How Do Superpixels Affect Image Segmentation?

Allan Hanbury

Pattern Recognition and Image Processing Group,
Institute of Computer-Aided Automation, Vienna University of Technology,
Favoritenstraße 9/1832, A-1040 Vienna, Austria

Abstract. Computationally intensive segmentation algorithms often operate on an image pre-segmented into small regions referred to as "superpixels". We investigate the effect of the choice of the pre-segmentation algorithm and its parameters on the outcome of the final segmentation. Three pre-segmentation algorithms are compared. To avoid the particularities of sophisticated segmentation algorithms, the final segmentations are built using agglomerative hierarchical clustering. These segmentations are evaluated using 300 images from the Berkeley Segmentation Dataset. This leads to useful insights about the variations in the final segmentation caused by the choice of the pre-segmentation algorithm.

Keywords: image segmentation, clustering, segmentation evaluation.

1 Introduction

Many recent segmentation algorithms are formulated as optimisation problems and solved by relaxation or energy minimisation techniques [1,2]. As these techniques are computationally intensive, applying them directly to the pixels in an image usually leads to long computation times. One possibility to reduce the computation time is to scale the image so as to reduce the total number of pixels, but a better alternative, which should lead to less information loss, is to use an initial over-segmentation to produce a set of small regions which we refer to as "superpixels" (as done in [3]). Superpixels should be local, coherent and preserve most of the information necessary for segmentation at the scale of interest [3]. They have the advantage that they should adapt themselves to the image structure, being larger where the colour remains similar over a large area. Various techniques have been adopted to produce this over-segmentation: the mean shift algorithm [4] has been used in [1], the normalised cuts [5] in [3], the graph-based segmentation technique of [6] in [2], and the watershed segmentation [7] in [8].

While the pre-segmentation into superpixels reduces the computational burden, an aspect that has not been considered is the effect of the choice of pre-segmentation algorithm on the final segmentation. We concentrate on this aspect in this paper, in which three algorithms for generating superpixels are compared: the watershed using volume extinction values [7], the mean shift algorithm [4] and the efficient graph-based segmentation technique [6]. As we do not wish

J. Ruiz-Shulcloper and W.G. Kropatsch (Eds.): CIARP 2008, LNCS 5197, pp. 178–186, 2008.

the results to be affected by the particularities of sophisticated optimisation algorithms, we produce the final segmentations using agglomerative hierarchical clustering [9]. This algorithm has the advantage that there is only one parameter to set to produce different segmentations. We evaluate the segmentation results on the 300 images from the Berkeley segmentation dataset [10].

The structure of the paper is as follows. Section 2 describes the algorithms for pre-segmentation into superpixels, while Section 3 describes the hierarchical clustering algorithm. The evaluation of image segmentations is discussed in Section 4 and applied to the segmentations in Section 5. Section 6 concludes.

2 Pre-segmentation into Superpixels

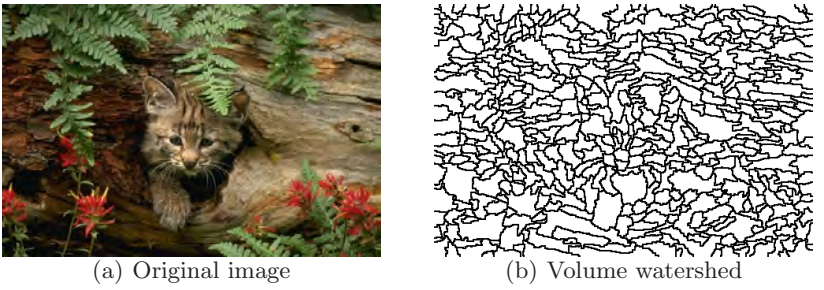
We examine three algorithms for doing the initial segmentation into superpixels.

The *watershed using volume extinction values*, abbreviated as *volume watershed*, is closely related to the standard morphological watershed algorithm [11]. However, in this version of the watershed, the lakes merge when they meet. A record is kept of the merging in the form of a graph [12]. It has been found that flooding so that each lake has the same volume provides the most useful segmentation [12]. Based on the graph, one can obtain a segmentation into a specified number of regions. The only parameter of this segmentation algorithm is the number of regions required. The flooding is done on the gradient of the colour image. We use the *saturation weighing-based colour gradient* that was found to give the best results in a morphological waterfall segmentation [13]. To simplify the image before segmenting it, we make use of the morphological *leveling* [14]. The filter used to produce the marker for the leveling operator is the morphological alternating sequential filter [11], where the size of the filter refers to the number of subsequent opening and closing operations. An example of a volume watershed with 500 regions and a pre-filtering of size 3 is shown in Figure 1b.

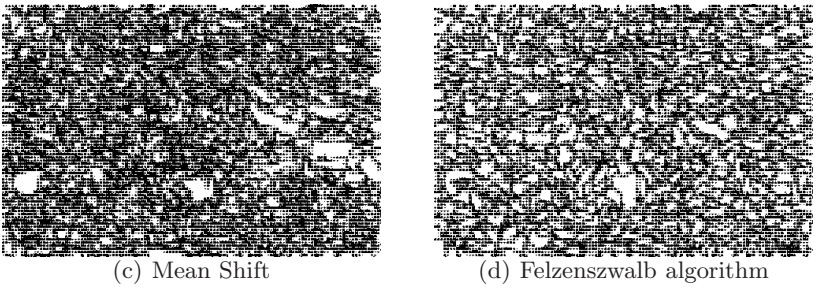
The *mean shift* algorithm is an iterative statistical approach to mode detection and clustering based on gradient estimation [4]. It has the advantage of being able to find non-spherical clusters. The mean shift procedure used[1] segments an image by clustering in a five-dimensional space, where each vector consists of the colour coordinates and the spatial coordinates of each pixel. The number and size of the resulting regions is controlled by two bandwidth parameters: the spatial bandwidth h_s related to the two spatial features and the range bandwidth h_r related to the colour coordinate part of the feature vector. The implementation also includes a parameter M, the minimum size of a region in pixels. An example of a mean shift segmentation is shown in Figure 1c.

This *efficient graph-based segmentation algorithm*, introduced by Felzenszwalb and Huttenlocher [6], is abbreviated here as the *Felzenszwalb algorithm*. The implementation used[2] starts from a 4-connected pixel neighbourhood graph. Neighbouring regions are fused based on a predicate which compares inter-component differences to within component differences [6]. Each colour channel is processed

[1] EDISON software: `http://www.caip.rutgers.edu/riul/research/code/EDISON/`
[2] `http://people.cs.uchicago.edu/~pff/segment/`



(a) Original image (b) Volume watershed



(c) Mean Shift (d) Felzenszwalb algorithm

Fig. 1. Pre-segmentations (superpixels) of the original image (a) using (b) the volume watershed with 500 regions and a pre-filtering of size 3. (c) the mean shift with $h_s = 5$, $h_r = 4$, $M = 50$. The number of regions is 1187. (d) The Felzenszwalb algorithm with $\sigma = 0.01$, $k = 8$, $M = 50$. The number of regions is 793.

separately and the resulting segmentations are fused. The number of regions is controlled by two parameters: σ is the size of a Gaussian filter used to smooth the image before segmentation, and k is the decision threshold on the value of the predicate. As for the mean shift algorithm, M is the minimum size of a region. An example of a segmentation is shown in Figure 1d.

3 Agglomerative Hierarchical Clustering

A hierarchical clustering method is a procedure for transforming a proximity matrix into a sequence of nested partitions [9]. We use an agglomerative clustering method on the Euclidean distances between the coordinates of the colours in the CIELAB space. As the transformation to CIELAB coordinates via CIEXYZ coordinates requires the original RGB coordinates to not be gamma-corrected (linear light coordinates) [15], we first apply an inverse gamma-correction to each of the RGB channels. We assume the commonly used value of $\gamma = 2.2$.

Once the tree has been built by the clustering algorithm, a final set of clusters (partition) is obtained by specifying a cutoff distance d — all cluster fusions corresponding to a distance larger than the given cutoff distance are removed. After this step, each superpixel in the image is assigned an index value indicating the colour cluster to which it belongs. This results in an image labelled by colour cluster membership. Figure 2 shows the segmentations resulting from clustering

(a) vol $d : 18.4$, $c : 8$, $s : 130$ (b) ms $d : 18.4$, $c : 10$, $s : 316$

(c) vol $d : 27.6$, $c : 3$, $s : 34$ (d) ms $d : 27.6$, $c : 4$, $s : 85$

(e) vol $d : 36.8$, $c : 2$, $s : 7$ (f) ms $d : 36.8$, $c : 2$, $s : 15$

Fig. 2. Results of clustering the superpixels generated by the 1000 region volume watershed (vol, left column) and mean shift (ms, right column). Below each image is the cutoff distance for the hierarchical clustering d, the number of resulting colour clusters c and the number of regions in the segmentation s.

the volume watershed and mean shift superpixels obtained from the cat image shown in Figure 1a for different values of the cutoff distance.

4 Segmentation Evaluation

As segmentation ground truth, we use 300 colour images from the Berkeley segmentation dataset[3] [10], where each image has been manually segmented by at least five human test subjects. For comparing an automatic segmentation to the manual segmentations, we use the boundary based comparison methodology outlined in [16]. This produces a precision-recall curve comparing the boundaries of the regions of each segmentation with the boundaries created by the human

[3] `http://www.cs.berkeley.edu/projects/vision/grouping/segbench/`

test subjects. In the original paper [16], an automatically determined boundary image, in which each pixel encodes the probability of it being a boundary pixel is evaluated. A series of n thresholds is applied to the boundary image, producing a set of binary images containing all pixels which have probabilities above the corresponding threshold. Exact correspondences between the pixels in each of these binary images and the ground truth are found by solving a minimum cost bipartite assignment problem. A pixel in the automatically determined boundary image is considered to be matched to a pixel in the ground truth image if the distance between them is less than 0.0075 times the length of the image diagonal. As the segmentation dataset images have a size of 321×481 pixels, this is a distance of 4.3 pixels. The *precision* is then the fraction of detections that are true positives (i.e. correspond to points in the ground truth) rather than false positives, while *recall* is the fraction of true positives that are detected rather than missed. A plot of the precision and recall values at each threshold is finally produced. As in [16], we use the maximum F-measure along a precision-recall curve to characterise the curve by a single value. For a precision-recall pair (P, R), the F-measure is defined as $F = PR/[\alpha R + (1 - \alpha) P]$ with $\alpha = 0.5$.

We use the same approach to evaluate our segmentations, except that the precision and recall are calculated for the region boundaries obtained for each value of the cutoff distance d. At each cutoff distance, the F-measure is also calculated. The best F-measure and the cutoff distance producing it are found by interpolating between the calculated F-measures.

5 Experiments and Results

We tested hierarchical clustering on superpixels produced by the volume watershed, mean shift and Felzenszwalb algorithm using the following parameters:

1. Volume watershed with 250 regions, pre-filtering of size 3 (v250 f3).
2. Volume watershed with 500 regions, pre-filtering of size 3 (v500 f3).
3. Volume watershed with 1000 regions, pre-filtering of size 3 (v1000 f3).
4. Volume watershed with 500 regions, no pre-filtering (v500 f0).
5. Mean shift with $h_s = 5$, $h_r = 4$, $M = 50$ (ms s5 r4 m50).
6. Felzenszwalb algorithm with $\sigma = 0.01$, $k = 8$ and $M = 50$ (fz s0.01 k8 m50)

The parameters for the last two algorithms were chosen to give a number of superpixels in the same range as those tested for the volume watershed. The mean shift parameters produce pre-segmentations with between 21 and 1296 regions, with a mean of 633 regions, while the Felzenszwalb algorithm parameters produce pre-segmentations with between 501 and 943 regions, with a mean of 750 regions. On a Pentium 4 computer, the fast version of the mean shift algorithm took an average of 4.3 seconds per image, the Felzenszwalb algorithm took an average of 1.7 seconds, while the volume watershed with pre-filtering required an average of 2.2 seconds irrespective of the number of regions specified.

For each image, we calculate segmentations corresponding to nine values of the cutoff distances d. We took the cutoff distances to be multiples of twice the

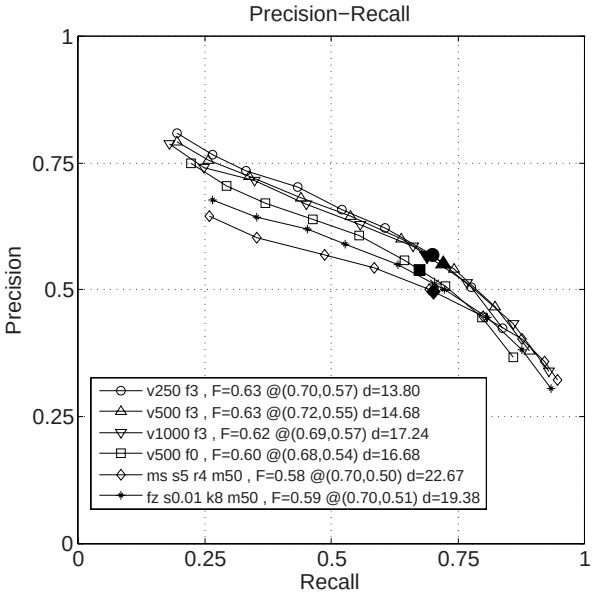


Fig. 3. Mean precision-recall curves for the segmentation results obtained by the hierarchical clustering algorithm applied to the superpixels obtained by the methods listed in the legend. The large points represent the positions of maximum F-measure, where these maximum F-measures, their coordinates and the estimated cutoff distance producing them are given in the legend. The rightmost point of each curve results from a cutoff distance of 4.6, and the leftmost point from a cutoff distance of 41.4.

just noticeable colour difference distance in the CIELAB space [17], or $d_i = 4.6i\,(i = 1, 2, \ldots, 9)$. The mean precision-recall curves over all 300 images for all of the tested pre-segmentation methods are shown in Figure 3. The three volume watershed pre-segmentations with pre-filtering produce precision-recall curves that lie above the other three curves over the whole domain except for recall above 0.8 and have the highest F-measures. It can also be seen that the number of superpixels chosen has little effect on the precision-recall curves. Removing the pre-filtering step results in a consistently lower precision-recall curve. The mean shift pre-segmentation method results in a precision-recall curve that remains below those of the other methods except at high recall values. It also has the lowest F-measure. The Felzenszwalb algorithm curve lies above the mean shift curve, except at high recall values.

Examining the estimated cutoff distances producing the highest F-measures given in Figure 3, one can see that these cutoff distances are larger for the pre-segmentation methods producing larger numbers of superpixels. The estimated optimal cutoff distance is largest for the pre-segmentation by the mean shift.

As it is instructive to look at the results on individual images, the results for all 300 images are available on the web[4]. For each image this includes the

[4] `http://muscle.prip.tuwien.ac.at/CIARP_segresult`

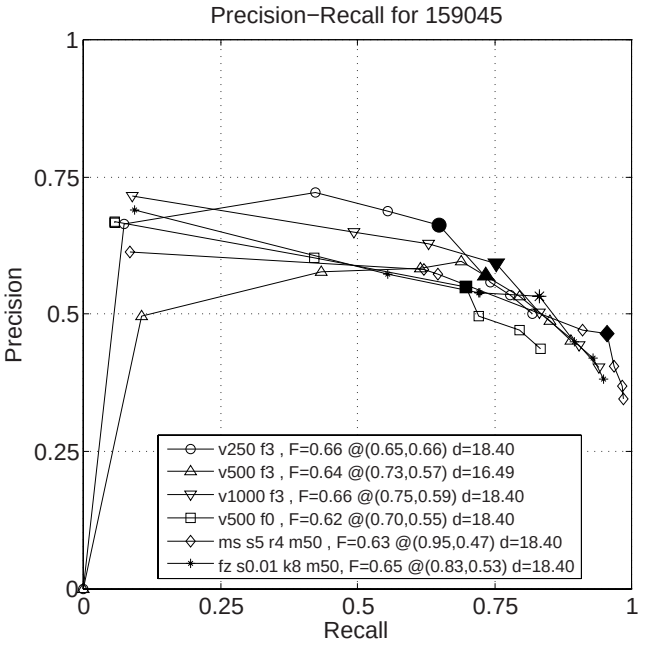


Fig. 4. Precision-recall curves for the segmentation results shown in Figure 2

precision-recall curve and segmented images. We examine one image in detail here. Consider the segmentations of the cat image shown in Figure 2, for which the precision-recall curves are shown in Figure 4. The highest F-measures of 0.66 are produced by the (v250 f3) and (v1000 f3) methods. These highest F-measures all occur at a cutoff distance of 18.4. From the images, it can be seen that the pre-segmentation by the mean shift tends to preserve much more detail than the volume watershed. The fact that the highest recall values for this image result from the mean shift superpixels also indicates this. When the cutoff distance takes on the high value of 36.8, then for both pre-segmentation methods shown, only the very contrasted red flowers are separated from the rest of the image. In the precision-recall graph, the curves diving to the origin indicate that all the superpixels have been merged into one region covering the whole image.

6 Conclusion

We investigated the effect of a pre-segmentation into superpixels on the segmentation obtained by clustering superpixels, and compared three algorithms for producing superpixels: the watershed using volume extinction values, the mean shift algorithm and the Felzenszwalb graph-based algorithm. An advantage of the watershed algorithm is that the only input parameter is the number of regions required, while for the latter two algorithms, the same parameters lead to a widely varying number of regions depending on the image.

The correlation between the number of superpixels and the cutoff distance producing the highest F-measure shows that the choice of a pre-segmentation algorithm and of its parameters should be taken into account in the design of algorithms which create segmentations based on superpixels. A rather surprising conclusion to be drawn from the results is that clustering on the superpixels of less complex shape generated by the volume watershed algorithm produces better segmentations (as measured by the F-measure). This could however be due to the widely varying number of regions produced by the same parameters for different images with the other two methods. This means that some images could already have salient contours removed by the pre-segmentation. The number of superpixels provided as a parameter to the volume watershed has less of an effect on the maximum F-measure. Further experiments will test if this result also holds for more complex superpixel grouping algorithms such as the minimum cut/maximum flow algorithm.

References

1. Keuchel, J., Heiler, M., Schnörr, C.: Hierarchical image segmentation based on semidefinite programming. In: Rasmussen, C.E., Bülthoff, H.H., Schölkopf, B., Giese, M.A. (eds.) DAGM 2004. LNCS, vol. 3175, pp. 120–128. Springer, Heidelberg (2004)
2. Mičušík, B., Pajdla, T.: Multi-label image segmentation via max-sum solver. In: Proc. of the Conf. on Computer Vision and Pattern Recognition (CVPR) (2007)
3. Ren, X., Malik, J.: Learning a classification model for segmentation. In: International Conference on Computer Vision, pp. 10–17 (2003)
4. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. IEEE Trans. on PAMI 24, 603–619 (2002)
5. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. on PAMI 22(8), 888–905 (2000)
6. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient graph-based image segmentation. International Journal of Computer Vision 59(2), 167–181 (2004)
7. Meyer, F.: An overview of morphological segmentation. International Journal of Pattern Recognition and Artificial Intelligence 15(7), 1089–1118 (2001)
8. Stawiaski, J., Decenciére, E.: Region merging via graph-cuts. Image Analysis and Stereology 27(1), 39–45 (2008)
9. Jain, A.K., Dubes, R.C.: Algorithms for Clustering Data. Prentice-Hall, Englewood Cliffs (1988)
10. 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. II, pp. 416–423 (2001)
11. Soille, P.: Morphological Image Analysis, 2nd edn. Springer, Heidelberg (2002)
12. Meyer, F.: Graph based morphological segmentation. In: Proceedings of the second IAPR-TC-15 Workshop on Graph-based Representations, pp. 51–60 (1999)
13. Angulo, J., Serra, J.: Color segmentation by ordered mergings. In: Proc. of the Int. Conf. on Image Processing, vol. II, pp. 125–128 (2003)

14. Meyer, F.: Levelings, image simplification filters for segmentation. Journal of Mathematical Imaging and Vision 20, 59–72 (2004)
15. Poynton, C.: A Technical Introduction to Digital Video. Wiley, New York (1996)
16. Martin, D., Fowlkes, C., Malik, J.: Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Transactions on Pattern Analysis and Machine Intelligence 26(5), 530–549 (2004)
17. Mahy, M., van Eyckden, L., Oosterlinck, A.: Evaluation of uniform color spaces developed after the adoption of CIELAB and CIELUV. Color Res. Appl. 19(2), 105–121 (1994)