Proc. Int. Conf. Image Processing (ICIP), vol. 2, (Atlanta, GA), pp. 2921 - 2924, Octt. 2006.

PERCEPTUAL FEATURE SELECTION FOR SEMANTIC IMAGE CLASSIFICATION

Dejan Depalov, Thrasyvoulos N. Pappas

EECS Department, Northwestern University
2145 Sheridan Rd., Evanston, IL 60208
{depalov,pappas}@ece.northwestern.edu

Dongge Li, Bhavan Gandhi

Motorola Labs, Motorola
1303 E. Algonquin Rd., Schaumburg, IL 60196
{dongge.li,bhavan.gandhi}@motorola.com

ABSTRACT

Content-based image retrieval has become an indispensable tool for managing the rapidly growing collections of digital images. The goal is to organize the contents semantically, according to meaningful categories. In recent papers we introduced a new approach for semantic image classification that relies on the adaptive perceptual color-texture segmentation algorithm proposed by Chen et al. This algorithm combines knowledge of human perception and signal characteristics to segment natural scenes into perceptually uniform regions. The resulting segments can be classified into semantic categories using region-wide features as medium level descriptors. Such descriptors are the key to bridging the gap between low-level image primitives and high-level image semantics. The segment classification is based on linear discriminant analysis techniques. In this paper, we examine the classification performance (precision and recall rates) when different sets of region-wide features are used. These include different color composition features, spatial texture, and segment location. We demonstrate the effectiveness of the proposed techniques on a database that includes 9000 segments from approximately 2500 photographs of natural scenes.

Index Terms— Content-based image retrieval, semantic analysis, segment classification, feature extraction, adaptive perceptual color texture segmentation.

1. INTRODUCTION

The field of content-based image retrieval (CBIR) has been quite active in recent years, as the number and size of digital image repositories have been rapidly growing. The primary emphasis has been on query by example techniques, which attempt to match low-level image features, such as color and texture, with or without relevance feedback by the user. A comprehensive review can be found in [1]. A more challenging problem is to first assign semantic labels to each image, so that retrieval can be based on semantic information. However, it has been difficult to infer high-level semantics from low-level features. This is known as the *semantic gap.*

Several approaches have been proposed recently that attempt to bridge the semantic gap. Most of them incorporate an image segmentation scheme, and then use the segment features and their content within the image to derive semantic information [2–6].

In an attempt to obtain a CBIR system that better approximates the performance of the human visual system (HVS), we have proposed a novel approach for image indexing that utilizes perceptual models for image segmentation and classification [7, 8]. It relies on the adaptive perceptual color-texture segmentation algorithm proposed by Chen et al. [9]. This algorithm combines knowledge of human perception and signal characteristics to segment natural scenes into perceptually uniform regions. The resulting segments can be classified into semantic categories using region-wide features as medium level descriptors. Such descriptors are the key to bridging the gap between low-level image primitives and high-level image semantics. In [7, 8], we presented linear discriminant analysis techniques for assigning labels to image segments based on color composition and spatial texture descriptors. However, several questions regarding the necessity and effectiveness of each feature remain unanswered. In this paper, we examine the classification performance (precision and recall rates) when different sets of region-wide features are used. These include different color composition features, spatial texture, and segment location. We show that all three types of features are necessary. We also show that more precise information about the first and second most dominant color is more important than more accurate information about the color composition at the expense of coarser color quantization. We demonstrate the effectiveness of the proposed techniques on a database that includes 9000 segments (obtained using the algorithm in [9]) from approximately 2500 photographs of natural scenes.

While further improvements can be achieved by incorporating the segment size and boundary shape, as well as the properties of the neighboring segments, the results of this papers demonstrate that color composition and spatial texture alone can achieve quite impressive results. We also show that as we add more features, such as segment location, the results keep further improving.

The focus of this paper is on still images. The techniques we discuss, however, can also form the basis for content-based analysis of video sequences. We consider the domain of photographic images with a wide range of content (indoor and outdoor natural and man-made scenes).

2. COLOR-TEXTURE FEATURE SELECTION

In this section, we briefly review the perceptual color-texture features that form the basis of both image segmentation [9] and segment classification. There are two types of spatially adaptive features. The first provides a localized description of the color composition of the texture and the second models the spatial characteristics of its grayscale component.

The color composition feature consists of a small number of spatially adaptive dominant colors and the corresponding percent occurrence of each color in the vicinity of a pixel:

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

This work was supported by the National Science Foundation (NSF) under Grant No.CCR-0209006 and the Defense Intelligence Agency (DIA) under DIA/NSF Grant No.IIS-0515929. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF or DIA. This work was also supported by the Motorola Center for Telecommunications at Northwestern University.

(a) (b) (c) (d)

Fig. 1. Color-texture image features and segmentation. (a) original color image (b) adaptive dominant colors (c) texture classes (smooth regions are shown in black, horizontal in gray, and complex in white) (d) final segmentation

where c_i is a 3-D color vector and p_i is the corresponding percentage. $N_{x,y}$ denotes the neighborhood of the pixel at (x, y) and M is the number of dominant colors in $N_{x,y}$; a typical value is $M = 4$. The spatially adaptive dominant colors are obtained using the adaptive clustering algorithm (ACA) [10]. An example is shown in Figure 1(b).

The spatial texture feature extraction is based on a steerable filter decomposition with four orientation subbands (horizontal, vertical, +45°, -45°). Here, we use a one-level decomposition. The local energy of the subband coefficients provides a simple but effective characterization of spatial texture. At each pixel location, the maximum of the four subband coefficients determines the texture orientation. A median filtering operation boosts the response to texture within uniform regions and suppresses the response resulting from to transitions between regions. The pixels are then classified into smooth, horizontal, vertical, +45°, -45°, and complex (i.e., no dominant orientation) categories. An example is shown in Figure 1(c).

The segmentation algorithm combines the color composition and spatial-texture features to obtain segments of uniform texture. Several critical parameters of the texture features and segmentation algorithm can be determined by subjective tests [11].

3. SEGMENT-WIDE FEATURE EXTRACTION

We now review the development of medium level color and spatial texture descriptors. While image segmentation requires a combination of local and global features [9], region classification requires segment-wide features [7]. Thus, for each segment, the color composition and spatial texture features must be recalculated using only information from within the segment, that is, the local averages and medians are computed across and strictly within the segment. The texture features of the segment can be similarly described by the percentage of smooth, horizontal, vertical, +45°, -45°, and complex pixels . An example is shown in Fig. 2, where (a) shows a segmented image, (b) shows a selected segment, (c) shows the segment-wide color composition (dominant colors and percentages), and (d) shows the region-wide spatial texture features (percentage of smooth, horizontal, vertical, +45°, -45°, and complex pixels).

As we saw in [7], there is an asymmetry between the two types of features. While the spatial texture features consist of six labels and the corresponding percentages, the color composition features consist of up to four dominant colors (each with a continuum of values) and the associated percentages. One approach to reducing the dimensionality of the color composition features, is by quantizing the colors. In [7], we assigned color names to the dominant colors of each region using the procedure proposed in [12]. The syntax contains color names for 267 regions in color space, and is summarized in Table 1. If we assign labels based on hue only, we end up with 14 labels (and corresponding percentages) instead of a continuum of color values, which establishes a symmetry with the spatial texture features. The use of a limited number of colors is consistent with Boynton's study, which found that when people are asked to categorize colors, the number of perceptually distinguishable color categories is small. (See his 1989 paper "Eleven colors which are almost never confused" [13].) The eleven color can be obtained by eliminating the olive, magenta, and beige categories. Further reductions in the number of colors are possible, for example, by eliminating the pink and orange categories.

Prim. Hue	Sec. Hue	Saturation	Lightness	Achromatic
Red	Reddish	Grayish	Blackish	Black
Orange	Brownish	Moderate	Very-dark	Gray
Brown	Yellowish	Medium	Dark	White
Yellow	Greenish	Strong	Medium	
Green	Bluish	Vivid	Light	
Blue	Purplish		Very-light	
Purple	Pinkish		Whitish	
Pink				
Beige				
Magenta				
Olive				

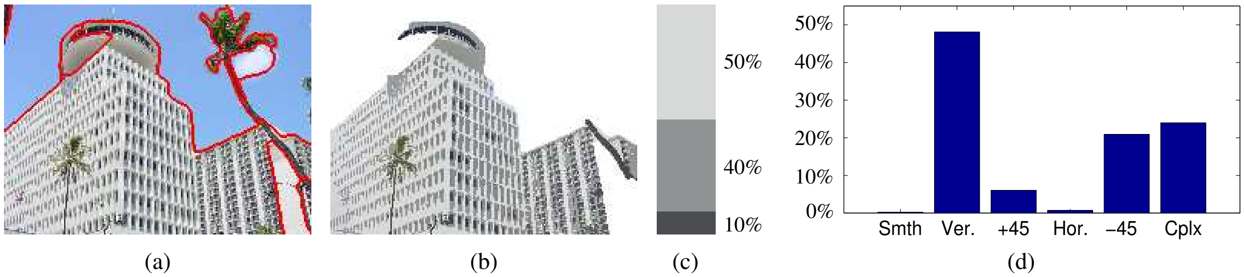
Table 1. Color Naming Syntax

Another approach for dominant color representation is as a coordinate in L*a*b* color space. Our statistical analysis of dominant colors [8] revealed that using the two dominant colors with highest percentage (the first and second dominant colors) is sufficient to describe the segment color composition. In our implementation we use the first dominant color and difference between first and second dominant colors for the total of six features.

4. SEGMENT CLASSIFICATION

We performed several experiments using approximately 2500 photographs. The majority of the images were obtained from the Corel Stock Photo Library. Additional images were obtained from a Key Photos Library and the investigators' personal repository. The images cover a variety of outdoor scenes, with a wide range of themes. The images were segmented using the adaptive perceptual color-texture image segmentation algorithm [9], and the segments were manually labeled to be used as the ground truth in supervised learning. Each segment was assigned exactly one label. Segments with area less than 3% of the total image area were not considered. This resulted in approximately 9000 labeled segments, 80% of which were used for training and the rest for testing.

For the training and classification we used the Linear Discriminant method (LDA) [14]. LDA is the method that belongs to the class of linear classifiers, which try to find a projection to a lower dimensional space such that samples from the different classes are well separated. LDA attempts to find directions that maximize variance among the means of different classes (between class scatter) and at the same time minimize the variance within each class (within class scatter). To achieve this goal, LDA maximizes the following objec-



(a) (b) (c) (d)

Fig. 2. Segment-wide feature extraction. (a) Segmented image. (b) Selected segment. (c) Its color composition. (d) Its texture composition.

tive function:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}} \qquad (2)$$

where S_B is the "between the classes scatter matrix" and S_W is the "within the classes scatter matrix" and are defined as:

$$S_B = \sum_c N_c(\mu_c - \bar{\mathbf{x}})(\mu_c - \bar{\mathbf{x}})^T \qquad (3)$$

$$S_W = \sum_c \sum_{i \in c} (\mathbf{x}_i - \mu_c)(\mathbf{x}_i - \mu_c)^T \qquad (4)$$

where,

$$\mu_c = \frac{1}{N_c} \sum_{i \in c} x_i \qquad \bar{\mathbf{x}} = \frac{1}{N} \sum_c N_c \mu_c \qquad (5)$$

The objective function $J(\mathbf{w})$ is maximized by solving the generalized eigenvalue problem $S_B \mathbf{w} = \lambda S_W \mathbf{w}$ We should also note, that for LDA to work, the data for each class has to form a single cluster. Furthermore, although not a requirement, LDA assumes that the underlying class distribution can be approximated with a Gaussian.

It is reasonable to expect that a particular label may consist of more than one cluster. For example, the label "water" may be represented by blue dominant color and smooth or horizontal texture, thus resulting in two clusters. To deal with such cases, we experimented with applying the **K-means** algorithm to each category in the training set to create additional clusters before applying LDA. Note that misclassification among clusters that belong to the same label is not recorded as a classification error.

5. RESULTS

We evaluated the performance of the proposed techniques using standard measures that are used in the literature. The **recall** is the ratio of the correctly classified segments to the total number of segments with the given label in the database. The **precision** is the ratio of the correctly classified segments to the total number of segments that the algorithm assigned to the particular label (correct and incorrect). Both performance measures are expressed as percentages.

The goal of our experiments is to identify the most suitable set of features for segment classification. In [8], we found that using the texture features presented in Section 2 and the precise color value of the first dominant color outperforms the texture features combined with 14 quantized dominant colors. This is a somewhat unexpected result. Our goal here is to try different sets of features in order to establish their necessity for the classification task.

The recall and precision rates for the several feature sets and techniques are shown in Fig.3. The first two plots show the classification results for LDA based only on color compositions features with eleven and fourteen perceptually quantized colors. Our results indicate that there is no significant difference between the two cases. We also considered reducing the number of dominant colors to nine, and again, found no significant difference in performance. On the other hand, increasing the number of colors beyond 14 by including information about the lightness/saturation or secondary hue, leads to dependencies between the feature vector components (perceptually similar colors) and performance degradation.

The next two plots present the classification results using only spatial texture and spatial texture with eleven perceptually quantized colors, respectively. It can be concluded that spatial texture plays an important role in classification feature by itself or in combination with color, resulting in significant improvement in recall and precision rates The following two plots show that using spatial texture with the unquantized color value of the first dominant color (in L*a*b* color space) or first and second dominant color outperforms the use of all the dominant colors perceptually quantized. Here, the first dominant color is expressed as an L*a*b* coordinate, while second is expressed as a difference. Note that adding information about the second dominant color improves the classification even further.

The last two plots show the effect of adding the segment position to the feature vector. It is expressed as the centroid normalized by the size of the image. As expected, adding position improves classification performance, especially for separating the sky and water categories. Finally, the last plot shows the result obtained by applying the K-means algorithm followed by LDA, in order to add within-class clusters. This yields a modest improvement in precision (6% on the average) while the recall remains the same.

Overall, our segment classification results compare well to the methods described in the literature (e.g., [15–20]). The precision and recall rates are quite good and serve as validation of the selected features. They also indicate that the adaptive perceptual color-texture segmentation algorithm does indeed provide semantic information.

6. REFERENCES

[1] W. M. Smeulders, *et al.*, "Content-based image retrieval at the end of the early years," *IEEE Tr. PAMI*, v. 22, p. 1349–1379, Dec. 2000.

[2] L. Zhu, *et al.*, "Keyblock: An approach for content-based image retrieval," *ACM Multimedia 2000*, p. 157–166, Oct. 2000.

[3] W. Wang, Y. Song, A. Zhang, "Semantics retrieval by content and context of image regions," *Proc. 15th Int. Conf. Vision Interface*, p. 17–24, May 2002.

[4] J. Y. Pan, *et al.*, "Automatic image captioning," *ICME*, 2004.

[5] J. Li, J. Wang, "Automatic linguistic indexing of pictures by a atatistical modeling approach," *IEEE Tr. PAMI*, v. 25, Sept. 2003.

[6] A. Mojsilović, B. E. Rogowitz, "Semantic metric for image library exploration," *IEEE Tr. MM*, v. 6, p. 828–838, Dec. 2004.

[7] D. Depalov, *et al.*, "A perceptual approach for semantic image retrieval," *Proc. ICASSP*, May 2006.

[8] D. Depalov, *et al.*, "Perceptually based techniques for semantic image classification and retrieval," *Human Vision and Electronic Imaging XI*, Proc. SPIE V. 6057, Jan. 2006.

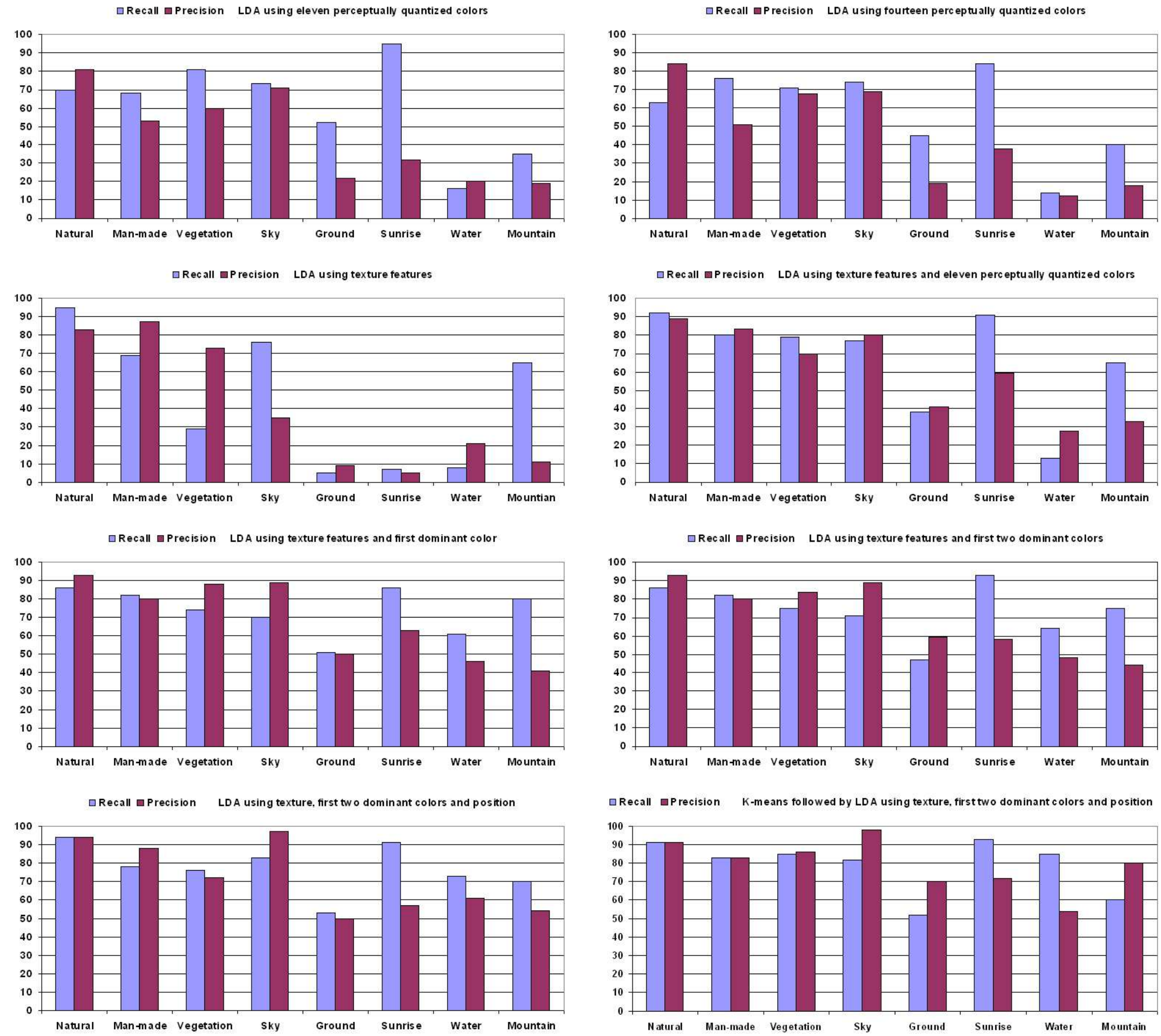


Fig. 3. Classification Results

[9] J. Chen, *et al.*, "Adaptive perceptual color-texture image segmentation," *IEEE Tr. Im. Proc.*, v. 14, p. 1524–1536, Oct. 2005.

[10] T. N. Pappas, "An adaptive clustering algorithm for image segmentation," *IEEE Tr. Sig. Proc.*, v. 40, p. 901–914, Apr. 1992.

[11] J. Chen, T. N. Pappas, "Experimental determination of visual color and texture statistics for image segmentation," *Human Vision and Electronic Imaging X*, Proc. SPIE V. 5666, p. 227–236, Jan. 2005.

[12] A. Mojsilović, "A computational model for color naming and describing color composition of images," *IEEE Tr. Im. Proc.*, v. 14, p. 690–699, May 2005.

[13] R. M. Boynton, "Eleven colors that are almost never confused," *Human Vision, Visual Proc., and Digital Display*, Proc. SPIE V. 1077, p. 322–332, Jan. 1989.

[14] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. Wiley-Interscience, 2nd ed., Oct. 2000.

[15] J. Pan, *et al.*, "Automatic image captioning," *ICME*, June 2004.

[16] K. Barnard, *et al.*, "Matching words and pictures," *J. Machine Learning Res.*, v. 3, p. 1107–1135, 2003.

[17] J. Li, J. Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Tr. PAMI*, v. 25, no. 9, p. 1075–1088, 2003.

[18] J. Jeon, *et al.*, "Automatic image annotation and retrieval using cross-media relevance models," *SIGIR*, p. 119–126, ACM, 2003.

[19] J. Z. Wang, *et al.*, "SIMPLIcity: Semantics-sensitive Integrated Matching for Picture LIbraries," *IEEE Tr. PAMI*, v. 23, p. 947–963, Sept. 2001.

[20] C. Carson, *et al.*, "Blobworld: A system for region-based image indexing and retrieval," *Proc. Third Int. Conf. Visual Info. Systems*, p. 509–516, Springer, June 1999.