Learning a Nonlinear Distance Metric for Supervised Region-Merging Image Segmentation

Antonio Carlos Sobieranski, Eros Comunello, Aldo von Wangenheim

Federal University of Santa Catarina – UFSC
Laboratory of Image Processing and Computer Graphics – LAPIX
Brazil
{asobieranski,eros}@cyclops.ufsc.br, awangenh@inf.ufsc.br

Abstract

In this paper a novel region-merging image segmentation approach is presented. This approach is based on a two-step procedure: a distance metric is learned from some features on the image, then a piecewise approximation function for the Mumford-Shah model is optimized by this metric. The global optimum of the approximation function is inductively achieved under high polynomial terms of the Mahalanobis distance, extracting the nonlinear features of the pattern distributions into topological maps. The penalizer terms of the Mumford-Shah equation are based on new similarity criteria, computed from the topological maps and the class label information. The results we obtained show a better discrimination of object boundaries and the location of regions when compared with the conventional Mumford-Shah algorithm, even when supplied with other well-known similarity functions.

Key words: image segmentation, distance metric learning, global optimization, Mumford-Shah model

1. Introduction

Image segmentation is one of the most important tools in computer vision and image understanding. Its use can be extended to different application areas, such as medical imaging, robot navigation, aerospace industry, security control and many others. Among the most popular image segmentation concepts are the variational models, more specifically the Mumford-Shah model [1]. This model uses low-level vision to segment a domain $W$ of an image $I$ into segments $W_i$. The basic idea is merging the regions on which the signal $I$ is homogeneous and is delimited by a system of regular discontinuities $K$, in order to make visible the objects on the scene.

The homogeneity of the signal $I$ in the Mumford-Shah model is computed by an approximation function, which measures how similar a segment $W_i$ is to the original image. In a practical context, an approximation function has some kind of similarity function, used to verify the similarity between any two adjacent pixels/regions $u$ and $g$ [2]. The most similar regions are then merged into new meta-regions, until some scale-space parameter is reached.
There are different ways to compute a similarity function for an image segmentation method [3]. The most common similarity functions are distance metrics based on the $L^2$ norm computed from two color vectors ($\|u - g\|$) [4][5]. Other similarity functions take into account the geometry of the color space, playing a very important role in the similarity measuring [6]. For each color space model, an appropriate distance metric should be used to optimize the relation between geometry and similarity. In some cases, the large amount of possible combinations between color space models and distance metrics may make it difficult to choose a good similarity function. Some image segmentation methods need many input parameters in order to compensate for the poverty of the similarity function used to merge the regions [2].

One way to increase the expressiveness of a similarity function is by making use of some knowledge about the target on the image. This knowledge may be considered a kind of side information [7] generally in the form of pairwise constraints\(^1\). The pairwise constraints are used as input data points into $Knn$-based classifiers in order to learn a distance metric. The most adequate approaches are the relevance component analysis (RCA) [8], the discriminative component analysis (DCA) [9], Xing’s method [7], and the Xiang classifier [4]. The latter learns a Mahalanobis distance from the pairwise constraints for data clustering and pattern classification. Xiang also compares his method against the aforementioned methods (RCA, DCA, and Xing’s) using a 5-dimensional vector composed of color and spatial information ($x_p = [r, g, b, x, y]^T$), showing a better classification for the color vectors when used in interactive image segmentation.

Interactive image segmentation approaches can be graph based. The methods are frameworks specifically developed to reduce the complexity of current image segmentation methods [10][11][12][13]. These frameworks compute the likelihood values for each pixel under some similarity criteria and merge the most similar regions. A connection between graph-based methods and distance metric learning is also found in the literature. In [14] a standard graph-based semi-supervised learning method is adapted to segment an image (called Linear Neighbor Propagation). In [15] distance metric learning is introduced into a graph-based method using a gradient descent method to find a locally optimal solution. Sapiro [16] presents an interactive algorithm for soft segmentation of natural images, where the user selects different regions of interest and then adaptive weights are obtained from Gabor filters.

In this paper a new supervised image segmentation approach is presented, where in contrast to the presented methods above the proposed approach is a region-merging method based on the Mumford-Shah energy functional [1]. In this approach a new approximation function and an equivalent boundary verification method is computed over a $q$-order polynomial map (or topological map), obtained from the polynomial Mahalanobis distance metric [5]. This distance metric is able to describe nonlinear pattern distributions by using only similar data points in the training step, instead of pairwise ones. The achieved results show a better delimitation of the region boundaries on the scene, having as input parameter only the $q$-order map and the number of desired regions. We compared our results with varia-

\(^1\)The observer defines which pairs of data points are similar ($S$) and dissimilar ($D$), in the form of $S : (p_i, p_j)$ and $D : (p_i, p_j)$.
tions of the Mumford-Shah model using well-known similarity functions developed for color similarity, such as HSV, Lab2000, LabCMC. Recently, we demonstrated the employment of a learned distance metric given by the Mahalanobis distance, replacing the vector norm commonly used in image segmentation [17]. In this novel approach we show the expressiveness and robustness of the nonlinear topological maps for image segmentation, even when different sets of input data points are used (e.g., composed by the region of interest or the background).

This paper is organized as follows: in Section 2 we present a brief mathematical description on similarity functions and distance metrics. In Section 3 we introduce the Mumford-Shah model. The proposed region-merging image segmentation method is detailed in Section 4. We compare our segmentation results to some Mumford-Shah variants in Section 5. Finally, Section 6 presents the conclusions about this work.

2. Mathematical Background

Finding an adequate similarity function is a key issue in any image segmentation method. Generally it is a composition between a color space model and a distance metric acting in this color space [6]. The most commonly used color space models are RGB, HSV, HSI and CIELab. RGB is the most simple color space, having all color components regularly and equally distributed. HSV or HSI are color space models directly obtained from RGB, whose colors are located as points in a cylinder with its range axis varying from black at the bottom to white at the top [18]. These color models attempt to describe perceptual color relationships more accurately than RGB, while remaining computationally simple. The CIELab is a color space with dimension $L$ for lightness and $a$ and $b$ for the color-opponent dimensions [19]. It is based on nonlinearly-compressed CIE XYZ color space coordinates, and is designed to approximate human vision, aspiring to represent perceptual uniformity.

In a color space model a suitable distance metric should be used to measure the distances between the color vectors. The most common distance metrics are the $L^1$ and $L^2$ norms, but they are properly used to indicate similarity only if variances among dimensions are the same and normally distributed [20]. Although its use should be restricted to the RGB color space, in some “perceptual” color spaces (such as CIELab) they can provide usable solutions due the natural color organization in these spaces. However, each color space has a suitable distance metric that can improve the relation between geometry and similarity. Related work can be found on HSV/HSI color spaces, where a “cylindric” measure was proposed [18]. For the CIELab color models, there are the so-called Lab94, Lab2000 and LabCMC distance metrics [19], specifically developed for similarity measures in these color spaces.

Other distance metrics are based on multivariate data distributions, such as the Mahalanobis distance (MD), the Kullback-Leibler, and the Frobenius norm [21]. These measures use some elements of the tensorial algebra, and they can be used to compute the similarity between any two statistical distributions. The Mahalanobis distance, for example, weights the distance calculation according to the statistical variation of each component, given by:

$$d_M(x, y) = \|x - y\|_A = \sqrt{(x - y)^T A^{-1} (x - y)},$$

(1)
where \( d_M(x, y) \) is the Mahalanobis distance between two color vectors \( x \) and \( y \) and \( A^{-1} \) is the inverted covariance matrix computed from a multivariate distribution. As shown in (1), the MD also reduces to the vector norm if \( A \) is an identity matrix.

These multivariate or statistical metrics have their performance reduced if the pattern is distributed nonlinearly over the space. A category of distance metrics able to describe complex pattern distributions take into account some kind of previous knowledge about the target on the image. These metrics are known as distance metric learning (DML) methods, and are supplied with known class label information. The most relevant DML methods are summarized in a survey by Yang [22] and can be classified into non-supervised and supervised. The non-supervised methods are a family of algorithms developed with known class label from the training data points, which improves the commonly used Euclidean distance. For the supervised ones, a large number of DML are \( Knn \)-based methods using pairwise constraints, a kind of side information [7]. Intermediate categories of DML are the semi-supervised methods, which reduce the human inference in the learning phase [23]. One advantage of the supervised or semi-supervised DMLs is the ability to describe complex patterns in contrast to the pattern distributions used in the learning phase.

In this paper we achieved the global optimum by using only similar data points in the training step instead of pairwise ones. It is more intuitive for the user to identify only the set of data points which belong to the same class. It also obtains a more robust nonlinear discrimination between the color vectors, improving the response of the approximation function for the Mumford-Shah model. This is obtained by the distance metric explained in the next subsection.

### 2.1. The Polynomial Mahalanobis Distance

In this work we avoid some limitations of the pairwise constraints by using the polynomial Mahalanobis distance (PMD) [5]. This distance metric is able to capture the nonlinear features of the pattern distribution by using only similar data points in the learning step. The nonlinear discrimination is guided by a polynomial degree \((q\text{-order})\), which determines how rigorous the similarity must be over a pattern distribution. Its \(q\)-order also provides for better outlier elimination for high polynomial orders, increasing the robustness of the measuring.

The PMD can be obtained from the Mahalanobis distance described in (1), which is also the first of \(q\)-orders in a PMD sequence. The main concept of the PMD is mapping the input data into higher order polynomial terms, creating very high polynomial maps. Let \( S : p_1, p_2, p_3, \ldots, p_N \) be a set of \(m\)-dimensional color vectors, where the \( p_i \in \mathbb{R}^m \) correspond to the similar data points (training set), and \( N \) is the cardinality of \( S \). The first step is to compute the MD metric between two color vectors \( x \) and \( y \) as in (1) and to use the input data \( S \) in the covariance calculation, which is the first \(q\)-order (1st-order, or the MD between \( x \) and \( y \)). The second step is to calculate the second polynomial projection (2nd-order), by mapping all \(m\)-dimensional \( p_k = \{p_{k1}, p_{k2}, \ldots, p_{km}\} \) of \( S \), for \( k = 1, \ldots, N \) basis, into all polynomial terms of order \(q\) or less. For example, consider a two-dimensional vector \( p_k = \{p_{k1}, p_{k2}\} \). The mapping of this two-dimensional vector into 2nd-order polynomial terms is \((p_{k1}, p_{k2}, p_{k1}^2, p_{k2}^2, p_{k1}p_{k2})\).
From a computational point of view, the $q$-order PMD could be obtained by directly mapping all data points into their respective polynomial terms, and so using it in (1). However, for large $m$ and $q$, the number of term combinations makes this distance metric computationally infeasible. A way to compute very large polynomial terms is by using the following framework proposed by [5]:

$$d_{PM}(x, y) = d_{M^2}(x, y) + \sum_{l=1}^{L} d_{M^2}(g_i^l, g_j^l),$$

(2)

where $d_{M^2}(x, y)$ is the Mahalanobis distance with a small positive value $\sigma^2$ used to avoid inversion limitation, $L > 0$ is the maximum number of $q$-order maps ($q = 2^L$), and the arguments $g_i^l$ and $g_j^l$ are the next projections of $x$ and $y$ into their polynomial terms. The PMD is detailed in [5].

3. The Mumford-Shah Model

Mumford-Shah is a variational model that optimizes the manner in which neighboring pixels can be merged into homogeneous regions, and at the same time separated by qualitative boundaries $K$. The Mumford-Shah model uses the energy functional

$$E(u, K) = \beta \int_W (u - I)^2 + \int_{W/K} |\nabla u|^2 dx + \lambda \int_K d\sigma,$$

(3)

which contains a data fidelity term and two regularity terms:

- The first term forces the approximation of a segment $u$ to the image $I$.
- The second term measures and controls the smoothness of $u$ on the open-connected components $W_i$ of $W/K$ ($W$ without the boundaries $K$).
- The third term controls the length, the smoothness, the parsimony, and the location of the boundaries $K$, inhibiting the phenomenon of over-segmentation.

The Mumford-Shah model tries to minimize the functional energy in order to merge different parts of a domain $W$. The coefficients $\beta$ and $\lambda$ make the Mumford-Shah a multi-scale model. The coefficient $\beta$ weights the approximation of $I$ by $u$, related to the scale. $\lambda$ is a regularization contrast parameter which controls the length of the boundaries, limiting the total boundary length of the segmentation. When $\lambda$ is small the minimization forces the results to have many boundaries, whereas large values tend to reduce the boundary length and consequently the total number of regions.

The simplified Mumford-Shah model can be obtained by restricting $u$ to be a local constant function in each $W_i$. Then $|\nabla u|^2 = 0$ and the Mumford-Shah equation is reduced to

$$E(u, K) = \beta \int_W (u - I)^2 + \lambda \int_K d\sigma,$$

(4)

where $u$ is determined by the boundary length $K$ and the homogeneity of the signal, since $u$ is the mean value of $I$ in each $W_i$ and the piecewise function is constant.
4. Supervised Region-Merging Image Segmentation

The proposed supervised approach is a region-merging algorithm based on the Mumford-Shah energy functional model [1]. We explore the simplified Mumford-Shah model in three aspects: a) learning the distance metric to create $q$-order polynomial maps; b) adapting a $q$-order polynomial map as a piecewise approximation function; and c) measuring the equivalent boundary between every adjacent region in order to minimize the boundary length term. These three aspects are explained below.

4.1. Learning the Distance Metric

The learning of the distance metric is composed of two parts: training set definition and the model calibration.

The training set definition is a process performed by the observer, where some features are selected from the image in order to be used as input data points to the model calibration. The input data points are composed by multivariate color distributions which should be approximated in some manner by the distance metric. In Figure 1(a) we illustrate how this process is performed with the starfish image. The white path is the selection defined by the observer, where different kinds of structures were selected (varying from the extremity to the center of the object). The three-dimensional distributions of these colors are also illustrated in (b) and (c) by the RGB decompositions (RG, RB, and GB faces respectively), used here for better visualization of the training set.

Once the training set is defined by the observer, the second part of the learning phase, model calibration, takes place. In this part the distance metric is effectively established. This is performed as follows:

- The set corresponding to the input data points defined in Figure 1(a) is used in the PMD metric as $S$, as discussed in Section 2.1.

- The parameter $L$ of Equation (2) needs to be defined. This parameter controls the number of polynomial map projections that should be created. More complex pattern distributions may require higher polynomial maps and consequently higher values of $L$.

- By modeling the distance metric, the polynomial maps are then obtained, one for each $q$-order, starting from $2^0$ up to $2^L$. The PMD metric is a projective method. All lower orders built from $q \leq L$ are available to be used without new training requirements.

- The total number of polynomial maps available is $L + 1$. These polynomial maps will be later used in the new approximation function and also to verify equivalent boundaries in the merging process.

In Figure 2 we show the polynomial map projections obtained by the PMD metric using the training set defined in Figure 1(a). We also compared the Euclidean ($L^2$), Mahalanobis, and polynomial Mahalanobis distances of different $q$-orders. Columns 1, 2 and 3 show the
similarity responses for each topological map in its RGB decompositions (RG, RB, and GB faces). The blue dots are the pattern distributions defined by the user, as in Figure 1, having a center at the red square, which is also the mean reference vector (µ) where the color vectors are being measured. The responses of the distance metrics are shown in each row of Figure 2 by their topological surfaces, where the more close to white the grouping is, the better is the similarity to µ.

In Figure 2(a) the result of the Euclidean distance metric is presented. The grey levels represent the grouped similarity ranges starting from µ. One notices that the similarity in this topological map is equally distributed over all directions. In (b) the Mahalanobis distance described in (1) is presented, where the distances are weighted according to the statistical variation in each color component of the training set S. The grey level surfaces indicate elliptic discrimination surfaces which provide a more robust discrimination between color vectors when compared to the $L^2$ norm. When used to guide a region-merging image segmentation algorithm, this kind of discrimination function can provides good segmentation results, including a considerable number of outlier points, in particular if S is nonlinearly distributed over the space.

However, with the polynomial Mahalanobis distance we frequently obtain a better discrimination surface for a given training set. This is illustrated by the 2nd-, 4th-, 8th- and 16th-order polynomial maps, shown in Figure 2(c)–(f), respectively. The 2nd-order is directly obtained from the original Mahalanobis distance shown in Figure 2(b), which is the 1st order of a PMD sequence. This $q$-order polynomial map is able to describe nonlinear surfaces, and, as the $q$-order increases, the response of the PMD metric to the training set S is tighter, as demonstrated by the 4th-, 8th-, and 16th-order maps. Higher $q$-order polynomial maps become more restricted to specific domains since they are more closed in S.

The polynomial maps above were created using $L = 4$ and modeled using all dimen-
Figure 2: Topological Maps. In (a) $L^2$ norm. In (b,c,d,e,f) the Polynomial Maps of $q$-order $2^{(0,1,2,3,4)}$, respectively.
sionality information from the set $S$. The RGB cube decompositions are only used for visualization purposes and a better understanding of the learning process. In the next subsection we present the use of these polynomial maps as piecewise approximation function for the Mumford-Shah model.

### 4.2. The Piecewise Approximation Function

The Mumford-Shah model can be considered complex to be implemented. One of the major challenges is to develop efficient algorithms to compute high quality minimizers of this functional [24] due to the non-regularity of the edge term. The difficulty is that regions are two-dimensional entities and the boundaries one-dimensional ones, interacting in a very subtle way. In [25][26] it is shown that the boundaries can be two-dimensional if two adjacent segment pairs $O_i, O_j$ have a common boundary $\partial(O_i, O_j)$ between them. The simplified Mumford-Shah equation can be defined as

$$
\lambda \cdot \partial(O_i, O_j) \leq |O_i| \cdot |O_j| \cdot \|u_i - u_j\|, 
$$

where $|O|$ is a scale parameter corresponding to the surface measure, the vectors $u_i$ and $u_j$ are the mean values of $I$ in the regions $O_i$ and $O_j$, and $\|u_i - u_j\|$ is the similarity computed from the mean values of the segments. The algorithm uses an ordered stack of energy estimations, where the top element contains the boundaries between regions where the minimization is optimal. The algorithm starts by merging the top-list boundaries and updating the stack at each merging performed. When the parameter $\lambda$ is achieved by the best energy minimization of the stack, the algorithm stops.

Equation (5) is equivalent to the Mumford-Shah model, but the exact optimization of this kind of functional is very difficult [27]. This equation computes the similarity between the regions $u_i$ and $u_j$ considering only the mean values on each open-connected region, since the piecewise approximation function is constant. For color images, this process is considered as a higher dimensionality problem and the $L^2$ norm is used. However, inconsistent merges when computing the real perceptual difference between the color vectors may occur. Another limitation is related to the boundary length, as discussed later.

In order to guide the energy minimization of the Mumford-Shah model, we use a $q$-order polynomial map to measure the quality of the approximation between every adjacent region,

$$
\lambda \leq \frac{|O_i| \cdot |O_j|}{|O_i| + |O_j|} \cdot \frac{\delta(u_i, u_j)}{\phi \cdot K}, 
$$

where $\phi$ is an equivalent boundary computed between the regions $O_i$ and $O_j$ (next subsection), $K$ is the boundary size $\partial(O_i, O_j)$, and $\delta(u_i, u_j)$ measures the similarity between $u_i$ and $u_j$ in the topological map as follows:

$$
\delta(u_i, u_j) = \|u_i - u_j\| \cdot \left(1 - e^{-d_{PM}(u_i, u_j, q)}\right). 
$$

The equation above is a normalization process used to change a $q$-order polynomial map into non-binary values varying in the range $[0-1]$. Similar color vectors are as near as possible
Figure 3: In (a) the boundaries length between two adjacent regions are shown in the spatial domain $\mathbb{R}^2$. In (b) we show the class label pertinence method used to find an equivalent boundary between the regions in $\Omega$ ($\mathbb{R}^3$).

to 0, and non-similar ones to 1, depending on the $q$-order. The variable $q \leq L$ defines a polynomial map degree used to measure the similarity between $u_i$ and $u_j$, and $\gamma > 0$ is a contrast parameter which controls the compactness level for the topological maps (the grey grouped ranges starting from $\mu$, shown in Figure 2). If both color vectors $u_i$ and $u_j$ are near to $S$, the distance between them is as small as possible, and the estimated energy is minimized even when high values of surface measure are obtained. Otherwise, the energy is normally computed by the $L^2$ norm. Small regions are merged because their surface area is lower, while the large ones remain unchanged.

4.3. Computing the Equivalent Boundary

The minimization of the Mumford-Shah energy functional is a compromise between the similarity computed from every adjacent region and the boundary length which divides these regions. The boundary length $K$ is computed by $\partial (O_i, O_j)$, but it presents some limitations when contrasted to the continuity of a segment pair. The estimated energy on each boundary increases according to the surface measure of these regions, but at the same time the boundary length runs in a different scale. In Figure 3 some limitations when trying to minimize the energy are shown:

- In Figure 3(a), at the top, the boundary $K$ dividing the regions $O_i$ and $O_j$ is large. By Equation (5), large boundaries contribute to merge adjacent regions, even when the similarity is not good.

- In Figure 3(b), at the bottom, the counterpart is presented, where the merging is inhibited due to the small boundary $K$ which divides these regions $O_i$ and $O_j$, even if they are similar.

Clearly, during the merging large adjacent regions connected by small boundaries are frequently lost to the background and the homogeneity and continuity of the segmentation result is ignored. In our approach the observer is part of the segmentation process and we can use some extra information about the target on the image. We compute a class label pertinence to find an equivalent boundary in the topological map between every adjacent region. Let $\Omega \subseteq \mathbb{R}^3$ be a region defined in the space and bounded by the intensities of a
a $q$-order polynomial map, where the set $S \in \Omega$ (the training set), and consider two adjacent regions $O_i$ and $O_j$ whose multivariate color distributions are scattered in $\mathbb{R}^3$. These two regions can be considered equivalent between them if and only if both regions are located in $\Omega$, or both are not in $\Omega$. These statements suggest that if two regions belong to the same class the boundary length which is computed between them also needs to be minimized to reduce the energy functional. This is computed as

$$
\phi = \begin{cases} 
0, & (O_i, O_j) \in \Omega \\
0, & (O_i, O_j) \notin \Omega \\
1, & O_i \in \Omega, O_j \notin \Omega \\
1, & O_i \notin \Omega, O_j \in \Omega 
\end{cases}
$$

where the values are obtained by a probability density function which measures the divergence between two multivariate distributions in a topological map, given by

$$
\phi = |P(O_i) - P(O_j)|,
$$

where $P = 1/n \sum_{x=1}^{n} \Omega(O_x)$, which yields the intensities in the topological map for each color vector $O_x$. In Figure 3(b) we illustrate how this process is performed. The region $R_1$ has two adjacent regions ($R_2$ and $R_3$). If both regions $R_1$ and $R_2$ are located in $\Omega$, then the term which computes the boundary length is avoided, and the same procedure is performed if both regions are located outside of $\Omega$. These regions will have priority on the merging stack, no matter how similar $R_3$ is to $R_1$. This statement does not ensure that a region like $R_3$ will never be merged with a region located in $\Omega$. $R_3$ should be merged when the regions whose similarities are better were already merged and there is no other adjacent regions having better similarity.

5. Experimental Results

In this Section we present the results we obtained by the proposed supervised region-merging image segmentation method, as well as a comparison against some Mumford-Shah variants with approximation functions based on color distance metrics. We call our approach “supervised” because the observer is able to guide the merging process towards similar or even completely unrelated regions. This is possible because the training set contains input data points selected from the input image $I$. The method creates the $q$-orders polynomial maps and they are used to compute the similarity between the color vectors.

In Figure 4 some results achieved by this method are presented, where different sets of input data points were selected from the same image. For each case the segmentation results are varying in a range of two to eighty regions.

- Figure 4(a): By using this training set the red and the yellow areas have a natural tendency to be merged earlier than the other regions. The target region is then merged into a homogeneous region and the background is “fine grained” and broken into many subregions. By decreasing the total number of regions, the background is also merged into a unique homogeneous region. To produce these segmentation results we used a 2nd-order polynomial map.
Figure 4: Merging evolution using the proposed approach. In (a), (b) and (c) three different kinds of training sets (white paths). At the right side, their respective segmentation results varying from 2 to 80 regions.

Figure 5: Merging evolution using the proposed approach. The white paths are the patterns defined by the observer. At the right side their respective segmentation results varying from 2 to 80 regions are shown.
Figure 4(b): In this selection we are interested in preserving petals (red) and the core (yellow) separated. With the 5 most significant regions we achieved good segmentation results and as the number of regions decreases the regions are also merged into a unique region formed by the object of interest. To produce these segmentation results we used a 4th-order polynomial map.

Figure 4(c): In this case we force the merging among the background areas. This was already obtained by the 80 most significant regions, while the flower is “fine grained” in many regions. To produce these segmentation results we used a 2nd-order polynomial map.

For these three examples the segmentation results containing the minimum number of regions (2 regions) are very similar. In addition, two very important functionalities are demonstrated: orienting the merging for unrelated regions (in similarity aspects) and preserving the regions unchanged until a minimum number of segments is reached.

In Figure 5 we demonstrate the use of the proposed approach in other application domains. The object of interest was seeded in (a) and (b), and the background in (c). The q-orders used in these examples were 1st, 4th and 1st, respectively. These results were obtained from textured images (like in a and b), whose gradient variations are strong.

In Figure 6 we present a brief comparison between the proposed approach and some variants of the Mumford-Shah model. From the left to right, the original image, the Mumford-
Shah model based on HSV cylindrical measure [18], the CIELab 2000 metric [19], the CIELab CMC metric [19], the Mahalanobis distance [17], and our proposed approach. The figures used in our experiments were selected from the Berkeley Image Dataset [28], and the training set was defined over the object of interest (except in (a), where the background was used). For each input image a fixed number of desired regions was used, being 2, 4, 10, 4, 5, and 8 regions for the rows (a) to (f), respectively. The $q$-orders used to produce these results were the $2^{nd}$, $2^{nd}$, $4^{th}$, $1^{st}$, $4^{th}$ and $2^{nd}$ for the images (a) to (f), respectively. The results we obtained have in general shown a better delimitation for the region of interest and the background, while in other methods we frequently lost some region to the background.

The amount of execution parameters in this approach was reduced to the number of desired regions and the $q$-order polynomial map, using RGB as the color space. Similar results, however, can also be achieved in other $q$-order maps depending on the pattern distribution of $S$. In variants supplied by the Lab2000 and LabCMC the combination of the parameters may vary depending on lightness and chromacity.

6. Conclusion

Variational models, such as the Mumford-Shah, are widely used in computer vision and image understanding. While this energy functional already yields a quite general of the image segmentation problem as an optimization problem, the way how to minimize the energy has been left open [27]. The major challenge is to develop efficient algorithms to compute high quality minimizers of this functional [24].

In this paper we demonstrated the use of induction to try minimize this energy functional. This process is performed by a distance metric learning method, where topological maps are obtained to gradually discriminate the set of colors that are imperceptibly/acceptably close to a given reference. These topological maps are then used as an approximation function in the Mumford-Shah model, minimizing the energy functional and approximating the similar regions. The segmentation results we obtained show a better matching to the original image for a minimum number of regions using RGB as the color space model. We compared our results against other Mumford-Shah variants using approximation functions supplied by well-known color distance metrics (HSV, Lab2000, LabCMC, and Mahalanobis). The main advantage of our method is the possibility to select different kinds of input data points from the same image, varying the segmentation results.

Interactive image segmentation methods found in the literature learn a specific distance metric from the pairwise constraints to classify the input image into a bimodal segmentation such as presented in [4]. In our approach we orient the merging process starting from low-level cues by a variational method using a nonlinear topological map.

Acknowledgments

We would like to thank Adiel Mittmann, Sylvio Luiz Mantelli Neto, Leandro Coser and Daniel Aguiar, from the Federal University of Santa Catarina, in Brazil, and Peter Barth from Hochschule RheinMain, in Germany, for their valuable insights into our work.
References


