

# A System Which Retrieves Images Using a Query Image

Mr. Eshwar Erva\*,  
Aurora's Technological Research Institute,  
Hyderabad

Mr. K Chandra Shekar\*\*  
Aurora's Technological Research Institute,  
Hyderabad

**Abstract:-** *It's a new technology to retrieve images without using any verbal descriptions. If an image equals to 1000 words, then imagine the difficulty of cataloging a database of up to 1, 00,000 images. For this providing a written description for each image, not only for image, the cataloger have to describe the color, shape and texture of every element within each picture and its relationship with other element, but if anyone searching for an image would have to guess what exactly words has to be use in the description. If the computers are capable of displaying millions of colors and an effectively infinite number of shapes and textures, the task would be undesirable, but not impossible. Image databases can be sorted and queried by color, shape and texture. It's a big and time consuming process, to avoid this we were using image instead of verbal descriptions, so that image work as a query and match this query to retrieve images from databases. For this we use color, shape, texture and histogram techniques.*

**Keywords:** *Color Descriptors, Content, Image Retrieval, Shape, Texture.*

## 1. Introduction

If a system retrieves images from databases without using text or any verbal descriptions, then this a quite new technology in retrieving images. The technology behind this process is converting image content into their respective properties like texture, shape, content of the image, and color. At first glance, content based querying appears deceptively simple because we humans seem to be so good at it. If a program can be written to extract semantically relevant text phrases from images, the problem may be solved by using currently available text search technology. But, in an unconstrained environment, the task of writing this program is beyond the reach of current technology in image understanding. At an article intelligence conference several years ago, a challenge was issued to the audience to write a program that would identify all the dogs pictured in a children's book, a task most 3 years old can easily accomplished. Nobody in the audience accepted the challenge and this remains an open problem. The process of grouping image features into meaningful objects and attaching semantic descriptions to scenes through model matching is an unsolved problem in image understanding.

Humans are much better then computers, however, are better than humans at measuring properties [1] and retaining then in long term memory. We know that an image is equals to thousand words. But, these thousand words may differ from one individual to another depending on their respective knowledge of the image context. For example, fig.1 gives a familiar demonstration that an image can have multiple and quite different interpretations. Thus, even if a thousand word image descriptions were available, it is not certain that the image could be retrieved by a user with a different description. The problem is fundamentally one of communication between information or user and the image retrieval system. Here we go for the feature space analysis; it is a widely used tool for solving low-level images for understanding tasks. If given a image, feature vectors then they are extracted from local neighbourhoods and mapped into the space spanned by their components. Significant features in the image then they correspond to high density regions in this space. Feature space analysis is the procedure of recovering the centers of the high density regions, i.e., the representations of the significant image features, histogram based techniques [2], Hough transform are examples of the approach. For detecting humans in images is a challenging task owing to their variable appearance and the wide range of poses that they can adopt. The first need is a robust feature set that allows the human form to be discriminated cleanly, even under the cluttered backgrounds under difficult illumination.



Figure.1. One or two faces

Here we study the issue of feature sets in detection of humans in images; shows that locally normalized Histogram of Oriented Gradient descriptors provide excellent performance relative to other existing feature sets including

wavelets. The proposed descriptors are reminiscent in edge orientation histograms [3], SIFT descriptors and shape contexts, but they are computed on a dense grid of uniformly spaced cells and they use overlapping local contrast normalizations for improving performance. We make a detailed study of the effects of various implementation choices on detector performance, taking pedestrian detection as a test case. To make simple and fast, we use linear SVM as a baseline classifier throughout the study. In these new detectors give essentially perfect results on the master installation tape pedestrian test set, so we have created a more challenging set containing over 1900 pedestrian images with a large range of poses and backgrounds. Our ongoing work suggests that feature set performs equally well for other shape based object classes. Since the user may have different needs and knowledge about the image collection, an image retrieval system must support various forms for query formulation. While computationally expensive, the results are far more accurate than conventional image indexing. Hence, there exists a tradeoff between accuracy and computational cost. This tradeoff decreases as more efficient algorithms are utilized and increased computational power becomes inexpensive. This involves entering an image as a query into a software application that is designed to employ CBIR [4] techniques in extracting visual properties, and matching them. This is done to retrieve images in the database that are visually similar to the query image.

## 2. Related work

A huge number of data had to be managed, processed and to be stored in database while spreading the information technology. It was also useful for textual and visual information parallel for the appearance and quick evolution of computers an increasing measure of data had to be managed. The growing of data storages and revolution of internet had changed the world. The efficiency of searching for information is a very important point of view. In case of texts we can search flexibly using keywords, but if we use images, we cannot apply dynamic methods. For this two questions may arrive. The first is who yields the keywords. And the second is an image can be well represented by keywords.

### 2.1 Query by image content

At first glance, content based querying appears deceptively simple because we humans seem to be so good at it. If a program can be written to extract semantically relevant text phrases from images, the problem may be solved by using currently available text search technology. But, in an unconstrained environment, the task of writing this program is beyond the reach of current technology in image understanding. At an article intelligence conference several years ago, a challenge was issued to the audience to write a program that would identify all the dogs pictured in a children's book, a task most 3 years old can easily accomplished. Nobody in the audience accepted the challenge and this remains an open problem. The process of grouping image features into meaningful objects and attaching semantic

descriptions to scenes through model matching is an unsolved problem in image understanding. Humans are much better than computers, however, are better than humans at measuring properties [1] and retaining them in long term memory.

### 2.2. Content Based Image Retrieval System

In CBIR - Content Based Image Retrieval, as currently defined by the image retrieval community, the term 'image content' refers to the implementation content of the image, i.e. to its pixel content. The objective of CBIR [4] research and development activities is to develop automated routines that can analyze a digital image or image stream and identify the objects, actions and events that it portrays. Currently, CBIR routines are relatively adept at identifying color and texture distributions, as well as primitive shapes. This information is the basis for specification of one or more image signatures that can act as a surrogate for the image and that can be indexed to provide rapid access to elements in the image collection. To date, there is little ability to connect this information to the semantics of the image, i.e. to object, action(s) and event identification. Notable exceptions include image collections from very specific domains, such as face and fingerprint recognition, recognition of faults in specific structures, e.g. pipe systems or bridge foundations, and to some extent LOGO identification.

### 2.3 Content, Descriptors, Color and Texture

Basically image is a long string of pixels for which each pixel is identified by its place in the image matrix, its color and its intensity. Analysis of the pixel set can give information about the distribution of dominant colors, the image texture, and the shapes formed by marked change in neighboring colors. Although there are many techniques used for color and texture extraction, but they have their variations and refinements of a basic color profile calculation that:

- It determines the number of colors to use and divides the color spectrum into related color sets.
- It determines the grid characteristics for the image i.e., number, shape and placement of the grid cells to be used for image analysis.
- It sums the number of pixels of each color set in each grid cell into a feature vector, and alternatively called an image signature or feature histogram.
- It indexes the feature vector or signature for each image in the database.

Query image is given as an example image or seed image or a sketch [5] for which a feature vector is calculated in the same way as that used for the database image set. The query signature is then compared to the database image signatures, using a distance measure. Low or pixel level feature-based queries tend to compare whole images and are typically formulated as "find images similar to "this one"". As for text, weights can be given to the feature set used in an image query. Commonly a user uses a query formation that is: category browsing, query by concept, query by sketch, and query by

example. Category browsing is to browse generally speaking about image content that may include both visual descriptors and semantic descriptors [6].

## 2.4 Identifying Shapes, Image Objects

Identification of shape and recognition of shape is the most difficult challenge to image analysis since it relies on:

- Isolation of the different objects or shapes within the image, which may not necessarily be whole or standardized, but are likely 'hidden' in the perspectives of the image.
- Normalizing the object's size and rotation Identification and possible connection of object parts, for example 'completing' the car which has a person standing in front of it.
- The semantic identification of the image components or objects.

Up to date, automatic objects recognition has only been accomplished for well defined domains where objects of interest are well known and well defined within the image, such as:

- Police images of faces and fingerprints,
- Medical images from domain x-rays,
- MRT and CT scans, and
- Industrial surveillance of building structures, such as bridges, tunnels or pipelines.

An object-based visual query requires a query language that can accept a visual object/image as an example, asking a question like: find images that contain "this image".

These contains operator for shape or object identification needs to be adapted for image retrieval such that in addition to the Boolean operators AND, OR and NOT, the location/spatial operators should include near, overlapping, within, in foreground or background. The shape of thesaurus can be developed for image collections from specific domains that can be used in a way, similar to that done for the terms in text document collections. This thesaurus can be used to identify shapes in both the image database and the visual query, increasing the likelihood of a good or relevant result list.

## 3. Our Proposed

Target search is to retrieve specific target image such as a particular image or a registered logo or a specific vintage model or a car or historical photograph. There are some existing techniques, designed around query refinement based on relevance feedback, this approach suffer from slow convergence, and do not guarantee to find intended targets. To overcome these limitations, we propose several efficient query point movement methods. We prove that our approach is able to reach any given target image with fewer iterations in the

worst and average cases. We propose a new index structure and query processing technique to improve retrieval effectiveness and efficiency. We also consider strategies to minimize the effects of users' inaccurate relevance feedback. Extensive experiments in simulated and realistic environments show that our approach significantly reduces the number of required iterations and improves overall retrieval performance. The experimental results also confirm that our approach can always retrieve intended targets even with poor selection of initial query points.

### 3.1 Query Image

In this we specifying what kind of images a user wishes to retrieve from the database can be done. Commonly a user uses a query formation that is: category browsing, query by concept, query by sketch, and query by example. Category browsing is to browse generally speaking about image content that may include both visual and semantic content. Visual content can be very general or domain specific. General visual content include color, texture, shape, spatial relationship, etc. Domain specific visual content, like human faces, is application dependent and may involve domain knowledge. A visual content descriptor can be either global or local. A global descriptor uses the visual features of the whole image, whereas a local descriptor uses the visual features of regions or objects to describe the image content. To obtain the local visual descriptors, an image is often divided into parts first. The simplest way of dividing an image is to use a partition, which cuts the image into tiles of equal size and shape. A simple partition does not generate perceptually meaningful regions but is a way of representing the global features of the image at a finer resolution. A better method is to divide the image into homogenous regions according to some criterion using region segmentation algorithms that have been extensively investigated in computer vision. A more complex way of dividing an image, is to undertake a complete object segmentation to obtain semantically meaningful objects (like ball, car, horse). Currently, automatic object segmentation for broad domains of general images is unlikely to succeed. Semantic contents obtained either by textual annotation or by complex inference procedures based on visual content. But we concentrate on general visual contents descriptions and then semantic contents. A good visual content descriptor [6] should be invariant to the accidental variance introduced by the imaging process (e.g., the variation of the illuminant of the scene). However, there is a tradeoff between the invariance and the discriminative power of visual features, since a very wide class of invariance loses the ability to discriminate between essential differences. Invariant description has been largely investigated in computer vision (like object recognition), but is relatively new in image retrieval.

#### 3.1.1 Indexing scheme

This is a effective indexing and fast searching of images based on visual features. Because the feature vectors of images tend to have high dimensionality and not well suited to traditional indexing structures, dimension reduction is

usually used before setting up an efficient indexing scheme. One of the techniques commonly used for dimension reduction is Principal Component Analysis. It is an optimal technique that linearly maps input data to a coordinate space such that the axes are aligned to reflect the maximum variations in the data. In addition to PCA, many researchers have used KL transform to reduce the dimensions of the feature space. Although the KL transform has some useful properties such as the ability to locate the most important subspace, the feature properties that are important for identifying the pattern similarity may be destroyed during blind dimensionality reduction [7]. Apart from PCA and KL transformation, neural network has also been demonstrated to be a useful tool for dimension reduction of features [8]. After dimension reduction, the multi-dimensional data are indexed. A number of approaches have been proposed for this purpose, including R-tree (particularly, R\*-tree), linear quad-trees, K-d-B tree and grid files. Most of these multi-dimensional indexing methods have reasonable performance for a small number of dimensions, but explore exponentially with the increasing of the dimensionality and eventually reduce to sequential searching. Furthermore, these indexing schemes assume that the underlying feature comparison is based on the Euclidean distance, which is not necessarily true for many image retrieval applications. One attempt to solve the indexing problems is to use hierarchical indexing scheme based on the Self-Organization Map [9]. In addition to benefiting indexing, SOM provides users a useful tool to browse the representative images of each type.

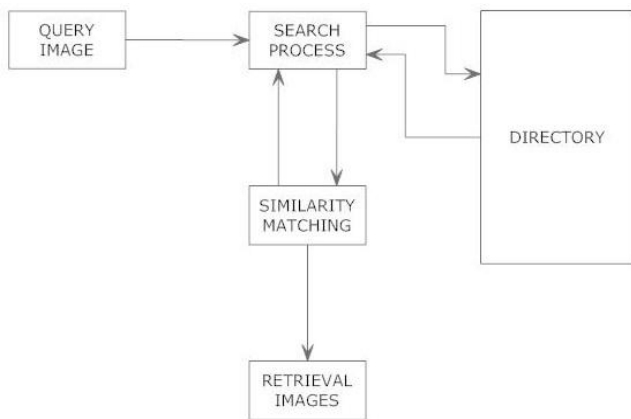


Figure.2. Architecture

### 3.2 Similarity matching

Similarity matching is technique, instead of exact matching; this retrieval system calculates visual similarities between a query image and images in a database. While coming to the retrieval process the result is not a single image but a list of images ranked by their similarities with the query image. Many similarity measures have been developed for image retrieval based on empirical estimates of the distribution of features in recent years. Different similarity or distance measures will affect retrieval performances of an image retrieval system significantly. In this, we will introduce

some commonly used similarity measures. We denote  $D(I, J)$  as the distance measure between the query image  $i$  and the image  $J$  in the database; and  $f_i(I)$  as the number of pixels in bin  $i(I)$ .

### Minkowski-Form Distance

If each dimension of image feature vector is independent of each other and is of equal importance, the Minkowski-form distance  $L_p$  is appropriate for calculating the distance between two images. This distance is defined as:

$$D(I, J) = (\sum_i |f_i(I) - f_i(J)|^p)^{1/p}$$

When  $p=1, 2,$  and  $\infty$ ,  $D(I, J)$  is the  $L_1, L_2$  (also called Euclidean distance), and  $L_\infty$  distance respectively. Minkowski-form distance is the most widely used metric for image retrieval. For instance, MARS system [10] used Euclidean distance to compute the similarity between texture features; Netra [11, 12] used Euclidean distance for color and shape feature, and  $L_1$  distance for texture feature; Blob world [13] used Euclidean distance for texture and shape feature. In addition, Voorhees and Poggio [14] used  $L_\infty$  distance to compute the similarity between texture images. The Histogram intersection can be taken as a special case of  $L_1$  distance, which is used by Swain and Ballard [15] to compute the similarity between color images. The intersection of the two histograms of  $I$  and  $J$  is defined as:

$$S(I, J) = \sum_{i=1}^N \min(f_i(I), f_i(J)) / \sum_{i=1}^N f_i(J)$$

It has been shown that histogram intersection is fairly insensitive to changes in image resolution, histogram size, occlusion, depth, and viewing point.

### 3.3 Retrieval Images

In this the results of the search is displayed and in displaying images in a specified format is the thumbnails. It can be formed by Prima facie the first  $n$  pieces of results can be displayed, which conveniently can be placed in the user interface. This number depends on the resolution of the monitor, and for as much the large resolution monitors are widely used, so this number can move between 20 and 40. Another approach is to define the maximum number of results ( $n$ ), but we also observe that how the goodness of individual results can vary. If the retrieval effectiveness is worse by only a given ratio, the image can be included in the display list. In our system the possible results are classified. Hence the solution set is more ordered and transparent. By default the results are displayed by relevance, but false-positive results can be occurred, which worsen the retrieval results. If the results are reclassified in according to some criterion, then the number of false positive results decreases. Thus the user perception is better. Since the color-based clustering for us is the best solution, so our choice was the k-means clustering method [2], which is perfectly suited for this purpose. The implemented user interface can be seen in Figure 3 and Figure



4. Our program has been written in JAVA, and during the implementation some new idea was considered.

### 4. Results

Process of the System Which Retrieves Images Using a Query Image can be started by indexing the contents after that user gives a query image for searching the related images, in this process the various techniques which are described above are used to transform the query to respective matching techniques with these techniques the retrieval of images are resulted

Our system was tested with more than one database to obtain a more extensive description of its positive and negative properties. The Microsoft Research Cambridge Object Recognition Image Database was used, which contains 209 realistic objects. All objects have been taken from 14 different orientations with 450x450 resolution. The images are stored in TIF format with 24 bits. This database is most often used in computer and psychology studies. Some images of this database can be seen in Fig 3.

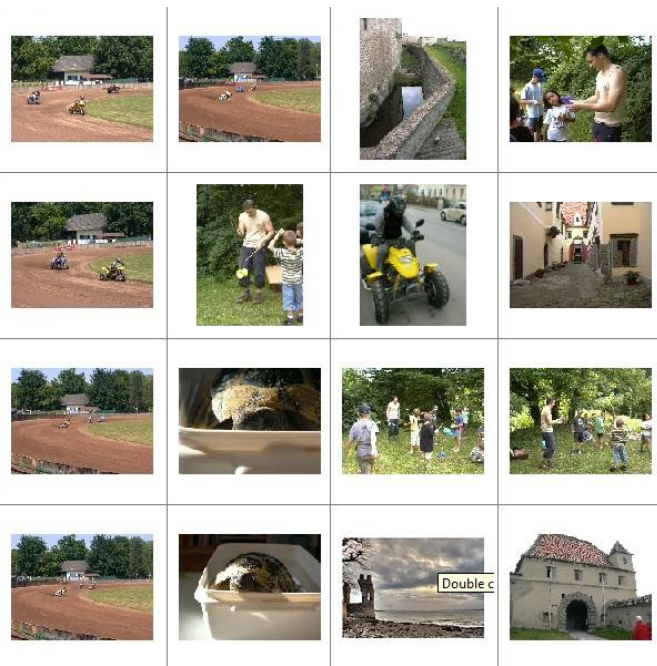


Figure.3. Search results performed on Microsoft Research Cambridge Object database.

Our database was the Flickr 160. This database was used before for measuring of a dictionary-based retrieval system [8]. 160 pieces of general-themed pictures have sorted from the photo sharing website called Flickr. These images can be classified in to 5 classes based on their shape. A lot of images contain the same building and moments. The database is accompanied by examples, which is based on the retrieval. Since the test result are documented and the retrieved sketches are also available, so the two systems can be compared with

each other. Some images of Flickr 160 database can be seen in Figure 4

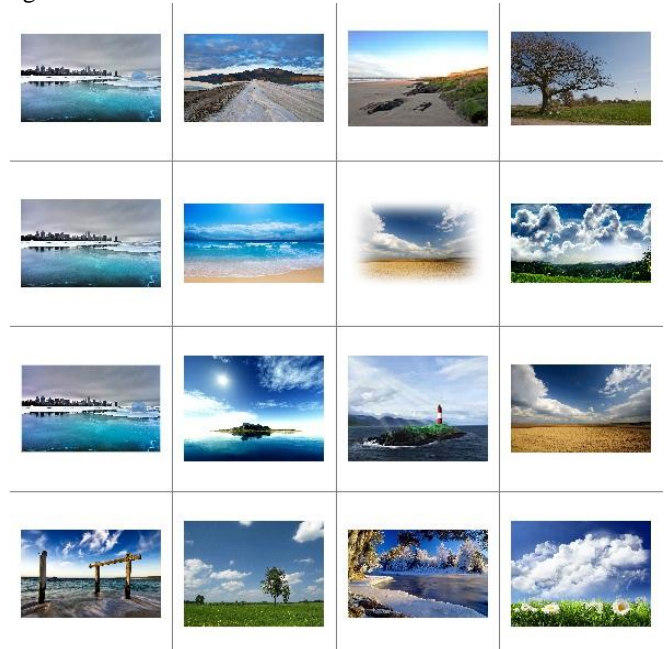


Figure.4. Search results performed on Flickr 160 database.

### 5. Conclusions

Our main objective is to retrieve images using a image. In this three main aspects were taken into consideration: Retrieval process, time and some degree of noise. The quality of the image result set is dependent on the descriptive quality of the metadata, including the algorithms for object identification and low-level feature extraction. Manual definition of attribute and text metadata can be used to capture the semantic content of the image material and thus can support good, i.e. relevant image retrieval if the: Metadata attribute values are well chosen for future user information requirements, and the Indexer selects descriptive terminology that the future user will use, and Associated text is actually descriptive of the semantic content of the image material.

### Acknowledgements

I would like to have a thankful to the various refer comments.

### References

[1] 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.

[2] D. Comaniciu, and P. Meer, "Robust analysis of feature spaces: color image segmentation", IEEE Conference on Computer Vision and Pattern Recognition, pp. 750-755, June 1997.

[3] N. Dalal, and B. Triggs, "Histograms of oriented gradients for human detection", IEEE Conference on Computer Vision and Pattern Recognition, pp. 886–893, July 2005.

[4] A.K. Jain, J.E. Lee, R. Jin, and N. Gregg, "Content based image retrieval: an application to tattoo images", IEEE International Conference on Image Processing, pp. 2745–2748, November 2009.

[5] 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.

[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] W. J. Krzanowski, Recent Advances in Descriptive Multivariate Analysis, Chapter 2, Oxford science publications, 1995.

[8] J.A. Catalan, and J.S. Jin, "Dimension reduction of texture features for image retrieval using hybrid associative neural networks", IEEE International Conference on Multimedia and Expo, Vol.2, pp. 1211-1214, 2000.

[9] H. J. Zhang, and D. Zhong, "A Scheme for visual feature-based image indexing", Proc. of SPIE conf. on Storage and Retrieval for Image and Video Databases III, pp. 36-46, San Jose, Feb. 1995.

[10] Y. Rui, T.S.Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS", Proceedings of International Conference on Image Processing, Vol.2, pp. 815 -818, 1997.

[11] W. Y. Ma, and B. S. Manjunath, "Edge flow: a framework of boundary detection and image segmentation", IEEE Int. Conf. on Computer Vision and Pattern Recognition, pp. 744-749, Puerto Rico, June 1997.

[12] W. Y. Ma, and B. S. Manjunath, "Netra: A toolbox for navigating large image databases", Multimedia Systems, Vol.7, No.3, pp.:184-198, 1999.

[13] C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein, and J. Malik, "Blobworld: A system for region-based image indexing and retrieval", In D. P. Huijsmans and A. W. M. Smeulders, ed. Visual Information and Information System, Proceedings of the Third International Conference VISUAL'99, Amsterdam, The Netherlands, June 1999, Lecture Notes in Computer Science 1614. Springer, 1999.

[14] H. Voorhees and T. Poggio. "Computing texture boundaries from images", Nature, 333:364-367, 1988.

[15] M. J. Swain, and D. H. Ballard, "Color indexing", International Journal of Computer Vision, Vol. 7, No. 1, pp.11-32, 1991.

#### Author Biographies



Mr. Eshwar Erva received his Bachelors Degree in Technology in Computer science and Engineering from Samskruthi college Of Engineering and Technology, JNTU, Hyderabad and Pursuing Masters in Technology in Computer science and Engineering from Aurora's Technological And Research Institute, JNTU, Hyderabad.

E-mail: erva\_eshwar@yahoo.com



Mr. K. Chandra Shekar working as Associate Professor in the Department of Computer Science and Engineering in Aurora's Technological and Research Institute with a teaching experience of 9 years and worked as Project Trainee. He had received his Masters Degree in Software Engineering. His areas of interest include Data mining and Information Security and his Specialization in Associative Classification Mining in Non-Binary Data.

E-mail: chandhra2k7@gmail.com