Survey Paper of Image Matching and Retrieval using Discrete Sine Transform

Vivek Kumar Soni¹
M. Tech Scholar, Department of Computer Science and Engg, CSIT, Durg (C.G.) INDIA
Mrs. Shikha Agrawal²
Assistant professor, Department of Computer Science and Engg, CSIT, Durg (C.G.) INDIA

Abstract- Management of the visual information is challenging as the quantity of data available is very huge. Current utilization of available databases is limited due to the issues in retrieval of the necessary images from a huge repository. This paper provides a comprehensive survey of the technical achievements in the research area of image retrieval, especially content-based image retrieval, an area that has been so active and prosperous in the past few years. A novel image matching scheme using Discrete Sine Transform and classification technique for classification of the given input image. Our system classifies the images based on the similarity of the images.

Index Terms- CBIR, Precision, Recall, Euclidian Distance, FFT sectors, Discrete Sine Transform, Classifier,

I INTRODUCTION

Based on the similarity of visual content for an given input query image, the task of image retrieval is to retrieve relevant images from an image repository. Query images are given as input by the user as example of his search. Low level features are one of the most popular method for Image Retrieval systems[1] with each model using a combination of low-level features, and then define a distance metric that is used to quantify the similarity between image models. A disadvantage of this method is that low-level image features cannot always capture the human perception of image similarity. In other words, semantic content of an image are found to be difficult for feature extraction using only low-level image features. This is known in as the semantic gap problem. Most of the image retrieval systems found in focuses on general purpose imagery of outdoor scenes. Advantage using such images are they tend to be rich in color and texture and can be characterized by global signatures of such properties. Examples of these earlier systems have been reported in various papers [2].

Transforming the image from spatial domain to frequency domain is one of the popular methods used in Image Retrieval. Various methodology have been proposed in the literature[3]. In this paper we focused on various image retrieval systems and use the Discrete sine transform with down sampling techniques to extract relevant features and apply classification techniques on them.

II. IMAGE RETRIEVAL SYSTEMS

Since the early 1990s, content-based image retrieval has become a very active research area. Many image retrieval systems, both commercial and research, have been built. Most image retrieval systems support one or more of the following options [4):

- random browsing
- search by example
- search by sketch
- search by text (including key word or speech)
- navigation with customized image categories.

We have seen the provision of a rich set of search options today, but systematic studies involving actual users in practical applications still need to be done to explore the trade-offs among the different options mentioned above. Here, we will select a few representative systems and highlight their distinct characteristics.

A. QBIC

QBIC [5, 6, 7, 8, 9, 10, 11], standing for query by image content, is the first commercial content-based image retrieval system. Its system framework and techniques have profound effects on later image retrieval systems. QBIC supports queries based on example images, user-constructed sketches and drawings, and selected color and texture patterns, etc. The color feature used in QBIC are the average (R,G,B), (Y,i,q), (L,a,b), and MTM (mathematical transform to Munsell) coordinates, and a k-element color histogram [12]. Its texture feature is an improved version of the Tamura texture representation [13]; i.e. combinations of coarseness, contrast, and directionality. Its shape feature consists of shape area, circularity, eccentricity, major axis orientation, and a set of algebraic moment invariants [9,7]. QBIC is one of the few systems which takes into account the high dimensional feature indexing. In its indexing subsystem, KLT is first used to perform dimension reduction and then R+-tree is used as the multidimensional indexing structure [10, 7]. In its new system, text-based key word search can be combined with content-based similarity search.

B. Virage

Virage is a content-based image search engine developed at Virage Inc. Similar to QBIC, Virage [14, 15] supports visual queries based on color, composition (color layout), texture, and structure (object boundary information). But Virage goes one step further than QBIC. It also supports arbitrary combinations of the above four atomic queries. The users can adjust the weights associated with the atomic features according to their own emphasis. In Jeffrey et al. further proposed an open framework for image management. They...
classified the visual features (“primitive”) as general (such as color, shape, or texture) and domain specific (face recognition, cancer cell detection, etc.). Various useful “primitives” can be added to the open structure, depending on the domain requirements. To go beyond the query by-example mode, Gupta and Jain proposed a nine-component query language framework.

C RetrievalWare

RetrievalWare is a content-based image retrieval engine developed by Excalibur Technologies Corp. [16, 17]. From one of its early publications, we can see that its emphasis was in neural nets to image retrieval [16]. Its more recent search engine uses color, shape, texture, brightness, color layout, and aspect ratio of the image, as the query features [17]. It also supports the combinations of these features and allows the users to adjust the weights associated with each feature.

D Photobook

Photobook is a set of interactive tools for browsing and searching images developed at the MIT Media Lab. Photobook consists of three sub books from which shape, texture, and face features are extracted, respectively. Users can then query, based on the corresponding features in each of the three sub books. In its more recent version of Photobook, Four Eyes, Picard et al. proposed including human in the image annotation and retrieval loop [18,19]. The motivation of this was based on the observation that there was no single feature which can best model images from each and every domain. Furthermore, a human’s perception is subjective. They proposed a “society of model” approach to incorporate the human factor. Experimental results show that this approach is effective in interactive image annotation.

E VisualSEEk and WebSEEk

VisualSEEk [20, 21] is a visual feature search engine and WebSEEk is a World Wide Web oriented text/image search engine, both of which are developed at Columbia University. Main research features are spatial relationship query of image regions and visual feature extraction from compressed domain. The visual features used in their systems are color set and wavelet transform based texture feature. To speed up the retrieval process, they also developed binary tree based indexing algorithms. VisualSEEk supports queries based on both visual features and their spatial relationships. This enables a user to submit a “sunset” query as red-orange color region on top and blue or green region at the bottom as its “sketch.” WebSEEk is a web oriented search50 RUI, HUANG, AND CHANG engine. It consists of three main modules, i.e. image/video collecting module, subject classification and indexing module, and search, browse, and retrieval module. It supports queries based on both keywords and visual content. Netra is a prototype image retrieval system developed in the UCSB Alexandria Digital Library (ADL) project. Netra uses color, texture, shape, and spatial location information in the segmented image regions to search and retrieve similar regions from the database. Main research features of the Netra system are its Gabor filter based texture analysis [22,23], neural net-based image thesaurus construction and edge flow-based region segmentation.

G MARS

MARS (multimedia analysis and retrieval system) was developed at University of Illino is at Urbana-Champaign [24,25]. MARS differs from systems in both the research scope and the techniques used. It is an interdisciplinary research effort involving multiple research communities: computer vision, database management system (DBMS), and information retrieval (IR). The research features of MARS are the integration of DBMS and IR (exact match with ranked retrieval integration of indexing and retrieval (how the retrieval algorithm can take advantage of the underline indexing structure), and integration of computer and human. The main focus of MARS is not on finding a single “best” feature representation, but rather on how to organize various visual features into a meaningful retrieval architecture which can dynamically adapt to different applications and different users. MARS formally proposes a relevance feedback architecture in image retrieval and integrates such a technique at various levels during retrieval, including query vector refinement, automatic matching tool selection, and automatic feature adaption.

III IMAGE SEARCH AND RETRIEVAL

Digital image search and retrieval is very popularly known as Content-based image retrieval (CBIR)[26][27], also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) is the application of computer vision to the image retrieval problem, that is, the problem of searching for digital images in large databases. “Content based” means that the search will analyze the actual contents of the image. The term ‘content’ in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself. Without the ability to examine image content, searches must rely on metadata such as captions or keywords, which may be laborious, ambiguous or expensive to produce. There is a growing interest in CBIR because of the limitations inherent in metadata-based systems, as well as the large range of possible uses for efficient image retrieval. Textual information about images can be easily searched using existing technology,[28][29] but requires humans to personally describe every image in the database. This is impractical for very large databases, or for images that are generated automatically, e.g. from surveillance cameras. It is also possible to miss images that use different synonyms in their descriptions. Systems based on categorizing images in semantic classes like “cat” as a subclass of “animal” avoid this problem but still face the same scaling issues. Potential uses for CBIR include: Art collections, Photograph archives, Retail catalogues, Medical diagnosis, Crime prevention, The military Intellectual property, Architectural and engineering design,Geographical information and remote sensing systems.

IV. ALGORITHMS

The method of image search and retrieval proposed here mainly focuses on the generation of the feature vector of search based on the real and imaginary parts of the complex numbers of the image transform generated by the Fast Fourier
transform (FFT). Steps of the algorithm are given below.

**Step1:** Fast Fourier Transform of each components i.e. R,G and B of an RGB image is calculated separately.

**Step2:** The frequency domain plane of each color components i.e. R,G and B are divided into 8 equal polar sectors by calculating the angle theta of every element in frequency domain.

**Step3:** The real and imaginary parts of the Fourier complex numbers in each sector are calculated and their average value is taken as one of the parameter of the feature vector. For this purpose we have selected 4 sectors of upper half of complex plane.

**Step4:** The Euclidian distance between the feature vectors of query image and the feature vectors of images in the database are calculated.

**Step5:** The algorithm performance is measured based on the average precision and average recall of each class of images and their average across the class.

V FEATURE VECTOR GENERATION

Every complex number can be represented as a point in the complex plane, and can therefore be expressed by specifying either the point's Cartesian coordinates (called rectangular or Cartesian form) or the point's polar coordinates (called polar form). A complex number x represented in cartesian coordinates as

\[ x = a + jb \]

is represented as \( Re_x \),

where \( R = (a^2 + b^2)^{1/2} \)

and \( \phi = \tan^{-1}(b/a) \)

This helps us to generate eight components of a feature vector based on the complex plane as mentioned above. The real and imaginary parts of complex numbers of images generated by the Fast Fourier transform are checked for the angle of complex plane to allocate them to different sectors each of \( \Pi/4 \) radians. The real and imaginary parts of the complex numbers lying in the range of angles 0 to \( \Pi \) is taken into consideration to generate feature vector of dimension 8.

The feature vectors are generated by taking mean of real and imaginary parts of the complex numbers in following ranges (0-\( \Pi/4 \), \( \Pi/4-\Pi/2 \), \( \Pi/2 - 3\Pi/4 \), \( 3\Pi/4 - \Pi \)). [18]

VI EUCLIDIAN DISTANCE

Similarity measure is a key component in image retrieval. Traditionally, Euclidean distance is used to measure the similarity between the query and the images in the database. The smaller the distance, the more similar the pattern to the query. However, this metric is sensitive to the sample topology, as illustrated in Fig. 1 (a).

Assume point “A” is the query, the Euclidean distance based similarity measure can be viewed as drawing a hyper-sphere in the high dimensional feature space (or a circle in 2-D), centered at point “A”. The larger the radius of the hyper sphere, the more images are enclosed in the hyper-sphere, as shown in Fig. 1 (a). The radius is determined indirectly by the number of retrieved images. For different queries, the center of the circles move accordingly. As a result, the retrieved images enclosed by the hyper-sphere are different although these query images are perceptually similar. Furthermore, many irrelevant images could be enclosed by the regular hyper-sphere and presented to the user. To solve these problems, we propose to use an “irregular” non-spherical boundary to separate the similar images from the dissimilar ones, and the Euclidean distance measure is applied only to a limited number of images. As shown in Fig. 1 (b), the Euclidean similarity measure for query “A” is only done with respect to the black rectangular patterns. The boundary can be learned from the positive and negative examples provided by the user in image retrieval. We decide to use learning techniques that are non-parametric and do not need a large number of examples to learn a decision boundary. Large margin classifiers, such as SVM [30][31] and AdaBoost [3], can be used for such purpose.

![Figure 1. Perceptually similar patterns as rectangular ones. See the text for description.](image)

Can we directly use the distances of the images to the boundary to define the similarity measure? The answer is “no”. Suppose a query image “B” is given by the user, which is very similar to image “C”, as shown in figure(a) and figure (b). Both “B” and “C” are at the positive side of the boundary, and yet close to the boundary. In such a case, other images with large distances to the boundary will always be ranked in the top matches when the distance from- boundary (DFB) metric is used for similarity measure, while image “C” can only be retrieved for example after top 20 matches or even more. In an extreme case, image “C” is the same as “B”, but cannot be retrieved in the top 1 or 2 matches. On the contrary, if we use Euclidean distance measure for the small number of images filtered by the boundary, the image “C” can usually be retrieved in the top matches. In other words, the merit of the Euclidean distance measure is lost if merely the DFB metric is used.

VII IMAGE REPRESENTATION

Color information is one of the important features for image retrieval [68]. We use the HSV color space since it provides the best retrieval performance for color histograms. The color histogram is quantized to 256 levels, which results in 256 features for each image. Color moments constitute another kind of color features, which are simple yet effective for image retrieval and do not require quantization. The first three order moments are calculated in the HSV space of each image, resulting in a feature vector of dimension 9. In addition, color coherence vectors (CCV) is used to incorporate spatial information into color histogram representation. The CCV features with 64 quantization result in a 128-dimensional feature vector. Texture is another type of low-level image feature. The Tamura features are designed based on the psychological studies in human visual

All Rights Reserved © 2012 IJARCET
perceptions of textures. We compute the coarseness histogram with 10 quantization levels, and the histogram of directionality with 8 quantization levels. Another one is the wavelet coefficients. The pyramidal wavelet transform (PWT) is used and the mean and standard deviation of the energy distribution are calculated corresponding to each of the sub-bands at each decomposition level. For three-level decomposition, PWT results in a feature vector with () components. We concatenate all color and texture features into one vector (with normalization) to represent each image.

VIII DISCRETE SINE TRANSFORM

In mathematics, the discrete sine transform (DST) is a Fourier-related transform similar to the discrete Fourier transform (DFT), but using a purely real matrix. It is equivalent to the imaginary parts of a DFT of roughly twice the length, operating on real data with odd symmetry (since the Fourier transform of a real and odd function is imaginary and odd), where in some variants the input and/or output data are shifted by half a sample.

\[ X_k = \sum_{n=0}^{N-1} x_n \sin \left( \frac{\pi}{N+1} (n+1)(k+1) \right) \quad k = 0, \ldots, N - 1 \]

IX CLASSIFIERS

In simple terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 4" in diameter. Even if these features depend on each other or upon the existence of the other features, a naive Bayes classifier considers all of these properties to independently contribute to the probability that this fruit is an apple. Abstractly, the probability model for a classifier is a conditional model

\[ p(C|F_1, \ldots, F_n) \]

over a dependent class variable \( C \) with a small number of outcomes or classes, conditional on several feature variables \( F_1 \) through \( F_n \). The problem is that if the number of features \( n \) is large or when a feature can take on a large number of values, then basing such a model on probability tables is infeasible. We therefore reformulate the model to make it more tractable.

Using Bayes’ theorem, we write

\[ p(C|F_1, \ldots, F_n) = \frac{p(C)p(F_1, \ldots, F_n|C)}{p(F_1, \ldots, F_n)} \]

In plain English the above equation can be written as

\[ \text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}} \]

X. PRECISION AND RECALL

Once the feature vectors are generated for all images in the database, they are stored in a feature database. A feature vector of query image of each class is calculated to search the feature database. The image with exact match with the minimum Euclidian distance \([32][33][34][35]\) is considered. The sorted Euclidian distance between the query image and the database images feature vectors are used to calculate the precision \([36]\) and recall to measure the retrieval performance of the algorithm. As it is shown in the equation (2) and (3). Once the query image of a class is taken The retrieved images are sorted in terms of increasing Euclidian distance between the feature vectors of the query image and the database images. Using equations (2) and (3) the precision and recall is plotted against the number of images retrieved.

\[ \text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}} \tag{2} \]

\[ \text{Recall} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images in database}} \tag{3} \]

**Figure 2** Four Possible Outcomes of CBIR Experiment

The cartoon image shown in the Fig.3 is taken as the query Image to search the given image database. The algorithm applied to generate feature vector of complex numbers for each image in the database and the query image.

XI. CONCLUSION

We investigate retrieval of images from a database containing noisy and regular images using down sampled discrete sine transform. Images were also rotated to investigate the performance of our methodology. Results obtained by classifying using Naïve Bayes classifier is fairly good. The performance of DST sectorization with augmentation for both planes gives good result of retrieval on average 45% when
using the Euclidean distance as similarity measure and 46% when using the sum of absolute difference as similarity measure. Thus it is advisable to use sum of absolute difference as similarity measure because of its simplicity and less computational complexity as compared to Euclidean distance.

REFERENCES


[33] Frank Y. Shih and Christopher C. Pu, “A maxima-


AUTHORS

[1] VIVEK KUMAR SONI

He received his MCA. Degree in Computer Application from Bhilai Institute of Technology ,Durg(C.G) under Chhattisgarh Swami vivekanand Technical University Bhilai (C.G) in 2009 and pursuing M.Tech in Computer Science and Engineering from Chhatrapati Shivaji Institute of Technology Durg(C.G) under Chhattisgarh Swami Vivekanand Technical University, Bhilai (C.G).

[2] MRS SHIKHA AGRAWAL

She received her B.E. degree Computer Science and Engineering from Chhatrapati Shivaji Institute of Technology Durg (C.G) under Chhattisgarh Swami Vivekanand Technical University, Bhilai (C.G) in 2009. M.E. Computer Science and Engineering from Shri Shankaracharya College of Engineering and Technology Bhilai (C.G) under Chhattisgarh Swami Vivekanand Technical University, Bhilai (C.G) in 2011. Presently working as Assistant professor at Department of Computer Science and Engineering CSIT, Durg (CG) with an experience of 1 years of teaching field. Her research area of interests includes Digital Image Processing and Data Mining.