Color Image Segmentation using an Enhanced Gradient Network Method

A. v. Wangenheim¹, R.F. Bertoldi¹, D. D. Abdala², A. Sobieranski¹, L. Coser¹, X. Jiang², M. M. Richter³, L. Priese⁴, F. Schmitt⁴

¹Image Proc. and Computer Graphics Lab – LAPIX, Federal University of Santa Catarina, Florianópolis, Brazil
{awangen, fogo, asobieranski, coser}@inf.ufsc.br

²CS Department, University of Münster, Münster, Germany
{abdalad, xjiang}@math.uni-muenster.de

³CS Department, University of Calgary, Alberta, Canada
mrichter@cpsc.ucalgary.ca

⁴CS Department, University of Koblenz, Koblenz, Germany
{priese, fschmitt@uni-koblenz.de

Abstract

The objective of this paper is to evaluate a new combined approach intended for reliable color image segmentation, in particular images presenting color structures with strong but continuous color or luminosity changes, such as commonly found in outdoors scenes. The approach combines an enhanced version of the Gradient Network Method, GNM2, with common region-growing approaches used as pre-segmentation steps. The GNM2 is an post-segmentation procedure based on graph analysis of global color and luminosity gradients in conjunction with a segmentation algorithm to produce a reliable segmentation result. The approach was automatically evaluated using a close/open world approach. Two different region-growing segmentation methods, CSC and Mumford & Shah with and without the GNM post-processing were compared against ground truth images using segmentation evaluation indices Rand and Bipartite Graph Matching. These results were also confronted with other well established segmentation methods (RHSEG, Watershed, EDISON, JSEG and Blobworld).

Keywords: color image segmentation, region-growing, outdoors scenes, Gradient Network Method

1 Introduction

Natural color scenes, such as outdoors images composed by many colored objects that are acquired under uncontrolled conditions can show complex illumination patterns across one same object in the picture. Examples are variations in lightness and specular effects. State-of-the-art region-growing segmentation methods present two main features that limit their applicability for dealing efficiently with natural scenes:

(a) A static region similarity concept, where pixels or textures within a region are expected to be homogeneous. Typical natural scenes, however, show strong continuous variations of color, presenting a different, dynamic order that is not taken into account by such algorithms. They divide a sky region with different intensities of blue into segments or will represent an irregularly illuminated surface as a set of different regions. When the parameters of such algorithms are stressed in order to try to accomplish a correct segmentation of a large object showing a long continuous gradient of color, typically with a gradual but large color variation, a region leakage into other objects in the image is likely to occur. Then the algorithm will become unstable and even inapplicable;
(b) Increase in complexity to present more stable results, which usually demands complex computations to detect segment-correlation clues, or are built upon additional texture information. This slows down considerably the processing time without being much more stable when extreme color variations are present.

A number of segmentation approaches already tried to cope with one or both problems: Rehrmann and Priese (1998) propose a hierarchical model to improve the segmentation of color scenes, while Deng and Manjunath (2001), Dupuis and Vasseur (2006), Kato and Pong (2006) address the problem by the additional analysis of object textures. Some authors, as Dony and Wesolkowski (1999) and Schneider et al. (2000), have tried to overcome the problem using different methods of chromatic and luminescence evaluation. Finally, Klinker et al. (1990), Tsang and Tsang (1996), Healey (1992) proposed illumination models for the formulation of segmentation algorithms.

We addressed this problem before, developing the Gradient Network Method (GNM) (von Wangenheim et al., 2007). GNM was originally developed intending to solve the problem of luminance and reflection variations by searching for structured gradients along the various neighboring segments, providing a fast and reliable post-segmentation approach for images otherwise difficult to segment (von Wangenheim et al., 2008). The method was also developed to be used as a framework, where different structured gradient analysis strategies could be embedded. The conceptual ideas, the general philosophy underlying the GNM algorithm and a detailed algorithmic description of the GNM are detailed in (von Wangenheim et al. 2007, von Wangenheim et al. 2008).

Our purpose in this paper is to present an enhanced version of this algorithm with some new features, which were included into the method following the intended framework philosophy. The new heuristics try to improve the results while still keeping a similar approach towards the problem of dealing with illumination effects. As its former version, it works together with a segmentation algorithm, which provides a pre-segmentation as a starting point. The general approach is to rely on segmentation algorithms such as CSC (Rehrmann and Priese, 1998) or any other producing as output an over-segmentation with edge preservation and subsequently apply our algorithm, which will iteratively merge segments logically connected through organized gradient patterns improving the final segmentation. Besides, in this paper we also provide objective quantitative segmentation validation results through ground-truth-based segmentation quality measures, which were not present in our former work, where we only discussed our results qualitatively.

2 Objectives

In our former papers, we presented the GNM approach, discussed that it was developed intending it to be used as a framework for different post processing heuristics (von Wangenheim et al., 2007) and offered a set of results that were empirically evaluated from the performance point of view (von Wangenheim et al. 2008). In this paper we pursued two main objectives:

1. Objective segmentation quality validation. For this purpose, we extended the qualitative approach used in (von Wangenheim et al. 2008) and selected two different objective ground-truth-based segmentation quality measures and developed a validation strategy to compare our results against a large set of standard segmentation approaches.
2. **Heuristic-usage suitability testing.** In order to validate the GNM approach as a framework to support different segmentation post-processing heuristics, we devised a specific heuristic that was considered adequate to solve some observed segmentation problems and validated the obtained results against the data produced during pursuit of Objective 1.

3. For the purpose of fulfilling these objectives we devised the following procedure, which included an evaluation using a close/open world approach:

   (a) First, different well-known segmentation methods to be used as pre-segmentation procedures were chosen: CSC and Mumford-Shah (Mumford and Shah, 1989).

   (b) Full segmentations performed with each of these algorithms were compared against ground-truth images using Rand (Rand, 1971) and Bipartite Graph Matching (BGM) (Jiang et al., 2006) indexes. For each segmentation algorithm, we executed it with a wide range of segmentation parameters and then selected the segmentation result considered to be the best one for every pair of image set and segmentation algorithm and generated Rand and BGM scores for the complete set of ground-truths for each image.

   (c) These same segmentation algorithms where also selected to generate over-segmented images to be used in combination with both the original GNM and the version presented in this paper. For each algorithm we selected a set of segmentation parameters that produced oversegmented images where no segment leakage with respect to any ground truth was allowed. Each of these results was used as an input for both GNM algorithms, which were also run with a set of different parameters. The resulting segmentations after post-processing with the GNM and its new GNM2 version also underwent a selection of the segmentation considered to be the best one, when compared against ground-truth images using Rand and BGM indexes.

   (d) We compared these results to three other well-established segmentation methods: EDISON (Comaniciu and Meer, 2002), Watershed (Vincent and Soille, 1991), RHSEG (Tilton, 2006), JSEG (Deng and Manjunath, 2001) and Blobworld (Carson et al., 2002) also using the ground-truth images and the Rand and BGM indexes.

The structure of this paper is as follows: In section 3 we review briefly the GNM main concepts and present in detail the new features added to the algorithm. Section 4 describes the experiment and further discusses the objective evaluation of segmentation results compared against ground-truths. Section 5 reviews the evaluation indexes used, discussing the main characteristics and drawbacks of each of them. Finally, section 6 discusses the results of our close/open world evaluation. All data and results are available in detail under http://www.lapix.ufsc.br/gnm.

3. **Enhancing the Gradient Network Method Segmentation with additional Heuristics: description of the new features**

The Gradient Network Method previously described in von Wangenheim et al. (2007) was developed to deal with segmentation problems where objects in the scene will be represented by several different but similar and gradually varying color shades, as they often are found in outdoors scenes. The GNM looks for a higher degree of organization in the structure of the scene through search and identification of continuous and smooth color gradients. Following
these same goals and principles, a new method based on the GNM algorithm including several new features was developed. Henceforth this new method will be referred as GNM2.

Analog to its first version, for the GNM2 to be able to process an image and identify these characteristics, a pre-segmentation of the image must be performed. The goal is to obtain groups of segments with a high degree of similarity represented in a simple way, avoiding possible problems with local noise induced by high granularity (e.g. at pixel level). The clusters in equation I are sets of pairs of coordinates in the $\Omega$ space of $\mathbb{R}^n$ values and color ranges in the $\mathbb{R}^m$, similar according to the segmentation approach used to realized the rough verification of the image and with dimensionality $m$ determined by the color space selected. Differences between the versions of algorithm start here, as GNM2 uses CIE L*ab as the colorspace of choice, instead of HSL as was the case in the former method.

Every cluster $C = \{c_1, ..., c_n\}$ is a discrete cluster, where each pixel is unequivocally assigned to a specific cluster, such that $\cap^n_{i=1} X_i = \emptyset$ and $\cup^n_{i=1} X_i = \Omega$ for the every $X = \{x_1, ..., x_p \subset C\}$.

The objective of the algorithm remains being to minimize the cardinality of the vertices $|G|$ through a merging procedure that searches sub graphs $S_i \subseteq G$ which present $\Phi < \min\left(\sum^N_k \nabla_k\right)$ where $N$ is the cardinality of the edges. After creating the graph $G(V, E)$, the next step is checking all the neighborhood relations if they comply with the similarity measure and provide continuous and smooth color gradients. The following steps of the algorithm differ from the former version, so they will be detailed below.

### 3.1 Cost function $f_s$

The function $f_s$ takes two vertices $v_1, v_2 \in V$ of $G_i(V, E)$ connected by an edge $e \in E$ as its parameters and outputs a cost related to these two vertices according to perceptive conditions. We consider perception as the capacity of being able to recognize differences of luminosity and light conditions such as reflections when comparing different colors (von Wangenheim et.al.
This function emulates the concept of perception function $f_p$ from the former GNM algorithm while improving its results, in a more intuitive way. Thus, to represent different perception conditions, given two vertices $v_1$ and $v_2$ as input, a vector $\vec{v}_p$ is obtained. The vector $\vec{v}_p = \mu_1, \mu_2$ connects the points $\mu_1, \mu_2 \in \mathbb{R}^3$ corresponding to the average color in CIE $L^*ab$ color space of both $v_1$ and $v_2$ vertices. An angle $\alpha$ between the vector $\vec{v}_p$ and a vector parallel to the L axis of CIE $L^*ab$ color space is calculated. The angle $\alpha$ is used in the following form:

$$\cos^2 \alpha \cdot \sin^2 \alpha \cdot (\mu_a(v_1) - \mu_a(v_2))^2 + (\mu_b(v_1) - \mu_b(v_2))^2,$$

where respectively $\mu_a = \frac{1}{n_p} \sum_{i=1}^{n_p} a_i(v)$ and $\mu_b = \frac{1}{n_p} \sum_{i=1}^{n_p} b_i(v)$, are multiplied by $\cos^2 \alpha$.

Both multipliers are complementary to 1 as $\sin^2 \alpha + \cos^2 \alpha = 1$, in order to reflect the similar nature of the relationship that luminance and chromaticity share. Depending on the relation between two different colors, one characteristic might be more interesting to evaluate the similarity between two colors. For instance, when comparing two color points $k_1$ and $k_2$ in CIE $L^*ab$ color space, such as the vector $\vec{k}_1, \vec{k}_2$ is parallel to the vector $L$. The more important property here would be the luminance of these colors. The purpose of the function $f_s$ is to identify this kind of relationship between different colors. In the old version of our algorithm the color perception was manually tuned.

### 3.2 Additional Heuristics

We also extended the semantics of our cost function heuristics, in order to reflect the fact that two regions represented by vertices $v \in V$ that share long borders probably will be related strongly to each other. We enhanced the function $f_s$, not relying only on color values, as other works like Zlatoff et al. (2008) have done. Instead, we employed properties like symmetry and closure. To achieve this, the sum of the differences is raised by $1 + n_b$, where $n_b$ is a ratio related to the number of pixels shared by the borders of vertices $v_1$ and $v_2$ divided by the smallest border cardinality of these same vertices (see equation IV).

The symbol $\Gamma$ in $f_s$ represents the border set of pixels for each vertex and $|\Gamma|$ their cardinality.

$$n_b = \frac{|\Gamma(v_1) \cap \Gamma(v_2)|}{\min(|\Gamma(v_1)|, |\Gamma(v_2)|)} \quad \text{IV}$$

The purpose is that this property only lessens the costs of merging regions sharing large part of their borders or even a region completely surrounded by another. It is never completely zeroed by this property alone. There is some resemblance to the border component $|v| \Gamma$ from the Mumford and Shah functional presented in Mumford and Shah (1989). The purpose of this component is to keep the borders the smallest possible, while the property used in GNM2 intends to privilege the union between regions sharing a good proportion of their total borders, especially when regions are surrounded by others. Dividing the cardinality of the border shared
by \( v_1 \) and \( v_2 \) by the smallest border cardinality of \( v_1 \) and \( v_2 \) causes the value of \( n_b \) to proportionally increase the more a smaller region is surrounded by a bigger region, reaching the maximum point when \( \min(\Gamma(v_1), \Gamma(v_2)) \leq \max(\Gamma(v_1), \Gamma(v_2)) \). In our approach, a smaller region completely or almost totally surrounded is usually considered part of the same object and the term \( n_b \) was developed to represent this condition in the function \( f_s \) evaluation. One last point to be observed is that the Mumford-Shah property is used in an additive way, reinforcing the point of keeping borders small, and our approach is used as the power to raise the resulting difference between regions. Since this property only lessens the cost of merging embedded regions, but never zeroes the cost, we avoid merging regions by default just because they are totally or partially surrounded by others.

The function \( f \) also uses two more coefficients \( c_1 \) and \( c_2 \), as a way to provide a manual tuning of the importance given to luminance and chromaticity. We give bigger values to the chromaticity coefficient, as we consider that difference in the hue of different colors should weigh more in the cost calculation based on our experiments, usually using a proportion of \( 3 * c_1 = c_2 \) for these coefficients. The function \( f_s \) can be seen in equation V.

\[
f_s(v_1, v_2) = (c_1(\mu_1 - \mu_2)^2 \sin^2 \alpha + c_2((\mu a_1 - \mu a_2)^2 + (\mu b_1 - \mu b_2)^2) \cos^2 \alpha)^{1+n_b}
\]

\( f_s(v_1, v_2) \rightarrow \mathbb{R}^+ \)

3.3 Substitution of the heuristics defined by \( f \)

Various other heuristics could be used as the function \( f_s \). The function presented here was developed to reflect what we expect from our segmentation, but there is no problem whatsoever to have a \( f_s \) with a different approach as long as \( f_s(v_1, v_2) \rightarrow \mathbb{R}^+ \).

Following this requirement, \( f_s \) can be used as a cost function and the GNM2 algorithm can be considered as a framework for other approaches intending the global evaluation of the color variation patterns in an image, no matter how different they are from the heuristic we presented here.

3.4 Vertex Path \( P_v \)

As stated before, in GNM2 the vertices \( v \in V \) are represented in the color space CIE \( L^*ab \). The coordinates of each vertex are determined by \( \frac{1}{n_p} \sum_{i=1}^{n_p} L_i(v) \), \( \frac{1}{n_p} \sum_{i=1}^{n_p} a_i(v) \), \( \frac{1}{n_p} \sum_{i=1}^{n_p} b_i(v) \). They can be linked in a vertex path \( P_v = \{v_1, \ldots, v_m \in V \mid v_{k-1}, v_k \subset e_1 \land v_k, v_{k+1} \subset e_2 \land e_1, e_2 \in E \} \)

This occurs if the transition from \( v_1 \) to \( v_2 \) is the less costly while maintaining similar perceptive conditions (see 3.2) between the regions these vertices represent. These paths represent disjoint partitions of regions, where \( \bigcup_{P_v} P_v = V \) and \( \bigcap_{P_v} P_v = \emptyset \). Initially, all paths \( P_v \) are empty and vertices will be united to them when acceptable conditions are met.

The cost for every candidate path \( P_v \) is tested to see if it fits the expected conditions by the sum of the results of function \( f_s \) from two of the edges connecting three vertices \( \langle v_{n-1}, v_n, v_{n+1} \rangle \), corresponding to a vertex in one of the ends of the current path \( (v_n) \), the former vertex in one of
the ends of the path \((v_{n,i})\) and a vertex that can possibly join the path \((v_{n+1})\). When \(P_i\) are first created, the three vertices will be simply vertices that share a path along the graph \(G(V,E)\).

In the case of \(P_i\) connected by only two edges, the calculation is straightforward. There will be however cases with three vertices connected by three edges, i.e. a cycle. To decide which two edges will be selected, first for each edge \(e \in E\) an angle \(\gamma\) is calculated. Every angle \(\gamma_1, \gamma_2\) and \(\gamma_3\) correspond to each one of the angles between different pairs of these three vectors

\[ \vec{u}_1 = v_{n-1}v_n, \quad \vec{u}_2 = v_nv_{n+1} \quad \text{and} \quad \vec{u}_3 = v_{n+1}v_{n-1}, \quad \text{i.e.} \quad \gamma_1 = \arccos\left(\frac{\vec{u}_1 \cdot \vec{u}_2}{\|\vec{u}_1\| \|\vec{u}_2\|}\right), \]

\[ \gamma_2 = \arccos\left(\frac{\vec{u}_2 \cdot \vec{u}_3}{\|\vec{u}_2\| \|\vec{u}_3\|}\right) \quad \text{and} \quad \gamma_3 = \arccos\left(\frac{\vec{u}_3 \cdot \vec{u}_1}{\|\vec{u}_3\| \|\vec{u}_1\|}\right). \]

The two edges selected to be summed will be those that have the biggest angle \(\gamma\) among the three angles \(\gamma_1, \gamma_2\) and \(\gamma_3\) \((\max(\gamma_1, \gamma_2, \gamma_3))\). This cycle with the three vertices can be seen as a triangle most of the times except when the three vertices are collinear or the two ends of the path are at the same position, then the cycle will be represented by a line. When cycles fall in this exception case, though the rule of choice is the same, it won't matter as all angles will be create equivalent multipliers to be used along the function \(f_i\) (since \(\sin^2 0 = \sin^2 \pi\) and \(\cos^2 0 = \cos^2 \pi\)). Whatever the case, \(\sum_{i=1}^{3} \gamma_i = \pi\).

The decision to use \(\max(\gamma_1, \gamma_2, \gamma_3)\) was based on the fact that \(f_i\) is a cost function but not a distance function. Even though it presents two properties that are expected from a distance function: I: \(f_i(v_1, v_2) \geq 0\) and \(f_i(v_1, v_2) = 0\) only if \(v_1 = v_2\), II: \(f_i(v_1, v_2) = f_i(v_2, v_1)\), it doesn't satisfy the triangle inequality property \(f_i(v_1, v_3) \leq f_i(v_1, v_2) + f_i(v_2, v_3)\). The dependence on the angle \(\alpha\) between \(v_v\) created from the vertices \(v_1\) and \(v_2\) with the axis \(L\) makes the distance between two points not always be the shortest distance along any path. This property reinforces our decision of not relying on the sum of \(f_i\) results alone to guide a decision when cycles are found. The biggest angle strategy also fits well the general idea of following a path, since selecting more obtuse angles make the paths created seem more continuous and less prone to dead ends.

### 3.4 Iterative region growing process

The starting point of each iteration \(i\) is the less costly path \(P_i\) among three vertices \(\{v_{n,i}, v_{n+1}\}\), as long they are similar enough to be acceptable according to \(t_{si}\). If there exists at least one specific \(P_{vi}\) satisfying the conditions required by \(t_{si}\), the vertices \(\{v_{i-1}, ..., v_i\} \cup P_{vi}\) will be considered equivalent, which will result in their merging at the end of this iteration. The next step is searching for acceptable color gradients through the next less costly and acceptable vertex \(v_c \in V\) connected to one of the two the endpoints of the current \(P_{vi}\). Every non-equivalent neighbor vertex \(v_c = v_{n+1}\) to a endpoint of \(P_{vi}\) is a candidate and is tested to see if the result of the sum of the \(f_i\) with the selected edges. If a candidate \(v_c\) proves to be acceptable according to \(t_{si}\) while being the less costly, \(v_c\) will be considered equivalent to all other vertices \(\{v_i, ..., v_{n+1}\} \cup P_{vi}\) and will be merged at the end of this iteration step, i.e. \(P_{vi} \cup v_c\). If \(v_c\) happens to be already in other path, all vertices of both paths will be united, resulting in \(P_{vi} \cup P_{vi}\). These steps are repeated until there are no vertices \(v_c\) acceptable enough to be merged to the \(P_{vi}\) or there are no more vertices to be tested. However, if there are still vertices non-connected to \(P_{vi}\) paths

\[ \sum_{i=1}^{3} \gamma_i = \pi. \]
remaining that are acceptable according to $t_{si}$, the one with the smallest cost is chosen and the search for vertices to connect to the new path $P_{vi}$ starts again.

When it is not anymore possible to keep paths growing, the vertices found to be equivalent and its respective regions are merged. The new vertices resulting from the merging and their new neighborhood relationships will then form the graph $G_{i+1}(V,E)$. The threshold $t_{si}$ is incremented and a new iteration $i = i + 1$ is started. Once no more iterations are possible, the remaining vertices will represent the partitions resulting from the first part of the GNM2 segmentation algorithm.

### 3.5 Additional Steps

The second part of our algorithm consists of 1) avoiding small and undesirable regions and 2) making sure some regions that show good similarity but were rejected by the first part of the algorithm are still merged.

The first step is eliminating the undesired regions in terms of size. We adopt a simple measure here: given a minimum size $m_s$, regions smaller than this threshold will be merged to the largest neighboring region. This decision is taken based solely on the size of the regions because it is considered that regions with a size considered too small might not have reliable color information, being possibly one of the reasons they could not be merged in the first part of the algorithm. Every vertex $v \in V$ has all its neighbor vertices $v_c$ for $\min(|v_c|) < m$. The vertices left after the merges are done and their new neighborhood relationships created an updated graph $G_{i+1}(V,E)$ which will be used by the next step.

The second step assures that regions that show great similarity to others but were rejected by the first part of the algorithm are still merged in one same partition in the final segmentation. This is achieved by defining a desired value $m_c$ for which the remaining regions will be tested. A vertex is merged to another vertex if both show the smallest value resulting from function $f_s$ for all neighbor vertices, while also being smaller than the desired parameter. Every vertex $v \in V$ has all its neighbor vertices $v_c$ for $\min(f_s(v,v_c)) < m_c$. The graph $G_{i+1}(V,E)$ with the regions merged after this second step will then correspond to the image partitioned in perceptively similar regions. The figure 2 depicts the workflow of GNM2.

### 4 Experiment

In order to be able to objectively evaluate we have proposed an objective and quantitative evaluation of GNM2 approach. Then, to choose the optimal segmentation parameters, we accepted that ground-truths (GTs), or hand-made segmentations, representing the judgment of a human observer, should play the role of golden standards. This process of evaluation of image segmentation results was investigated by a number of researchers (Unnikrishnan et al., 2007; Sahasrabundhe, 1999; Di Gesu and Starovoitov, 1999).

We adopted the Berkeley segmentation benchmark image dataset (Martin et al., 2001). This dataset is well known in the image processing community and each sample image contains various ground-truths generated by different subjects. The evaluation schema adopted in this paper works over a subset of 35 images selected from this dataset.

#### 4.1 Evaluation Procedure
In order to objectively evaluate the quality results of the improved GNM2 algorithm we have envisioned an evaluation procedure. For every image in our dataset, we performed hundreds of segmentations with different combinations of parameters combined with pre-segmentations obtained with two algorithms: Color Structure Code (CSC), Mumford & Shah (MS).

The segmentation results of the GNM2 were compared to the results of several other algorithms, including the previous version of the GNM itself. We generated a series of segmentation results for the Color Structure Code (CSC), Mumford&Shah (MS), Watershed (WS), RHSEG, JSEG, Blobworld and EDISON algorithms, together with the GNM combined with the same pre-segmentations as GNM2, CSC and MS.

The Mumford&Shah segmentation we employed is an implementation of the Mumford&Shah functional based segmentation technique found in the Megawave image processing package (Megawave, 2006). Color Structure Code is an algorithm developed for fast and robust segmentation, especially with outdoors images. EDISON (EDISON, 2007) is the implementation of the mean-shift image segmentation described by Comaniciu and Meer (2002). Blobworld is a segmentation algorithm using expectation maximization, especially designed to work in image querying. RHSEG is a hierarchical segmentation technique developed by NASA Goddard Space Flight Center. JSEG is an unsupervised segmentation based in a concept called 'j-image' considered especially robust when applied to scenes where texture predominates.

Our purpose is to compare both GNM and GNM2 results with those obtained by several state-of-art algorithms, looking to validate our algorithm quality when compared to these already well-known algorithms.

For the tests performed with the GNM2 and GNM, all pre-segmentations are obtained with a set of different parameters for each pre-segmentation procedure. The goal is to achieve an over-segmented image avoiding leakages and preserving smooth gradients to suit and allow a better segmentation by our algorithm. Both GNM algorithms are post processing region-growing algorithms, so if there is information loss in the preprocessing step, like regions wrongly merged, it is not possible to recover from this kind of error. For this reason, we generate pre-segmentations with cautious parameters. All algorithms selected provide ways to parametrically control the segmentation results and achieve pre-segmented images according to our algorithm expectations. The parameter used for CSC was threshold = 30 and for Mumford & Shah images were generated with lambda = 600.

On each of the pre-segmented images was performed a series of GNM post-segmentations. The original GNM algorithm varied only two parameters, \( t_{cp} \) and \( t_{rp} \), which varied both in the [0.01, 0.07] range, with an offset of 0.005. All other GNM parameters were kept with the standard values described in Wangenheim et al. (2007). GNM2 has 2 parameters for each of its two parts: \( in \) (initial ts value), \( it \) (number of iterations), \( ms \) (minimum size to be relevant) and \( mc \) (minimum color similarity). All these parameters were varied for each image looking to produce a large set of results, showing various levels and kinds of segmentations. The ranges used were: \( in = [0.015, 0.055] \), step \( in = 0.001 \); \( it = [2,15] \), step \( it = 1 \); \( ms = [20, 100] \), step \( ms = 10 \); \( mc = [0.03, 0.2] \), step \( mc = 0.01 \).

A series of segmentation results was also produced using those segmentation techniques alone and used to be evaluated against the combined GNM2 and GNM approach. The parameter ranges and increment steps used for these segmentation methods were the following: 1) CSC:
20 ≤ \( \text{threshold} \) ≤ 100, \( \text{step}_{\text{threshold}} = 10 \); 2) Mumford & Shah: 1000 ≤ \( \lambda \) ≤ 15000, \( \text{step}_{\lambda} = 500 \); 3) EDISON: 3 ≤ \( SS \) ≤ 30, \( \text{step}_{SS} = 1 \), \( SR = 8 \); 4) Watershed: 0 ≤ \( threshold \) ≤ 0.5, \( \text{step}_{threshold} = 0.05 \). Blobworld and JSEG are unsupervised techniques and do not require parameters. RHSEG, however, is supervised but, as it is intrinsically a hierarchical segmentation, was not necessary to be tested over a range of values. All RHSEG segmentation results were created with the similarity “entropy” (segmentation parameter number 9 in the \text{params} file), with a factor of convergence equal to 1.75 and with a 0.1 importance to spectral clustering.

In order to evaluate the quality of the results obtained with the GNM and GNM2 combinations, the resulting segmentations were objectively validated against ground truths using distance measures properly 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 object identification. The aim is to achieve the closest segmentation to the expected regions defined by the ground-truth. There are several approaches to these distance measures. According to Cheng et al. (2001), two possible kinds of distances are by counting of pairs and by set matching. In our tests, we used one measure of each these kinds: \text{Rand Index} and \text{Bipartite Graph Matching} (BGM), respectively a pair-counting and a set-matching measure.

The ground-truth images provided by the Berkeley image dataset were tested with both distance measures. Both Rand and BGM produce a result ranging from 0 to 1, where 0 means a perfect match between segmentations and 1 means no relation at all between them. The fact that both measures are in the same range facilitates comparing algorithms with different approaches and features. Details about these two evaluation measures can be found in the appendix A.

All the parameters presented so far were used in order to produce the best possible segmentation with each algorithm within a limited range chosen by the authors after experiments with all algorithms. The result selected from each resulting set of image segmentations produced automatically by each algorithm was the one that obtained the highest score for each image with the segmentation evaluation measures Rand and BGM.

Since each image presented a set of diverging GTs produced by different observers, the resulting Rand and BGM scores for each individual segmentation result were calculated as the mean Rand and BGM scores obtained after processing all GTs of a given image.

4.2 Image Selection Criteria

The images for our evaluation experiment were randomly selected from the Berkeley dataset and afterwards, images considered inadequate were eliminated from the evaluation set. The Berkeley dataset presents, for each image, a set of GTs sketched by different observers. In some images there is a high degree of accordance between GTs, whereas in others observers disagree strongly, such as depicted in Figure 3, where observer: #1130 identifies 208 objects while observer 1102 sees only 3 objects in the same image. This means that to validate the segmentation quality of such an image, while employing a multi-GT validation procedure such as the one we used, always produces a bad result because it is impossible to agree with several GTs that are in strong discordance between themselves. For this reason we manually discarded images presenting such GTs from our test set and accepted images where the difference between GTs consisted in less than one order of magnitude.
5 Results

The GNM and GNM2 combined segmentation results in general obtained a better average score with both quality measures pixel counting and set matching (see Table 1). The GNM2 together with the Mumford and Shah algorithm as its pre-segmentation step presented the best results, with a mean Rand index of 0.1177 and a mean BGM index of 0.1678. It also showed the lowest mean variances and mean standard deviations for both Rand and BGM indexes. These last results show that the GNM2 algorithm was also more robust and stable, when used in combination with the Mumford and Shah algorithm as its pre-segmentation step.

Comparing the GNM and GNM2 combined segmentation results with CSC and Mumford and Shah, GNM and GNM2 shows a sensible improvement in Rand scores for both pre-segmentation techniques.

The BGM scores showed clearly an improvement of GNM2 approach over GNM in terms of set matching evaluation. While the original GNM produced results that were behind those of state-of-the-art algorithms under the BGM index, the GNM2 produced the best BGM result when employing MS as a pre-segmentation algorithm. When employing CSC as its pre-segmentation, GNM2 (mean BGM = 0.1741) lays only slightly behind the Watershed implementation we employed (mean BGM = 0.1734).

There are some reasons that can explain why our current algorithm improved in terms of image quality stability when compared to its former version. We first note some details about the measures used. The BGM evaluation index and its set matching approach tends to penalize a segmentation score because of small residual, unrelated regions, rewarding more aggressive approaches, even if the results are somewhat under-segmented, showing strong segment leakage. BGM, as other set matching indices like GCE and LCE (respectively, Global and Local Consistency Error) (Martin, 2002), isn't reliable when the segmentation is formed by segments of only one pixel or an only segment is found in the image. Thus, as long good enough segmentations are provided to be evaluated with the respective ground-truths, BGM and other set matching techniques are reliable and interesting measures. The Rand approach of pixel counting results in only small penalizations for small regions that are not predicted by the ground-truth, so scores do not suffer as much as in BGM because of segmentations that contain small and unexpected segmented objects. Though this might be a desirable condition sometimes, segmentations with an excess of objects may cause an overload of information and slowing following processes when used by following pattern recognition algorithms, for instance. Our first algorithm, GNM, had a tendency of leaving small fragments unmerged, suffering from lower scores in BGM. Our new approach with GNM2, with the added features of using post processing correction steps and facilitating the union between regions sharing meaningful and long borders, corrected that and still kept similar scores according to the Rand index. The addition of some characteristics not depending only on the colors of the image proved to help the algorithm improve. Both GNM approaches also were usually among those with the smallest standard deviation values in the tests, showing better stability among the results within the same image sets than other approaches.

As the tests we realized resulted in more data than we would be able to properly display in this paper, so we show a sample in Figure 4 and have made all the test data available in a website at http://www.lapix.ufsc.br/gnm.
6 Discussion

We have shown through well known validation measures that the quality of the segmentations generated by our two-step approach is very promising and comparable to segmentations generated by state-of-the-art segmentation methods that were available for comparison when this paper was being written. The Gradient Network Method and its new version are segmentation post-processing methods that are independent of the region-growing method applied to generate the super-segmented input image, as shown by the comparison between the results produced by the different segmentation methods used as the pre-processing step for both our new algorithm and its previous version. We do not intend to single out one segmentation method as the most appropriate to be used with our approach, as we expect that any technique able to provide the features expected by our algorithm as a pre-segmentation step should perform well using our approach. Even though MS+GNM2 showed the best results within our tested image set, proving to be a very viable option, certain kind of images or problems could possibly benefit from pre-segmentations obtained from different algorithms.

We would also like to note that in any evaluation measure that intends to point the “best” possible result in such a complex field, as actually is color image segmentation, the resulting scores have to be analyzed carefully. These measures may not provide an absolute way to determine which actually the best segmentation is, but they provide an interesting objective resource where one can compare and evaluate different segmentations under the optics of the specific approach the measure takes. Both selected measures, Rand and BGM, apply very distinct approaches to what is a better segmentation and, as it is tough to define how correct a segmentation is, it is hard to argue which one proves to be more correct. What these measures actually provide are ways of finding if under expected conditions, reflected by the evaluation approach, a segmentation is able to perform at a desired level of quality. We selected these measures for our tests because of the very reason of looking and accessing the quality of the results through very different but at the same acceptable views of the same problem. Thus observing the results presented here we demonstrated that our new algorithm GNM2, under the conditions expected by both selected measures, can provide segmentations of at least similar quality as several state-of-art algorithms by either a pixel counting or set matching approach. Our new algorithm also improved in terms of stability concerning these different kinds of evaluations, with GNM2 averaging better scores with Rand and BGM than our former algorithm GNM.

Our method could, at a first glance, be viewed as a graph-based method. In this context, graph partitioning has been popular for image segmentation; see (Shi and Malik 2000) and (Wang and Siskind 2003) for two different formulations of graph cut cost functions. As a general concept, graphs cuts have also found applications in energy minimization (Boykov et.al. 2001) and active contours (Sumengen and Manjunath 2006). In contrast to these top-down methods, our approach is a bottom-up method. It intends to provide for efficiency through the usage of a pre-segmentation as a starting condition, which is performed employing extremely conservative parameters to avoid segment leakage. In addition to this fundamental difference between our bottom-up and the above top-down graph-based segmentation methods, our approach explicitly deals with objects which are represented by several different but similar and gradually varying color shades, as they often are found in outdoors scenes.

The new algorithm presented here brought new features as improved perception detection among different colors and reliance on other characteristics as shared borders size besides the
color information of the image. These different heuristics were used in GNM2 with a similar framework used by the original GNM. The enhanced framework presented was created with the purpose of being easily coupled with many different heuristics. The approach our similarity function took is a representation of our goals when dealing with color image segmentation and we intended with this new algorithm that other research activities could be performed with this same framework.

We plan other future works for our algorithm besides coupling it with new heuristics. We are looking to test it with a bigger set of images, to present even more data regarding the quality of GNM2. Evaluating our results with other measures as LCE, GCE and NPR (Normalized Probabilistic Rand) (Unnikrishnan et al., 2007) are also planned. We also look to increase the parameter stability of our algorithm, reducing the number of parameters necessary and making them vary in a smaller range of value for results of similar or close quality. We consider the number of parameters used by GNM2 still massive and we will look to decrease them, making our algorithm more suitable for applications as video segmentation. We intend to make the step concerning the elimination of small, undesirable regions more reliable and efficient quality-wise, because the current approach is naive and can certainly be improved. Even with several new features and tests ahead, the algorithm here presented already shows advance in relation to its former version, while still maintaining the focus on our perceptive differentiation of colors already used by our previous work.

7. Acknowledgments

D.D. Abdala enjoys a CNPq-Brazil Ph.D scholarship under the process number 290101/2006-9.

8. References


14


Appendix

A Distance Measures

A.1 Rand Index

The Rand index first reported in [15] and reviewed in Jiang et al. (2006) is a similarity measure specially developed to evaluate the quality of clustering algorithms by comparison with other clustering results or with a golden standard (in our case, ground-truths). To compare two clustering results $C_1$={c1, c12, ..., c1N} and $C_2$={c21, c22, ..., c2M} over the same image $P$={p1, p2, ..., pK} where each element of $C_1$ or $C_2$ is a subset of $P$ and $c_{1j} = \{p_{1j}, p_{2j}, ..., p_{Lj}\}$, 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 $C_1$ and $C_2$.

The rand index is so defined by eq. A.I

$$R(C_1, C_2) = 1 - \frac{N_{11} + N_{00}}{n(n-1)}$$  \hspace{1cm} A.I

To compute the quantities $N_{11}$ and $N_{00}$ one must iterate over the entire image for each pixel in order to evaluate the conditions defined above given an $O(n^4)$ algorithm. A clever approach is to use the method described in Jiang et al. (2006) where a matching matrix is used to summarize the occurrences of pixels in the respective classes. The matching matrix is constructed allocating each cluster from the clustering $C_1$ to a row and each cluster from clustering $C_2$ 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 kxl size each cell can be defined as $m_{ij} = |c_i \cap c_j|$, $c_i \in C_1$, $c_j \in C_2$.

The quantities $N_{11}$ and $N_{00}$ can be computed in terms of the matching matrix as follows:

$$N_{11} = \frac{1}{2} \left( \sum_{i=1}^{k} \sum_{j=1}^{l} m_{ij}^2 - n \right)$$  \hspace{1cm} A.II
\[ N_{00} = \frac{1}{2} \left( n^2 - \sum_{i=1}^{k} n_i^2 - \sum_{j=1}^{l} n_j^2 + \sum_{i=1}^{k} \sum_{j=1}^{l} m_{ij} \right) \]

where \( n \) is the cardinality of \( P \) and \( n_i \) and \( n_j \) are the cardinality of the clusters \( c_1 \) and \( c_2 \).

### A.2 Bipartite Graph Matching

The BGM index Jiang et al. (2006) compute an one-to-one correlation between clusters at the same time trying to maximize the relationship. It considers each cluster of the \( C_1 \) and \( C_2 \) clustering as vertices of a bipartite graph. Edges are added between each vertex of the two partitions and they are valued as \( |c_1_i \cap c_2_j| \), a value that can be directly extract from the matching matrix. Then the maximum-weight bipartite graph is defined as the subgraph \( \{(c_1_i_1, c_2_i_1), \ldots, (c_1_i_r, c_2_i_r)\} \) where only the edges from \( c_1_i \) to \( c_2_j \) with maximum weight are present. After all max-valued edges were found the overall graph weight is calculated by sum of all remaining edge weights.

\[
BGM(C_1, C_2) = 1 - \frac{W}{n}
\]
Figure Captions

(a) A color image     (b) After pre-segmentation     (c) After GNM2

**Figure 1.** Some images representing the stages GNM2 segmentation goes through. The column at the left is the original image, a pair of flowers. The image at the center is a segmentation done with very careful parameters, creating what is expected from a pre-segmentation for the GNM2 algorithm. The column in the right is the result of GNM2 segmentation with the pre-segmented image. It is easy to notice that the final segmentation is able to separate the flowers from the background while keeping some other relevant features of the image intact.

**Figure 2.** The workflow shows the steps to partition an image into similar objects with GNM2 algorithm. Our goal is that the algorithm serves as a framework for other possible segmentation algorithms. The heuristic used with function $f_s$ in this paper can easily be replaced by other as long the expected conditions are met.

**Figure 3.** Two ground truths of a same image showing strong inter-observer variation.

<table>
<thead>
<tr>
<th>Original</th>
<th>CSC+GNM2</th>
<th>MS+GNM2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>(b)</td>
<td>(c)</td>
</tr>
<tr>
<td>(d)</td>
<td>(e)</td>
<td>(f)</td>
</tr>
</tbody>
</table>

**Figure 4.** Segmentation image examples obtained by the GNM2. The images here were postprocessed after CSC and Mumford-Shah pre-segmentations. The first row corresponds to image number 118035 and the second row corresponds to image 68077 from the Berkeley image dataset. The segmentation sample image is also provided by the Berkeley dataset. More image sets, results comparing several techniques and higher resolution images can be found in http://www.lapix.ufsc.br/gnm.
Table 1. This table shows the mean values of the best Rand and BGM indexes obtained for all 35 images with their respective variance and standard deviation, considering all ground truths of each image.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>RAND</th>
<th>BGM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>variance</td>
</tr>
<tr>
<td>CSC</td>
<td>0,1956</td>
<td>0,0172</td>
</tr>
<tr>
<td>CSC-GNM</td>
<td>0,1269</td>
<td>0,0075</td>
</tr>
<tr>
<td>CSC-GNM2</td>
<td>0,1347</td>
<td>0,0088</td>
</tr>
<tr>
<td>MS</td>
<td>0,2997</td>
<td>0,0320</td>
</tr>
<tr>
<td>MS-GNM</td>
<td>0,1262</td>
<td>0,0087</td>
</tr>
<tr>
<td>MS-GNM2</td>
<td>0,1177</td>
<td>0,0058</td>
</tr>
<tr>
<td>Watershed</td>
<td>0,2196</td>
<td>0,0313</td>
</tr>
<tr>
<td>Jseg</td>
<td>0,1578</td>
<td>0,0101</td>
</tr>
<tr>
<td>Edison</td>
<td>0,2174</td>
<td>0,0177</td>
</tr>
<tr>
<td>BlobWorld</td>
<td>0,2559</td>
<td>0,0070</td>
</tr>
</tbody>
</table>
Pattern Recognition Letters

Authorship Confirmation

Please save a copy of this file, complete and upload as the “Confirmation of Authorship” file.

As corresponding author, I Aldo von Wangenheim, hereby confirm on behalf of all authors that:

1. This manuscript, or a large part of it, has not been published, was not, and is not being submitted to any other journal. If presented at a conference, the conference is identified. If published in conference proceedings, the publication is identified below and substantial justification for re-publication must be presented.

2. All text and graphics, except for those marked with sources, are original works of the authors, and all necessary permissions for publication were secured prior to submission of the manuscript.

3. All authors each made a significant contribution to the research reported and have read and approved the submitted manuscript.

Date 04/08/2008

Previous conference presentation

Previous conference proceedings publication

Justification for re-publication
Figure 2
Click here to download high resolution image
Figure 4b
Click here to download high resolution image