A General Framework for Image Segmentation

Rui Luís V. P. M. Pires, Patrick De Smet and Wilfried Philips
e-mail: (rlpires,pds,philips)@telin.ugent.be
Department of Telecommunications and Information Processing
Ghent University
Sint-Pietersnieuwstraat 41, B-9000 Gent, BELGIUM

Abstract—**In this paper we describe a methodology to generate a partition of an image and how a hierarchical region merging scheme can be used to improve the quality of the segmentation. The segmentation method is based on the watershed transform. Prior to the actual segmentation, the image is smoothed to decrease the amount of detail detected by the watershed transform. To further improve the segmentation result, we use an iterative region merging process that sequentially merges the most similar pair of regions according to a pre-defined similarity metric. We propose the use of a combined similarity function that considers not only the intensity similarity of two neighboring regions but also the gradient magnitude at their dividing boundary. Results obtained illustrate the overall good performance of this segmentation methodology and the usefulness of the combined similarity function.**

Keywords— **Image segmentation, region merging, region adjacency graph.**

I. Introduction

Image segmentation remains a very important topic in the domains of image and video processing and analysis with numerous applications ranging from image coding to object recognition and tracking. In this paper we describe a method to generate an accurate partition of an image and a region merging scheme that is used to improve the final segmentation result.

The initial image partition is obtained using the watershed transform [1], [2], [3]. Well known for the good accuracy in delineating the object borders, the watershed transform also has the major disadvantage of yielding oversegmented results when used with real images, making inappropriate its use without some form of pre-processing of the images to segment.

A significant reduction of this inherent watershed oversegmentation can be achieved with the introduction of smoothing before the transform is actually applied. For this step we use nonlinear diffusion [4], [5], [6], which is able to largely preserve the edges present in the image while introducing the required smoothing.

The segmentation results obtained with the use of nonlinear diffusion as the smoothing filter are indeed better [6], as oversegmentation is to some extent reduced and still maintaining a very good accuracy in terms of edge location. However, images often remain excessively fragmented, especially in textured areas where the smoothing step may not be particularly efficient.

A post-processing step is thus required to further improve the segmentation results. In this paper we describe a hierarchical region merging method that makes use of a graph to represent the image partitions. In this graph, the image segments are viewed as the graph vertices and segment adjacency is represented by the graph edges connecting different vertices. A sequential merging of adjacent regions is then performed taking into consideration a pre-defined region similarity measure.

The paper is organized as follows: Section II gives a description of the edge-preserving smoothing step and the algorithm we use to implement the watershed transform. In Section III the hierarchical region merging method is explained and an analysis of the similarity measures used to guide the merging process is given. Experiments carried out and results obtained are described in Section IV. A summary and conclusions are presented in Section V.

II. Initial Image Partitioning

A. Image Smoothing

As already mentioned in Section I, prior to the generation of an image partition using the watershed algorithm, the image must be smoothed to avoid an otherwise severe oversegmentation. Traditionally, linear techniques such as Gaussian blurring were used to achieve these goals. However, since the filtering is place independent, they show the major disadvantage of blurring image edges, making the accurate determination of their location a far from trivial task.

This problem is circumvented with the use of nonlinear diffusion [4], [5] which is able to introduce the required smoothing while preserving most of the edges. In a continuous setting, the nonlinear diffusion process can be described as follows:

$$div\left(c(\vec{r},t)\nabla I(\vec{r},t)\right) = \frac{\partial I(\vec{r},t)}{\partial t}\ . \qquad (1)$$

In this equation, I represents the image, $\vec{r}$ is a vector with spatial pixel coordinates and t is the time or scale parameter. The diffusion coefficient $c(\vec{r},t)$, is a function of the image gradient magnitude $\nabla I(\vec{r},t)$, assuming large values for low gradients (i.e. interior regions), and low values in the vicinity of object boundaries, with large gradient magnitude, and defined as follows:

$$c(\vec{r},t) = \frac{c_{max}}{1 + \frac{\|\nabla I(\vec{r},t)\|^2}{\theta}} \; . \tag{2}$$

The parameters c_{max} and θ are, respectively, the maximum admissible value of a diffusion coefficient and a gradient threshold. This results in diffusion being inhibited in the neighborhood of edges and therefore preserving the edge sharpness, but allowed to take place in interior regions.

B. Watershed Segmentation

To build the initial image partition we use the watershed transform applied to the gradient magnitude of the image. The transform views its input as a topographic landscape in which "valleys" correspond to interior pixels whereas "mountains" correspond to pixels in the neighborhood of edges.

We use a rainfalling implementation that takes as input a gradient magnitude image represented with floating point accuracy, which shows a number of advantages when compared to other implementations (see [2], [3] and references therein.) The algorithm basically consists of two logical steps: first, we define a threshold, i.e. a flooding or drowning level, and merge all neighboring pixels whose gradient magnitude lies below this level; this yields a number of "lakes" in the topographic relief. In a second step, for each pixel not yet classified we determine the steepest descent direction, i.e. the direction in which a drop of rain would flow should it fall on the topographic surface formed by the gradient image (hence the designation rainfalling.) The pixel under consideration and the steepest descent neighbor are then merged. In this second step, the segments formed earlier (the "lakes") grow and new regions may be formed too. Once all pixels are classified the segments formed cover the entire image. Hence, a region of the generated image partition, R_i, is defined as the set of pixels belonging to the image domain N^2 such that a vector L holding the segment labels shows the label i at those pixel locations:

$$R_i = \{(x,y) \in N^2 : L[x,y] = i\}. \tag{3}$$

III. Region merging

A. Graph representation

As explained above, a segmented image can be easily represented by a *label image*, i.e. an image with the same dimensions as the original and that, for each pixel position, holds an identifier associated with the segment to which that pixel belongs.

Though simple and extremely compact, this representation of the segmentation poses some problems when one needs further manipulate the current image partition. For instance, with such a label image one cannot directly tell how many pixels constitute a segment. Similarly, it is not straightforward to tell which segments some particular segment is neighboring.

A solution that circumvents these and other problems is a data structure such as the region adjacency graph (RAG) of the image [7]. In this type of graph, each vertex represents a different segment of the image partition while the edge connecting two vertices of the graph represents the neighboring or adjacency relation between the segments associated with those vertices.

The output of the watershed transform, typically a label image, must therefore be converted to this graph representation. This is achieved with a raster scan of the label image, during which the adjacency relations are established and a number of additional data fields in the vertices and edges of the graph are filled in. Once this conversion is complete, the manipulation of the graph itself (and consequently the image partition) is fairly straightforward.

B. Hierarchical merging

Starting from an initial image partition of size M obtained by the watershed transform, for a given *target* partition size $M - k$, there exists an ordered sequence of k region merges that optimises a pre-defined global criterion. As pointed out in [7], finding such a globally optimal merging sequence is a very difficult task, as this is a combinatorial optimisation problem. Even for very short merging sequences it would lead to prohibitively large search spaces. Instead, as it is also the case in [7], we use a suboptimal solution: for each merging step we look for the pair of adjacent regions that optimises the pre-defined criterion. Even though such a sequence cannot guarantee that the final image partition is optimal with respect to the defined criterion (unless this sequence consists of a single merging step!), good preliminary results support its use, not to mention the gains yielded in terms of computation time.

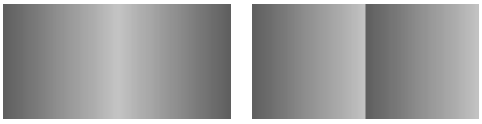
C. Merging criteria

In [7] a dissimilarity function is used such that the mean squared error (*MSE*) of the image partition is optimised. It can be shown that the increase of the *MSE* can be minimised with the merging of the pair of adjacent regions showing the smallest value for the following dissimilarity function,

$$\delta_{MSE}\left(R_M^i, R_M^j\right) = \frac{\|R_M^i\| \cdot \|R_M^j\|}{\|R_M^i\| + \|R_M^j\|} \cdot \left[\mu\left(R_M^i\right) - \mu\left(R_M^j\right)\right]^2, \quad (4)$$

where $\|R_M^i\|$ and $\|R_M^j\|$ represent the number of pixels contained, respectively, in the regions i and j of a certain image partition M, and $\mu\left(R_M^i\right)$ and $\mu\left(R_M^j\right)$, the mean intensities of these regions.

In [8] we introduced a similarity criterion resembling the *Fisher's discriminant* that allows the merging of neighboring regions with a larger region mean intensity difference when their variances are also larger. It yields better results especially in the presence of heavily textured areas [8].



(a) (b)

Fig. 1. The images above illustrate why mean intensity might not always be appropriate as measurement of segment similarity: both left and right halves of these images show identical mean intensity. However (a) seems to contain only one segment while (b), due to the strong contrast in the middle of the picture, seems to contain two segments.

In either case, the intensity contrast at the dividing boundary is not taken into consideration. As clearly illustrated in fig. 1, the exclusive use of the region mean intensity difference as a merging criterion will not always be appropriate to obtain a visually correct segmentation. This evidence suggests that the use of the contrast information at the boundary could yield qualitatively better image partitions. An obvious choice is to use the average gradient magnitude at the common boundary B_{ij} of adjacent regions (*AGB*),

$$\delta_{AGB}\left(R_M^i, R_M^j\right) = \frac{1}{\|B_{ij}\|} \cdot \sum_{Bij} \|\vec{\nabla}\left(I_{B_{ij}}\right)\|, \quad (5)$$

in which $\|B_{ij}\|$ represents the number of pixels constituting the common boundary B_{ij}.

This criterion obviously does not stand on its own because, considering only the region boundaries, it is too local. However, it can be used in combination with other criteria. In this work we have investigated its use in combination with the criterion shown in eq. 4, yielding the following combined criterion:

$$\delta_C\left(R_M^i, R_M^j\right) = \alpha \cdot \delta_{MSE}(.) + \beta \cdot \delta_{AGB}(.). \quad (6)$$

Since the two individual criteria do not show the same scale, it is necessary to re-scale them and hence the need of the additional coefficients α and β. Due to the nature of the image partitions generated by the watershed transform, these separate criteria show histograms of the initial set of dissimilarities which are identical in shape, with a pronounced peak for low values (low dissimilarity) and a smoothly elongated tail for higher values (higher dissimilarity.) Hence, each of the coefficients can be assigned to the inverse of the standard deviation of the correspondent set of initial dissimilarity values, giving them a comparable scale. Other settings are possible, particularly if one wants to give different weights to either sub-criterion.

IV. Experimental results

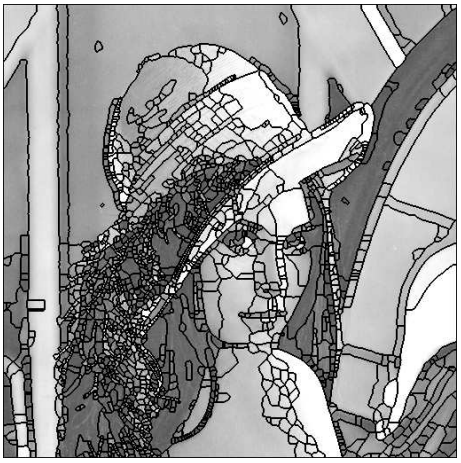
In this section we show a number of results when using the methodologies described in the previous sections.

In fig. 2 we show the original LENA as well as the boundaries of the partition yielded by the watershed transform. For better visualisation and inspection, the segment boundaries are overlaid on a brightened version of the image. These results were obtained after smoothing the original image with 40 iterations of the nonlinear diffusion process, using a c_{max} of 0.1 and a diffusion threshold set to 50% the average squared gradient magnitude of the original image. The drowning level was in this case set to 1×10^{-5}. This example illustrates the typical performance of the watershed transform with real images. The resulting partition is indeed oversegmented but it must be pointed out that the boundaries are very well located.

The partition shown in fig. 2(b) is then used as the starting point of the region merging scheme. In order to make a comparison possible, a merging sequence was forced on this partition until only 60 segments were left. The resulting partitions using with either criteria are shown in fig. 3. Both partitions show a good preservation of the most relevant features but 3(b) seems to be less cluttered.

The image EYE and the boundaries of the partition generated by the watershed transform are shown in fig. 4. The setup for the nonlinear diffusion step was identical to that

(a)



(b)

Fig. 2. The (a) original image LENA (512×512) and (b) the segment boundaries yielded by the watershed transform, 1950 segments.

(a)

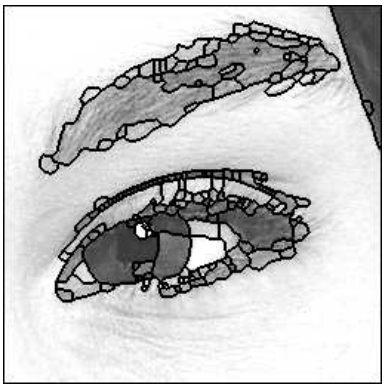
(b)

Fig. 3. The resulting partitions of the image LENA, after region merging using either (a) minimisation of *MSE* or (b) combined criteria, both with 60 segments.

used with the image LENA but in this case the drowning level of the watershed transform was set to 1×10^{-4}. The partitions resulting of region merging with either criteria are included in fig. 5. Similarly to the procedure followed with the image LENA, a merging sequence was forced on the image until a number of segments was left, in this case 40. It becomes even more apparent with this image that the partition yielded by the combined criteria *seems* to show a smaller number of segments. This example image shows more clearly that most regions in this partition are larger than those obtained with the *MSE* minimising criterion. However there are also very small segments that were not merged, located next to a high contrast boundary. This occurs because the average of the gradient magnitude at these very short boundaries is highly biased by the presence of a very high contrasting boundary in their vicinity. However, this potential shortcoming of the combined criteria can be easily suppressed with a size constraint on these very small segments.

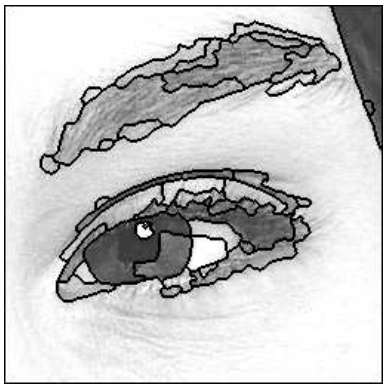
Figure 6 shows a trend of the evolution of the average gradient magnitude on the boundaries and the image mean squared error during the merging process. It is clear that with the combined criteria the mean squared error of the merged partition increases at a faster rate than with the criterion described in eq. 4 (see fig. 6(a)). However, it is also clearly noticeable that the combined criteria always yields a higher average gradient magnitude on the boundaries (see fig. 6(b)) which indicates that more contrasting

(a)

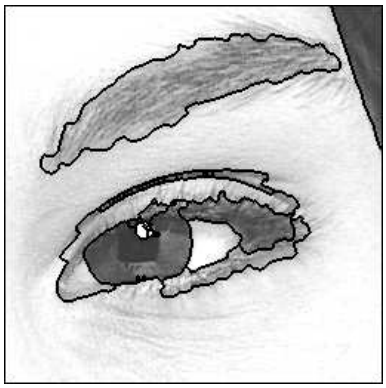


(b)

Fig. 4. The (a) original image EYE (256×256) and (b) the segment boundaries yielded by the watershed transform, 179 segments.



(a)



(b)

Fig. 5. The resulting partitions of the image EYE, after region merging using either (a) minimisation of *MSE* or (b) combined criteria, both with 40 segments.

region boundaries are being preserved.

V. CONCLUSION

In this paper we described an image segmentation method based on the watershed transform and a region merging scheme to further improve the segmentation result. The watershed transform, when provided with a reliable edge indicator as input, yields very well delineated region boundaries. However, even with a pre-smoothed image, it typically results in over fragmented image partitions. The region merging method makes use of a graph to represent the image partitions. We investigated a new region dissimilarity measure that considers the minimisation of the increase of the mean squared error of a partition as well as the intensity contrast on the boundary between adjacent regions. The segmentation results shown illustrate the good performance of these processes.

REFERENCES

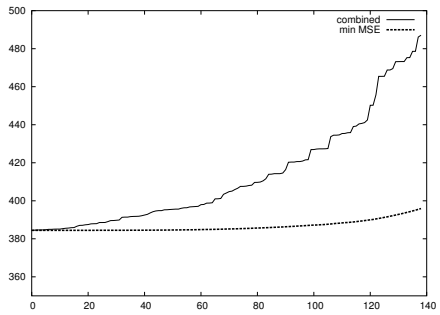
[1] L. Vincent and P. Soille, *Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 13 No. 6, pp. 583–589, 1991.

[2] P. De Smet and R. L. Pires, *On the Implementation of a Highly Optimized Rainfalling Watershed Algorithm*, Proc. of the International Workshop on Very Low Bitrate Video Coding, VLBV99, pp. 101–104, Kyoto, Japan, October 29–30, 1999.

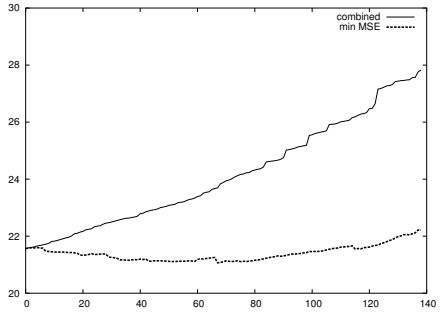
[3] P. De Smet and R. L. Pires, *Implementation and Analysis of an Optimized Rainfalling Watershed Algorithm*, IS&T/SPIE 12th Annual Symposium Electronic Imaging 2000: Science and Technology Conference: Image and Video Communications and Processing (ei26, volume/conference 3974), pp. 759–766, San José, California, USA, January 23–28, 2000.

[4] P. Perona and M. Malik, *Scale-Space and Edge Detection Using Anisotropic Diffusion*, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 12 No. 7, pp. 629–639, 1990.

[5] J. Weickert, *A Review on Nonlinear Diffusion Filtering*, B. ter Haar Romeny, L. Florack, J. Koenderink, M. Viergever (Eds.), Scale-Space Theory in Computer Vision, Lecture Notes in Computer Science, pp. 3–28, 1997.



(a)



(b)

Fig. 6. Evolution of the (a) image mean squared error (*MSE*) and (b) the average gradient magnitude on the boundaries (*AGB*) (both normalised) during the merging process for the image EYE when using either similarity criteria, i.e. minimisation of the increase of the *MSE* or combined criterion.

[6] P. De Smet, R. L. Pires, D. De Vleeschauwer and I. Bruyland, *Activity Driven Non-Linear Diffusion for Color Image Watershed Segmentation*, SPIE - Society for Photo-Optical Engineers, Journal of Electronic Imaging, special section on Nonlinear and Model-Based Image Analysis, pp. 270–278, Vol. 08 (03), July 1999.

[7] K. Haris, S. N. Efstratiadis, N. Maglaveras and A. K. Katsaggelos, *Hybrid Image Segmentation Using Watersheds and Fast Region Merging*, IEEE Trans. on Image Processing, Vol. 7, No. 12, pp. 1684–1699, 1998.

[8] R.L. Pires, P. De Smet and I. Bruyland, *A Region Adjacency Graph based Merging Scheme for the Improvement of Watershed*, Proc. of the IASTED International Conference on Signal Processing, Pattern Recognition and Applications SPPRA'2001, pp. 252-256, Rhodes, Greece, July 3-6, 2001.