Toward Objective Segmentation Evaluation

Štěpán Šrubař
Department of Computer Science, FEI, VŠB - Technical University of Ostrava,
17. listopadu 15, 708 33, Ostrava, Czech Republic
stepan.srubar@vsb.cz

ABSTRACT
Two different segmentations of the same image can be evaluated in many ways. One resulting number hardly generalizes all differences in segmentations. Moreover, common methods can evaluate similar segmentations as quite different. Proposed evaluation divides dissimilarity into granularity and border difference. Granularity difference represents number of segments while border difference evaluates a rate of agreement in delineating of objects in the image. Such approach evaluates segmentations more precisely and keeps natural meaning of both resulting values.

Keywords
Segmentation evaluation, probabilistic Rand index.

1. INTRODUCTION
Images are segmented for detection of objects and their separation. There is no best segmentation for an image. Still, many people will segment the same image similarly to each other, thus there exist some common rules for segmentation. Evaluation of segmentation can be divided into two main classes. The first takes image and its corresponding segmentation, the second takes only two different segmentation of the same image. We will be interested here in the second category only.

Segmentation evaluation methods are often based on the number of pixels (or probability of pixels) that were incorrectly classified or differs in these two segmentations. Some methods computes distance of mis-classified pixels or border pixels to the nearest correct place. Number of segments for evaluation was also proposed. We can define evaluation using some feature of the image or segments, namely size of segments or its eccentricity. Mentioned methods were closely described and compared in [Fer06a, Jia05a, Zha96a].

The newest inventions were made in the first class of evaluation methods. It was shown recently in [Unn07a] that probabilistic Rand index (PRI) outperforms some other methods from the same class. That is the reason why I take PRI for comparison with proposed method.

Three images (see fig. 1) and their segmentations was used for testing and comparison. Figures 2 and 3

Figure 1: Images 100075, 100098 and 103041 from Berkeley segmentation database [Mar01a].

show segmentations of two different images. PRI of segmentations in figure 3 is 0.54 and PRI of the first segmentations in figure 2 and 3 is 0.607. The higher number, the higher correspondence. Evidently PRI is not suitable for segmentations where the same object is segmented into different number of segments. Many other methods suffers the same problem.

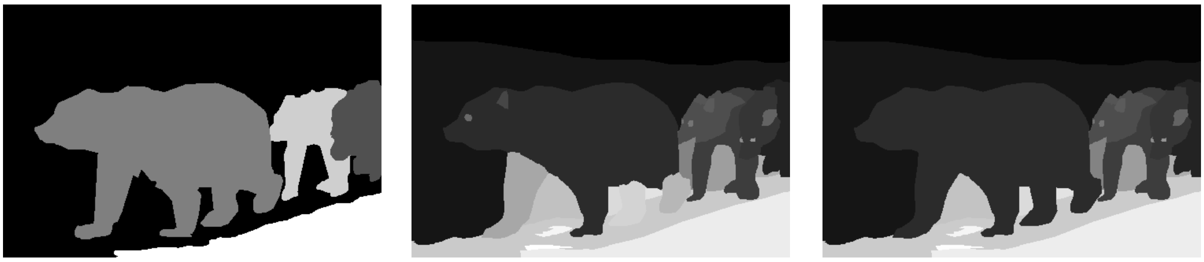

Figure 2: Two segmentations of the image 100075. Third segmentation is partial result of processing of second segmentation.

Figure 3: Two segmentations of the image 100098.

2. SEGMENTATION DIFFERENCE

We cannot say, objectively, what the correct number of segments in segmentation should be. On the other hand the borders of the segments are created according to some rules which correspond to the borders of objects. Therefore, segmentation evaluation should measure the precision of borders and suppress the number of segments. Still, the number of the segments should be expressed by another number.

For simplicity, first segmenter can detect only main objects while the second segmenter can separate these objects into smaller ones (see figures 2 and 3). Still, the borders of main objects should correspond in both segmentations. First task is to group segments from the second segmentation to create equivalent representation of objects from the first segmentation. Such grouping expresses granularity difference. Having coarse second segmentation with the same number of segments as in the first segmentation, difference of borders can be then evaluated. Segmentation difference is then expressed as two numbers. Border difference corresponds to inaccuracy of borders of objects, while granularity difference expresses difference in resolution of objects.

Granularity Difference

We can assume that bigger segments are better noticeable, thus they should have higher weight than smaller segments. Taking logarithm of normalized weighted sum of sizes of segments, we get following formula of granularity:

$$g(i) = -\log \frac{\sum_j |s_{ij}|^2}{\left(\sum_j |s_{ij}|\right)^2},$$

where $|s_{ij}|$ is the size of j-th segment of the image i. Logarithm converted unnatural geometric progression into more suitable arithmetic progression. This formula can be also used for arbitrary shaped part of an image, which will be used later.

Result of $g(i)$ is between zero and plus infinity. In numerator, the sum of sizes of segments is multiplied by their weights which are the sizes of the segments, thus the use of square. It is normalized by sum of weights (sizes of segments) and number of pixels of image. Both values are also identical which is represented by a square.

Say, we have one segment in the first segmentation representing some object. In the second segmentation, the same object is represented by more than one segment (see left bear in figure 2). We will call these segments in the second segmentation as joint segment. Such correspondence on an object in image will be called binding one to many segments or equivalently binding one segment to one joint segment. Another allowed bindings are one to one (trivial case of one to many), null to many and null to one (trivial of previous). For clearness see figure 4. Trivial cases will not be explicitly mentioned hereafter. Many to many binding is forbidden because it could lead to zero border difference for totaly different segmentations.

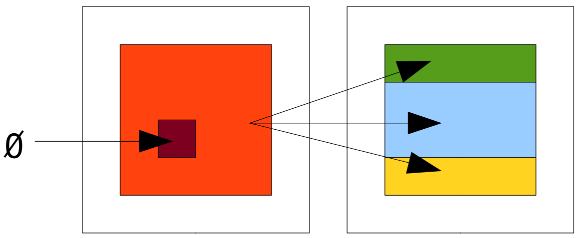

Figure 4: An example of binding in two images. Null to one is on the left, one to many binding is in the middle.

Pseudocode of searching of bindings for two images follows:

$I \leftarrow$ find all intersections of segments

$B \leftarrow$ null to many binding for each image

while I not empty

    $i \leftarrow$ remove the biggest intersection from I

    if both segments from i unprocessed then

        $B \leftarrow$ create new binding from i

    else if one segment from i unprocessed

        if putting unprocessed segment into binding of processed segment will not create many to many binding then

            put unprocessed segment into that binding

        else

            put unprocessed segment into corresponding null to many binding

    mark segments from i as processed

Now we have all segments in one to many or null to many bindings. We can compute granularity difference and border difference. First, we need intersection in bindings to be able of computing granularity:

$$b_k = s_{1k} \cap s_{2k},$$

where s_{1k} and s_{2k} are joint segments that are bound. For null to many binding we make intersection of segments with whole image.

Suppose joint segment as an irregular shaped segmented image. Granularity of binding with that joint segment is defined as granularity of such segmented image. Resulting granularity difference is weighted sum of granularity of intersections:

$$gd(s_1, s_2) = \frac{\sum_k |b_k| \cdot g(b_k)}{\sum_k |b_k|},$$

where $|b_k|$ is number of pixels of binding k and $g(b_k)$ is its granularity.

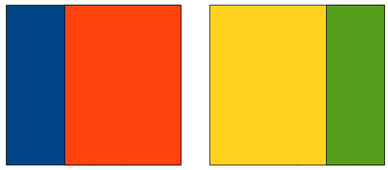

Figure 6: Segmentations for pseudonormalization of BD.

Time complexity, according to pseudocode and formulas, is $O(n)$ where n is number of pixels of the image.

Border Difference

Second value representing segmentation difference evaluates some type of precision of borders. For this purpose, we union segments in each joint segment to get one to one bindings only (see the union of a bear in the figure 2). Both null to many bindings are omitted in evaluation of border distance.

Each binding has now two corresponding segments. Border distance is based on sum of all pixels from one segment to the nearest pixels of the other segment. Proposed pseudonormalized border distance can be computed using following formula:

$$bd(s_1, s_2) = \frac{9 \cdot \sum_{b_i} [\sum_{x \in s_{1i}} d(x, s_{2i}) + \sum_{x \in s_{2i}} d(x, s_{1i})]}{w \cdot h \cdot max(w, h)},$$

where b_i is binding, s_{1i} is a segment from the segmentation s_1 and from the binding b_i, $d(x,s)$ is a distance of pixel x to the nearest point of a segment s. I propose euclid distances due to radial symmetry. w and h represents width and height of the image respectively.

The outer sum is pseudonormalized. Special case, that was chosen for pseudonormalization, consists of two segments in both segmentations as seen in figure 6. One segment takes left third while the second segment takes the rest. The second segmentation is horizontally flipped case of the first segmentation. Border difference of such segmentation pair is $\frac{1}{9}\, w \cdot h^2$ (see the number 9 in previous fomula). Function max is used because similar case is created by rotation by 90° and we take the worse of these two cases. Such segmentation pair is not the worst case, thus we call it

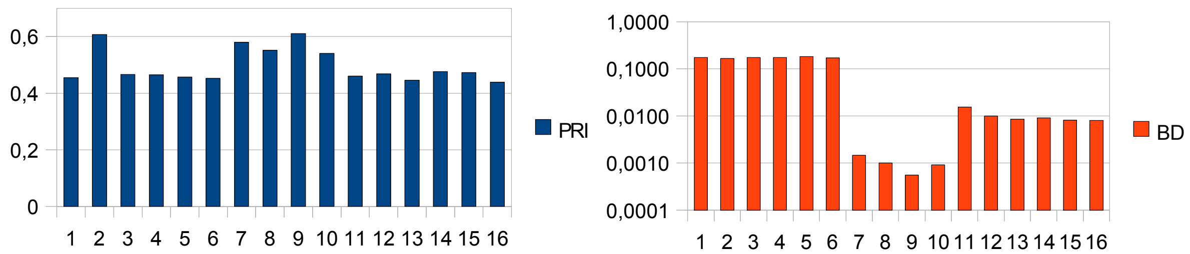

Figure 5: PRI and Border difference between 100098-1116 segmentation and 16 others. 1-6 are segmentations of image 100075, 7-10 are segmentations of image 100098 and 11-16 are segmentations of image 103041. PRI represents rate of correspondence, BD the rate of difference.

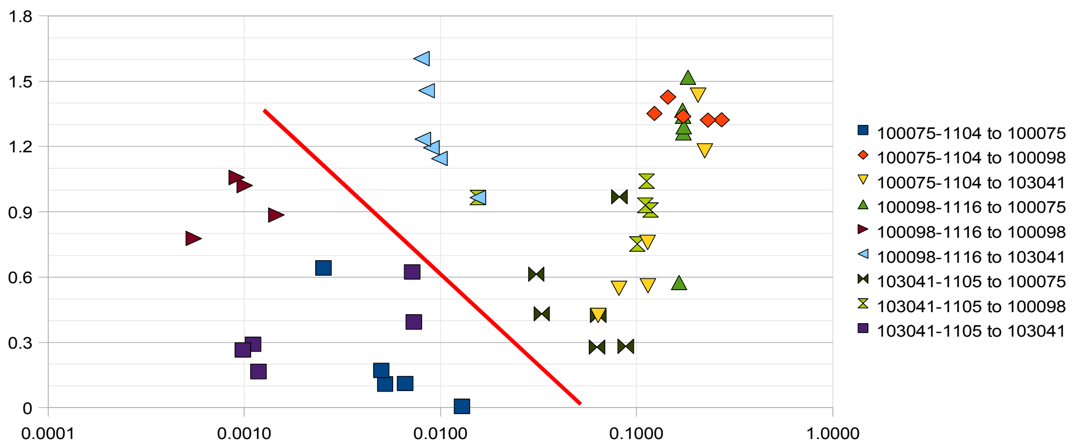

Figure 7: Similarity difference between three selected reference segmentations and segmentations of images 100075, 100098 and 103041. Horizontal axis represents border difference (in logarithmic scale) and vertical axis represents granularity difference. Red line separates points representing pairs of segmentations that belong to the same image. Points representing pairs from different images are on the right side of the red line.

pseudonormalization. On the other hand, typical values are not higher than pseudonormalization value.

Time complexity of border difference is *O(mn)*, where *n* is number of pixels and *m* is number of segments from both segmentations. Typically, the number of segments is much smaller than the number of pixels, thus the expected time complexity is *O(n)*. Running time of evaluation of granularity and border differences are in milliseconds on segmentations like in figures 2 and 3. Reference PRI method has time complexity $O(n^2)$ and runs tens of seconds on these segmentations to be computed precisely. For shorter computation time, PRI must use randomization algorithm Monte Carlo to estimate the result.

3. COMPARISON

I chose manually segmented images (see fig. 1) and all its segmentations. Proposed border difference (BD) is compared to probabilistic rand index (PRI) [Unn07a] in the figure 5. Images indexed as 7-10 should maximaly differ from others. Differences by BD are much greater than PRI and without any error (see index 2). Moreover, BD has to be represented in logarithmic scale.

Figure 7 shows robustness of proposed method. Similarity difference is measured for three chosen reference segmentations and the rest of the test set. All segmentations corresponding to their references are split from the other results by red line. The red line is diagonal, thus we need both granularity and border distance to make correct separation. Single value is evidently insufficient.

4. CONCLUSION

Two segmentations cannot be easily compared using a single number. Different number of segments does not necessarily mean that segmentations are from different images as the PRI could present. In fact, this can be caused by different granularity only. Thus, as was shown, granularity should be evaluated separately from precision of borders of segments. Such separation of properties in evaluation of segmentations leads to more robust results.

5. REFERENCES

[Fer06a] Fernando C. Monteiro and Aurlio C. Campilho. Performance Evaluation of Image Segmentation. ICIAR 2006, LNCS 4141, p. 248–259, 2006

[Jia05a] Xiaoyi Jiang, CyrilMarti, Christophe Irniger, and Horst Bunke. Distance Measures for Image Segmentation Evaluation. EURASIP Journal on Applied Signal Processing, vol. 2006, pp. 1–10, July 2005.

[Mar01a] D. 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. Proceedings of 8th IEEE International Conference on Computer Vision (ICCV ’01), vol. 2, pp. 416–423, Vancouver, BC, Canada, July 2001.

[Unn07a] Ranjith Unnikrishnan, Caroline Pantofaru, and Martial Hebert. Toward Objective Evaluation of Image Segmentation Algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 6, June 2007.

[Zha96a] Y. J. Zhang. A survey on evaluation methods for image segmentation. Pattern Recognition, vol. 29, pp. 1335-1346, 1996.