REGION-BASED IMAGE ANNOTATION USING COLOR AND TEXTURE CUES

Eli Saber
Xerox Corporation
435 W. Commercial St.
East Rochester, NY 14445
saber@roch803.mc.xerox.com

A. Murat Tekalp
Dept. of Electrical Engineering
University of Rochester
Rochester, NY 14627
tekalp@ee.rochester.edu

ABSTRACT

We present algorithms for automatic image annotation and retrieval based on pixel-based color, and block- or region-based texture features. Region formation has been accomplished by utilizing Gibbs random fields or morphological based operations. Color, and texture indexing may be knowledge-based (using appropriate training sets) or by example. The algorithms are designed to: i) offer the user a wide range of options and flexibilities in order to enhance the outcome of the search and retrieval operations, and ii) provide a compromise between accuracy and computational complexity.

1 INTRODUCTION

In recent years, we have seen a growing interest in accessing, searching, retrieving, and handling of digital images, with an emphasis on intelligent, content-based processing. The present trend is to develop systems that assist a user in accessing and retrieving images based on some low- or mid-level features, such as color, texture, etc. in an automatic or semi-automatic fashion. Among the systems, found in the literature (see [1, 2] for surveys), are IBM's Query by Image Content (QBIC) [3] which can index and retrieve images based on color, texture, shape, and sketches; MIT's Photobook [4] which employes eigenimages, finite elements, and Wold decomposition to represent appearances, shapes, and textures respectively; Columbia University's Multimedia/VOD testbed system [5] where color, texture, and shape sets can be integrated using logical AND/OR operations; and MIT's FourEyes system [6] which selects the most suitable models for retrieval from a "society of models."

In this paper, we describe two classification algorithms (A and B) that utilize color and texture cues to obtain image descriptive keywords, such as skin, sky, grass, outdoor, etc., in a QBIC fashion. Algorithm A is based on a sequential integration of color and texture information, while Algorithm B encompasses a simultaneous integration in a Bayesian framework using a maximum *a posteriori* probability (MAP) approach. Algorithm B provides generally a more accurate segmentation at the expense of an increase in computational complexity. The final classification, resulting from the use of Algorithm A or B, is employed as a basis for obtaining keyword annotations.

2 COLOR SPACE

The YES color space [7] has been selected as a medium for classification. It is defined as follows:

$$\begin{bmatrix} Y \\ E \\ S \end{bmatrix} = \begin{bmatrix} 0.253 & 0.684 & 0.063 \\ 0.500 & -0.500 & 0.000 \\ 0.250 & 0.250 & -0.500 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

where the Y channel represents the luminance component, and E, S channels represent the chrominance. It has been chosen because: 1) it constitutes a linear transformation from RGB free of singularities, and 2) it provides some computational efficiencies. (The E and S channels can be computed from RGB by shifting bits rather than multiplication). However, it is generally agreed that there does not exist a single color space which is good for all images [8]. The E and S channels provide a suitable space for recognition of the classes under consideration based on color. The Y channel will be mainly utilized to model the texture information.

3 TEXTURE MODEL

Texture is a useful cue for image classification and annotation. It is modeled, based on the luminance (Y) channel, by a set of predefined features. The features employed are the "Haralick" measures [9] computed from co-occurrence matrices. The co-occurrence matrices provide information about the relative frequency of occurrence of pairs of gray levels, separated by a certain distance in a certain direction. Here, we choose five of these features: contrast (CON); angular second momentum (ASM); inverse difference moment (IDM); entropy (ENT); and information measures of correlation (IMC). However, the proposed texture classification algorithm is generic in the sense that it can be utilized with any type and number of features, such as Tamura features, wavelet domain features, etc. [1]. Our choice is influenced by the work of Ohanian *et al.*

[10], who showed that co-occurrence based features performed best in their texture classification experiments.

The feature vector $\boldsymbol{f}_{ij} = [CON_{ij}\ ASM_{ij}\ IDM_{ij}\ ENT_{ij}\ IMC_{ij}]^T$ at pixel (i,j) is computed based on a local $M \times N$ window (typically a 16×16) centered at that pixel, under the assumption that the block contains a single texture class. We model the conditional distribution of features over all blocks belonging to the class by a Gaussian probability density function (pdf), whose mean vector and covariance matrix are computed from appropriate training sets or examples.

4 COLOR AND TEXTURE INTEGRATION ALGORITHMS

It is clear that we are faced with certain limitations when employing color, and texture cues individually. For instance, color alone cannot be used to distinguish between a "blue car" and a "blue sky." Therefore, combination of these cues can serve to improve annotation and retrieval results. Hence, in this section, we introduce classification algorithms that are based on the combined use of color, and texture information.

4.1 Algorithm A: Parallel Integration of Pixel-Based Color and Block-Based Texture

The simplest approach is to combine pixel-based color and block or region-based texture by logical "AND/OR" operations. A block diagram of Algorithm A is shown in Figure 1. The color classification is performed using a pixel-based approach followed by smoothing via morphological operations (three erosions followed by three dilations using a 3×3 kernel) or Gibbs random field based filtering to obtain contiguous label clusters. The advantages of this approach lie in its computational efficiency, and its ability to form smooth regions without any segmentation. The details are described below.

4.1.1 Pixel-based Color

The pixel-based color classification module [11] consists of two steps, which are briefly reviewed below: classification by fixed thresholding, and reassignment by adaptive thresholding. The second step can be bypassed, thereby improving the computational efficiency. In summary, the class conditional pdf of the chrominance components $\boldsymbol{w}_{ij} = [E_{ij}\ S_{ij}]^T$ belonging to each class is modeled by a 2-D Gaussian, where the mean vector and the covariance matrix for each class are estimated from appropriate training sets or examples. Then, a succession of binary hypothesis tests with image-adaptive thresholds are employed to decide whether each pixel in an image belongs to one of the predetermined classes or not. The thresholds can be estimated either at run time from user specified confidence bounds, or pre-computed by using receiver operating characteristic (ROC) analysis on a set of training images. The main advantage of this approach is its computational efficiency since a segmentation is not required prior to color classification. Furthermore,the user may choose to employ morphological operations or GRF smoothing to form contiguous decision regions, as an alternative supervised region formation procedure. The performance of this approach was demonstrated in [11], where skin, sky, and grass were selected as the target classes. These results indicate that color can serve as a powerful initial classifier for image access and retrieval.

4.1.2 Smoothing via Gibbs Random Fields

A classification of the image into N classes is attempted by maximizing the *a posteriori* probability of the class labels given the observed chrominance data. The proposed approach is different from many available MAP color image segmentation methods because we target recognition of specific color cues obtained from a training set of images. Let $\boldsymbol{w}$ denote a specific realization of a random field, where $w_{ij} = [E_{ij}\ S_{ij}]^T$ denote the chrominance vector of a pixel at location (i,j). According to Bayes theorem, a MAP estimate of the class labels $\boldsymbol{x}$ is formed by maximizing the *a posteriori* probability:

$$p(\boldsymbol{x} \mid \boldsymbol{w}) \propto p(\boldsymbol{w} \mid \boldsymbol{x}) p(\boldsymbol{x}) \tag{2}$$

The term $p(\boldsymbol{x})$ represents the *a priori* pdf of the region process. It is modeled by the Gibbs distribution:

$$p(\boldsymbol{x}) = \frac{1}{Z}\ \exp\left\{-\sum_{c \in C} V_c(\boldsymbol{x})\right\} \tag{3}$$

where Z is a normalizing constant, C is the set of all cliques, and V_c is the clique potential for clique c. We will consider the second order neighborhood with pairwise cliques. The clique potentials are defined as:

$$V_c(i,j;k,l) = \begin{cases} -\beta_1 & if\ x_{ij} = x_{kl}\ and\ (i,j),(k,l) \in c \\ +\beta_1 & if\ x_{ij} \neq x_{kl}\ and\ (i,j),(k,l) \in c. \end{cases} \tag{4}$$

where β_1 is chosen as a positive quantity indicating that two neighboring pixels are more likely to belong to the same class than to different classes.

On the other hand, the term $p(\boldsymbol{w} \mid \boldsymbol{x})$ represents the conditional pdf of the observed chrominance vectors of an image given the region labels. It is modeled by a Gaussian pdf using a space-varying image intensity model similar to [12] for color images except that our model does not assume independence among the color channels. As a result, the *a posteriori* pdf (2) becomes

$$p(\boldsymbol{x} \mid \boldsymbol{w}) \propto \exp\{e_1 + e_2\} \tag{5}$$

where:

$$e_1 = -\sum_{(i,j)} \frac{1}{2}[w_{ij} - \boldsymbol{\mu}_{c,ij}^{x_{ij}}]^T\ \left[\boldsymbol{K}_{c,ij}^{x_{ij}}\right]^{-1}\ [w_{ij} - \boldsymbol{\mu}_{c,ij}^{x_{ij}}]$$

$$e_2 = -\sum_{C} V_C(\boldsymbol{x})$$

and $\boldsymbol{\mu}_{c,ij}^{x_{ij}}$, $\boldsymbol{K}_{c,ij}^{x_{ij}}$ denote the chrominance mean vector and covariance matrix, respectively, for pixel (i,j) with respect to the class x_{ij}. The MAP classifications can then be obtained by maximizing Eq. (5) iteratively through the use of iterated conditional modes (ICM), where the sites are visited one by one in a raster scan, and the label that yields the MAP is accepted as the estimate for the site.

4.1.3 Block-based Texture

The block-based texture classification approach is dependent on the features computed from the co-occurence matrices as discussed above. To this effect, the image is first divided into blocks of size $M \times N$. Note that the block size utilized during classification should correspond to the size employed during the training stage. Each block is then classified by using a minimum distance classifier. The threshold utilized during the classification is computed at run time based on some user specified confidence bound, or pre-computed from a representative training set using an ROC curve analysis. The output of the classification corresponds to the blocks whose texture is similar to the one represented by the training set or example. The advantage of the approach lie in its computational efficiency since a region formation process is not required prior to performing the texture analysis.

4.1.4 Region-based Texture

The region-based approach classifies selected regions as a representative of a particular texture class (e.g., in the training set or provided by the example) or not. The method followed here is similar to that of the block-based texture classifier, except: i) a denser set of feature vectors have been computed within each region by allowing a 50 % overlapping of the blocks in each direction, and ii) a majority voting decision is utilized to classify the overall region.

To this effect, each region is divided into several rectangular blocks whose size is chosen to match the size employed in the training set to compute the corresponding statistics for each texture class. Each block is then classified using an appropriate threshold computed at run time or pre-computed based on the training set as in the case of block-based classification. Finally, a decision for the region is made based on majority voting of the results of all blocks. The outcome is a set of regions that possess a texture "similar" to those in the training set or to the example.

4.2 Algorithm B: Bayesian Integration of Color and Texture

Color and texture classifications can also be integrated within a Bayesian framework using a single classification map $\boldsymbol{x}$ to indicate regions with similar color and texture features. This approach determines $\boldsymbol{x}$ by using the MAP criterion rather than simple logical operations; thereby resulting in enhanced classification at the expense of an increase in computational complexity. By assuming conditional independence between the color and texture features, we can express

$$p(\boldsymbol{x} \mid \boldsymbol{w}, \boldsymbol{f}) \propto p(\boldsymbol{w}, \boldsymbol{f} \mid \boldsymbol{x}) p(\boldsymbol{x}) = p(\boldsymbol{w} \mid \boldsymbol{x}) p(\boldsymbol{f} \mid \boldsymbol{x}) p(\boldsymbol{x}) \tag{6}$$

where $\boldsymbol{w} = [\boldsymbol{E}\ \boldsymbol{S}]^T$ denote the vector of chrominance components and
$\boldsymbol{f} = [\boldsymbol{CON}\ \boldsymbol{ASM}\ \boldsymbol{IDM}\ \boldsymbol{ENT}\ \boldsymbol{IMC}]^T$ denote the vector of texture features.

The class-conditional pdf of color and textures features are modeled by Gaussian pdfs for the class of interest. The *a priori* probability is modeled by the Gibbs distribution with a second order neighborhood and pairwise cliques. Assuming conditional independence among the pixels within the image, the *a posteriori* pdf:

$$p(\boldsymbol{x} \mid \boldsymbol{w}, \boldsymbol{f}) \propto \exp\{e_1 + e_2 + e_3\} \tag{7}$$

where:

$$e_1 = -\sum_{(i,j)} \frac{1}{2} [\boldsymbol{w}_{ij} - \boldsymbol{\mu}_{c,ij}^{x_{ij}}]^T \ [\boldsymbol{K}_{c,ij}^{x_{ij}}]^{-1} \ [\boldsymbol{w}_{ij} - \boldsymbol{\mu}_{c,ij}^{x_{ij}}]$$

$$e_2 = -\sum_{(i,j)} \frac{1}{2} [\boldsymbol{f}_{ij} - \boldsymbol{\mu}_{t,ij}^{x_{ij}}]^T \ [\boldsymbol{K}_{t,ij}^{x_{ij}}]^{-1} \ [\boldsymbol{f}_{ij} - \boldsymbol{\mu}_{t,ij}^{x_{ij}}]$$

$$e_3 = -\sum_{C} V_C(\boldsymbol{x})$$

Eq. (7) has three terms: The first and second terms are the color and texture data consistency terms, and the third imposes spatial continuity of the segmentation labels. ICM is utilized to maximize Eq. (7) iteratively by alternating between computing the mean vectors and covariance matrices for the appropriate classes, and estimating $\boldsymbol{x}$ until a convergence criterion is satisfied.

5 RESULTS

Figure 2 demonstrates an example of integrating color, and texture using algorithms A and B to detect the presence of "grass" regions within a given image. Figure 2b shows the results of applying the pixel-based color classification algorithm. Note that, throughout the results, the classifications are shown as "white" on a "gray" background, where the "white" portions indicate the regions that resulted in a positive match. Figures 2c, and 2d demonstrate the results of smoothing Figure 2b by applying GRF and morphology, respectively. Figures 2e, 2f, and 2g demonstrate the results of combining color and texture using the block and region-based texture options of Algorithm A, and Algorithm B respectively. These classification are suitable for generating keyword annotations such as "grass" or "outdoor."

We also compiled the annotation results using the keywords "Outdoor" and "People" with a true positive (TP) - false positive (FP) approach on a database of

31 images. The annotations obtained manually (ground truth) are: true Outdoor 15, and true People 14. The results obtained using the pixel-based color classifier alone are: TP for Outdoor 15, FP for Outdoor 4, TP for People 14, and FP for People 1. Those obtained using color and texture are: TP for Outdoor 15, FP for Outdoor 0, TP for People 14, and FP for People 1.

6 CONCLUSIONS

This paper presented color and texture based classification algorithms designed to: 1) offer the user a wide range of options and flexibilities to enhance the outcome of the classification, and 2) provide a compromise between accuracy and computational complexity. The flexibility of allowing the user to select the appropriate tool or set of tools must be, in our opinion, an underlying requirement of any system designed for image access and retrieval. For instance, a user may be interested in retrieving all "grass" images and subsequently locating the "grass" in a given image. Certainly, keyword searching would be sufficient to retrieve all "grass" images in a "real time" fashion, while algorithm A or B, for instance, would help to locate the "grass" in the scene.

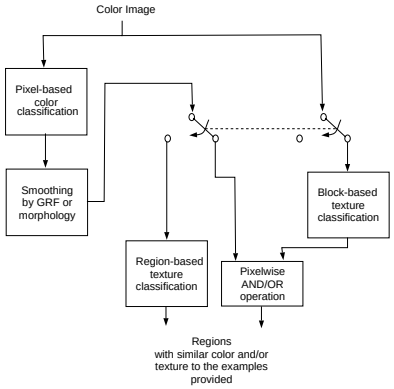

Figure 1: Block diagram of integration algorithm A

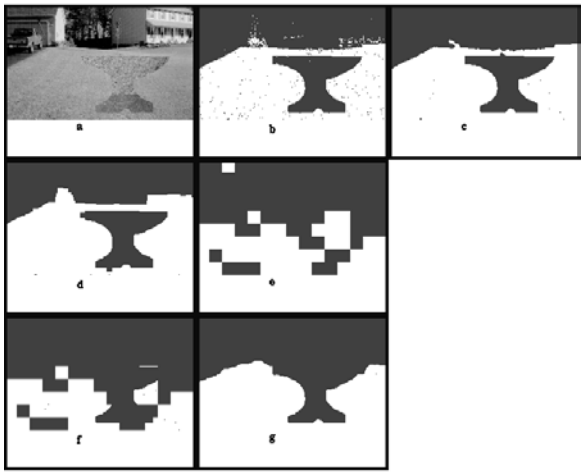

Figure 2: Integration of color and texture information

References

[1] B. Furht, S. W. Smoliar, and H. Zhang, *Image and video indexing and retrieval techniques*. Kluwer Academic Publishers, 1995.

[2] V. N. Gudivada and V. V. Raghavan (Eds.), "Special issue on content-based image retrieval systems," *Computer*, September 1995.

[3] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz, "Efficient and effective querying by image content," *Journal of Intelligent Information Systems*, vol. 3, pp. 231–262, 1994.

[4] A. Pentland, R. W. Picard, and S. Sclaroff, "Photobook: content-based manipulation of image databases," *International Journal on Computer Vision*, Fall 1995.

[5] S. F. Chang and J. R. Smith, "Extracting multidimensional signal features for content-based visual query," in *SPIE*, vol. 2501, pp. 995–1006, 1995.

[6] T. P. Minka and R. W. Picard, "Interactive learning using a society of models," *to appear in the special issue of Pattern Recognition on image databases.*

[7] "Xerox color encoding standards," tech. rep., Xerox Systems Institute, Sunnyvale, CA, 1989.

[8] J. Liu and Y. H. Yang, "Multiresolution color image segmentation," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 16, pp. 689–700, July 1994.

[9] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. on System, Man, and Cyber.*, vol. 3, pp. 610–621, November 1973.

[10] P. P. Ohanian and R. C. Dubes, "Performance evaluation for four classes of textural features," *Pattern Recognition*, vol. 25, no. 8, pp. 819–833, 1992.

[11] E. Saber, A. M. Tekalp, R. Eschbach, and K. Knox, "Annotation of natural scenes using adaptive color segmentation," in *IS & T/SPIE Symposium on Electronic Imaging: Science and Technology*, (San Jose, California), February 1995.

[12] M. M. Chang, M. I. Sezan, and A. M. Tekalp, "Adaptive Bayesian segmentation of color images," *Journal of Electronic Imaging*, vol. 3, pp. 404–414, October 1994.