A SURVEY ON CONTENT BASED IMAGE RETRIEVAL
Komal Mirchandani 1, T.Rajani Mangala 2
1Department of Electronics and Telecommunication Engineering
2Department of Electronics Engineering VESIT University of Mumbai, India

ABSTRACT:
Content based image retrieval is very popular for browsing, searching and retrieving images from a large database of images as it requires less human intervention. Extraction of features and its demonstration from large database is the major issue in CBIR. In order to improve the retrieval accuracy of CBIR system, research focus has been shifted from designing sophisticated low-level feature extraction algorithm to reduce semantic gap between visual feature and richness of human semantics. In this paper different category in reducing semantic gap is discussed. In addition, some image test bed and retrieval performance evaluation is also discussed. This paper provides a comprehensive survey of all the aspects.

Keywords: CBIR, semantic gap, similarity measures, precision, recall

[1] INTRODUCTION
As the increasing use and availability of devices that capture images like digital cameras, image scanners the collection of digital images is also increasing. Various domains include features like remote sensing, fashion, crime prevention, publishing, architecture which give rise to image retrieval system As manually enter keywords may not capture each and every detail of the image, it makes Content Based Image Retrieval very important [1]. Database management system (DBMS) can also be used to perform image retrieval but has the disadvantage that it includes certain level of human labour for manual annotation and annotation accuracy due to the change in subjectivity of human perception [2, 3, 78].

A new idea was published by Chang which represented picture indexing and abstraction approach for retrieval of images [4]. Few systems have developed like Query by image content (QBIC) [5], Photo book [6], Virage [7], Visual SEEK [8], Netra [9], Simplicity [10].

[1.1] The semantic gap
Low level features include colour, texture, shape, spatial layout etc. [11]. High level features are used to measure similarity. Colour, texture and shape features find it difficult to model image semantics which deals with large image databases [12]. In [2] Eakins has discussed three levels of queries: level 1 includes retrieval of features such as color, texture,
shape or spatial location of image elements. For example, ‘find pictures like this’. Level 2 includes retrieval of objects to be identified by derived features with some degree of logical inference. For example, ‘find a picture of a flower’. And level 3 includes retrieval by abstract attributes with some amount of high level reasoning like named events or pictures with religious significance. For example, ‘find pictures of joyful crowd’. The gap in levels 1 and 2 is semantic gap and level 2 and 3 are semantic image retrieval [74]. There is a need to bridge the “semantic gap” between image features and human semantics to support high level concepts [13].

1.2 High level semantic based image retrieval

There are five techniques to reduce semantic gap which are mentioned: (1) object ontology to define high level concepts, (2) machine learning tools to connect low level features with query images, (3) relevance feedback(RF), (4) semantic template to support high level images, (5) WWW(web) image retrieval [1]. Retrieval is generally performed at level 2 whereas at level 3 it is difficult and less common. Level 3 is generally found at specific areas of domain like art museums, newspaper gallery etc.

2 Low- level image features

Currently CBIR systems are region based where it is found that users are more interested in specific regions. Image features can either be extracted from entire regions or from specific one. The first step to perform region based image retrieval (RBIR) is to implement image segmentation. Low level features like color, texture, shape or spatial location can also be extracted from segmented regions. Then the similarity between two images is defined which is based on region features. In this paper a brief discussion of what are used in Region based image retrieval (RBIR) system with high level semantics is mentioned.

2.1 Image segmentation

Many techniques of image segmentation is used in the past, few of them mentioned as (1) curve evolution [14], (2) energy diffusion [15], (3) graph partitioning [16]. Images which are high in definition to both colour and texture are considered as mosaic. Texture plays an important role in high level features but is difficult when segmentation method is used.

Figure: 1. JSEG segmentation results

But these problems are overcome by ‘JSEG’ (Japan Society of Engineering Geology) segmentation which is explained in two steps. In first, image color consists of several classes which are replaced by color class labels to obtain class map of an image. Then spatial segmentation is performed to obtain texture composition and this is used in many systems [17]. In second, Blobworld segmentation which is obtained by clustering of pixels in color texture feature and which is modelled with a mixture of Gaussians and the expectation maximization
(EM) is used which in result provide the segmentation of the image [18]. JSEG and Blobworld segmentation are widely used but also are difficult to conclude which algorithm gives better results.

[2.2] Low-level image features

In this topic we discuss about features used in RBIR systems with high level semantics.

[2.2.1] Color feature

Color plays an important role in searching images of arbitrary subject matter and in human perception mechanism. Color features are classified in two categories: one is color histogram and other is statistical method. Color spaces used in RBIR system are RGB (red, green and blue used in colour monitors and cameras), LAB (L*a*b, lightness, a and b are two colour dimensions from green to red and from blue to yellow), HSV (hue, saturation, value), YCrCb (luminance, chroma red, chroma blue) and the hue-min-max-difference (HMMD) [17, 19, 71, 75]. Common color features in RBIR system include color covariance matrix, color histogram, color moments and color coherence vector [20]. Some color features have effectively described colors but are not directly related to high level semantics. Selection of color feature is important as it depends on segmentation results. Like, if the segmentation gives a color which does not have homogeneous color, then the color is not a good choice. In [17] HSV space is defined as color of region.

[2.2.2] Texture feature

In image retrieval texture feature is used in high level semantics and which describes the content of real world images like sky, trees etc. One of the few methods for representing texture features was proposed by Haralick et al. [21] called Grey level co-occurrence matrices (GLCM). Commonly used texture features in image retrieval are obtained using Gabor filtering [9], wavelet transform [10], and local statistical features like six Tamura texture features [22], and Wold features which is proposed by Liu et al. [23]. Six features include: coarseness, directionality, regularity are more significant while contrast, line likeness, contrast and roughness are not much effectiveness. First three features are employed by MPEG-7 as texture browsing descriptors [24]. Periodicity, randomness and directionality are considered to be a part of Wold features which have proved to work effectively on Brodatz textures [25, 72]. Brodatz textures are defined as the most commonly used texture dataset, especially in computer vision and signal processing community [72].

Due to its limitation, wold features are also affected by image distortion like scale which results in distortion. But as these are working on Brodatz features, they are proved to be less effective [26]. Among all the texture features only Gabor and wavelet features are used for image retrieval, also they live upto the standard to match the results of human vision study [9,10,24]. In RBIR systems, regions are of arbitrary shapes but Gabor filter and wavelet transform are designed for rectangular images. So the question arises as how the texture features will be extracted from arbitrary shaped RBIR systems?

Texture features are obtained by the property of pixels contained in the region [15,80]. Huang and Dai et al. [27] have proposed a method for gradient vector of images as texture feature.
[2.2.3] Shape

Shape features include aspect ratio, circularity, Fourier descriptors, moment invariants, consecutive boundary segments [28]. Shape features are specially used in manmade objects. Despite having difficulty in using shape features it has shown benefits in RBIR systems. As an example in shape features eccentricity and orientation are used [29]. As stated in [30] second order moments and gross shape inertia is used. MPEG-7 consist of three shape descriptors for image retrieval where one is 3D shape descriptors which is derived from 3D meshes, second is original shape descriptor which is derived from Zernik moments and third is contour based shape which is derived from curvature scale space(CSS) [24]. Mokhtarian and Abbasi et al. [30] have extended the idea of CSS descriptor to be robust which is close to general shape descriptors.

[2.2.4] Spatial location

Spatial location also plays an important role in region classification. For example, ‘sky’ and ‘sea’ do have their color and texture same but their spatial location differs as sky is towards up and sea is directly at the bottom [76]. Spatial location includes top, bottom, up, down according to the location of an image [31,32]. In spatial location, centre of its region is represented by its spatial location. Two types of relationships exist in spatial location: one is relative which tends to be more important than absolute which helps in deriving semantic features. 2D strings are used to represent relationships between objects such as ‘left, right, below, above’ [33].

However these directional relationships are not sufficient to represent the content of images. So to support image retrieval based on semantic context modeling algorithm is explained in [34]. This modeling consist of six spatial relationships: left, right, up, down, touch, front. Smith et al. [35] proposed a method of Composite region template (CRT) where each semantic class is characterized by CRT to obtain the spatial location from collection of images.

[2.3] Similarity measure

Similarity of an image is measured at two levels in RBIR system: one is region level which is use to measure the distance between two images and second is image level which is use to measure similarity of images which contain different number of regions. Researchers prefer Minkowski-type metric to define region distance [77]. Let us consider two regions \( x \) and \( y \) represented by \( p \) dimensional vector \((x_1, x_2, ..., x_p)\)\((y_1, y_2, ..., y_p)\). So the Minkowski metric is defined by

\[
d(x, Y) = \left( \sum_{i=1}^{p} |x_i - y_i|^r \right)^{1/r}
\]

Where \( r=2 \) it is known as Euclidean distance and when \( r=1 \) it is known as Manhattan distance. Canberra distance, Angular distance, Czekanowski coefficient [36], inner product, dice coefficient, cosine coefficient and Jaccard coefficient [37]. There are two ways to measure the similarity of an image: one is One-One match [38] and second is Many-Many match [39].

Li et al. [40] proposed a method on integrated region matching (IRM) which is used for matching a region of one image to other images and results in decreasing the impact of inaccurate segmentation. Minkowski metric is used in systems to measure region distance. But few questions still remain unsolved as one of it is how to measure perceptual similarity. In [41], by choosing smaller amount of dimensions Dynamic partial distance function (DPF) reduces the dimension of feature vectors. However DPF proved to give more accurate results than
Minkowski metrics [77] but to be used in image retrieval research is still going on, to compare its performance in various applications.

Perceptual distance for shape similarity measure is discussed in [42] which consist of set of tokens. First the metric distance is defined between tokens then the non-metric distance. Vasconcelos and Lippmann et al. [43] proposed a method of Multiresolution Manifold Distance (MRMD) for face recognition. In [44] similarity measure is explained and the advantage of this measurement is that it can be combined into integrated future.


According to number of users semantic gap is classified in different ways. For example in semantic gap if we consider the application domain, they can be classified at artwork retrieval [45], scenery image retrieval [46,47,48], WWW images retrieval [49], etc. Different techniques used to derive high level semantics are mentioned: (1) using object ontology to define high level concepts [1,48,50,51,52]. (2) Machine learning methods to associate low level features with query concepts [1,3,46,53]. (3) Relevance feedback (RF) [1,48,54,55,79]. (4) Semantic template [1,35,56,57]. (5) Web image retrieval [1,53,56,58].

Object ontology is easy in designing and suitable for applications. Generally machine learning technique is preferred for complex semantics. Due to its simple design in mapping from low level features to high level semantics, decision tree is applied for image retrieval if the learning methods are well organized. RF is effective to enhance the accuracy in image retrieval. Current systems require five or more repetition before it can deliver a stable performance, but many users give up at second or third trial [54,55,59,60]. ST technique is used to reduce the ‘semantic gap’. Web image retrieval is a search engine. Techniques are incorporated with one another in order to implement semantic based image retrieval. In Ref. [46,48,56,79] RF is combined with machine learning and object ontology.

In Ref. [40], interesting work is done, where the database images are classified in texture and non-texture, graph and auto-graph. In Ref. [73], an approach is done to human perception by central object of an image as it is considered that the most interesting object is always taken in the centre of the frame while clicking.

[4] Image database and performance evaluation

[4.1] Image databases

In past few decades many systems use Corel image database [61] while others use set of self-collected images like Kodak database [63], Brodatz texture [25, 72] to test the performance. Many researchers use natural scenery as it is easier to analyze and the reason is two-fold. Corel database contains a large database as it is categorized by size 100 because of its large size and ground truth [62]. Though Corel image database is not suitable for CBIR performance evaluation but is still widely used. In [63] a new algorithm is used for image retrieval to overcome the problems of Corel dataset.

The main focus in selecting the categories are they should be well defined and be meaningful. In [47] few categories are mentioned like: brick, grass, road, skin. In [64] semantic categories like: desert, beach, building, park. In Ref. [65] high level features are mentioned like: tree, water wave, ground. It is found that humans can only recognize 5000-30,000 categories. In
[66] an approach is made by Bayesian algorithm to recognize 101 object categories where it is believed that images are classified not more than 10-20 categories.

[4.2] Performance evaluation

Precision and recall are used to measure the performance evaluation in CBIR systems [67]. Recall measures the ability of the system to all the models that are relevant and is given by,

\[ \text{Recall} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images}} = \frac{A}{A + C} \]

......(2)

Precision measures the ability of the system to retrieve only the models that are relevant and is given by,

\[ \text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}} = \frac{A}{A + E} \]

......(3)

Where A represent the number of relevant images that are retrieved, B represents number of irrelevant items and C represents number of relevant items those were not retrieved [75].

It is ‘ideal’ that both precision and recall should be high and because of it they are represented as joint precision recall curve Pr(Re)curve to evaluate the performance of CBIR [62]. Many researchers prefer precision scope curve as many a times recall is often low in color images when retrieved [68]. Precision scope (PrSc) is defined as,

\[ \text{Pr (Sc)} = \frac{\text{Nr}}{\text{Sc}} \]

......(4)

Where Nr represents number of relevant images retrieved and Sc presents number of images returned to user.

Another method to evaluate the performance is rank (Ra) measure [68-70] and is stated as average rank of retrieved images. It is noted that the smaller the rank the better the performance.

[5] CONCLUSION

It is concluded that the main focus is generally on image processing and low level feature extraction but these features find it difficult to describe high level semantic concepts. The main focus of CBIR should be bridging the semantic gap between low level features and human semantics. This paper provides the study of reducing the semantic gap by different techniques. Current techniques in similarity measures are also discussed. Minkowski metric cannot capture human perception so perceptual image similarity is to be focussed further. Dataset and performance evaluation are also discussed.

To implement a good image retrieval system for users few requirements like integration of low level feature extraction, high level semantics, user friendly interface, indexing tool. A CBIR system with all the features which provide a balance view is a need.
REFERENCES


52. S. Kulkarni, B. Verma, Fuzzy logic for texture queries in CBIR, Proceedings of the International Conference on Computational Intelligence and Multimedia Applications (ICCIMA), Xi’an, China, 2003, pp. 223–226.


64. Z. Yang, C.-C. Jay Kuo, Learning image similarities and categories from content analysis and relevance feedback, Proceedings of the ACM Multimedia Workshops, 2000, pp. 175–178.


Author’s brief Introduction

Ms Komal Mirchandani has done diploma in the field of Industrial Electronics in 2010 and Electronics Engineering in 2013 from North Maharashtra University, Dhule. Currently I am pursuing Masters in Electronics and Telecommunication Engineering from Mumbai University and my area of interest includes image processing, feature extraction, and image searching in large database

T. Rajani Mangala, received the B.E. and M.E. degree in Electronics and Telecommunication from University of Mumbai, in 1994 and 2005. Ph.D. degree in the field of Digital Image Processing from NMIMS, Mumbai in May 2013. She is currently working as Professor in Electronics department of VES Institute of Technology. Areas of interest include digital image processing, wireless networks, 3G and 4G.