This paper was presented at the *2003 Asilomar Conference on Signals, Systems, and Computers*, (Pacific Grove, CA), November 10-12, 2003, in the special session on "Perceptual Models in Image and Video Processing."
It will appear in the *Proc. 2004 Asilomar Conference on Signals, Systems, and Computers.*

Perceptual Tuning of Low-level Color and Texture Features for Image Segmentation*

Junqing Chen, Thrasyvoulos N. Pappas
Department Electrical & Comp. Engineering,
Northwestern University, Evanston, IL 60208

Aleksandra Mojsilovic, Bernice E. Rogowitz
IBM T. J. Watson Research Center,
Yorktown Heights, NY 10598

Abstract

We perform subjective tests to determine the key parameters of low-level texture and color features for a previously proposed image segmentation algorithm. The parameters include thresholds for texture classification and feature similarity, as well as the window size for texture estimation. The subjective tests use small isolated patches of textures that correspond to homogeneous texture and color distributions. The goal is to determine what information such small image patches convey to human observers, and to relate those to image statistics. We show that this perceptual tuning of the segmentation algorithm leads to significant performance improvements.

1 Introduction

The focus of this paper is on segmentation of images of natural scenes based on color and texture. One of the challenging aspects of image segmentation is the extraction of perceptually relevant information. An important step towards accomplishing this goal is the development of low-level image features and segmentation techniques that are based on perceptual models and principles about the processing of color and texture information. Although significant effort has been devoted to understanding perceptual issues in image analysis(*e.g.,* [?, ?, ?]), relatively little work has been done in applying perceptual principles to complex scene segmentation(*e.g.,* [?]).

Segmentation of images of natural scenes is particularly difficult because, unlike artificial images that are composed of more or less pure textures, the texture characteristics are not uniform due to effects of lighting, perspective, scale changes, etc. To account for such characteristics, we have proposed a new approach for color-texture segmentation [?, ?, ?] that is based on spatially adaptive color and texture features. In addition, the proposed approach incorporates perceptual knowledge in the feature extraction techniques and the design of the segmentation algorithm. In this paper, we design and conduct subjective tests for determining the key parameters of the algorithms, including thresholds for texture classification and feature similarity, as well as the window size for texture estimation. The main purpose is to link statistics of textures that are found in natural images with human perception of such textures. We show that this perceptual tuning of the segmentation approach leads to significant improvements in performance.

The paper is organized as follows. We first give a brief overview of the segmentation algorithm. The subjective tests are then discussed, followed by analysis of the results. We then compare algorithm performance before and after the perceptual tuning.

2 Segmentation Algorithm Overview

In this section, we overview the segmentation algorithm presented in [?, ?]. The algorithm is based on spatially adaptive color and texture features. As illustrated in Fig. 1, two types of features are developed, one describes the local color composition, and the other the spatial characteristics of the grayscale component of the texture. These features are first developed independently, and then combined to obtain an overall segmentation.

The color features describe the color composition in terms of the dominant colors and associated percentages in the vicinity of each pixel. They are based on the estimation of the spatially adaptive dominant colors, which on one hand, reflects the fact that the human visual system cannot simultaneously perceive a large number of colors, and on the other, the fact that image colors are spatially varying. The spatially adaptive dominant colors are obtained using the adap-

*This work was supported by the National Science Foundation (NSF) under Grant No.CCR-0209006. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

tive clustering algorithm (ACA) for segmentation [?]. The color feature representation is as follows:

$$f_c(x, y, N_{x,y}) = \{(c_i(x, y, N_{x,y}), p_i(x, y, N_{x,y})), \quad i = 1, \ldots, M, p_i(x, y, N_{x,y}) \in [0, 1]\} \quad (1)$$

where each of the dominant colors, $c_i(x, y, N_{x,y})$, is a three dimensional vector in *Lab* space and $p_i(x, y, N_{x,y})$ is the corresponding percentage. $N_{x,y}$ denotes the neighborhood around the pixel at location (x, y) and M is the total number of colors in the neighborhood. A typical value is $M = 4$. Finally, a perceptual metric (OCCD) [?] is used to determine the similarity of two color feature vectors.

The spatial texture features describe the spatial characteristics of the grayscale component of the texture, and are based on a multiscale frequency decomposition such as the steerable pyramid [?] or the Gabor transform [?]. We use the local median energy of the subband coefficients as a simple but effective characterization of spatial texture. Median operators tend to respond to texture within uniform regions and suppress responses associated with transitions between regions. The texture features consist of a classification of each pixel into one of the following categories: *smooth, horizontal, vertical, +45°, -45°*, and *complex.*

Let $s_0(x, y)$, $s_1(x, y)$, $s_2(x, y)$, and $s_3(x, y)$ represent the subband coefficient at location (x, y) that corresponds to the horizontal, diagonal with positive slope, vertical, and diagonal with negative slope directions, respectively. We will use $s_{\max}(x, y)$ to denote the maximum absolute value of the four coefficients, and $s_i(x, y)$ to denote the subband index that corresponds to that maximum. A pixel (x, y) is classified as smooth if the median of $s_{\max}(x', y')$ over a neighborhood of (x, y) is below a threshold T_0. In [?] this threshold is determined using a 2 level K-means over the image. In the next session, we will see how this threshold can be determined by subjective tests. If the pixel is nonsmooth, then it is further classified as follows. We compute the percentage for each value (orientation) of the index $s_i(x', y')$ in the neighborhood of (x, y). If the maximum of the percentages is higher than a threshold T_1 (*e.g.,* 36%) and the difference between the first and second maxima is greater than a threshold T_2, (*e.g.,* 15%), then there is a dominant orientation in the window and the pixel is classified accordingly. Otherwise, the pixel is classified as complex. Note that the first threshold ensures the existence of a dominant orientation and the second ensures its uniqueness. Again, these thresholds can be determined by subjective tests. In [?] we showed that, while the proposed approach depends on the structure of the frequency decomposition, it is relatively independent of the detailed filter characteristics.

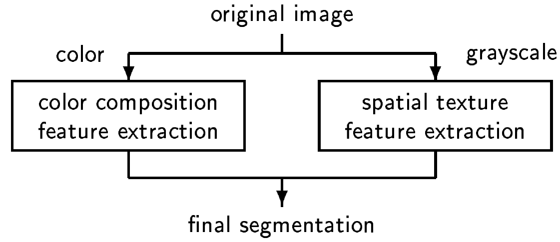

Fig. 1: Schematic of segmentation algorithm

The segmentation algorithm then combines the color and spatial texture features to obtain segments of uniform texture within two steps. The first relies on a multigrid region growing algorithm to obtain a crude segmentation. The segmentation is crude due to the fact that the estimation of the spatial and color texture features requires a finite window. The second uses an elaborate border refinement procedure, which extends the idea of the ACA [?] to color texture, and progressively relies on the color segmentation to obtain accurate and precise border localization.

3 Subjective Experiments

Several key parameters of the segmentation algorithm described in the previous section can be determined by subjective tests. The first such parameter is the threshold T_0 for the smooth/nonsmooth classification. Instead of using the K-means algorithm (which relies solely on individual image statistics) to determine this threshold, we base this decision on a combination of texture statistics and how humans perceive textures. Two additional parameters (T_1 and T_2) are necessary in order to determine whether there is a dominant orientation. Another critical parameter is the window size that is used for the determination of the texture features. In order to allow accurate border localization and adaptation to local texture characteristics, it is important to keep this parameter as small as possible. On the other hand, the window size should be big in order to obtain accurate estimates of the texture characteristics. Thus, it is necessary to select the smallest window size that captures the texture characteristics at a given scale. This can also be obtained through subjective tests. Finally, another important parameter is the threshold for the color composition feature similarity.

The subjective experiments we describe below isolate small patches of images corresponding to homogeneous texture and color distributions. It is important that such patches are considered out of context, just

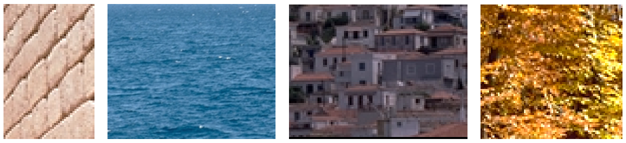

Fig. 2: Examples of Color Texture Patterns used

as the algorithms do not make use of any context information. Our goal is to determine what information such small texture patches convey to human observers, and to relate those to the image statistics.

Experimental setup

The setup for our subjective experiments has been implemented in JAVA and has been published on the web.[1] The subjects were people with normal or corrected vision and normal color vision. The viewing distance was about two feet from the computer display. The subjects were advised not to move their head too close to the display. There were no other restrictions on viewing conditions. Throughout the experiment, the images were displayed against a neutral gray background. However, the subjects were allowed to adjust the background color for the best/clearest view of the color textures. The stimuli were 37 uniform color texture segments of images from a photo CD, and were available at four or five scales. Some examples are shown in Fig. 2. At the time of the writing of this paper, 20 subjects had participated in the study. Their ages ranged from 22 to 50 and included both experts and non-experts in image processing. We used a square shaped window throughout our experiments. We now describe the subjective experiments in more detail.

Experiment 1: Texture classification

In this experiment, the subjects were asked to classify a color texture pattern into one of the following three categories:

- *S*mooth: Images with uniform or slowly varying intensity that contain no objects or sharp boundaries.
- *T*exture: Images of approximately uniform texture patterns. Since natural textures are often statistically nonuniform, slowly varying texture patterns should be included in this category.
- *O*ther: Neither smooth nor texture, such as images with multiple objects or regions.

The subjects were also asked to further classify the "texture" images into one of the following categories based on the perceived dominant orientations: *horizontal, vertical, +45°, -45°*, and *complex.*

The size of texture window is an important parameter of this experiment. The window must be large enough for a human observer to perceive any texture. On the other hand, it must be kept small in order to avoid significant changes in the spatially varying texture characteristics. The window size that we used for the subjective experiments was 23×23 pixels. In addition, the windows should not contain any region boundaries.

Experiment 1a: Minimum window size

Humans perceive texture at different scales. At each scale a minimal window size is required in order to identify a texture. This is true for both human perception and computer-based texture recognition. In Experiment 1, we used a fixed window size for all scales. At that window size, several texture scales can be perceived. However, by displaying several texture scales, we can find the minimum scale that can be perceived at that window size. Conversely, since the minimum window size at which a texture can be perceived is inversely proportional to the scale, this experiment can be used to determine the minimum window size. However, at the writing of this paper we did not have enough data to make a reliable determination of the minimum window size.

Experiment 2: Texture similarity

The goal of this experiment is to establish a threshold for the similarity of the color composition texture features. In the test, two color texture segments were displayed side by side. The subjects were asked to provide a similarity score for the displayed texture patterns. The options were: *same texture, very similar, similar, somewhat similar,* and *totally different.* No definition of similarity was given. The test included segments from the same texture and segments that the subject classified into the same category in Experiment 1. It is highly unlikely, of course, that textures belonging to different categories will be classified as anything but "totally different." This was critical in reducing the length of the test.

4 Analysis of Experimental Results

We now analyze the results of the experiments and use them to tune the segmentation algorithm.

Smooth vs. nonsmooth classification

The smooth/nonsmooth classification is based on models of natural image statistics, which we extract

[1] http://peacock.ece.utk.edu/FeatureTest/

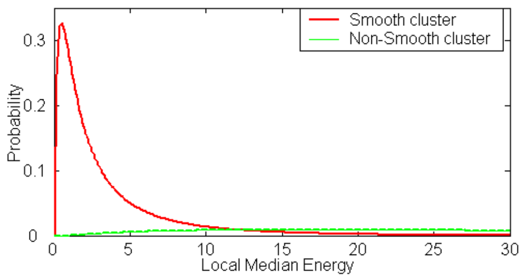

Fig. 3: Distribution of smooth and nonsmooth classes

from data obtained in Experiment 1. As the subject judgments varied, the image category was determined using a "majority wins" rule, *i.e.,* the category that receives more than half of the votes is chosen. Note that the majority can be defined in a stricter sense (higher than 50%). Once the texture category was determined, we analyzed the image segments in the smooth and nonsmooth categories. We obtained the steerable pyramid decomposition of each image and calculated the median energy of $s_{\max}$ over the image. We then collected the median energy values for each of the textures that were classified as smooth and nonsmooth, and tried different distributions in order to find the best fit. We found that the Log Normal model is the best in terms of accuracy and simplicity. Both the Kolmogorov-Smirnov test and the Chi-square test indicated that the difference between the empirical and theoretical cumulative distributions is not significant at the significance level of $\alpha = 0.05$. The models we obtained for the smooth and nonsmooth classes are $LogN(0.73, 1.20)$ and $LogN(4.27, 1.23)$, respectively, where the first parameter denotes the mean and the second the standard deviation of the distribution. Figure 3 shows the fitted Log Normal distribution for the smooth and nonsmooth classes. When the two classes are equiprobable, the threshold below which a pixel is classified as smooth is 12.11, which is the point where the two models intersect. The threshold is a function of the means and standard deviations of the two distributions and the probability of occurrence of each class.

The smooth/nonsmooth determination can now be based on the threshold provided by the above subjective experiment. As we saw above, this threshold also depends on the probability of occurrence of each class. This probability could be determined for each image using the following iterative scheme. First, an initial classification is obtained assuming equal probability for the smooth and nonsmooth classes. The probability of the smooth class is then recalculated based on current classification and the threshold updated.

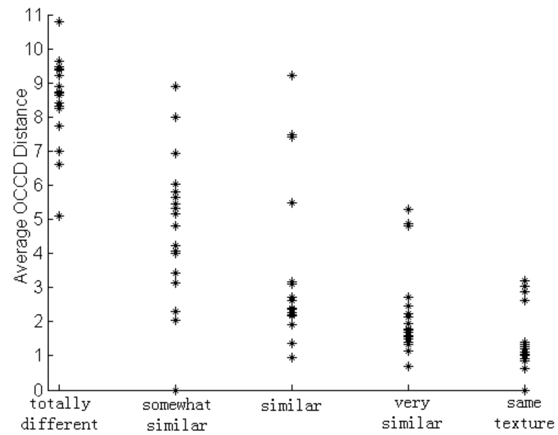

Fig. 4: Scatter plot of average OCCD distance for all subjects.

The thresholding and probability updating procedure is then repeated until convergence. Note that there is no need for a cluster validation test, in contrast to the K-means classification procedure used in [?, ?, ?]. We also avoid other biases that can arise in the K-means clustering procedure, for example, when one cluster is much stronger than the other. Overall, based on experimentation with hundreds of images, the use of models of natural images statistics, provides a more accurate and robust classification.

Texture orientation determination

As we saw Section 2, the determination of the texture orientation of the nonsmooth regions is based on two threshold T_1 and T_2. To obtain these thresholds, we collected all the images that were classified as having one dominant orientation in Experiment 1. We then calculated the histogram of maximal indices over the image, and computed the values of T_1 and T_2. We then found the smallest value of T_1 and T_2 over all subjects and all images, and used those as the thresholds. The values we obtained based on the available data were $T_1 = 42\%$ and $T_2 = 10\%$.

Color feature similarity threshold

To obtain the color feature similarity threshold, we calculated the average OCCD color feature distance of all image pairs in each similarity category, for each subject. In Fig. 4, each star represents the average OCCD color feature distance over all image pairs classified into the similarity category by one subject. Post-test interviews with the subjects indicate that they used the "similar" category for image pairs about whose similarity they were most uncertain.

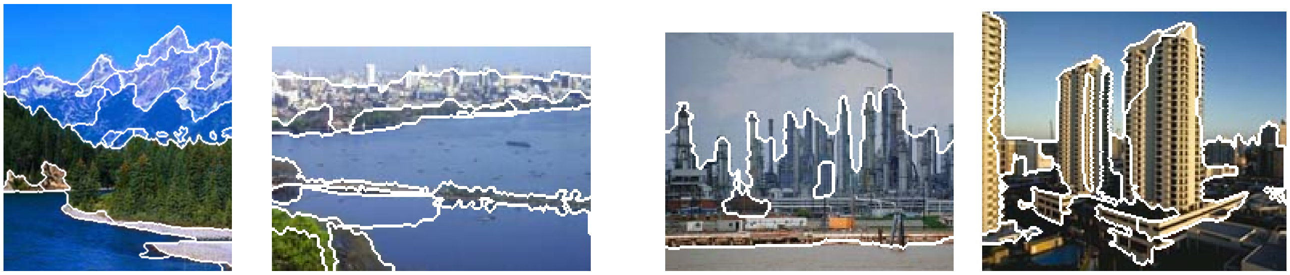

Fig. 5: Image Segmentation without perceptual tuning.

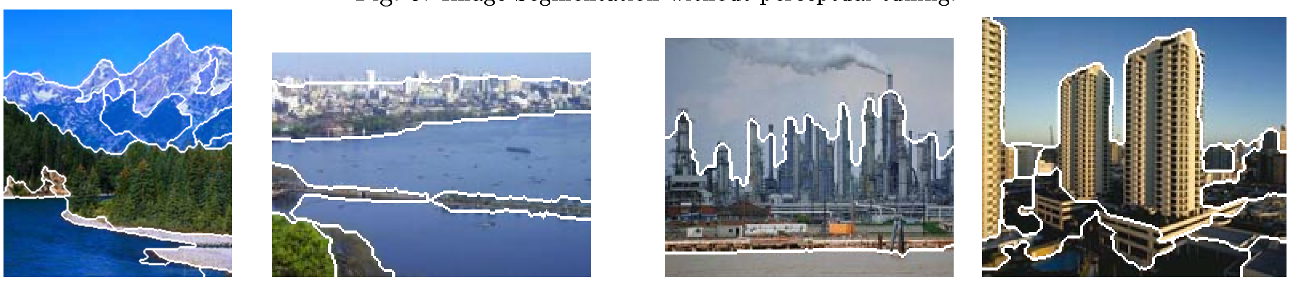

Fig. 6: Image Segmentation with perceptual tuning.

Thus, we combined the data from the "same texture" and "very similar" categories into one group and the data from the "totally different" and "somewhat similar" categories into another. We then fitted distributions to the two groups using procedures similar to those that were used in the smooth/nonsmooth classification. The fitted distributions were $LogN(0.486, 0.243)$ and $N(6.65, 6.53)$ for the two clusters, where $N(\mu, \sigma)$ represents the Normal distribution with mean μ and standard deviation σ. Assuming that the two clusters are equally likely, the threshold then becomes 2.78.

5 Perceptually Tuned Segmentation

Based on the experimental results of the previous sections, we can now obtain a perceptually tuned version of the color-texture segmentation algorithm. The segmentation results before and after tuning are compared in Figs. 5 and 6, respectively. As expected, perceptual tuning results in considerable improvement in image segmentation. As we collect data from more subjects, more accurate statistical models can be obtained, which in turn can lead to further improvements in performance.