Content-Based Image Retrieval- Literature Survey

EE 381K: Multi-Dimensional Digital Signal Processing

Michele Saad

March 18, 2008

Abstract

We survey feature extraction and selection techniques adopted in content-based image retrieval (CBIR); a technique that uses the visual content of a still image to search for similar images in large-scale image databases, according to a user's interest. The CBIR problem is motivated by the need to search the exponentially increasing space of image and video databases efficiently and effectively. The visual content of an image is analyzed primarily in terms of low-level features extracted from the image. These constitute color, texture and shape features. We present a survey of the most popular of these feature extraction algorithms. We then briefly discuss a feature selection algorithm based on a fuzzy approach and relevance feedback; an approach that attempts to bridge the gap between low-level features extracted from an image and high-level semantic features.

1 Introduction

The increase in computing power and electronic storage capacity has lead to an exponential increase in the amount of digital content available to users in the form of images and video, which form the bases of many entertainment, educational and commercial applications [1]. Consequently, the search for the relevant information in the large space of image and video databases has become more challenging. How to achieve accurate retrieval results is still an unsolved problem and an active research area. A typical image retrieval system includes three major components: i) feature extraction (usually in conjunction with feature selection), ii) high dimensional indexing and iii) system design [2]. In the following, we discuss the first component, that of feature extraction and selection. An image can be represented as a set of low-level visual features such as color, texture and shape features. While several image retrieval systems rely on only one feature for the extraction of relevant images, it has been shown that an appropriate combination of relevant features can yield better retrieval performance [3]. The process of determining the combination of features that is most representative of a particular query image is called feature selection. In this work, we survey color and texture feature extraction algorithms, as well as discuss a feature selection algorithm based on a fuzzy approach and relevance feedback. In section 2, we discuss color features along with the pros and cons of each. In section 3, we survey texture feature extraction techniques in a 'compare and contrast' manner. In section 4, we move on to a fuzzy feature selection algorithm, and we conclude in section 5.

2 Color Feature Extraction

Color features include the conventional color histogram (CCH), the fuzzy color histogram (FCH), the color correlogram (CC) and a more recent color-shape-based feature. The extraction of the color-based features follows a similar progression in each of the four methods: i) Selection of the color space, ii) quantization of the color space, iii) extraction of the color feature, iv) derivation of an appropriate distance function [4].

2.1 The Conventional Color Histogram

The conventional color histogram (CCH) of an image indicates the frequency of occurrence of every color in an image. From a probabilistic perspective, it refers to the probability mass function of the image intensities. It captures the joint probabilities of the intensities of the color channels (R, G and B in the RGB color-space, or H, S and V in the HSV color-space, and similarly for other color spaces). The CCH can be represented as $h_{A,B,C}(a, b, c) = N \cdot Prob(A=a, B=b, C=c)$, where A, B and C are the three color channels, and N is the number of pixels in the image [4]. Computationally, it is constructed by counting the number of pixels of each color (in the quantized color space). The appealing aspect of the CCH is its simplicity and ease of computation. There are however, several difficulties associated with the CCH. The first of these is the high dimensionality of the CCH, even after drastic quantization of the color space. Another downside of the CCH is that it does not take into consideration color similarity across different bins. Further, the CCH is a global image feature that does not encode any color-spatial information.

2.2 The Fuzzy Color Histogram

In the fuzzy color histogram (FCH) approach, a pixel color belongs to all histogram bins with different degrees of memberships to each bin. More formally, given a color space with K color bins, the FCH of an image I is defined as follows: $F(I)=[f_1, f_2, \dots f_K]$ where $f_i = \sum_{j=1}^{N} \mu_{ij} P_j = \frac{1}{N}\sum_{j=1}^{N} \mu_{ij}$. N is the number of pixels in an image, and μ_{ij} is the membership value of the j^{th} pixel in the i^{th} color bin [5] (key paper #1). The primary advantage of the FCH is that it encodes the degree of similarity of each pixel color to all other histogram bins through a fuzzy-set membership function (namely the μ_{ij}). Taking into account color pixel similarity makes the FCH more robust to quantization errors as well as to changes in light intensity [5]. The FCH however, still embeds several drawbacks. Like the CCH, the FCH delineates only the global color properties of the image, and the dimensionality of the FCH features is as high as that of the CCH. In addition, the FCH approach introduces the additional challenge of computing the appropriate fuzzy membership function μ_{ij}.

2.3 The Color Correlogram

The color Correlogram (CC) expresses how the spatial correlation of pairs of colors changes with distance. A CC for an image is defined as a table indexed by color pairs, where the d^{th} entry for row (i,j) specifies the probability of finding a pixel of color j at a distance d from a pixel of color i in the image [6]. In general, since local correlations between different colors are more significant than global correlations in an image, a small value of d is sufficient to capture special correlations. An efficient algorithm for computing the CC exists and is described in [6]. The computation is linear in the image size. The highlights of the CC method are that it encodes local as well as global spatial information, and it has been shown to work well for coarse color images [6]. Huang et al. have shown in [6] that 8-color CC perform better than 64-color CCH in the CBIR problem. The major drawback of this method is the high dimensionality of the feature space.

2.4 The Color-Shape Based Method

Howe et al. have proposed in [7] (key paper #2) a color-shape based method (CSBM) based on color, area and perimeter-intercepted lengths of segmented objects in an image. The algorithm starts by clustering image pixels into K clusters according to the K-means algorithm. The mean value of each cluster is regarded as a representative color for the cluster. A quantized color image I' is obtained from the original image I by quantizing pixel colors in the original image into K colors. A connected region having pixels of identical color is regarded as an object. The area of each object is encoded as the number of pixels in the object. Further, the shape of an object is characterized by 'perimeter-intercepted lengths' (PILs), obtained by intercepting the object perimeter with eight line segments having eight different orientations and passing through the object center. The PILs have been shown to be a good characterization of object shapes. The immediate advantage of this method is that it encodes object shapes as well as colors. The drawback on the other hand, is more involved computation, and the need to determine appropriate color thresholds for the quantization of the colors. Another drawback of CSBM is its impressionability to contrast and noise variation.

3 Texture Feature Extraction

The notion of texture generally refers to the presence of a spatial pattern that has some properties of homogeneity [4]. Directional features are extracted to capture image texture information. Texture feature extraction methods include the steerable pyramid, the contourlet transform, the Gabor wavelet transform and the complex directional filter bank (CDFB).

3.1 The Steerable Pyramid

The steerable pyramid generates a multi-scale, multi-directional representation of the image. The basic filters are translations and rotations of a single function. The image is decomposed into one decimated low-pass sub-band and a set of undecimated directional sub-bands. The decomposition is iterated in the low-pass sub-band. Because the directional sub-bands are undecimated, there are 4K/3 times as many coefficients in the representation as the original image, where K is the number of orientations [8] (key paper #3).

3.2 The Contourlet Transform

The contourlet transform provides a multi-scale, multi-directional decomposition of an image. It is a combination of a Laplacian pyramid and a directional filter bank (DFB). Bandpass images from the Laplacian pyramid are fed into the DFB so that directional information can be captured. The low frequency components are separated from the directional components. After decimation, the decomposition is iterated using the same DFB. Its redundancy ratio is less than 4/3 because the directional sub-bands are also decimated [8].

3.3 The Gabor Wavelet Transform

To obtain a Gabor filter bank with K orientations and S scales, the two-dimensional Gabor function is dilated and rotated appropriately by setting the parameters of the Gabor function (thus obtaining K*S Gabor functions). The image is then convolved with each of the obtained Gabor functions. It has been shown that the Gabor Transform for texture image retrieval yields the highest texture retrieval results [8]. However, it results in an over-complete representation of the original image with a redundant ratio of K*S.

3.4 The Complex Directional Filter Bank

The shift-invariant complex directional filter bank (CDFB) has recently been proposed by Oraintara et al. in [8]. The transform consists of a Laplacian pyramid and a pair of DFBs, designated as primal and dual filter banks. The filters of these filer banks are designed to have special phase functions so that the overall filter is the Hilbert transform of the primal filter bank. A multi-resolution representation is obtained by reiterating the decomposition at the lowpass branch. The attractive features of the CDFB are its shift invariance, its comparably high texture retrieval performance and its relatively low redundancy ratio. The over-complete ratio of the CDFB is bounded by 8/3, whereas those of the Gabor transform and steerable pyramid increase linearly with the number of directional sub-bands [8].

A plot of the frequency response of the four methods is shown in figure 1. The 4 texture retrieval methods were compared in [8] for texture image retrieval and the results are shown in figure 2.

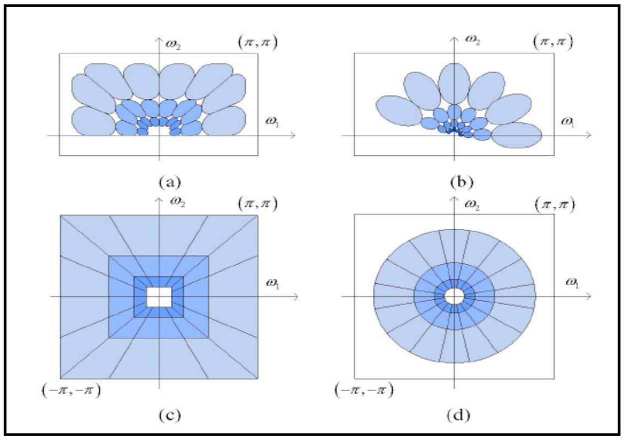

Fig. 1: Frequency response of the 4 DFBs. (a) CDFB (S=3, K=8), (b) Gabor Wavelet (S=4, K=6), (c) Contourlet Transform (S=3, K=8), (d) Steerable Pyramid (S=3, K=8)

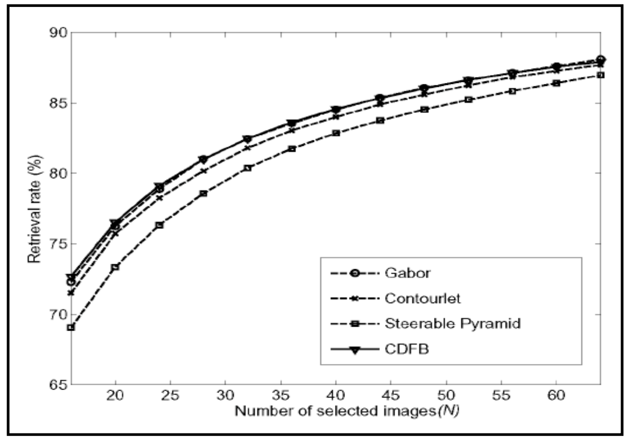

Fig. 2: Average texture retrieval rate for the 4 DFB methods as a function of the number of top images considered.

4 Fuzzy Feature Selection with Relevance Feedback

The goal of feature selection is to find the optimal feature subspace where the 'relevant' and 'irrelevant' feature sets are best separated. In an attempt to bridge the gap between high-level user semantics and low-level visual features, Jiang et al. proposed in [9] (key paper #4) an online feature selection algorithm in the relevance feedback learning process. The online feature selection algorithm is implemented in a boosting manner by combining incrementally learned classifiers over the selected features into a strong ensemble classifier. The learning phase involves acquiring feedback from users that are asked to label the initially returned images as 'relevant' or 'irrelevant'.

5 Implementation and Future Work

We propose to compare the surveyed color and texture feature extraction techniques as well as implement the proposed feature selection algorithm in [9] to study the combined effect of the multiple feature extraction techniques in image retrieval. The code is currently being written in MATLAB, and the simulations will be done on an image dataset of 10,000 images obtained from the web-crawled *misc* database [10].

References

[1]. A. Bovik, *Handbook of Image and Video Processing*, 2nd Edition, Elsevier Academic Press, ISBN 0-12-119792-1, 2005.

[2]. Y. Rui, T. S. Huang and S. Chang, "Image Retrieval: Current Techniques, Promising Directions and Open Issues ", *Journal of Visual Communication and Image Representation,* vol. 10, pp. 39-62, March 1999.

[3]. P Liu, K. Jia, Z. Wang and Z. Lv, "A New and Effective Image Retrieval Method Based on Combined Features", *Proc. IEEE Int. Conf. on Image and Graphics,* vol. I, pp. 786-790, August 2007.

[4] J. R. Smith and S.-F. Chang, " Automated image retrieval using color and texture", Columbia University, Tech. Rep CU/CTR 408-95-14, July 1995.

[5]. J. Han and K. Ma, "Fuzzy Color Histogram and Its Use in Color Image Retrieval", *IEEE Trans. On Image Processing,* vol. 11, pp. 944 – 952, Aug. 2002.

[6]. J. Huang, S. R. Kumar, M. Mitra and W. J. Zhu, R. Zabih, "Image Indexing Using Color Correlograms", *Proc. IEEE Conf. on Computer Vision and Pattern Recognition,* pp. 762 – 768, June 1997.

[7]. N. R. Howe and D. P. Huttenlocher, "Integrating Color, Texture and Geometry for Image Retrieval", *Proc. IEEE Conf. on Computer Vision and Pattern Recognition,* vol. II, pp. 239-246, June 2000.

[8]. S. Oraintara and T. T. Nguyen, "Using Phase and Magnitude Information of the Complex directional Filter Bank for Texture Image Retrieval", *Proc. IEEE Int. Conf. on Image Processing,* vol. 4, pp. 61-64, Oct. 2007.

[9]. W. Jiang, G. Er, Q. Dai and J. Gu, "Similarity-Based Online Feature Selection in Content-Based Image Retrieval", *IEEE Trans. on Image Processing*, vol. 15, no. 3, pp. 101-104, March 2006.

[10]. http://wang.ist.psu.edu/docs/related/