MULTI-STAGE REGION MERGING FOR IMAGE SEGMENTATION

Thomas Brox[1], Dirk Farin[1] and Peter H.N. de With[2]
farin@ti.uni-mannheim.de / p.h.n.de.with@tue.nl

[1]Dept. Circuitry and Simulation, University Mannheim, Germany
[2]CMG Eindhoven / University of Technology Eindhoven, Netherlands

The region-merging algorithm is a widely used segmentation technique for still-image segmentation. This paper evaluates the properties of several merging criteria when applied to real-world images. These criteria properties have been exploited to develop a novel algorithm, which is a multi-stage generalization of conventional region merging. The new algorithm uses a sequence of different criteria to achieve a semantically and subjectively superior segmentation result.

1 INTRODUCTION

The partitioning of natural images into a set of semantically meaningful objects (*segmentation*) is an important issue in image understanding and object-oriented video coding, such as in MPEG-4. A semantically correct segmentation is difficult to achieve and requires high-level *a-priori* knowledge of the image. Conventionally, this knowledge is integrated into the analysis steps subsequently following the segmentation. However, if knowledge about the input image contents is added into early low-level segmentation steps too, oversegmentation and incorrectly merged regions are decreased. By dividing the segmentation process into independent stages, which are each optimized for a specific image property, the selection of optimal segmentation parameters is alleviated and more robust to support a wide range of image types.

We will first present the principles of region merging. The remainder of this paper concentrates on several merging criteria and their performance, in order to come to a multi-stage algorithm.

2 REGION-MERGING PRINCIPLES

Let $P = \{p_i\}$ be the set of pixels in the input image with corresponding luminance $f(p_i)$ and let $R = \{r_i\}$ be a partition of the image pixels into disjoint regions $r_i \subseteq P$. From this partition, a region adjacency graph $G = (R, E)$ can be derived with $E = \{(r_i, r_j) | \exists p_k \in r_i, p_l \in r_j \text{ with } p_k \text{ being adjacent to } p_l\}$. Each edge can be attributed with a weight, providing a measure of dissimilarity between the two regions. We will refer to this measure as the *merging criterion*.

The region-merging algorithm starts with an initial segmentation. This can be a trivial segmentation with each image pixel forming a region of its own, or it can be the output of a preceding segmentation step. Then the algorithm proceeds by continuously searching for the edge with the lowest dissimilarity value and merging the two regions until a stopping criterion is satisfied (usually a threshold is defined and the merging process is terminated as soon as the minimum edge-weight exceeds this threshold). Note that after each step, the edge-weights of all edges outgoing from the merged region have to be recalculated according to the selected criterion (see [2] for an overview of terms and [1, 3] for efficient data structures and algorithms).

3 MERGING CRITERIA

A merging criterion consists of two parts: a *region model*, describing each image region with a set of features, and a *dissimilarity measure*, defining a metric on the features of the region model. The range of possible region models reaches from simple models like uniform luminance up to texture, shape or motion parameters. In the following, we will concentrate on low-level features which are applied at early stages of the algorithm. Furthermore, we only consider greyscale images. However, all presented critera can be readily generalized to work on color images.

The better a region model matches the real image-data, the longer the minimum edge-weights remain small and the steeper is the relative increase as soon as the segmentation has reached its final state. This makes the segmentation process more robust to the selection of the fixed threshold for the stopping condition.

3.1 MEAN LUMINANCE DIFFERENCE

The simplest region model is to describe each region r_i by its mean luminance μ_i. A straightforward possibility to define a dissimilarity measure on this model is to use the squared difference, from now on referred to as the *Mean*-criterion

$$w_{ij}^M = \left(\mu_i - \mu_j\right)^2.$$

3.2 WARD'S CRITERION

Another measure which operates on the mean-luminance model is the *Ward*-criterion [4]. The idea is to consider the model error for a region r_i, defined as $\mathcal{E}_i = \sum_{p \in r_i} \left(f(p) - \mu_i\right)^2$. The dissimilarity associated with a pair of regions is defined as the additional total error that is introduced by merging the two regions: $w_{ij}^W = \mathcal{E}_{ij} - \mathcal{E}_i - \mathcal{E}_j$ (with $\mathcal{E}_{ij}$ being the error after a hypothetical merge of r_i and r_j). After elementary simplifications, this can be expressed as:

$$w_{ij}^W = \frac{|r_i| \cdot |r_j|}{|r_i| + |r_j|} \left(\mu_i - \mu_j\right)^2.$$

3.3 MEAN/WARD MIXTURE

As will be discussed in the following section, neither the Mean-criterion nor the Ward-criterion produce a subjectively approporiate segmentation. A better criterion may be a compromise between the characteristics of Mean and Ward. For this reason, we introduce the geometrical mean of both criteria ($w_{ij}^G = \sqrt{w_{ij}^M \cdot w_{ij}^W}$) as a new *Mean-Ward* criterion. Since the absolute value of the criterion is not important, the square-root can be ignored, resulting in

$$w_{ij}^G = \frac{|r_i| \cdot |r_j|}{|r_i| + |r_j|} \left(\mu_i - \mu_j\right)^4.$$

3.4 LINEAR-LUMINANCE MODEL

Because of illumination effects, natural images seldomly consist of completely homogeneous regions. Almost all regions that we perceive as homogeneous, contain a small luminance gradient. Therefore, it is sensible to use a region model that is capable of describing slowly varying luminance gradients. A possible region model defines the luminance distribution as $f'(x,y) = \alpha + \beta x + \gamma y$ with the three parameters α, β, γ. For each individual region, these parameters are estimated from the image data, using a least-squares approach. Comparable to the Ward-criterion, we define the model error as $\mathcal{E}_i^L = \sum_{p \in r_i} \left(f(x,y) - f'(x,y)\right)^2$ and the region dissimilarity as

$$w_{ij}^L = \mathcal{E}_{ij}^L - \mathcal{E}_i^L - \mathcal{E}_j^L.$$

3.5 BORDER CRITERION

Although the linear-luminance model handles well most regions occurring in natural images, the model has two main drawbacks: it is rather computationally intensive, and it still cannot handle all cases of small luminance variations. Especially curved surfaces have complicated luminance distributions. Both problems can be circumvented by using the following *Border* criterion.

Let $B_{ij} = \{(p_k, p_l)\}$ be the set of pairs of pixels along the common boundary between region r_i and r_j (with $p_k \in r_i$ and $p_l \in r_j$). We define the *Border*-criterion as the sum of squared differences along the boundary

$$w_{ij}^B = \frac{1}{|B_{ij}|} \sum_{(p_k, p_l) \in B_{ij}} \left(f(p_k) - f(p_l)\right)^2.$$

4 CRITERIA PROPERTIES

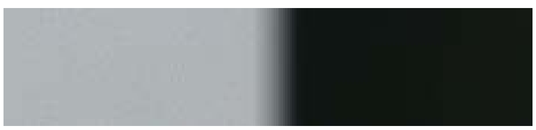
4.1 GENERAL BEHAVIOUR

Figure 1 depicts a detail view of an image containing an unsharp edge. As the image is part of a real-world image, it contains a hardly visible luminance gradient in the "flat" image regions and some camera noise. The image has been

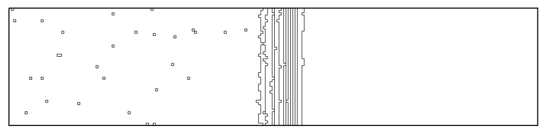
criterion / problem class	Mean	Ward	Mean-Ward	Linear-Lum.	Border	Watershed presegm.
noise	−	++	+	−	−	−
blurred edges	−	+	+	+	−−	+
double edges	−	+	+	−	−	+
illumination effects	−	−−	+	++	++	+
subjective evaluation	+	−−	++	+	+	−
stopping criterion	−−	+	++	+	+	N/A
comp. complexity	+	+	+	−−	+	++

Table 1: Comparison of the performance of different criteria on a number of typical problem classes.

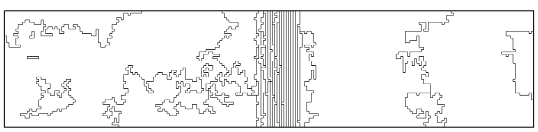
segmented independently with the Mean, Ward, and Linear-Luminance criterion until only 50 regions were remaining.



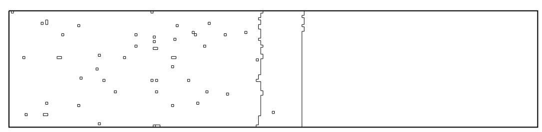
(a) input image



(b) Mean criterion



(c) Ward criterion



(d) Linear-Luminance criterion

Figure 1: Unsharp edge, 50 regions remaining

It is easily visible that the Ward criterion favours the removal of small noisy areas instead of combining large, but only slightly different regions. This is the case because the Ward criterion considers the total error and small differences in very large regions outweigh larger differences in very small regions. The Mean criterion does not show this effect, because it does not take region size into account. Similarly, the Linear-Luminance criterion can adapt its model to approximate the gradient with sufficient accuracy. Furthermore, it is also capable to model the unsharp edge itself and does not lead to the oversegmentation with many narrow regions, as it is the case with the other two criteria.

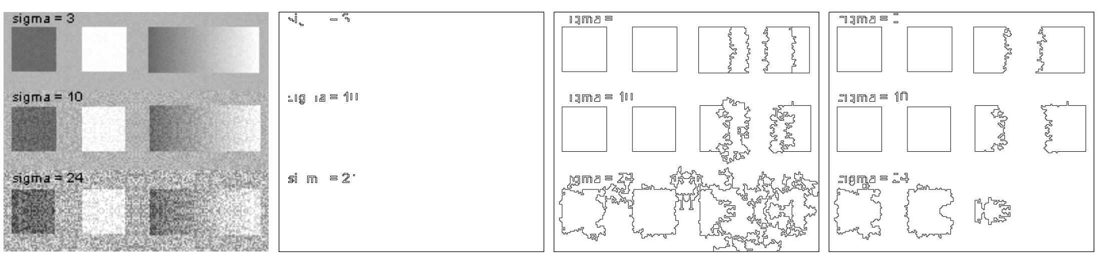
4.2 COMPARISON

In natural image segmentation, several classes of commonly occuring difficulties can be identified. The robustness of each criterion on the problem classes was evaluated and is depicted in Table 1. In the following, some problem classes are described in more detail[1].

- **Noise.** Camera noise has a well visible effect at the beginning of the segmentation process. Dissimilarity measures which are normalized to their region sizes, like Ward's criterion, give superior results, because single noisy pixels introduce no large overall error.
- **Blurred edges and double edges.** Objects which are out of camera focus appear with blurred edges in the image. This can lead to an oversegmentation into many thin rings around the object boundary. The Linear-Luminance criterion can approximate the blurred edge with a single region if the object boundaries are straight lines. Curved boundaries can be handled by the border criterion.

  However, the more general model of the Border criterion has the disadvantage to ignore the pixels inside a region. Thus, it is possible for an object to grow along its unsharp border until it is completely merged with the background (see Figure 4c).

  Almost all sharp edges in real-world images consist of pixels midway between the colors of the two regions. After a segmentation with a too low threshold, objects seem to have double edges. Because of its tendency to merge small regions, the Ward-criterion can remove these double edges well.
- **Subjective evaluation.** As can be seen in Figure 2, the Ward criterion has a tendency to split large regions into several segments, while the Mean criterion removes large regions equally likely as small regions. Figure 2d shows the segmentation using the Mean-Ward criterion. The result is much more subjectively pleasing as the large regions are preserved, and much of the text is kept.



(a) input image (b) Mean (c) Ward (d) Mean-Ward

Figure 2: Segmentation results, 50 regions remaining.

- **Stopping criterion.** In Figure 3, the minimum edge-weights prior to the last 10 merging steps of Figure 1a are plotted for each criterion. To ease the selection of robust stopping thresholds, it is desirable that the weights

[1] See section 5.1 for more information on the watershed-presegmentation column.

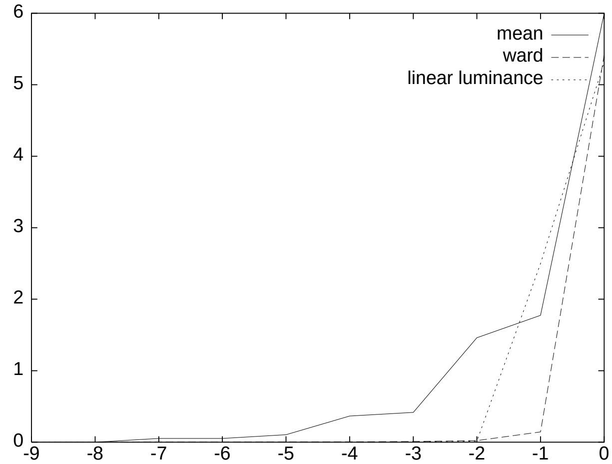


Figure 3: Minimum edge weights at last 10 merging steps. Weights have been scaled to fit into a single plot.

remain small for as long as possible and show large growth as soon as the model cannot adapt to the enlarged regions anymore.

The Mean criterion increases before a subjectively recognizable reason is visible. On the other hand, both Ward and Linear-Luminance show a very steep increase in model error.

5 MULTI-STAGE MERGING

As discussed in the last sections, each criterion shows both advantages and disadvantages. Choosing a single criterion for the complete segmentation process results in a dissatisfactory segmentation. This motivates a multi-stage approach. A criterion is used as long as it can well handle the current configuration. Then, the criterion is exchanged by another one. By using several stages, the selection of an appropriate threshold in the stopping criterion is not critical. It should be chosen sufficiently low to ensure that control is passed to the next criterion, before the situation exceeds the capabilities of the criterion's region model. A sequence of criteria which produced good results was:

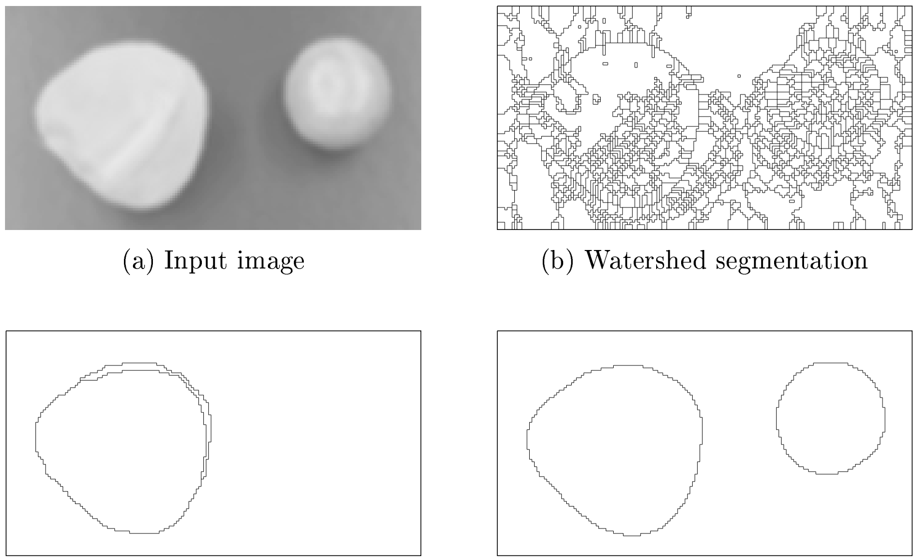
1. **Ward**, removing much of the image noise and eliminating double edges,
2. **Mean-Ward**, which does the main work, before finally
3. **Border** merges regions in which illumination effects play a central role.

5.1 APPLYING A WATERSHED PRESEGMENTATION

Instead of starting the algorithm with single pixel regions, it is possible to perform a presegmentation with the watershed algorithm on a gradient map of the input image.

This presegmentation has the advantage to considerably reduce the computational complexity as the watershed transform is a fast algorithm and reduces the amount of input regions. Furthermore, it alleviates the problem of the Border criterion to destroy complete objects with unsharp boundaries (see Figure 4). The watershed transform parts these blurred areas at the object boundaries at the position of maximum gradient into only two regions. For this reason, the threshold in the stopping condition for the Border criterion can be set lower.

The disadvantage of applying this presegmentation is that small image structures can get destroyed. Additionally, in the presence of camera noise, smooth edges in the image can get "fuzzy" in the segmentation.



(a) Input image

(b) Watershed segmentation

(c) Without watershed. The right object is removed by the Border criterion.

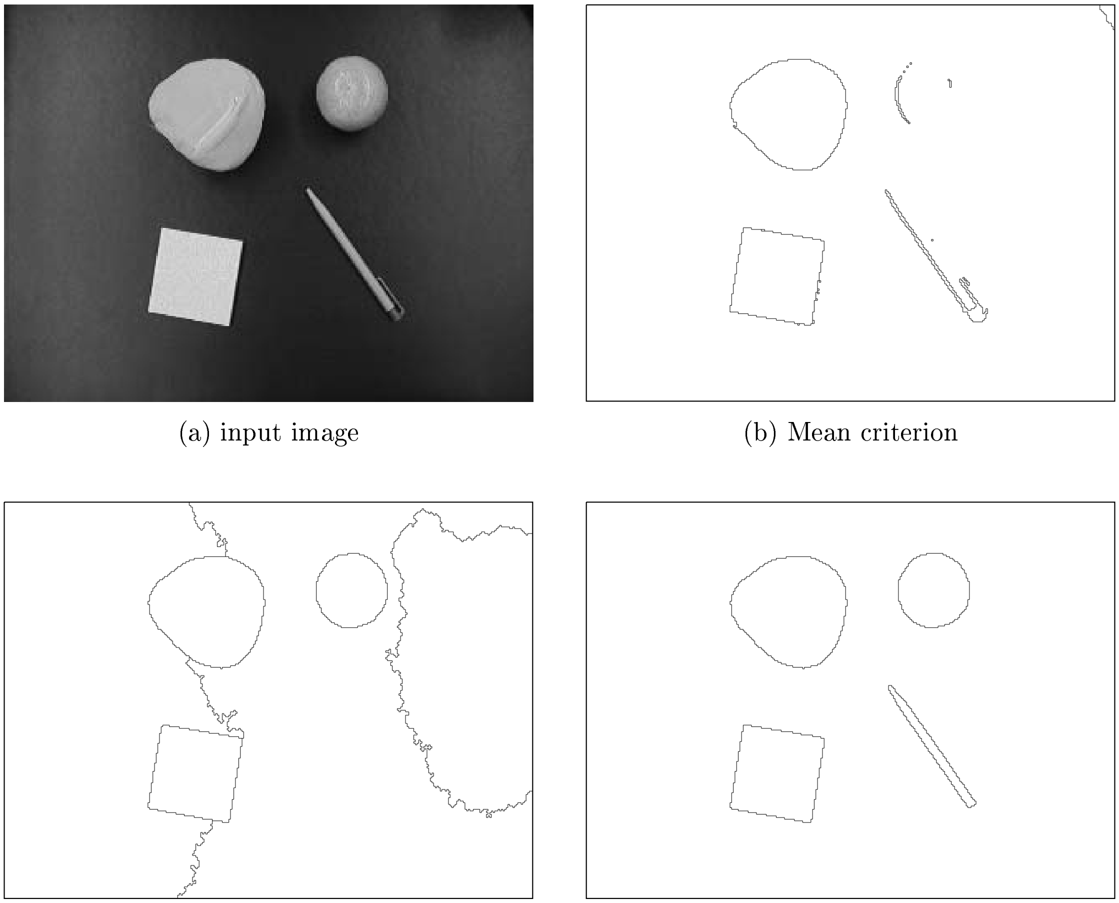
(d) With watershed. The right object is correctly segmented due to a lower threshold.

Figure 4: Effect of applying a watershed presegmentation.

6 RESULTS AND CONCLUSIONS

We have described region-merging as an image segmention algorithm where merging criteria play a key role for improving the segmentation result. Several low-level merging criteria have been evaluated for application in natural image segmentation. Based on the properties of the criteria, a multi-stage approach has been presented. The Ward criterion is used in the first stage to reduce the influence of image noise. The subsequent Mean-Ward stage performs the actual color segmentation, and finally, the Border criterion reduces oversegmentation due to illumination effects.

Figure 5 shows a sample image with the results of the multi-stage segmentation algorithm. Neither the Mean criterion, nor the Ward criterion alone achieves acceptable segmentation results. Only by using a multi-stage approach of Ward,



(a) input image (b) Mean criterion

(c) Ward criterion (d) Multi-stage

Figure 5: Several objects on a table.

Mean-Ward, and Border (Figure 5d), a subjectively superior object separation is obtained.

REFERENCES

[1] Dirk Farin and Peter H. N. de With. Towards real-time MPEG-4 segmentation: a fast implementation of region merging. In 21^{st} *Symposium on Information Theory in the Benelux*, pages 173–180, 2000.

[2] Kostas Haris, S. N. Efstratiadis, and N. Maglaveras. Watershed-based image segmentation with fast region merging. In *IEEE International Conference on Image Processing*, volume 3, pages 338–342, 1998.

[3] T. Kurita. An efficient clustering algorithm for region merging. In *IEICE Trans. of Information and Systems*, volume E78-D, No. 12, 1995.

[4] J. H. Ward. Hierarchical grouping to optimize an objective function. *J. American Stat. Assoc.*, 58:236–245, 1963.