COMPARATIVE STUDY OF SPATIAL COLOR AND SHAPE FEATURES FOR LOW LEVEL CONTENT BASED IMAGE RETRIEVAL SYSTEM

Jyoti Narwade, Ph.D Scholar, PACIFIC University Udaipur Rajasthan, India
Dr. M.M.Puri Director, JSPM Group of Institutes, Pune, India

Abstract— In multimedia technology, image data in various formats is becoming available at an explosive rate. Various multimedia systems need huge image data collection. Such data sources need search and retrieval of image databases to provide open access to relevant information source. Thus content-based information retrieval has become active research area. Development of Content Based Image Retrieval is enhanced due to growth in need of variety of images in multimedia technology. In particular texture, color and shape features of image are used to define similarity among images. In this paper a brief overview of local feature descriptors of image is given.

Index Terms— Color features, shape features, feature extraction, image feature similarity, image retrieval.

1. INTRODUCTION

Images have always been an inevitable part of human communication and its roots millennia ago. Images make the communication process more interesting, illustrative, elaborate, understandable and transparent. An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. Early techniques were not generally based on visual features but on the textual annotation of images. Images were first annotated with text and then searched using a text-based approach from traditional database management systems. Text-based image retrieval uses traditional database techniques to manage images. Through text descriptions, images can be organized by topical or semantic hierarchies to facilitate easy navigation and browsing based on text queries. Use of image annotation is time-consuming, laborious and expensive. Hence it is difficult for the traditional text-based methods to support a variety of task-dependent queries[1].

Content based Image Retrieval is the process of retrieving images from a database of digital images according to the visual contents of an image. Visual contents of the images in the database are extracted and described by multi-dimensional feature vectors. The feature vectors of the images in the database form a feature database. To retrieve images, users provide the retrieval system with example images or sketched figures. The system then changes these examples into its internal representation of feature vectors. The similarities/distances between the feature vectors of the query example and those of the images in the database are then calculated. Images which have minimum distance with feature vectors of query image are retrieved. The algorithms used in these systems include major task of feature extraction, selection and classification.

The extraction task transforms rich content of images into various content features. Feature extraction is the process of generating features to be used in the selection and classification tasks. Feature selection reduces the number of features provided to the classification task. Those features which are likely to assist in discrimination are selected and used in the classification task. Features which are not selected are discarded [19]. Of these three activities, feature extraction is most critical because the particular features made available for discrimination directly influence the efficacy of the classification task. The end result of the extraction task is a set of features, commonly called a feature vector, which constitutes a representation of the image. Figure 1 shows block diagram of typical CBIR system.

A visual content descriptor can be either global or local. A global descriptor uses the visual features of the whole image, whereas a local descriptor uses the visual features of regions or objects to describe the image content.

2. REVIEW OF LITERATURE

D.A. Kumar and J. Esther [8] discuss content based image retrieval system using texture and color feature by Color Histogram, Gabor and Wavelet Transform. The Gabor filter [17] is frequently used filter method in texture extraction. A variety of Gabor filters in different degree and their relative positions captures value at that specific frequency and direction. Texture can be extracted from this group of value distributions [8]. Other texture feature extraction methods are co-occurrence matrix, wavelet decomposition, Fourier filters, etc

K. E. A. van de Sande, T. Gevers and C. G. M. Snoek[18] have evaluated performance of various color descriptors for object and scene recognition. Color histogram describes the different colors distribution in an image in a simple and computationally efficient manner.
D.A. Kumar and J. Esther[8] mentions drawback of Traditional color histogram representation as it is dependent of the color of the object being studied, ignoring its shape and texture. Color histograms can be identical for two images with different object content which share color information. Other color feature extraction techniques are region histogram, color coherence vector, color moments, correlation histogram etc.

Thepade, S.D.[19] compared performance of content based video retrieval technique on different color spaces. Lianghai Jin, Enmin Song, Lei Li, Xiang Li[3] discuss various methods for shape feature extraction. They used gradient operators for edge detection. The problem with edge extraction using gradient operators is detection of edges in only either horizontal or vertical directions. Shape feature extraction in image retrieval requires the extracted edges to be connected in order to reflect the boundaries of objects present in the image.

S. Varadarajan, C. Chakrabarti, L. J. Karam, and J. M. Bauza[16] have used canny edge detector for edge histogram computation. Haibin Ling, David W. Jacobs[1] discuss shape Classification Using the Inner-Distance. Hai Shan Wu[15] used shape distributions for 2-D shape matching. Shape descriptors such as Smooth Curve Decomposition, Convex Hull, Triangle-area representation (TAR) and Curvature Scale Space (CSS) have drawback as descriptor has to correspond to key points such as maxima of curvature or inflection points.

To quantify the similarity between two histograms, there are many methods being reported: Euclidean distance, histogram intersection, histogram quadratic (cross) distance and Canberra distance[6] Minkowski, Kullback-Leibler Divergence, Jeffrey Divergence, Quadratic-form, Earth Mover’s Distance, \( \chi^2 \) statistics, Hausdorff distance, etc.[12].

The difference between the approaches proposed in this paper and above works is that color space selection and quantization parameters play important role as that of various color and shape features during image retrieval. We propose to implement combination of shape and color features for content based image retrieval system. Chord context is effective descriptor[10] for shape feature to extract contour attributes and slope magnitude method along with the gradient operator such as canny edge detector to extract the shape features in form of connected boundaries[16]. Spatial color feature such color coherence vector and hybrid histogram[13] are to be implemented.

3. Feature Extraction

Image content includes both visual and semantic content. Visual contents are very general or domain specific. It includes color, texture, shape, spatial relationship, etc.

Domain specific visual contents like human faces, fingerprints are application dependent. Feature extraction system was developed by J. Smith [6]. Color feature extraction system has main 3 stages- selection of a color space, quantization of the color space, computation of color feature [12]. This chapter concentrates on general visual content descriptors.

3.1 Color Features

In CBIR system, Color is the most common visual feature, primarily because of the simplicity of extracting color information from images [1, 8]. Color is a property that depends on the reflection of light to the eye and the processing of that information by the brain. We use color to tell the difference between objects, places, and the time of day. Generally color features are extracted using the color histogram technique[19].

3.1.1 Color Models

A Color Space is a model for representing color in terms of intensity values. Each pixel of the image is represented as a point in a 3D color space. Commonly used color space for image retrieval include RGB, Munsell, CIE L*a*b*, CIE L*u*v*, HSV (or HSL, HSB), and opponent color space. There is no agreement on which is the best. However, one of the desirable characteristics of an appropriate color space for image retrieval is its uniformity [13]. Uniformity means that two color pairs that are equal in similarity distance in a color space are perceived as equal by viewers. In other words, the measured proximity among the colors must be directly related to the psychological similarity among them.

RGB space is a widely used color space for image display. It is composed of three color components red, green, and blue. These components are called "additive primaries" since a color in RGB space is produced by adding them together. In contrast, CMY space is a color space primarily used for printing. The three color components are cyan, magenta, and yellow. These three components are called "subtractive primaries" since a color in CMY space is produced through light absorption. Both RGB and CMY space are device-dependent and perceptually non-uniform.

CIE L*a*b* and CIE L*u*v* spaces are device independent and considered to be perceptually uniform. They consist of a luminance or lightness component (L) and two chromatic components a and b or u and v. CIE L*a*b* is designed to deal with subtractive colorant mixtures, while CIE L*u*v* is designed to deal with additive colorant mixtures. The transformation of RGB space to CIE L*u*v* or CIE L*a*b* space can be found in [13].

In HSV (or HSL, or HSB) space is widely used in computer graphics and is a more intuitive way of describing color. The three color components are hue, saturation (lightness)
and value (brightness). The hue is invariant to the changes in illumination and camera direction and hence more suited to object retrieval. RGB coordinates can be easily translated to the HSV (or HLS, or HSB) coordinates [14].

3.1.2 Color Coherence Vector

A color coherence vector (CCV) stores the number of coherent versus incoherent pixels with each color. By separating coherent pixels from incoherent pixels, CCV’s provide finer distinctions than color histograms. Color’s coherence is defined as the degree to which pixels of that color are members of large similarly-colored regions. These significant regions are referred as coherent regions [12].

Color coherence vector is represented as \((\alpha_i, \beta_i)\) for each color \(i\). \(\alpha\) is number of uniformly colored pixels. \(\beta\) is number of non uniformly colored pixels. For each color \(i\) we compute the pair \((\alpha_i, \beta_i)\). This pair is called as the coherence pair for the \(j^{th}\) color. The color coherence vector for the image consists of \((\alpha_1, \beta_1), \ldots, (\alpha_n, \beta_n)\). This is a vector of coherence pairs, one for each color.

3.1.3 Hybrid Histogram

It is combination of annular and angular histogram. Angular regions are formed by forming both annular and angular regions. Hybrid distribution density is a vector \([|R_1|, |R_2|, \ldots, |R_N|]\). Hybrid distribution density vector [13] is calculated by counting the number of points in each sector region.

3.2 Shape Feature

Shape is an important visual feature and it is one of the basic features used to describe image content. It is an attribute that must not change when the original object is submitted to a certain set of geometric transformations such as translation, scaling and rotations. Shape descriptors can be divided into two main categories as region based and contour-based methods. Region-based methods use the whole area of an object for shape description, while contour-based methods use only the information present in the contour of an object. Both boundary-based and region-based descriptions are perceptually meaningful and interchangeable in the sense that each one can be used as a basis to compute the other (e.g. by filling-in the interior region or by tracing the boundary). But the explicit shape features available in each type of description are quite different, so that an ideal description should include both boundaries and regions in order to obtain more efficient retrieval.

3.2.1 Chord Context

Chord context analysis corresponds to finding the distribution of all chord lengths in different directions in a given shape. For discrete binary image data, we consider each object point as one and the background as zero. In the shape recognition field, it is common to consider the case where the general function \(f(x, y)\) is given by equation 1.

\[
f(x, y) = f(x) = \begin{cases} 
1, & \text{if } f(x, y) \in D \\
0, & \text{otherwise}
\end{cases}
\]

where \(D\) is the domain of the binary shape. In each direction, we can find all the chords in the shape. Fig. 1 shows an example of chords in direction \(\theta\). A set of lines \(T(\rho, \theta)\) is defined by

\[
\rho = x \cos \left(\theta - \frac{\pi}{2}\right) + y \sin \left(\theta - \frac{\pi}{2}\right),
\]

\[\theta, \rho \in [0, \pi], \text{ and } \rho \in [-\infty, \infty].\]

The chords are defined by the parts of these lines within the domain of the binary shape. So a shape can be represented by a discrete set of chords sampled from its silhouette. Considering different angles \(\theta\), the number and length of chords obtained in different directions may not be the same, except in the case of a circle. One way to capture this information is to use the distribution of chord lengths in the same direction in a spatial histogram. Concretely, let us assume that the set of chords in directions \(\theta_i\) are represented by

\[
C = \{c_{i,n} | n \in [1, l]\}
\]

where \(Ni\) is the number of the chords in direction \(\theta_i\). Let \(L(ci,n)\) be the length of chord \(ci,n\). Equation (4) is used to compute a histogram \(hi\) in direction \(\theta_i\) by

\[
\text{for } l \in [1, L_{i,m}], \text{ and } l \in [1, L_{i,m}]
\]

where \(L_{i,max}\) is the longest chord in direction \(\theta_i\). A low pass filter is applied to discard too short chords. If we normalize a shape in an image with 128×128 pixels, i.e. the largest size of the shape is 128 pixels, and the shorter size transforms in proportion, then we can consider the set of chords whose length is shorter than 4 to be too short chords. So they should be discarded. Figure 3.2.1 shows too short chords.

Fig. 3.2.1: short and long chords

With \(\theta\) increasing from 0 to 179 degrees, all the chords in different directions in the silhouette can be recorded. If we
divide the orientation range [0, 179] into D’, then we can obtain D’ histograms \( h_i \), \( i \in [0, D’-1] \). Matrix \( M \) arranged by a set of histograms with column vector \( h \) according to the order of angles represented as follows:

\[
M = [h_0, h_1, h_2, \ldots, h_{D’-1}] \quad (5)
\]

The matrix element is the number of equal-length chords whose direction and length are given by the value of abscissa and y-axis, respectively. The abscissa is the orientation angle \( \theta \), and the y-axis is the length of the chords. The value in each row of the matrix \( M \) is the number of the chords with same length in different directions; and each column is the chord length histogram in the same direction.

3.2.2 Canny Edge detection

Edge detection is a very important in image analysis. Edges give idea about the shapes of objects present in the image. Canny edge detector[11] uses a filter based on the first derivative of a Gaussian, because it is susceptible to noise present on raw unprocessed image data, the raw image is convolved with a Gaussian filter. Canny operator is nothing but gradient of Gaussian filtered image. Following are the steps to generate edge histograms[16].

1. **Smoothing**: Blurring image to remove noise.
2. **Finding gradients**: The edges should be marked where the gradients of the image have large magnitudes.
3. **Non - maximum suppression**: Only local maxima should be marked as edges.
4. **Double thresholding**: Determination of Potential edges
5. **Edge tracking by hysteresis**: Final edges are determined by suppressing all edges that are not connected to a very certain (strong) edge.

The figure 3.2.2 depicts the results[16] of before and after feature extraction of image.

4. SIMILARITY MEASUREMENT METRICS

To define similarity among images based on appearance of image distances of query image with each of the image in database are computed using distance measures. These distances are used to rank images in the database. Image in the database which is having less difference value with query image as compared to others is given a top most rank. Finally from database ten most similar images with query image are displayed to user. Various similarity measures that can be used are as follows[12].

4.1 Euclidean Distance Measure

In this distance formula, there is only comparison between the identical bins in the respective. histograms. Two different bins may represent perceptually similar colors but are not compared crosswise. All bins contribute equally to the distance.

\[
D = \sum_{q=0}^{D-1} (d(q, \xi))
\]

where D stands for distance, q is bin number. Each bin is calculated independent of other. Finally all differences are summed together to get distance in images.

4.2 Histogram Intersection Distance

The color histogram intersection was proposed for color image retrieval in [4]. The intersection of histograms \( h \) and \( g \) is given by:

\[
d(h, g) = \sum_{q=0}^{D-1} \left[ \min (h(q), g(q)) \right] \] (7)

where \( |h| \) and \( |g| \) gives the magnitude of each histogram, which is equal to the number of samples. Colors not present in the user’s query image do not contribute to the intersection distance. This reduces the contribution of background colors. The sum is normalized by the histogram with fewest samples.

4.3 Character Matrix Distance

Equation 8 gives metric used to define similarity between short context shape descriptor for two different images[22].

\[
CMD_{Fq,Fm} = \min \left[ \frac{1}{2} \text{Dist}Fq,Fm \right]
\]

5. METHODOLOGY

We propose to analyze shape and color feature extraction techniques one by one using a query image. Techniques to be used are CIELab color space, Color coherence vector, hybrid histogram, canny edge detection and chord context shape feature. These techniques will be evaluated using parameters, Time, Accuracy and Redundancy Factor. The
COMPARATIVE STUDY OF SPATIAL COLOR AND SHAPE FEATURES FOR LOW LEVEL CONTENT BASED IMAGE RETRIEVAL SYSTEM (79-84)

goal is to find the optimum combination of techniques to be used for better image retrieval.

5.1 Time
It is the time taken in seconds for the retrieval task to complete, at the end of which the system returns the images which are matched with the features of the query images, according to the technique used.

5.2 Accuracy
Accuracy of an image retrieval task as given in equation 9, defined as the ratio of the number of relevant images retrieved to the total number of images retrieved expressed in percentage.

\[
\text{Accuracy} = \frac{\text{No of Relevant images}}{\text{Total number of images retrieved}} \times 1 \tag{9}
\]

Where, total number of images retrieved = number of relevant images + number of irrelevant images

5.3 Redundancy Factor
Redundancy Factor (RF) is one aspect which has been largely neglected in the analysis of CBIR techniques. It is a measure to take into account the extent of irrelevant images returned upon completion of a retrieval process. It is expressed as:

\[
\text{RF} = \frac{\text{images retrieved – total images in a class}}{\text{total number of images in a class}} \tag{10}
\]

6. Conclusion
Limitation of RGB color space as discussed in [7] that the proximity of colors does not indicate color similarity. So it is necessary to select a uniform color space[9]. Robust and accurate image segmentation is difficult to achieve, comparing only shape features for image retrieval has been limited[16]. Hybrid histogram and color coherence vector gives better results in terms of accuracy and time in quantized color space[8].

References


Author
COMPARATIVE STUDY OF SPATIAL COLOR AND SHAPE FEATURES FOR LOW LEVEL CONTENT BASED IMAGE RETRIEVAL SYSTEM (79-84)

Jyoti A. Manoorkar is Assistant Professor in the Department of IT at MIT College of Engineering, Pune and is having 5 Years of experience in teaching field. She is PhD scholar at PACIFIC University Udaipur Rajasthan. She is the member of CSI Communications. Her interest is in the area of Image processing.

Dr M.M.Puri works as Director at JSPM Group of Institutes Pune. Dr M.M.Puri has over 24 years of teaching as well as administrative experience in the field of technical education. She is M.Tech from N.I.T Kurukshetra and PhD (IT) from Guru Gobind Singh Indraprastha University ,Delhi. She has had the distinction of working at All India Council for Technical Education (A.I.C.T.E) as Directo. Dr M.M.Puri is a member of various professional bodies like I.S.T.E ,C.S.I ,AIMA and I.E.E.E. She was also member Board of Studies (Electronics) Pune University and member Board of Examinations II.I.E. She has worked as Vice President of Poona Management Association and treasurer of AMMI. She has also been invited to contribute a chapter on “Quality in Engineering Education” being published by Melbourne University Australia. She is an expert member on the various teams of N.B.A(National Board of Accreditation). She is a recognized PhD guide under university of Pune. Her research interest lies in the field of Software Engineering and application of soft computing approach i.e Neural Networks and Fuzzy Logic to various software engineering problems. She has published four books in the field of electronics for under graduate engineering students.