A Scalable Sketch Based Image Retrieval System

Kathy Khaing, SaiMaungMaungZaw, Nyein Aye
kathrine.truth@gmail.com, saisaimmz@gmail.com, nyeinaye@gmail.com

University of Computer Studies, Mandalay, Myanmar

Abstract- Due to the progress in digital imaging technology, image retrieval (IR) has become a very active research area in computer science. Although many researches are increased in Sketch Based Image Retrieval (SBIR) field, it is still difficult to bridge the gap between image and sketch matching problem. Therefore, this paper presents a scalable SBIR system and contributes to get more efficient retrieval result. The features of both the query sketch and database images are extracted by Scale Invariant Feature Transform (SIFT) algorithm. Then the cropped keypoint images are processed by Canny edge detection. After blocking the edge image, the matched feature values are get by pixel count ratio. The retrieved images similar with query sketch are displayed by rank. Mean Average Precision (MAP) and Recall rates is measured as evaluation criteria. To evaluate the performance of this system, the benchmark sketch dataset of Eitz et al. is used.

Index Terms- Canny, IR, keypoint, MAP, SBIR, SIFT,

I. Introduction

.
With the development of the technology and availability of the image capturing devices such as scanners, digital cameras, the size of the digital image collection increases rapidly. It is important to efficiently store and retrieve images for different applications, for this purpose many Image Retrieval (IR) system have been developed. The image retrieval system returns a set of images which are ranked in a certain order under a similarity function in response to a query given by a user. Early IR researches focused on the content of image such as color, texture and shape and called Content based Image Retrieval (CBIR) system. In CBIR system, example image is required as a query input. If the user has no image to show as an example, it becomes a problem. An easy way to express a user query is a line-based hand drawing, a sketch. Therefore, the recent researches are focused on Sketch based Image Retrieval (SBIR) system but those researches still keep low performance. The most important task is to bridge the gap between database image and query sketch.

The major properties of an image are color, shape, texture and saliency. Among them, shape is the key feature for SBIR system due to the lack of other properties in binary drawings. In sketch based image retrieval system user provided a drawing sketch as input. The feature vector of input sketch compares with feature vector of database images and retrieves the matched image from the database and display on screen as output of the system. The SBIR system is essential and effective in real life such as medical diagnosis, digital library, search engines, crime prevention, geographical information, art gallery and remote sensing systems. In some cases, we can recall our minds with the help of figures or drawings.

This paper is organized in 5 sections.Section 2 gives the literature survey, which gives the information about previously used systems for image retrieval. The background theory of image retrieval is described in section 3. In section 4, the proposed system architecture is explained and discusses the evaluation criteria of the system. Finally, section 5 provides the conclusion of the system.

II. Related Works

This section reviews the suitable background literature and describes the concept of an image retrieval system. Scientific publications included in the literature survey have been chosen in order to build a sufficient background that would help out in solving the research sub-difficulties.

Since Sketch Based Image Retrieval is a young research area, there are a few works on SBIR. J. D. Fendarkar and K. A. Gulve [2] presented the problems and challenges concerned with the design and the creation of CBIR systems, which is based on a free hand sketch (i.e. SBIR). The use of the existing methods, described a possible result, how to design and implement a task specific descriptor, which can handle the informational gap between a sketch and a colored image to make an opportunity for the efficient search. Another edge-based approach is Edge Histogram Descriptor (EHD) that was proposed in the visual part of MPEG-7. The idea is to get a local distribution of five types of edges from the local regions of the image. Eitz, et al. [4] introduced a benchmark for evaluating the performance of large-scale sketch-based image retrieval systems. The necessary data are acquired in a controlled user study where subjects rate how well given sketch/image pairs match. They suggest how to use the data for evaluating performance of sketch based image retrieval systems the benchmark data as well as the large image database are made publicly available for further studies of this type. The dataset of Eitz, et al. consisting of 31 sketches, each with 40 photos ranked by similarity. Furthermore, they develop new descriptors based on the bag-of-features approach and use the benchmark to demonstrate that they significantly outperform other descriptors in the literature.

The majority of SBIR methods are based on histograms of orientations either to compute a global representation or a local one. RuiHua and John Collomossea [7] described HOG seems to be the favorite descriptor in the community. However, because of the sparseness of sketches, HOG descriptors may be also sparse

which may impact negatively in the final effectiveness. To address this problem, Saavedra and Bustos [3] proposed the HELO (histogram of edge local orientations) descriptor where the orientation histogram is formed by local orientations that are estimated by grouping pixels in cells and determining just one representative orientation for each cell. The representative gradient is computed using the square gradient approach.
The proposed SBIR system in this paper is used the SIFT algorithm for feature extraction. Most of SBIR systems used K-means clustering algorithm. Thus, the fuzzy C-means clustering algorithm is developed in this work.

III. Image Retrieval System

The efficiency of searching in information set is a very important point of view. In case of texts user can search flexibly using keywords, but if use images, user cannot apply dynamic methods. Two questions can come up. The first is who yields the keywords. And the second is an image can be well represented by keywords. In many cases if user wants to search efficiently some data have to be recalled. Since the human is visual type, looking for images using other images, and follow this approach also at the categorizing. In this case, users search using some features of images, and these features are the keywords. At this moment unfortunately there are not frequently used retrieval systems, which retrieve images using the non-textual information of a sample image. What can be the reason? One reason may be that the text is a human abstraction of the image. To give some unique and identifiable information to a text is not too difficult. At the images the huge number of data and the management of those cause the problem.

A. Content Based Image Retrieval System (CBIR)

Nowadays the application of internet and www is increasing exponentially and the collection of image accessible by the users is also growing in numbers. During the last decade there has been a rapid increase in volume of image and video collections. A huge amount of information is available, and daily gigabytes of new visual information is generated, stored, and transmitted. However, it is difficult to access this visual information unless it is organized in a way that allows efficient browsing, searching, and retrieval. Traditional methods of indexing images in databases rely on anumber ofdescriptive keywords, associated with each image. However, this manual annotation approach is subjective and recently, due to the rapidly growing database sizes, it is becoming outdated. To overcome these difficulties in the early 1990s, Content-Based Image Retrieval (CBIR) emerged as a promising means for describing and retrieving images. According to its objective, instead of being manually annotated by text-based keywords, images are indexed by their visual contents such as color, shape, texture, and spatial layout. Content Based Image Retrieval is an automatic process based on user input to search relevant images. In the digital image processing, CBIR is most popular and rising research area. The CBIR system first computes the similarity between the query and the images stored in the database. The development of content based image retrieval, for indexing libraries have usually used manual image annotation and then later their image collections are retrieved. However, manual image annotation procedure is an expensive and labor intensive and hence there has been great interest in coming up with retrieving images based on content in an automatic way. To extract the visual content of an image like texture, color, shape or sketch is the goal of CBIR. A typical CBIR process first the image features extracts and efficiently store them. Then it compares with the database images and returns the results [2].The information extracted from the content of query is used for the content based image retrieval information systems. In these systems the keywords are annotated with images and then using text based search method retrieve images. Query by Image Retrieval (QBIR) is also known as content based image retrieval.

The aim of this paper is to develop a content based image retrieval system, which can retrieves images using sketches in frequently used databases.

B. Sketch Based Image Retrieval System (SBIR)

A sketch is s free hand-drawing consisting of a set of strokes. A sketch lacks texture and color. SBIR is part of the image retrieval field. In SBIR system, the input is a simple sketch representing one or more objects. Although a vast number of researchers on multimedia image retrieval are mainly focused on content based image retrieval systems using a regular image as query, in the last few years the interest in the SBIR problem has been increased. This interest may be owed to the emerging touch screen technology that allow users to draw a query directly on a screen, turning the process of making a query easy and accessible.

The Sketch-based image retrieval (SBIR) was introduced in QBIC and Visual SEEK systems. In these systems the user draws color sketches and blobs on the drawing area. The user has a drawing area where he can draw those sketches, which are the base of the retrieval method. The retrieval system using sketches can be essential and effective in our daily life such as Medical diagnosis, digital library, search engines, crime prevention, geographical information, art gallery and remote sensing systems [4].Feature extraction and similarity measure are very dependent on the used features. There would be more than one representation in every feature. Among these representations, to describe features histogram is the most commonly used technique.Describing the edge–based features of shape and texture is important for image retrieval using graphical rough sketches. Although edge-based feature extraction methods that used a Fourier Descriptor can be applied in order to realize image retrieval systems, these methods can be only applicable of features from limited images consisting of close-curve boundaries. Sufficient edge-based feature extractions of shape information have not yet been developed, because effectively extracting the edge-based features of shape information is difficult.

IV. Proposed SBIR System

This system is proposed an approach based on SIFT features for sketch based image retrieval and consists of feature extraction, clustering features and similarity measure parts. The contribution of this work is comprised as post feature extraction.
The stages of proposed system are described as shown in figure 1.

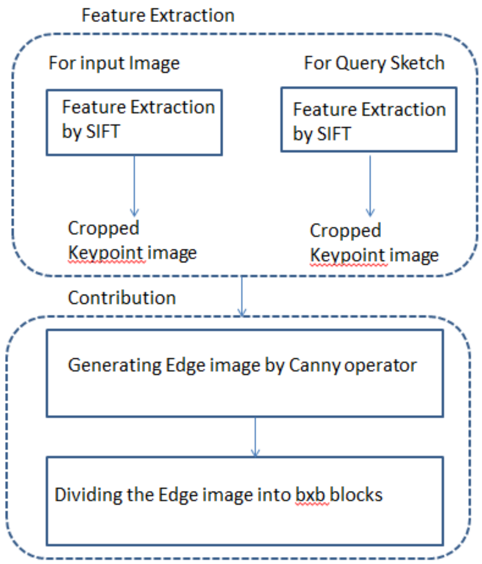

Fig.1.The stages of proposed system

A. Feature Extraction

Feature extraction is the main part in SBIR system. It extracts the visual information from the image and save them as the feature vectors in database. The feature extraction finds the image description in the form of feature values called feature vector for each pixel. These feature vectors are used to compare the query with the other images and retrieval.
Scale Invariant Feature Transform (SIFT) algorithm is used for feature extraction in this system. SIFT features are features extracted from images to help in reliable matching between different views of the same object. The extracted features are invariant to scale and orientation, and are highly distinctive of the image.
The procedures of SIFT algorithm are:

1. Constructing a scale space
2. Laplacian of Gaussian (LoG) approximation
3. Finding keypoints
4. Get rid of bad keypoints
5. Assigning an orientation to the keypoints
6. Generate SIFT features

For constructing a scale space, firstly the original image is progressively blurred and resized. Mathematically, blurring is referred to as the convolution of the Gaussian operator and the image. Gaussian blur is applied to each pixel.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

Where,
L = the blurred image
G = Gaussian operator
I = an image
x, y = the location coordinates
σ = the "scale" parameter
The actual Gaussian Blur operator is

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$$

The amount of blurring in each image is important. Assume the amount of blur in a particular image is σ. Then, the amount of blur in the next image will be k*σ. Here k is constant. The first step of SIFT generated several octaves of the original image. Each octave's image size is half the previous one. Within an octave, images are progressively blurred using the Gaussian Blur operator.
The Laplacian of Gaussian (LoG) operation calculates the second order derivatives on the blurred image. This locates edges and corners on the image. These edges and corners are good for finding keypoints.
Finding keypoints is a two part process.

- Locate maxima/minima in DoG (Difference of Gaussian) images
- Find sub_pixel maxima/minima

The first step is to coarsely locate the maxima and minima. This is iterated through each pixel and checks all its neighbors. The check is done within the current image, and also the one above and below it.
Using the available pixel data, sub_pixel values are generated. This is done by the Taylor expansion of the image around the approximate key point.
Mathematically,

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x}$$

It can be easily found the extreme points of this equation (differentiate and equate to zero). On solving, sub_pixel key point locations are getting. These sub_pixel values increase chances of matching and stability of the algorithm.
If the magnitude of the intensity (i.e., without sign) at the current pixel in the DoG image (that is being checked for minima/maxima) is less than a certain value, it is rejected. In this step, the number of keypoints was reduced. This helps increase efficiency and also the robustness of the algorithm. Keypoints are rejected if they had a low contrast or if they were located on an edge.
The next thing is to assign an orientation to each keypoint. This orientation provides rotation invariance. The size of the "orientation collection region" around the keypoint depends on its scale. The bigger the scale, the bigger the collection region. To assign an orientation a histogram is used and a small region around it. Using the histogram, the most prominent gradient orientations are identified. If there is only one peak, it is assigned to the keypoint. If there are multiple peaks above the 80% mark, they are all converted into a new keypoint (with their respective orientations).
The final step of SIFT algorithm is to create a feature vector of each keypoint. Take a 16x16 window of "in-between" pixels around the keypoint and split that window into sixteen 4x4

windows. From each 4x4 window it can be generated a histogram of 8 bins. Each bin corresponding to 0-44 degrees, 45-89 degrees, etc. Gradient orientations from the 4x4 are put into these bins. This is done for all 4x4 blocks. Finally, it is normalized the 128 values is getting.
The result of feature extraction is as shown in figure 2.

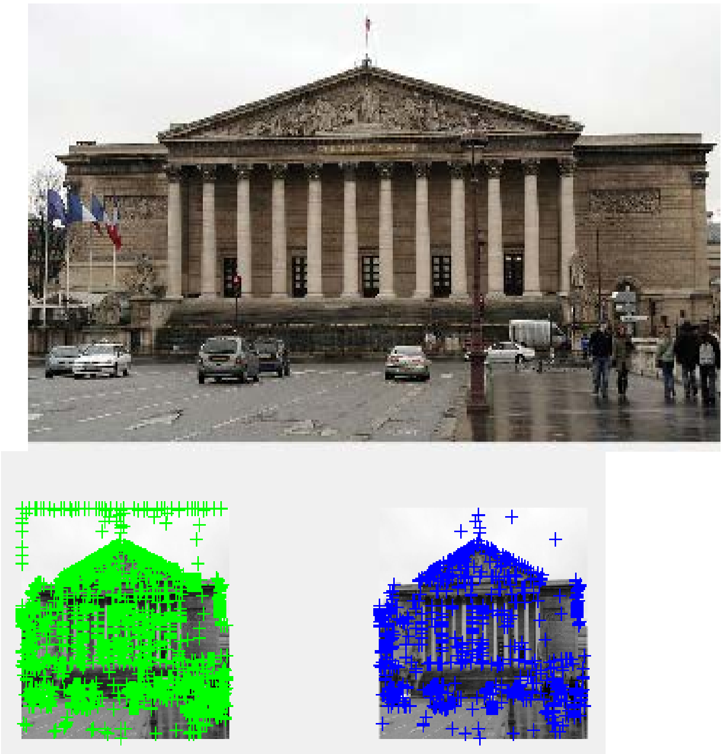

Fig.2. Thekeypoint image and its original

B. *Clustering Features*

Most of the SBIR systems used K-Means algorithm for clustering feature vectors. In this system Fuzzy C-Means algorithm is used to get more efficient retrieval result.

The Fuzzy C-Means (FCM) is a clustering which allows one piece of data to belong to two or moreclusters. This method is frequently used in pattern recognition. The FCM objective function and itsgeneralizations are the most heavily studied fuzzy model in Pattern Recognition. There is an infinite range ofpossible fuzzy partitions. Therefore, an optimization model or objective function must be devised to searchfor the optimal partition according to the chosen objective function. The way that most researchers havesolved the optimization problem has been through an iterative locally optimal technique, called the FCMalgorithm. The FCM objective function weighted the distance between a given data point and a givenprototype by the corresponding degree of membership between the two. Thus, partitions that minimize thisfunction are those that weight small distances by high membership values and large distances by lowmembership values.

The FCM algorithm attempts to partition a finite collection of n elements $X = \{x_1,\ldots\ldots,x_n\}$ into a collection of c fuzzy clusters with respect to some given criterion.Given a finite set of data, the algorithm returns a list of c cluster centers.$C = \{c_1,\ldots\ldots,c_n\}$. And a partition matrix $U = u_{ij} \in [0,1]$, i= 1, . . ,n , j = 1,….,c , where each elementu_{ij} tells the degree to which element x_i belongs to clusterc_j. Like the k-means algorithm, theFCM aims to minimize an objective function. The standard function is:

$$u_k(x) = \frac{1}{\sum_j \left(\frac{d(center\ _k, x)}{d(center\ _k, x)}\right)^{2/(m-1)}}$$

This differs from the k-means objective function by the addition of the membershipvalues u_{ij} and the fuzzifier m. The fuzzifier m determines the level of cluster fuzziness. A large m results in smaller memberships u_{ij} hence fuzzifier clusters. This algorithm is used for clustering of images for retrieval after calculating feature vectors.

C. *Feature Matching*

Features extracted from the whole image are called global features. Local features are extracted from an object or a segment of an image. Global features cannot provide enough information to estimate the similarity between images. Therefore, local feature descriptor is used in this system.

After extracting the feature vector by SIFT algorithm, the cropped keypoint image is getting. To achieve efficient matched result, the edge of cropped keypoint image is generated by Canny operator. Then, the edge image is divided in bxb blocks. As the last stage of feature matching, the pixels in edge of each block are counted to match those of query sketch.

D. *Datasets*

Evaluating a sketch based image retrieval system is a momentous task. It is even more difficult to match images and sketches due to the vague nature of the sketch. A sketch can depict shapes or symbols or an imaginary scene, thus semantic convergence with photographic images is not always the case.

So far, three publicly available benchmark datasets have been published in the literature: EitzSBIRdataset , Flickr160 and Flickr15k. Among those, EitzSBIR dataset is used in this system.

EitzSBIR benchmark was published by Eitz*et al.* and is based on a controlled user study of 28 subjects. It consists of 31 hand-drawn sketches, 1,240 images related to these sketches and 100,000 distractor images. It is also available online 1. The authors establish sketch/image ratings based on user ratings in a controlled environment. Generation of input sketches was designed with focus on shape based retrieval. Users were prompted to avoid too much abstraction and symbols in their drawings and encouraged to generate sketchesdepicting objects or scenes in a way that they would expect to perform well for an image retrieval system. This process generated 164 sketches. Authors selected the 31 more precise and coherent sketches that reasonably matched a sufficient number of images in the database. Some of the input sketches are depicted in Figure 3.

Each sketch was associated with 40 visually similar images according to the user rating, for a total of 1,240 images. 100,000 distractor images were also provided as noise and mixed with the 1,240 images. The author's aim was to create a benchmark to quantitatively compare a machine's result with respect to the human performance. That is, how correlated is the ranking produced by a human, in this case the mean score of all the participants, to that of a computer.

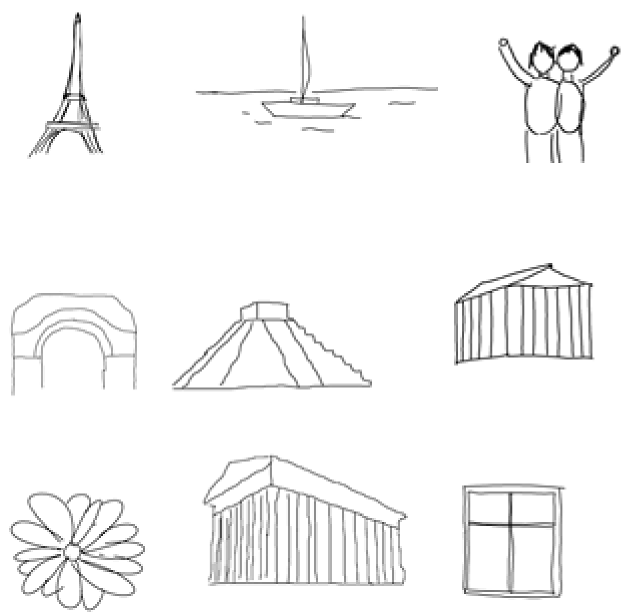

Fig.3. Some sketch queries from EitzSBIR Dataset

V. Performance Evaluation

To evaluate the performance of the Sketch Based Image Retrieval System, the Mean Average Precision (MAP) and Recall rate are calculated.

The Precision provides information related to effectiveness of the system.

Mean Average Precision (MAP) = $\sum$Avg P (q)/Q

Where,
Q= no: of queries images displayed with similar shape

Recall = Q / Z

Where,
Q = No. of images displayed with similar shape.
Z = No. of images with similar shape in whole database.

VI. Conclusion

This paper is presented the scalable sketch based image retrieval system using SIFT algorithm and fuzzy clustering. It is still difficult to bridge the gap between image and sketch matching problem. Thus, the main contribution of this work is to improve the effectiveness of image retrieval by querying as the hand drawn sketch. This system is an ongoing research work. Therefore, the overall experimental result cannot be described. Similarity score will be calculated by pixel count ratio as the contribution of this work. Mean Average Precision (MAP) and Recall rates will be measured as evaluation criteria.

References

[1] B. Szanto. P. Pozsegovics, Z. samossySz. Sergyan, "Sketch4Match Content-based Image Retrieval System Using Sketches" SAMI 2011 _9th IEEE International Symposium on Applied Machine Intelligence and Informatics January 27-29, 2011 Smolenice, Slovakia.

[2] Fendarkar J. D., Gulve K. A. "Utilizing Effective Way of Sketches for Content-based Image Retrieval System" International Journal of Computer Applications Volume: 116 – No.: 15, pages: 23-28,April 2015.

[3] Jose M. Saavedra. "Sketch based image retrieval using a soft computation of the histogram of edge local orientations (s-helo)". In International Conference on Image Processing, ICIP'2014.

[4] Mathias Eitz, Kristian Hildebrand, TamyBoubekeur, and Marc Alexa. "Sketch-based image retrieval: Benchmark and bag-of-features descriptors" IEEE Transactions on Visualization and Computer Graphics, 17(11):1624–1636, 2011.

[5] N. Prajapati, G.S.Prajapti, "Sketch base Image Retrieval System for the Web - Survey", International Journal of Computer Science and Information Technologies, Volume 6, Issue 4, April 2015 ISSN: 0975-9646.

[6] Prof. BalramPuruswani, Jyoti Jain, "A Preliminary Study on Sketch Based Image Retrieval System", International Journal of Modern Engineering & Management Research, Volume 1, Issue 1, April 2013 ISSN: 2320-9984

[7] RuiHua and John Collomossea, "A Performance Evaluation of Gradient Field HOG Descriptor for Sketch Based Image Retrieval", Centre for Vision, Speech and Signal Processing, University of Surrey, UK.

[8] http://www.aishack.in/2010/05/sift-scale-invariant-feature-transform/2/

[9] R. Datta, D. Joshi, J. Li and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age", ACM computing Survey, vol.40, no.2, pp.1-60, 2008.