Image Segmentation and Similarity of Color-Texture Objects

Theo Gevers, *member, IEEE*

Abstract— **We aim for content-based image retrieval of textured objects in natural scenes under varying illumination and viewing conditions. To achieve this, image retrieval is based on matching feature distributions derived from color invariant gradients. To cope with object cluttering, region-based texture segmentation is applied on the target images prior to the actual image retrieval process. The retrieval scheme is empirically verified on color images taken from textured objects under different lighting conditions.**

Keywords— **Texture, image segmentation, image similarity, retrieval, color constancy, color invariance, image databases, digital libraries.**

I. Introduction

Content-based image retrieval from digital image libraries is a difficult but interesting problem. Recently, a large number of image databases have been created in various narrow domains such as flowers, fish, pills, and tiles. Also more general image databases are being published such as photo-stocks of consumer photography and the WWW [2], [14].

In this paper, we aim for content-based image retrieval of non-uniformly textured objects in natural scenes under varying illumination and viewing conditions. Non-uniformly textured objects are specified as objects composed of irregularly distributed texture elements such a sand, wood, brick, and grass.

It is known that texture can be described by its color primitives and their spatial layout. The spatial layout can be periodic, quasi-periodic or random [3], [7]. As the goal is to retrieve images containing objects having irregular texture organization, in this paper, the spatial organization of these texture primitives is, in worst case, random. Therefore, we focus on statistical texture measures. It has been demonstrated that for irregular texture, the comparison of gradient distributions achieves satisfactory accuracy [10], [11] as opposed to fractal or wavelet features. Therefore, most of the work on texture image segmentation are stochastic from nature [1], [8], [16]. However, these methods rely on grey-value information which is very sensitive to the imaging conditions.

Therefore, in this paper, computational methods are proposed to measure texture similarity based on comparing feature distributions derived from color constant ratio gradients. To cope with object cluttering, region-based texture segmentation is applied on the target images prior to the actual image retrieval process. The region-based segmentation algorithm computes image regions having roughly the same texture content as the query texture image. After segmenting the target images, the retrieval process is based on comparing color ratio gradient distributions of query texture image and target regions.

This paper is organized as follows. In Section II, related work is discussed. In Section III, robust color constant gradients are presented. In Section IV, texture measures are given. Further, in Section V, the segmentation and retrieval methods are presented. Finally, the proposed segmentation and retrieval methods are tested in Section VI.

II. Related Work

Various color-based image search systems have been proposed based on various representation schemes such as color histograms, color moments, color edge orientation, color texture, and color correlograms [2], [14]. In general, these retrieval systems are based on comparing global color feature distributions computed from the query image and target images in the image database. However, a drawback of global image matching is object cluttering and occlusion, introducing and discarding respectively color feature values. To cope with this, it has been shown [5] that comparing color distributions based on histogram intersection is to some degree robust to object occlusion and cluttering. However, mismatches still occur when the color feature values are coming from different non-connected regions, corresponding to different objects in an image, producing together roughly the same color distribution as that of the query object. Mismatches may decrease when only color feature dis-

Theo Gevers is with the Computer Science Institute, University of Amsterdam, Kruislaan 403, 1098 SJ, Amsterdam, The Netherlands, gevers@science.uva.nl.

tributions of connected pixels (i.e. blobs) in images are considered during the retrieval process [1]. To this end, our approach is to segment target images into textured regions or blobs prior to the actual retrieval process. In fact, image segmentation is based on finding connected regions with similar color feature distributions as the query object image. After segmentation, these regions are then used for image retrieval.

For both segmentation and image similarity, color features are required which are robust to a change in illumination as well as a change in object pose. Therefore, in the next section, color ratio gradients are computed. These color ratio gradients are derived from the RGB channels from a color ccd camera.

III. Color Ratio Gradients

In this section, various color ratio gradients are presented. These color ratio gradients are independent of the spectral power of the light source (color constancy) and object geometry (normalization). First, in Section III-A, we discus the image formation model. Then, in Section III-B, color ratio gradients are presented robust to a variation in the spectral power distribution of the light source. In Section III-C, we tighten the restriction on the SPD of the light source by assuming that the SPD is relatively smooth.

A. Image Formation

Consider the body reflection term of the dichromatic reflection model with narrow-band filters [12]:

$$C_k(\vec{x}) = G_B(\vec{x}, \vec{n}, \vec{s}) E(\vec{x}, \lambda_k) B(\vec{x}, \lambda_k), \qquad (1)$$

giving the measured sensor pulse response of a matte, infinitesimal surface patch of an inhomogeneous dielectric object under unknown spectral power distribution of the illumination. Although standard video cameras are not equipped with narrow-band filters, spectral sharpening could be applied [4] to achieve this to a large extent. Note that the material is matte ignoring specular or surface reflection. $G_B(\vec{x}, \vec{n}, \vec{s})$ is the geometric term dependent on the surface orientation $\vec{n}$ and illumination direction $\vec{s}$. For example, assuming Lambertian reflection we have $\cos(\vec{n} \cdot \vec{s})$. Further, $E(\vec{x}, \lambda_k)$ is the illumination and $B(\vec{x}, \lambda_k)$ is the surface albedo at wavelength λ_k. The aim is to derive expressions only based on $B(\vec{x}, \lambda_k)$ discounting dependencies on object geometry G_B (i.e. normalization) and illumination E (i.e. color constancy).

B. Color Constant Ratio Gradients

Illumination-independent color ratios have been proposed by Nayar and Bolle [9]. A drawback, however, is that these color ratios might be negatively affected by the geometry and pose of the object.

We propose the following color constant color ratio:

$$M(C_1^{\vec{x}_1}, C_1^{\vec{x}_2}, C_2^{\vec{x}_1}, C_2^{\vec{x}_2}) = \frac{C_1^{\vec{x}_1} C_2^{\vec{x}_2} - C_1^{\vec{x}_2} C_2^{\vec{x}_1}}{C_1^{\vec{x}_2} C_2^{\vec{x}_1} + C_1^{\vec{x}_1} C_2^{\vec{x}_2}}, C_1 \neq C_2, \qquad (2)$$

ranging from $[-1, 1]$ expressing the color ratio between two neighboring image locations, for $C_1, C_2 \in \{C_1, C_2, ..., C_N\}$ giving the measured sensor pulse response at different wavelengths, where $\vec{x}_1$ and $\vec{x}_2$ denote the image locations of the two neighboring pixels.

For a standard RGB color camera, we have:

$$M_1(R^{\vec{x}_1}, R^{\vec{x}_2}, G^{\vec{x}_1}, G^{\vec{x}_2}) = \frac{R^{\vec{x}_1} G^{\vec{x}_2} - R^{\vec{x}_2} G^{\vec{x}_1}}{R^{\vec{x}_2} G^{\vec{x}_1} + R^{\vec{x}_1} G^{\vec{x}_2}}, \qquad (3)$$

$$M_2(R^{\vec{x}_1}, R^{\vec{x}_2}, B^{\vec{x}_1}, B^{\vec{x}_2}) = \frac{R^{\vec{x}_1} B^{\vec{x}_2} - R^{\vec{x}_2} B^{\vec{x}_1}}{R^{\vec{x}_2} B^{\vec{x}_1} + R^{\vec{x}_1} B^{\vec{x}_2}}, \qquad (4)$$

$$M_3(G_{\vec{x}_1}, G^{\vec{x}_2}, B^{\vec{x}_1}, B^{\vec{x}_2}) = \frac{G^{\vec{x}_1} B^{\vec{x}_2} - G^{\vec{x}_2} B^{\vec{x}_1}}{G^{\vec{x}_2} B^{\vec{x}_1} + G^{\vec{x}_1} B^{\vec{x}_2}}. \qquad (5)$$

The color ratio difference is independent of the illumination intensity and color, and also to a change in viewpoint, object geometry, and illumination direction as shown by substituting equation (1) in equation (2):

$$M(C_1^{\vec{x}_1}, C_1^{\vec{x}_2}, C_2^{\vec{x}_1}, C_2^{\vec{x}_2}) = \frac{C_1^{\vec{x}_1} C_2^{\vec{x}_2} - C_1^{\vec{x}_2} C_2^{\vec{x}_1}}{C_1^{\vec{x}_2} C_2^{\vec{x}_1} + C_1^{\vec{x}_1} C_2^{\vec{x}_2}} =$$

$$\frac{B(\vec{x_1}, \lambda_{C_1}) B(\vec{x_2}, \lambda_{C_2}) - B(\vec{x_2}, \lambda_{C_1}) B(\vec{x_1}, \lambda_{C_2})}{B(\vec{x_2}, \lambda_{C_1}) B(\vec{x_1}, \lambda_{C_2}) + B(\vec{x_1}, \lambda_{C_1}) B(\vec{x_2}, \lambda_{C_2})}. \qquad (6)$$

The gradient of the color constant color ratio is given by:

$$\nabla \mathcal{C}_M(C_1^{\vec{x}_1}, C_1^{\vec{x}_2}, C_2^{\vec{x}_1}, C_2^{\vec{x}_2}) =$$

$$(M(C_1^{(x-1,y)}, C_1^{(x+1,y)}, C_2^{(x-1,y)}, C_2^{(x+1,y)})^2 +$$

$$M(C_1^{(x,y-1)}, C_1^{(x,y+1)}, C_2^{(x,y-1)}, C_2^{(x,y+1)})^2)^{1/2}, \qquad (7)$$

where the locations $(x-1, y), (x+1, y), (x, y-1), (x, y+1)$ are image locations with respect to $\vec{x} = (x, y)$.

C. Color Invariance and Normalization

Color ratio gradient $\nabla \mathcal{C}_M$ requires narrow-band filters to achieve full color constancy. Although, general purpose color CCD cameras do not contain narrow-band filters, spectral sharpening could be applied [4]

to achieve this to a large extent. An alternative way is to assume that the illumination has a smoothly distributed spectral power over the wavelengths (e.g. white light). Therefore, in Section III-C.1, color ratios are discussed which are independent of the imaging conditions assuming a smooth SPD. In Section III-C.2, gradients are computed from these color ratios.

C.1 Color Invariant Ratios

Reconsider the body reflection term of the dichromatic reflection under a smooth SPD such as white illumination but without narrow-band filters:

$$\beta_k(\vec{x}) = G_B(\vec{x}, \vec{n}, \vec{s})E(\vec{x}) \int_\lambda B(\vec{x}, \lambda)F_k(\lambda)d\lambda, \quad (8)$$

giving the kth sensor response of an infinitesimal matte surface patch.

Let's focus on the 3-dimensional RGB-space given by:

$$C_b = G_B(\vec{n}, \vec{s})E \int_\lambda B(\lambda)F_R(\lambda)d\lambda, \quad (9)$$

leaving out the spatial dependency $\vec{x}$, where $C_b = \{R_b, G_b, B_b\}$ i.e. R_b, G_b, and B_b denote the red, green, and blue sensor response of an infinitesimal matte surface patch under the assumption of a white and smooth SPD of the light source.

To parameterize, the color invariant model is based on the polar coordinates $\theta_1\theta_2$ derived from RGB given by [5]:

$$\theta_1 = \arctan(\frac{R}{G}), \quad (10)$$

$$\theta_2 = \arctan(\frac{R}{B}), \quad (11)$$

which are insensitive to surface orientation, illumination direction and illumination intensity derived by substituting equation (9) in equation (10) - (11):

$$\theta_1(R_b, G_b, B_b) = \arctan(\frac{G_B(\vec{n}, \vec{s})E \int_\lambda B(\lambda)F_R(\lambda)d\lambda}{G_B(\vec{n}, \vec{s})E \int_\lambda B(\lambda)F_G(\lambda)d\lambda}) =$$

$$\arctan(\frac{\int_\lambda B(\lambda)F_R(\lambda)d\lambda}{\int_\lambda B(\lambda)F_G(\lambda)d\lambda}), \quad (12)$$

only dependent on the sensors and the surface albedo. Equal argument holds for θ_2.

As $\theta_1\theta_2$ is computed from the same position and consequently do not contain any local (spatial) information, in the next section, we present a way of taken gradients in the $\theta_1\theta_2$ domain.

C.2 Color Invariant Ratio Gradients

If each color channel is considered separately, the amount of color change can be estimated by summing up the gradient magnitudes. For notational simplicity, in this paper, the color channels of an image are differentiated in the x and y direction using the Prewitt filter giving the gradient as $\left(\frac{\partial c_i}{\partial x}, \frac{\partial c_i}{\partial y}\right)$. Here, c_i is the notation for a particular color channel. The modulus of the gradient ∇F of the color planes is obtained by taking the Euclidean distance:

$$\nabla F = \sqrt{\sum_{i=1}^{N} \left[\left(\frac{\partial c_i}{\partial x}\right)^2 + \left(\frac{\partial c_i}{\partial y}\right)^2 \right]}, \quad (13)$$

where N is the dimensionality of the color space (i.e. $N = 3$ for a RGB camera).

Often false gradients are introduced due to sensor noise. These false gradients are usually eliminated by using a threshold value determining the minimum acceptable gradient modulus. We aim at providing a computational framework to automatically determine a local threshold value.

Let the result of a measurement of a quantity u be:

$$\hat{u} = u_{\text{est}} \pm \sigma_u, \quad (14)$$

where u_{est} is the average value and σ_u the standard deviation. Suppose that $u, \cdots, w$ are measured with corresponding uncertainties $\sigma_u, \cdots, \sigma_w$, and the measured values are used to compute the function $q(u, \cdots, w)$. If the uncertainties in $u, \cdots, w$ are independent, random and relatively small, then the predicted uncertainty in $\hat{q}$ [15] is:

$$\sigma_q = \sqrt{\left(\frac{\partial q}{\partial u}\sigma_u\right)^2 + \cdots + \left(\frac{\partial q}{\partial w}\sigma_w\right)^2}, \quad (15)$$

where $\partial q/\partial u$ and $\partial q/\partial w$ are the partial derivatives of q with respect to u and w.

In any case, the uncertainty in q is never larger than the city block distance

$$\sigma_q \leq \left|\frac{\partial q}{\partial u}\right| \sigma_u + \cdots \left|\frac{\partial q}{\partial w}\right| \sigma_w. \quad (16)$$

Substitution of (10) - (11) in (15) gives the uncertainty for the $\theta_1\theta_2$ coordinates

$$\sigma_{\theta_1} = \sec^2 \frac{RG\sigma_R^2 - R^2\sigma_G^2}{G^3},$$
$$\sigma_{\theta_2} = \sec^2 \frac{BG\sigma_B^2 - B^2\sigma_G^2}{G^3}, \quad (17)$$

where σ_R, σ_G and σ_B denote the uncertainties in RGB-space. Further, to propagate the uncertainties from these color components through the gradient modulus, the uncertainties are determined using (16) because the transformed color components are dependent. Using (16), the propagation of uncertainty of the Prewitt filter can be implemented by filtering the uncertainty planes of the different color spaces with the absolute masks yielding the uncertainties in the gradient $\sigma_{\partial c/\partial x}$ and $\sigma_{\partial c/\partial y}$. Then, the uncertainty in the gradient modulus is determined using (16) as

$$\sigma_{\nabla F} \leq \frac{\sum_i \left[(\partial c_i/\partial x) \cdot \sigma_{\partial c_i/\partial x} + (\partial c_i/\partial y) \cdot \sigma_{\partial c_i/\partial y}\right]}{\sqrt{\sum_i \left[(\partial c_i/\partial x) + (\partial c_i/\partial)\right]}}, \tag{18}$$

where i is the dimensionality of the color space. In this way, the effect of measurement uncertainty due to photon noise is propagated through the color invariant gradient.

If the average number of counts is large, then the photon counting noise (Poisson distribution) is well-approximated by the Gaussian distribution [15]. For a Gaussian distribution, 99% of the values fall within a 3σ margin. If a gradient modulus is detected which exceeds $3\sigma_{\nabla F}$, we assume that there is 1% chance that this gradient modulus corresponds to a non-color transition. Therefore, for the $\theta_1\theta_2$ color model, we have the following gradient

$$\nabla C_{\theta_1\theta_2}(\vec{x}) = \begin{cases} \nabla F_{\theta_1\theta_2}(\vec{x}) & \text{if } \nabla F_{\theta_1\theta_2}(\vec{x}) > 3\sigma_{\nabla F_{\theta_1\theta_2}}(\vec{x}) \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

and for the standard RGB color space we obtain

$$\nabla C_{RGB}(\vec{x}) = \begin{cases} \nabla F_{RGB}(\vec{x}) & \text{if } \nabla F_{RGB}(\vec{x}) > 3\sigma_{\nabla F_{RGB}}(\vec{x}) \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

The novelty of our approach is that the threshold value is automatically and locally adapted to the amount of uncertainty of the color (invariant) gradient.

IV. Texture Measures

Color ratio gradient histograms are created for each image in the image database by counting the number of times a discrete color ratio gradient occurs in the image. The color constant histogram from the query image is created in a similar way. Then, image retrieval is reduced to the problem to what extent histogram $\mathcal{H}^{\mathcal{Q}}$ derived from the query image $\mathcal{Q}$ is similar to a histogram $\mathcal{H}^{I_k}$ constructed for each image I_k in the image database. A similarity function $\mathcal{D}(\mathcal{H}^{\mathcal{Q}}, \mathcal{H}^{I_k})$ is required returning a numerical measure of similarity between $\mathcal{H}^{\mathcal{Q}}$ and $\mathcal{H}^{I_k}$.

Firstly, $\mathcal{D}$ is expressed by histogram intersection [13]:

$$\mathcal{D}_a(\mathcal{H}^{\mathcal{Q}}, \mathcal{H}^{I_i}) = \frac{\sum_{\vec{k}=1}^{N_d} \min\{\mathcal{H}^{\mathcal{Q}}(\vec{k}), \mathcal{H}^{I_i}(\vec{k})\}}{\sum_{\vec{k}=1}^{N_d} \mathcal{H}^{\mathcal{Q}}(\vec{k})}, \tag{21}$$

where $\vec{k}$ denotes the bin index and N_d the number of bins.

Alternatively, histogram matching can be defined by normalized cross correlation:

$$\mathcal{D}_x(\mathcal{H}^{\mathcal{Q}}, \mathcal{H}^{I_i}) = \frac{\sum_{\vec{k}=1}^{N_d} \mathcal{H}^{I_i}(\vec{k})\mathcal{H}^{\mathcal{Q}}(\vec{k})}{\sum_{\vec{k}=1}^{N_d} (\mathcal{H}^{I_i}(\vec{k}))^2}, \tag{22}$$

The property of histogram intersection, as opposed to histogram normalized cross correlation, is that the measure is robust to a large degree of object occlusion and cluttering [5], [13].

V. Image Retrieval

Our approach is to segment target images into regions prior to the actual retrieval process for two reasons: 1. To reduce the number of false matches; and 2. To cope with object occlusion and cluttering. Therefore, in Section V-A, the region-based segmentation method is presented. In Section V-B, the image retrieval process is described.

A. Segmentation into Texture Regions

In this paper, we focus on quadtree-based split-and-merge segmentation [6]. The steps of the texture segmentation algorithm are the following.

Initialization: The input of the segmentation algorithm is the query object image $\mathcal{Q}$ and the target image I_k. Then, from the query and target image, the color ratio gradients are computed. Color ratio gradient distributions are represented by histogram $\mathcal{H}^{\mathcal{Q}}$ derived from query image $\mathcal{Q}$ and histogram $\mathcal{H}^{I_k}$ derived from target image I_k.

Splitting: The second step will split a quadrant of the target image I_k into four quadrants if the quadrant does not satisfy the similarity measure. The similarity measure, during splitting, is given by the histogram intersection $\mathcal{D}_a$ cf. eq. (21). When the histogram intersection is below a given threshold, then the quadrant is considered to contain roughly the same texture as the query. Above the threshold, the region will be split up. The threshold is application dependent. The splitting phase continues until all quadrants satisfy

the similarity measure or until a minimum block-size is reached. A minimum block size is required to ensure enough pixels to enable the derivation of meaningful gradient information. To achieve this, the minimum block size has been fixed to 8 x 8 pixels. This minimum block size has been derived empirically through experimentation. It has been proven to be useful for our application.

Merging: Once the image I_k has been split into squared regions of roughly the same texture as the query image, the merging phase will join regions until a stopping criterion is achieved. Grouping of adjacent regions is necessary because textured objects might not exactly fit within a squared region. We allow adjacent regions to merge if their resulting region is below a threshold for the normalized cross correlation $\mathcal{D}_x$ cf. eq. (22) with respect to the query image. The merging phase starts by computing the start block from which the merging process will begin. To determine the start block, various adjacent blocks are considered. The neighboring blocks with highest similarity with the query image are taken as starting point. By considering adjacent blocks, the method is more robust to noise. After selecting the start block, adjacent regions are merged until all adjacent regions of the growing block violate the similarity expressed by the normalized cross correlation $\mathcal{D}_x$. Again this threshold is application dependent.

Region Growing: Because there are many small regions on the borders of texture elements, the region growing method allows large texture regions to grow simultaneously, meeting each other somewhere in the middle of it. Hence, the method allows large regions to group simultaneously with their neighbor regions, if their resulting regions satisfy the texture measure. This process continues until there will be no more neighbor regions that can be grouped. If no large regions can grow any further, the region growing process stops.

Output: A labeled image is constructed by assigning a label to each pixel according to the label of the region where the pixel belongs to. Each region has an unique label j, $1 \leq j \leq$ (*number of final regions*).

B. Color Constant Texture Retrieval

As stated above, we focus on image retrieval by image example, where an example query image is given by the user on input. Further, we consider the application of retrieving images containing a specific textured object. Then, the query is specified by an example image taken from the object at hand.

As the image segmentation algorithm is computationally expensive, we propose a two-step retrieval process. First, the images in the image database are filtered to obtain proper candidate images. Then, these candidate images are taken as input to the image retrieval algorithm. Hence, only a limited number of images within the image database are considered for blob-based retrieval as follows.

During the *filtering* stage, images are pre-selected to get to the most similar images with respect to the query image on the basis of *global* histogram intersection. Global histogram intersection is, to a large degree, robust to object cluttering and occlusion [5]. To refine the retrieval process, only highly ranked candidates are used during the region-based retrieval process.

During the *refinement* stage, color ratio gradient histograms are created for the query image and each candidate image. The query image, together with the most similar candidate images, are used as input to the segmentation algorithm. After image segmentation, the region having the highest similarity with the query object is selected to represent each candidate image. Hence, for the purpose of image retrieval, only a single region is used for matching. Then, the normalized cross correlation $\mathcal{D}_x$ is taken as the similarity function during query-region matching. After matching, images are shown to the user in descending order of resemblance.

VI. Experiments

In this section, experiments are conducted on color images. In Section VI-A, the segmentation scheme is evaluated with respect to its robustness against varying imaging conditions. In Section VI-B, we compare the retrieval accuracy of the proposed retrieval scheme with respect to the global histogram matching.

A. Texture Segmentation

This section describes the experiments that have been performed on the region-based segmentation algorithm. This section is organized as follows. In Section VI-A.1, we study the accuracy of the texture segmentation algorithm on RGB color images with respect to shading and varying illumination intensity. In Section VI-A.2, we simulate a change in illumination color by multiplying each RGB-color band by an independent scalar.

A.1 *RGB* Texture Images: Color Normalization

In Figure 1.b, a natural image of an outdoor scene is shown. The image comes from the MIT VisTex © image database. The image has 512x512 spatial resolution with 8 bits for each color channel. The image shows a coast border of grass near the sea. As one can see, the texture elements of the image are nonuniformly distributed. Also, boundaries between adjacent textures are not well-defined but are changing smoothly from one texture to another. Furthermore, the image is contaminated by a substantial amount of shadows, shading, and changes in color intensity fields.

To evaluate the robustness of the texture segmentation method with respect to the varying imaging and viewing conditions, the query image, shown in Figure 1.a, is composed of a part of the (textured) grass area. In Figure 1.c, segmentation result is shown obtained by the texture segmentation technique based on $\nabla\mathcal{C}_{RGB}$. Clearly, the segmentation result is negatively affected by shadows and shading. In contrast, the segmentation results based on $\nabla\mathcal{C}_{\theta_1\theta_2}$, shown in Figure 1.d, and color ratio gradient $\nabla\mathcal{C}_M$, shown in Figure 1.e, are insensitive for shadows, illumination and surface orientation changes. Computed region borders correspond properly to the grass-sea border.

In Figure 2.b, a natural image of an indoor scene is shown. The image is taken by a digital camrecorder. The image has 380x282 spatial resolution with 8 bits for each color channel. The image shows a bookcase against a wall composed of bricks. The texture elements of the wall are nonuniform with a periodic component. Again, the image is contaminated by a substantial amount of shadows, shading, and changes in color intensity fields. Note that three parts of the wall are present due to occlusion of the wall by the bookcase. The query image, shown in Figure 2.a, consists of a patch of a brick from the same wall. Note that the query image is a different recording then the original image. Hence, illumination conditions are different. In Figure 2.c, segmentation result is shown obtained by the region based segmentation technique based on $\nabla\mathcal{C}_{RGB}$. The texture segmentation result is only slightly affected. This is because the wall is flat and does not cause much shading. However, only one of the three wall parts have been detected. In contrast, segmentation results for polar color ratio gradient $\nabla\mathcal{C}_{\theta_1\theta_2}$, shown in Figure 2.d, and color constant ratio gradient $\nabla\mathcal{C}_M$, shown in Figure 2.e, are robust to the imaging conditions.

A.2 *RGB* Texture Images: Color Constancy

In this section, we simulate a change in illumination color for the color images. It is known that that the variation in SPD of the illumination can be approximated by the coefficient rule or diagonal model, where the change in the illumination color corresponds to the multiplication of each RGB-color band by an independent scalar. The diagonal model of illumination change holds exactly in the case of narrow-band sensors. To evaluate the sensitivity with respect to a change in the color of the illumination, the R, G and B-images of the query image are multiplied by a factor 1.2, 1 and 0.8 respectively, see Figure 3.a and 4.a.

The segmentation results based on $\nabla\mathcal{C}_{RGB}$ is shown in Figure 3.c and 4.c. Further, the segmentation results based on $\nabla\mathcal{C}_{\theta_1\theta_2}$ is shown in Figure 3.d and 4.d. From the results, we can observe that color ratio gradient $\nabla\mathcal{C}_{\theta_1\theta_2}$, which achieved best segmentation results under white illumination, is highly sensitive to a change in illumination color. As expected, only $\nabla\mathcal{C}_M$ is insensitive to a change in illumination color, see Figure 3.e and 4.e.

B. Image Retrieval

In this section, the proposed retrieval process is evaluated with respect to global histogram matching. The dataset consists of 600 target images taken by a digital color camera (Casio). Images have sizes 256x256 with 8 bits per color. Images were taken at different places in the city of Amsterdam. Images were taken with different viewpoints and illumination conditions, see Figure 5. Note that the color of the sun changes by day and month due to latitude and atmospheric conditions.

In this section, we focus on detecting and localizing information signs. This is an important application in our society (e.g. traffic, surveillance, robotics, license plate recognition). Moreover, the application has all the difficult problems such as cluttered street scenes where the illuminant color can vary greatly, where the pose and position of the signs are unknown, and where the signs can be partially obscured or in shadow. To this end, an independent set (the query set) of query or test recordings was made of each traffic sign already in the database which were recorded under a new, arbitrary position and orientation with respect to the camera, some rotated, some at different distances.

For a measure of match quality, we compute the recall versus precision curve as follows. Let $|R|$ be

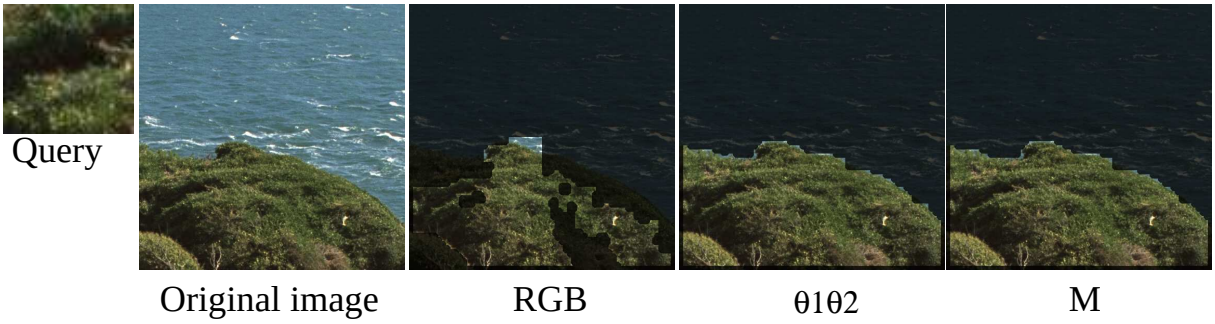


Fig. 1. *a. Query texture b. Target image c. Segmentation result based on $\nabla \mathcal{C}_{RGB}$. d. Segmentation result based on $\nabla \mathcal{C}_{\theta_1\theta_2}$. e. Segmentation result based on color ratio gradient $\nabla \mathcal{C}_M$.*

Fig. 2. *a. Query texture b. Target image c. Segmentation result based on $\nabla \mathcal{C}_{RGB}$. d. Segmentation result based on $\nabla \mathcal{C}_{\theta_1\theta_2}$. e. Segmentation result based on color ratio gradient $\nabla \mathcal{C}_M$.*

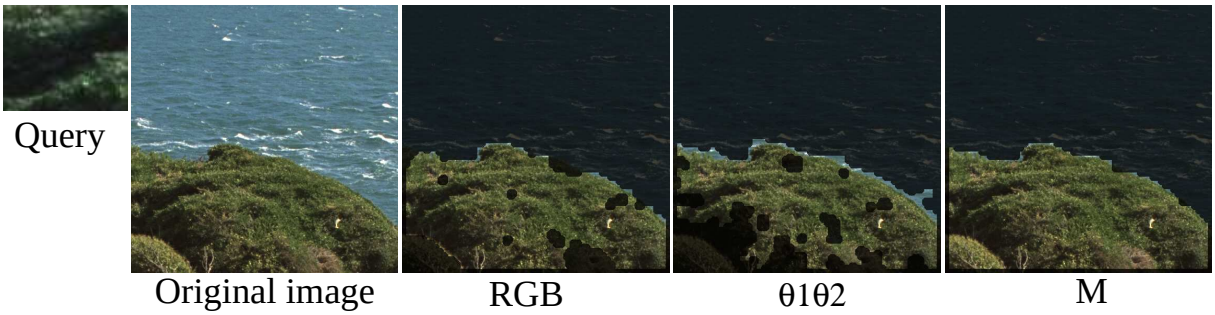


Fig. 3. *a. Query texture under different illumination b. Target image c. Segmentation result based on $\nabla \mathcal{C}_{RGB}$. d. Segmentation result based on $\nabla \mathcal{C}_{\theta_1\theta_2}$. e. Segmentation result based on color ratio gradient $\nabla \mathcal{C}_M$.*

Fig. 4. *a. Query texture under different illumination b. Target image c. Segmentation result based on $\nabla \mathcal{C}_{RGB}$. d. Segmentation result based on $\nabla \mathcal{C}_{\theta_1\theta_2}$. e. Segmentation result based on color ratio gradient $\nabla \mathcal{C}_M$.*

the number of relevant images i.e. images containing the specific textured object one is looking for. Let $|A|$ denote the answer set i.e. the number of images shown to the user. Let $|Ra|$ be the number of images in the intersection of the sets R and A. Then, recall is the fraction of the relevant images (R) which has

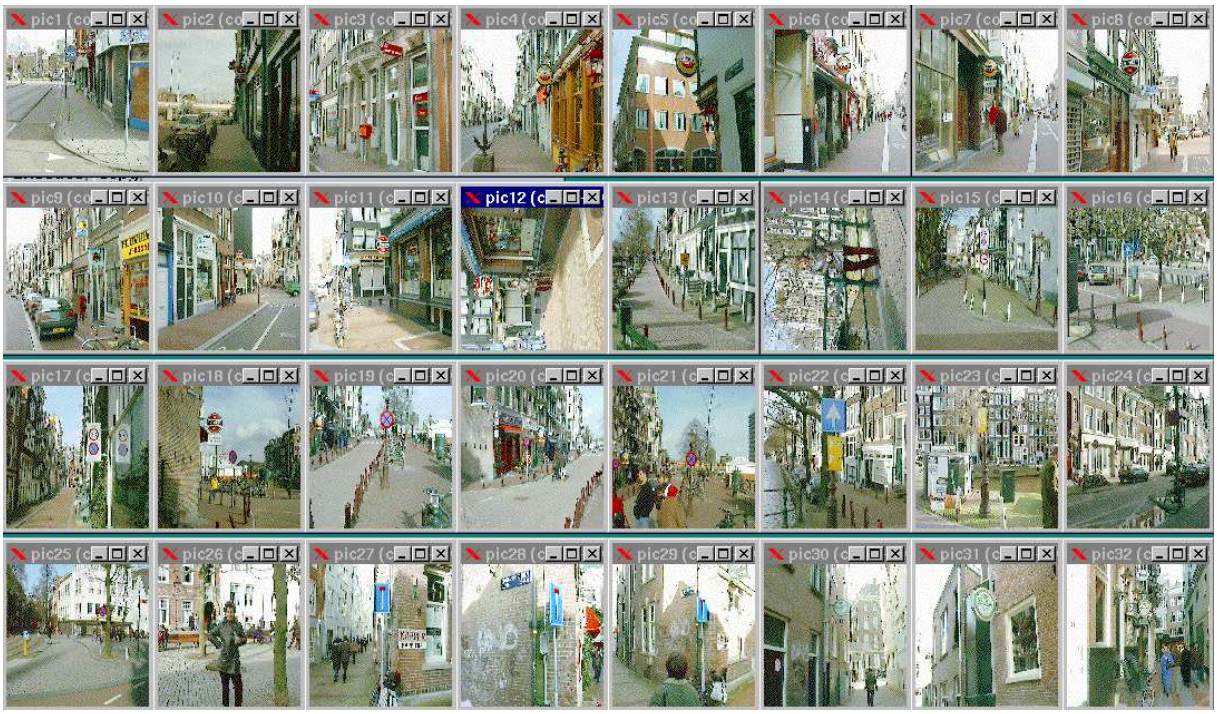


Fig. 5. *Subset of 32 images representative for the dataset consisting of 600 images taken by a digital color camera (Casio). Images were taken at different places in a city with different viewpoints and illumination conditions*

been retrieved:

$$\text{Recall} = \frac{|RA|}{|R|}, \tag{23}$$

and precision is the fraction of retrieved images (A) which is relevant:

$$\text{Precision} = \frac{|Ra|}{|A|}. \tag{24}$$

Retrieval results for searching specific traffic signs are shown in Figure 6 in descending order of resemblance based on comparing distributions taken from the polar color coordinates $\theta_1\theta_2$. The number of relevant traffic signs is 8 ($|R| = 8$). Further, the precision versus recall curve is shown graphically in Figure 7, computed on the basis of $\theta_1\theta_2$ for global histogram intersection and for the proposed region-based retrieval process. It is shown that region-based matching approach achieves high retrieval accuracy where all 8 traffic signs are within the first 15 rankings. Further, global based matching provides that the 8 traffic sign are within the first 82 images. The proposed image retrieval process clearly outperforms the global based image retrieval scheme.

A different experiment has been performed for searching warehouse signs based on $\theta_1\theta_2$ (visual results not shown here). The number of relevant images is 5 i.e. $|R| = 5$. The proposed image retrieval achieves satisfactory retrieval accuracy where 5 traffic signs are within the first 16 rankings. Further, precision versus recall graphs are shown graphically in Figure 8, for global histogram intersection and for the proposed retrieval scheme. It is now shown that the proposed image retrieval process only slightly outperforms the global based image retrieval scheme. This is due to the fact that the warehouse sign consists of large portion of the image. In conclusion, for large objects occupying a large portion of the image, the

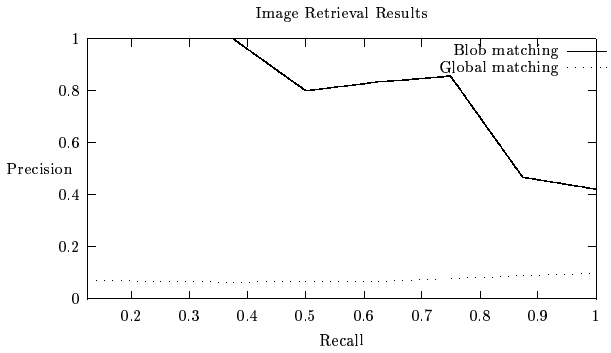


Fig. 7. *Precision versus recall graphs. Global vs local based image retrieval.*

Fig. 6. *Results of finding images containing a specific traffic sign. All 5 traffic signs are within the first 8 highest rankings.*

proposed method is similar to global-based matching.

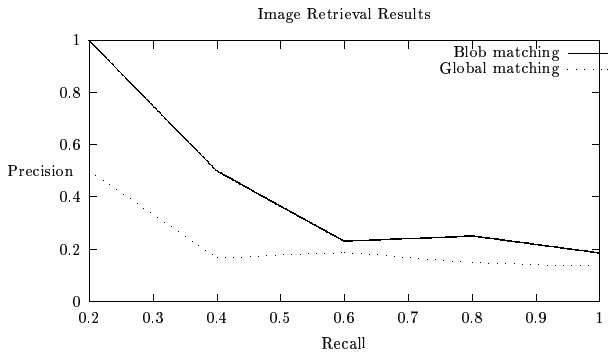


Fig. 8. *Precision versus recall graphs. Global vs local based image retrieval.*

VII. Conclusion

From the theoretical and experimental results, it is concluded that color constant texture matching in image libraries provides high retrieval accuracy and is robust to varying illumination and viewing conditions. Further, it is concluded that the proposed retrieval method outperform global-based retrieval. However, when objects to be searched for consists of a large portion of the image, the proposed retrieval method achieves similar performance as global-based retrieval.

Acknowledgments

The author is grateful to Arnold Smeulders for advise and suggestions. Peter Vreman is acknowledged for partly implementing the algorithm and for the discussions we had.

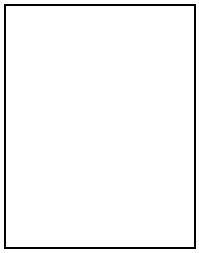
References

[1] S. Belongie, C. Carson, H. Greenspan, and J. Malik, Color- and Texture-based Image Segmentation using EM and Its Application to Content-based Image Retrieval, Proc. Int'l Conf. Computer Vision, pp. 675-682, 1998.

[2] Proceedings of Proc. IEEE Workshop on Content-based Access and Video Libraries, CVPR, 2001.

[3] Y. Deng and B.S. Manjunath, Unsupervised Segmentation of Color-Texture Regions in Images and Video, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 23, no. 8, pp. 800-810, 2001.

[4] G.D. Finlayson, M.S. Drew and B.V. Funt, Spectral Sharpening: Sensor Transformation for Improved Color Constancy, JOSA, 11, pp. 1553-1563, 1994.
[5] T. Gevers and Arnold W.M. Smeulders, PicToSeek: Combining Color and Shape Invariant Features for Image Retrieval, IEEE Trans. on Image Processing, vol. 9, no. 1, pp. 102-119, 2000.
[6] S.L. Horowitz and T. Pavlidis, Picture Segmentation by a Directed Split-and-Merge Procedure, In Proc. 2nd Int'l Conf. on Pattern Recognition. pp. 424-433, 1974.
[7] M. Mirmehdi and M. Petrou, Segmentation of Color Textures, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 22, no. 2, pp. 142-159, 2000.
[8] D.K. Panjwani and G. Healey, Markov Random Field Models for Unsupervised Segmentation of Textured Color Images, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 17, no. 10, pp. 939-954, 1995.
[9] S. K. Nayar, and R. M. Bolle, Reflectance Based Object Recognition, Int'l Journal of Computer Vision, Vol. 17, No. 3, pp. 219-240, 1996.
[10] T. Ojala, M. Pietikainen and D. Harwood, A Comparison Study of Texture Measures with Classification based on Feature Distributions, Pattern Recognition, 29, pp. 51-59, 1996.
[11] M. Pietikainen, S. Nieminen, E. Marszalec and T. Ojala, Accurate Color Discrimination with Classification based on Feature Distributions, Proc. Int'l Conf. Pattern Recognition, Vol. 3, pp. 833-838, 1996.
[12] S.A. Shafer, Using Color to Separate Reflection Components, COLOR Res. Appl., 10(4), pp 210-218, 1985.
[13] M. J. Swain and D. H. Ballard, Color Indexing, Int'l Journal of Computer Vision, vol. 7, no. 1, 11-32, 1991.
[14] Proceedings of Visual99, The Third International Conference on Visual Information Systems, Amsterdam, The Netherlands, 1999.
[15] J.R. Taylor, An Introduction to Error Analysis, University Science Books, 1982.
[16] S.C. Zhu and A. Yuille, Region Competition: Unifying Snakes, Region Growing, and Bayes/MDL for Multiband Image Segmentation, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 18, no. 9, pp. 884-900, 1996.



Theo Gevers is an assistant professor of Computer Science at the University of Amsterdam, The Netherlands. His main research interests are in the fundamentals of image database system design, image retrieval by content, theoretical foundation of geometric and photometric invariants and color image processing. Http://www.science.uva.nl/research/isis/