to generate a robust and a more efficient image segmentation. This motivation was mainly guided by the *object class segmentation* problem where the focus is to build a segmentation that segments semantic objects inside the image.

Our approach is briefly described as follows: (i) Select a certain segment from the set of segments generated from all image segmentations, (ii) Include the segment into our final segmentation if it passes a certain criteria. We give this criteria the "segment goodness" name. Repeat (i) and (ii) until the image is fully segmented. From the brief explanation of our approach, for each task that requires an efficient image segmentation, we aim to focus on two important issues. The first one is how to select the segments from a set of segments. In addition, we have to determine a criteria for determining whether this segment passes the "goodness criteria" or not.

Our technique is inspired by the work proposed in [8]. The authors enhance the *object class segmentation* accuracy by generating an over-segmentation for each image resulting into the image's "superpixels" [15]. Afterwards, they try to aggregate the histograms from neighboring superpixels to give more information about each superpixel based on its neighbors. In other words, the work in [8] mainly increases the information around each segment (superpixel in this case) by increasing the segment's size (by aggregating neighboring superpixels).

We applied more experiments using the same framework in [8] on the PASCAL VOC 2007 segmentation dataset. In our experiments we used the segments generated from the different generated segmentations (see section 2) instead of superpixels. For these segmentations, we also investigated in increasing the neighborhood values. The results are illustrated in figure 1

Figure 1 illustrates two important conclusions. First, large segments improve the results. This conclusion is based on the observation of increasing the overall average accuracy in case of (i) the large segments generated by the normalized

4

cut technique, and (ii) the generated segments using the superpixels or mean-shift methods when taking larger neighborhood size into account.

The second conclusion is that up to a certain limit, increasing the size of the segments or the neighborhood, decrease the final average accuracy. We attribute this to the loss of semantic information due to the combination of segments that are no longer related to each others. Consequently, our target is to find a criteria to decide whether combining smaller segments or splitting larger ones is more useful.

After we had these conclusions we aim to apply our proposed approach for combining multiple segmentations. Assume we have N segmentations S_i, where $i = 1, ..., N$ and each of the segmentations correspond to $S_i = \bigcup_{j=1}^{C_i} T_{ij}$, where T_{ij} is a segment in the segmentation S_i and C_i is the number of segments obtained from the segmentation S_i. Our objective is to obtain a better segmentation $S^* = \bigcup_{k=1}^{C^*} T_k^*$, where T_k^* is a segment in our final set of optimized segments and C^* is the total number of optimized segments. Segmentation S^* should satisfy the following criteria:

1. The segments of the final segmentation should be as large as possible. In other words, we need to minimize C^*.
2. The segments should be "good" segments. In our context, we defined a "good segment" as the largest set of connected pixels that belong to the same object.

Consequently, our motivation for combining segmentations is to start looking for the large segments and then check if this segment is a "good" segment. We applied a greedy algorithm that favors large segments and check if they pass a certain criteria of "goodness". Our proposed approach is described as follows: Examine the largest segment obtained from all segmentations methods and check whether this segment is a "good" segment. If the segment is "good", we include it in our final segmentation (Add to S^*). Otherwise, we try to segment it recursively using the least number of segments from the other segmentations. Similarly, we segment the remaining part of the image.

To this end, we need to determine whether the generated segments are "good" or not. To do this, we investigate several "goodness" evaluation criteria. First, we studied the possibility of assuming that the color of each region should follow a normal distribution. However, this "goodness" criteria didn't hold on real images as shown in figure 2(c). Second, we examined the unimodality test to check whether the distribution of the histogram of each RGB channel follow a unimodal function. We use Dip test [10] for unimodality. Although the resulting segments were slightly better than those obtained using the normality test, it also fails to handle real images' segments as in figure 2(c). In addition, we also investigated the outliers detection and the Ridge based Distribution Analysis (RAD) methods. We'll describe these techniques in more details in the next sections.
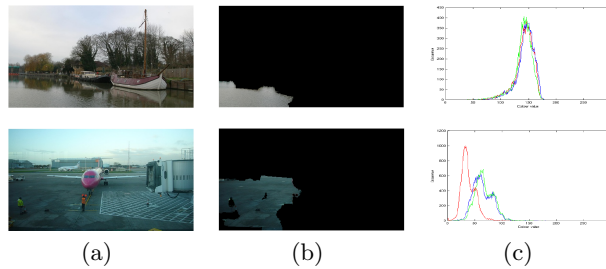
Fig. 2. (a) Original images (b) Current largest segments (c) Histograms from every channel. As we can see the first image can be approximated into a normal distribution or a unimodal distribution while the second one fails in that test.

### 3.1 Good Segments Evaluation Using Outliers Finding

As defined by [9], "An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs." In our work, we consider a segment to be "good" if the fraction of outliers inside this segment is below a certain threshold. We performed a series of experiments to empirically determine the best value for the tolerance of outliers inside the segment. Results have shown that a segment is considered "good" if the number of outliers lying inside is less than 2% of the number of pixels in the segment.

Although outliers detection worked well in detecting the segments shown in Figure 2(b), they didn't work on detecting the segments that contain variations in shadows and highlights. Consequently, we examined other evaluation methods that work well in the presence of shadows and illuminations. For this purpose, we use the RAD evaluation method due to its robustness to shadows and highlights.

3.2 Good Segments Evaluation Using RAD

We use the RAD method [20] for obtaining image segmentations in the presence of shadows and highlights. This method is based on the insight that the distributions formed by a single-colored object have a physically determined shape (i.e. follow the same ridge) in the color histogram space [17]. To capture these ridges Vazquez et al. [20] proposed a new Ridge based Distribution Analysis (RAD) to find the set of ridges representative of the dominant colors in the image.

We model the region "segment" by its set of dominant colors (DC). This DC is described by a distribution in histogram-space. We assume that if this segment contains only 1 DC then this segment is a "good" segment as this segment contains only 1 semantic object. Based on this technique, the evaluation method becomes very robust to shadows and highlights and becomes closely related to the physical features of each segment than the previously investigated methods.
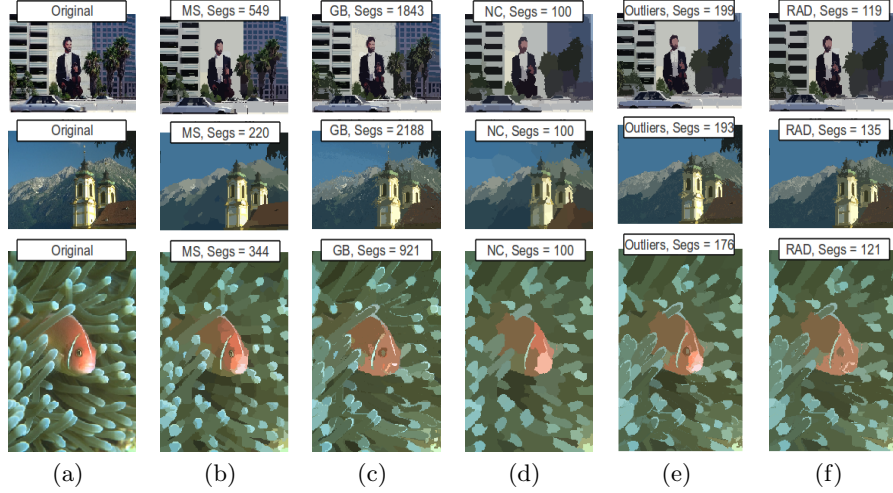
Fig. 3. (a) Original image. (b) Mean-shift segmentation. (c) Graph-Based segmentation. (d) Normalized-Cut segmentation. (e) Evaluation with Outliers. (f) Evaluation with RAD.

4 Experimental results

In figure 3 we show some qualitative results to the images resulting from the various segmentation techniques as well as the two ways discussed for combining segments and evaluating segments quality.

We also perform a series of experiments on the standard Berkeley Segmentation Dataset and Benchmark [12]. We evaluate the performance of the three considered segmentation methods with our newly proposed segmentation methods. Our method combines the complementary information obtained by all the considered segmentation methods using the Outliers and RAD criteria for segments "goodness" evaluation. Our combination method's main motivation is getting the largest possible segments as a motivation to improve the task of *object class segmentation*.

We evaluate four different error measurements for evaluating the quality of the segments generated by our approach. These four error measures are: First, the Probabilistic Rand Index (PRI) [19], which counts the fraction of pairs of pixels whose labellings are consistent between the computed segmentation and the ground truth, averaging across multiple ground truth segmentations to account for scale variation in human perception. The second error measure is the Variation of Information (VoI) [13]. It defines the distance between two segmentations as the average conditional entropy of one segmentation given the other, and thus roughly measures the amount of randomness in one segmentation which cannot be explained by the other. The third is the Global Consistency Error (GCE) [12] that measures the extent to which one segmentation can be viewed as a refinement of the other. Segmentations which are related in this manner are considered

	PRI	VoI	GCE	BDE
Mean-Shift	0.7424	4.5293	**0.0842**	**14.2716**
Graph-Based	0.7082	5.1148	0.1150	17.2428
Normalized-cut	0.7079	4.1370	0.1153	14.7337
Combine + Outliers Evaluation	0.7462	3.7264	0.1380	14.7730
Combine + RAD Evaluation	**0.7483**	**3.5364**	0.1433	14.8047

Table 1. Segmentation benchmarks results for the Berkeley Segmentation Dataset and Benchmark. 4 error measures were considered for 4 different segmentations.

to be consistent, since they could represent the same natural image segmented at different scales. The fourth measure is the Boundary Displacement Error (BDE) by [7]. It measures the average displacement error of boundary pixels between two segmented images. Particularly, it defines the error of one boundary pixel as the distance between the pixel and the closest pixel in the other boundary image.

Table 1 demonstrates that our proposed segmentation method improves both of the PRI and the VoI error measures. Whether evaluating our segments quality with outliers or RAD evaluation criteria. We attribute this to the fact that PRI and VoI are clearly concerned about comparing segments entirely with the ground truth segments. This justifies our motivation behind our proposed segmentation method which is based on improving the size and the quality of the segment. In contrary, GCE and BDE measures don't perform well on our approach. These measures are mainly concerned about the boundaries of the segments which our approach doesn't take into concern.

5 Conclusions

We have demonstrated the effect of using multiple segmentations on improving the segments and hence improving the applications that use image segmentations as a prerequisite. Our approach aims to select the "good" segments from amongst a set of segmentations. We've investigated two criteria for measuring segments' goodness. We believe we can still improve our resulting segmentations by investigating more criteria for measuring goodness. For instance, for the JSEG segmentation [5], the criterion chosen for good segmentations evaluation is the spatial relation existing between the pixels in the image space. Another way is to use the color boosting algorithm introduced in [21] and consider a segment "good" if the number of salient points is below a certain threshold.

References

1. Agarwal, S., Madasu, S., Hanmandlu, M., Vasikarla, S.: A comparison of some clustering techniques via color segmentation. In: Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'05) (2005)

2. Borenstein, E., Ullman, S.: Class-specific, top-down segmentation. In: ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part II (2002)
3. Cheng, H.D., Jiang, X.H., Sun, Y., Wang, J.L.: Color image segmentation: Advances and prospects. Pattern Recognition 34, 2259–2281 (2001)
4. Comaniciu, D., Meer, P., Member, S.: Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 603–619 (2002)
5. Deng, Y., Manjunath, B.S.: Unsupervised segmentation of color-texture regions in images and video. Pattern Analysis and Machine Intelligence 23, 800–810 (2001)
6. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. International Journal of Computer Vision 59 (2004)
7. Freixenet, J., noz Xavier Mu Raba, D., Martí, J., Cufí, X.: Yet another survey on image segmentation: Region and boundary information integration. In: ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part III. pp. 408–422 (2002)
8. Fulkerson, B., Vedaldi, A., Soatto, S.: Class segmentation and object localization with superpixel neighborhoods. In: Proc. ICCV (2009)
9. Grubbs, F.E.: Procedures for detecting outlying observations in samples. Technometrics 11, 1–21 (1969)
10. Hartigan, J.A., Hartigan, P.M.: The dip test of unimodality. The Annals of Statistics 13 (1985)
11. Leibe, B., Leonardis, A., Schiele, B.: Combined object categorization and segmentation with an implicit shape model. In: ECCV workshop on statistical learning in computer vision. pp. 17–32 (2004)
12. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proc. 8th Int'l Conf. Computer Vision. vol. 2, pp. 416–423 (July 2001)
13. Meila, M.: Comparing clusterings by the variation of information. pp. 173–187 (2003)
14. Pantofaru, C., Schmid, C., Hebert, M.: Object recognition by integrating multiple image segmentations. In: European Conference on Computer Vision. vol. III, pp. 481–494 (2008)
15. Ren, X., Malik, J.: Learning a classification model for segmentation. In: In Proc. 9th Int. Conf. Computer Vision. pp. 10–17 (2003)
16. Russell, B.C., Efros, A.A., Sivic, J., Freeman, W.T., Zisserman, A.: Using multiple segmentations to discover objects and their extent in image collections. In: Proceedings of CVPR (June 2006)
17. Shafer, S.A.: Using color to separate reflection components. Color Research & Application 10(4), 210–218 (1985)
18. Shi, J., Malik, J.: Normalized cuts and image segmentation. Pattern Analysis and Machine Intelligence, IEEE Transactions on 22 (2000)
19. Unnikrishnan, R., Pantofaru, C., Hebert, M.: Toward objective evaluation of image segmentation algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 929–944 (2007)
20. Vazquez, E., Weijer, J., Baldrich, R.: Image segmentation in the presence of shadows and highlights. In: ECCV '08: Proceedings of the 10th European Conference on Computer Vision. pp. 1–14 (2008)
21. Weijer, J.V.D., Gevers, T., Bagdanov, A., Edge, I.A., Detection, F., Statistical, I.D.: Boosting color saliency in image feature detection. IEEE Trans. Pattern Analysis and Machine Intelligence 28, 150–156 (2005)