FOURIER TRANSFORM BASED SALIENCY DETECTION FOR SKETCH BASED IMAGE RETRIEVAL SYSTEMS

K. Durga Prasad*[1], Dr. K. Manjunathachari[2], Dr. M. N. Giriprasad[3]

[1]Asst. Prof, Vardhaman College of Engineering, Samsabad, Hyderabad, Telangana, India.

[2]Head of the Department, Dept. of ECE, GITAM University, Hyderabad, Telangana, India.

[3]Prof. & HOD, Department of ECE, JNTUA College of Engg, Anantapur (A.P), India.

ABSTRACT

Visual Saliency is the perceptual quality that makes an object or pixel region stand out relative to its neighbors and thus capture viewer attention. Thus Saliency models indicate the region of interest. This paper aims at improving the performance of sketch based image retrieval using saliency detection approach. Several methods have been developed to extract the saliency information from an image. We use the Fourier transform to detect the saliency. Experimental results prove that the proposed technique outperforms the existing techniques and produce better retrieval results.

1. INTODUCTION

The content-based image retrieval systems retrieve images from database using visual information such as color, texture, or shape."Content-based" means that the search analyzes the contents of the image rather than the metadata such as keywords, tags, or descriptions associated with the image. The term "content" in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself [1]. In most of the systems, the user queries by presenting a name or an example image that has the desired features.Before the actual application is ready, a large amount data has to be managed, processed and stored. The growth of data storages and revolution of internet has changed the world. The efficiency of searching an information set is a very important. In case of texts we can search flexibly using keywords, but if we use images, we cannot apply dynamic methods. There are two issues in this aspect. The first is who yields the keywords. And the second is how well can an image be represented by keywords. Having humans manually annotate images by entering keywords or metadata in a large database can be time consuming and may not capture the keywords desired to describe the image. In many cases if we want to search efficiently some data have to be recalled. Human beings are able to recall visual information more easily using features like the shape of an object or arrangement of colors on objects and based on other contextual information. In the case of query image based image retrieval systems, the features of the image are used for searching similar images.But at this moment unfortunately there are not frequently used retrieval systems, which retrieve images using the non-textual information of a sample image. Our purpose is to develop a content based image retrieval system, which can retrieve using images in frequently used databases[2]. Whenever the user provides a query image, the features of the image are extracted. These features are then matched with prior extracted features of the images from the database. Based on the matching indices, the results are displayed. The interaction between user and CBIR system can help in achieving better retrieval results. The interaction ranges from simply allowing the user to submit a new query based on a existing one to giving the user thepossibility to select part of the result image as relevant and non-relevant to allow the user visually arrange a small set of the database images.

2. LITERATURE SURVEY

There exist different approaches to saliency detection. Before describing state-of-art methods it is good to introduce main approaches to saliency detection. Saliency detection methods can be grouped according to the model inspiration source. For instance, Itti’s approach is referred as a biological inspired method. Such methods explore peculiarities of human vision and attention operation and try to mimic the processes taking place while a

*Corresponding Author

human observes a scene. Another group of methods explore natural statistics found in images. For instance, in Hou et al proposed spectral residual approach that exploits spectral histogram singularities to detect salient regions in images [9].

Another common approach of saliency is computational. These methods exploit information domain properties to detect salient regions. Another grouping can be made considering what type of task the authors addressed in their works. Here, two groups are possible: human fixations and region of interest. Although these two tasks may sound similarly, there is a noticeable difference in output maps. Human vision system works with very sparse data and fixations are usually found only on a small portion of object's area. Thus a method trying to predict human fixations would highlight edges and contrast spots of an object.

Methods aimed at detection of salient regions should provide a map highlighting the whole object of interest. Often, in region-based methods human fixations map is an interim product that is further developed into region-based map by means of region growing or segmentation. Output map can also differ in their representation. Here two options are possible: binary and grayscale. The former draws salient pixels/regions in white and non-salient in black. The later usually represents probability of a region to be salient by different tones of gray. For region of interest detection methods output maps can also differ in how the salient region is highlighted. Some authors prefer to use rectangular windows, others use segmentation mask and paint different segments according to their value of saliency and lastly region masks can be represented by raw pixels.

3. PROPOSED APPROACH

The saliency detection model proposed by Itti et al resulted in an avalanche of different saliency detection algorithm. One of the most recent and well-excepted extensions of Itti's saliency detection approach was presented by Judd et al in . Although Itti's approach was taken as a basis the authors have combined a much broader set of features. The authors proposed to use three levels of features: low level, mid-level and high-level [3]. Low level is formed by intensity, orientation and colour contrast features as they were defined in the Itti's work. In addition, the authors included distance to centre, local energy pyramids, and probability of colour channels computed from 3D colour histograms with median filter at 6 scales. The mid-level is formed by horizon line detector.

Finally high level features are formed by Viola-Jones face and person detectors. The classification is done using SVM with linear kernel. Before features are extracted from an image it is resized to 200 × 200 px, thus original aspect ratio of the image is loosen. For the training and evaluation of their model the authors collected a database of 1003 images from photo sharing services and collected human fixations from 15 users [4]. It is worth mentioning the training settings the authors used. Instead of direct parsing training images data and ground truth labels to SVM, the authors performed selection of data for training.

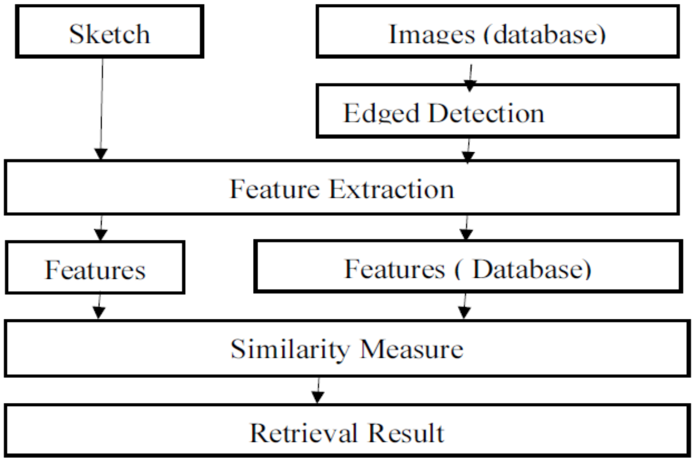


Fig 1: Sketch Based Image Retrieval

4. FOURIER TRANSFORM APPROACH

Some processes performed on an image in the spatial domain may be very computationally expensive. These same processes may be significantly easier to perform after transforming an image to a different domain. These transformations are the basis for many image filters, applied to remove noise, to sharpen, or extract features. Domain transformations also provide additional information about an image and can offer compression benefits. The most common representation of a pixel's value and location is spatial, where it appears in three dimensions (x, y, and z). Pixel value and location in this space is usually referred to by column (x), row (y), and value (z), and is known as the spatial domain. However, a pixel's value and location can be represented in other domains [7]. In the frequency or Fourier domain, the value and location are represented by sinusoidal relationships that depend upon the frequency of a pixel occurring within an image. In this domain, pixel location is represented by its x- and y-frequencies and its value is represented by an amplitude. Images can be transformed into the frequency domain to determine which pixels contain more important information and whether repeating patterns occur.

In the time-frequency or wavelet domain, the value and location are represented by sinusoidal relationships that only partially transform the image into the frequency domain. Like the transformation to the full frequency domain, the transformation to the time-frequency domain helps to determine the important information in an image. In the Hough domain, pixels are presented by sinusoidal lines. Since straight lines within an image are transformed into the Hough domain as intersecting sinusoidal lines, these intersections can be used to determine if and where straight lines occur within an image. In the Radon domain, a line of pixels occurring in an image is represented by a single point. This transformation is useful for detecting specific features and image compression. Since transforming images to and from the Hough and Radon domains use similar methods, the Radon image representation is described in the same section as the Hough representation.

The Fast Fourier Transform (FFT) is used in numerical analysis to transform an image between spatial and frequency domains. The FFT decomposes an image into sines and cosines of varying amplitudes and phases. The values of the resulting transform represent the amplitudes of particular horizontal and vertical frequencies. This image information in the frequency domain shows how often patterns are repeated within an image. Low frequencies represent gradual variations in an image, while high frequencies correspond to abrupt variations in the image.

Low frequencies tend to contain the most information because they determine the overall shape or pattern in the image. High frequencies provide detail in the image, but they are often contaminated by the spurious effects of noise. Masks can be easily applied to the image within the frequency domain to remove the noise. The FFT process is the basis for many filters used in image processing. One of the easiest FFT filters to understand is the one used for background noise removal. This filter is simply a mask applied to the image in the frequency domain. The Fourier Transform is an important image processing tool which is used to decompose an image into its sine and cosine components. The output of the transformation represents the image in the *Fourier* or frequency domain, while the input image is the spatial domain equivalent.

In the Fourier domain image, each point represents a particular frequency contained in the spatial domain image. The DFT is the sampled Fourier Transform and therefore does not contain all frequencies forming an image, but only a set of samples which is large enough to fully describe the spatial domain image. The number of frequencies corresponds to the number of pixels in the spatial domain image, *i.e.* the image in the spatial and Fourier domain are of the same size.

For a square image of size N×N, the two-dimensional DFT is given by equ.

$$F(k,l) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} f(i,j)\, e^{-\iota 2\pi\left(\frac{ki}{N}+\frac{lj}{N}\right)}$$

where *f(a,b)* is the image in the spatial domain and the exponential term is the basis function corresponding to each point *F(k,l)* in the Fourier space. The equation can be interpreted as: the value of each point *F(k,l)* is obtained by multiplying the spatial image with the corresponding base function and summing the result.

The basis functions are sine and cosine waves with increasing frequencies, *i.e. F(0,0)* represents the DC-component of the image which corresponds to the average brightness and *F(N-1,N-1)* represents the highest frequency.In a similar way, the Fourier image can be re-transformed to the spatial domain. The inverse Fourier transform is given by equ

$$f(a,b) = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} F(k,l)\, e^{\iota 2\pi \left(\frac{ka}{N} + \frac{lb}{N}\right)}$$

Note the $\frac{1}{N^2}$ normalization term in the inverse transformation. This normalization is sometimes applied to the forward transform instead of the inverse transform, but it should not be used for both. To obtain the result for the above equations, a double sum has to be calculated for each image point. However, because the Fourier Transform is *separable*, it can be written as equations

$$F(k,l) = \frac{1}{N} \sum_{b=0}^{N-1} P(k,b)\, e^{-\iota 2\pi \frac{lb}{N}}$$

where

$$P(k,b) = \frac{1}{N} \sum_{a=0}^{N-1} f(a,b)\, e^{-\iota 2\pi \frac{ka}{N}}$$

Using these two formulas, the spatial domain image is first transformed into an intermediate image using *N* one-dimensional Fourier Transforms [8]. This intermediate image is then transformed into the final image, again using *N*one-dimensional Fourier Transforms. Expressing the two-dimensional Fourier Transform in terms of a series of *2N* one-dimensional transforms decreases the number of required computations.

Even with these computational savings, the ordinary one-dimensional DFT has N^2 complexity. This can be reduced to $N \log_2 N$ if we employ the Fast Fourier Transform (FFT) to compute the one-dimensional DFTs. This is a significant improvement, in particular for large images. There are various forms of the FFT and most of them restrict the size of the input image that may be transformed, often to $N = 2^n$ where *n* is an integer. The mathematical details are well described in the literature.

5. EXPERIMENTAL RESULTS

The following images constitute the database.

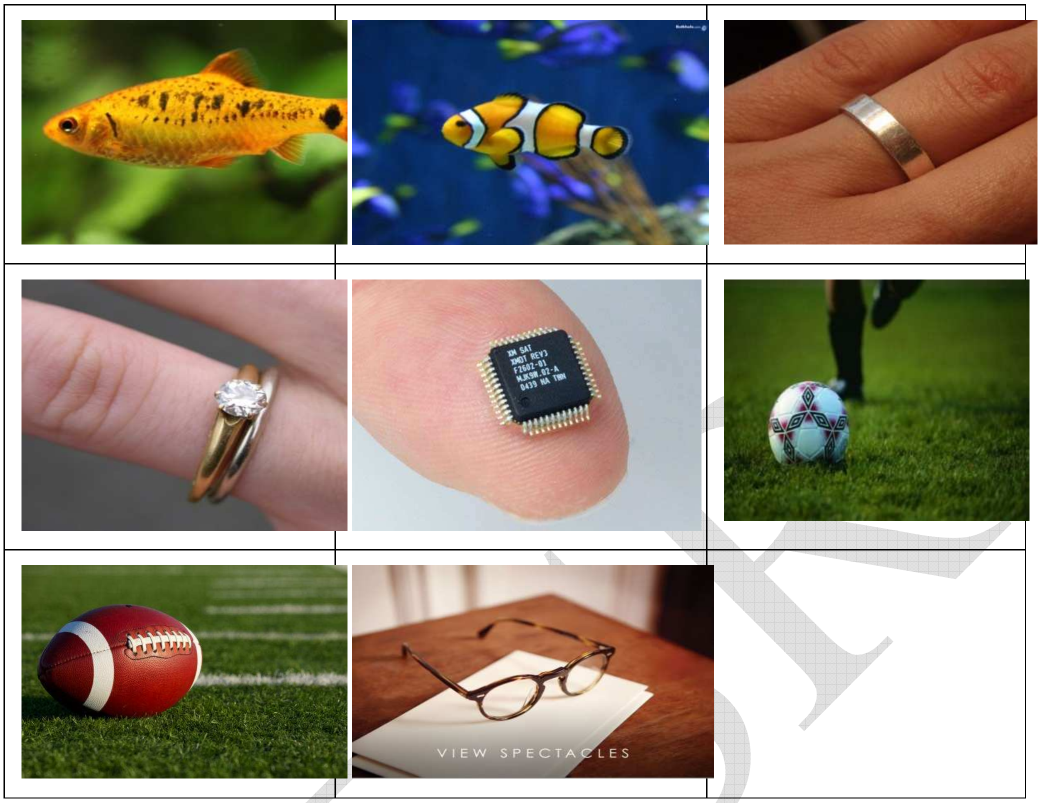


Fig 2: Database images

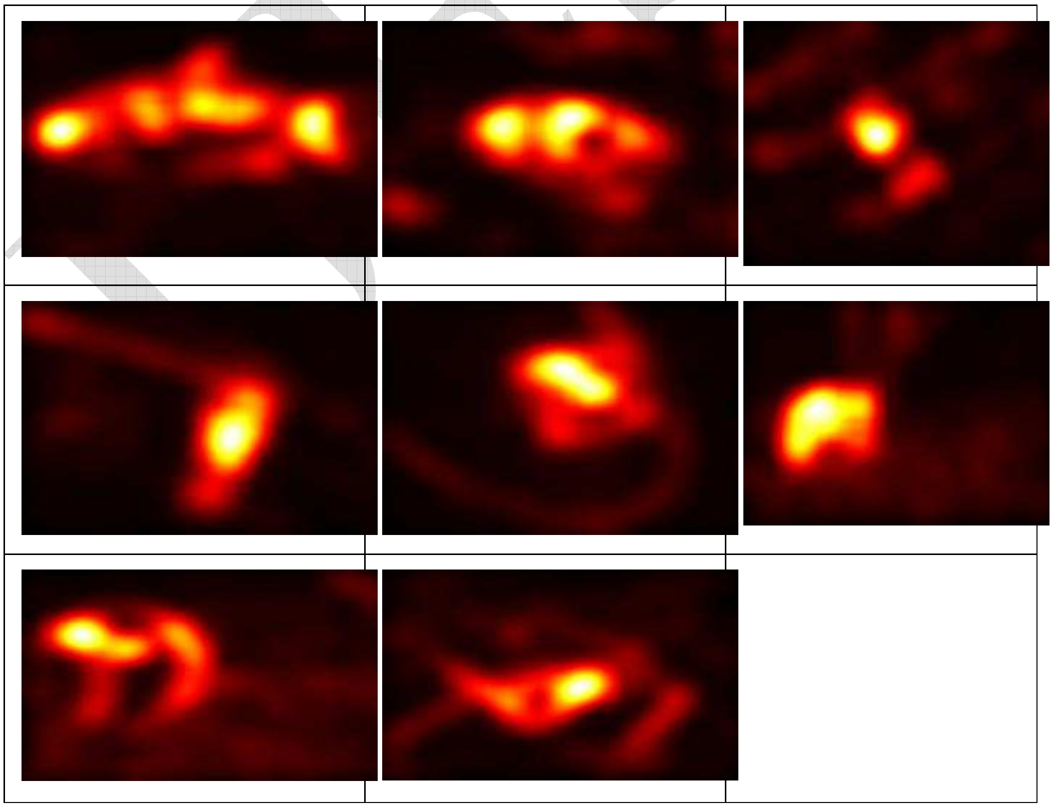


Fig 3: Saliency detection output of the DB images

The following images illustrate how saliency extraction would affect the edge detection process.

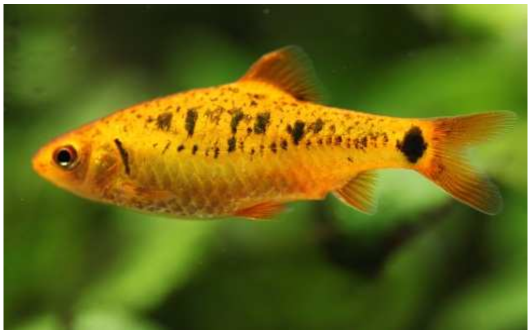


Fig 4: Input image

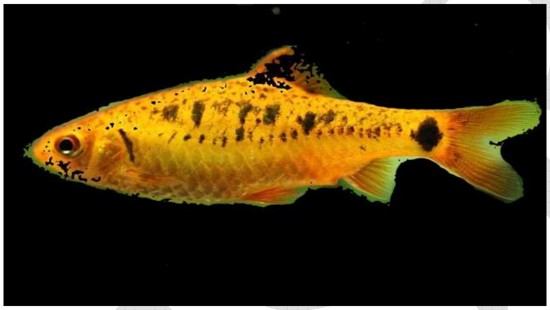


Fig 5: Output Image - saliency

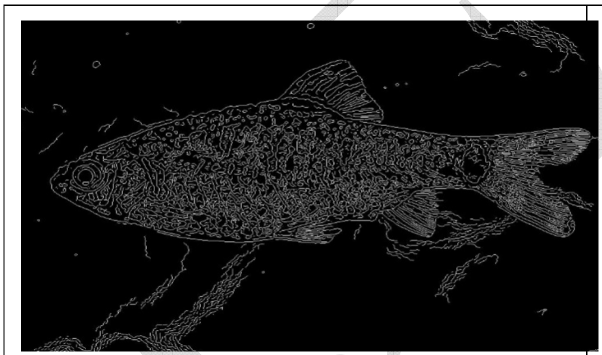
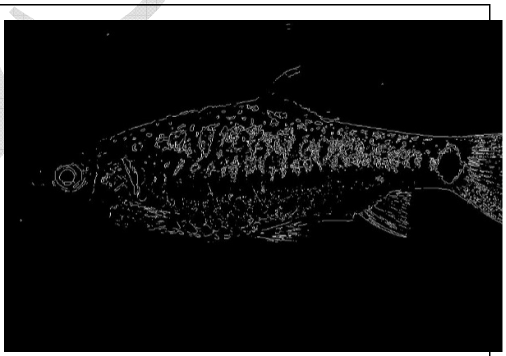
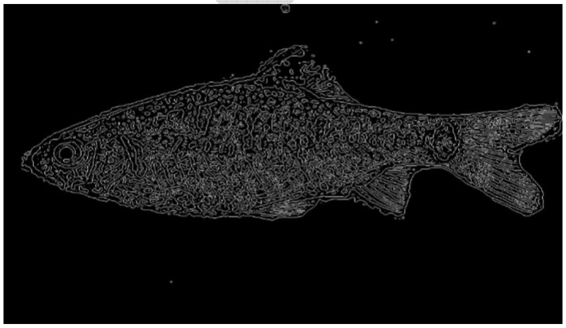
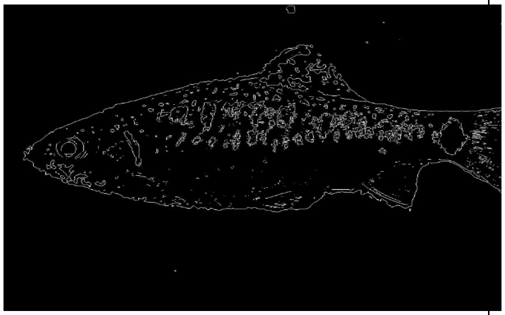
	
Canny edge detection of input	Sobel edge detection of input
	
Canny edge detection of output	Sobel edge detection of output

Fig 6: Outputs of edge detection

The edges extracted from the image vary in both the images drastically. The number of extra edges are reduced clearly. This would have a direct effect on efficiency of the sketch based image retrieval. The non-salient regions can be removed from the image thus increasing the accuracy of the image retrieval system.

6. CONCLUSION

This paper presents an image retrieval system based on image saliency detection using Fourier transform. As the saliency output produces only the regions in the ROI, preprocessing techniques like region segmentation extract accurate features by eliminating false ones. Thus improving the accuracy of the image retrieval system. The experimental results performed have proved that the performance of the image retrieval system with complex background images has been improved.

REFERENCES

[1] Eitz M, Hildebrand K, Boubekeur T, Alexa M. Sketch-Based Image Retrieval: Benchmarkand Bag-of-Features Descriptors. IEEE Transactions On Visualization And Computer Graphics 2011; 17(11).

[2] Oliva A, Torralba A. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. Int’l J. ComputerVision 2001; 42(3): 145-175.

[3] Lowe DG. Distinctive Image Features from Scale-Invariant Keypoints. Int’l J. Computer Vision 2004; 60(2): 91-110.

[4] Squire D, Mueller W, Mueller H, Raki J. Content-Based Query of Image Databases, Inspirations from Text Retrieval: Inverted Files, Frequency-Based Weights and Relevance Feedback. Proc. Scandinavian Conf. Image Analysis, 1999; 7-11.

[5] Sivic J, Zisserman A. Video Google: A Text Retrieval Approach to Object Matching in Videos. Proc. IEEE Int’l Conf. Computer Vision, 2003; 1470-1477.

[6] Jegou H, Douze M, Schmid C. Hamming Embedding and Weak Geometric Consistency for Large Scale Image Search. Proc.European Conf. Computer Vision, 2008; 304-317.

[7] Wu Z, Ke Q, Isard M, Sun J. Bundling Features for Large Scale Partial-Duplicate Web Image Search. Proc. IEEE Conf.Computer Vision and Pattern Recognition, 2009; 25-32.

[8] Forsyth DA. Benchmarks for Storage and Retrieval in Multimedia Databases. Proc. Storage and Retrieval for Media Databases 2002; 240-247.

[9] Scale, Saliency and Image Description. Timor Kadir and Michael Brady. International Journal of Computer Vision 2001; 45 (2): 83–105.

[10] Kadir T, Zisserman A, Brady M. An affine invariant salient region detector. Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic, 2004.

[11] Shao L, Kadir T, Brady M. Geometric and Photometric Invariant Distinctive Regions Detection. Information Sciences 2007; 177 (4): 1088-1122.

[12] Li W, Bebis G, Bourbakis NG. 3-D Object Recognition Using 2-D Views. IEEE Transactions on Image Processing 2008; 17(11): 2236–2255.