Survey paper on Sketch Based and Content Based Image Retrieval

Prachi A. Gaidhani\(^1\), Prof. Bagal S. B.\(^2\)

\(^1\)Department of Electronics & Telecommunication, KCT’s Late G.N. Sapkal College of Engineering, Ajaneri, Nashik, India.

\(^2\)Department of Electronics & Telecommunication, KCT’s Late G.N. Sapkal College of Engineering, Ajaneri, Nashik, India

Abstract: This survey paper presents an overview of development of Sketch Based Image Retrieval (SBIR) and Content based image retrieval (CBIR) in the past few years. There is awful growth in bulk of images as well as the far-flung application in too many fields. The main attributes to represent as well index the images are color, shape, texture, spatial layout. These features of images are extracted to check similarity among the images. Generation of special query is the main problem of content based image retrieval. Features are extracted from the images and the similarity between them is check by using different algorithms. Sketch based image retrieval is effective and chief method which are not necessarily having a high skill to draw the query sketch. Over last twenty years different algorithms and models are explored for the retrieval of images. In concluding section we present the limitations of the current image retrieval algorithms.

Keywords: Sketch based image retrieval, Content based image retrieval, Feature extraction, Gradient field histogram of oriented graph, image descriptor

1. Introduction

Considerable amount of research efforts have been paid in covering the Sketch Based Image Retrieval (SBIR) and Content Based Image Retrieval (CBIR) problems. An image retrieval system returns a set of images from a collection of images in the database to meet user's demand with similarity evaluations such as image content similarity, color similarity, edge pattern similarity, etc. An image retrieval system offers an efficient way to access, browse, and retrieve a set of similar images in the real-time applications. Long before the invention of writing systems, people drew and sketched to communicate with each other, and the oldest-known forms of writing were primarily logo-graphic in nature. Generally, a sketch depicts the rough shape of an object and provides a conceptual representation to facilitate communication. We can easily recognize objects from other people’s sketches, and this form of expression is arguably the most universal communication tool for people who speak different languages. Compared with keywords, a sketch is generally more natural and more informative, breaking down the language barrier. Sketch-based image retrieval (SBIR) can therefore be a very valuable information search tool, supplementary to keywords-based search. The benefits of SBIR are becoming obvious with the development of touch-screen devices such as smart phones & tablets. SBIR performance cannot meet the requirements of practical application, however, which drives the need for further improvement.

Content-based image retrieval (CBIR) is the application of computer sight to the image retrieval trouble, so as to the difficulty of searching for digital image in big databases. "Content-based" means with the purpose of explore will analyze the actual contents of the image. Here the term content in this context might submit to shade, figure, surface, or some other information that can be derived from the image itself [1]. For CBIR technology few strong applications could be identified as architecture design, art & craft museums, archaeology, medical imaging and geographic info system, trademark databases, weather forecast, criminal investigations, image classification, image search over the internet and remote sensing field for indexing biomedical images by contents[2,3,4].

2. Literature Survey

The main aspect of image retrieval system is to offer an efficient way to access, browse, and find similar images in the real-time applications. Various methods have been proposed in the literature for sketch-based and content based Retrieval. Since the core of SBIR is shape matching, many classic descriptors which are widely used in matching (recognition) tasks can be adopted in SBIR with minor adjustments. A substantial amount of research attempts have been committed in addressing the Content Based Image Retrieval (CBIR) problem [5].

2.1 Sketch-based image retrieval (SBIR)

Sketch-based image retrieval (SBIR) is a relevant means of querying large image databases. All of researches focus on how to solve the gap between sketch and image matching problem. A lot of ways are discussed or discovered about this gap. Recently, we have reviewed bellow a method that deal with the main three points of sketch based image retrieval.

Application area of sketch based information retrieval is the searching of analog circuit graphs from big database [19]. The user has to make a sketch of the analog circuit, and the system can provide many similar circuits from the database. The Sketch-based image retrieval (SBIR) was introduced in QBIC [20] and Visual SEEK [21] systems. In these systems the user draws color sketches and blobs on the drawing area. The images were divided into grids, and the color and texture features were determined in these grids. The applications of...
grids were also used in other algorithms, for example in the edge histogram descriptor (EHD) method [22]. The disadvantage of these methods is that they are not invariant opposite rotation, scaling and translation. Lately the development of difficult and robust descriptors was emphasized. Another research approach is the application of fuzzy logic or neural networks. In these cases the purpose of the investment is the determination of suitable weights of image features [23].

2.1.1 Sketch-based image retrieval System

Still even if the compute of research in sketch-based image retrieval increases, there is no widely old SBIR scheme. Our aim is to enlarge a content-based associative investigate engine, which records are accessible for anybody looking back to unguided sketch. The client has a diagram area, where he preserves all outline and instant, which are predictable to take place in the given place and with a given size. The retrieval results are grouped by color for superior clearness. Our mainly vital task is to overpass the information gap between the drawing and the image, which is assist by own preprocessing alteration process. In our organization the iteration of the consumption process is probable, by the existing outcome looking again, thus increasing the precision. The system building blocks include a preprocessing subsystem, which remove the troubles caused by the multiplicity of metaphors. Using the attribute vector generating subsystem our image can be represented by numbers considering a given property.

![Figure 1: The global structure of the system [6]](image)

The database management subsystem provides an interface between the database and the agenda. Bottom on the feature vectors and the model image the retrieval subsystem provides the response list for the user using the displaying subsystem (GUI). The global structure of the system is shown in Fig. 1.

Early sketch based image retrieval systems were typically driven by queries comprising blobs of color or predefined texture [24] [25]. Later systems explored shape descriptors [26] and spectral descriptors such as wavelets [27]. Eitz et al. [28] introduced a grid based approach to shape retrieval, dividing the image into regular grids and locate photos using sketched depiction of object shape. Descriptors from each cell were concatenated to form a global image feature. However this offers limited invariance to changes in position, scale or orientation. A depiction invariant descriptor which encapsulates local spatial structure in the sketch and facilitates efficient codebook based retrieval was proposed by Hu et al. [29]. This descriptor is able to mitigate the lack of spatial information within a BoVW representation by capturing structure from surrounding regions using a multi-scale HoG descriptor computed over a gradient field interpolated from the orientations of strong Canny edges (GF-HOG). Eitz et al. [30] later computed HoG over Canny edges (SHOG) for BoVW though did not interpolate orientations from edges.

Sciascio et al. [26] investigate extracting shape feature in photo-realistic images using image segmentation. However, in addition to the problem of unstable regions produced by texture-based segmentation algorithm, a particular object can often be either under segmented or over-segmented, failing to produce an ideal semantically coherent region. Hoiem et al. [31] estimate the coarse geometric properties of a single image by learning local appearance and geometric cues on super-pixels even in cluttered natural scenes. Russell et al. [32] developed an algorithm that finds and segments visual topics within an unlabeled collection of images by combining multiple candidate segmentations with probabilistic document analysis methods. Todorovic and Ahuja [33] represent objects as region trees and combine structural cues of the trees for matching. Guet al. [34] present a unified framework for object detection, segmentation, and classification using robust overlaid regions produced by a novel region segmentation algorithm [35]. To the best of our knowledge, our approach is the first technique to apply multi-scale region segmentation within a BoVW framework for sketch based image retrieval.

2.1.2 Gradient Field Descriptor for Sketch Based Retrieval

This system accepts monochrome free-hand sketched queries describing a shape, and returns images that contain similar shapes. This requires a matching process robust to depictive inaccuracy (e.g. in location, scale, or shape deformation) and photometric variation. The approach is to transform database images into canny edge maps, and capture local structure in the map using a novel descriptor. We recommend setting an appropriate scale and hysteresis threshold for the canny operator by searching the parameter space for a binary edge map in which a small, fixed percent of pixels are confidential edging. These easy heuristics remove central boundaries and discourages response at the scale of finer texture. This paper introduces the Gradient Field HoG (GF-HOG) descriptor; an adaptation of HoG that mitigates the lack of relative spatial information within BoW by capturing structure from surrounding regions. We are inspired by work on image completion (in-painting) capable of propagating image structure into voids, and use a similar “Poisson filling” approach to improve the richness of information in the gradient field prior to sampling with the HoG descriptor [7]. This simple technique yields significant improvements in performance when matching sketches to photos, compared to three leading descriptors: Self-Similarity Descriptor (SSIM); SIFT; and HoG. Furthermore we show how the descriptor can be applied to localize sketched objects within the retrieved images, and demonstrate this functionality through
a sketch driven photo montage application. The success of the descriptor is dependent on correct selection of scale during edge extraction, and use of image salience measures may benefit this process. The system could be enhanced by exploring colored sketches, or incorporate more flexible models for object localization.

2.2 Content Based Image Retrieval (CBIR)

(CBIR) is a technique that used to view image features like (color, shape, texture) to find a query image in a large size of the database. The retrieval processes image, including, low level (content based features) and high level (semantic based features). The difficulties of CBIR lie in reducing the differences of contents based feature and the semantic based features. This problem in giving efficient retrieval images guide the researchers to use (CBIR) system, to take global color and texture features to reach, the better retrieval, where others used local color and texture features [8].

A new type of CBIR approach is presented in [9], in which the spatial pyramid and order less bag-of features image representation were employed for recognizing the scene categories of images from a huge database. This method offers a promising result and outperforms the former existing methods in terms of the natural scene classification. The method in [10] presented the holistic representation of spatial envelop with a very low dimensionality for representing the scene image. This approach presented an outstanding result in the scene categorization. The method in [11] proposed a new approach for image classification with the receptive field design and the concept of over-completeness methodology to achieve a preferable result. As reported in [11], this method achieved the best classification performance with much lower feature dimensionality compared to that of the former schemes in image classification task.

Tandon et al. developed a CBIR system called FISH- Fast Image Search in Huge Databases which learns the relevance of image features based on user feedback. Experimental results prove that the FISH system can seamlessly scale to huge image databases and maintain the interactive response time. Short and Long term learning of the intent of a user within a session is used to retrieve images with higher accuracy and efficiency. In this system the users provide the query image through the user interface, which is a web-front end. The query image is then processed and an appropriate index structure is used for a quick search of similar images. The retrieved images are then displayed to the user. User can give a feedback whether the retrieved images are similar to the query image. The feedback provided by the user is used for short term learning by the system to provide better results in the next iteration. This learned knowledge is stored in the long term memory to provide better retrieval results in the future queries [36].

A CBIR system [PATSEEK] for US-based patent database was developed by Tiwari et al. as a patent always consists of an image along with textual information. The user has to provide keywords along with the query image that might appear in the text of patents whose images would be searched for similarity. The image grabber searches the patent database on a certain criteria and provides a page image which consists of more than one image. Connected components or blocks are identified using horizontal scans and the vertical scans. Feature vectors are then calculated from these separated images using Edge Orientation Autocorrelogram (EOAC). Magnitude and gradient of the edges are found using the canny edge operator. The gradient of edges is then quantized and the direction fundamensal orientation autocorrelogram is stored in the database along with the patent number and the page number where the image was found. These feature vectors stored in the database are then used in searching the similarity using the L1 and L2 distance measures. Experimentally, PATSEEK obtained 100% recall rate on 61% of query images [37].

Xiang-Yang Wang et al. developed an effective color image retrieval method combining the color, shape and texture features of the image. They used the modified version of Dominant color descriptors to extract the color feature. In this method some dominant colors and their percentages are obtained after predetermining the image using the Fast Color Quantization algorithm with clusters merging. Steerable Filters were used to extract the spatial texture feature of the image. In this method the filters having arbitrary orientations are synthesized. Then the image is split into orientation subbands to detect the edges located at various orientations in the image. Pseudo-Zernike Moments based shape descriptors were used to extract the shape feature as their magnitudes are invariant to the image rotation and are less sensitive to image noise. Also, they have multilevel representation capabilities. The performance of the developed CBIR system was tested using the Corel image database and the results prove that the method has higher accuracy and efficiency levels compared to methods such as Color Histogram and Color Histogram of subblocks [39].

Support vector machines (SVM) are extensively used to learn from relevance feedback due to their capability of effectively tackling the above difficulties. However, the performances of...
SVM depend on the tuning of a number of parameters. It is a different approach based on the nearest neighbor paradigm. Each image is ranked according to a relevance score depending on nearest neighbor distances. This approach allows recalling a higher percentage of images with respect to SVM-based techniques [14] there after quotient space granularity computing theory into image retrieval field, clarify the granularity thinking in image retrieval, and a novel image retrieval method is imported. Firstly, aiming at the Different behaviors under different granularities, obtain color features under different granularities, achieve different quotient spaces; secondly, do the attribute combination to the obtained quotient spaces according to the quotient space granularity combination principle; and then realize image retrieval using the combined attribute function. [15] Then a combination of three feature extraction methods namely color, texture, and edge histogram descriptor is reviewed. There is a provision to add new features in future for better retrieval efficiency. Any combination of these methods, which is more appropriate for the application, can be used for retrieval. This is provided through User Interface (UI) in the form of relevance feedback. The image properties analyzed in this work are by using computer vision and image processing algorithms. For color the histogram of images are computed, for texture co occurrence matrix based entropy, energy, etc, are calculated and for edge density it is Edge Histogram Descriptor (EHD) that is found[16] After that local patterns constrained image histograms (LPCIH) for efficient image retrieval is presented. Extracting information through combining local texture patterns with global image histogram, LPCIH is an effective image feature representation method with a flexible image segmentation process. This kind of feature representation is robust and invariant for several images transforms, such as rotation, scaling and damaging [17]. In another system the image is represented by a Fuzzy Attributed Relational Graph (FARG) that describes each object in the image, its attributes and spatial relation. The texture and color attributes are computed in a way that model the Human Vision System (HSV) [18].

2.2.1 Yet another Content Based Image Retrieval
A Content Based Image Retrieval system generally consists of two main phases as an indexation phase and a retrieval phase. Here indexation is done off-line and Retrieval process is on-line. Images are indexed using the physical characteristics like color, texture and shape of each image in the database. These descriptors are extracted automatically from the image content. The query is an image example. The results are images from the database similar to the query image according to predefined criteria. Choosing good indexes is thus a very important matter.

The Yet another Content Based Image Retrieval system combines three characteristics like color, texture and points of interest of an image to compute a weighted similarity measure [12]. Color and texture characteristics are global while points of interest are shape local characteristics.

The color characteristic is widely used in generalist CBIR systems. The indexer module quantifies colors and creates histogram specific to each image. The similarity measure will be made on histograms and then color similarity $S_c$ is calculated. If the value of $S_c$ is 1 that means the two images has same color. For the texture characteristic four descriptors like contrast, entropy, energy and inverse differential moment are used. In this they have used Euclidian distance to find out the similarity of texture $S_t$. The value $S_t = 1$ means that the two images have same texture. The point of interest analyzer is based on the Harris detector. The number of correct matching points between two images will quantify the similarity measure $S_s$. If this value is low, we have a bad similarity. If this value is high relatively to the total number of interest points, we have a good similarity. The similarity measure used in YACBIR is a sum of weighted color, texture and points of interest (shape) similarity measures. This similarity measure is given by:

$$S = \alpha . S_c + \beta . S_t + \gamma . S_s \text{with } (\alpha + \beta + \gamma) = 1$$

CBIR systems are various and diverse. There is a variety of physical characteristics used to index images. A system can use region histogram while another uses color coherence vector. For some systems there are no available details. The collection of images used in the tests can influence the results. With a given CBIR system, searching for example for a cat image in a database containing only dogs always yields images of dogs. To have a more powerful and efficient retrieval system for image and compact disk record, contented stand question must be shared with wording and keyword predicates.

2.2.2 User Oriented Image Retrieval System Based on Interactive Genetic Algorithm
In this method, a user-oriented mechanism for CBIR method based on an interactive genetic algorithm (IGA) is anticipated. Color characteristic like the signify rate, the ordinary departure, and the figure bitmap of a color image are used as the features for recovery. In count, the entropy support on the older point co-occurrence matrix and the edge histogram of an image is also considered as the surface features. In addition, to decrease the break between the recovery outcome and the users’ hope, the IGA is working to assist the users recognize the images that are most satisfied to the users’ want. Untried outcome and judgment show the feasibility of the proposed approach [13].

3. Discussions
The above studied systems for Content-based image retrieval (CBIR) and Sketch-based image retrieval (SBIR) in the literature survey are giving good results but they are carrying few disadvantages with them like for Content-based image retrieval (CBIR) the traditional block truncation coding (BTC) derives the low and high mean values by preserving the first-order moment and second-order moment over each image block, which requires additional computational time. Most existing descriptors for Sketch based image retrieval (SBIR), whether derived from classic descriptors, such as SHoG and Gradient field histogram of oriented graph (GF-HOG), or specifically designed for Sketch-based image retrieval (SBIR), such as APAI and Structure Tensor, treat sketches/extracted edges as a set of points. They were designed to capture the pixel-level features from image
patches. It is difficult to isolate the features that correspond to certain edges from the whole patch feature vector. For SBIR current methods show poor effectiveness under scale, translation, and rotation changes and Content based image retrieval systems have two disadvantages i.e. it requires huge amount of human labor in the manual annotation when the image collection is large. It is hard to precisely annotate the rich content of an image by humans due to perception subjectivity.

4. Conclusion

The concept of content-based and sketch based image retrieval has been discussed at length in this paper. The past and the current systems for SBIR and CBIR have been reviewed in terms of the techniques used, their accuracy and actual results. Many journal papers have been researched to explain in deep the fundamental aspects of SBIR and CBIR. Future research directions have been suggested and open research issues identified in this paper.

References


[17] Zhi-Gang Fan, Jilin Li, Bo Wu, and Yadong Wu “Local Patterns Constrained Image Histograms for Image Retrieval” ICIP 2008 IEEE.


