Image segmentation towards natural clusters

Zhigang Tan, Nelson H.C. Yung
Dept. EEE, University of Hong Kong
{zgtan, nyung}@eee.hku.hk

Abstract

To find how many clusters in a sample set is an old yet unsolved problem in unsupervised clustering. Many segmentation methods require the user to specify the number of regions in the image or some delicate thresholds to get a sensible segmentation. In this paper, we propose a segmentation method that is able to automatically determine the number of regions in an image. The method effectively discerns distinct regions by analyzing the properties of the joint boundary between neighboring regions. By requiring that every region should be distinct from each other, it is able to choose a natural partition from the partition set which contains all possible partitions. Results are given at the end of this paper to demonstrate the effectiveness of this approach.

1. Introduction

Image segmentation, a technique that divides the input image into non-overlapping homogeneous regions, is a fundamental operation in many content-based applications of computer vision.

Among myriads of existing image segmentation techniques, many can be tagged as unsupervised clustering methods. For example, region merging is the analogue of agglomerative/hierarchical clustering[1]. Graph cut methods such as minimal cut and normalized cut represent the clustering problem in a graph theoretic manner. A notorious problem known as the problem of validity for this kind of methods is how to determine the number of clusters in an image. Since the problem is largely unresolved, most techniques need the user to provide a stop condition.

Hierarchical merging methods [2-4] need to explicitly specify the number of regions in the input image or alternatively to set a threshold. In our investigation, we find that the threshold is often related to size of the image, size of the objects, as well as noise level, which are often unknown without the knowledge of the ground truth and is hard to control.

In normalized cut[5, 6], the method bipartition the image gradually until a certain stop cut value is met. However, the stop cut value, i.e. the ratio of the sum of all inter edges to that of inter and intra edges, is also determined by the structure of the image and cannot be determined before ground truth is obtained.

This paper is an attempt to solve the problem of validity in the field of image segmentation. The idea is to resolve an easier case to tell when two regions are "distinct enough". Once such a method is established, it is possible to obtain a natural grouping by requiring every region to be distinct from their neighbors,

The proposed method starts with M initial segments of an over-segmented image. Then a sequence of merging operations is applied on the initial segments and iteratively generates a series of K-regions partitions, $K=1, 2...M$. At each merging, two adjacent regions are joined towards stepwise optimizing a global energy function. Afterwards, the method analyzes joint boundaries of each partition and determines whether each region is distinct from each other (see the details in Section 4). Then a partition with a largest number of regions and within which every region is distinct from each other is chosen as the final segmentation.

The paper is organized in the following way. In Section 2, the mathematical framework for our method is defined. Section 3 introduces how and why the initial segments are generated. Then, in section 4, we explicate how the natural clusters can be obtained. In Section 5, some segmentation results are depicted and compared with graph cut to show its effectiveness.

2. Problem formulation

Let $I:\Lambda \rightarrow R, \Lambda \subset \mathbf{Z}^2$ represents an image. The objective of image segmentation is to find a partition $\{R_1,\cdots,R_K \mid \Lambda = \bigcup_i^K R_i,\ R_i \bigcap R_j = \varnothing,\ \forall i \neq j\}$ which optimizes an objective function or functional. Our functional is

$$E[f,\Gamma \mid K] = \alpha \sum_{i=1}^{K} \sum_{(x,y)\in R_i} (f_i(x,y) - I(x,y))^2 - \lambda \sum_{\Gamma_m \in \Gamma} \sum_{(x,y)\in \Gamma_m} \frac{1}{|\Gamma_m|} \|\nabla I(x,y)\|$$

(1)

where $\Gamma = \{\Gamma \mid \Gamma_m = \partial R_i \cap \partial R_j \neq \varnothing, R_i, R_j \in R\}$ is the segmentation boundaries of the entire image with Γ_m be any joint boundary between two regions. Suppose the optimal number of segment is K *(K is unknown)* and $f=\{f_i\}$ be the output image and is piecewise

constant.

Since K ($K \in N$ and $K \leq M$) is unknown, the proposed method iteratively generates a best partition P_i according to (1) for all K from M to 1, and then chooses one from $\{P_i, \ i=1,2,...,M\}$ based on the criterion that each region is distinct from others.

The partition set is generated by merging two regions at a time towards stepwise optimizing (1). Each time, two regions with minimal value of the merging criterion in the K region partition are merged to generate a K-1 region partition. The merging criterion shown in (2) is obtained by doing just that.

$$\delta(R_i, R_j) = \alpha \frac{\|R_i\| \cdot \|R_j\|}{\|R_i\| + \|R_j\|} (f_i - f_j)^2 + \lambda \frac{1}{|\Gamma|} \sum_{(x,y) \in \Gamma_i} \|\nabla I(x,y)\| \qquad (2)$$

3. Initial segments

In our implementation, the initial segments are produced by a marker-controlled morphological watershed transform of the gradient magnitude image. Please note that the initial oversegmentation is not limited by watershed algorithms. Other methods like quadrilaterals[7] and superpixel[8] can also be used.

Indeed, the purpose of this step is to trim down the searching space for a global optimum, while keeping sufficient boundary information. Exactly, the searching space is reduced from $2^{|A|}$ to $2^{|\Gamma|}$, ($|A|$ is the size of the input image and $|\Gamma|$ is the cardinality of Γ)[1]. For convenience, an initial segment is also called an atom segment in the later part and they are used interchangeably. Atom segments are combined to form regions.

4. When regions are distinct?

In real images, noises and uneven illumination are common. These lead to parts of the same object having different intensities i.e. making them distinct in the intensity space thus produce false boundaries. Moreover, conventional threshold methods on this kind of images usually do not work very well.

To overcome these difficulties, we use a divide-and-conquer strategy to handle the problem. We observe that the initial segments of the oversegmentation result are usually tight and compact in the feature space i.e. their variations are often less than the final regions. Therefore, it should be easier to analyze them at an atom region level. Here, we propose an adaptive method based on the aforementioned idea.

The method first checks all atom segments along the region boundary. Then all neighboring atom segments on the boundary are grouped into atom-segment pairs.

---

[1] In most cases, $|A| >> |\Gamma|$. For a 300×200 image, $|A|$=60000, $|\Gamma|$=200, which means the later searching space is only 2^{-58000} of the original!

For every atom-segment pair, it delineates a variable-width narrow-band-shaped support along the common edges and analyzes properties of the gradient and intensity distribution in the support and classifies whether the common edge is "reliable". Finally, the ratio of the reliable edge pixels can be used to judge whether the two regions are reliably distinct. By dividing the long joint boundary into pieces, this analysis has an adaptive nature to the local content.

An example is given in Figure 1 (a) to show the division of joint boundary, where along the joint boundary of R_i and R_j, both have four atom segments (a, b, c, d for R_i and e, f, h, g for R_j) marked by red dash lines. The joint boundary comprises the common edges of the neighboring atom segments pairs, which are marked by solid blue and yellow lines alternately in Figure 1 (a). An illustration of a w-wide narrow band support of a common edge between atom-segment pair {a1, a2} is shown in Figure 1 (b). The details of analysis of the common edges are given in Section 4.1.
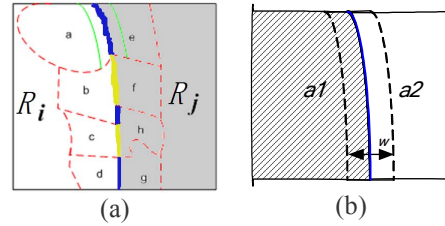

(a)   (b)

Figure 1: (a) Along the joint boundary of R_i and region R_j, shows eight atom segments. The 5 atom-segment pairs are {a, e}, {b, f}, {b, h}, {c, h}, {d, g}. (b) Illustration of support.

4.1. Analysis of supports of common edges

Suppose the true image $I_o(x, y)$ is affected by noise $\eta(x, y)$, then the input image is

$I(x, y) = I_o(x, y) + \eta(x, y)$. (4)

The linear difference gradient along x-axis is

$\Delta I_x(x, y) = I_o(x+1, y) - I_o(x, y) + \eta(x+1, y) - \eta(x, y)$. (5)

Assume $\eta(x, y) \sim N(\mu, \delta^2)$ is uncorrelated with $I(x, y)$ and coordinates (x, y). Equation (5) can be written as

$\Delta I_x = \Delta I_{ox} + \eta_d$, (6)

where $\Delta I_{ox} = I_o(x+1, y) - I_o(x, y)$ and $\eta_d = \eta(x+1, y) - \eta(x, y)$.

It is straightforward to show that $\eta_d \sim N(0, \ 2\sigma^2)$ and ΔI_{ox} is dependent on the image content. There are two kinds of pixels in the support area: edge pixels and non-edge pixels. If pixel (x, y) is not a edge pixel, i.e. it is inside the region and $I(x+1, y) = I(x, y)$, the gradient along x axis should be $\Delta I_{x(non-edge)} = \eta_d$. If pixel (x,y) is an edge pixel, i.e. the adjacent pixel belongs to another region such that $I(x+1, y) \neq I(x, y)$, then $\Delta I_{x(edge)} = \Delta I_{ox} + \eta_d$.

This analysis reveals that $p(\Delta I_x | edge) \sim N(\mu, \ 2\sigma^2)$

and $p(\Delta I_x|non\text{-}edge) \sim N(0, 2\sigma^2)$, where μ is the difference of the intensity of the two regions i.e. $\mu = I(x+1, y) - I(x, y)$. The distribution of gradient ΔI_x in the support area is a mixture of the above two.

$$P(\Delta I_x) = p(\Delta I_{x(non\text{-}edge)}|non\text{-}edge)\, p_{non\text{-}edge}$$
$$+ p(\Delta I_{x(edge)}|edge)\, p_{edge}. \qquad (7)$$

Usually, there are more non-edge pixels than edge pixels i.e. $p(non\text{-}edge) > p(edge)$. Figure 2 (a), (b) show the distributions of $p(\Delta I_x)$ and their the mixture distribution.
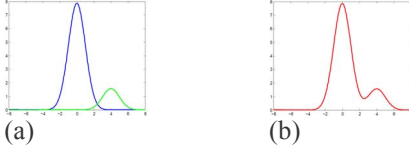

(a)   (b)

Figure 2: Distribution of ΔI_x in the support. (a) $p(\Delta I_x|edge)$ and $p(\Delta I_x|non\text{-}edge)$ *are* shown green and blue. (b) Mixtures of the two.

The parameters $\mu,\ \sigma^2$ can be estimated as

$$\mu = f_i - f_j, \qquad (8)$$

$$\sigma^2 = \frac{1}{|R_i|}\sum_{f\in R_i}(f-f_i)^2 + \frac{1}{|R_j|}\sum_{f\in R_j}(f-f_i)^2, \qquad (9)$$

where f_i and f_j are intensity means of the support of the adjacent regions. p_{edge} is estimated as the ratio of the number of pixels on the common edge to that of the entire support and $p_{non\text{-}edge} = 1 - p_{edge}$.

Let t be the intersection point of the two pdf, then t shown in (10) represents the point where an optimal statistical classifier distinguishes edge pixels and non-edge pixels.

$$t = \frac{\mu}{2} - \frac{\sigma^2(\ln P_{edge} - \ln P_{non\text{-}edge})}{\mu} \qquad (10)$$

It is easy to see that t is linear with σ^2. If σ^2 is small i.e. the two distributions concentrate on their individual centers, t is close to $\mu/2$ and the two distributions are clearly separated. As noises increase, the two peaks move toward each other and the non-edges peak gradually clouds the edge peak. When $|t| > |\mu|$, the edge peak is so seriously affected by the non-edge peak that it is no longer distinguishable from the non-edge peak. Graphically, this equals that the two pdf intersect on the left of the center of the edge peak in Figure 2 (a). When this occurs, the error rate, see (11), of mistaking edge pixels as non-edge pixels is greater than 50%. At this point, we can safely say the edge is not clear or the common edge is not a reliable edge.

$$e = P_{edge}\frac{2}{\sigma\sqrt{\pi}}\int_{|t|}^{\infty}e^{-(x-\mu)^2/\sigma^2}dx + P_{non\text{-}edge}\frac{2}{\sigma\sqrt{\pi}}\int_{|t|}^{\infty}e^{-x^2/\sigma^2}dx. \qquad (11)$$

The condition $|t| > |\mu|$ is equal to

$$\left|\frac{\mu}{2} - \frac{\sigma^2(\ln P_{edge} - \ln P_{non\text{-}edge})}{\mu}\right| > \mu, \qquad (12)$$

which can be further simplified to

$$|\mu/\sigma|^2 < 2\ln(p_{non\text{-}edge}/p_{edge}) \qquad (13)$$

This is to say the signal noise ratio (SNR) μ^2/σ^2 and the estimated relative ratio $p_{edge}/p_{non\text{-}edge}$ shed light on the presence of bimodality. In other words, large SNR and small relative ratio indicates existence of two peaks in the distribution. Note the inequality (13) is deduced from gradient along x-axis. It is easy to see that the same condition is arrived by considering other directions.

Also, let $|\Delta I_o|$ denotes average gradient magnitude along the common edge and it can be shown that $|\Delta I_o| > t$.

Finally, the edge classifier is formulated as: let s_r denote "reliable" edges, s_u denote "unreliable" edges. A common edge is as follows,

$$label = \begin{cases} s_r, |\mu/\sigma|^2 \geq 2\ln(p_{non\text{-}edge}/p_{edge}) \ \&\&\ |\Delta I_o| > t \\ s_u, otherwise \end{cases} \qquad (14)$$

In addition, the width w of the support is varied from 2 to 16 pixels for each common edge. The width corresponding to the largest estimated SNR is set as the best width. The analysis above is done only for grayscale images. For color images, we simply combine the reliable edges detected in different channels.
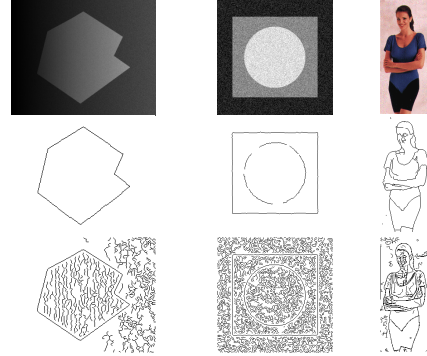

Figure 3: Comparison of reliable edge map and Canny edge map. First row, input images. Second row, reliable edge maps. Third row, Canny edge map using the default parameter setting of Matlab function edge('canny').

Our experiments show that this method is effective in finding reliable edges, even in noisy environment and non-uniform illumination cases. Some examples are shown in Figure 3. Canny method detects more false edges than our method (The Canny output can be improved by tuning the parameters. Ours is superior due to its adaptive nature.). However, our method makes mistakes when noise level is intolerantly large. Note the reliable edges of the second column have gaps along the circle and become discontinuous.

4.2. Goodness measure and final results

Though the adaptive classifier mentioned in 4.1 may make some errors, by combining the detected reliable edges together, we can obtain a powerful goodness measure of every joint boundary to judge whether two regions are distinct from each other. The goodness measure is defined in (15) as a reliable edge ratio,

$$g(\Gamma_i) = \sum_{l \in s} |l| \Big/ |\Gamma_i| \qquad (15)$$

where Γ_i is any considered boundary, S denote the set of common edges labeled as reliable on Γ_i.

The decision of whether R_i is distinct from R_j is made due to (16).

$$l(R_i, R_j) = \begin{cases} distinct, & g(\Gamma_{ij}) \geq T_0 \\ indistinct, & g(\Gamma_{ij}) < T_0 \end{cases} \qquad (16)$$

The threshold T_0 is related to the success rate of reliable edge detection on real boundaries. The optimal value of T_0 can be found statistically. Experiments on our test images find that setting T_0 from 0.6 to 0.7 renders good results. And in fact, the obtained segmentation results are quite stable by setting T_0 to any values between 0.6 and 0.7.

The method checks every joint boundary between two adjacent regions during merging and a final result is found if all joint boundaries of the segmentation are labeled as distinct.

5. Experimental results and conclusion

In all our following experiments on different images, we set T_0=0.6. Please note that the different nature of the test images shown in the first flow of Figure 3. The first image is an example of non-uniform illumination, the second is affected by severe Gaussian noise and the third is a natural image of human.

Figure 4: some elements in a partition set of the 1st image of Figure 3, α=0, λ=1.

The parameter α stresses small error between output image and input image and λ emphasize gradient information. Figure 4 shows some elements in a partition set. A comparison between the proposed method and a graph cut method[6] is shown in Figure 5. Since the graph cut method needs to specify a region number, in order to do a comparison, we set it to be the same as the number found by our method. For the second row of images, the graph cut method fails to detect the object in the first image and fails to segment major parts of the woman image; while our method correctly segments the object in the first image and divides the third into sensible regions.

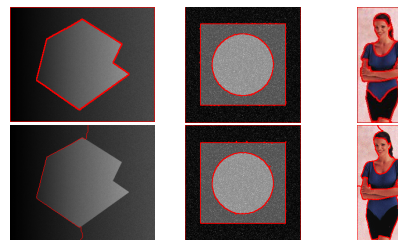More results are shown in Figure 6.

Figure 5: results comparison. 1st row, the proposed method; 2nd row, graph cut.

The experiments confirm that the proposed method is able to find a natural segmentation from the partition set. And the result is encouraging with every region clearly separated by a distinct region boundary.

Though our analysis is done in the intensity feature space, the idea can be extended to other features if a distance measure exists in other feature space. Now, we are extending our segmentation method to textures.
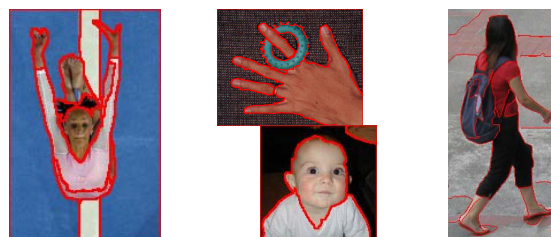

Figure 6: more results

References

[1]   R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed., 2003.

[2]   J. M. Beaulieu and M. Goldberg, "Hierarchy in Picture Segmentation - a Stepwise Optimization Approach," *IEEE Transactions on PAMI,* Feb 1989.

[3]   S. E. Hernandez and K. Barner, "Region merging using homogeneity and edge integrity for watershed-based image segmentation," *Optical Engineering,* 2005.

[4]   H. K, "Hybrid image segmentation using watersheds and fast region merging," *Image Processing, IEEE Transactions on,* vol. 7, 1998.

[5]   J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on PAMI,* 2000.

[6]   T. Cour, F. Benezit, and J. Shi, "Spectral segmentation with multiscale graph decomposition," in *CVPR*, 2005, pp. 1124-1131 vol. 2.

[7]   R. H. Y. Chung, N. H. C. Yung, and P. Y. S. Cheung, "An efficient parameterless quadrilateral-based image segmentation method," *IEEE Transactions on PAMI,* Sep 2005.

[8]   X. Ren and J. Malik, "Learning a classification model for segmentation," in *ICCV. ,* 2003, pp. 10-17 vol.1.