

Color-Based Watershed Segmentation of Low-Altitude Aerial Images

Adriano B. Huguet¹, Marcos C. de Andrade²

¹Univ. Federal de Minas Gerais

Departamento de Ciência da Computação
Av. Antônio Carlos, 6627, Belo Horizonte, MG
huguet@dcc.ufmg.br, mca@cdtn.br

Rodrigo L. Carceroni¹, Arnaldo de A. Araújo¹

²Centro de Desenv. da Tecnologia Nuclear

Caixa Postal 941, Belo Horizonte, MG
{carceron,arnaldo}@dcc.ufmg.br

Abstract

In this paper we introduce the Color-based Catchment Basin Merging Algorithm, an area-based segmentation approach that is specifically designed to segment low-altitude aerial images, as a preprocessing step to 3D reconstruction. This approach extends the watershed-based Catchment Basin Merging Algorithm, which is purely geometric, by introducing color similarity as an additional criterion to decide on the merging of evolving regions. Experiments performed with real aerial images of varied nature demonstrate that this modification eliminates over-segmentation problems of the existing algorithm, allowing large-scale urban scenes to be segmented in an accurate, reliable and fully automatic way.

1. Introduction

In this paper we deal with the specific problem of segmenting architectural elements such as roofs, walls and pavement in low-altitude aerial images, so that these segmented elements can later be used as the basis to build 3D reconstruction algorithms specifically tailored to recover the geometry of entire metropolitan areas in a fully automatic way.

Image segmentation is, by definition, the problem of decomposing images into regions that are semantically uniform (e.g., correspond to the visible projections of individual 3D objects). However, since images themselves provide only semantically poor information, image segmentation is essentially an application-oriented problem that demands either strong intervention of human experts or application-specific solutions.

In other words, no fully automatic, general-purpose segmentation method exists. The choice of a particular technique depends on the nature of the task to be performed (3D reconstruction, shape recognition, quality control, localization, measurement), on the nature of the images available (presence or not of non-homogeneous illumination, texture,

ill-defined contours, occlusions, shadows), on the primitives to be extracted (contours, straight segments, regions, shapes, textures) and on physical limitations (real-time constraints, limitations on computational power and storage capacity) [7]. Thus, the literature in this topic is quite eclectic.

One of the most traditional forms of segmentation is region growing and merging [17, 1]. These techniques work well in noisy images, but they are sensitive to seed initialization, which is hard to perform in a fully-automatic way, and are prone either to produce jagged boundaries, or to leak through narrow gaps or weak edges, generating under-segmentation.

Segmentation by deformable models overcomes some of these shortcomings by describing region boundaries as continuous, piecewise-smooth curves that evolve under a suitable energy functional until they (hopefully) converge to the semantically meaningful borders. Unfortunately, the most traditional techniques of this type, such as the snakes method [12], require very good prior knowledge about the topology and even the approximate position of the actual object boundaries in order to guarantee proper convergence and to avoid being trapped by local minima. Some work [28, 8] has been done towards ameliorating this difficulty, for instance by adding an inflation force to the snakes, which helps them escape from local minima. However, this inflation force often pushes the contour over weak edges, which leads back to under-segmentation.

Theoretically, the need for prior knowledge about topology can be avoided by modeling region boundaries within the level-set framework [23, 22], which allows merging of non-significant curves or even the splitting of under-segmented regions. In practice, existing level-set methods [6, 16, 15] require initialization steps that are difficult and time-consuming, such as the manual introduction of polygons around the features of interest. Convergence is also difficult since some curves are still evolving while others have finished their evolution or, worse, have leaked through weak boundaries.

Thus, to some extent or another, all techniques mentioned so far have initialization and convergence problems. In this paper, to avoid this kind of difficulty, we use watershed segmentation [4, 5] - a class of techniques that treat the image as a relief (3D surface), where elevations are correlated with the likelihood of finding an intensity edge at each pixel.

In the Watershed Transformation [27] as proposed by Vincent and Soile (VSWT), region growing starts from the minima of this surface and expands following geodesic paths (paths of minimal energy), employing an efficient sorting-by-address calculation [11]. The region growing process continues until neighbor regions touch each other. This provides a powerful stopping criterion difficult to achieve in the other region evolution techniques. However, VSWT usually leads to a strong over-segmentation of images having significant high-frequency content. Elaborated solutions to this problem exist [3, 2, 10, 9], but they require fine-tuning of parameters related to the geometry of each region to be segmented.

The watershed transformation as proposed by Meyer [20, 18] avoids the need of fine-tuning by starting region growing from pre-defined markers. But this leads back to difficulties in initialization. So, in this paper, we stick with the VSWT, and we study approaches based on it that are specifically tailored to the task of reconstructing large-scale urban scenes from low-altitude aerial images.

Complete 3D reconstructions of scenes of this type can be accomplished using dense stereo techniques, which try to find the mapping between individual pixels of one image and corresponding pixels of a second image that minimizes some pre-defined global image-to-image matching error. Methods that are specifically designed to allow discontinuities in the mapping at occlusion boundaries tend to be more accurate than those that enforce smoothness everywhere, but also tend to be much more expensive computationally, because they have to "search" a much larger space of possible mappings to find the optimal solution to the problem. We want to establish a compromise between these two extremes, in which discontinuities are allowed, but only near previously computed object boundaries.

Although identifying all such boundaries automatically in general scenes is perhaps an unsolvable task, aerial images of urban areas possess certain regularities that facilitate automatic segmentation. Architectural elements such as roofs, walls and pavement (and even canopies of trees) in general possess homogeneous photometric properties and well-defined borders. Scene illumination is also quite homogeneous, and very strong, since the data acquisition flights are almost always performed in cloudless days. In the work described in this paper, we take full advantage of these regularities, by introducing similarity between the av-

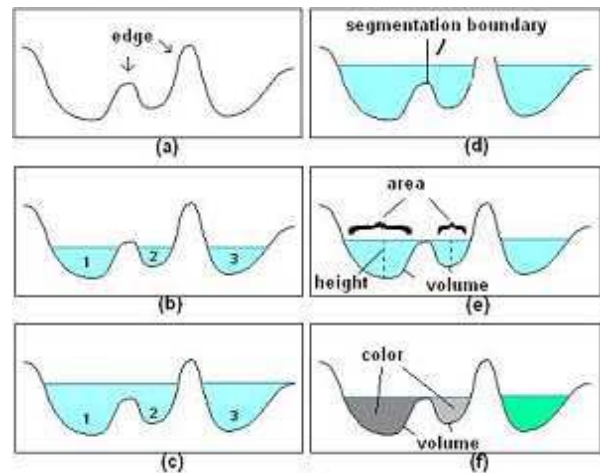


Figure 1. Watershed segmentation: (a) "edgeness" profile, (b) early stage of flooding simulation, (c) later stage of flooding simulation, (d) determination of catchment basin boundaries when a watershed submerges, (e) geometric criteria for deciding on region merging, and (f) color similarity criterion.

erage colors of two segmented regions as an additional criterion to decide if they are joined or kept separate either after or during the execution of watershed segmentation.

As we shall see in the following sections, this idea leads to segmentation algorithms that have no initialization or convergence problems, require little or no parameter tuning and, within the specific context of our intended application, identify object boundaries in a fully automatic and reliable way.

2. Watershed Segmentation with Merging

The watershed transformation (VSWT) segments images through a flooding simulation process (Figure 1). Initially, an edge-detection filter is applied on the input image, producing an "edgeness" profile, i.e., a monochromatic image where the intensity of each pixel is proportional to the likelihood of existence of an edge at that position in the original image. Watershed segmentation treats this "edgeness" profile as digital elevation map, where graylevels are interpreted as elevations.

A simulation is then performed where the terrain represented by this digital elevation map is gradually flooded from all its minima. As the flooding occurs, each minimum gives rise to a unique catchment basin (segmented region). In the moment when a watershed between two adjacent catchment basins submerges, the basins are not joined,

and their boundaries in the previous iteration are used to create a new segmentation edge (Figure 1.d).

A simulation is then performed where the terrain represented by this digital elevation map is gradually flooded from all its minima. As the flooding occurs, each minimum gives rise to a unique catchment basin (segmented region). In the moment when a watershed between two adjacent catchment basins submerges, the basins are not joined, and their boundaries in the previous iteration are used to create a new segmentation edge (Figure 1.d).

Thus, in VSWT there is a one-to-one correspondence between each minimum and its catchment basin. Because an excessive number of local minima are usually found in "edgeness" profiles of real images, VSWT per se tends to generate strong over-segmentation unless careful image denoising is performed beforehand. However, since VSWT segments the image through flooding simulation, avoiding over-segmentation is, at least in principle, only a matter of identifying and merging non-significant regions as the immersion occurs, which can be done with the Catchment Basin Merging Algorithm (CBMA) [3, 2].

CBMA embeds an on-the-fly merging mechanism into VSWT, in order to reduce over-segmentation. The merging process is guided by a set of rules that take into account geometric attributes of the catchment basins such as depth, area and volume (see Figure 1.e). A typical CBMA, when called upon the situation shown in Figure 1.d, would create new segmentation edges only between catchment basins whose volumes (or depths, or areas) in the previous iteration are larger than pre-defined thresholds. Since noise often increases the catchment basins depth, a special rule uses the current area of a basin to delay the computation of its depth (and volume) thus embedding a denoising capability into the algorithm.

CBMA explores the relative difference between local and regional catchment basins properties, using geometric thresholds to decide on merging. The relatively smaller depths, areas and volumes of local (non-significant) catchment basins make it possible to distinguish them from regional (significant) ones, as long as the actual semantic boundaries present in the input image are uniform in scale, which is usually the case.

Moreover, CBMA allows the user to adjust the geometric thresholds interactively, so as to produce the desired segmentation. Automatic selection of thresholds is also possible using double immersion. During the first immersion depth and area distributions are computed. Statistics of these distributions such as their means are then used to establish the thresholds for the second immersion, in which the actual segmentation takes place.

3. Color-Based Watershed Segmentation

Watershed segmentation with CBMA is an elegant, easy-to-use algorithm that works well in many domains, but it is not adequate to the specific application that we target in this paper. Its shortcomings in the segmentation of low-altitude aerial images of urban areas can be observed clearly in Figure 2. Urban scenes usually contain architectural elements that differ greatly in size, such as the mall ceiling that appears on the left part of the images and roofs of the residential neighborhood that is depicted in the center. Since CBMA uses, for each attribute, one-size-fits-all thresholds that are purely geometric, it is bound to either under-segment regions with a lot of small elements or to over-segment the larger architectural structures, as can be seen in Figure 2.b.

In response to this difficulty, we propose two watershed segmentation (CBMA) approaches that use information about the colors within the segmented regions on the input images (not in the "edgeness" profile), in order to perform adaptive region merging. These approaches, which produce the segmentation depicted in Figures 2c and 2d, when given as input Figure 2.a, are described in Sections 3.1 and 3.2, respectively.

3.1. Using color a posteriori

The simplest way to overcome CBMA's limitations is to include region merging based on color similarity as an a posteriori step, performed after the termination of the entire flooding simulation. Since the CBMA assigns unique labels to each region during immersion, the mean color of a region can be easily computed from the resulting labelled image. All pixels pertaining to each segmented region have exactly the same label. Thus, a single scan through the labelled image suffices to compute the mean color of each region. A second scan of the labelled image then allows the deletion of edges between any two adjacent regions whose mean colors differ by less than a preset threshold. This solution merges the regions not yet merged by the rules based on the regions' geometry, dramatically reducing over-segmentation.

3.2. Using color on-the-fly

A more elaborated, statistically-better way of introducing color-based region merging into VSWT is to modify the CBMA to use the differences between average colors of adjacent catchment basins as an additional criterion to decide on region merging, during the flooding simulation. A Color-based CBMA is presented in Algorithm 1. In this algorithm, the average color of the pixels within each segmented region - calculated on the input image - is the attribute that is more relevant to decide whether merging of regions covered



Figure 2. Segmentation of a suburban scene: (a) input image, (b) colorless segmentation, (c) segmentation with color a posteriori, and (d) segmentation with color on-the-fly.

by the same catchment basin is going to take place or not. Only in cases in which at least two such regions within a unique basin have significantly distinct colors, will the volumes of these regions be taken into account at all, in order to avoid the creation of tiny regions in the final segmentation.

3.3. Comparison with related approaches

Color-based watershed segmentation was first introduced by Meyer [19], but that particular work is seed-based - thus, not appropriate to applications such as ours, in which fully-automatic segmentation is required. Even though Lezoray et al. [13] have recently proposed a methodology to compute segmentation seeds automatically in color images, their methodology depends on strong prior knowledge about which colors are more likely to be found in the regions of interest of the scene to identify the seeds reliably [13, 14]. While this works fine in restricted domains such as segmentation of cytologi-

cal images, the accuracy of this kind of technique in more unpredictable domains such as the analysis of aerial images of urban scenes is questionable.

In addition, two of the most popular forms of hierarchical watershed segmentation - the Waterfall algorithm [5] and the segmentation using Dynamics of Contours [21] - have been generalized to color [24, 26]. In both cases, however, region merging is performed after the flooding simulation is finished. Thus, these techniques are quite similar to our a posteriori approach, but not to our approach that uses color similarity on-the-fly. Trémeau and Colantoni [25] have proposed a way to extend essentially any watershed segmentation algorithm to color images through the use of Regions Adjacency Graphs, but their technique, too, works as a post-processing step to the flooding simulation.

4. Experiments

Due to the task-specific nature of segmentation, it is not possible to define a priori how good a segmentation method really is. Although human judgment of segmented images

Algorithm 1 Catchment Basin Merging Algorithm based on the use of color comparisons on-the-fly

Inputs:

$S(k)$, the set of segmented regions at height k ;
 $C(k+1)$, the catchment basins at height $k+1$;
 T_C , the color threshold;
 T_V , the volume threshold;

Output: $S(k+1)$, the segmented regions at height $k+1$;
 Calibrate the cameras and rectify the images geometrically and radiometrically;

$S(k+1) \leftarrow \emptyset$;

for each catchment basin $b \in C(k+1)$: **do**

$r \leftarrow \emptyset$;

if (b intersects more than one region in $S(k)$ and at least a pair of these intersected regions have average colors that differ by more than T_C) **then**

for each intersected region r' : **do**

if ($\text{volume}(r') > T_V$) **then**

$S(k+1) \leftarrow S(k+1) \cup r'$;

else

$r \leftarrow r \cup r'$; (*Merge region*)

end if

end for

else

$r \leftarrow b$;

$S(k+1) \leftarrow S(k+1) \cup r$;

end if

end for

is usually a reasonable estimate of a method's quality, rigorous empirical evaluation of segmentation techniques must be based on the accuracy, reliability and/or performance obtained when the segmented primitives are used in the execution of a given target task, or, if that is not practical, this evaluation must be based on a well-defined, measurable, task-relevant quality metric. In the current paper, we compare three watershed segmentation approaches - a colorless, volume-based one, and the color a posteriori and color on-the-fly approaches described in Sections 3.1 and subsec:3.2, respectively - using the quality metric described in Figure 3.

This metric assigns a value in the range $[0,1]$ to every pixel in the input image. More specifically, it compares the segmented region that contains the pixel with the actual semantic unit to which the pixel belongs. The closer the former is to the latter, the closer the metric's value for that particular pixel is to one, and the larger is the relative discrepancy between these areas, the closer the pixel's quality is to zero. By averaging the values assigned to all pixels, a unique metric is then defined for the entire input image, where a value of one means perfect segmentation, and anything close to zero means a very bad segmentation.

This metric is clearly task-relevant, since it is based on the definition of a semantic unit, which is itself task-

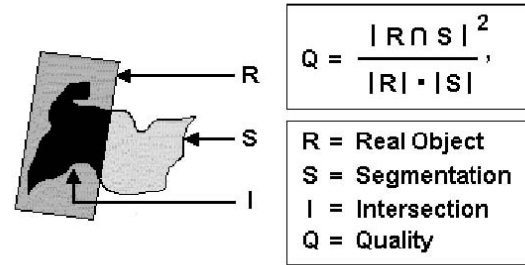


Figure 3. Definition of the quality metric Q , where $|S|$ denotes the area of a region S .

dependent. However, in spite of being well-defined, the metric is difficult to measure in practice, because in the case of our target task, semantic units (i.e., architectural elements) can only be identified by manual segmentation. So, instead of computing the actual value of the metric, we used an approximation to it, based on a manual semantic analysis of the input images.

More specifically, on each image a set of random pixels is drawn from a uniform distribution and each real object that contains at least one of these pixels is manually segmented. Within each of these "ground-truth" objects, the metric's value is then computed on a pixel-by-pixel basis. Finally, the average of all pixel values is calculated and used as a final estimate of the segmentation quality over the entire image.

In addition to the input image displayed in Figure 2.a, we used the images displayed in Figures 5.a and 5 as the input data on which the three approaches being compared were tested. A mere visual analysis of Figures 2.b and 5.b is already sufficient to show that the colorless approach performs severe over-segmentation of the larger objects, which does not happen with the color-based approaches. The distinction between these is less dramatic, but nonetheless, it can be visually observed, especially in Figure 2. The approach that uses color a posteriori misses quite a few relevant boundaries - for instance, those that separate the pavement of the highway that appears on the left of the image from neighboring vegetation areas and houses - that are correctly picked by its color-on-the-fly counterpart. This fact is stressed in Figure 6, where under-segmentation errors that contribute to the evaluated metric are displayed as regions that protrude out of the true objects (shown in black).

Finally, in Table 1 and Figure 7 we display, respectively, the global segmentation quality, Q , of each approach on each input image, and a precision-recovation plot of the pixel-wise quality obtained by the three approaches, for all pixels where the metric was evaluated. Comparing the num-



Figure 4. Segmentation of a downtown scene: (a) input image, (b) colorless segmentation, (c) segmentation with color a posteriori, and (d) segmentation with color on-the-fly.



Figure 5. Other three real images used in the experiments.

bers in Table 1 with the input images in Figures 2, 4 and 5, it can be observed that the advantage of the on-the-fly approach over the a posteriori one is larger in images where object sizes vary more. The precision-recovation plot is consistent with this fact. For each fixed segmentation quality threshold, the ratio of pixels where that threshold is reached

with either one of the color-based approaches is considerably higher than the ratio of pixels where it is reached with the colorless approach. The difference between the two color-based approaches is more subtle, but nonetheless the on-the-fly curve is almost always above the a posteriori curve.

IMAGES							
Approach		Figure 2	Figure 5	Figure 6.a	Figure 6.b	Figure 6.c	Mean
	On-the-fly	0.62	0.39	0.42	0.42	0.30	0.43
	A posteriori	0.47	0.39	0.36	0.31	0.32	0.37
	Colorless	0.10	0.12	0.16	0.24	0.14	0.14

Table 1. Segmentation accuracy in 5 urban images, for each of the three approaches studied in this paper.

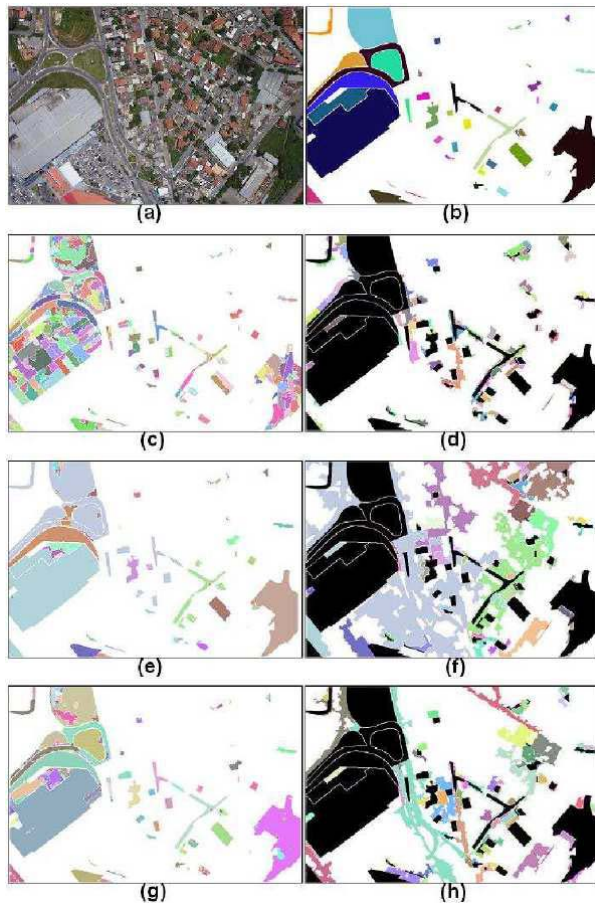


Figure 6. Segmentation errors: (a) input image, (b) manually selected objects, (c), (e), and (g) segmentation inside these objects by the colorless, a posteriori and on-the-fly approaches, respectively, (d), (f), and (h) regions under-segmented by the colorless, a posteriori and on-the-fly approaches, respectively.

5. Conclusions

In this paper we have demonstrated that watershed segmentation can be used as the basis to construct fully auto-

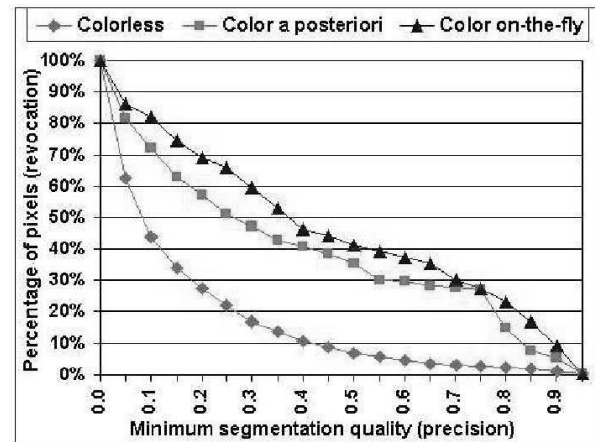


Figure 7. Percentage of pixels with quality higher than each precision threshold, for each approach in study.

matic, reliable techniques for segmenting architectural elements in low-altitude aerial images of urban scenes. Although the basic watershed segmentation algorithm found in the literature suffers from severe over-segmentation in the context of this specific application, regularities that exist in this class of images - especially the homogeneity of radiometric properties within most semantic units - led us to use alternative approaches that add color similarity as a criterion to decide if neighboring regions should be kept separate or should be joined.

As demonstrated in experiments with aerial images of varied nature (suburbs, downtown areas), these modifications led to segmentations of significantly higher quality (according to a well-defined metric), especially if color information is used on-the-fly, during the execution of the watershed algorithm (CBMA) - rather than as an a posteriori step. Currently, we are continuing this work through the development of especially-tailored dense stereo techniques that will use the results of watershed segmentation as the basis to obtain 3D models of large-scale architectural scenes in a fully automatic fashion.

Acknowledgments

The authors would like to thank CNPq, CAPES, and FAPEMIG for their support to this work

References

- [1] R. Adams and L. Bischof. Seeded region growing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15(6):641–647, 1994.
- [2] M. Andrade, G. Bertrand, and A. Araujo. Segmentation of microscopic images by flooding simulation: A catchment basins merging algorithm. In *Proc. SPIE Nonlinear Image Processing VIII*, pages 164–175, 1997.
- [3] M. C. Andrade. *Um algoritmo topológico de segmentação de imagens por atributos*. Phd thesis, Universidade Federal de Minas Gerais, 1998.
- [4] S. Beucher. *Segmentation d'image et morphologie mathématique*. Phd thesis, École Nationale Supérieure de Mines de Paris, 1990.
- [5] S. Beucher. Watershed, hierarchical segmentation and waterfall algorithm. In *Proc. Mathematical Morphology and its Applications to Image Processing*, pages 69–76, 1994.
- [6] F. Caselles. . image selective smoothing and edge detection by nonlinear diffusion. *SIAM J. Numerical Analysis*, 29:183–193, 1992.
- [7] J. P. Cocquerez and S. Philipp. *Analyse d'images: filtrage et segmentation*. Masson, 1995.
- [8] L. Cohen and I. Cohen. Finite element methods for active contour models and balloons for 2d and 3d images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15:1131–1147, 1993.
- [9] M. Grimaud. *La geodesie numerique en morphologie mathématique. Application a la detection automatique de microcalcifications en mammographie numerique*. Phd thesis, École Nationale Supérieure de Mines de Paris, 1991.
- [10] M. Grimaud. A new measure of contrast: the dynamics. In *Proc. SPIE. Image Algebra and Morphological Image Processing*, 1992.
- [11] E. J. Isaac and R. C. Singleton. Sorting by address calculation. *J. of the ACM*, 3:169–174, 1954.
- [12] M. Kass, A. Witkin, and D. Terzopoulos. Snakes - active contour models. *Int. J. Computer Vision*, 1:321–331, 1987.
- [13] O. Lezoray and H. Cardot. Bayesian marker extraction for color watershed in segmenting microscopic images. In *Proc. Int. Conf. Pattern Recognition*, pages 739–742, 2002.
- [14] O. Lezoray, A. Elmoataz, H. Cardot, G. Gougeon, M. Lecluse, H. Elie, and M. Revenu. Segmentation of cytological images using color and mathematical morphology. *Acta Stereologica*, 18(1):1–14, 1999.
- [15] R. Malladi, J. A. Sethian, and B. C. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(2):158–175, 1995.
- [16] R. Malladi, J. A. Sethian, and B. C. Vemuri. A fast level set based algorithm for topology-independent shape modeling. *J. Math. Imaging and Vision*, 6(2/3):269–290, 1996.
- [17] A. Mehnert and P. Jackway. An improved seeded region growing algorithm. *Pattern Recognition Letters*, 18:1065–1071, 1997.
- [18] F. Meyer. Un algorithme optimal de ligne de partage des eaux. In *VIII Congrès de Reconnaissance de Forme et d'Intelligence Artificielle*, pages 847–857, 1991.
- [19] F. Meyer. Color image segmentation. In *Proc. Int. Conf. on Image Processing*, pages 303–306, 1992.
- [20] F. Meyer and S. Beucher. Morphological segmentation. *Visual Communication and Image Representation*, 1:21–46, 1990.
- [21] L. Najman and M. Schmitt. Geodesic saliency of watershed contours and hierarchical segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(12):1163–1173, 1996.
- [22] J. A. Sethian. Tracking interfaces with level sets. *American Scientist*, pages 254–263, 1997.
- [23] J. A. Sethian. *Level Set Methods and Fast Marching Methods*. Cambridge U. Press, second edition, 1999.
- [24] L. Shafarenko, M. Petrou, and J. Kittler. Automatic watershed segmentation of randomly textured color images. *IEEE Trans Image Processing*, 6(11):1530–1544, 1997.
- [25] A. Tremeau and P. Colantoni. Regions adjacency graph applied to color image segmentation. *IEEE Trans Image Processing*, 9(4):735–744, 2000.
- [26] I. Vanhamel, I. Pratikakis, and H. Sahli. Hierarchical segmentation using dynamics of multi-scale color gradient watersheds. In *Proc. Int. Conf. Scale-Space and Morphology in Computer Vision*, 2001.
- [27] L. Vincent and P. Soille. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(6):583–598, 1991.
- [28] C. Xu and J. L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Trans Image Processing*, 7(3):359–369, 1988.