

Adaptive Image Segmentation by Combining Photometric Invariant Region and Edge Information

Theo Gevers, *Member, IEEE*

Abstract—An adaptive image segmentation scheme is proposed employing the Delaunay triangulation for image splitting. The tessellation grid of the Delaunay triangulation is adapted to the semantics of the image data by combining region and edge information. To achieve robustness against imaging conditions (e.g., shading, shadows, illumination, and highlights), photometric invariant similarity measures, and edge computation is proposed. Experimental results on synthetic and real images show that the segmentation method is robust to edge orientation, partially weak object boundaries, and noisy, but homogeneous regions. Furthermore, the method is robust to a large degree to varying imaging conditions.

Index Terms—Image segmentation, adaptive splitting, integrating region and edge information, photometric color invariance, noise robustness.

1 INTRODUCTION

In this paper, image segmentation is considered as the process of partitioning an image into constituent regions of the same surface albedo, i.e., independent of the imaging conditions (e.g., shading, shadows, illumination, and highlights). In the past, to achieve robustness against varying imaging conditions, segmentation methods have been proposed that account for the process of image formation as represented by the contributions collected by Wolff et al. [1]. One of the first methods based on physics considerations is Klinker et al. [2] based on the dichromatic reflection model proposed by Shafer [3]. Bajscy et al. [4] developed a similar image segmentation method using the $H - S$ color space instead of the RGB -color space. Healey [5] proposed a method to segment images on the basis of normalized color.

The above described methods are based on two basic properties of pixels in relation to their local neighborhood: *Discontinuity* or *similarity*. However, segmentation based only on discontinuity often fails when a portion of the edge has a small value difference or when regions are homogeneous but very noisy. Further, segmentation based only on similarity allows for reliable computation of region statistics but often fails to localize region outlines accurately. To this end, segmentation methods have been proposed to combine region and boundary information [6], [7], [8]. Tabb and Ahuja [6] proposes a method of multiscale image segmentation. Chakraborty and Duncan [7] use a game-theoretic approach to integrate information for the segmentation of MRI images. Wu [8] proposes a segmentation method in which splitting of a nonuniform region is performed along the strongest edge in it. However, these segmentation methods do not account for the process of image formation. Consequently, intensity edges or region outlines may appear in the scene without an albedo transition to support them.

Therefore, in this paper, an adaptive image segmentation scheme is proposed which is robust against varying imaging conditions. The segmentation scheme is based on the Delaunay triangulation for image splitting. The novelty of the paper is that

- The author is with the Intelligent Sensory Information Systems, Department of Computer Science, Faculty of Science, University of Amsterdam, Kruislaan-403, 1098-SJ Amsterdam, The Netherlands.
E-mail: gevers@science.uva.nl.

Manuscript received 30 June 2000; revised 22 Apr. 2001; accepted 5 Oct. 2001.

Recommended for acceptance by S. Sclaroff.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 112370.

the tessellation grid of the Delaunay triangulation is adapted to the structure of the image data by combining photometric invariant region and edge information. To this end, new computational models are proposed to obtain photometric invariant similarity measures and edges.

The paper is organized as follows: The adaptive image segmentation scheme is introduced in Section 2. Image segmentation on gray-value and color images is proposed, respectively, in Sections 3 and 4. Experimental results are given in Sections 5 and 6.

2 ADAPTIVE IMAGE SEGMENTATION

Although for neighborhood referencing the quad-tree is simple to implement and computationally efficient, its major drawback is that the segmentation process is unable to adapt the quadtree tessellation grid to arbitrary shaped region outlines. Therefore, in this paper, we consider the Delaunay triangulation as the geometric datastructure for image segmentation. The Delaunay triangulation maximizes the minimum angle, minimizes the maximum circumscribing circle, and minimizes the maximum smallest enclosing circle for each triangle [9]. Therefore, the Delaunay triangulation of a set of points tends to generate regularly shaped triangles and is preferred over alternative triangulations for image segmentation [10], [11].

The adaptive image segmentation method is proposed as follows:

Initialization. Let \mathcal{D}^j denote the incremental Delaunay triangulation after j insertions of points in \mathcal{R}^2 . Let d_i^j be the i th triangle of the j th triangulation. Further, consider the function $g: \mathcal{R}^2 \rightarrow \mathcal{R}$ defining an image surface $g(x, y)$. $g_i^j(x, y)$ is a compact area of g which is bounded by the vertices of triangle d_i^j . Because it is assumed that the image data points are limited to a rectangular image domain, the image segmentation method starts with the construction of the initial triangulation \mathcal{D}^0 consisting of two triangles d_i^0 for $i = 1, 2$ whose vertices are the corners of g .

Splitting. After the construction of \mathcal{D}^0 , the algorithm successively examines triangles d_i^j by computing the similarity predicate $H()$. The similarity predicate is defined on g_i^j denoting the underlying image data of triangle d_i^j . If the similarity predicate is false, edge pixels in g_i^j are classified topographically based on their local neighborhood by the difference function $D()$. Then, the splitting function $S()$ assigns a *transition error* to every edge point. The goal to adapt the image tessellation grid properly to the underlying structure of the image data. As a consequence, the edge point with the lowest transition error is taken and entered into \mathcal{D}^j to generate the next triangulation \mathcal{D}^{j+1} . The splitting phase continues until all triangles satisfy $H()$.

Merging. Let R_i be a point set in \mathcal{R}^2 forming the i th polygon with corresponding $r_i \subset \mathcal{R}^2$ which is a compact area of the plane by merging triangular areas of the final Delaunay triangulation (\mathcal{D}^N). In fact, $r_i = g_1^N \cup g_2^N \cup \dots \cup g_n^N$, where all n triangular image regions are adjacent. The merging phase starts with the triangulation produced by the splitting phase $R_i = d_i^N$ for all i . Function $H()$ provides the criterion by which two adjacent polygons are merged into one.

The algorithm is determined by functions $H()$ ($H_I()$ for intensity images and H_C for color images), $D()$ (i.e., $D_I()$ and $D_C()$) and $S()$. They are discussed in Section 3 for gray-value images and in Section 4 for color images.

3 GRAY-VALUE IMAGES

In this section, we focus on gray-value images. The goal is to get to image segmentation robust to varying imaging conditions. Therefore, a fitting model is proposed which approximates shading and illumination characteristics. The assumption is that the image data

exhibits *surface coherence* within object regions meaning that the image data can be interpreted as a (noisy) piecewise-smooth surface function. Hence, our model is not suited to textured surfaces.

3.1 Similarity Predicate $H_I()$

We consider the similarity predicate of a region to be true if all image pixels in the region can be modeled by an approximating (fitting) function. We follow the assumption that (simple) shading characteristics can be approximated, to a large extent, by low-order bivariate polynomial functions [12]. In this way, uniformly colored regions are approximated by planar, quadric, or cubic surfaces making the segmentation scheme more robust against the imaging conditions. Consequently, we concentrate on bivariate polynomials up to order $m = 3$ mathematically specified as:

$$f_m(x, y) = \sum_{i+j \leq m} a_{ij} x^i y^j, m \leq 3. \quad (1)$$

A least-square solver calculates vector \vec{a} and mean fitting error for an arbitrarily shaped image region T defined as:

$$H_m(T) = \sqrt{\frac{\sum_{x,y \in T} (g(x, y) - f_m(x, y))^2}{N}}, \quad (2)$$

where $g(x, y)$ is the image surface and N is the total number of surface interior pixels of T contributing to the sum.

If the mean noise standard deviation σ in T is estimated, $H_m(T)$ can be compared to this. If the noise is additive, stationary, and has zero mean, an estimate of σ is obtained by a local least-square linear fit to an approximately constant image patch (e.g., 5x5 mask), where edge pixels are discarded from contributing to the sum. Then, the functional homogeneity predicate defined over image region T , to return a Boolean value, is as follows:

$$H_I(T) = \begin{cases} \text{true}, & \text{if } H_m(T) \leq \sigma \\ \text{false}, & \text{otherwise.} \end{cases} \quad (3)$$

3.2 Difference Function $D_I()$

When the homogeneity predicate $H_I(g_i^j)$, defined over a region given by the vertices of d_i^j , is false, the image data in g_i^j deviates from the fitting function and should be broken at *surface salient points* into two or more regions. Surface salient points are defined to be on locations where the image data exhibit significant changes of pixel values causing d_i^j to be heterogeneous. To classify pixels in g_i^j by their topographic characterization, the output of the difference function $D_I()$ is an edge type image $t(x, y) : \mathcal{R}^2 \rightarrow \mathcal{R}$ consisting of:

Edges. Edges are found by differentiating the image region domain. Canny's edge detector is used to compute edges.

Corners. Corners are defined on locations where the image surface g exhibits high values of isophote. We use the corner detector of [13] which employs Gaussian-based fuzzy derivatives.

3.3 Splitting Function $S()$

For each edge point in t_i^j , a *transition error* is computed. The transition error expresses the mean fitting error induced by the next triangulation \mathcal{D}^{j+1} . Let $\hat{p}_l \in \mathcal{R}^2$ be the location (x, y) of an edge pixel in $t_i^j(x, y) = l$, where l denotes the edge type (i.e., edges or corners). Triangles d_i^j of \mathcal{D}^j , for which the circumcircle contains \hat{p}_l , form a region R_r in the plane which is called the *influence region* of \hat{p}_l . Edges not shared by any two of the triangles in R_r compose the *influence polygon* of \hat{p}_l in \mathcal{D}^j and is denoted by R_p . Then, we have the following [9]:

1. The insertion of a new vertex in \mathcal{D}^j modifies triangle d_i^j in \mathcal{D}^j if and only if d_i^j is in R_r .
2. Let \mathcal{D}^{j+1} be the Delaunay triangulation obtained from \mathcal{D}^j by the insertion of \hat{p}_l . \mathcal{D}^{j+1} is obtained from \mathcal{D}^j by deleting all edges and triangles that are internal to R_p and by connecting all vertices of R_p to \hat{p}_l .

Let the newly created triangles of \mathcal{D}^{j+1} within R_p be denoted by $d_{i'}^{j+1}$. We want the fitting errors over all triangles $d_{i'}^{j+1} \in \mathcal{D}^{j+1}$ to be small. Therefore, the *transition error* is defined as follows:

$$e_t^{\hat{p}_l} = \|g_{i'}^{j+1} - f_{i'}\|_{d_{i'}^{j+1}}, \quad (4)$$

where $d_{i'}^{j+1}$ are the newly created triangles of \mathcal{D}^{j+1} obtained by the insertion of \hat{p}_l in \mathcal{D}^j , and $f_{i'}$ is the approximating function evaluated over $g_{i'}^{j+1}$ which is the image surface bounded by vertices of $d_{i'}^{j+1}$. In other words, $e_t^{\hat{p}_l}$ is the fitting error of \mathcal{D}^{j+1} computed over the newly defined triangles. To minimize the fitting error, for location \hat{p}_l for each edge point in $t_i^j(x, y)$, transition error $e_t^{\hat{p}_l}$ is computed and \hat{p}_l yielding lowest transition error

$$\epsilon_{\min} = \min_{\hat{p}_l} e_t^{\hat{p}_l} \quad (5)$$

is taken to generate \mathcal{D}^{j+1} . This is done by erasing from \mathcal{D}^j all triangles d_i^j whose circumcircle contains \hat{p}_l . Then, for each triangle-edge which is not shared by any two of the erased triangles, a new triangle is created with \hat{p}_l and added into \mathcal{D}^j to obtain \mathcal{D}^{j+1} .

3.4 Merging Function $M()$

The merging stage consists of the following steps at each iteration. First, the largest triangular region $R_i = \max_{R_k \in \mathcal{D}^N} \eta(R_k)$, where η denotes the triangle size, is selected as a seed for growing. Let R_j , $j = 1, \dots, M$ be a neighboring polygon of seed R_i . The functional homogeneity predicate $H(R_i \cup R_j)$ (i.e., merge score) is computed. If $H(R_i \cup R_j)$ is below the mean noise standard deviation σ (i.e., merge threshold), R_j is merged with the seed. The merging continues with this newly merged region $R_i \cup R_j$ as seed until all merge-scores of neighbor triangles are larger than the merge threshold. Then, the next largest unmerged triangle is taken as a new seed to grow for the next iteration. This process continues until there are no unmerged triangles left.

4 COLOR IMAGES

In this section, we focus on color images. First, in Section 4.1, photometric color invariance is briefly reviewed. Then, in Section 4.2, the similarity function is given. In Section 4.3, color invariant edges are proposed.

4.1 Color Invariance

To provide robustness against imaging conditions (e.g., illumination, shading, highlights, and interreflections), photometric color invariants have been introduced [14], [15], [16]. For example, effective illumination-independent color ratios have been proposed by Funt and Finlayson [14] and Nayar and Bolle [16]. Further, for the dichromatic reflection model, Gevers and Smeulders [15] showed that normalized color rgb ($c_1 c_2 c_3$) is to a large extent invariant to a change in camera viewpoint, object pose, and for the direction and intensity of the incident light. In addition, the hue color space H ($l_1 l_2 l_3$) is insensitive to highlights under the restriction of white illumination or a white-balanced camera.

In this paper, we focus on:

$$c_1(R, G, B) = \arctan\left(\frac{R}{G}\right), \quad (6)$$

$$c_2(R, G, B) = \arctan\left(\frac{R}{B}\right), \quad (7)$$

$$c_3(R, G, B) = \arctan\left(\frac{G}{B}\right), \quad (8)$$

where R , G , and B are the red, green, and blue channels of a color camera. $c_1c_2c_3$ is insensitive to a large extent to a change in camera viewpoint, object pose, and for the direction and intensity of the incident light [15]. Further, when shadows correspond to a change in intensity, which is often the case, then $c_1c_2c_3$ is also insensitive to shadows. When shadows are strongly colored, then $c_1c_2c_3$ is not shadow-invariant.

Further, we focus on:

$$l_1(R, G, B) = \frac{|R - G|}{|R - G| + |B - R| + |G - B|}, \quad (9)$$

$$l_2(R, G, B) = \frac{|R - B|}{|R - G| + |B - R| + |G - B|}, \quad (10)$$

$$l_3(R, G, B) = \frac{|G - B|}{|R - G| + |B - R| + |G - B|} \quad (11)$$

also insensitive to highlights under the restriction of white illumination or a white-balanced camera [15].

4.2 Similarity Predicate $H_C()$

We define region \mathcal{R} to be homogeneous when the observed color invariant values of the region can be approximated in the color invariant space by a Gaussian distribution with mean and standard deviation due to noise. If the standard deviation is below a predefined threshold, region \mathcal{R} is considered to be homogeneous. Again, the mean noise standard deviation $\hat{\sigma}$ in the image is estimated by applying a least-squares fit to a uniformly colored region (e.g., derived from a 5x5 mask). A more detailed study on noise and instability of color invariants is given by Gevers [17]. Then, the similarity predicate $H_C()$, returning a Boolean value, is given by:

$$H_C(\mathcal{R}) = \begin{cases} \text{true}, & \text{if } \epsilon \leq \hat{\sigma} \\ \text{false}, & \text{otherwise,} \end{cases} \quad (12)$$

where region \mathcal{R} is considered to be homogeneous if the color invariant values of \mathcal{R} form a Gaussian distribution, which falls within the limit of the noise standard deviation.

4.3 Difference Measure $D_C()$

In this section, the principled way is taken to compute gradients in vector images as described by di Zenzo [18] and further used in [19], which is summarized as follows.

Let $\Theta(x_1, x_2) : \mathbb{R}^2 \rightarrow \mathbb{R}^m$ be a m -band image with components $\Theta_i(x_1, x_2) : \mathbb{R}^2 \rightarrow \mathbb{R}$ for $i = 1, 2, \dots, m$. For color images, we have $m = 3$. Hence, at a given image location the image value is a vector in \mathbb{R}^m . The difference at two nearby points $P = (x_1^0, x_2^0)$ and $Q = (x_1^1, x_2^1)$ is given by $\Delta\Theta = \Theta(P) - \Theta(Q)$. Considering an infinitesimal displacement, the difference becomes the differential

$$d\Theta = \sum_{i=1}^2 \frac{\partial\Theta}{\partial x_i} dx_i$$

and its squared norm is given by:

$$d\Theta^2 = \sum_{i=1}^2 \sum_{k=1}^2 \frac{\partial\Theta}{\partial x_i} \frac{\partial\Theta}{\partial x_k} dx_i dx_k = \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix}^T \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} dx_1 \\ dx_2 \end{bmatrix}, \quad (13)$$

where $g_{ik} := \frac{\partial\Theta}{\partial x_i} \cdot \frac{\partial\Theta}{\partial x_k}$ and the extrema of the quadratic form are obtained in the direction of the eigenvectors of the matrix $[g_{ik}]$ and the values at these locations correspond with the eigenvalues given by:

$$\lambda_{\pm} = \frac{g_{11} + g_{22} \pm \sqrt{(g_{11} - g_{22})^2 + 4g_{12}^2}}{2} \quad (14)$$

with corresponding eigenvectors given by $(\cos\theta_{\pm}, \sin\theta_{\pm})$, where $\theta_+ = \frac{1}{2} \arctan \frac{2g_{12}}{g_{11} - g_{22}}$ and $\theta_- = \theta_+ + \frac{\pi}{2}$. Hence, the direction of the minimal and maximal changes at a given image location is expressed by the eigenvectors θ_- and θ_+ respectively, and the corresponding magnitude is given by the eigenvalues λ_- and λ_+ respectively. Note that λ_- may be different than zero and that the strength of a multivalued edge should be expressed by how λ_+ compares to λ_- , for example by subtraction $\lambda_+ - \lambda_-$ as proposed by [19].

Then, the color gradient for RGB is as follows:

$$\nabla C_{RGB} = \sqrt{\lambda_+^{RGB} - \lambda_-^{RGB}} \quad (15)$$

for

$$\lambda_{\pm} = \frac{g_{11}^{RGB} + g_{22}^{RGB} \pm \sqrt{(g_{11}^{RGB} - g_{22}^{RGB})^2 + 4(g_{12}^{RGB})^2}}{2}$$

$$g_{11}^{RGB} = \left| \frac{\partial R}{\partial x} \right|^2 + \left| \frac{\partial G}{\partial x} \right|^2 + \left| \frac{\partial B}{\partial x} \right|^2$$

$$g_{22}^{RGB} = \left| \frac{\partial R}{\partial y} \right|^2 + \left| \frac{\partial G}{\partial y} \right|^2 + \left| \frac{\partial B}{\partial y} \right|^2$$

$$g_{12}^{RGB} = \frac{\partial R \partial R}{\partial x \partial y} + \frac{\partial G \partial G}{\partial x \partial y} + \frac{\partial B \partial B}{\partial x \partial y},$$

where the partial derivatives are computed through Gaussian smoothed derivatives.

Further, we propose that the color invariant gradient (based on $c_1c_2c_3$) for matte objects is given by:

$$\nabla C_{c_1c_2c_3} = \sqrt{\lambda_+^{c_1c_2c_3} - \lambda_-^{c_1c_2c_3}} \quad (16)$$

for

$$\lambda_{\pm} = \frac{g_{11}^{c_1c_2c_3} + g_{22}^{c_1c_2c_3} \pm \sqrt{(g_{11}^{c_1c_2c_3} - g_{22}^{c_1c_2c_3})^2 + 4(g_{12}^{c_1c_2c_3})^2}}{2},$$

where

$$g_{11}^{c_1c_2c_3} = \left| \frac{\partial c_1}{\partial x} \right|^2 + \left| \frac{\partial c_2}{\partial x} \right|^2 + \left| \frac{\partial c_3}{\partial x} \right|^2,$$

$$g_{22}^{c_1c_2c_3} = \left| \frac{\partial c_1}{\partial y} \right|^2 + \left| \frac{\partial c_2}{\partial y} \right|^2 + \left| \frac{\partial c_3}{\partial y} \right|^2,$$

$$g_{12}^{c_1c_2c_3} = \frac{\partial c_1 \partial c_1}{\partial x \partial y} + \frac{\partial c_2 \partial c_2}{\partial x \partial y} + \frac{\partial c_3 \partial c_3}{\partial x \partial y}.$$

Similarly, we propose that the color invariant gradient (based on $l_1l_2l_3$) for shiny objects is given by:

$$\nabla C_{l_1l_2l_3} = \sqrt{\lambda_+^{l_1l_2l_3} - \lambda_-^{l_1l_2l_3}} \quad (17)$$

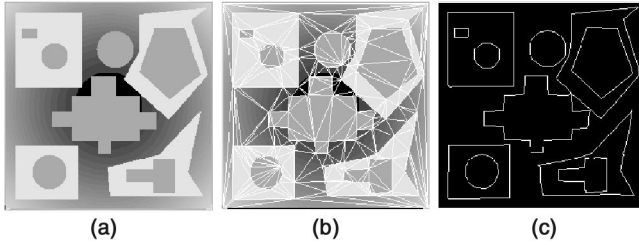


Fig. 1. (a) Synthetic image composed of randomly placed geometric shapes on a background of quadratically varying intensity. (b) Splitting result superimposed on the image. (c) Final result of the segmentation method.

for

$$\lambda_{\pm} = \frac{g_{11}^{l_1 l_2 l_3} + g_{22}^{l_1 l_2 l_3} \pm \sqrt{(g_{11}^{l_1 l_2 l_3} - g_{22}^{l_1 l_2 l_3})^2 + 4(g_{12}^{l_1 l_2 l_3})^2}}{2},$$

where

$$g_{11}^{l_1 l_2 l_3} = \left| \frac{\partial l_1}{\partial x} \right|^2 + \left| \frac{\partial l_2}{\partial x} \right|^2 + \left| \frac{\partial l_3}{\partial x} \right|^2,$$

$$g_{22}^{l_1 l_2 l_3} = \left| \frac{\partial l_1}{\partial y} \right|^2 + \left| \frac{\partial l_2}{\partial y} \right|^2 + \left| \frac{\partial l_3}{\partial y} \right|^2,$$

$$g_{12}^{l_1 l_2 l_3} = \frac{\partial l_1}{\partial x} \frac{\partial l_1}{\partial y} + \frac{\partial l_2}{\partial x} \frac{\partial l_2}{\partial y} + \frac{\partial l_3}{\partial x} \frac{\partial l_3}{\partial y}.$$

5 EXPERIMENTS ON GRAY-VALUE IMAGES

In Section 5.1, we focus on the sensitivity of the method to various object edge orientations and positions. Further, the segmentation method is tested on a real image with respect to shading and illumination variations. Finally, in Section 5.2, robustness of the method is tested with respect to partially weak object edges and noisy homogeneous regions.

5.1 Sensitivity to Edge Orientations and Positions

A polynomial image has been synthesized, see Fig. 1a. The size of the image is 128x128 with 256 gray levels. The image consists of a background of quadratically varying brightness and randomly placed geometric shapes with different constant brightness including a variety of circles and polygons. To simulate ramp edges obtained by a camera, Gaussian smoothing with $\sigma_g = 1.0$ has been performed on the image. In Fig. 1b, the result is shown of the splitting phase. The method is effective in the correct adjustment of the tessellation grid to low-level image statistics where vertices of the tessellation grid neatly correspond to topographically most promising edge pixels. In Fig. 1c, the image segmentation result is presented. Although the test image consists of varying background brightness and objects with different boundary orientations and positions, the method yields a geometrically good and topologically sound image segmentation result. The number of split operations was 110 resulting in a run time of 1.6 seconds on a Ultra 10 Sparc station.

Further, we illustrate the segmentation method for a real image, see Fig. 2a. The image consists of three objects on a background. Objects consist of plastic material and painted wood. The objects are painted homogeneously. Note that, in this paper, image segmentation is considered as the process of partitioning an image into constituent regions of the same surface albedo. Therefore, our method is not suited to textured surfaces. Objects were recorded

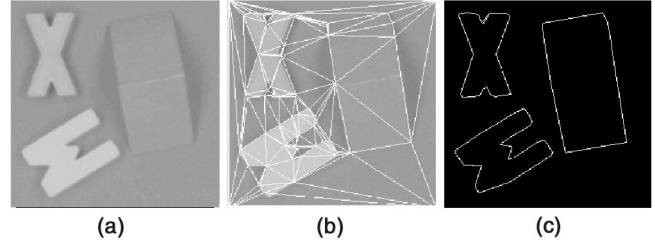


Fig. 2. (a) Real image consisting of three homogeneously colored objects on a homogeneously painted background. (b) The splitting result superimposed on the image. (c) The final segmentation result.

with the aid of a SONY XC-003P camera. The digitization was done in 8 bits. Two light sources of average day-light color were used to illuminate the objects in the scene. The size of the image was 256x256 with 256 gray levels. The order of the polynomial approximating criterion was set to $m = 3$. Hence, we considered the case where the shading and illumination variations can be approximated by a polynomial function up to order 3. The measured mean noise standard deviation is $\sigma = 3.84$ and used for the surface fitting process defined by $H_I(\cdot)$.

In Fig. 2b, the splitting result is superimposed on the original image. As one can see, proper conformation is established between tessellation grid and low-level image characteristics, where grid vertices correspond to corners placing grid edges along region borders. The method is efficient where the number of split operations is 85 yielding a run time of 0.9 seconds.

5.2 Robustness to Noise

Two basic sets of synthetic images are generated to simulate objects of different shapes and sizes [20]. The size of the images are 128x128 with 256 gray values. The first set of images is composed of centered squares of different sizes s . The object value is $v_o = 144$ and the background value is $v_b = 112$. Images have been created for $s \in \{60, 40, 30, 12\}$, where the area of the rectangular objects varies approximately from 20 percent of the image area, through 10 percent, 5 percent down to 1 percent. The second set consists of centered circular discs with different radii r with object value v_o on the middle of a homogeneous background v_b . Images of discs of radii $r \in \{32, 22, 16, 7\}$ are generated to obtain disc objects with areas from approximately 20 percent, 10 percent, and 5 percent to 1 percent of the image area. Again, to simulate ramp edges obtained by a camera, Gaussian smoothing with $\sigma_g = 1.0$ has been performed on the images.

The effect of noise is produced by adding independent zero-mean additive Gaussian noise with $\sigma = \sigma_n$ to the images. The Signal-to Noise Ratio (SNR) is defined as:

$$SNR = \left(\frac{v_o - v_b}{\sigma_n} \right)^2. \quad (18)$$

We use $SNR \in \{100, 64, 16, 4, 2, 1\}$ in the experiments.

To quantify the geometric precision of the segmentation method, an error measure \mathcal{E} is defined. Let X be the image raster and a a binary image containing the "true" shape A defined by $A = \{\vec{x} \in X : a(\vec{x}) = 1\}$. Further, let b be a binary image, called the segmented image, containing the image segmentation result $B = \{\vec{x} \in X : b(\vec{x}) = 1\}$. Evaluation measure \mathcal{E} compares B with A to return a numerical measure of discrepancy. Let $d(\vec{x}, A)$ denote the shortest distance from pixel $\vec{x} \in X$ to $A \subseteq X$, then Pratt's figure of merit FOM [20], is defined as follows:

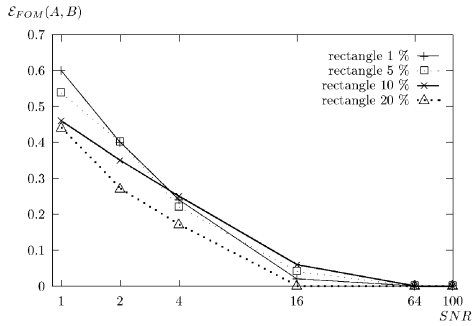


Fig. 3. The average value of \mathcal{E}_{FOM} differentiated for rectangles corresponding to 20 percent, 10 percent, 5 percent, and 1 percent of the image size against the SNR .

$$\mathcal{E}_{PRATT}(A, B) = \left(\frac{1}{\max(\eta(A), \eta(B))} \sum_{\vec{x} \in B} \frac{1}{1 + \alpha d(\vec{x}, A)^2} \right), \quad (19)$$

where α is usually set to $1/9$ and we will follow suit. We prefer to use $\mathcal{E}_{FOM}(A, B) = 1 - \mathcal{E}_{PRATT}(A, B)$ for the ease of graphical illustration.

For each SNR level, we have generated $N = 10$ images. In other words, independent Gaussian noise has been added separately at level $SNR \in \{100, 64, 16, 4, 2, 1\}$, for each noise-free image 10 times. Then, after segmentation, for each of the 60 realizations, \mathcal{E}_{FOM} is computed over the largest region and the average value over $N = 10$ is taken as the final result. In Figs. 3 and 4, accuracy graphs are shown for respectively rectangles and circles of different sizes for the SNR values. First, we concentrate on the quality of the segmentation results with respect to different SNR levels. For $SNR < 4$, results show a rapid decrease in the performance of the method with respect to the noise. For $SNR > 16$, the results get close to 0 (for rectangles) or to a constant value (for circles), an artifact caused by the ill-defined representation of the object's boundary on the rectangular grid. In conclusion, the method gives good results up to considerable amounts of noise ($SNR = 2$), where object boundaries are on average within two pixels from the reference boundary, even for the smallest object. Second, attention is focused on the performance of the method for objects of different sizes in response to noise. Obviously, the method performs better on images containing larger objects. Finally, the performance of the method is studied for objects of different shapes with respect to the noise. The shape and slope of the curves for rectangles and circles do not differ significantly except for the error introduced by the ill-defined representation of the circular image objects on the rectangular grid. Apart from this bias, the performance of the method is approximately the same for the two different geometric shapes.

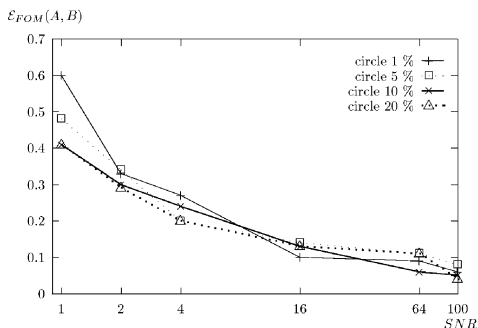


Fig. 4. The average value of \mathcal{E}_{FOM} differentiated for circles corresponding to 20 percent, 10 percent, 5 percent, and 1 percent of the image size against the

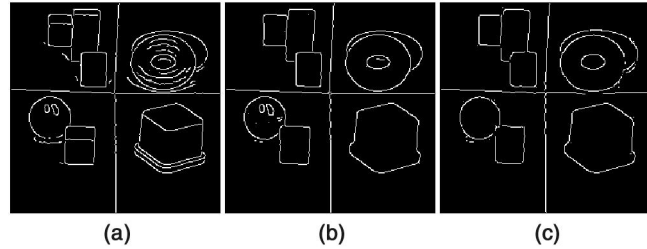


Fig. 5. Edge maps of the various color models computed from the first recorded color image shown in Fig. 6a. (a) Edge map based on RGB gradient field ∇C_{RGB} with nonmaximum suppression. (b) Edge map based on $c_1c_2c_3$ gradient field $\nabla C_{c_1c_2c_3}$ with nonmaximum suppression. (c) Edge map based on $l_1l_2l_3$ gradient field $\nabla C_{l_1l_2l_3}$ with nonmaximum suppression.

6 EXPERIMENTS ON COLOR IMAGES

Fig. 6a is an image of several objects against a background consisting of four squares. The size of the image is 256×256 . The image has been recorded by the SONY XC-003P and the Matrox Magic Color frame grabber. The digitization was done in 8 bits per color. Two light sources of average day-light color are used to illuminate the objects in the scene. The image is clearly contaminated by shadows, shading, highlights, and inter-reflections. Inter-reflections occur when an object receives the reflected light from other objects. In Fig. 5a, edges are shown obtained from the RGB image with nonmaximum suppression with $\sigma_g = 1.0$ used for the Gaussian-based fuzzy derivatives. Clearly, edges are introduced by abrupt surface orientations, shadows, inter-reflections and highlights. In contrast, computed edges for $c_1c_2c_3$ and $l_1l_2l_3$ defined by $\nabla C_{c_1c_2c_3}$ and $\nabla C_{l_1l_2l_3}$, respectively, shown in Fig. 5b and 5c, are insensitive for shadows, surface orientation changes, and highlights (only for $\nabla C_{l_1l_2l_3}$).

To avoid edge grouping, to obtain proper region outlines with closed contours, the $l_1l_2l_3$ edge map is used as the input of the region-based segmentation method. Again, in Fig. 6a, the recorded color image is shown. The mean noise standard deviation is estimated by applying a least-squares fit to a uniformly colored region (5×5 mask). The measured mean noise standard deviation is $\hat{\sigma} = 3.1$ and used as the threshold for the similarity predicate $H_C(\cdot)$. The splitting result is shown in Fig. 6b. The final segmentation result is shown in Fig. 6c. Despite the various radio-metrical and geometrical variations caused by the imaging process, region outlines correspond neatly to material boundaries.

7 CONCLUSION

In this paper, an adaptive image segmentation method has been proposed based on Delaunay subdivision. The method is robust to various object edge orientations and positions, a favorable

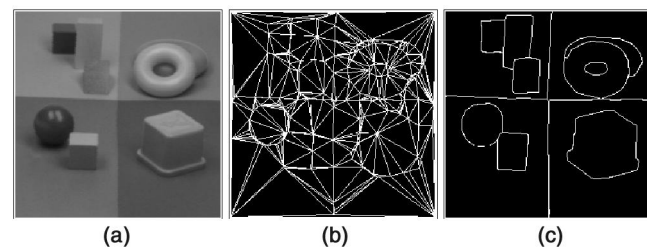


Fig. 6. (a) First recorded color image. (b) Splitting result based on Delaunay splitting. (c) The final segmentation result of the region-based segmentation method.

property over quadtree-based split and merge strategies. Further, the method gave robust results even with considerable amounts of noise. Experiments on real images indicates that the method yields geometrically good and topologically sound image segmentation results robust to the varying imaging conditions. The segmentation method is restricted to white illumination or a white-balanced camera.

REFERENCES

- [1] L. Wolff, S.A. Shafer, and G.E. Healey, *Physics-Based Vision: Principles and Practice*. Jones and Bartlett, 1992.
- [2] G.J. Klinker, A. Shafer, and T. Kanada, "A Physical Approach to Color Image Understanding," *Int'l J. Computer Vision*, vol. 4, pp. 7-38, 1990.
- [3] S.A. Shafer, "Using Color to Separate Reflection Components," *Color Resolution Applications*, vol. 10, no. 4, pp. 210-218, 1985.
- [4] R.K. Bajcsy, S.W. Lee, and A. Leonardis, "Detection of Diffuse and Specular Interface Reflections and Interreflections by Color Image Segmentation," *Int'l J. Computer Vision*, no. 3, pp. 241-272, 1996.
- [5] G. Healey, "Segmenting Images Using Normalized Color," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 22, pp. 64-73, 1992.
- [6] M. Tabb and N. Ahuja, "Multi-Scale Image Segmentation by Integrated Edge and Region Detection," *IEEE Trans. Image Processing*, vol. 6, no. 5, pp. 642-655, 1997.
- [7] A. Chakraborty and Y.S. Duncan, "Game-Theoretic Integration for Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 2, no. 1, pp. 12-30, Jan. 1999.
- [8] A. Wu, "Adaptive Split-and-Merge Segmentation Based on Piecewise Least-Square Approximation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 8, Aug. 1993.
- [9] D.L. Lee and B.J. Schachter, "Two Algorithms for Constructing a Delaunay Triangulation," *Int'l J. Computer and Information Science*, vol. 9, no. 3, pp. 219-424, 1980.
- [10] F. Davoine, M. Antonini, J.M. Chassery, and M. Barlaud, "Fractal Image Compression Based on Delaunay Triangulation and Vector Quantization," *IEEE Trans. Image Processing*, vol. 5, no. 2, pp. 338-346, 1996.
- [11] M. Tuceryan and A.K. Jain, "Texture Segmentation Using Voronoi Polygons," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 2, pp. 211-216, Feb. 1990.
- [12] P.J. Besl and R.C. Jain, "Segmentation through Variable-Order Surface Fitting," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 10, no. 2, pp. 167-192, Feb. 1988.
- [13] B.M. ter Haar Romeny, "Scale-Space: its Natural Operators and Differential Invariants," *Proc. and Medical Imaging*, pp. 239-255, 1991.
- [14] B.V. Funt and G.D. Finlayson, "Color Constant Color Indexing," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 5, pp. 522-529, May 1995.
- [15] T. Gevers and A.W.M. Smeulders, "Color Based Object Recognition," *Pattern Recognition*, no. 32, pp. 453-464, 1999.
- [16] S.K. Nayar and R.M. Bolle, "Reflectance Based Object Recognition," *Int'l J. Computer Vision*, vol. 17, no. 3, pp. 219-240, 1996.
- [17] T. Gevers, "Robust Histogram Construction from Color Invariants," *Proc. IEEE Int'l Conf. Computer Vision (ICCV)*, pp. 126-132, 2001.
- [18] S. di Zenzo, "Gradient of a Multiimages," *Computer Vision and Graphic Image Processing*, vol. 33, pp. 116-125, 1986.
- [19] G. Sapiro and D.L. Ringach, "Anisotropic Diffusion of Multivalued Images with Applications to Color Filtering," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 5, no. 11, pp. 1582-1586, Nov. 1996.
- [20] Y.J. Zhang and J.J. Gerbrands, "Comparison of Thresholding Techniques Using Synthetic Images and Ultimate Measurement Accuracy," *Proc. IAPR, Int'l Conf. Pattern Recognition (ICPR)*, vol. B, pp. 209-213, 1992.

► For more information in this or any computing topic, please visit our Digital Library at <http://computer.org/publications/dilib>.