International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 7 (2015) pp. 18195-18210

http://www.ripublication.com

Annotation Based Algorithm For Content Based Image Retrieval of Digital Images

***Pathur Nisha S, **Ilango Krishnamurthi**
**Department of Computer science and Engineering, Nehru Institute of Technology, Coimbatore, TN, India.*
*** Dept. Of Computer Science and Engineering, Sri Krishna College of Engg & Technology, Coimbatore. TN, India*

Abstract

Computer aided manufacturing environment needs CAD images for production planning and manufacturing of new products. Those images can be retrieved from the database instead of designing every time. To reduce the searching time and to increase the accuracy (99%) this enhanced model is designed for retrieving images with automatic annotation. This study considers all primitive features of an image viz., colour, shape, texture and logical features and abstract attributes. The proposed framework includes automatic annotation and similar image retrieval based on relevance factor to narrow down the search for modifications to be performed on newly developed design of CAD images for Computer aided engineering environment. This improved model enhances the speed of retrieval with best possible accuracy in the range up to 0.96 relevance factor, when compared to the present algorithms. This integrated module has a capability to be integrated with any of the design, drawing and development software module.

Key words: Geometrical Image Retrieval, Color, Shape, Texture, Automatic Annotation, Relevance factor.

Introduction

Content based image retrieval models are used in many areas of science, technology and business to perform search for a specific image with its title/ semantics associated for the use large size of data base of images for various applications of engineering. Searching and locating an image with accuracy inside a larger image database is considered tedious and time consuming, even many a times lacks in accuracy. Hence retrieval of images with accuracy inside large database of stored images is very difficult; the only solution for easy and efficient image retrieval is content-based image retrieval (CBIR). This paper describes integrated method to use CBIR is used for image identification, comparison and retrieval simultaneously with different techniques like Intensity-based (color and texture) retrieval and Geometry-based (shape) retrieval. The detailed literature study carried out on the area have yielded the major challenges and scope for research is the potential for data mining and knowledge extraction to integrate manufacturing, product characteristics, and the engineering design processes requires retraction algorithms in huge data bases and yet to be explored and addressed, J. A. Harding et.al, in the field of computer aided engineering (CAE) indicates there is large scope for the retrieval and integration. Though image retrieval is desired for all the domains of CAE, the need for establishment of semantic relationship of images and the requirements is needed to be addressed in the field of engineering. The data on many leading CAE technologies is in textual form, but the proposed system uses the data is in the form of image, text and combination of both. This paper proposes to improve retrieval accuracy based on the relevance factor with automatic annotation considering low level features along with and abstract attributes of an image. The text data retrieval and processing semantics of the images which can support for a series of complex operations in the field of processing images and videos, such as recognizing shapes and objects. It is also important to have annotation metadata for images for semantic retrieval based on image content.

Related Works

The work done by Andrew Kusiak, Matthew Smith (2) gave insight in image based data-mining for product and manufacturing system, one of the major challenge is for development of an efficient algorithm to process image, geometry, audio, and video data. The difficulty in designing geometrical image retrieval system is still unresolved and needs a suitable retrieval algorithm. Danushka Bollegala et.al. (7) and Roland Kwitt.et.al (15) Edward Kim. Et.al (8) have demonstrated the applicability of the context of colour texture retrieval on four texture image databases and compared retrieval performance to collection approaches for image retrieval but did not address the issues related to combine the shape based retrieval. Sathyabama. et.al (16) have developed computer-aided Image Retrieval method based on images of plant leaf shapes has lead to the challenge of combining the color texture and shape, which is discussed and solved with our enhanced model. Jianping Fan.et.al (13) Jean Francois aujol, Tony F.Chan (12) Subramanian Appavu (17) developed algorithms to classify the image into geometrical information and texture information but the relevance factor was not incorporated and addressed in this enhanced algorithm. Cosmin Grigorescu (6) Alberto J. Alvares.et.al (1) Balafar (3). Brian Kulis, (4) have reported the usage of data mining to design and manufacturing integration as a possibility of usage of STEP based environment which will be the improvement area of this model. Koen Deschacht (14) Tao Jiang (19) Yang Mingqiang (20) had given a deep insight of various shape feature extraction techniques using the edge detection algorithms and methods. Mehwish Rehman.et.al., used feature extraction and relevance feedback. Research papers by Aman Chadha.et.al and Manimala Singha et.al., discussed the ways to improve CBIR by discriminating power of color indexing techniques using Canberra distance as similarity measure.

Based on the literature findings, the idea is to develop an enhanced algorithm capable of providing the best possible retrieval (speed and accuracy) of image content, based on shape, colour, texture and semantic information with relevance value. The new algorithm should be accommodative of geometrical, CAD, CAM and CAE data sources. Since the complex geometric image and its contents, retrieval and its results required in manufacturing environment to be integrated with computer integrated systems, Machine learning algorithms, computational intelligence tools, which depend more on the large data bases for their applications. The large data collection, storage and retrieval system for process control and execution purposes is resolved and addressed by inter disciplinary approach of this research work. Though there are some customized tools developed so far in development laboratories for the above requirement but an exclusive and integrated solution is yet to be arrived. This capability is the most valuable asset of Decision Support/ learning Systems, which is supported by the proposed algorithm for digital/geometrical images. This algorithm supports in decision making process of similar parts identification in computer aided manufacturing environment by reducing manual/ judgmental decisions on high technology applications. Hence the manual intervention of the closest part selection out of the large data base is reduced and eliminated if integrated with the Enterprise Resources planning tool platform.

Proposed Work

The mining methods use low level functions of color, texture and shape (e.g., shape recognition and pattern extraction) during the process of image analysis and information extraction, segmentation algorithms are used to partition the image into regions related to the relevant areas according to the application criteria. Hence automatic annotation tool is inevitable and developed in this research for image retrieval, to simplify the process of searching the image database. In this paper, we introduce a novel approach for feature selection in high dimensional data using a new content based image retrieval algorithm. The dependent attributes of a given image are annotated, identified, compared and enlisted for the geometrical images in XML schema. Dependency between attributes are calculated by first grouping them and then by calculating the distance of value of class attribute using Euclidean distance method. The procedure is repeated for all possible combinations of attributes, and the dependencies between the whole attribute set in a dataset are found. The proposed method shows better results in terms of number of selected features, classification accuracy, and running time than most of the existing algorithms. The newly developed image retrieval algorithm uses automatic annotation based on array similarity distance matrix, color layout distance calculation, visual descriptor distance calculation, edge detection and texture segmentation methods, also used as a retrieval tool of keyword or image and both. The idea is to list similar images of the large image database based on the relevance. Proposed system encompasses image annotation and image retrieval with the help of automatic annotation and retrieval tool for geometrical digital images.

Frame Work of Enhanced Image Retrieval Algorithm

The images of a dataset are selected based on related application (ie.criteria) and stored in the data base and trained by the manual or automatic annotation frame work incorporated in this Enhanced image Retrieval Algorithm (EIRA). Preprocessing phase of EIRA extracts the features of all images and stores in indexed order as Meta data of the object as trained set of data. The search process based on query or by image used to extract the objects and match up with the query data for similar pattern of color, shape and texture and lists based on relevance factor. The highest relevance factor image is listed first and the next follows in the long list of images under consideration up to the lowest factor of the entire image data base. The relevance may either be color, shape and texture and combination of any these low level features. The results are then interpreted to generate intelligence to the DSS that can be applied in problem understanding, decision making and other activities. The annotation results can be mapped to a XML schema, once stored using XML (to represent object properties and relationships), the images could be retrieved through descriptions to express selection criteria that should be satisfied.

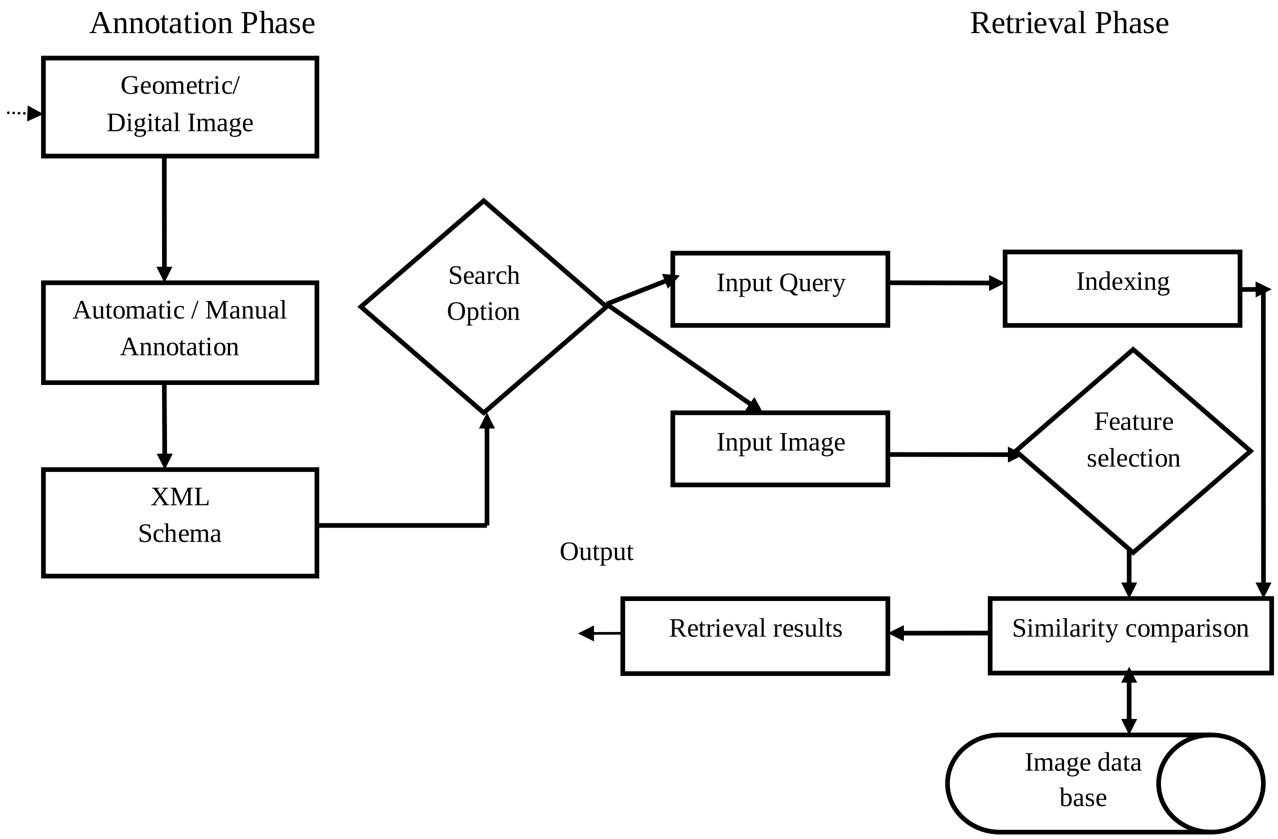

Figure 1: Frame Work of Proposed Image Retrieval Methodology

Figure 1 explains the proposed work which consists of two modules. Automatic or Manual Annotation module and Similar Image retrieval module based on the levels of Primitive features such as color, texture, shape of image elements, Derived (sometimes known as logical) features, involving some degree of logical inference about the identity of the objects depicted in the image.

A. Enhanced Image Retrieval Algorithm (EIRA)

1. Collection of Image database, containing 50, 100, 200, 500… 10000 etc Geometrical /Digital images.
2. Manual, Automatic Annotation and storage of XML schema of the image. Content-based image annotation starts from a set of candidate annotations obtained by an existing image annotation algorithm. Then, the query-biased transition probability matrix is constructed for the query image using both the content feature of the query image and the corpus information. Feature Extraction is carried out by using color, texture or by using shapes. Color feature extraction is done by Color Layout Descriptor (CLD), Texture extraction using co-occurrence matrices. The geometrical images are stored as metadata with their corresponding features such as color, texture, shape.
3. Similarity Searching

3.1 Query based

Done by image indexing technique,

3.2 Image based (Color, Shape, and Texture)

Done by using color histogram for color. The texture based similarity retrieval is done using Wavelet transform procedure with segmentation methods. The process of Shape feature retrieval is done with moment similar image retrieval can be done using the similarity measure procedures.

B. Automatic Annotation

Step 1: Input the Image.
Step 2: Segmentation is done based on the region.
Step 3: Extract the feature based on color, shape, Texture.
Step 4: Annotate the image based on the pattern.
Step 5: Storage of semantic information as XML schema.

Perform Manual annotation if needed. (Who, Where, What.etc) and store semantic information as XML schema.

The input image is taken for clustering using Euclidean distance method by taking ordinary distance between two points that one would measure with a ruler and is given by the Pythagorean formula. By using this formula as distance Euclidean space becomes metric space for comparison between object slicing of the images. The Euclidean distance between points p and q is the length of the line segment connecting them (pq)

In general for an n-dimensional space, the distance is

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \ldots + (p_i - q_i)^2 + \ldots + (p_n - q_n)^2} \quad (1)$$

Here we are using 2 dimensional geometrical objects for classification and clustering. The two dimensional Euclidean plane,

if p= (p_1, p_2) and q= (q_1, q_2) then the distance is given by

$$d(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2} \quad (2)$$

The clustered objects are stored as binary values inside the database, and based on the values of the input image the fast mapping operation done by visual descriptor distance and color layout distance for annotation. This calculator uses the distance between two objects for fast mapping using the following formulae for distance function for symmetric and must obey the triangle inequality Distance in k is:

$$d[k+1](O_1, O_2)^2 = d[k](O_1, O_2)^2 - (x_1[k] - x_2[k])^2 \quad (3)$$

Color Layout Descriptor (CLD) is designed to capture the spatial distribution of color in an image. This descriptor effectively represents the spatial distribution of color of visual signals in a very compact form. This compactness allows visual signal matching functionality with high retrieval efficiency at very small computational costs. It provides image-to-image matching as well as ultra high-speed sequence-to-sequence matching, which requires so many repetitions of similarity calculations.

The distance between the two descriptors can be computed as:

$$D = \sqrt{\sum_i w_{yi}(DY_i - DY_i')^2} + \sqrt{\sum_i w_{bi}(DCb_i - DCb_i')^2} + \sqrt{\sum_i w_{ri}(DCr_i - DCr_i')^2} \quad (4)$$

A suffix tree for a string *S* of length *n* can be built in Θ(*n*) time, if the letters come from an alphabet of integers in a polynomial range (in particular, this is true for constant-sized alphabets). For larger alphabets, the running time is dominated by first sorting the letters to bring them into a range of size *O*(*n*); in general, this takes *O*(*n*log*n*) time. Assume that a suffix tree has been built for the string *S* of length *n*, or that a generalized suffix tree has been built for the set of strings

$$D = \{S_1, S_2, \ldots S_k\} \text{ of total length } n = |n_1| + |n_2| + \ldots + |n_k| \quad (5)$$

Search for strings

- Check if a string *P* of length *m* is a substring in *O*(*m*) time.
- Find the first occurrence of the patterns $P_1, \ldots, P_q$ of total length *m* as substrings in *O*(*m*) time.
- Find all *z* occurrences of the patterns $P_1, \ldots, P_q$ of total length *m* as substrings in *O*(*m* + *z*) time..
- Search for a regular expression *P* in time expected sub linear in *n*.
- Find for each suffix of a pattern *P*, the length of the longest match between a prefix of $P[i \ldots m]$ and a substring in *D* in Θ(*m*) time . This is termed the **matching statistics** for *P*.

The fast mapped output is stored as XML schema along with the properties of the image, which is used for searching during the similarity search.

At end of the automatic annotation process a detailed description of the image-Meta data of the image is stored in XML schema. For our extraction, we adopt MPEG-7. It is expected that this standard would be used in searching and retrieving for all types of media objects. If we have images stored with metadata, it would be easier to do semantic retrieval. The stored files contain a reference to the location of the corresponding image file with different schemas. We propose to combine XML schema integration techniques and image retrieval techniques using low-level features with automatic annotations. It also uses XML as the language of choice for the textual representation of content description, as XML Schema has been the base for the DDL (Description Definition Language) used for the syntactic definition of MPEG-7 Description Tools and for allowing extensibility of Description Tools.

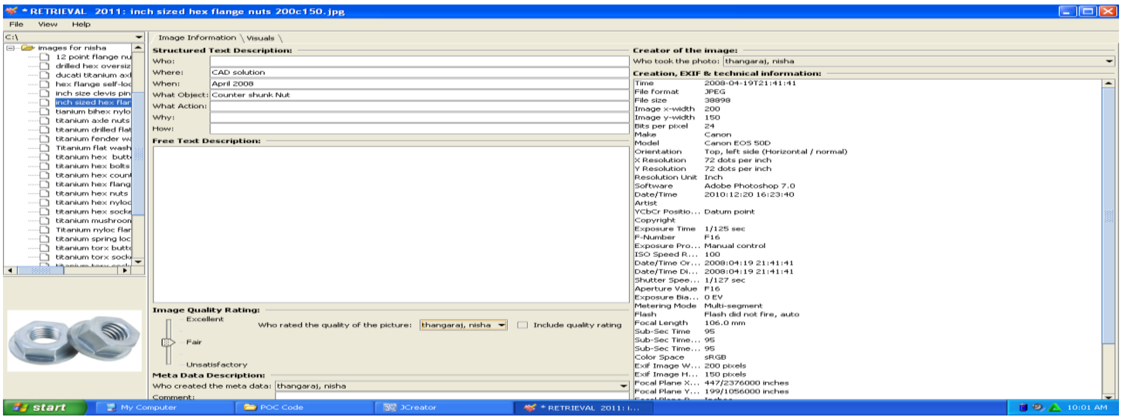

Figure 2: Automatic Annotation of Image

C. Similarity Searching

Several methods for retrieving images on the basis of color, texture and shape by using any one of the features have been described in the literature; most have given same idea of similarity search algorithms and its usages. The color histogram technique is used for similarity search of color feature. At search time, the user can either specify the desired proportion of each color (75% olive green and 25% red, for example), or submit an example image from which a color histogram is calculated. Either way, the matching process then retrieves those images whose color histograms match those of the query most closely.

Texture segmentation is used in this paradigm for retrieving similar image based on texture feature of the image. To know the texture details a set of metrics are calculated in image processing to quantify the perceived texture of the image. It gives the information about spatial arrangement of color or intensities in an image or selected region of an image. It is used to help in segmentation or classification of images.

Shape matching of three-dimensional objects is a more challenging task – particularly where only a single 2-D view of the object in question is available. One approach is to generate a series of alternative 2-D views of each database object, each of which is matched with the query image. Related research issues in this area include defining 3-D shape similarity measures, and providing a means for users to formulate 3-D shape queries.

1) Query Based search:

A query is given to retrieve the similar geometrical image from the image dataset based on the description that is already stored as XML schema during the annotation. This is done by creating an index for the data set that is to be stored. Based on the index the tool retrieves the similar images for the given query.

2) Image Based Search:

A geometrical image is given as a query to retrieve the similar image from the image dataset based on the description that is already stored as XML schema during the annotation. The Color feature of the image, the color histogram serves as an effective representation of the color content of an image if the color pattern is unique compared with the rest of the data set. The color histogram is easy to compute and effective in characterizing both the global and local distribution of colors in an image. The color histogram, i.e., the distribution of the number of pixels for each quantized bin, can be defined for each component. Clustering methods used to determine the K best colors in a given space for a given set of images, option is to use the bins that have the largest pixel numbers since a small number of histogram bins capture the majority of pixels of an image To take the spatial information of pixels into consideration, thus very different images can have similar color distributions. To increase discrimination power, a simple approach is to divide an image into sub-areas and calculate a histogram for each of those sub-areas and the division can be as simple as a rectangular partition, or as complex as a region or even object segmentation.

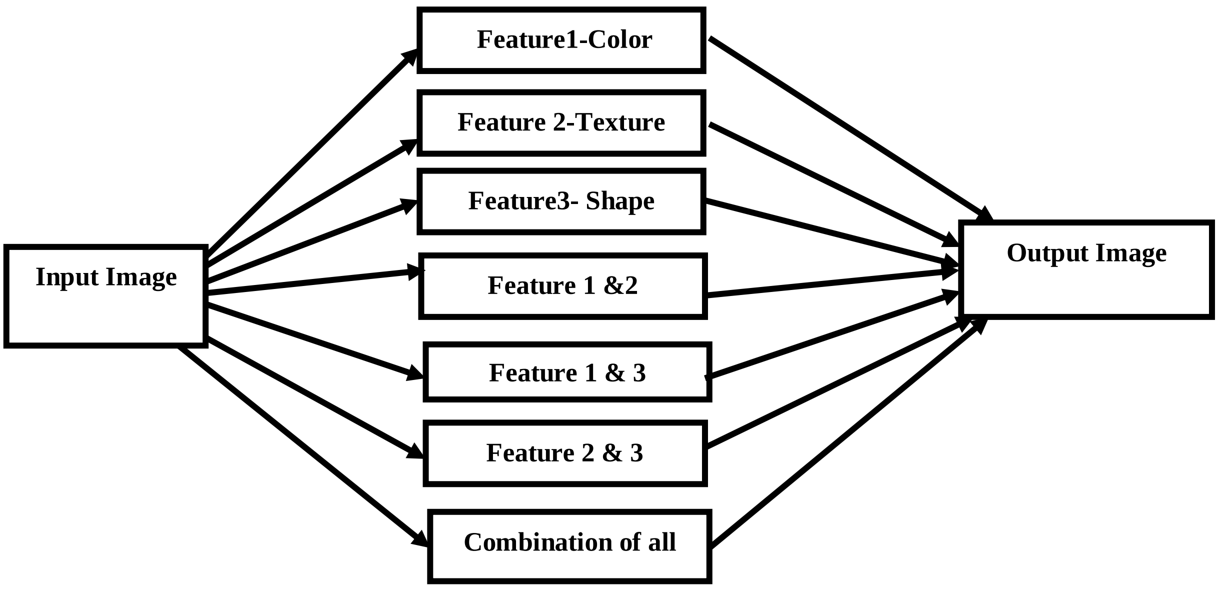

Figure 3: Feature selection Combinations for similarity comparison

The texture based similarity retrieval is done using Edge detection procedure with co-occurrence matrices and segmentation methods. Segmentation based on image texture, region based and boundary based. Region based attempts to group or cluster pixels based on texture properties together. Boundary based attempts to group or cluster pixels based on edges between pixels that come from different texture properties. Segmenting the objects is based on edge detection algorithm and in order to avoid an extreme segmentation, the description is mapped to a XML. Edge detection is used to determine the number of edge pixels in a specified region helps determine a characteristic of texture complexity. The direction of edges can also be applied as a characteristic of texture and can be useful in determining patterns in the texture. These directions can be represented as an average or in histogram.

Consider a region with N pixels the gradient –based edge detection is applied to this region by producing two outputs for each pixel p: the gradient magnitude Mag (p) and the gradient direction Dir (p). The edginess per unit area can be defined by

Fedgeness=|{p|Mag(p) > T}|/N for some threshold T. (6)

The Co-occurrence Matrices formula is used to capture properties of a texture. Numeric features computed from co-occurrence matrices can be used to represent and compare textures. Standard features from a normalized co-occurrence matrix is

$$\text{Energy} = \sum_{i\,j} N2dd[i,j] \tag{7}$$

$$\text{Entropy} = -\sum_{i}\sum_{j} Ndd(i,j)\log 2Nd(i,j) \tag{8}$$

$$\text{Contrast} = \sum_{i}\sum_{j} (i-j)2Ndd(i,j) \tag{9}$$

$$\text{Homogeneity} = \sum_{i}\sum_{j} \frac{Nd(i,j)}{1+|i-j|} \tag{10}$$

The process of Shape based similar image retrieval can be done in two ways, by using the relevance. Classical shape representation uses a set of moment invariants. If the object R is represented as a binary image, then the central moments of order p+q for the shape of object R is defined as:

$$\mu_{p,q} = \sum_{(x,y)\in R} (x-x_c)^p (y-y_c)^q \tag{11}$$

Where $x_c = m_{10}/m_{00}$, $y_c = m_{01}/m_{00}$

and x_c, y_c are called the center of object. Hence the central moments of order up to 3 can be computed as

$$\left.\begin{array}{l}
\mu_{00} = m_{00} \\
\mu_{01} = 0 \\
\mu_{10} = 0 \\
\mu_{11} = m_{11} - y_c m_{10} \\
\mu_{20} = m_{20} - x_c m_{10} \\
\mu_{02} = m_{02} - y_c m_{01} \\
\mu_{30} = m_{30} - 3x_c m_{20} + 2m_{10}\, x_c^2 \\
\mu_{21} = m_{21} - 2x_c m_{11} - y_c m_{20} + 2x_c^2 m_{01} \\
\mu_{12} = m_{12} - 2y_c m_{11} - x_c m_{02} + 2y_c^2 m_{10} \\
\mu_{03} = m_{03} - 3y_c m_{02} + 2y_c^2 m_{01}
\end{array}\right\} \quad (12)$$

This central moment can be normalized to be scale invariant

$$\eta_{p,q} = \frac{\mu_{p,q}}{\mu_{o,o}^{\gamma}}, \quad \gamma = \frac{p+q+2}{2} \quad (13)$$

Based on these moments, a set of moment invariants to translation, rotation, and scale can be derived. Information about image orientation can be derived by first using the second order central moments to construct a covariance matrix.

$$\left.\begin{array}{l}
\mu'_{20} = \mu_{20}/\mu_{00} = M_{20}/M_{00} - x\square^2 \\
\mu'_{02} = \mu_{02}/\mu_{00} = M_{02}/M_{00} - y\square^2 \\
\mu'_{11} = \mu_{11}/\mu_{00} = M_{11}/M_{00} - x\square y\square
\end{array}\right\} \quad (14)$$

The input image is taken for clustering using Euclidean distance method by taking ordinary distance between two points that one would measure with a ruler and is given by the Pythagorean formula. By using this formula as distance Euclidean space becomes metric space for comparison between object slicing of the images. The Euclidean distance between point's p and q is the length of the line segment connecting them (pq).

In general for an n-dimensional space, the distance is

$$d(p,q) = \sqrt{}(p_1-q_1)^2+(p_2-q_2)^2+\ldots+(p_i-q_i)^2+\ldots+(p_n-q_n)^2 \quad . \quad (15)$$

Here we are using 2 dimensional geometrical objects for classification and clustering. The two dimensional Euclidean plane, if p= (p_1, p_2) and q= (q_1, q_2) then the distance is given by

$$d\,(p, q) = \sqrt{}(p_1-q_1)^2+(p_2-q_2)^2 \quad (16)$$

The clustered objects are stored as binary values inside the database, and based on the values of the input image the fast mapping operation done by visual descriptor distance and color layout distance for annotation. This calculator uses the distance between two objects for fast mapping using the following formulae for distance function for symmetric and must obey the triangle inequality

Distance in k is: $d[k+1](O_1,O_2)^2 = d[k](O_1,O_2)^2 - (x_1[k]-x_2[k])^2$ (17)

The key property of this evolution is the order of the substitution. The substitution is done according to a relevance measure K given by

$$K(S_1,S_2) = \beta(S_1, S_2)l(S_1,)l(S_2) / l(S_1)+l(S_2) \tag{18}$$

where $\beta(S_1, S_2)$ is the turn angle at the common vertex of segments S_1, S_2 and $l(\alpha\ )$ is the length of α, $\alpha = S_1$ or S_2 normalized with respect to the total length of a polygonal curve. The similarity search is performed by query based and image based methods.

Query Based Searching

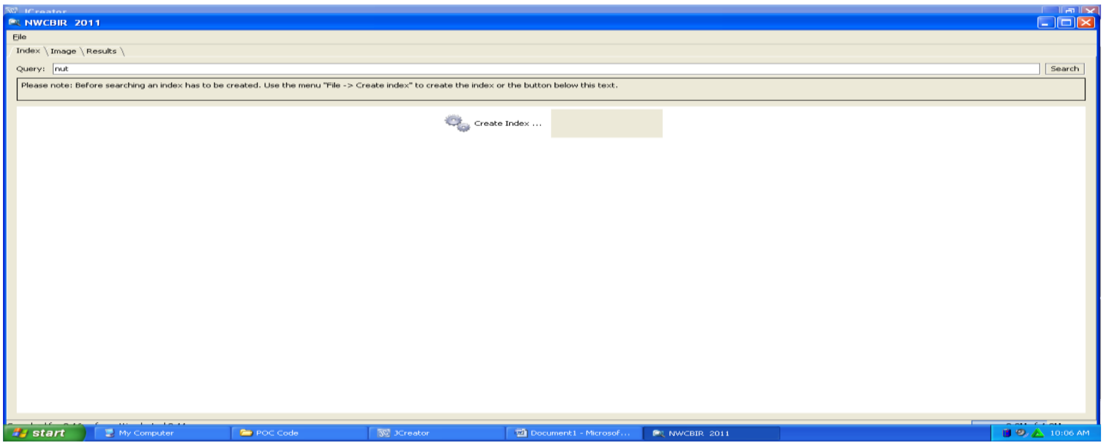

Figure 4: Image Retrieval Based On The Input Query

Figure 5: Results For Automatic Annotation Based Search (Query Search)

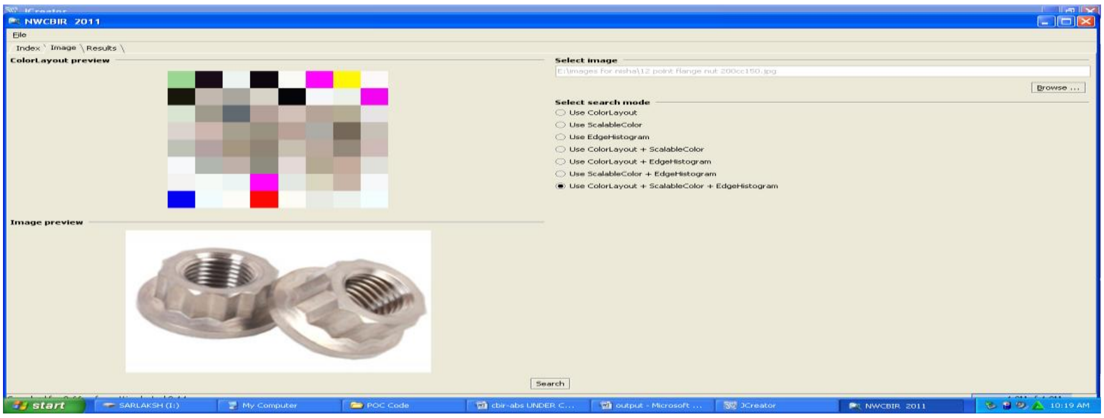

Figure 6: Image Retrieval Based on The Query Image

Figure 7: Results For Automatic Annotation Based Search (Semantic Search)

The above integrated approach is proposed for digital image databases combining text annotations. This approach combines a conceptual and visual content description together with an improved relevance scheme and performs better than the prevailing data mining techniques for the image data retrieval (19). Our retrieval framework is primarily intended for databases where the geometrical / digital images do not have text annotations and are automatically captured and stored as Meta data of the image for the retrieval (21). The integrated scheme for relevance factor using automatically annotated databases improves to bring better results by performing both queries (20). The different schemes in the prevalent data mining methods and their features used in this System and its performance comparison are as listed in Table 1. The IIRA method is also checked for the effectiveness of the retrieval and proved faster than the methods discussed in the reference texts. The time required to perform the retrieval is computed and plotted in Graph.1 for the different numbers and heterogenic nature of digital images in the database by steps.

Discussion on The Results

Our retrieval framework is primarily intended for databases where the geometrical / digital images do not have text annotations and are automatically captured and stored as Meta data of the image for the retrieval (21). The integrated scheme for relevance factor using automatically annotated databases improves to bring better results by performing both queries (20). The different schemes in the prevalent data mining methods and their features used in this System and its performance comparison are as listed in Table 1. The IIRA method is also checked for the effectiveness of the retrieval and proved faster than the methods discussed in the reference texts. The time required to perform the retrieval is computed and plotted in Graph.1 for the different numbers and heterogenic nature of digital images in the database by steps.

In our experiments, we evaluated how effectively our integrated tool could be used to collect annotation data and to retrieve the similar geometrical images from the image set. We were interested in several factors, including how easy is our tool to use, how satisfied are the users with their annotated/ retrieval result, and how does our tool more efficiently collect data of the image. The digital images annotated automatically for the database to get stored as XML schema, the data semantics is now serves as the key word for the search. The accuracy level and retrieval speed were compared from the result screen shots displayed here. Here also the relevance value plays the vital role in telling the user, the nearest of the databse images despite listing all the possible items in the database.

Irina Mocanu (11) has compared the cycle time for retrieval of the images by the famous methods like Centroid Radii Turning angle (CRTA) method, Distance Histogram (DH) method, Centroid-Radii model (CR), Fourier Descriptors (FD), Turning Angle (TA) methods and the comparison has yielded the method proposed (EIRA) gave the best even when operated for the most complex (geometrical) of the shapes being compared for less than 2000 images in a database.

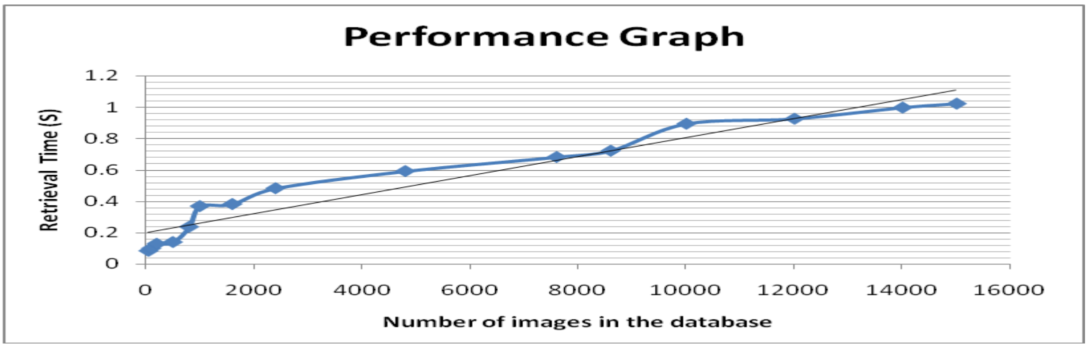

Figure 8: Performance of The Algorithm Retrieval Time

Considering the average precision and recall obtained for the test databases, the EIRA method outperforms the rest of the analyzed methods. The low retrieval performance of turning angle method indicates it as an inaccurate shape representation. There are few situations in which distance histograms method outperforms precision and recall of the centroid radii and turning angle method (these situations appear for shapes with a lot of small edges).

Table 1: Comparision of Performance of Various Methods

Method	**Retrieval Time (Secs)**	**Comparison Time (Secs)**	**Total Time (Secs)**
CR	0.02	0.6	0.8
FD	0.3	0.01	0.31
TA	0.06	0.05	0.11
DH	0.9	0.8	1.7
CRTA	1	1	2
EIRA	**0.02**	**0.03**	**0.05**

Conclusions

This work executes the connection between an annotation model and retrieval method for any given geometrical image family. The major contribution to the technology is

1. The use of clustering allows automatic annotating the details of image and stores image data into the database as XML Schema of the image, which is used to the image in better accuracy with reasonably rapid way.
2. A new search criterion (automatically annotated information) for an image is established and easy to retrieve any image of complex geometrical shapes.
3. Our results show that the tool greatly simplifies and fastens the process of annotation for large geometrical image.
4. This prototype is able to annotate and retrieve geometrical data using methods, compare, classify and group datasets, analyze similarity and perform data mining in different levels.

References

[1] Alberto J. Alvares & Joao Carlos Espindola Ferreira,” A system for the design and manufacture of feature-based parts through the Internet” Int J Adv Manuf Technol (2008) 35:646–664.

[2] Aman Chadha, Sushmit malik, Ravdeep johar, “Comparative Study and Optimization of Feature-Extraction Techniques for Content based Image Retrieval”, International Journal of Computer Applications pp0975 – 8887 Vol 52– No.20, 2012.

[3] Andrew Kusiak, Matthew Smith,” Data mining in design of products and production systems”, Elsevier, Annual Reviews in Control 31 (2007) 147–156.

[4] Balafar M A ,. Ramli A. R and. Mashohor S, “Edge-preserving Clustering Algorithms and Their Application for MRI Image Segmentation”, Proceedings of the International MultiConference of Engineers and Computer Scientists 2010 Vol-I.

[5] Brian Kulis, Member, IEEE, Prateek Jain, Student Member, IEEE, and Kristen Grauman, Member, IEEE,” Fast Similarity Search for Learned Metrics”, IEEE Transactions on Pattern Analysis And Machine Intelligence, Vol. 31, No. 12, December 2009.

[6] Carol J Romanowski, Rakesh Nagi, Moises Sudit,” Data mining in an engineering design environment: OR applications from graph matching”, Computers & Operations Research, December 2004.

[7] Cosmin Grigorescu, Student Member, IEEE, and Nicolai Petkov,” Distance Sets for Shape Filters and Shape Recognition”, IEEE Transactions On Image Processing, Vol. 12, No. 10, October 2003.

[8] Danushka Bollegala, Yutaka Matsuo, and Mitsuru Ishizuka, Member, IEEE,” A Web Search Engine-Based Approach to Measure Semantic Similarity between Words”, IEEE transactions on Knowledge and Data Engineering, vol. 23, no. 7, July 2011.

[9] Edward Kim, Student Member, IEEE, Xiaolei Huang, Member, IEEE, and Gang Tan, Markup SVG An Online Content-Aware Image Abstraction and Annotation Tool,” IEEE Transactions on Multimedia, vol. 13, no. 5, October 2011.

[10] Fuhui Long, Hongjiang Zhang and David Dagan Feng,” Fundamentals of Content-Based Image Retrieval” – Book.

[11] Harding J. A , Shahbaz, Srinivas M, Andrew Kusiak,” Data Mining in Manufacturing: A review”. Journal of Manufacturing Science and Engineering November2006, Vol. 128 / 969.

[12] Irina MOCANU,” From Content-Based Image Retrieval by Shape to Image Annotation,Advances in Electrical and Computer Engineering Volume 10, Number 4, 2010.

[13] Jean- Francois aujol, tony F.Chan Combining Geometrical and textured information to perform image classification. Journal of visual communication and Image representation. Vol- 17(2006) p1004-1023.

[14] Jianping Fan, Yi Shen, Chunlei Yang, and Ning Zhou,” Structured Max-Margin Learning for Inter-Related Classifier Training and Multilevel Image Annotation”, IEEE Transactions on Image Processing, vol. 20, no. 3, March 2011.

[15] Koen Deschacht ,Marie-Francine Moens, “Text Analysis for Automatic Image Annotation”, and Proceedings of the 19th Belgian-Dutch Conference on Artificial Intelligence pages 343-344, Utrecht, November 2007.

[16] Mangijao Singh.S and Hemachandran.K.,”Content Based Image Retrieval using Color Moment and Gabor Texture Feature”, International Journal of Computer Science Issues, Vol. 9, Issue 5, No 1, 2012

[17] Manimala Singha and Hemachandran.K., "Content Based Image Retrieval using Color and Texture Signal & Image Processing", An International Journal (SIPIJ) Vol.3, 2012.

[18] Mehwish Rehman, Muhammad Iqbal, Muhammad Sharif and Mudassar Raza,"Content Based Image Retrieval: Survey", World Applied Sciences Journal pp 404-412, Vol-19 (3), 2012.

[19] Roland Kwitt, Peter Meerwald, and Andreas Uhl," Efficient Texture Image Retrieval Using Copulas in a Bayesian Framework", IEEE Transactions on Image processing, vol. 20, no. 7, July 2011.

[20] Sathya Bama B," Content Based Leaf Image Retrieval (CBLIR) Using Shape, Color and Texture Features," Indian Journal of Computer Science and Engineering.

[21] Subramanian Appavu Alias Balamurugan, Ramasamy Rajaram," Effective and Efficient Feature Selection for Large-scale Data Using Bayes Theorem", February 2009.

[22] Swati V. Sakhare and Vrushali G. Nasre," Design of Feature Extraction in Content Based Image Retrieval (CBIR) using Color and Texture", International Journal of Computer Science & Informatics, Volume-I, Issue-II, 2011.

[23] Tao Jiang and Ah-Hwee Tan, Senior Member, IEEE," Learning Image-Text Associations", IEEE Transaction on Knowledge and Data Engineering, June 2008.

[24] Yang Mingqiang, Kpalma Kidiyo1 and Ronsin Joseph," A Survey of Shape Feature Extraction Techniques", Open Access database**,** www.i-techonline.com.