Image Segmentation and Evaluation

Ning Tang

College of Information Science and Engineering, Ocean University of China, Qingdao, China

Abstract—Image segmentation is a fundamental problem in computer vision. Despite many years of research, general purpose image segmentation is still a challenging task. Traditional image segmentation schemes are categorized into several classes: threshold based techniques, edge based techniques, region based techniques, graph based techniques, active contour based techniques, neural network based techniques, etc. In this paper, we adopt five image segmentation methods, including Otsu's thresholding method, Canny based segmentation method, Watershed method, Mean Shift (MS), and Normailzed cut (NCut). To evaluate and compare the performance of these five methods, we apply three evaluation measures — Probabilistic Rand Index (PRI), Variation of Information (VoI), and Global Consistency Error (GCE).

1 Introduction

Image segmentation is a classical and fundamental problem in computer vision. It refers to partitioning an image into several disjoint subsets such that each subset corresponds to a meaningful part of the image. As an integral step of many computer vision problems, the quality of segmentation output largely influences the performance of the whole vision system. A rich amount of literature on image segmentation has been published over the past decades. Some of them have achieved an extraordinary success and become popular in a wide range of applications, such as medical image processing, object tracking, recognition, image reconstruction and so on.

In recent years, plenty of efforts have been focusing on the segmentation process. Numbers of different segmentation techniques are presented, but there is not even a one single method to be considered as a best method for different kind of images, only suitable for one specific type of images. Hence, image segmentation is still a difficult task in computer vision.

The outline of the paper is as follows. In Section 2, the theoretics of the five methods we adopt are briefly reviewed. In Section 3, some quantitative metrics of the segmentation quality are described. The performances of five segmentation techniques are analyzed in Section 4. Section 5 draws the conclusion.

2 Methods

In this Section, we briefly review the following five image segmentation methods:

- Otsu's thresholding method [6]
- Canny based segmentation method [1]
- Watershed method
- Mean Shift (MS) [2]
- Normailzed Cut (NCut) [7]

2.1 Otsu's Thresholding Method

Otsu's thresholding [6] is used to automatically perform clustering-based image thresholding, or, the reduction of a graylevel image to a binary image. The algorithm assumes that the image contains two classes of pixels following bi-modal histogram (foreground pixels and background pixels), it then calculates the optimum threshold separating the two classes so that their intra-class variance is minimal.

In Otsu's method, the intra-class variance is defined as a weighted sum of variances of the two classes:

$$\sigma_w^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t) \tag{1}$$

Weights ω_0 and ω_1 are the probabilities of the two classes separated by a threshold t and σ_0^2 and σ_1^2 are variances of these two classes.

The class probability $\omega_{0,1}(t)$ is computed from the L histograms:

$$\omega_0(t) = \sum_{i=0}^{t-1} P(i) \tag{2}$$

$$\omega_1(t) = \sum_{i=t}^{L-1} P(i) \tag{3}$$

Otsu shows that minimizing the intra-class variance is the same as maximizing inter-class variance:

$$\begin{aligned}\sigma_b^2 &= \sigma^2 - \sigma_w^2(t) = \omega_0(\mu_0 - \mu_T)^2 + \omega_1(\mu_1 - \mu_T)^2 \\ &= \omega_0(t)\omega_1(t)[\mu_1(t) - \mu_0(t)]^2\end{aligned} \tag{4}$$

which is expressed in terms of class probabilities ω and class means μ.
While the class mean $\mu_{0,1,T}(t)$ is:

$$\mu_0(t) = \sum_{i=0}^{t-1} iP(i)/\omega_0 \tag{5}$$

$$\mu_1(t) = \sum_{i=t}^{L-1} iP(i)/\omega_1 \tag{6}$$

$$\mu_T = \sum_{i=0}^{L-1} iP(i) \tag{7}$$

The following relations can be easily verified:

$$\omega_0\mu_0 + \omega_1\mu_1 = \mu_T \tag{8}$$

$$\omega_0 + \omega_1 = 1 \tag{9}$$

The class probabilities and class means can be computed iteratively.

2.2 Canny Based Segmentation Method

The key of Canny based segmentation method is extracting useful edges by using Canny edge detection[1]. Edge detection methods are carried out based on abrupt changes in intensity levels or grey levels of an image, in these methods our interest mainly focus on identification of isolated points, lines and edges. Among the edge detection methods developed so far, Canny edge detection algorithm is one of the most strictly defined methods that provides good and reliable detection. Once we get continuous edges, some morphological operations could be used to get the results in the form of regions.

2.3 Watershed Method

Watershed techniques considered the gradient of an image (GMI) as a topographic surface. Pixels having the highest GMI correspond to watershed lines, which represents region boundaries some positive points of watersheds are by this method segmentation results are stable, they do not depend on any threshold and secondly the region boundaries are formed naturally out of the process. The boundaries are continuous and there are no gaps. Negative point considered over segmentation.

2.4 Mean Shift

Mean Shift (MS) [2] is considered a robust technique used for image segmentation, visual tracking etc. Mean shift method is an iterative mode detection algorithm in the density distribution space or a tool for finding modes in a set of data samples.

Mean shift procedure is as follows:

1. Find a window around each data point.
2. Compute the mean of data with in the window.
3. Translate density estimation window.
4. Shift the window to the mean and repeat till convergence

2.5 Normailzed Cut

Normailzed Cut (NCut) [7] is a well-known graph based segmentation method. The graph cut is measured by the weights of $vol(.)$, which is the total connection from vertices in a set (e.g., A) to all the vertices in the graph. Formally we have $vol(A) = \sum_{v_i \in A, vj \in V} w(vi, vj)$, where weight $w(vi, vj)$ measures a certain image quantity (e.g., intensity, color, etc.) between the two vertices connected by that edge. Then Normailzed Cut (NCut) cost function is defined as follows:

$$\begin{aligned} NCut(A,B) &= \frac{Cut(A,B)}{vol(A)} + \frac{Cut(A,B)}{vol(B)} \\ &= \frac{\sum_{(x_i>0,xj<0)} -w(vi,vj)x_i x_j}{\sum_{x_i>0} d_i} + \frac{\sum_{(x_i<0,xj>0)} -w(vi,vj)x_i x_j}{\sum_{x_i<0} d_i} \end{aligned} \tag{10}$$

where x_i is the indicator variable, $x_i = 1$ if vertex v_i is in A and $x_i = 1$ otherwise. $di = \sum_j w(vi, vj)$ is the total connection from v_i to all the other vertices. Note that with this definition, the partitions containing small set of vertices will not have small Ncut value, and hence the minimal cut bias is circumvented. The minimization of Eq. (10) can be formulated into a generalized eigenvalue problem, which has been well-studied in the field of spectral graph theory. After a common matrix transformation, the NCut problem can be re-written into:

$$\min NCut(A,B) = \min_y \frac{\mathbf{y}^T(\mathbf{D}-\mathbf{W})\mathbf{y}}{\mathbf{y}^T\mathbf{D}\mathbf{y}} \tag{11}$$

subject to $\mathbf{y}(i) \in \{1, -b\}$, $b = \frac{\sum_{x_i>0} d_i}{\sum_{x_i<0} d_i}$ and $\mathbf{y}^T\mathbf{D}\mathbf{1} = 0$, where $\mathbf{D}$ and $\mathbf{W}$ are the degree matrix and the adjacency matrix of G, respectively. We call $\mathbf{L} = \mathbf{D} - \mathbf{W}$ the graph Laplacian of G. It can be seen that $-b$ represents the ratio of connections which are from v_i to vertices inside and outside the same set, respectively. The relaxed optimization of Eq. (11) is obtained by discarding the discreteness condition but allowing $\mathbf{y}$ to take arbitrary real values.

3 Evaluation of Image Segmentation Methods

In previous sections, we have reviewed briefly five image segmentation methods. It is well-known that image segmentation is an ill-posed problem, which makes the evaluation of a candidate algorithm a very challenging work. The most usual way of evaluation is to visually observe different segmentation results by the user. However, it is time consuming and may result in different outcomes by users. Quantitative evaluation of segmentation is hence more preferable in practice. In supervised evaluation, the task is performed by measuring the similarity between the segmentation results and some ground truth images, which are provided by human observers. This has been widely used by researchers.

To quantitatively evaluate the segmentation results, we use three well-known indices: Probabilistic Rand Index(PRI) [8], Variation of Information (VoI) [5], and Global Consistency Error (GCE) [3].

Probabilistic Rand Index(PRI). The PRI defines the correctness of segmentations under a statistical point of view. It is supposed that the segmentation of an image can be described in the form of binary numbers $\mathbf{I}(l_i^{S_k} = l_j^{S_k})$ on each pair of pixels (x_i, x_j). The distribution of these numbers follows a Bernoulli distribution and gives a random variable with expected value denoted by p_{ij}. The PRI of two segmentations is then defined as:

$$PR(S_{test}, \{S_k\}) = \frac{1}{\binom{N}{2}} \sum [\mathbf{I}(l_i^{S_{test}} = l_j^{S_{test}}) p_{ij} + \mathbf{I}(l_i^{S_{test}} \neq l_j^{S_{test}})(1 - p_{ij})] \tag{12}$$

where N is the number of pixels, $\{S_k\}$ is the set of ground truth segmentations, p_{ij} is the ground truth probability that the labels of (x_i, x_j) are the same. In practice, the mean pixel pair relationship in all ground truth segmentations is used to compute p_{ij}. We could see that the penalization of segmentation for being/not-being in the same region is dependent on the fraction of disagreeing with the ground truth data. The PRI takes values in the range $[0, 1]$, where a score of zero indicates the labeling of test image is totally opposite to the ground truth segmentation and 1 indicates that they are the same on every pixel pair. The PRI accommodates the region refinements appropriately in that it accepts refinement only in regions that human observers find ambiguous. This property is more preferable than the refinement-invariant measures for preventing the degenerate cases.

Variation of Information (VoI). Meila [5] proposed an information-theoretic distance of clustering. For segmentations, it can be interpreted as the average conditional entropy of one segmentation given the other:

$$VoI(S_{test}, S_k) = H(S_{test}|S_k) + H(S_k|S_{test}) \tag{13}$$

The first term in Eq. (13) measures the amount of information about S_{test} that we loose, while the second term measures the amount of information about S_k that we have to gain, when going from segmentation S_{test} to ground truth S_k . An equivalent expression of Eq. (13) is:

$$VoI(S_{test}, S_k) = H(S_{test}) + H(S_k) - 2I(S_{test}, S_k) \tag{14}$$

where H and I are respectively the entropies of and the mutual information between the segmentation S_{test} and the ground truth S_k. VoI is a distance metric since it satisfies the properties of non-negativity, symmetry and triangle inequality. If two segmentations are identical, the VoI value will be zero. The upper bound of VoI is finite and depends on the number of elements in the segments.

Global Consistency Error (GCE). This evaluation criterion is designed for computing the degree of overlap of regions. Martin et al. [3] proposed the GCE measure to quantify the segmentation quality in different granularities. This measure allows for refinement, but suffers from degeneracy. Let $R(S, p_i)$ be the set of pixels in segmentation S that contains pixel p_i, the local refinement error is defined as:

$$E(S_1, S_2, p_i) = \frac{|R(S_1, p_i) \backslash R(S_2, p_i)|}{R(S_1, p_i)} \tag{15}$$

This error is not symmetric w.r.t. the compared segmentations, and takes the value of zero when S_1 is a refinement of S_2 at pixel p_i. GCE is then defined as:

$$GCE(S_1, S_2) = \frac{1}{n} \min \left\{ \sum_i E(S_1, S_2, p_i), \sum_i E(S_2, S_1, p_i) \right\} \tag{16}$$

From the above introduction of the three indices, one should note that it is not possible to define a criterion for comparing segmentations that fits every problem optimally. For example, PRI is based on examining the relationship between pairs of pixels. As a result, segmentation algorithms which are concerned with pairs (e.g., graph partitioning) can better use PRI for evaluation. While for clustering algorithms (e.g., Mean Shift) focus on the relationship between a point and its clustering centroid, VoI will be a better choice. A good segmentation will achieve large value of PRI while small values of GCE, and VoI.

4 Experiments on Image Segmentation

In this section, five well-known segmentation methods are selected for our experiments: Otsu's thresholding method [6], Canny based segmentation method [1], Watershed method, Mean Shift [2], and Normailzed cut [7]. The evaluation also shows the consistency of segmentation quality produced by them.

All the experiments were performed on the Berkeley test set [4], where all of the 200 images are used for our evaluation. These images have 5 to 8 human-marked ground truths on each one of them.

Particularly, for each image, we choose the same parameters in MS and NCut. In MS method, the spatial band width was set to 30, the color band width was set to 20. For NCut method, the number of segments was the only parameter and it was set to 10. Fig. 1 shows the segmentation results on 5 test images. We can see that MS and NCut could get better segmentation results than others. Table 1 presents the quantitative measures of segmentation quality on all results produced by five segmentation methods. We see that NCut and MS have the highest PRI score above 0.74, which demonstrates it has best performance on image segmentation. When evaluating the relationship of pixel pairs, NCut has stronger ability to segment the given images than other methods. However, results by GCE show that MS outperforms NCut for producing more good segmentations, and NCut has the worst performance.

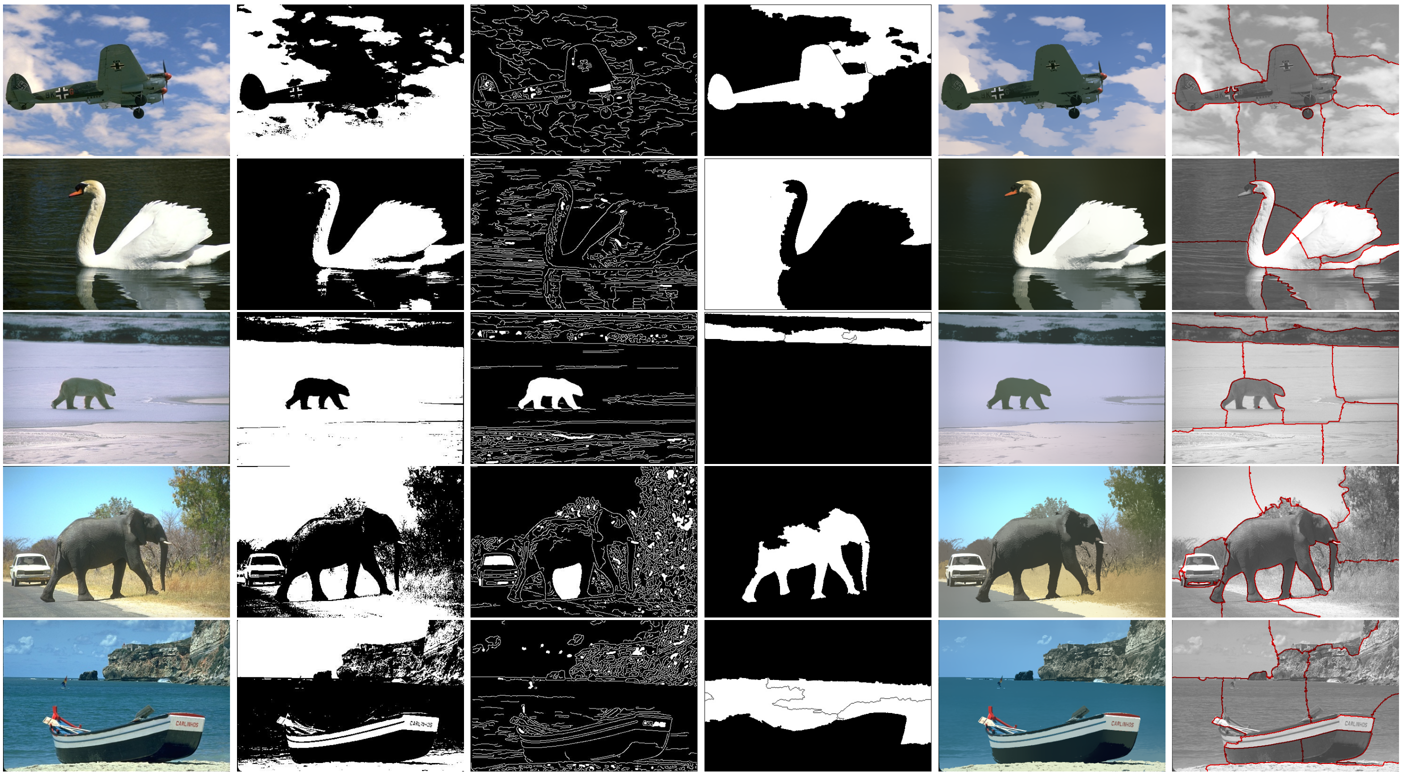

Fig. 1: Some segmentation results on the BSDS500 test set. From Left to Right: Image, and results obtained by Otsu's thresholding method, Canny based segmentation method, Watershed method, Mean Shift, and Normailzed cut.

5 Conclusion

In this paper, we use five well-known methods for image segmentation and three measures for evaluation of segmentation quality. The final evaluation results show that it is not easy to find a single quantitative measure for evaluating the

Algorithms	PRI	VoI	GCE
Otsu's thresholding method	0.6018	2.7267	0.2757
Canny based segmentation method	0.4116	3.4520	0.2465
Watershed method	0.5051	2.4327	0.1510
Mean Shift	0.7109	10.3534	0.0946
Normailzed cut	0.7465	2.5726	0.3051

Table 1: PRI, VoI, and GCE scores on the total segmentation results by Otsu's thresholding method, Canny based segmentation method, Watershed method, Mean Shift, and Normailzed cut, respectively.

segmentation quality, even with a group of ground truth given beforehand. To some extent, the evaluation criterion might vary in different applications.

References

[1] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986.

[2] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[3] D. R. Martin. An empirical approach to grouping and segmentation. In *International Symposium on Physical Design*. University of California, Berkeley, 2002.

[4] D. R. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *Proc. Intl Conf. Computer Vision*, 2:416–423, 2002.

[5] M. Meil. Comparing clusterings: an axiomatic view. In *Proceedings of the 22Nd International Conference on Machine Learning*, ICML '05, pages 577–584. ACM, 2005.

[6] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.

[7] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[8] R. Unnikrishnan and M. Hebert. Measures of similarity. In *Proceedings of Seventh IEEE Workshops on Application of Computer Vision*, volume 1 of *WACV/MOTION '05*, pages 394–400. IEEE Computer Society, 2005.