Learning a color distance metric for region-based image segmentation

Aldo von Wangenheim¹, Antonio C. Sobieranski¹, Daniel D. Abdala², Eros Comunello³

¹Image Proc. and Computer Graphics Lab – LAPIX, Federal UFSC - University of Santa Catarina - Brazil
awangenhi@inf.ufsc.br
asobieranski@cyclops.ufsc.br

²CS Department, University of Münster, Münster, Germany
abdalad@uni-muenster.de

³Graduate Program in Applied Computer Science, CS Department, Universidade do Vale do Itajaí – Univali, São José, Brazil
eros.com@univali.br

Abstract

In this paper we describe an experiment where we studied empirically and in depth the application of a global adaptive color space to be used in the similarity function of an established region-growing color image segmentation algorithm. To perform this experiment we chose the Mumford-Shah energy functional and the Mahalanobis distance metric. The objective was to test the approach empirically and in an objective and quantifiable way on this specific algorithm when using this particular distance model, without making any generalization claims. The empirical validation of the results was performed applying the resulting segmentation method on a subset of the Berkeley Image Database, an exemplar image set possessing ground-truths and validating the results against the ground-truths using two well-known segmentation validation methods, the Rand and BGM indexes. The obtained results suggest that to employ this adaptive color distance approach provides better and more robust segmentations, even if no other modification of the segmentation algorithm is performed.

Keywords: Region-Growing Color Image Segmentation, Mumford-Shah Algorithm, Mahalanobis Distance, Non-Isotropic Distance Measures.

1. Introduction

Color similarity computation in region-growing segmentation algorithms is usually based on an Euclidean norm calculated from the color vectors, where all channels are regular and equally weighted. If an object is composed by multicolored information dispersed into the color space, such as the starfish shown in Figure 1, the scale space parameter needs to be higher to allow the merging of the entire object, leading to segment leakages. Other color spaces could be used to improve the quality of similarity (e.g. HSI or YPbPr), but these are also linear color spaces and the segmentation in some cases will still require post-processing. Some examples of post-processing algorithms presenting interesting approaches are [1, 2, 3 and 4].

The usage of distance metrics and similarity functions is addressed in many image processing works. In [5] an active contours method applied Bhattacharya distance to drive the gradient flow. Renner [6] states that a color space is inherently non-linear and uses Riemannian geometry and harmonic maps to develop an anisotropic diffusion filter. Shah [7] studied earlier the usage of anisotropic diffusion aiming image segmentation. Shapiro [8] developed a semi-automated image segmentation method using Gabor Filters that, through manually selected image features, computes adaptive weights to drive a supervised segmentation. These features compose a connected network that defines the segmentation regions, and the number of objects is then defined by manually selected seeds. A general methodology for the application of a supervised region growing based on non-isotropic properties of the color space for color image segmentation, however, has still not been established.

In this paper we describe an experiment where we studied empirically the application of a global, image-specific, adaptive non-isotropic color space to be used directly as the similarity function of a traditional region-growing color image segmentation algorithm. To perform this experiment we chose a well-known region-growing segmentation algorithm and an also well-known distance metric. The objective was to first test the approach empirically on this specific algorithm when using this particular distance metric, without making any generalization claims. The empirical validation of the results was performed applying the resulting segmentation method on an exemplar image set presenting ground-truths and validating the results, which were repeatedly calculated for a large set of different parameters. For each segmentation result, the ground-truth evaluation was applied using two well-known segmentation quality measures, developed especially for image segmentation evaluation.
Our approach is based on a global non-isotropic color space defined by the Mahalanobis distance [9], which is used to guide the Mumford-Shah segmentation algorithm [10]. We choose to select this algorithm because it is an extremely simple and robust segmentation algorithm that employs a similarity criterion based on the signal homogeneity computed within each region. In our approach, we have changed the similarity function to accommodate a non-isotropic color space and express the quality of approximation between adjacent regions in this space.

In order to properly evaluate the output of the similarity function used, we designed a validation experiment that confronts results from the original Mumford-Shah model using the simple vector norm in the RGB color space, against our approach. This experiment compares the results of processing 22 images selected from the Berkeley Image Processing Dataset [12] against their ground truths using a quantitative segmentation evaluation computed using the Rand and BGM metrics [13, 14].

2. Methods
2.1. Color Distance Computation
In this experiment we adopted an adaptive non-isotropic color space generated from features selected from the input image in the RGB color space. The use of the RGB as our initial color space can be justified by the regularity and color constancy presented in all color vectors. The properties of the adaptive color space are based some similarity or dissimilarity function captured from the original image by means of seed points. These seed points are mapped into a matrix-valued second order tensor, preserving the properties of positive-definite matrix, and describing the magnitude and direction over the color space. The Mahalanobis distance weights the color space into a new adapted color space, based on the eigensystem produced by the properties previously described.

The Mahalanobis distance [9] is a statistical metric that allows the computation of the distance between arbitrary coordinates defined upon multivariate data. Taking the mean color values of two regions \( u \) and \( g \), where \( u \) and \( g \) are \( d \)-dimensional column vectors, the similarity between these two regions can be expressed as:

\[
d_{mg} = \left( u - g \right)^T \mathbf{C}^{-1} \left( u - g \right),
\]

where \( \mathbf{C} \) is the covariance matrix. If \( \mathbf{C} \) is the identity matrix \( d \times d \), the Mahalanobis distance is reduced to the Euclidean norm \( || u - g || \).

After taking a set \( Z \) of \( n \) observations of the image to be segmented (seeds) \( \{z_1, z_2, \ldots, z_n\} \), the covariance matrix can be estimated as in Eq.2:

\[
\mathbf{C} = \frac{1}{N-1} \mathbf{A}^T \mathbf{A},
\]

where \( \mathbf{A} \) is a \( n \times d \) dimensional matrix. The column vector \( \mathbf{A} \) can be computed as show in Eq.3:
Learning a color distance metric for region-based image segmentation

\[ A = \begin{bmatrix}
z_1^T - z_0^T \\
z_2^T - z_0^T \\
\vdots \\
z_n^T - z_0^T
\end{bmatrix}. \quad (3) \]

Now, let \( z_i \) be column vectors with dimension \( d \) (\( z_i \in \mathbb{R}^d \) for all \( i \)) of \( Z \), the reference point to verify the covariance level is \( z_0 \), which can be expressed by the mean of the vector \( z_1, z_2, \ldots, z_n \). It also can use any other arbitrary point to define the center of the color space. In this paper we have used the mean of \( Z \) as reference point. The matrix representing \( C \) is

\[ C = \begin{pmatrix}
\sigma_{z_1}^2 & \cdots & \sigma_{z_1 z_n} \\
\sigma_{z_2}^2 & \cdots & \sigma_{z_2 z_n} \\
\vdots & \ddots & \vdots \\
\sigma_{z_n}^2 & \cdots & \sigma_{z_n}^2
\end{pmatrix}. \quad (4) \]

Eq.1 shows that the computation of the Mahalanobis distance requires the inversion of the covariance matrix \( C \), which can be obtained through the singular value decomposition – SVD of \( A^T A \).

2.2. Similarity Function of the Segmentation Method

In order to apply the Mahalanobis distance to build an adaptive color space and use it as similarity function in our tests, we had to take into consideration that the Mumford-Shah method works comparing a region with the adjacent ones, where the comparison occurs between the mean of color vectors of two regions \( u \) and \( g \) [10]. During the merging process, Eq.5 represents the partial energy resulting from the similarity between regions \( u \) and \( g \) regardless of their shared boundaries [10]:

\[ \beta \int_{\Omega} (u - g)^2 \, dx \quad (5) \]

Based on this term, the Mumford-Shah method computes the quality of approximation between two color vectors using the Eq. 5. The modified Mumford-Shah energy functional is given by Eq.6, where the similarity is computed under the covariance given by \( C^{-1} \).

\[ E(u, K, C^{-1}) = \beta \int_{\Omega} (u - g)^T C^{-1} (u - g)^T dx + \lambda \ast L(K), \quad (6) \]

The Mumford-Shah implementation used in our experiments was based on the work originally developed by [11] based upon a previous implementation from [15]. The version embedding the adaptive color space was developed by our workgroup as described in [15]. Mostly only the similarity function used to compare regions during the region merging process was changed in order to accommodate the adaptive color space.

2.3. Validation Method

In order to objectively validate the quality of the results obtained with our approach, the image segmentations were evaluated using quality measures especially developed for this task. The goal of these measures is to compare an image segmentation to a given ground-truth image and quantify the quality of the segmentation results. There are several approaches to compute this kind of similarity. According to [16], there are metrics that evaluate ground-truth-conformity through counting of pairs and set matching. In our tests, we used one exemplar measure of each kind: Rand Index [13] and Bipartite Graph Matching (BGM) [14], respectively a pairs-counting and a set-matching measure. They are briefly explained below.

2.3.1. Rand Index

The Rand index first reported in [13] and reviewed in [14] is a similarity measure developed to the evaluation of the quality of clustering algorithms through comparison against each other or against ground-truths. To compare two different clustering results \( C_1 = \{c_{11}, c_{12}, \ldots, c_{1N}\} \) and \( C_2 = \{c_{21}, c_{22}, \ldots, c_{2M}\} \) of the same image \( P = \{p_1, p_2, \ldots, p_k\} \) where each element of \( C_1 \) or \( C_2 \) is a subset of \( P \) and \( c_{1j} = \{p_{i_1}, p_{i_2}, \ldots, p_{i_L}\} \), the following quantities are calculated:

- \( N_{11} \) - the number of pixels in the same cluster in both \( C_1 \) and \( C_2 \).
• $N_{00}$ - the number of pixels in different clusters both in C1 and C2.

The Rand index is defined by Eq. 7:

$$R(C1, C2) = 1 - \frac{N_{11} + N_{00}}{n(n-1)/2}$$

To compute the quantities $N_{11}$ and $N_{00}$ it is necessary to iterate over the entire image for each pixel in order to evaluate the conditions defined before, given an $O(n^4)$ algorithm. A clever approach is to use the method described in [14], where a matching matrix is used to summarize the pixel occurrences in their respective classes. The matching matrix is constructed allocating each cluster from the clustering C1 to a row and each cluster from clustering C2 to a column. The matrix cells are then defined as the intersection of the clusters specifying each row and column. If the matching matrix has $k \times l$ size, each cell can be defined as $m_{ij} = |c_i \cap c_j|$, $c_i \in C1, c_j \in C2$.

The quantities $N_{11}$ and $N_{00}$ are computed as follows:

$$N_{11} = \frac{1}{2} \left( \sum_{i=1}^{k} \sum_{j=1}^{l} m_{ij}^2 - n \right)$$

$$N_{00} = \frac{1}{2} \left( n^2 - \sum_{i=1}^{k} n_i^2 - \sum_{j=1}^{l} n_j^2 + \sum_{i=1}^{k} \sum_{j=1}^{l} m_{ij}^2 \right)$$

where $n$ is the cardinality of P and $n_i$ and $n_j$ are the cardinality of the clusters $c_1i$ and $c_2j$.

### 2.3.2. Bipartite Graph Matching

The BGM index [14] computes a one-to-one correlation between image element clusters trying to maximize their relationship. There are two clusterings: C1, representing the segmentation, and C2, representing the ground truth. It considers each cluster of the C1 and C2 clustering as vertices of a bipartite graph. Edges are added between each vertex of the two partitions and they are valued as $|c_1i \cap c_2j|$, a value that can be directly extracted from the matching matrix. The maximum-weight bipartite graph is then defined as the sub graph $\{ (c_1i_1, c_2j_1), \ldots, (c_1i_r, c_2j_r) \}$ where only the edges from $c_1i$ to $c_2j$ with maximum weight are present:

$$BGM (C1, C2) = 1 - \frac{\sum_{i=1}^{r} \sum_{j=1}^{j} \max(v_{ij})}{n}$$

Where $v_{ij}$ are the edge weights, $i$ the number of clusters in C1 and $j$ the number of clusters in C2. The total number of elements is given by $n$. The max($v_{ij}$) guarantees that only the maximum weight value for each partite connection should be computed in the final weight summation.

### 2.4. Validation Procedure

The segmentation results produced were validated against the ground-truth images provided by the Berkeley image dataset with both distance measures. Both Rand and BGM produce values between 0 and 1, where 0 denotes a perfect match between segmentation and ground truth and 1 stands for no relation at all between them. The fact that both measures are in the same range facilitates comparing the different approaches.

During validation each image was processed with both segmentation methods several times, each time with a different $\lambda$ parameter, as described in Section 3. All results were tested with both quality measures described below. The two results that showed the best index for each quality measure were selected as the best results for each approach. Results were plotted as bar graphs and are discussed in Section 4. Additionally, the whole dataset was made available online at: http://www.lapix.ufsc.br/noniso_imseg.

### 3. Experiment and Results

The experiment described here aimed at a direct comparison between the original Mumford-Shah algorithm, using a linear distance in the RGB color space, and the version using the non-isotropic color space, which in our experiment uses the Mahalanobis distance to generate it. We choose to perform a direct numerical comparison between the segmentation results produced as described in [14]. Since ground truths are needed to perform an objective numerical comparison, we have chosen to use an image subset extracted from the public available Berkeley Image Database [12].
These images were selected given their low variance and standard deviation with respect to human-made segmentations.

The RGB implementation was run over all 22 selected images with $\lambda$ varying from 512 to 65,536 with an exponential (base 2) increment, producing 8 segmentation results for each image: $\lambda = 512, 1024, 2048$ and so on. For the proposed approach, the chosen parameter range for $\lambda$ was from 2 up to 65,536, also with an exponential increment. A second parameter was used in order to execute in the proposed approach, namely the adapted color space, represented by $C'$, obtained from a set of seed points selected for each image by the user with focus in collect samples of the most representative regions of the image. The number of seed points can vary greatly from image to image, having an average of 125 points within the scope of this study.

Figure 2 – a) original image #113016. b) ground truth-based border image. c) seeded image. d, e) segmentation results using the RGB as color space presenting their best Rand index and best BGM, respectively. f, g) segmentation results using the adaptive color space generated by the seed points (c) presenting respectively the results with best Rand index and best BGM index.

In both cases, a given $\lambda$ produces a single region, the batch process stops itself preventing further processing. In some cases, ideal $\lambda$ values can lay between the exponential base 2 sequences, an empirical sub-processing routine reprocess these intermediate parameters trying to improve the next step of evaluation analysis. The intervals with better visual results were selected and additional segmentation results were produced with the interval given by Eq. 11.

$$\lambda_{\text{step}} = \frac{\lambda_{\text{max}} - \lambda_{\text{min}}}{10}. \quad (11)$$
The Fig.3 gives an example of how the experiment was conducted to validate the Mahalanobis distance to generate the adaptive color space for the color image segmentation. An input image together with its GT (Figure 3-b) is segmented by linear and adaptive approaches. The Figure 3-c shows the seeded image used to model the adaptive color space. White circles are the collected features or points used to represent it. As results, we have produced two categories of validation: a) by the best rand index and b) the best BGM index.

The Figures 2-d and 2-e show segmentation results obtained with the traditional Mumford-Shah segmentation using the RGB as color space. Results of the employment of the non-isotropic color space to drive the Mumford-Shah segmentation (as described by the Eq.6) are demonstrated in Figures 2-f and 2-g. Some other results selected using the Rand index are shown in Figure 3 and the same original images, but with results selected using the BGM index are shown in Figure 4. The whole dataset we processed, containing all results for all parameters and also all used seed points, can be accessed online at: http://www.lapix.ufsc.br/noniso_imseg.

The graphics presented in Figures 5 and 6 show respectively the best Rand index and the best BGM index results obtained for each image, after running both segmentation algorithms for all $\lambda$ values in the selected range. In 95.4% of the images, the best Rand index was that of a segmentation generated with the proposed approach. For the BGM index validation, this rate was of 74.2% of better segmentations produced with the proposed approach.
Figure 3 – Segmentation results selected by the best rand index. a) original image. b) segmentation result using RGB color space. c) segmentation result using the adaptive color space. Applied $\lambda$-values are shown on each image.
Figure 4 – Segmentation results selected by the best BGM index. a) original image. b) Segmentation result using RGB color space. c) segmentation result using adaptive color space. Applied $\lambda$-values are shown on each image.

An exemplar plot for the image #295087 of the behavior of the Rand and BGM indexes for both linear and adaptive color spaces as the $\lambda$ parameter varies is shown in Figure 7.

The $\lambda$ parameter for which best results were achieved varied both between linear (RGB) and adaptive color spaces and also between images, as can be seen in Figures 5 and 6. However, there was observed a systematic behavior of the Rand and BGM indexes, which was similar for the linear and the adaptive color models and that can be described as follows:

- **Rand index**: the graphic has the general form of a bowl and Rand results become better as the $\lambda$ increases, until some sub segmented regions start to appear in the resulting image, when the Rand index starts to worsen again.
- BGM index: the graphic performs a continuous descent, with better BGM values for each increase of $\lambda$, until a strongly sub segmented result is produced. It occurs very often that there is a plateau before this strongly sub segmented result appears, where the BGM index does not change significantly anymore.

Graphic representations of the behavior of the Rand and BGM indexes with different $\lambda$ values for all images used in this work can be seen online at: http://www.lapix.ufsc.br/noniso_imseg.

Figure 5 – Segmentation quality validation by the best Rand index.

Figure 6 – Segmentation quality validation by the best BGM index.
4. Discussion and Conclusion

Our approach generated segmentation results that were generally better than those generated with the same algorithm using a standard RGB as color space. This can be observed through visual inspection of the resulting segmentation images. We also have shown that, using two different objective segmentation quality metrics, Rand and BGM, based upon two very different paradigms, our approach also proved to be systematically better than the traditional linear color model.

A point that has to be observed, and that could be considered a weakness of this approach, is that our color metric was manually generated specifically for each image. The question if a tailored metric is a good choice for a segmentation algorithm can be addressed straightforwardly: segmentation algorithms, when applied to a specific task, seldom appear isolated, normally they will be embedded in an application that was specially developed for the application in some field where it will be used to segment an open set of images that are very similar to each other. This implies that a tool for an application area has to be specifically developed; generic image processing tools for generic usage are applied only in fields where there are no available specific applications. If a tool has to be developed for a specific application domain, then to provide this tool with a specific color distance metric specially tailored for this application domain is not a step that turns the development of this tool unfeasible. On the contrary: it is just another parameter. From this point of view, since our approach postulates that there is the need to develop domain-specific color metrics and not general metrics, the generation of a specific color metric for an application domain will be just another parameter generation step that will be part of the development cycle of an application, improving so the final results of the segmentation procedure.

The achieved results show that a specific adaptive color space featured by a set of initial seeds is a feasible solution when applied as a similarity measure in the Mumford-Shah model. These results give us reasons to believe that a metric generation technique like the Mahalanobis distance and its variations, such as the Bhattacharya \[17\] distance could be applied not only to the Mumford-Shah algorithm, but to most segmentation techniques to improve their results.

On the other side, the manually selected seeds used in this work provide an interesting research field where there is a need to develop automation procedures, allowing better generalization of this approach. Some interesting approaches about the use of auto-adaptable seeds are described in \[18\], and eventually could also find application here.

5. Acknowledgments

Daniel D. Abdala thanks the CNPq-Brazil-Brasilia for granting him a Ph.D scholarship under the process number 290101/2006-9.

6. References


