

Analysis of Segment Statistics for Semantic Classification of Natural Images

Dejan Depalov and Thrasyvoulos N. Pappas

Electrical Engineering and Computer Science, Northwestern University
2145 Sheridan Road, Evanston, IL, 60208

ABSTRACT

A major challenge facing content-based image retrieval is bridging the gap between low-level image primitives and high-level semantics. We have proposed a new approach for semantic image classification that utilizes the adaptive perceptual color-texture segmentation algorithm by Chen et al., which segments natural scenes into perceptually uniform regions. The color composition and spatial texture features of the regions are used as medium level descriptors, based on which the segments are classified into semantic categories. The segment features consist of spatial texture orientation information and color composition in terms of a limited number of spatially adapted dominant colors. The feature selection and the performance of the classification algorithms are based on the segment statistics. We investigate the dependence of the segment statistics on the segmentation algorithm. For this, we compare the statistics of the segment features obtained using the Chen et al. algorithm to those that correspond to human segmentations, and show that they are remarkably similar. We also show that when human segmentations are used instead of the automatically detected segments, the performance of the semantic classification approach remains approximately the same.

Keywords: Image analysis, retrieval, segmentation, semantic classification, perceptual models, human segmentations

1. INTRODUCTION

A major challenge for content-based image retrieval (CBIR) is bridging the *semantic gap* between low-level image primitives and high-level semantics. Several approaches have been proposed in the recent literature that attempt to bridge this gap. Many of these approaches use an image segmentation scheme as an intermediate step, and then rely on the content of the segmented regions as well as their context within an image to obtain semantic information. However, the segmentation problem is quite challenging, and the unavailability of semantically meaningful segmentations has been cited (e.g., in Ref. 1) as a significant obstacle to the success of CBIR systems, and it is one of the reasons that they are still a long way from matching the performance of the human visual system (HVS). In this paper, we consider a new approach for semantic classification of images of natural scenes^{2,3} that relies on automatic segmentation into perceptually uniform regions. Our goal is to test the sensitivity of the approach to the performance of the segmentation algorithm. We show that when human segmentations are used instead of the automatically detected segments, the performance of the approach remains approximately the same. We also show that the segment statistics that result from the automatic and human segmentations are remarkably similar.

As we argued above, the choice of the segmentation technique is critical to the success of the CBIR system. To avoid the problems with image segmentation, several approaches partition the image into fixed-size rectangular blocks, and use the features (typically color and texture) of these blocks for semantic analysis. Such approaches range from codebook-based approaches^{4,5} to probabilistic approaches,⁶ to high-dimensional feature extraction and discriminative classifier learning.⁷ Others rely on more sophisticated segmentation techniques. Carson *et al.*⁸ model the distribution of pixel features with a mixture of Gaussians and use an expectation-maximization algorithm to group the pixels into regions. Barnard *et al.*⁹ rely on the normalized cut algorithm¹⁰ to segment the image into blobs. Li and Wang¹¹ segment images based on color and texture features using a K-means algorithm. Finally, some authors attempt to link low-level image features directly to image semantics (e.g., in Ref. 12).

In Ref. 2 we proposed a new approach for semantic image classification that utilizes the adaptive perceptual color-texture segmentation algorithm by Chen *et al.*,¹³ which segments natural scenes into perceptually uniform regions. The

Further author information: Send correspondence to T. N. Pappas at email: pappas@eecs.northwestern.edu



Figure 1. Color-texture image features and segmentation. (a) Original color image. (b) Adaptive dominant colors. (c) Texture classes (smooth regions shown in black, horizontal in gray, and complex in white). (d) Final segmentation

color composition and spatial texture features of the regions are used as medium level descriptors, based on which the segments are classified into semantic categories. The segment features consist of spatial texture orientation information and color composition in terms of a small number of spatially adapted dominant colors. The feature selection and the performance of the classification algorithms are based on the segment statistics.

In this paper, we investigate the dependence of the segment statistics on the segmentation algorithm. For this, we extract and compare the statistics of segment features obtained using the Chen *et al.* algorithm¹³ to those that correspond to human segmentations, and show that there is a remarkable similarity in the two sets of segment statistics. We also show that the algorithm performance does not change significantly when human segmentations are used instead of automatic segmentations. This indicates that perhaps there is not much to be gained from further improvements in the segmentation algorithm, and also, that it is not critical to closely match the semantic classification technique to the technique used for image segmentation. In our experiments, we used a database of approximately 3300 images of natural scenes, which were segmented using the algorithm in Ref. 13. We also used a database of 400 images, obtained from the University of California at Berkeley (UCB),¹⁴ segmented by 30 human subjects.

The paper is organized as follows. In Section 2, we review the feature selection for both the adaptive perceptual color-texture segmentation algorithm and the semantic classification algorithms. In Section 3, we review the basic elements of the semantic analysis approach. The statistics of automatic and human segmentations are presented in 3, while in Section 5, we compare classification performance based on automatic and human segmentations.

2. COLOR-TEXTURE FEATURE SELECTION

The selection of appropriate color-texture features is critical for both image segmentation and segment classification. The segmentation approach proposed in Ref. 13 is based on two types of spatially adaptive texture features, color composition and spatial texture.

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:

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

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).¹⁵ An example is shown in Fig. 1(b). The perceptual similarity between two color composition feature vectors is based on the ‘‘Optimal Color Composition Distance (OCCD)’’.¹⁶

The spatial texture feature extraction is based on a steerable filter decomposition with four orientation subbands (horizontal, vertical, $+45^\circ$, -45°).¹⁷ In this paper, we use a one-level steerable filter decomposition. The local energy of the subband coefficients is used, first to classify the image pixels into smooth and non-smooth classes, and then to further classify the non-smooth pixels on the basis of dominant orientation, as horizontal, vertical, $+45^\circ$, -45° , and complex (i.e., no dominant orientation). A median filtering operation boosts the response to texture within uniform regions and suppresses the response resulting from transitions between regions. An example is shown in Fig. 1(c).

The segmentation algorithm combines the color composition and spatial texture features to obtain segments of uniform texture. It is a fairly elaborate algorithm that relies on spatial texture to determine the major structural composition of the

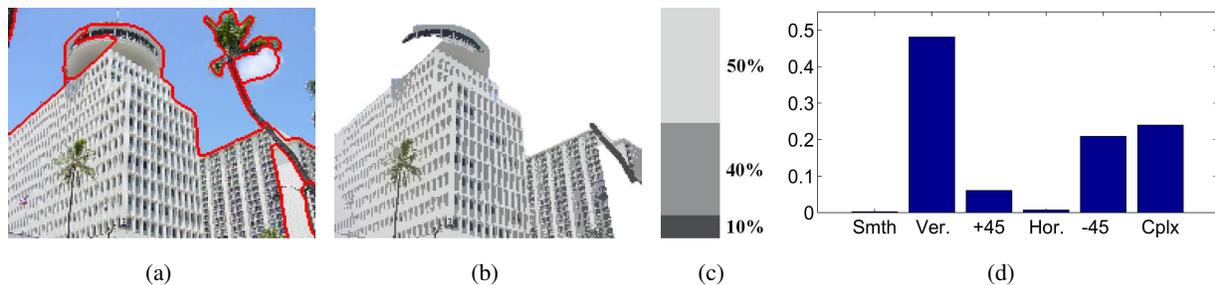


Figure 2. Segment-wide feature extraction. (a) Segmented image. (b) Selected segment. (c) Its color composition. (d) Its texture composition.

image and combines it with color, first to obtain a crude estimate of the major segments, and then to refine it in order to obtain accurate and precise localization of the border between regions. Several critical parameters of the texture features and segmentation algorithm can be determined by subjective tests.¹⁸

2.1. Segment Wide Feature Extraction

We now review the selection of segment-wide color composition and spatial texture features. These are derived from the features we described above, once the segments have been obtained. They are calculated on a segment by segment basis, using only information from within the segment. An example is shown in Fig.2, where Fig.2(a) shows a segmented image, Fig.2(b) shows a selected segment, and Fig.2(c) shows the color composition of the segment (dominant colors and percentages). The texture features of the segment can be similarly described by the percentage of smooth, horizontal, vertical, +45°, -45°, and complex pixels as shown in Fig.2(d)

The spatial texture features are represented as a six-dimensional vector that consists of the percentages for each texture category. For the color composition features, we have considered two possibilities.

The first¹⁹ uses a perceptually quantized color space, whereby a relatively small number of perceptually distinct representative colors (color names) are selected (e.g., Boynton’s “eleven colors which are almost never confused”²⁰), and the feature vector consists of the percentage of each color for the segment. One justification for this approach is that humans use a few color names to describe the color composition of a scene. Note that this representation provides a symmetry between the two types of features,¹⁹ in that they both consist of percentages that correspond to a fixed set of categories. The procedure we use to assign color names to the dominant colors of a region is described in Ref. 21. The color names (labels) are consistent with a National Bureau of Standards recommendation for color names. The syntax, shown in Table 1, contains names for 267 regions in color space, and employs English terms to describe colors along the three dimensions of the color space: hue, lightness and saturation. Of these, we select a basic set of twelve prototypical hues and three achromatic colors to end up with 15 labels. Thus, we obtain a 15-dimensional color feature vector. Note, however, that the feature vector of each segment will have up to $M = 4$ nonzero components.

The second approach is to use only one or two dominant colors, the ones with highest percentages, and to represent them by their (unquantized) $L^*a^*b^*$ coordinates. We will refer to these as the first and second dominant colors. In our implementation, we used the first dominant color and the difference between the first and second dominant color. In Ref. 2 we showed that this representation results in better classification performance than the color naming approach. To explain this result, we looked at the segment statistics and found that, for the majority of segments, the percentages of the third and fourth dominant colors were insignificant. We also found that, for the majority of segments, the second dominant color is similar to the first.

In Section 4, we compare the statistics that correspond to the automatic segmentations to those that are obtained when human segmentations are used.

Hue-primary	Hue-secondary	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				
Violet				
Magenta				
Olive				

Table 1. Color Naming Syntax

3. SEMANTIC ANALYSIS

3.1. Semantic Labeling

The assignment of semantic labels to segments is based on the subjective experiments described in Refs. 12, 22–24, which were conducted in order to identify important semantic categories that humans use for image organization and retrieval. We have assembled a vocabulary of labels consistent with their findings, and also, those used in annotation of the NIST TRECVID 2003 development set.²⁵ The set of labels we selected is a subset of NIST lexicon, and is shown in Table 2. The labels are arranged in a hierarchical manner, with natural, man-made, human, and animal at the top. In our classification algorithms, we use only leaf nodes; the higher order categories are then derived from the leaves.

The main focus of our experiments was on natural vs. man-made classification. The human detection problem, and especially face detection, is well-studied in the literature,²⁶ and the existing techniques can easily be combined with the proposed approach. The problem of animal detection is more complicated because of natural camouflage. Scenes containing animals could be handled by our current approach as natural scenes without any explicit detection of animals.

Natural				Man-made	Human	Animal
Vegetation	Sky	Landform	Water			
Grass	Day-sky	Snow		Building/House	Face	
Trees/bushes	Night-sky	Mountain/Hill		Bridge	Person	
Forest	Sun	Ground		Car	People	
Flowers	Clouds	Pavement/Road*		Boat		
	Sunrise/Sunset			Airplane		
				Other		

Table 2. Segment Labels

3.2. Learning and Classification

For the training and classification we considered both unsupervised (clustering) and supervised learning techniques. Among unsupervised techniques we experimented with K-means, K-nearest neighbors, agglomerative, and conglomerative clustering methods, while in the supervised learning experiments we used Gaussian Mixture Models (GMM),²⁷ Support Vector Machines (SVM),²⁸ and Linear Discriminant Analysis (LDA).²⁹ It quickly became clear that supervised techniques are

*While “pavement/road” is man-made, its features are almost identical to those for “ground.”

best suited for the problem at hand, primarily because of the complexity of the cluster configurations. Supervised techniques require the existence of ground truth for a large database of segments.

After performing several sets of experiments using supervised learning techniques, and comparison of the recall and precision rates, we concluded that LDA is best suited for the problem at hand because it works better with noisy data and is less sensitive to data redundancies than the other two approaches.³⁰ Furthermore, LDA is not critically dependent on the correct choice of the kernel function like the SVM approach. GMM achieved lower recall and precision rates than LDA due to inherent limitations in the mixture approach and the fact that the EM algorithm is only locally convergent.

We now summarize the basics for the Linear Discriminant Analysis (LDA) method.²⁹ LDA belongs to the class linear classifiers, which try to find a subspace projection such that samples from the different classes are well separated, i.e., to find directions in the data space that facilitate data classification. This is done by finding a direction that maximizes the variance between the class means and at the same time minimizes the variance within each class.

The measure of separation between the class means is defined as:

$$S_B = \sum_c N_c (\mu_c - \bar{\mathbf{x}})(\mu_c - \bar{\mathbf{x}})^T \quad (2)$$

where,

$$\mu_c = \frac{1}{N_c} \sum_{i \in c} x_i \quad (3)$$

and

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_c N_c \mu_c \quad (4)$$

The measure of within class variance is defined as:

$$S_W = \sum_c \sum_{i \in c} (\mathbf{x}_i - \mu_c)(\mathbf{x}_i - \mu_c)^T \quad (5)$$

S_B is usually referred to as the “between the classes scatter matrix,” while S_W is known as the “within the classes scatter matrix.”

Using these measures, the objective function can be represented as:

$$J(\mathbf{w}) = \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}} \quad (6)$$

where \mathbf{w} is a transformation matrix. The objective function $J(\mathbf{w})$ is maximized by solving the generalized eigenvalue problem:

$$S_B \mathbf{w} = \lambda S_W \mathbf{w} \quad (7)$$

so that the columns of an optimal \mathbf{w} are the eigenvectors associated with the largest eigenvalues. One of the underlying assumptions that are necessary for LDA to work is that the data for each class form a single cluster.

4. SEGMENT STATISTICS FOR NATURAL IMAGES

In this section we compare the segment statistics that are obtained from automatic and human segmentations.

Automatic segmentations were performed on a image database containing approximately 3300 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 in the database cover a variety of outdoor scenes, with a wide range of themes. The images were segmented using the algorithm in Ref. 13 described above. Figure 3 shows several images segmented using this algorithm. The resulting segments were manually labeled to be used as the ground truth for collecting statistics and supervised learning. Each segment was assigned exactly one label. Segments whose area was less than three percent of total image area were not considered.

We also used a database of human segmentation from the University of California at Berkeley (UCB).¹⁴ The database contained approximately 400 images, with approximately 1600 labeled segments. Each image was segmented by an



Figure 3. Adaptive perceptual color-texture segmentation results.

average of seven subjects. There was a total of 30 subjects. For reasons we explained in Section 3.1, we did not consider segments containing humans or animals. As with the automatic segmentations, the segments were manually labeled; each segment was assigned exactly one label.

Figure 4 shows two images from the UCB database segmented by different humans. Observe that there are substantial differences in the segmentations. Martin *et al.*¹⁴ have shown that, if we allow for mutual refinement, the segmentations by different humans are consistent. What is of more interest to us is whether the segment statistics for the human segmentations are significantly different to those obtained from the human segmentations.

The statistics for automatic segmentations using the Chen *et al.* algorithm¹³ are shown in the top row of Fig. 5, which shows histograms for the first, second, third, and fourth dominant color across all segments in the database. The horizontal axis represents the percentage of the area that the corresponding dominant color occupies in a segment, while the vertical axis represents the probability of occurrence for each bin. Based on these statistics, the great majority of segments could be described by the first two dominant colors with very little loss of information. We also looked at the distance between the dominant colors. The histogram of the distance between the first and second dominant colors is shown at the top left of Fig. 6, and the distance between the first and third is shown at the right. Observe, in the majority of cases, the second dominant color is less than twenty units away from the first. Here we should note that colors with $L^*a^*b^*$ distance of 10 units are quite similar. This means that for a great majority of segments the second dominant color is similar to the first. This explains why using the $L^*a^*b^*$ value of the first dominant color gives better classification results than using the 15

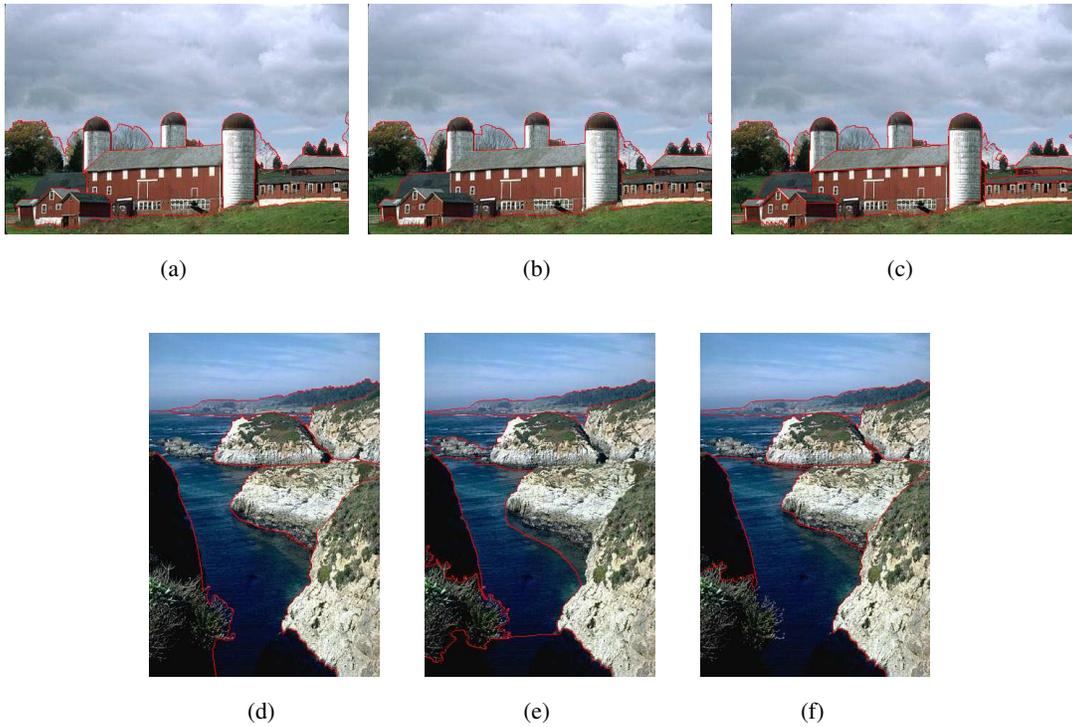


Figure 4. Human Segmentations

quantized colors, and also why adding the second dominant color improves performance by only a small amount.

The corresponding statistics for the human segmentations are shown in Figures 5 and 6. Note that, with the exception of the 100% bin for the first dominant color, the statistics are quite similar. A similar observation holds for the distribution of the distance between the first and second dominant colors and the distance between the first and third dominant colors. Thus, we can safely assume that ignoring the second and third dominant color will not have any significant effect on segment classification. Finally, we should point out that the two sets of statistics are quite similar, even though they were obtained over different sets of images.

5. RESULTS

In addition to the segment statistics, it is interesting to investigate whether the performance of the classification algorithm can improve if human segmentations are used instead of the automatic ones, or if it degrades due to the fact that the features we are using are matched to the segmentation algorithm and do not work with the human segments. For this, we compared the classification performance of the region-wide spatial texture and color composition features described in Section 2 on both human and automatic segmentations. We applied LDA to the same labeled sets of segmented images that we used for the segment statistics in Section 4. In both cases we used 80% of the segments for training and the rest for testing.

We evaluated the performance of the classification techniques using the standard measures that are used for evaluating search strategies in the literature. The **recall** is the ratio of the correctly labeled segments to the total number of relevant segments in the database (i.e., those with the particular label). The **precision** is the ratio of the correctly labeled segments to the total number of segments that the algorithm assigned to the particular label (both correct and incorrect). Both performance measures are expressed as percentages. Overall performance can be expressed as the accuracy over the whole database.

The results are shown in Fig. 7(a) for segment classification using the spatial texture features, the first two dominant colors and position, and Fig. 7(b) using the same set of features and the K-means preprocessing.³¹ We should note that

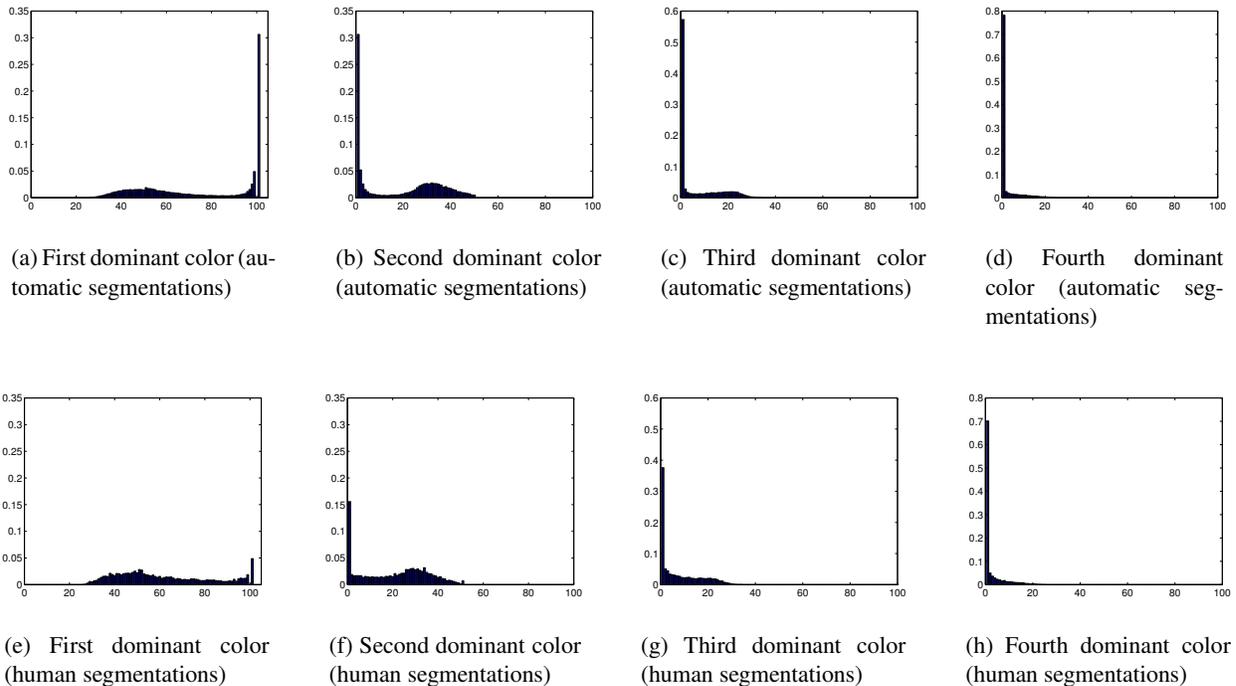


Figure 5. Statistics of dominant colors. The horizontal axis represents the percentage of the area that the dominant color occupies in a segment and the vertical axis represents the probability of occurrence for each bin.

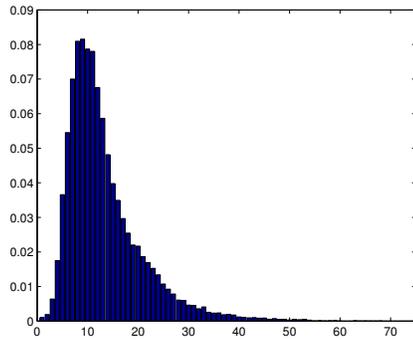
“Sunrise/Sunset” category was omitted from the human segmentations results because of insufficient number of samples for classification. Comparing the recall and precision rates from the two experiments we can see that on the average they are approximately the same. We also conducted experiments with different sets of features (only one most dominant color, 15 quantized colors, etc.) and got the same conclusion. Finally, we also verified that using a third dominant color does not increase performance, and that using a fourth dominant color actually reduces the classification ability of LDA.

6. CONCLUSIONS

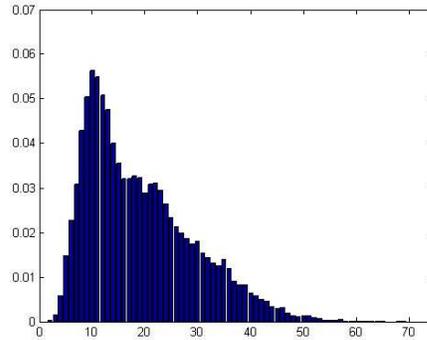
We considered a new approach for semantic classification of images of natural scenes that relies on automatic segmentation into perceptually uniform regions, and examined its sensitivity to the performance of the segmentation algorithm. We showed that when human segmentations are used instead of the automatically detected segments, the performance of the approach remains approximately the same. We also showed that the segment statistics that result from the automatic and human segmentations are remarkably similar. These results indicate that there is not much to be gained from further improvements in the segmentation algorithm, and also, that it is not critical to closely match the semantic classification technique to the technique used for image segmentation.

7. ACKNOWLEDGMENTS

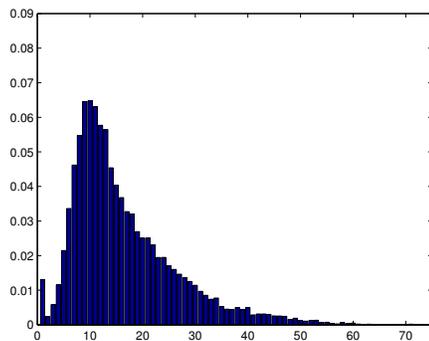
This work was supported by the National Science Foundation (NSF) under Grant No.CCR-0209006 and the Defense Intelligence Agency (DIA) under DIA/NSF Grant No.IIS-0515929. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily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



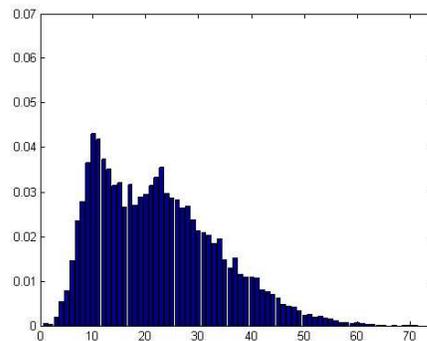
(a) Distance between first and second dominant color (automatic segmentations)



(b) Distance between first and third dominant color (automatic segmentations)



(c) Distance between first and second dominant color (human segmentations)

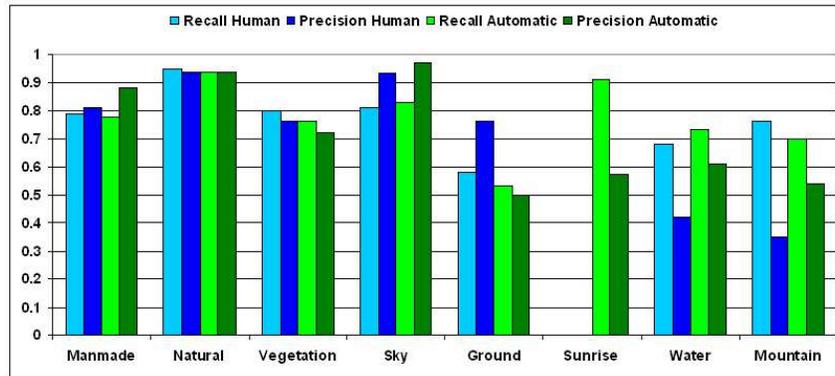


(d) Distance between first and third dominant color (human segmentations)

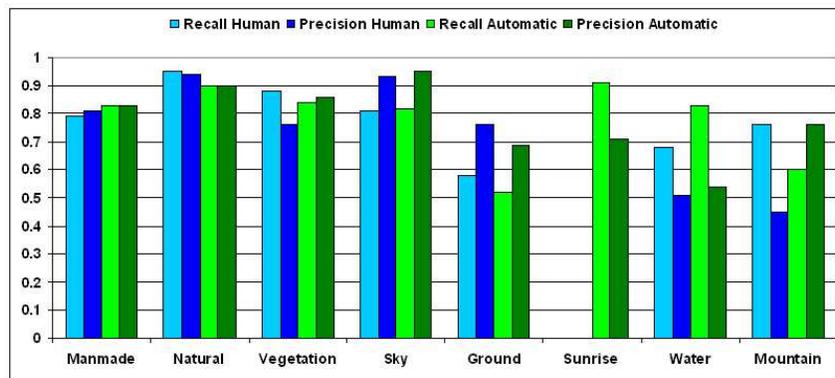
Figure 6. Histograms of distances between dominant colors in $L^*a^*b^*$ color space

REFERENCES

1. D. A. Forsyth and J. Ponce, *Computer Vision - A Modern Approach*, Prentice-Hall, 2002.
2. D. Depalov, T. N. Pappas, D. Li, and B. Gandhi, "Perceptually based techniques for semantic image classification and retrieval," in *Human Vision and Electronic Imaging XI*, B. E. Rogowitz, T. N. Pappas, and S. J. Daly, eds., **Proc. SPIE Vol. 6057**, pp. 6057OZ-1-6057OZ-10, (San Jose, CA), Jan. 2006.
3. T. N. Pappas, J. Chen, and D. Depalov, "Perceptually based techniques for image segmentation and semantic classification," *IEEE Commun. Mag.* **45**, pp. 44-51, Jan. 2007.
4. L. Zhu, A. Zhang, A. Rao, and R. Srihari, "Keyblock: An approach for content-based image retrieval," in *Proc. ACM Multimedia*, pp. 157-166, (Los Angeles, CA), Oct. 2000.
5. W. Wang, Y. Song, and A. Zhang, "Semantics retrieval by content and context of image regions," in *Proc. 15th Int. Conf. on Vision Interface*, pp. 17-24, (Calgary, Canada), May 2002.
6. S. L. Feng, R. Manmatha, and V. Lavrenko, "Multiple bernoulli relevance models for image and video annotation," in *Conference on Computer Vision and Pattern Recognition, (CVPR 2004)*, **2**, pp. 1002-1009, June 2004.
7. D. H. W. S. Gao and C. H. Lee, "Automatic image annotation through multi-topic text categorization," in *Int. Conference on Acoustics, Speech and Signal Processing, (ICASSP 2006)*, **2**, pp. 377-380, (Toulouse, France), May 2006.



(a) LDA using spatial texture features, the first two dominant colors, and position



(b) K-means preprocessed LDA using spatial texture features, the first two dominant colors, and position

Figure 7. Comparison of classification results using human and automatic segmentations

8. C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image segmentation using expectation-maximizations and its application to image querying," *IEEE Trans. Pattern Anal. Machine Intell.* **24**, pp. 1026–1038, Aug. 2002.
9. K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. Blei, and M. I. Jordan, "Matching words and pictures," *Journal of Machine Learning Research* **3**, pp. 1107–1135, 2003.
10. J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.* **22**, pp. 888–905, Aug. 2000.
11. J. Li and J. Z. Wang, "Automatic linguistic indexing of pictures by a statistical modeling approach," *IEEE Trans. Pattern Anal. Machine Intell.* **25**, pp. 1075–1088, Sept. 2003.
12. A. Mojsilović and B. E. Rogowitz, "Semantic metric for image library exploration," *IEEE Trans. Multimedia* **6**, pp. 828–838, Dec. 2004.
13. J. Chen, T. N. Pappas, A. Mojsilovic, and B. E. Rogowitz, "Adaptive perceptual color-texture image segmentation," *IEEE Trans. Image Processing* **14**, pp. 1524–1536, Oct. 2005.
14. D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. Int. Conf. Computer Vision (ICCV)*, **2**, pp. 416–423, (Vancouver, Canada), July 2001.

15. T. N. Pappas, "An adaptive clustering algorithm for image segmentation," *IEEE Trans. Signal Processing* **SP-40**, pp. 901–914, Apr. 1992.
16. A. Mojsilović, J. Hu, and E. Soljanin, "Extraction of perceptually important colors and similarity measurement for image matching, retrieval, and analysis," *IEEE Trans. Image Processing* **11**, pp. 1238–1248, Nov. 2002.
17. E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proc. ICIP-95, vol. III*, pp. 444–447, (Washington, DC), Oct. 1995.
18. J. Chen and T. N. Pappas, "Experimental determination of visual color and texture statistics for image segmentation," in *Human Vision and Electronic Imaging X*, B. E. Rogowitz, T. N. Pappas, and S. J. Daly, eds., **Proc. SPIE Vol. 5666**, pp. 227–236, (San Jose, CA), Jan. 2005.
19. D. Depalov, T. N. Pappas, D. Li, and B. Gandhi, "A perceptual approach for semantic image retrieval," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP-06)*, **II**, pp. 417–420, (Toulouse, France), May 2006.
20. R. M. Boynton, "Eleven colors that are almost never confused," in *Human Vision, Visual Proc., and Digital Display*, B. E. Rogowitz, ed., **Proc. SPIE Vol. 1077**, pp. 322–332, (Los Angeles, CA), Jan. 18–20 1989.
21. A. Mojsilović, "A computational model for color naming and describing color composition of images," *IEEE Trans. Image Processing* **14**, pp. 690–699, May 2005.
22. B. E. Rogowitz, T. Frese, J. R. Smith, C. A. Bouman, and E. Kalin, "Perceptual image similarity experiments," in *Human Vision and Electronic Imaging III*, B. E. Rogowitz and T. N. Pappas, eds., **Proc. SPIE, Vol. 3299**, pp. 576–590, (San Jose, CA), Jan. 1998.
23. A. Mojsilovic and B. Rogowitz, "A psychophysical approach to modeling image semantics," in *Human Vision and Electronic Imaging VI*, B. E. Rogowitz and T. N. Pappas, eds., **Proc. SPIE Vol. 4299**, pp. 470–477, (San Jose, CA), Jan. 2001.
24. A. Mojsilovic and B. Rogowitz, "Capturing image semantics with low-level descriptors," in *Proc. Int. Conf. Image Processing (ICIP-01)*, pp. 18–21, (Thessaloniki, Greece), Oct. 2001.
25. C. Y. Lin, B. L. Tseng, and J. R. Smith, "Video collaborative annotation forum: Establishing ground-truth labels on large multimedia datasets." NIST TRECVID, 2003.
26. P. Viola and M. Jones, "Robust real-time face detection," *International Journal of Computer Vision* **57(2)**, pp. 137–154, 2004.
27. A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal Royal Statistical Society* **39(1)**, pp. 1–21, 1977.
28. V. N. Vapnik, *Statistical Learning Theory*, Wiley, 1998.
29. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley-Interscience, 2nd ed., Oct. 2000.
30. D. Depalov, T. N. Pappas, D. Li, and B. Gandhi, "Perceptually based techniques for semantic image classification," *IEEE Trans. Image Processing*. Submitted to IEEE Trans. Image Processing.
31. D. Depalov, T. N. Pappas, D. Li, and B. Gandhi, "Perceptual feature selection for semantic image classification," in *Proc. Int. Conf. Image Processing (ICIP-06)*, pp. 2921–2924, (Atlanta, GA), Oct. 2006.