

Can the Use of a nonlinear Color Metric systematically improve Segmentation?

Pode o uso de uma métrica de cor não-linear melhorar sistematicamente a segmentação?

Carvalho, L.E.^{1 5*}, Mantelli Neto, S. L^{2 5}, Sobieranski, A.C^{3 5}, Comunello, E., A.C^{4 5}, von Wangenheim, A^{1 5}

Abstract: Image segmentation is a procedure where an image is split into its constituent parts, according to some criterion. In the literature, there are different well-known approaches for segmentation, such as clustering, thresholding, graph theory and region growing. Such approaches, additionally, can be combined with color distance metrics, playing an important role for color similarity computation. Aiming to investigate general approaches able to enhance the performance of segmentation methods, this work presents an empirical study of the effect of a nonlinear color metric on segmentation procedures. For this purpose, three algorithms were chosen: Mumford-Shah, Color Structure Code and Felzenszwalb and Huttenlocher Segmentation. The color similarity metric employed by these algorithms (L_2 -norm) was replaced by the Polynomial Mahalanobis Distance. This metric is an extension of the statistical Mahalanobis Distance used to measure the distance between coordinates and distribution centers. An evaluation based upon automated comparison of segmentation results against ground truths from the Berkeley Dataset was performed. All three segmentation approaches were compared to their traditional implementations, against each other and also to a large set of other segmentation methods. The statistical analysis performed has indicated a systematic improvement of segmentation results for all three segmentation approaches when the nonlinear metric was employed.

Keywords: Image Segmentation — Nonlinear Color Metrics — Polynomial Mahalanobis Distance — Split and Merge — Variational Methods — Graph-Based Segmentation

Resumo: Segmentação de imagens é um procedimento onde a imagem é dividida em suas partes constituintes, de acordo com algum critério. Na literatura, existem diferentes abordagens bem conhecidas para segmentação, tais como clusterização, limiarização, com base em teoria de grafos e crescimento de regiões. Tais abordagens, adicionalmente, podem ser combinadas com métricas de distância de cor, tendo um importante papel na computação da similaridade das cores. Visando investigar abordagens que possam melhorar a performance dos métodos de segmentação, este trabalho apresenta um estudo empírico do efeito de uma métrica de cor não-linear nos procedimentos de segmentação. Para este propósito, três algoritmos foram escolhidos: MumfordShah, Color Structure Code e Felzenszwalb and Huttenlocher Segmentation. A métrica de similaridade de cor empregada por estes algoritmos (L_2 -norm) foi substituída pela Distância polinomial de Mahalanobis. Esta métrica é uma extensão da distância estatística de Mahalanobis utilizada para medir a distância entre centros de coordenadas e distribuições. Uma avaliação com base na comparação automática dos resultados de segmentação contra seus respectivos padrões ouro, disponíveis na base de dados de Bekeley, foi realizada. Todas as três abordagens de segmentação foram comparadas com suas implementações tradicionais, umas contra as outras e também contra um grande conjunto de outros métodos de segmentação. A análise estatística realizada indicou uma melhora sistemática nos resultados de segmentação para as três abordagens quando a métrica não-linear foi empregada.

Palavras-Chave: Segmentação — Métricas de cor não-linear — Distância polinomial de Mahalanobis, Divisão e fusão — Métodos variacionais — Segmentação baseada em grafo

¹ Graduate Program in Computer Science - Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil

² Brazilian Institute for Space Research - INPE, São José dos Campos, São Paulo, Brazil

³ Department of Computing - Federal University of Santa Catarina, Ararangua, Brazil

⁴ University of Itajaí Valley - Univali, Itajaí, Santa Catarina, Brazil

⁵ National Brazilian Institute for Digital Convergence - Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil

*Corresponding author: lcarvalho@incod.ufsc.br

DOI: <https://doi.org/10.22456/2175-2745.79885> • Received: 23/01/2018 • Accepted: 04/08/2018

CC BY-NC-ND 4.0 - This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

1. Introduction

The segmentation of complex color images in a fast and reliable manner still poses a major challenge to the field of image processing and understanding [1]. There are several image segmentation approaches, each one with its own advantages and disadvantages. The principal challenge in the segmentation area is to create algorithms that circumvent the problem of over and under segmentation, by dividing a single object in too many regions, by leaking region boundaries on another object or even fusing various objects into a single one. To obtain results that overcome these problems, several approaches have been proposed in the literature, improving or combining the existing ones. One example is the substitution of the color similarity metric to achieve a better segmentation of patterns. These algorithms usually employ linear discriminating functions ignoring the nonlinear features of some cases on generic color spaces such as the RGB. It has been demonstrated however that using a customized similarity functions [2] based on an oriented nonlinear metric [3], has the potential to considerably enhance the quality and the robustness of different kinds of segmentation approaches [3, 4, 5, 6].

But can this customized feature be considered a widely employable approach as a segmentation algorithm, or is it only used for a specific niche problems? In this context, the objective of this study is to investigate whether the substitution of a linear color metric by a oriented nonlinear one, provides a significant change in the quality of segmentation results, regardless of the approach employed. In order to investigate the impact of this change, we select three different approaches: split and merge, variational region growing and graph-based segmentation, that can be adapted to an oriented nonlinear metric. To validate the proposed method results, in the next sections we outline the general methodologies employed. Afterward, it was established an statistical experiment to investigate if there is a significant improvement from the nonlinear approach against their original counterpart and from other additional approaches, evaluating their overall performance.

2. Material and Methods

In this section, a brief explanation of the methodology employed and the selection of segmentation approaches used is presented. Then, the changes made to adapt their nonlinear version and the criteria used to validate the images are described. In the Figure 1 a diagram summarizing the overall methodology used is shown.

From the analysis of the segmentation approaches, it was chosen a set of specific implementation which we understood, is widely accepted and represents properly each particular segmentation philosophy: Color Structure Code (CSC) [7, 8, 9], Mumford-Shah (MS) [10, 11] and Felzenszwalb and Huttenlocher (FH) [12] representing respectively split and merge, variational region growing and Graph-Based Segmentation. As nonlinear color similarity metric we employed the Poly-

nomial Mahalanobis Distance (PMD) [13] in the RGB space [14]. This color space was used to demonstrate the robustness of the nonlinear metric, once this space is regularly distributed over all dimensions, and color is represented as a dependency of these dimensions. In other words, RGB color space is not suitable to represent colors properly such as a smooth gradient variation when L_2 -norm is used to categorize color vectors. For each of these implementations (CSC, MS, FH) we present modified versions (WCSC, SMS, WFH) that employ the PMD as its similarity metric [5, 4, 6].

2.1 NonLinear Color Metric: the Polynomial Mahalanobis Distance

The Mahalanobis Distance (MD) is a statistical distance used to calculate the similarity between multi-variated distributions [13]. The MD is based upon the mean and co-variance matrix of the distribution, used as the reference and is formally defined by equation 1.

$$MD(x,y) = \sqrt{(x-y)^T A^{-1} (x-y)}, \quad (1)$$

where $MD(x,y)$ is the Mahalanobis Distance between two color vectors $x = (r_x, g_x, b_x)$ and $y = (r_y, g_y, b_y)$ on RGB color space and A is the co-variance matrix of the distribution used as reference, having A^{-1} as its inversion operation.

When the distribution used as reference presents variation on its dimensions (i.e. the co-variance between components and its relations RR, RG, RB), the Polynomial Mahalanobis Distance (PMD) can be used to allow a better discrimination of complex selected image patterns on the color space. The PMD is the polynomial combination of the input vectors and their successive projections to a higher order polynomial terms. While the MD is used for linear case, the PMD allows a nonlinear discrimination of data. For this purpose we used the framework presented in [14] that optimizes the mapping of polynomial terms eliminating non-representative terms in the projection of higher orders (i.e., Principal Component Analysis of variables). The PMD can be computed through equation 2:

$$PMD(x,y) = MD_{\sigma^2}(x,y) + \sum_{l=1}^{L-1} MD_{\sigma^2}(q_l^i, q_l^j). \quad (2)$$

The first order-term $MD_{\sigma^2}(x,y)$ is the MD with a lower positive value σ^2 used to avoid limitations of matrix inversion. This is necessary especially when $d \ll N$, since we use the Singular Value Decomposition (SVD) to obtain a system of Eigenvalues and vectors, and by adding a small positive σ^2 we make it invertible again. With $L = 1$ we calculate the original MD. The following $L > 1$ terms results on q-order ($q = 2^{L-1}$). For example, with $L = 2$ we obtain a q-order of 2, and so on. The terms composed by q_l^i and q_l^j are the next projections ($q + 1$) of x and y into their polynomial terms. If more precision is needed to discriminate vector and eliminate

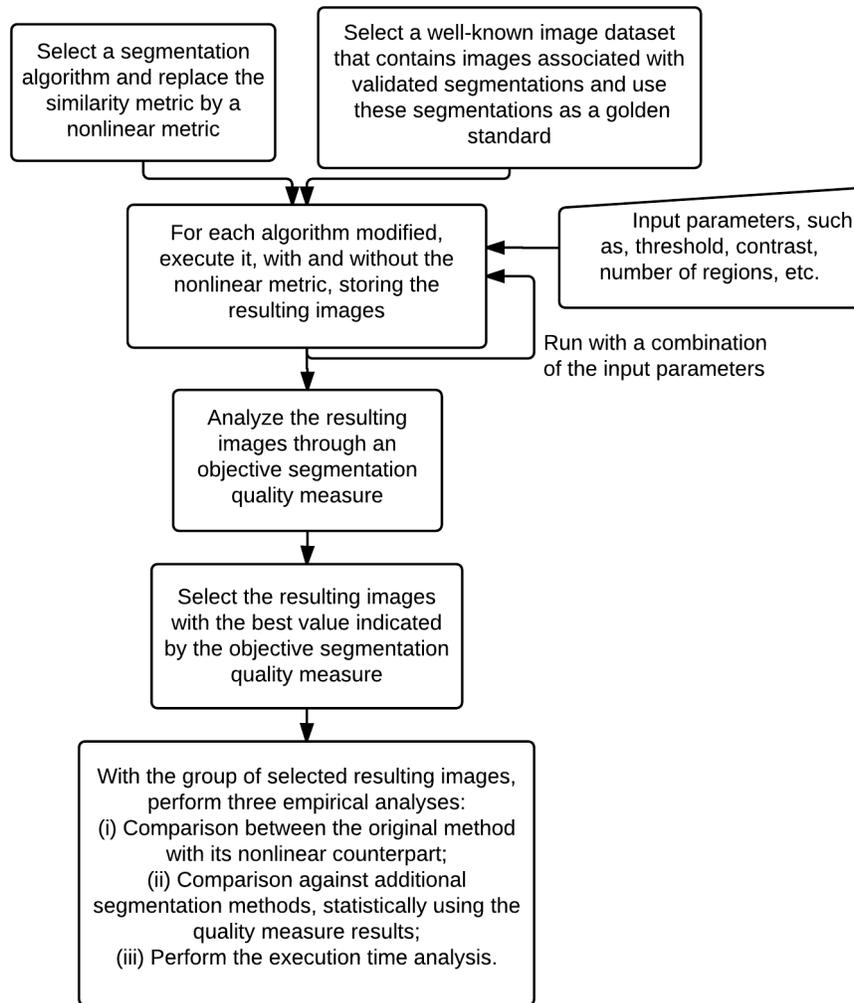


Figure 1. Flowchart of the methodology used.

outlier points, the order of projection of the polynomial can be expanded, increasing the value of L , restricting the influence area of the distribution used as reference. Interestingly, if the co-variance matrix is an identity matrix, the metric is reduced to the L_2 -norm, presenting itself as a dual alternative to a Weighted Nonlinear Euclidean Distance.

2.2 Split and Merge: Weighted Color Structure Code

The Color Structure Code (CSC) [7, 8] is a split-and-merge segmentation method that builds a hierarchical structure of hexagonal pixel islands, structured in levels, in order to perform the segmentation process. The hierarchical structure is an efficient form of keeping the partial results, allowing an easy navigation among distinct levels. The hierarchical structure also allows the algorithm to keep a record of the local and global characteristics of the image. The algorithm can be described in four steps: pre-processing, initialization, linking and splitting. The CSC uses a local variable criterion based upon the linear Euclidean Distance (ED) for the calculation of the similarity during linking.

The Weighted Color Structure Code (WCSC) [5] modifies the similarity measure employed in CSC through the insertion of a nonlinear metric into the Weighted Euclidean Distance (WED) calculation. From a practical point-of-view, PMD provides nonlinear topology which are used to weight the WED in such manner that the regular L_2 -norm values are combined into this topology. Larger the polynomial order used to construct the maps are, more restrictive the PMD is to the WED. An example of how the topology acts over a discrete domain and how it interacts with a set of training points is presented in [3].

Important features of WCSC are the threshold (th) and contrast (c) parameters. The threshold determines the similarity boundary for a region to be merged or not, and can lie in a range from 0 to $+\infty$. The contrast parameter c , used only on WCSC, determines how delimited the pattern will be and can vary in a range from 0 to $-\infty$.

The whole process executed by the nonlinear similarity function is illustrated in Figure 2. The input parameters for

WCSC are: the selected pattern P , the contrast and threshold values. In the first step, the Euclidean Distance (ED) between the regions A and B (B is the region being tested) is calculated. If the distance is lower than a desired threshold (th), the two regions are linked. Otherwise region B will be re-analyzed checking its WED to a selected pattern P by the PMD. The value is checked to see if it is lower than th , in order to re-check if the regions should be linked or not. Then the segmentation process continues.

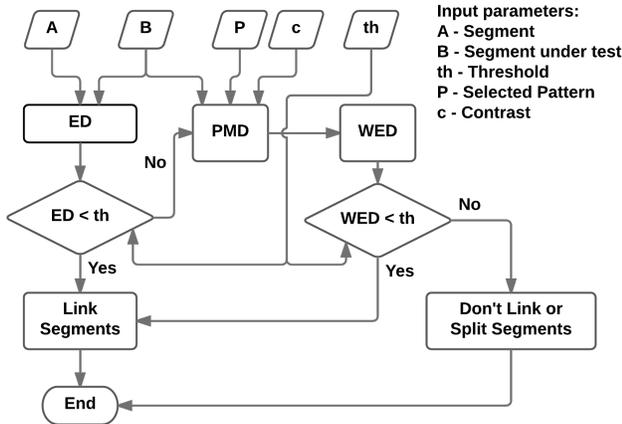


Figure 2. Flowchart of the WCSC method, illustrating the procedure used.

2.3 Variational Methods: Mumford Shah Energy Functional

The Mumford-Shah (MS) [10, 11] is a variational model based on the minimization of a functional energy E . This function employs partial differential equations to model properties of a phenomenon (A_1, \dots, A_n) and to find functions $u_1(x), \dots, u_n(x)$ that can effectively minimize the energy, generated by the sum of each particular property A_n . Each property is also called a term of the energy function, and the minimization of E is a compromise among each term. In our experiments we used a simplified version of this energy function [2] also described in the equation below 3:

$$E(u, K) = \beta \int_{\Omega} (u - I)^2 dx + \lambda \ell(K) \quad (3)$$

where:

- $\beta = \frac{|\Omega_i| \cdot |\Omega_j|}{|\Omega_i| + |\Omega_j|}$, is a coefficient of area used to provide regularity to the model when the region merging is being performed between adjacent regions $|\Omega_i|$ and $|\Omega_j|$;
- $(u - I)$ is a fidelity term which estimates the quality of the piece-wise approximation of a function $u(x, y)$ to $I(x, y)$;
- λ - is the scale space parameter for segment granularity, used to control sub and over segmentation;

- $\ell(K) = \int_K d\ell$ is the length of the frontier between $|\Omega_i|$ and $|\Omega_j|$ in the Hausdorff sense.

The simplified minimization energy $E(u, K)$ is given only as a function of common boundaries for adjacent regions. The coefficient β weights the approximation of regions by their mean values u and I . Its typical exponential growth makes large regions absorbing small ones, according to a pre-determined similarity criteria among the color vectors. The coefficient λ controls and limits the total number of regions. Starting with each pixel in the image being a single region, the algorithm successively merges any two adjacent regions Ω_i and Ω_j , which lead to the largest energy decrease. The constant approximation, regions ranking, fusion of two smallest energy regions Ω_i and Ω_j , the recalculation of energy and reordering of stack of regions, continues until the number of regions informed as parameter is reached. Mumford-Shah uses fusion as discrimination criteria for recalculating and re-ranking the regions stack.

Supervised Mumford-Shah (SMS) is a variant of the original Mumford-Shah equation, where a supervised discrimination function is used to replace the approximation between adjacent regions [2, 4]. This variant modifies the fusion and consequently the energy and reordering calculation of regions using the nonlinear metric, described as the following inequality 4:

$$\lambda \leq \beta \cdot \frac{\phi}{\ell(K)} \cdot PMD(u_i, u_j) \quad (4)$$

where:

- ϕ is the equivalent boundary between regions Ω_i and Ω_j [4];
- $\ell(K)$ is the boundary length;
- $PMD(u_i, u_j)$ is the new similarity measure between regions Ω_i and Ω_j .

The computational flow executed by the supervised Mumford-Shah variant is illustrated in the Figure 3. The input parameters of the SMS are the regions to be grouped, the selected pattern of the image and the scale parameter, respectively. Initially the execution parameters were determined from the input adjacent regions Ω_i and Ω_j . The distance between adjacent regions Ω_i and Ω_j , which will be grouped, is calculated using the mean value of each region (u_i and u_j), the topological map M (generated from the selected pattern of the image) and a contrast (c) value. Next step is the calculation of the area regularization term, i.e., β also from equation 4. For this purpose, the quantity of pixels bordering between regions Ω_i and Ω_j is used. After that, the last term of the function, ϕ , scalar ≥ 0 , is used to penalize the equation due to an equivalence border method, obtaining the final value of the Eq. 4. Based on this value the algorithm makes the update of the stack of adjacent segments and continues this process of calculation and update until the stop condition is reached.

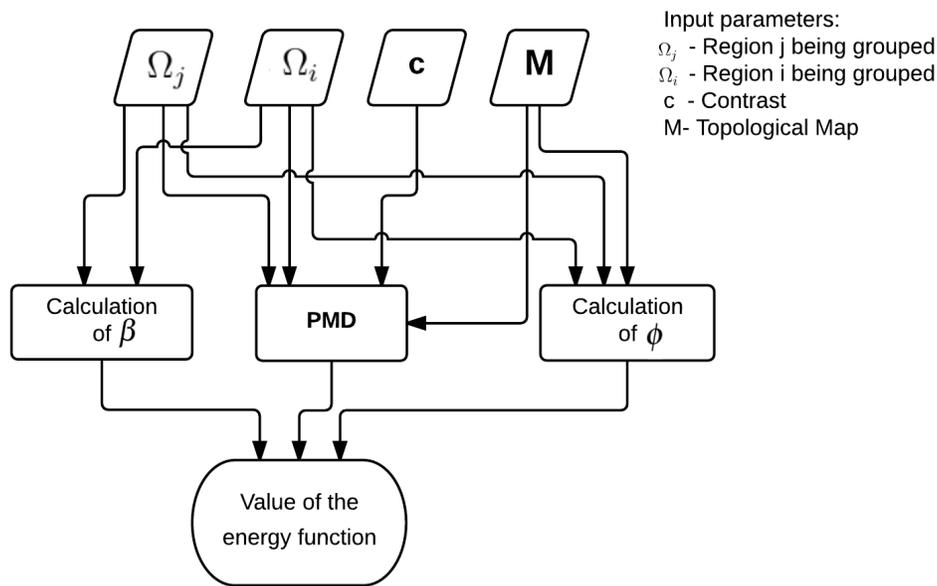


Figure 3. Flowchart of the SMS method, illustrating similarity functions used.

2.4 Graph-Based Segmentation: Felzenszwalb and Huttenlocher Method

The Felzenszwalb and Huttenlocher method (FH) [12] is based on a graph representation of the image as a non-directed graph $G = (V, A)$ where the nodes (or pixels) $v_i \in V$ are connected by edges $(v_i, v_j) \in A$. The segmentation S is the image partition in regions (or components) $C_n \in S$, where each edge has a positive weight $w(v_i, v_j)$ indicating the measure of dissimilarity between two neighboring elements connected by this edge. For determination of dissimilarity, FH employs a predicate that compares the differences between regions based on the ED of color attributes.

The Weighted Felzenszwalb and Huttenlocher method (WFH) [6] modifies the construction and grouping of the connected regions, replacing the linear ED by a nonlinear WED. Similarly as used on WCSC case, WED is also weighted by the Polynomial Mahalanobis Distance, but including the modifications for nonlinear measures. The WCSC tests the threshold th twice: if it fails the ED, the region is granted a second chance of linking on WED. The WFH tests th only once on WED. Another modification was made on the calculation of the weight factor by the PMD. It also considers not only the similarity of node B in relation to the selected pattern, but also the similarity of node A towards the same pattern. This modification is performed in order to strengthen the union of nodes that have a higher degree of similarity with the selected pattern.

Figure 4 illustrates the process performed by WFH. The input parameters for WFH are: regions A and B , selected pattern P and the threshold th . The first step, is the calculation of the PMD between the pattern P and the nodes A and B

separately. The second step is the calculation of WED using the two obtained results and the original nodes, resulting in a new weight of dissimilarity W associated to the connection of nodes A and B . If the value of W is lower than th , then the nodes A and B are grouped in the same connected region and the weight of the region is adjusted. Otherwise, the nodes are kept apart.

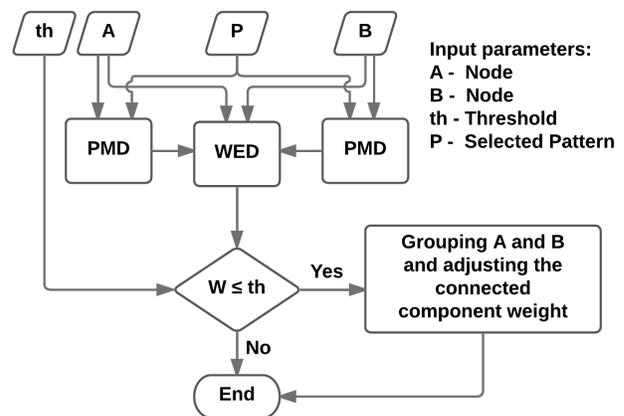


Figure 4. Flow diagram of the WFH method, illustrating the procedure used.

It is important to notice that original FH algorithm employs a pre-processing step before the segmentation. The authors used a smoothing function σ with a threshold th and determined a minimum size of a connected region min . They originally suggested the following values: 0.5, 500 and 20



Figure 5. Effects of two different pattern selection using a white line for WCSC segmentation. Original image on (A) with first pattern selection resulting on (B) regions. A second pattern selection (C) resulting on (D) regions.

respectively for σ , th and min . We had to test these parameters in order to obtain more meaningful results. Using again the Berkeley dataset, we tested a combination of parameters finding new th and min values [6].

2.5 Evaluation Data and Metric

The images used on the present work were obtained from the Berkeley segmentation dataset and benchmark ¹ [15]. This dataset is well known by the image processing community and is used in the present work as a gold standard segmentation reference. Each sample image contains several ground-truths (GT) generated by different human observers and a subset of 60 were selected and used. These images presented a high degree of consensus (or low discordance) on segment objects manually evaluated by observers, as can be noticed in the table 1 appendix A. Due to rationalization of space, a few images were used to illustrate the results, but the full dataset could be observed on the following address: <http://www.incod.ufsc.br/non-linear-color-metrics/>.

In order to evaluate the results, it was used a well-established automated quality metric the Rand Index (RI) [16]. The resulting images were compared against Berkeley's ground truth (GT), keeping the same strategy employed by previous works [3, 8, 5, 6].

3. Experimental Results

This section presents the segmentation results for the proposed approach comparison categorized into three experiments. The first one compares the nonlinear method against their original version. This will be performed by parametric hypothesis test on Rand index, to confirm if the differences between the algorithms are significant. The second experiment compares the Rand index of the methods using the nonlinear metric against commonly used segmentation methods. The last experiment compares all the methods average execution times, aiming to investigate the algorithms in terms of their efficiency. It is important to be outlined that the nonlinear method is a pattern oriented segmentation. That feature is better illustrated on figure 5 where for the same image, the selection of different patterns result on different regions.

¹Images are available at: (<https://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/>), accessed on: 06-19-2018.

3.1 Experiment 1: Comparison between the linear and nonlinear versions

In this section we investigate the performance differences between the linear and nonlinear versions of three segmentation algorithms. The Split and Merge comparison between CSC and WCSC is illustrated side by side in Figure 6. As it can be noticed by first row of the figure 6, fewer regions were generated by WCSC on the star and a better coherence of objects was achieved. This performance is directly related to the pattern selection illustrated on (B). Where the more similar the selection is to the target pattern, greater is its probability of being merged in the same region. In the second row, due to the background selection as pattern, we obtained a better coherence in the foreground elements. In the third row both algorithms identified two regions, but while the CSC misses the shape of the boat, the WCSC keeps it coherent. In the fourth row the selection presented variations in terms of color distribution. In this case WCSC showed a noticeable better segmentation of the mushroom, in one homogeneous region, which highlights the pattern oriented feature added by the PMD inclusion.

The variational region growing comparison between MS and SMS is illustrated side by side on figure 7. As it is presented, MS by itself presents limitations in all column (C) results. The first row presented some regions merged. On the second and third rows some elements were missed and in the fourth row some over segmentation was noticed. These limitations were almost mitigated on the SMS results, column (D). In the first row the star is coherently segmented in one unique segment, eliminating some inappropriate leakage previously presented. In the second row, the same aspects of under segmentation are mitigated and in the third and fourth row, a better coherence between segments are presented, avoiding the missing regions and reducing the over segmentation, respectively.

The Graph-Based Segmentation comparison between FH and WFH is illustrated side by side on figure 8. Similarly as in the previous comparisons, one can notice that FH presented problems in all column (C) results, specially over segmentation. The WFH results in all rows appear with less over segmentation and more region coherence, also highlighting the feature of pattern oriented segmentation.

A parametric hypothesis test was performed for each case to confirm if the differences between the two algorithms, linear and nonlinear, are significant or just caused by external

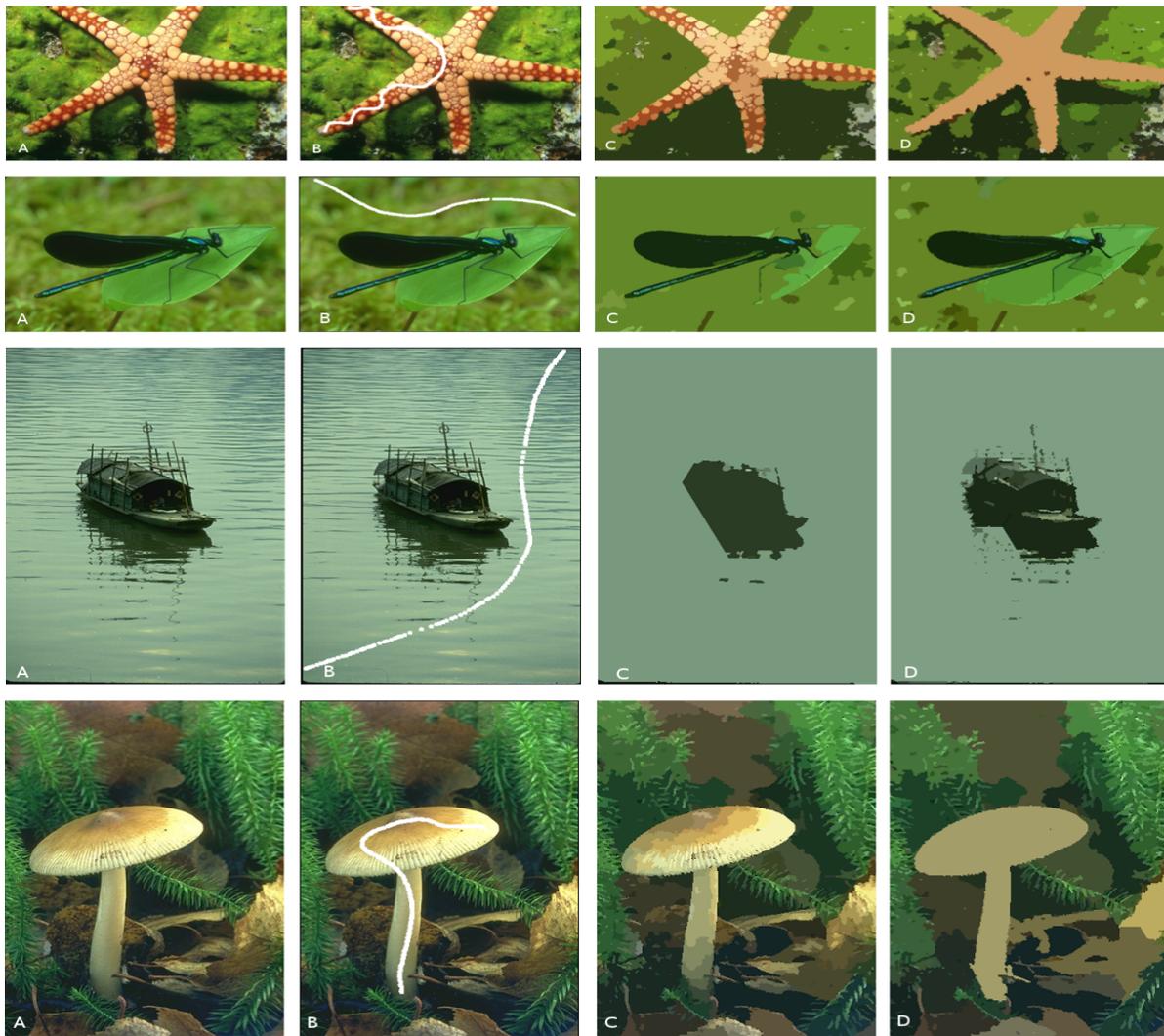


Figure 6. Comparison between CSC and WCSC segmentation. Original image on column (A), a white line showing the pattern selection on (B) and the segmentation results of CSC and WCSC on (C) and (D), respectively.

effects. We used the following assumptions: The Rand Index x paired differences ($x_{linear} - x_{nonlinear}$) between the two methods were considered to have a normal distribution with a sufficient large population of $n = 60$ images. An unilateral hypothesis test with a level of significance of $\alpha = 5\%$ was established, with null hypothesis indicating that algorithms have similar performance and alternative hypothesis indicating the proposed algorithm shows a significant better performance (smaller dissimilarity between segmentation and GT). During the analysis, the rejection of null hypothesis for all investigated cases, indicated with 95% of confidence that the nonlinear versions showed a significant improvement on performance. Table 1 described all parameters determined on the analysis and figure 9 illustrate the box-plot graphs obtained in the first experiment. From this figure it is possible to observe the improvement of the nonlinear version demonstrated by the hypothesis test in the following parameters. Reduction of Rand Index and variance parameters, which indicated a more

stable segmentation behavior for all three approaches tested. In [17], we presented a detailed description of Rand index calculation for each case.

3.2 Experiment 2: Broader comparison against other segmentation methods

In this section we describe the data comparison among the previous linear and nonlinear approaches against additional 8 commonly used segmentation methods using Rand Index, against a GT. The 8 added segmentation algorithms are: Edge Detection and Image Segmentation (EDISON) [19], Mumford-Shah (MS) [10], Watershed (WS) [20], JSEG [21], Recursive Hierarchical Segmentation (RHSEG) [22], Gradient Network Method (GNM and GNM2) combined with MS and CSC [8]. The graph of figure 10 shows values of Rand index for the images in the figure 12 for each algorithm tested. According to the graph, the nonlinear versions are among the smallest Rand Index values, indicating a higher similarity and smaller

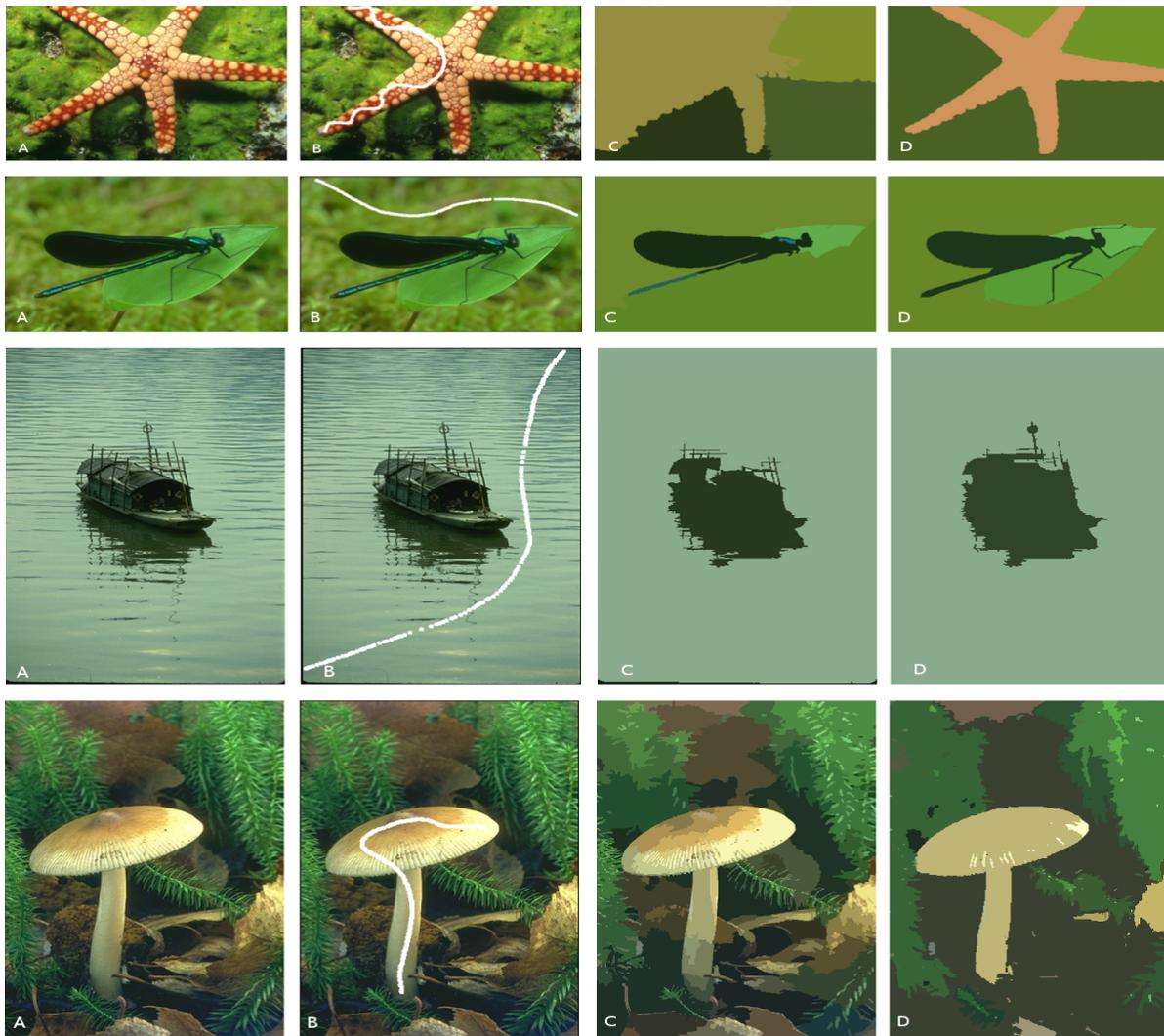


Figure 7. Comparison between MS and SMS segmentation. Original image on column (A), a white line showing the pattern selection on (B) and the results of MS and SMS on (C) and (D), respectively.

data variance.

Figure 12 shows the segmentation of image 368078 for each method used. As can be observed on this figure, the RHSEG segmentation presented regions grouped incorrectly. This segmentation results are in accordance to the Rand index value as shown on figure 10. On the other hand by observing the results obtained by EDISON, a small number of regions were generated and there is a higher coherence between the objects forming the image. Therefore, EDISON presented the lowest Rand index value as shown on figure 10.

Additional tests were performed to extend the analysis to a group of 16 images selected from the set with 60 images previously constructed. Each Rand Index resulted from segmentation comparison against the GT image, was grouped by method and illustrated on figure 11. On this figure, methods with lower Rand Index values, smaller variance and no outliers

are the ones that have a more stable segmentation behavior, which can be noticed in the three methods associated with the nonlinear metric.

3.3 Experiment 3: Comparison of methods based on execution time

In this section we compare the average execution times from each segmentation method with and without the proposed improvement, to evaluate the algorithms in terms of efficiency. For this analysis, the execution times for the methods were recorded running the algorithms on a Core 2 Duo CPU P8700 2.53GHz under Ubuntu Linux for all the 60 selected images used on the first experiment. The individual values obtained can be observed on the web page describing our experiments and the technical report [17]. The mean execution time obtained by the experiment for all the 60 images is illustrated on



Figure 8. Comparison between FH and WFH segmentation. Original image on column (A), a white line showing the pattern selection on (B) and the results of FH and WFH on (C) and (D), respectively.

figure 13.

Analyzing the mean execution times, one can notice that methods employing Polynomial Mahalanobis Distance (PMD) have higher execution times. This is justified by the introduction of the extra calculations of the PMD to the similarity calculation (training and evaluation). However, despite this increase on the execution time, we can observe a gain in terms of effectiveness as shown by the previous experiments. Table 2 gives a summary about the comparison between execution time and effectiveness. In this table we can highlight that although the execution time of all nonlinear methods has increased the results were improved, demonstrated by the smaller mean and standard deviation values when compared against their linear counterparts.

4. Discussion

Every study necessarily has limitations and imposes trade-offs. We discuss here if our method could represent a threat to a more general validity of the conclusions reached in this work. The first identified threat was the lower number of algorithms chosen to be tested within our approach. With the goal of verify the influence of a nonlinear metric as a central device for segmentation algorithms, looking for a general overall improvement of the segmentation results, we selected three algorithms (CSC, MS and FH). Those algorithms were selected because this central device is in some manner dissociated from their main concept, and a distance metric can be easily included to drive these algorithms. On the other hand, there are algorithms where this "device" is not completely clear and other dependencies can be found, making it hard or imprecise the addition of a nonlinear metric. Therefore, we adopted a strategy to use stable segmentation algorithms, where the

Table 1. Parameters obtained during hypothesis test experiment. Where s_{linear} and $s_{nonlinear}$ are the standard deviation for the algorithms with linear a nonlinear approaches, respectively, and $z_c = 1.65$ is the critical value obtained from statistical table from [18] for an accumulated value = 0.950529.

Parameter of differences	CSC/WCSC	MS/SMS	FH/WFH
Mean $x_{diff} = x_{linear} - x_{nonlinear}$	0.0285	0.1099	0.02930
Standard deviation $s_{diff} = \sqrt{\frac{s_{linear}^2}{n} + \frac{s_{nonlinear}^2}{n}}$	0.0149	0.0186	0.01749
Normalized z-value $z_{diff} = \frac{x_{diff}}{s_{diff}}$	1.9127	5.9086	1.6752
Critical z-value $z_c =$	1.65	1.65	1.65

Table 2. Table showing the methods execution time and their mean rand index value for all the 60 images.

Method	Mean execution time (ms)	Mean Rand value	Rand value standard deviation
CSC	112.287	0.1514	0.0893
WCSC	4616.29	0.1229	0.0742
MS	28970.40	0.2129	0.1228
SMS	169411.39	0.1030	0.0756
FH	123.12	0.1464	0.1082
WFH	15599.90	0.1171	0.0815

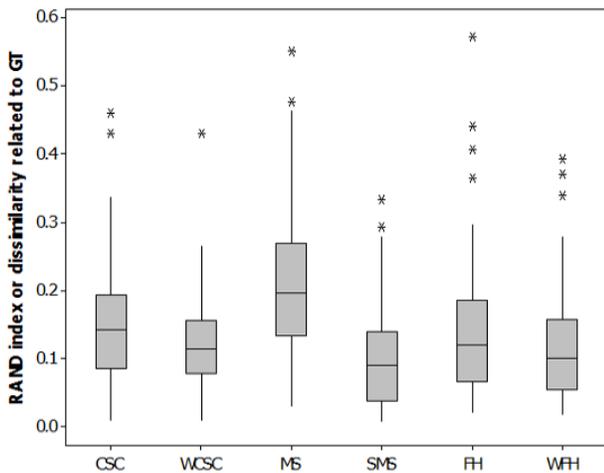


Figure 9. Box-plot graphs comparing statistic analysis of all data for CSC, WCSC, MS, SMS, FH and WFH for all the 60 selected images.

addition of a nonlinear metric can be easily performed without affecting other parts of the compared method. For the sake of completeness, we provided, in Experiment 2, a comparison of our results to a second set of other common segmentation algorithms in their original versions, i.e we did not extend these algorithms with nonlinear color spaces.

The second threat was the use of only a subset of Berkeley dataset. Our Berkeley dataset subset selection, containing only "consensus GTs", can appear to be subjective, but our

criterion selection was simple: the 60 images presenting GT's with the least variation between subjects. We did this in order to avoid introducing extra noise due to divergence between the different GTs of the same image. The resulting subset still represents a wide variety of outdoors images and a statistically significant n .

Another possible threat was the employment of only the Rand Index as validation approach. The choice of the Rand Index as the only method for the automated quality analysis of the results was based upon past experiments [4, 8], where other GT analysis methods were also used. During those experiments, the results obtained using Rand were the most similar to human subjective analysis when treating GT-deviant segmentations, compared to other well-known methods such as Fowlkes-Mallows [23], Jacard [24], BGM [25] and Dongen [26]. BGM, for example, tends to overrate under segmentations, giving false high scores to results with region leakage.

Also a possible threat was our choice to not use other generic color spaces and also customized linear color spaces. Our intention with this work was to provide a first analysis in a broader context, not limited to one homogeneous category of images and not limited to one kind of algorithm, of the question whether nonlinear, customized color spaces are generally a better choice than simple linear color spaces. There is a large palette of general-purpose linear color spaces, ranging from simple HSV to elaborate choices such as CIELab, that we did not take into consideration in this study. There is, however, evidence in the literature (e.g. [27]) that, even if some generic color spaces are better for specific categories of images, in general, there are no significant differences when they are compared in a broader context. Linear customized

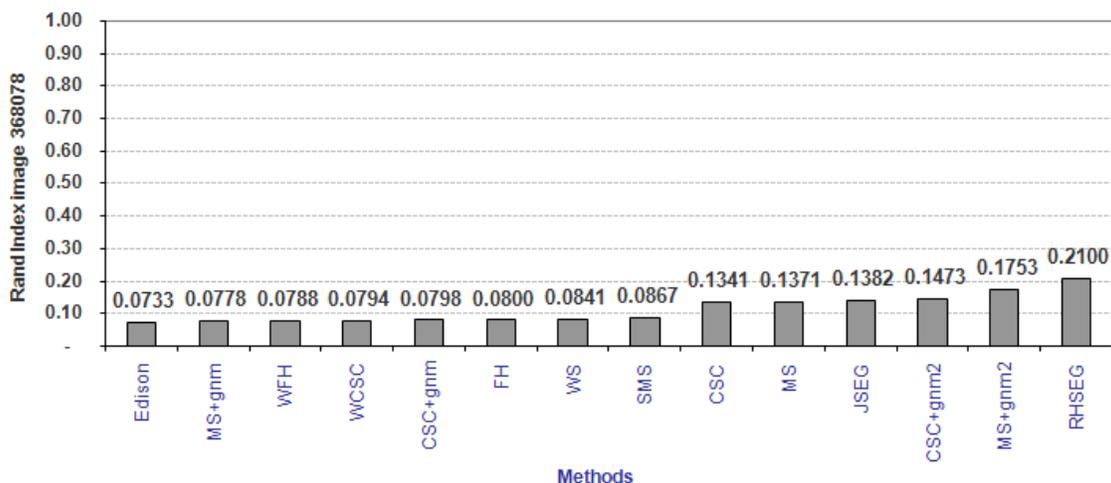


Figure 10. Graph showing the Rand index score for the image 368068 for each method tested.

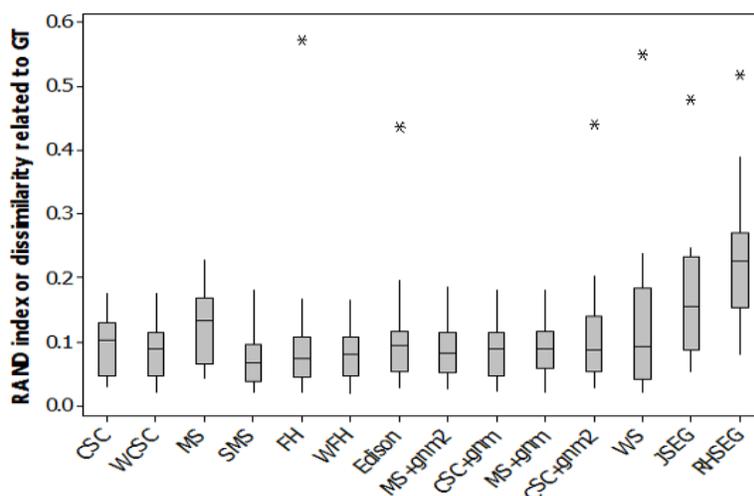


Figure 11. Comparison between all the methods tested based on the Rand index value.

color spaces, on the other side, were left out because evidence gathered in a former study [3] has shown that, at least for the Mahalanobis Distance (MD), to produce a customized linear metric involves the same sample-selecting effort for a given image category than for a nonlinear metric, but the nonlinear version of the customized metric produces much better results.

The last identified threat was on using customized learned color spaces the segmentation algorithms should always perform better because the color space is specific for a particular image and lacks general character. Furthermore, such a color space would characterize this approach as a semi-supervised segmentation, which is not comparable to an unsupervised segmentation employing generic color spaces. This would be true if a given color space could be used for only one image. Within a given context, however, a customized similarity function generated from a pattern obtained from only a few sample images from a category can indeed be generalized for

the whole category of images, as previously shown for the category of "outdoors horse images" [2] and was also shown for the "robot navigation in a forest environment" problem category in [14], especially because it is possible to target the color characteristics of the objects of interest that commonly appear in a particular category of images. This means that a customized similarity function generated, e.g. for "highway traffic scenes" would probably be useless for the segmentation of "pointillist artwork", but will probably work better for car navigation scenarios than a generic linear color space such as RGB. This assumption, however, still needs to be better supported by more experimental results.

5. Conclusion and Future work

The purpose of this study was to analyze whether the use of customized, nonlinear color metric can systematically im-

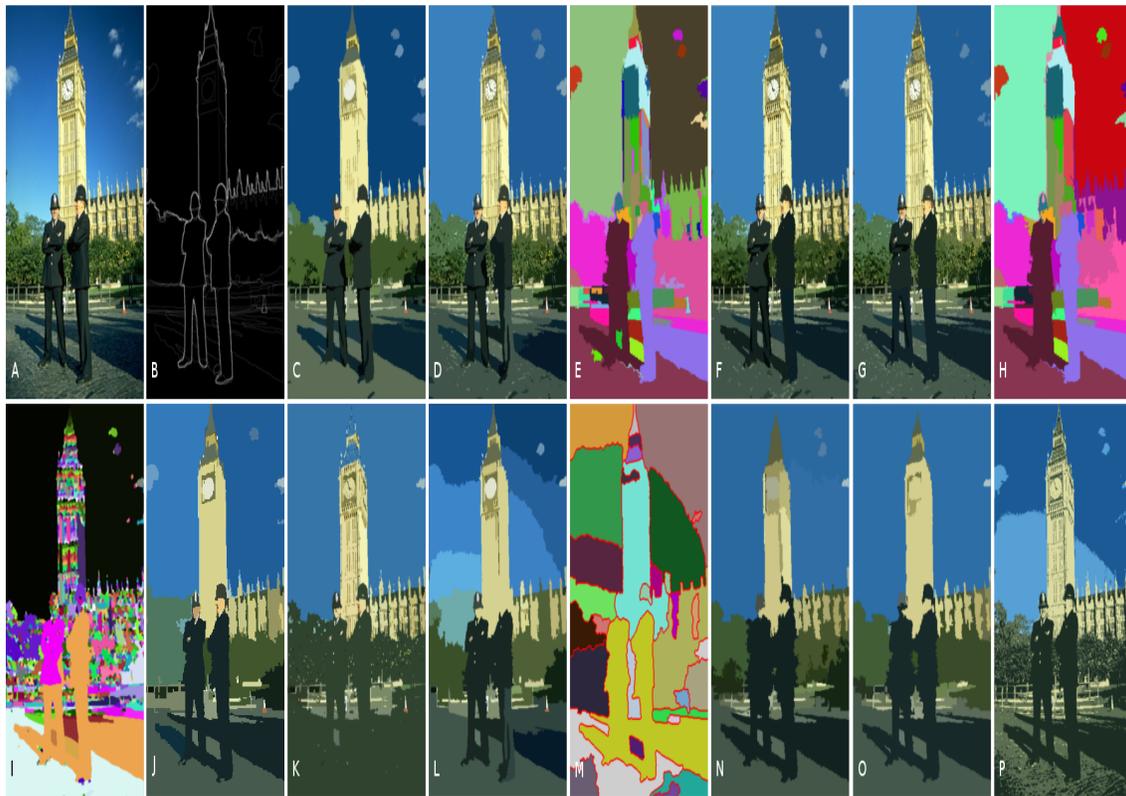


Figure 12. Visual comparison of the obtained results by the tested methods for the image 368078. (A) original image, (B) Ground Truth, (C) Edison, (D) MS+GNM, (E) WFH, (F) WCSC, (G) CSC+GNM, (H) FH, (I) WS, (J) SMS, (K) CSC, (L) MS, (M) JSEG, (N) CSC+GNM2, (O) MS+GNM2, (P) RHSEG.

prove the performance of segmentation algorithms, regardless of which kind of segmentation algorithm is employed. In order to demonstrate the potential of this approach, three experiments were performed on images of the Berkeley dataset, using the Rand index as an objective segmentation quality measure for the automated validation of segmentation results. The obtained results show, for these experiments, a decreasing of the dissimilarity index when a nonlinear discrimination function is used to conduct the segmentation process.

In the first experiment, we statistically demonstrated the improvement potential of our approach with the comparison between the selected algorithms original version against their nonlinear counterpart. In the second experiment, we positioned the methods associated with the nonlinear metric in an experiment environment with some commonly used segmentation algorithms. In this experiment, we showed that the methods with the customized approach produced results with less outliers and variance in comparison against the other approaches. In the last experiment, we compared the segmentation algorithms based on their execution time, where we showed that the nonlinear metric increases the execution time, making evident the trade-off between effectiveness and efficiency.

Even if this study is only a first step in a series of possible empirical evaluation studies employing other segmentation

methods and different nonlinear distance metrics, it has shown strong evidences that, at least for some classes of images, the use of a customized nonlinear color space metric can be considered as a general-purpose applicable improvement.

5.1 Future works

Despite the promising results obtained with analysis and validation of the developed method, there are still improvements to be made in order to achieve more consistent results. Some of these improvements are:

- One disadvantage resulting from the introduction of the Polynomial Mahalanobis Distance (PMD) is the increase of the execution time. To solve this problem an implementation using *Purpose Graphics Processing Unit* (GPGPU) could be performed.
- Aiming at the improvement of the results presented by segmentation methods due to the adding of the PMD as a metric, there are still optimization possibilities of the input parameters, finding the optimal values for threshold and contrast and performing a deeper analysis of the selected input pattern.
- Another field of study is the possibility of adding extra information, e.g. texture, to the similarity metric, which



Figure 13. Graph illustrating the algorithms execution time on a logarithmic scale.

is facilitated by the use of the PMD, given that the PMD can analyze the similarity between distributions in a n-dimensional space, including parameters describing texture attributes [14].

- More experiments can be performed extending the original versions of the algorithms employed in the experiment 2 and employing GT-associated datasets of images of different kinds, not only the outdoors scenes from the Berkeley dataset. Also, different color spaces would be experienced in the further experiments.

6. Author contributions

- Carvalho, L.E. - Researcher that developed two of the main methods evaluated in the manuscript and wrote the article.
- Mantelli Neto, S. L - Researcher that helped in the writing process and performed the statistical analysis.
- Sobieranski, A.C. - Researcher that developed the other main method evaluated in the manuscript and helped in the writing process.
- Comunello, E. - Researcher that performed a more general revision.
- von Wangenheim, A - Advisor that helped in the writing process and also performed a general revision.

References

- [1] UDUPA, J. K. et al. A framework for evaluating image segmentation algorithms. In: *Computerized Medical Imaging and Graphics*. New York, Usa: Pergamon Press, 2006. v. 30.
- [2] SOBIERANSKI, A. C. et al. Learning a color distance metric for region-based image segmentation. *Pattern Recognit. Lett.*, v. 30, n. 16, p. 1496–1506, 2009.
- [3] SOBIERANSKI, A. et al. Learning a nonlinear color distance metric for the identification of skin immunohistochemical staining. In: *Computer-Based Medical Systems, 2009. CBMS 2009. 22nd IEEE International Symposium on*. Albuquerque, USA: IEEE, 2009. v. 22.
- [4] SOBIERANSKI, A. C.; COMUNELLO, E.; WANGENHEIM, A. von. Learning a nonlinear distance metric for supervised region-merging image segmentation. *Comput. Vis. Image Underst.*, v. 115, n. 2, p. 127–139, 2011.
- [5] CARVALHO, L. E. et al. Hybrid color segmentation method using a customized nonlinear similarity function. *Int. J. Image Graph.*, v. 14, n. 01n02, p. 1–13, 2014.
- [6] CARVALHO, L. E. et al. Improving graph-based image segmentation using nonlinear color similarity metrics. *Int. J. Image Graph.*, v. 15, n. 04, p. 361–369, 2015.
- [7] REHRMANN, V.; PRIESE, L. Fast and robust segmentation of natural color scenes. In: PONG, R. C.-C. (Ed.). *Proceedings of the 3rd Asian Conference on Computer Vision*. Hong Kong, China: [s.n.], 1998. (LNCS, v. 1351), p. 598–606.
- [8] WANGENHEIM, A.; BERTOLDI, R.; ABDALA, D. a. Color image segmentation using an enhanced gradient network method. *Pattern Recognit. Lett.*, v. 30, n. 15, p. 1404–1412, 2009.
- [9] PRIESE, L.; STURM, P. *Color Structure Code source code*. Accessed 26 August 2014. Disponível em: http://www.uni-koblenz.de/~{ }lb/lb_download.

- [10] MUMFORD, D.; SHAH, J. Boundary detection by minimizing functionals. In: *IEEE Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE, 1985. v. 2.
- [11] MUMFORD, D.; SHAH, J. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.*, Wiley Subscription Services, Inc., A Wiley Company, v. 42, n. 5, p. 577–685, jul 1989.
- [12] FELZENSZWALB, P. F.; HUTTENLOCHER, D. P. Efficient graph-based image segmentation. *Int. J. Comput. Vision*, v. 59, n. 2, p. 167–181, 2004.
- [13] MAHALANOBIS, P. C. On the generalised distance in statistics. In: *Proceedings National Institute of Science, India*. New Delhi: National Institute of science, 1936. v. 2.
- [14] GRUDIC, G. Outdoor path labeling using polynomial mahalanobis distance. In: *in Proceedings of Robotics: Science and Systems*. Pennsylvania, USA: University of Pennsylvania, 2006. v. 1.
- [15] MARTIN, D. et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *Proc. 8th*. Vancouver, Canada: IEEE, 2001. v. 2.
- [16] RAND, W. M. Objective criteria for the evaluation of clustering methods. *J. Am. Stat. Assoc.*, v. 66, n. 336, p. 846–850, 1971.
- [17] CARVALHO, L. E. et al. *Application of non-linear metric on segmentation algorithms*. [S.l.], 2014.
- [18] MONTGOMERY, D. C.; RUNGER, G. C. *Applied Statistics and Probability for Engineers*. 6. ed. Hoboken, Usa: John Wiley & Sons Inc, 2013. v. 1.
- [19] CHRISTOUDIAS, C.; GEORGESCU, B.; MEER, P. Synergism in low level vision. In: *Pattern Recognition, 2002. Proceedings. 16th International Conference on*. Washington, DC, USA: IEEE, 2002. (ICPR '02, v. 4).
- [20] VINCENT, L.; SOILLE, P. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE Trans. Pattern Anal. Mach. Intell.*, v. 13, n. 6, p. 735–744, 1991.
- [21] DENG, Y.; MANJUNATH, B. Unsupervised segmentation of color-texture regions in images and video. *IEEE Trans. Pattern Anal. Machine Intell.*, v. 23, n. 8, p. 800–810, 2001.
- [22] TILTON, J. D-dimensional formulation and implementation of recursive hierarchical segmentation. *IEEE Trans. Pattern Anal. Machine Intell.*, v. 24, n. 5, p. 603–619, 2002.
- [23] FOWLKES, E. B.; MALLOWS, C. L. A method for comparing two hierarchical clusterings. *J. Am. Stat. Assoc.*, v. 78, n. 383, p. 553–569, 1983.
- [24] BEN-HUR, A.; ELISSEEFF, A.; GUYON, I. A stability based method for discovering structure in clustered data. *Pac Symp Biocomput.*, v. 1, n. 1, p. 6–17, 2002.
- [25] JIANG, X.; MARTI, C.; IRNIGER, C. a. Distance measures for image segmentation evaluation. *EURASIP J. Appl. Signal Process.*, v. 2006, n. 1, p. 209–209, 2006.
- [26] DONGEN, S. *Performance Criteria for Graph Clustering and Markov Cluster Experiments*. Amsterdam, The Netherlands, 2000.
- [27] WANG, X. et al. Comparison of different color spaces for image segmentation using graph-cut. In: BRAZ, J. (Ed.). *VISAPP 2014 - International Conference on Computer Vision - Theory and Applications*. Lisbon, Portugal: IEEE, 2014. v. 3.

Appendix

Table 1. Table showing Berkeley image number, the number of regions assigned for each observer (Obs), the Mean and the Standard Deviation (STDEV) determined for all the 60 images used in the present work.

Image	Obs1	Obs2	Obs3	Obs4	Obs5	Obs6	Obs7	Obs8	Mean	STDEV
2092	17	10	19	9	24	7	21		15.2857	6.6008
3096	3	11	6	6	6				6.4	2.8809
12003	6	6	98	6	7				24.6	41.0341
15088	20	15	23	12	17				17.4	4.2778
16052	8	6	4	2	5	10			5.8333	2.8577
22090	86	19	22	36	33	14			35	26.3514
24004	4	10	5	40	36				19	17.5499
24063	16	6	40	15	51				25.6	18.9815
35070	11	12	9	9	10				10.2	1.3038
46076	11	17	80	14	7				25.8	30.5237
48055	16	22	12	5	28	11			15.6667	8.2623
60079	3	3	31	5	4				9.2	12.2147
61060	12	10	6	6	20				10.8	5.7619
62096	17	17	28	13	11	13	9	30	17.25	7.7597
66075	4	8	17	7	7				8.6	4.9295
67079	6	9	39	11	51				23.2	20.4254
68077	14	10	12	14	24	32			17.6667	8.5244
69015	35	13	45	47	39	29	14		31.7142	13.8168
80099	2	2	2	4	6				3.2	1.7888
94079	8	6	2	6	11				6.6	3.2863
97017	15	12	26	11	18	34			19.3333	8.9814
97033	13	32	22	24	25				23.2	6.8337
100075	18	5	13	16	30	16			16.3333	8.1158
100080	7	35	4	27	12				17	13.3977
113009	13	61	41	26	14				31	20.2361
113016	3	31	7	11	24	13			14.8333	10.6285
113044	10	15	5	9	17	6			10.3333	4.8027
118035	14	19	47	12	21				22.6	14.1173
124084	5	49	9	11	46				24	21.5870
126007	21	31	32	17	24	7	18		21.4285	8.6575
135069	6	7	4	4	7				5.6	1.5165
143090	8	11	12	19	17				13.4	4.5055
151087	62	27	43	35	55				44.4	14.2758
159091	10	15	12	11	6	9			10.5	3.0166
161062	19	13	18	13	27				18	5.7445
163014	18	16	16	27	21	17			19.1667	4.2622
167062	5	28	6	8	5	5			9.5	9.1378
167083	5	7	8	7	6				6.6	1.1401
183055	7	52	7	7	18				18.2	19.4858
196073	8	7	2	2	12				6.2	4.2661
207056	4	4	5	5	10				5.6	2.5099
208001	37	19	84	57	25				44.4	26.4726
216053	15	20	36	61	25				31.4	18.2838
225017	12	10	9	11	9				10.2	1.3038
227092	8	62	10	13	10				20.6	23.2120
238011	12	12	5	15	14	12			11.6667	3.5023
247085	6	32	7	53	40	20			26.3333	18.7474

Continued on next page

Table 1 – continued from previous page

Image	Obs1	Obs2	Obs3	Obs4	Obs5	Obs6	Obs7	Obs8	Mean	STDEV
249061	8	14	16	10	25				14.6	6.6181
253036	31	37	31	50	37	37			37.1667	6.9402
260058	19	21	7	10	9				13.2	6.3403
291000	11	9	4	7	10	4			7.5	3.0166
295087	7	12	9	33	10	18	15		14.8571	8.8209
299091	4	5	3	11	11				6.8	3.8987
300091	7	8	11	6	20				10.4	5.6833
304034	3	7	6	10	12				7.6	3.5071
310007	14	13	8	9	9				10.6	2.7018
368078	12	7	43	14	35	31	34		25.1428	13.8735
374067	22	12	11	17	10				14.4	5.0299
388016	61	6	19	35	54	110			47.5	36.9364