ISSN 2320-9984

Volume 1, Issue 1, April 2013

ISSN: 2320-9984 (Online)

International Journal of Modern Engineering & Management Research

Website: www.ijmemr.org

A Preliminary Study on Sketch Based Image Retrieval System

Jyoti Jain
M.Tech (CTA)
Gyan Ganga Institute of Technology & Science
Jabalpur (M.P.) [INDIA]
Email: jainjyoti109@gmail.com

Prof. Balram Puruswani
Dept. of Computer Science,
Gyan Ganga College of Technology
Jabalpur (M.P.) [INDIA]

Abstract—*Content based image retrieval is the method of retrieving the image from the database on the basis of content. content may be refer as color, text, shape, sketch or any other information that can be derived the image itself[1]. Here we show various image retrieval method which is used in sketch content.*

We have studied EHD, HOG and SIFT. Overall, the results show that the sketch based system allows users an intuitive access to search-tools.[2]. In this paper show the comparison with the EHD, HOG and SIFT methods.

This paper aims to introduce the problems and challenges concerned with the design and the creation of SBIR systems, which is based on a free hand sketch (Sketch based image retrieval – SBIR). With the help of the existing methods, a possible solution is describe and represent how to design and implement a task specific descriptor, which can overcome the informational gap between a sketch and a colored image. It is an opportunity for the efficient search hereby.

Keywords: EHD, HOG, SIFT

1. INTRODUCTION

Content-based image retrieval (CBIR), also known as query by image content (QBIC) and content based visual information retrieval (CBVIR) is used for computer vision to the image retrieval problem. The problem is searching for digital images in large databases.

The Content based means that the search will analyze the actual contents of the image widely used to describe the process of retrieving desired images from a large collection on the basis of content or feature (such as color, texture, shape, sketch etc.). with the help of content can be extracted the images themselves. The features used for retrieval can be primitive or semantic, but the extraction process must be predominantly automatic.

A *CBIR* system consists of three major components and its variations are depend on features used.

i. ***Feature extraction*** – Examine raw image data to essence feature specific information.

ii. ***Feature storage*** – Provide efficient storage for the extracted information, also help to improve searching speed.

iii. ***Similarity measure*** – Measure the difference between images for determining the relevance between images.

2. IMAGE RETRIEVAL THROUGH SKETCH

A sketch based system can play very important and efficient roll in many areas of the real life. In some cases we can recall our minds with the help of figures or drawing. Because human is able to recall visual information using shape an object or an arrangement of color and object since human is a visual type, we find required images with the help of other image. In this case we search images using some features of image and these features are the keywords following paragraph some application possibilities are analyzed[3].

The CBIR systems have a big importance in the criminal investigation. The identification of petty images, tattoos and graffitist can be supported by these systems. Similar applications are implemented in [4], [5].

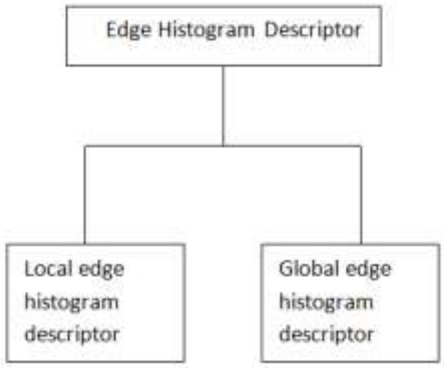
Another possible application area of sketch based information retrieval is the searching of analog circuit graphs from a big database [7]. The user has to make a sketch of the analog circuit, and the system can provide many similar circuits from the database.

The Sketch-based image retrieval (SBIR) was introduced in QBIC [8] and VisualSEEK [9] 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 [6].

3. EHD (EDGE HISTOGRAM DESCRIPTOR) CONCEPT

Histogram characteristic is to represent the global feature composition of an image. It is fixed to translation and rotation of the images and normalizing the histogram leads to scale invariance. Suppressed the above properties, the histogram is considered to be very essential for indexing and retrieving images [10][11].

In EHD method, it show how to made global and semi-global edge histogram bins from the local histogram bins. Through various possible clusters of sub-images, in semi-global histograms used 13 patterns. These 13 semi-global regions and the whole image space are adopted to define the semi-global and the global histograms respectively. These extra histogram information can be obtained directly from the local histogram bins without feature extraction process. Experimental results show that the semi-global and global histograms generated from the local histogram bins help to improve the retrieval performance[12].



Local Edge Histogram: The normative part of the edge histogram descriptor consists of 80 local edge histogram bins [13][14]. The semantics of those histogram bins are described in the following sub-sections.

To localize edge distribution to a certain area of the image, divide the image space into 4x4 sub-images as shown in Figure 1. Then, for each sub-image, generate an edge histogram to represent edge distribution in the sub-image. for define different edge types, the sub-image is further divided into small square blocks called image-blocks.

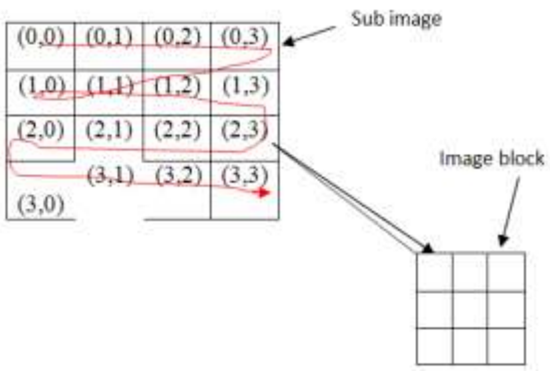


Figure 0. Definition of sub-image and image-block

Edge Type: there are different types of image

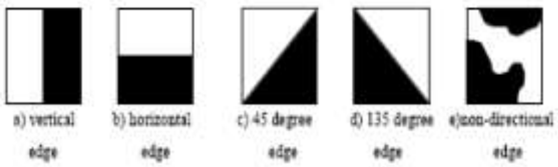


Figure 1: Types of Images

4. SEMANTICS OF LOCAL EDGE HISTOGRAM

After the edge extraction from image-blocks, we count the total number of edges for each edge type in each sub-image. Since there are five different edges, we can define five histogram bins for each sub-image. Then, since there are 4x4=16 sub-images, we have total 16x5=80 bins for the edge histogram. By scanning sub images[14]. according to the order shown in Figure 1, the semantics of the bins are defined as in Table 1

Histogram Bins	Semantics
Local_Edge [0]	Vertical edge of sub-image at (0,0)
Local_Edge [1]	Horizontal edge of sub-image at (0,0)
Local_Edge [2]	45 degree edge of sub-image at (0,0)
Local_Edge [3]	135 degree edge of sub -image at (0,0)
Local_Edge [4]	Non directional edge of sub-image at (0,0)
Local_Edge [5]	Vertical edge of sub-image at (0,1)
. . .	. . .
Local_Edge [78]	135 degree edge of sub -image at (3,3)
Local_Edge [79]	Non-directional edge of sub-image at (3,3)

5. SEMANTICS OF LOCAL EDGE HISTOGRAM

To achieve a high retrieval performance, the local histogram alone may not be sufficient Rather, we may need an edge distribution information for the whole image space and horizontal and vertical semi-global edge distributions also required.

There are 13 different clusters and for each cluster we generate edge distributions for five different edge types. Consequently, we have total 80 bins(local) + 5 bins(global) + 65 bins (13x5, semi-global) = 150 bins.

Note that the bin values for all global and semi-global histograms can be obtained directly from the local histogram [12].

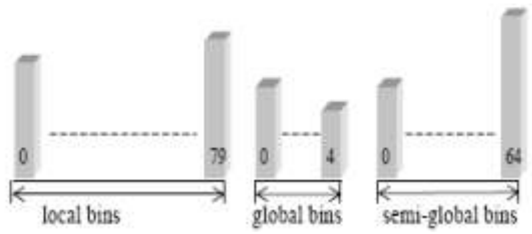


Figure 2: Overall Histogram Semantics

Advantages of EHD

1) show how to construct global and semi global edge histogram bins from the local histogram bins.

2) used 13 patterns for the semi-global histograms.

3) Extra histogram information can be obtained directly from the local histogram bins without feature extraction process.

4) Semi global and global histogram generated from the local histogram bins help to improve the retrieval performance.

Disadvantages of EHD

1) Not provide invariant opposite rotation, Scaling and translation.

2) The development is to difficult and robust descriptor is emphasized.

6. SIFT (THE SCALE INVARIANT FEATURE TRANSFORM) CONCEPT

Face recognition is becoming an increasingly important for many applications including human machine interfaces, multimedia, security, communication, visually mediated interaction and anthropomorphic environments. One of the most difficult problems is that the process of identifying a person from facial appearance has to be performed differently for each image, because there are so many conflicting factors altering facial appearance[19].

The approach of SIFT feature detection which is used for object recognition. the invariant features extracted from images can be used to perform reliable matching between different views of an object or scene. The features have been shown to be invariant to image rotation and scale and robust across a substantial range of affine distortion, addition of noise, and change in illumination.

For each sketch query, gallery photograph, and each sketch/photo correspondence in our dictionary, we compute a SIFT feature representation. SIFT based object matching is a popular method for finding correspondences between images. Introduced by Lowe,[20] SIFT object matching consists of both a scale invariant interest point detector as well as a feature-based similarity measure. Our method is not concerned with the interest point detector as well as a feature-based similarity measure. Our method is not concerned with the interest point detection aspect of the SIFT framework, but instead utilizes only the gradient-based feature descriptors (known as SIFT-features) [19].

SIFT method is implemented as the following stages:

1) Creating the Difference of Gaussian Pyramid,

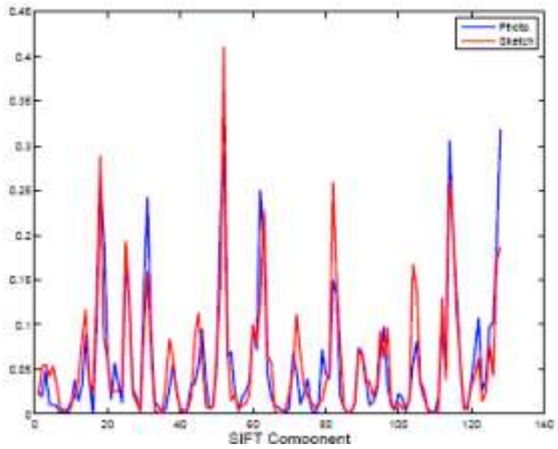
2) Extrema Detection,

3) Noise Elimination,

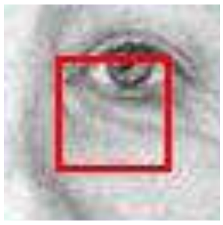
4) Orientation Assignment,

5) Descriptor Computation,

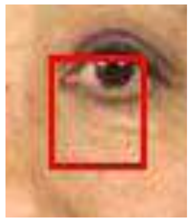
6) Key points Matching.



(a)



(b)



(c)

Figure 3: The similarity between sketch and photo image patches of the same person (s = 32). (a) Plots of the SIFT descriptor computed of the sketch image (b) and the photo image (c). The SIFT descriptor was computed within the solid box of the sketch and photo. The two descriptors exhibit high levels of similarity despite being computed in different image domains.

1) Creating the Difference of Gaussian Pyramid :

In first stage to construct a Gaussian "scale space" function from the input image [20]. This is formed by convolution (filtering) of the original image with Gaussian functions of varying widths.

The difference of Gaussian (DOG), D(x, y, σ), is calculated as the difference between two filtered images, one with k multiplied by scale of the other.

$$D(x, y, \sigma) = L(x, y, k\sigma)-L(x, y, \sigma)$$

These images, L(x, y, σ), are produced from the convolution of Gaussian functions, G(x, y, kσ), with an input image, I(x, y).

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

$$G(x, y, \sigma) = 1/2\pi\sigma^2 (\exp\{-(x2+y2)/2\sigma^2\})$$

First, the initial image, I, is convolved with a Gaussian function, *G*0, of width *σ* 0. Then we use this blurred image, *L*0, as the first image in the Gaussian pyramid and incrementally convolve it with a Gaussian, *Gi*, of width *σi* to create the *ith* image in the image pyramid, which is equivalent to the original image filtered with a Gaussian, *Gk*, of width *kσ* 0. The effect of convolving with two Gaussian functions of different widths is most easily found by converting to the Fourier domain, in which convolution becomes multiplication i.e.

$$G_{\sigma_i} * G_{\sigma_0} * f(x) \rightarrow \tilde{G}_{\sigma_i} \tilde{G}_{\sigma_0} \tilde{f}$$

The Fourier transform of a Gaussian function, *eax*2 is given by

$$F_x[e^{ax^2}](t) = \sqrt{\frac{\pi}{a}}\, e^{-\pi^2 (t)^2/a}$$

Substituting this into equation (4) and equating it to a convolution with a single Gaussian of width $k\sigma_0$ we get

$$e^{-t^2\sigma_i^2} e^{-t^2\sigma_0^2} = e^{-t^2 k^2 \sigma_0^2}$$

Performing the multiplication of the two exponentials on the left of this equation and comparing the coefficients of $-t^2$ gives:

$$\sigma_i^2 + \sigma_0^2 = k^2 \sigma_0^2$$

And so we get

$$\sigma_i = \sigma_0 \sqrt{k^2 - 1}$$

This subtle point is not made clear in the original paper, and it is important because after sub sampling of the low-passed filtered images to form the lower levels of the pyramid we no longer have access to the original image at the appropriate resolution, and so we cannot filter with *Gk* directly.

2) Extrema Detection

This stage is to find the extrema points in the DOG pyramid. To detect the local maxima and minima of D(x, y, σ), each point is compared with the pixels of all its 26 neighbors. If this value is the minimum or maximum this point is an extrema. We then improve the localization of the key point to sub pixel accuracy, by using a second order Taylor series expansion. This gives the true extrema location as:

$$z = -\left(\frac{\partial^2 \mathbf{D}}{\partial \mathbf{x}^2}\right)^{-1} \frac{\partial \mathbf{D}}{\partial \mathbf{x}}$$

where D and its derivatives are evaluated at the sample point and $x = (x, y, \sigma)^T$ is the offset from the sample point[19].

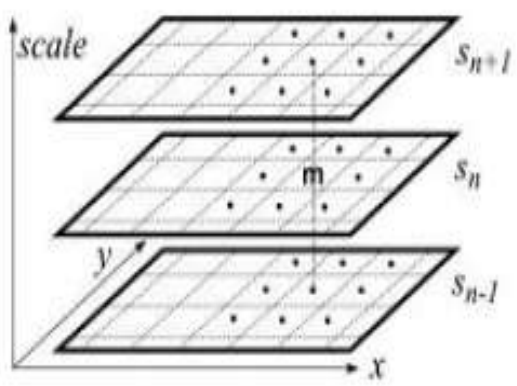


Figure 4: An extrema is defined as any value in the DOG greater than all its neighbors in scale-space

***Advantages* of SIFT:**

1) Perform reliable matching between different views of an object or scene

2) image rotation and scale and robust across a substantial range of affine distortion, addition of noise, and change in illumination.

3) The image gradient magnitudes orientations are sampled around the key point location

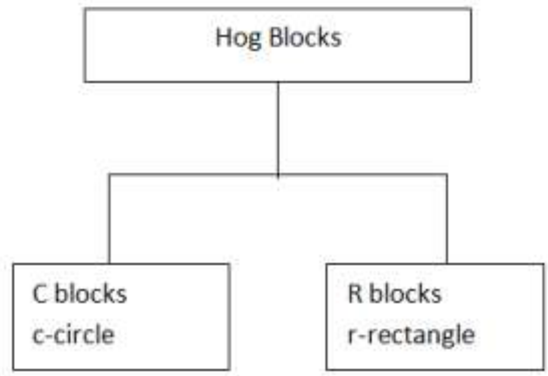
Disadvantages of SIFT:

1) Edges are poorly defined and usually hard to detect, but there are still large numbers of key points can be extracted from typical images.

2) Perform the feature matching even the faces are small

Hog (Histogram of Oriented Gradient descriptors) concept :

The essential thought behind the Histogram of Oriented Gradient descriptors is that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions. The implementation of these descriptors can be achieved by dividing the image into small connected regions, called cells, and for each cell compiling a histogram of gradient directions or edge orientations for the pixels within the cell. The combination of these histograms then represents the descriptor. For improved accuracy, the local histograms can be contrast-normalized by calculating a measure of the intensity across a larger region of the image, called a block, and then using this value to normalize all cells within the block. This normalization results in better invariance to changes in illumination or shadowing[15].



***Computing The Gradient* :**

Several gradient detectors tried

- [1,-1], [1,0,-1], [1,-8,0,8,-1]

- Unfiltered and Pre-filtered with Gaussian smoothing

- Simplest [1,0,-1] proved best

- Gaussian smoothing affected results negatively

For color images

- Compute each channel separately

- Choose the largest value as the gradient for that pixel

Advantages:

1) It is examine for more database.

2) The HOG in more cases it is much better than the EHD based retrieval.

3) The edge Histogram descriptor not mainly look better for information poor sketch, while other case show better result can be achieve for more detailed this problem can be overcome by the Hog method.

4) Capture edge or gradient structure that is very characteristic of local shape.

Disadvantages:

Working on HOG-based detectors that incorporate motion information using block matching or optical flow fields. Finally, although the current fixed-template-style detector has proven difficult to beat for fully visible pedestrians, humans are highly articulated and we believe that including a parts based model with a greater degree of local spatial invariance[16].

Performance comparison :

Method	HOG (with gradient map)	Hog (without gradient map)	SIFT	EHD	HOG (own)
Average precision	54 %	42%	41 %	43%	44%

Table 2: the performance of used method in sketch based system [2]

REFERENCES:

[1] A Review on Content Based Image Retrieval Mahesh Maurya, Chitvan Oza, Alok Shah,Sadiya Patel MPSTME, SVKM's NMIMS (Deemed-to be University), Vile Parle West, Mumbai 56 maheshkmaurya@yahoo.co.i chitvanoza@gmail.com, mrcool89@gmail.com, ash_cateyes@yahoo.co

[2] Sketch4Match – Content-based Image Retrieval System Using Sketches B. Szant o, P. Pozsegovics, Z. V amossy, Sz. Sergy an Obuda University/Institute of Software Technology, Budapest, Hungary szanto90balazs, pozsee1st @gmail.com, vamossy.zoltan, sergyan.szabolc @nik.uni-obuda.hu

[3] B.Szanto, P. Pozsegovics, Z.Vamossy, Sz.Sergyan, "Sketch4Match – Content Based Image Retrieval System Using Sketches", SAMI 2011, 9th IEEE International Symposium on Applied Machine Intelligence and Informatics, IEEE January 2011.

[4] A.K. Jain, J.E. Lee, and R. Jin, "Sketch to photo matching: a feature-based approach," Proc. SPIE, Biometric Technology for Human Identification VII vol. 7667, pp. 766702–766702, 2010.

[5] A.K. Jain, J.E. Lee, R. Jin, and N. Gregg, "Graf ti-ID: matching retrieval of graf ti images," ACM MM, MiFor'09 pp. 1–6, 2009.

[6] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa, "An evaluation of descriptors for large-scale image retrieval from sketched feature lines," Computers and Graphics, vol. 34, pp. 482–498, October 2010.

[7] Gy. Gyor ok, "Embedded hybrid controller with programmable analog circuit," IEEE 14th International Conference on Intelligent Systems pp. 59.1–59.4, May 2010.

[8] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Hiang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by image and video content: the QBIC system," IEEE Computer vol. 28, pp. 23 –32, 2002

[9] J.R. Smith, and S.F. Chang, "VisualSEEK: a fully automated content- based image query system," ACM Multimedia '96 'pp. 97–98, 1996.

[10] M.J. Swain and D.H. Ballard, "Color indexing," Int. J. of Computer Vision, vol.7-1, pp. 11-32, 1991.

[11] A. K. Jain and A. Vailaya, "Image retrieval using color and shape," Pattern Recognition, Vol. 29, No. 8, pp.1233-1244, 1996.

[12] Efficient Use of Local Edge Histogram Descriptor Dong Kwon Park Dep. of Electronic Engineering, Dongguk Univ. 26, 3-ga, Pil-dong, Jung-gu, Seoul, 100-715, KOREA +82-2-2260-3337 dkpark@dgu.ac.k Yoon Seok Jeon Dep. of Electronic Engineering, Dongguk Univ. 26, 3-ga, Pil-dong, Jung-gu, Seoul, 100-715, KOREA +82-2-2260-3337 hardjeon@dgu.ac.k Chee Sun Won Dep. of Electronic Engineering, Dongguk Univ. 26, 3-ga, Pil-dong, Jung-gu, Seoul, 100-715,KOREA +82-2-2260-3337 cswon@dgu.ac.

[13] S. J. Park, D. K. Park, C. S. Won, “Core experiments on MPEG-7 edge histogram descriptor,” MPEG document M5984, Geneva, May, 2000.=

[14] ISO/IEC/JTC1/SC29/WG11 : “Core Experiment Results for Edge Histogram Descriptor (CT4),” MPEG document M6174, Beijing, July 2000.

[15] http://en.wikipedia.org/wiki/Histogram_of_oriented_gradients.

[16] Histograms of Oriented Gradients for Human Detection Navneet Dalal and Bill Triggs INRIA Rh.one-Alps, 655 avenue de l'Europe, Montbonnot 38334, France Navneet.Dalal,Bill.Triggsg @inrialpes.fr, http://lear.inrialpes.f

[17] http://en.wikipedia.org/wiki/ Scale _Invariant _Feature Transform

[18] Lowe, D.G..2004. Distinctive Image Features from Scale-Invariant Keypoints. January 5, 2004

[19] Implementing the Scale Invariant Feature Transform(SIFT) Method YU MENG and Dr. Bernard Tiddeman (supervisor) Department of Computer Science University of St. Andrews yumeng@dcs. st-and.ac.u

[20] Sketch to Photo Matching: A Feature-based Approach Brendan Klarea and Anil K Jaina,b aDepartment of Computer Science and Engineering Michigan State University East Lansing, MI, U.S.A Department of Brain and Cognitive Engineering Korea University Seoul, Korea.