A Comprehensive Analysis on Content Based Image Retrieval

Prerna Sahariya[1] Deepesh Namdev[2]
[1,2]*Deptt. of Electronics & Communication Engineering,*
Gurukul Institute of Engineering & Technology, India

Abstract- This paper provides the details of how it becomes possible to use the storage capacity with the rapid increase in multimedia technology. Content Based Image Retrieval (CBIR) grow into an efficient technology against the massive storage of database. This paper illustrates techniques used to combine shape, color and texture for image retrieval. Further paper offers numerous commonly used image matching methods and algorithms for feature representation and their extraction. In this paper, the authors present the existing techniques with their advantages, disadvantages and applications regarding CBIR. This paper is quite helpful in find out the research gaps and challenges of CBIR faced by researchers. It provides a thorough examination in between the current available systems with the previous traditional techniques. Subsequent comparative analysis helps the researchers to find out their directions according to the problem they are facing.
Keywords- Scale Invariant Feature Transform (SIFT), Content Based Image Retrieval (CBIR), Feature Vector, HSV, Histogram

I. INTRODUCTION

In the field of image analysis, Content-based image retrieval (CBIR) emerges as the effective technology for image recovery. It is a well acclaimed and known as Content-Based Visual Information retrieval (CBVIR) and Query by Image Content (QBIC) [1]. Image retrieval process following the steps in the sequence of feature extraction, calculation of similarity based on features of image, feedback from semantic and image acquisition [2, 3]. These steps are the results of information retrieval studies, database technology, pattern recognition, machine vision. Color, layout, figure, shape, texture etc. are some important features to extract from general-purpose image retrieval. Shape, color and texture of the image are mostly preferred out of them in CBIR process. Nowadays, this methodology is widely commended due to its unique features for visual representation as well as to express the other mediums.

A rapid progress in multimedia world with the advancement of smart gadgets such as cell phones and computers, their storage capacity and access networks, system offers a high-speed transmission of the large number of images. In recent decades, it feels a need to develop image retrieval methods along with text retrieval processes. Accessing of visual data can be most effectively done by CBIR phenomenon [4]. CBIR offers to extract image structure, shape and color of image content as a replacement of annotated text. In traditional database technology, it does not fulfil the needs of image database that has a massive storage capacity as compared with text object database. It becomes a huge challenge to researchers to retrieve images in the same storing capacity as text database. The traditional way in adequacies the effective and automatic description of the image. It only uses text for an annotated image. There is a need for proper interpretations and understanding of the content of sequenced images in order to implement CBIR system. The retrieval index should be produced automatically, which provides more a visual retrieval interface to users. CBIR refers to image content that is retrieved directly, by which the images with certain features or containing specific content will be searched from image database. CBIR performs the tasks of information analysis from images by considering low level features of visuals [2]. These features are creating a relationship between objects and space, shape, texture, color etc. by indexing them in form of vectors. Retrieval methods are based on multidimensional features of an image.

Broad applications are present of CBIR including the fields such as education, agriculture, architectural design, medical science, military affairs etc. Visuals carry some important features like no language restrictions and abundant in content for facilitating international exchanges. A progress has been made on to develop CBIR systems. Some of them are named as Netra [5], SIMPLIcity [6], VisualSEEk [7], Photobook [8], QBIC [9] and Virage [10] etc. CBIR can further classify as low level and high level based on its features. Apart from shape, color and texture, features can be subdivided on the basis of either local for a small group of pixels or global for the entire image.

II. CATEGORIZATION OF IMAGE FEATURES

Content-based image retrieval (CBIR) technique retrieves visuals, shots from videos and images required a massive library for large pictures. CBIR plays a significant role in database organization and management. At present, CBIR research mainly focuses on image data model, feature describing of image content, image database technology, content retrieval system design, retrieval performance evaluation and many other fields. CBIR performed process in the following steps: retrieval content information, content retrieval system design, feature describing of image content and image data model. Feature describing image content step produces vectors for multi-dimensional picture. These vectors decides the content whether it is a low-level or a high-level. Low-level features efficiently remove the sensory gap between descriptive information from recorded scene and the object in the world. The high-level features remove the semantic gap between interpreted data and the extracted information from the images for a user in a given situation. Generally, reflecting color, shape, color and salient points are used as the most common low-level features in an image [2]. In almost all CBIR systems, the most effective feature of a visual is color due to the low storage requirements, simplicity, effectiveness and robustness advantages. Human perception prefers RGB space over color because of better results achieved by CIE Lab or HSV or CIE Lab and LUV spaces [10]. Histograms created for the representation of the color distribution and created visuals from featured vectors. Histogram can be defined as a process adopting for color improvement in CBIR as various pictures operated at identical or similar color histograms. Same visuals operated at different ambient lighting result as the generation of different histograms. After color, texture is most widely used feature in CBIR. Texture feature performs the tasks within an image by of capturing repetitive patterns and granularity of surfaces [2]. Interpretation of natural images in the MPEG-7 standard, performed by a set of texture and color descriptors consisting the spatial texture/color descriptor and histogram-based descriptors [11].

Ultimate aim of CBIR is to reduce semantic gap between the richness of human semantics and visual features [12]. High-level semantic derive the features of CBIR by object-ontology [13] to define their concepts. Query concepts associate with low-level feature, used for learning supervised or unsupervised methods [12]. Learning of user's intentions related with retrieval loop perform the relevance feedback [13]. Generation of semantic templates [13] support high-level image retrieval. Semantic gap is difficult to eliminate due to inconsistence understanding of visual data for distinct users.

III. MATHEMATICAL MODELS FOR EXTRACTION OF FEATURES

3.1 Shape Feature Extraction

Extraction based on shape feature of an object is one of the salient feature. Shape feature representations are very intuitive and used to distinguish objects. It is more challenging task as compared with color and texture features due to the semantic information of object that can be related with more high-level. Therefore, in essence, shape feature is expressing its characteristics in a way that is more complex. Extraction of shape feature can be performed by invariant moments algorithm to find out their centroid. For segmentation of visuals, canny operator performs the task successfully.

(p, q) -order moment of function $f(x, y)$ is defined as [12]:

$$m_{pq} = \iint x^p y^q f(x, y)\, dxdy$$

Centroid:

$x_c = m_{10}/m_{00}$, $y_c = m_{01}/m_{00}$, (p, q) -order moment centre:

$$\mu_{pq} = \iint (x - x_c)^p (y - y_c)^q f(x, y)\, dxdy$$

For digital images, the above formula can be expressed as:

$$\mu_{pq} = \sum_x \sum_y (x - x_c)^p (y - y_c)^q$$

3.2 Texture Feature Extraction

Statistical methods solve the texture feature from visuals. It mainly consists the spectrum and structure analysis. Most common methods to extract texture features are Gabor wavelet, co-occurrence matrix gray and Tamura texture. One premise assumption is present in all of these methods that the perspective to obtain images is same in all methods that is practically not possible. Therefore, the outcomes of above-mentioned methods have become relatively inefficient and poor. Some other methods are available like wavelet texture analysis method for geometric resistant invented by Pan Hui. It is proposed to enhance the capacity of anti-geometric deformation of the picture. Although, the demerit is that it is not good for the perception of human visual making the task difficult to understand. A solution is there to combine this methodology with Tamura texture characteristic. It certainly improve the shortcomings of previous methods. It makes the system more perceptual with human and more coincide with the human vision. Texture features can be categorized as;

3.2.1 Resist Geometry Distortion Feature

Main idea of texture behind image retrieval method is the use of anti-deformation. It makes the Fourier Transform (FT) applicable for signal directly. Let us assume that in (*m*, *n*) 2D sequence of limited region, *x*(*m*, *n*) is the surface. Then the received FT of 2D image, in this is,

$$X(k,l) = DFT[x(m,n)]$$

$$= \sum_{k=0}^{M-1}\sum_{n=0}^{N-1} x(k,l) W_N^{-mk} W_N^{-nk}$$

where, W_N' is weight function

(m = 0, 1, 2,..., M-1; l = 0, 1,..., N-1)

3.2.2 Tamura Texture Feature

Based on the research conduct on people for texture to check their visual sense psychology, Tamura is an advance way to express the texture. It features six components based on six properties in psychology depending on the texture features: roughness, regularity, line likeness, directionality, coarseness and contrast. Some of them are described as [13];

First: Directionality

Graded vector of each pixel are the building blocks of directionality and its computation. The direction and mode are illustrated as:

$$|\Delta G| = (|\Delta_H + \Delta_V|)/2$$

$$\theta = \tan^{-1}(\Delta_V/\Delta_H) + \pi/2$$

Δ_H and Δ_V denotes the convolution matrixes for horizontal and vertical direction variations respectively. After the grading vectors of all pixels are achieved, θ can be computed with the help of histogram H_D. depending on the weak values of sharpness in histogram, the complete directionality of image obtained.

$$F_{dir} = \sum_{p}^{n_p}\sum_{\varphi}(\varphi - \varphi_p)^2 H_D(\varphi)$$

Second: Contrast

Image intensity represents the contrast of an image and can be achieved by statistics the distributional condition. It has demonstrated as $\alpha_4 = \mu_4/\sigma_4$, where the σ4 is the variance and μ4 is the forth moment. Contrast can be illustrated as;

$$F_{con} = {}^{\sigma}/_{\alpha_4^{1/4}}$$

Third: Coarseness

Computation of coarseness can be done in several steps: first, it requires to calculate average intensity of the pixel in the activity pixel window that the image's size is $2^k \times 2^k$:

$$A_k(x,y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1}\sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} \frac{g(i,j)}{2^{2k}}$$

k = 0,1,2,3,4,5 , and g(i, j) is the pixel intensity value at point (i, j) . Then to find out the average value between the unconfident windows in vertical and horizontal direction for each pixel.

3.2.3 Inner Normalization

Inner components of feature vector computes the Gaussian normalization based on similarity compare for same status. The following steps can explain it as,

For the n-D feature vector: $F = [f_1, f_2, \ldots, f_n]$, n is the number of the feature elements. $I_1, I_2, \ldots, I_M$ represent *M* images in the base, for the image I_i, it's corresponding feature vector is $F_i=[f_{i1}, f_{i2}, \ldots, f_{in}]$. Gaining an $M\times N$ metric $F=f_{ij}$, and f_{ij} is the j-th feature element of the F_i. *F* is the feature sequence of each column which the length is *M*, that is ${}_jF'_j=[f_{1,j}, f_{2,j}, \ldots, f_{m,j}]$,and ${}_jF'_j$ is the Gaussian sequence, compute its average value m_j and the standard deviation σ_j, then use the formula to make the sequence normalize to the *N*(0,1) distributional sequence.

$$f_{(i,j)}^n = \frac{f_{i,j} - m_j}{\sigma_j}$$

Then, we can gain a vector having 13 texture feature components:
$F' = [f'_1, f'_2, \dots, f'_{13}]$

3.3 Color Feature Extraction
Next and very important factor is color feature for the image retrieval. It uses HSV model that works on the senses of human visuals [6]. Computation reduction can do the extraction of colors for expressing the colors without depressing the picture quality. Without lessening the quality, there is huge reduction in storage capacity that also results in boost up the velocity of processing. For quantization in HSV space, the representation is,

$$H = \begin{cases} 0, H\epsilon[0,60] \\ 1, H\epsilon[60,120] \\ 2, H\epsilon[120,180 \\ 3, H\epsilon[180,240] \\ 4, H\epsilon[240,300] \\ 5, H\epsilon[300,360] \end{cases}$$

Completion the quantization results the division of color into $L=L_h*L_s*L_v$ areas. Where, *L*h、*L*s、*L*v denote the three progressions of quantization respectively. These are obtained in the range for 36 color. 1D vector forms with the combination of three components in the following pattern: $L=Ls \times Lv \times H + Lv \times S+V$. Therefore the outcome of their combination will be L=6H+3S+V. After that, results obtained in form of 36 stems for a 1-D histogram. To extract the texture feature, color histogram is the important method by performing the tasks such as translation, scale and rotation invariance. Discrete function of 1D histogram can be represented as,

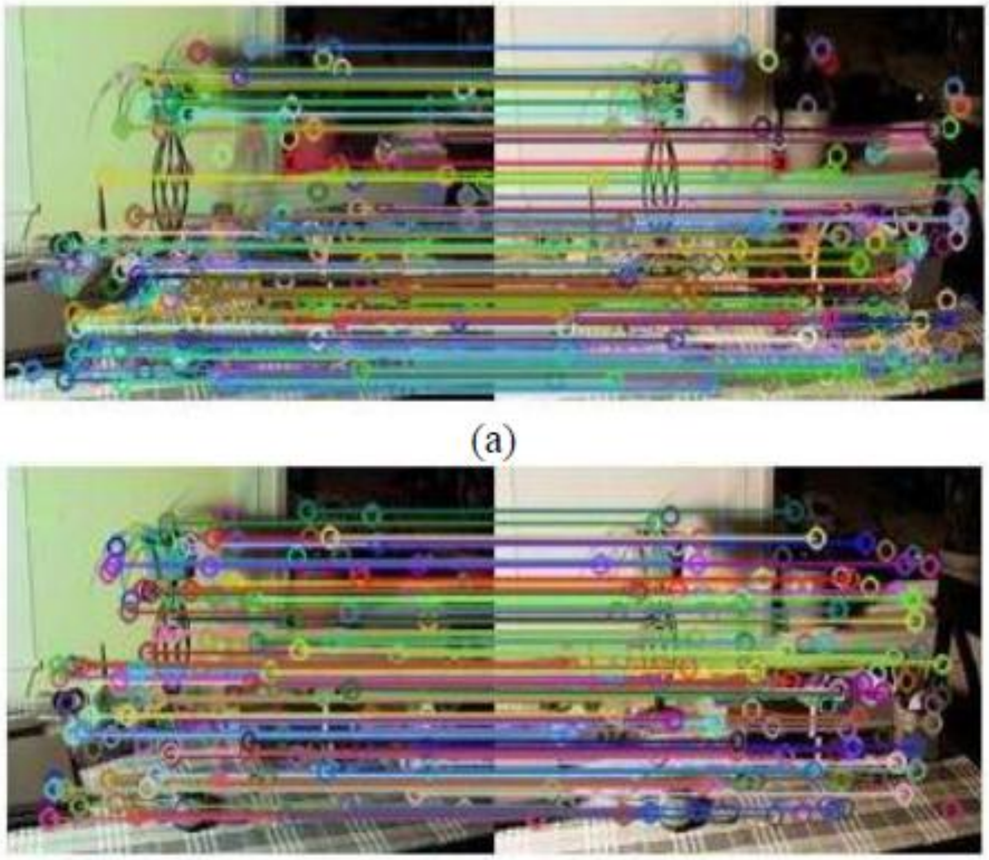
$$h_k = \frac{n_k}{n}, k = 0,1, \dots, L-1$$

Where, L is the number of features, which are 36, *k* is the feature value. Therefore, the histogram of image P: $H(p)=[h_1,\ h_2,\ \dots h_{36}]$ known as color feature vector.

IV. COMPARATIVE ANALYSIS OF SURF, SIFT & HOG

4.1 Surf & Sift
Speeded Up Robust Features (SURF) works as both feature descriptor as well as detector. It shows its significance in 3D reconstruction, their classification, image registration and object recognition. It is the result of feature extraction algorithm extracting from Scale invariant feature transform (SIFT). SIFT was coming into light in the late 90’s and was the first algorithm for image analysis. According to the developers who invented SURF, the proposed algorithm is several times faster than SIFT by fetching the results with high robustness. SURF can comute three integer operations designed by Hessian blob detector. It uses a precomputed integral image to find out the intersect points. Based on the combination of Haar wavelet, SURF’s feature descriptor details the vectors like a proficient for an internal image. SURF descriptors also helpful in recognizing, locating and tracking of objects, to extract point of interest, to reconstruct 3D scenes and human faces. The process is start with the image transformation into coordinates. Next step is to copy image with the help of multi-resolution technique with Laacian and Pyramidal Gaussian. It results the same size images but with lessen bandwidth. It causes special blurring effect in original image that is called scale space or scale invariant. The algorithm has three main parts:

1. Interest point detection
2. Local neighborhood description
3. Matching



(a)

(b)

Fig. 1 Matching of varying images by (a) SIFT and (b) SURF

4.2 Histogram of Oriented Gradients (HOG)

The histogram of oriented gradients (HOG) is another feature descriptor used for digital image processing as well as in computer vision. It is first proposed by Navneet Dalal and Trigg in year of 2005 for the purpose of object detection. In an image, it calculates the occurrences of gradient orientation in localized portions in visual. It is just a parallel methodology used for edge orientation histogram. Previous descriptors are only for shape contexts and scale-invariant transformations. Accuracy improvement is the next task and achieved by the computation on a compact of grid of uniformly spaced cells. It utilizes the overlying of local contrast normalization. Interconnection of cells is the next task resulting from subdivisions of the image. Compilation of HOG performed for each pixel within cells. After cells, working on local histogram performed the contrast-normalized operation. It can be done by computing the measure of intensity across a larger region of the image, called a block. After that, it uses obtained values for improving the accuracy by normalizing all the cells placed in the block. Merits of this procedure are better performance under strong illumination and shadowing. Testing of HOG performed with MIT data set, which features 200 datasets combine with 509 training set. This procedure mainly contains the pictures of humans with back and front face. It generated significant outcomes which made it well acclaimed. HOG consist of the following steps:

1. Gradient computation
2. Orientation binning
3. Descriptor blocks
4. Block normalization
5. SVM classifier
6. Neural Network Classifier

V. CONCLUSION

Important methodologies with their gaps are discussed in this paper. Digital image processing needs to reduce the storage capacities with the raid increase in multimedia visuals. Color, texture and shape of an object are the ultimate features to improve the extraction process of an image. Certain algorithms based on mathematical models have been developed to extract the information from color under the regions of low frequency after wavelet dissociation. When computation process is not great, this method makes the texture and color features together on same task. Resulting tests show that hybrid features overcome the issues arises due to human's sight. The further research on single as well as the combination of features and their characters is going on.

VI. REFERENCES

[1] B. Moghaddam, C. Nastar and A. Pentland, "A Bayesian similarity measure for deformable image matching," *Image and Vision Computing*, vol. 19, no. 5, pp. 235-244, 2001.

[2] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. "BRIEF: Binary robust independent elementary features," *In European Conference on Computer Vision*, 2010. \

[3] J. Faidit, "Planetariums in the world," *Proceedings of the International Astronomical Union*, vol. 5, no. 260, 2009..

[4] Song Yan, "Research of Image Retrieval Based on color and texture feature", Computer application and soft, 2007, 9(2),pp.42-50.

[5] Shi Yan, "The study of Image Retriveal Technique Based on Wavelet Transform", North University,2006, pp.24-25.

[6] Zeng Tai-sheng, "Study on the image retrieval of based on color and texture and feedback", Hua Qiao University,2006, pp.34-25.

[7] Hafner J, Sawhney H S , "Efficient color histogram indexing for quadratic form distance functions", IEEE Transactions on Pattern Analysis and Machine Intelligence, Intelligence,1995,17(7): pp.729- 736.

[8] Dong Wei-jun, Zhou Ming-quan, Geng Guo-hua, "Research of Image Retrieval Based on compositive feature ", Computer application and soft, 2005,11(11): pp. 34-50.

[9] Pan Hui, "A robust feature extraction arithmetic to geometric distortions for texture image", Zhejiang university, 2007, pp.14-36..

[10] Peter Howarth, Stefan M Roger, "Evaluation of Texture Features for Content-Based Image Retrieval", Third International Conference, CIVR, 2004, pp. 326-334..

[11] M. Quartulli and I. G Olaizola, "A review of EO image information mining," *ISPRS J. Photogramm. Remote Sens.*, vol. 75, pp. 11–28, 2013

[12] Xiang Fu, Jie-xian Zeng, "A novel image retrieval method based on interest points matching and distribution", Chinese Journal of Lasers, vol. 37(3), pp. 774-778, 2010.

[13] Meng Fan-jie, Guo Bao-long, "A novel image retrieval algorithm based on the color and distribution of interest points" Journal of Xidian University, vol.32, pp.256-259, Feb 2005.