

Visual Analytics for Semantic Based Image Retrieval (SBIR): Semantic Tool

R.Aravindhan

*Department of Computer Science and Engineering
Sri Eshwar College of Engineering, Coimbatore, Tamilnadu, India*

Dr.R.Shanmugalakshmi

*Department of Computer Science and Engineering
Government College of Technology, Coimbatore, Tamilnadu, India*

Abstract- Semantic-based image retrieval (SBIR) is a retrieval technique based on the intuition of user. However, implementing a meticulous strategy without incurring a semantic gap (disparity between the retrieved images and the user desired images) is a challenging task. The semantic gap is not trivial, when it comes to sensitive medical diagnostic images. Also, with rapid digitization of medical diagnostic images, statistics reveal that data from US health care system alone reached 150 Exabyte's and so with the presumption that this diagnostic medical data will shoot up, memory requirements and computational expenses should be reduced substantially. Furthermore, as only a region of image will be most relevant (ROI-Region of Interest) while retrieving diagnostic medical data, the feature dimensionality is reduced and thereby the speed of the retrieval process is increased. As, most SBIR algorithms do not intuitively explain why the retrieved images are considered relevant to the user query, there is a chance that the user may miss out relevant images which may be outliers. Semantic technologies like ontology offers auspicious approach to image retrieval as it tries to map the low level image features to high level ontology concepts. So a tool called Visual Analytics Tool for Semantic Retrieval (VATSR) is proposed which allows user to visually interact and refine the query and/or search results. Moreover, in this paper, a hierarchical classifier is proposed which classifies at different levels with different set of features, which ensures that a relevant image missed out in one classifier may be retrieved in another classifier. Along with this, a boosting algorithm is use to improve the precision of the search results.

Keywords – Image Retrieval, Ontology, Semantic Tool, Semantic gap

I. INTRODUCTION

It is truly and worthily claimed that an image is valuable than 1000 words. [1] However it can never and ever be presumed that an individual will predict these 1000 words in the same knowledge based on the image context and their perspective. It frequently differs. Assume, the images shown in the below Figure 1, it makes us so sure that an image may possess multiple interpretations. This scenario explains that the users retrieve images through different descriptions. Here arises a communication gap that no one can prevail over that easily. That is fundamental between the information/image seeker/user and the image retrieval system. Query formulation process must emerge with various forms in the image retrieval system with utmost necessity as the user may possess the fluctuation in the needs and the core knowledge about the image collection. Basically, image retrieval queries are classified into: Query 1: Attribute-based – To retrieve images, these queries make use of the context and the structural metadata values. Query 2: Textual – To retrieve the desired images they use the term-based specification that matches with the textual image descriptors. Query 3: Visual – Here, for image retrieval, visual characteristics (color, texture, shape) / sample image to be compared with visual descriptors are specified.[2] Various types of image retrieval show in Figure 2.

Text-Based Image Retrieval (TBIR):

What is a Text-Based Image Retrieval (TBIR)? For instance, usually the verbal description/specification of the obligated image will be given by the image requester.



Figure.1 Optical illusion

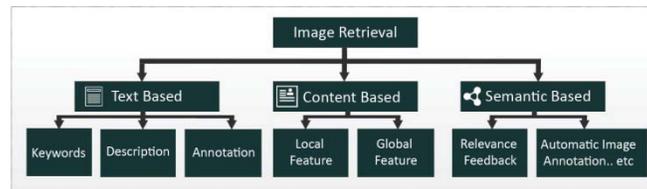


Figure.2 Types of Image Retrieval System

Working Process: Prior step included in the annotation of images are: i) Employs keywords describing the image, ii) Caption of an image iii) surrounded/embedded text within it iv) the containing pages complete text and v) exact filename. Now, the text based Database Management systems leave space to retrieve images[3].

Textual descriptors: Semantic reasoning is done in 2 ways i) to be done manually using tools ii) automatically relates semantic terms with relative descriptions. Images are tagged here using metadata (data about data). It consists of: i) Image Creation Date ii) Image Creator iii) Format and iv) Simple descriptions of the digital object.

Images in databases: Final step is indexing images for ease and efficient retrieval. Collections are derived into databases which lead to search queries and fetches images accordingly.

Disadvantages[4]:

- i) Time consuming – As this is done manually.
- ii) High probability of disagreement.
- iii) Lack of captions/ accompanying text.



Figure.3 A Typical Text Based Image Retrieval System

Content Based Image Retrieval (CBIR):

What is Content Based Image Retrieval (CBIR)? It is all about ‘image content’ i.e. its pixel content. Their objective of developing automated routines shows out the functionalities: i) Analyzing digital image/image stream and ii) Identifying the object, actions and events. CBIR identifies color and texture distributions and primitive shapes. This leads to image signatures and a space for providing rapid access to elements in the image collection. As illustrated in Figure 4, a typical CBIR System.[5] This can be semantically connected. The queries here are: image attributes and Visual examples (Sketch, texture palette, etc). [6-8] Distinguished below are common image needs:

Color Feature Based Retrieval: Color features and similarity measures are extracted for image retrieval.

Color Histogram

This is all about a color (calculates Red, Green and Blue) feature representation. Two different representations are: 3-D histogram (one) and 1-D histogram (three separate). The mathematical representations calculate the similarities in queried images and stored images based on two methods.

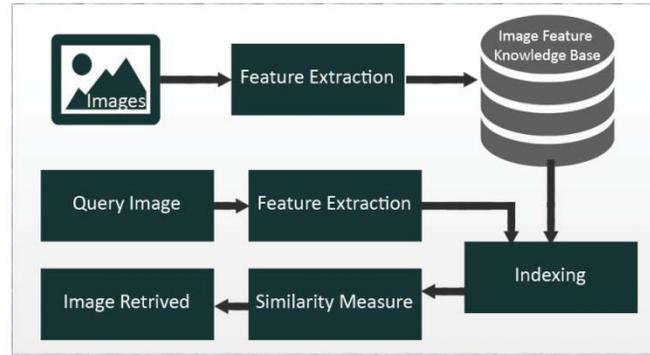


Figure.4A Typical Content Based Image Retrieval System

Histogram Intersection

Calculations are done between normalized color histograms of database image and query image.

Distance Method

Euclidean distance between the vectors is used to measure the similarity.

Shape based image retrieval: This further feature will be extracted only when the image colors are similar. Two methods mentioned are: Boundary based – uses the outer boundary and Region based – uses the entire region.

Semantic-Based Image Retrieval (SBIR)

The major limitation of CBIR, which is measured by the difference between the understanding and search intent of a human user and the computed (machine) understanding of the image content [6] is semantic gap. Even though human interpretation of the image retrieval is not an important part of CBIR workflow, it is still a forth coming research area [1], [7], [8].

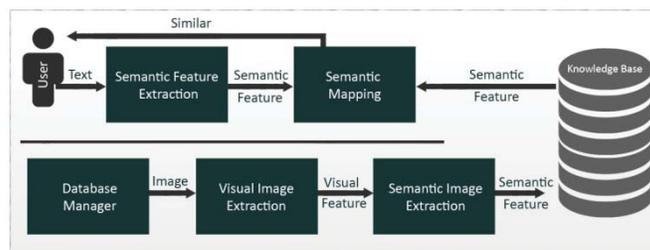


Figure.5 A Typical Semantic Based Image Retrieval System

To overcome the semantic gap, several medical CBIR systems [9]–[13] used the method of retrieving the m most similar images (as decided by the underlying algorithm) and the final arbiter is the user who figures out which of the images are pertinent to the query. Results are often presented in form of a order list to the user that is based on the computed similarity to the query. Any how the ordered lists only contain some inevitable images from the dataset most of the evitable images are discarded from the dataset while creating an ordered list. North less, optimizing precision is mainly focused by retrieval algorithms, the proportion of retrieved images which are relevant, low priority may still be given to some relevant images. Commonly used retrieval algorithms find in the range of 5 to 20 images, and so it is possible to discuss a relevant image in a large data base that could fall below the threshold. The ordered list never provides any explanation about the features which are used to order the images [6].

II. TECHNIQUES FOR FILLING SEMANTIC GAP

When the retrieval system is of low-level features, then the drawbacks are: Mismatches with human perception, unpredictable and insufficient. There arises the semantic gap [14]. There are several techniques innovated to reduce the ‘semantic gap’. It includes:

Object ontology: Here the semantics of the images are used to define them. The region attribute is used in each level to define the image into the intermediate level of the descriptor [15]. Texture Features and Color Quantization are the key features here. Color naming is the core node for the quantizing the colors.

Machine Learning: Though this is originated to handle the composite semantics it possesses difficulty in implementation. Based on the low-level features (query concepts) they are segmented into:

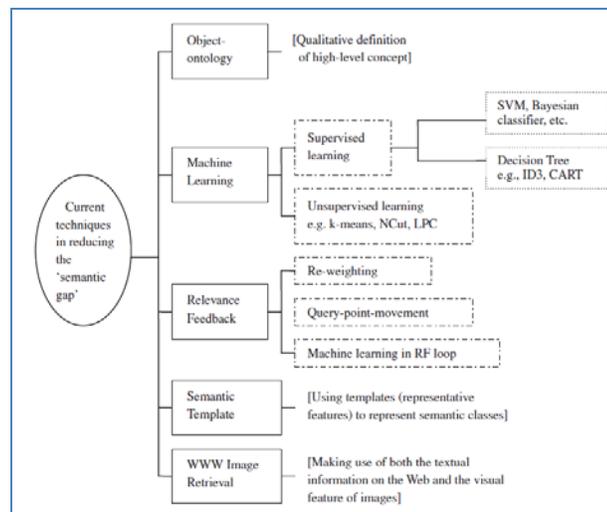


Figure 6. Techniques for Filling Semantic Gap

i) Supervised machine intelligence – the recognition of semantic category label is achieved through the binary classifier and numbers of images are collected for entry measures. ii) Bayesian Classifier: an automated classification of indoor and outdoor types from the image database is proposed in this important technique. iii) Neural network: users are enhanced to choose 11 concepts and then the introduced inputs establish the link between high-level semantics and low level features. iv) Decision tree: relevancy and irrelevancy of the image input is divided into two classes and thus the semantic features are achieved. v) Unsupervised learning – the input data will be clubbed up into groups and each group will be named. Thus enhances in obtaining the results within the definite group.

Relevance Feedback: This is online and user interaction/communication technique. The input query should be in text, picture or image. [16]It categorizes the images into "relevant" and "irrelevant". However, the user is the one who judges the relevancy of the retrieval and starts the cycle process till his/her target search is achieved.

Semantic Template: Through the number of images saved in the database their general characteristics are calculated and this is all about semantic template. Semantic visual template (SVT) is used to link the characteristics features (low-level and high-level). Semantic visual template (SVT) image also links the characteristics features (low-level and high-level) but for video retrieval. Users are forced to observe the image characteristics in depth.

WWW Image Retrieval: This varies from the previous techniques as they are slightly web dependant to reduce the semantic gap. [36] URL of the image file and HTML document for image title information are obtained here. Anyhow, the result may end up in vain as this is time-consuming. Researchers are working on the same for rectification.

This paper is systematized as follows. Techniques for filling semantic gap are analyzed in Section 2. Section 3 talks about literature surveys. The proposed Framework for Visual Analytics Approach for Semantic Based Bio Medical Image Retrieval is discussed in Section 4. The analysis of proposed experimental solution is constituted in Section 5 presented. The conclusion and future study are stated in Section 6.

III. LITERATURE REVIEW

Even after the groundwork of content similarity, a lot and lots of researched were conducted in Image Retrieval (IR). All the researched techniques were initiated focusing on the results retrieved in image search.

Mohsen SardariZarchi et al [17] this proposal comes out with an idea of considering user-interested concepts and semantic-based model is used in resourceful image retrieval. To identify the user-interested regions a algorithm under the name 'interactive image segmentation is carried out. These methods of recognizing through regions, becomes valid only when a neural network classifier is used. The elimination of surplus and inappropriate features becomes possible through the usage of ant colony optimization algorithm features (low-level) of the images. The semantics of the visual content is represented due to the proposal of a novel technique. This first index involves itself in: i) restructuring visual word vectors into an Ontology model, ii) calculating the distance between the features (Visual word and Concept) i.e. called out as concept range. The properties, instances, or concepts that are defined in Ontology are recognized and extracted in the second index using textual description

Camille Kurtz et al[19] suggests two different strategies for image retrieval that too based on the semantic features. Mutually image both ontological term and image-based relations are taken into account in the new term dissimilarity measure. This helps in the annotations similarity evaluation which leads to the retrieval of similar images. Multi-scale Riesz wavelets are used to automatically “soft” predict the ontological terms (describes image content). These two strategies work out themselves as a connecting factor in retrieving similar and accurate images with the support of image annotations. This pretends to be the probably good solution for the semantic gap problem.

GowriAllampalli-Nagaraj et al[20] this proposal mainly targets on bridging the high-level semantics and low-level features. Thus, they privilege themselves in the usage of leverage sophisticated image-processing techniques and the image content information is extracted automatically in to the MPEG-7 format. These are then associated into the domain ontologies that still exist and are developed by experts.

Ying Liua et al[21], the technical achievements that have been taken place in the high-level semantic-based image retrieval are provided as a comprehensive survey. This is his suggestion. The semantic gap is narrowed down through identifying the foremost major categories of the state-of-the-art techniques.

SohailSarwar al [22] suggests an image retrieval framework based on ontology from a collection of natural scene images by imparting human cognition in the retrieval process. To assess the authenticity of this approach, the authors decided that the framework would be tested over other types of images data like cityscapes etc.

Deb et al. [23] and Veltkamp et al. [24] discuss state of the art segmentation, indexing and retrieval for a number of CBIR systems. According to them, the hindrance to the access of multimedia data from databases is the chasm between low level image features and high level semantic expressions. An important feature revealed by these surveys is the dependency of almost all existing approaches on the usage of low level image feature, categorization and retrieval. The key to all content-based image categorization and retrieval systems is to understand the image.

N.Magesh et al.[25] recommends a retrieval method based on the keyword based annotation structure, and ontology guided reasoning. The authors’ inferred that to present more meaningful answers to queries than just to hit-lists, one can apply Ontology-enriched knowledge base of image Metadata.

EeroHyvonen et al.[26] suggests a circumstance when a user is faced with an image repository whose content is complicated and semantically unknown to some extent. In such cases, it is shown how ontologies can help the user to formulate the required information, the query, and the answers. The authors deduced that by using less detailed ontologies and annotations, a balance between annotation work and quality of information retrieval can be done, if needed.

Oleg Starostenko et al.[27] proposes a user may submit textual queries, which are converted to image characteristics providing in this way searching, indexing, interpretation, and retrieval. Both an image and sketch may be used in case of visual query. Approaches for image interpretation and retrieval are applied to color filtering, shape indexing and semantic. The authors came to a conclusion the following could be considered as possible extension of the proposed approaches - tolerance to deformation, to complex images with significant number of objects, to presence of noise, and feasibility of indexing.

Till date, the surviving study material have always proved them in bridging the semantic gap between the search results and user requirements and through involving them in the improvisation of the content based images retrieval results.

IV. PROPOSED FRAMEWORK

Semantic Based Image Retrieval (SBIR): This is basically Ontology based system architecture. It portrays the tasks of the various modules/ components that are available in the system; and also their interaction with one another. The system architecture of SBIR is detailed in Figure 7.

It consists of the following modules:

Tier I: User Query Module

Tier II: SBIR Engine Module

Tier III: Boosting Module

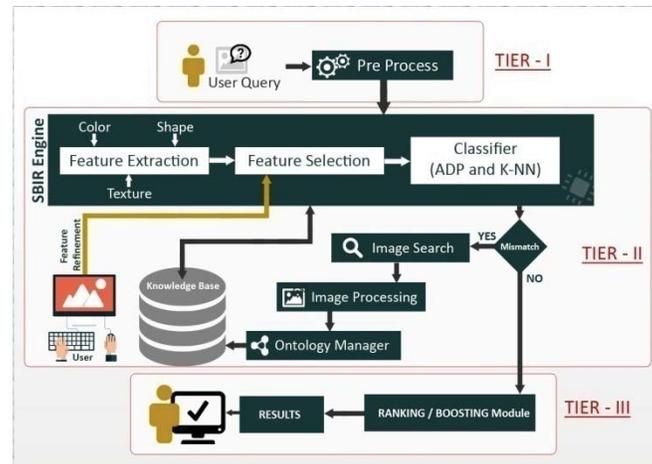


Figure.7 Proposed System Tier Architecture

Our proposed architecture is a 3 Tier process.

Previously, when the user types in an input, the pre-process will be undertaken by Tier I. Once the pre-process is complete; the whole query enters into Tier2. Here, it will be initiated with the SBIR Engine. There are three different processes it is made out of. First process of the SBIR engine is Feature Extraction stage. Speaking, the Shape, Color and Texture of the input are grasped. The very next stage will be the Feature Selection. And from there it moves into the Classifier stage. If at all, there comes any mismatching criteria in this process, then there is another prolonged process in Tier II itself. In this, firstly, the Image Search process will be done and then the Image Processing and it extends itself into the Ontology Manager. The altogether abbreviated and combined data are stored into the Knowledge Database from where it is again streamlined into the SBIR Engine until and unless the data retrieved matches the user query. If the user query matches all the searching processes, the whole process is taken-up by Tier III from Tier II. Then in Tier III, the Boosting will be done which proceeds in showing the best achieved results to the users.

Assume, if the user finds the results little annoying; and does not matches his/her search query, users can attain the search by facilitating them the feature refinement option, in spite of, filling in the semantic gap.

The schema of inducing ontologies into the content-based image processing retrieval is not an innovation indeed. In the previous decade, researchers have/had derived many schematics on these ontologies on content-based image processing retrieval [23-28]. Here, we differ in the hybrid approach.

Tier I: User Query Module: There are two distinct ways to specify the users input, once they begin to search for. i) Text Input: This is the very first method in which probably the text is given as an input into the SBIR. Of course, as each and every one are very well familiar with the same in the universally prompted search engines such as, e.g. Google, Yahoo, and AltaVista etc. The process hangs over in a way that once when the user enters into the text queries (MRI Scan, X-Ray CT scan etc.), it is carried over into the Text based Query module. ii) Image Input: Considerably, image will be given as an input into the SBIR engine. The only criteria are the instance of uploading the literal image that they are searching for.

The core functionality while we pretty well converse about this module is that it is entrusted on the regulation of images with sophistication to the user query. Then the role of the SBIR Engine Module is to exceed the matching precision that too as a resultant image set(s). This portrays the ontology concepts' sum that is as of user's query. Taking into consideration the matching value(s), the results are retrieved in the descending order. Prioritizing the accuracy, the top ten most apt images (equivalent to user query) will be retrieved. Besides, the minimal relevancy will be retrieved only on user's request that too in decreasing order.

Image pre-processing involves in performing some number of operations that can be done to an image. Literally, the operations done to the images are in the minor level without touching the main core of the image. In elaborating its functionality, it squeezes the undesired distortions and important image features are attracted for the future processes. However, the image core is not at all collapsed. It enhances and never increases. It helps in the pixel valuation of the images. That is, if the neighboring pixel consists of the same brightness of the real image that will not be disturbed on any cause. Whereas, if there is another pixel that does not matches its valuation then an average value will be in-taken.

In order to calculate the brightness of the new pixel, the size of the neighborhood pixel is to be calculated. That ended up in four image pre-processing method categorization: i. the brightness of the pixel is transformed. ii. Geometric transformations are done. iii. Local neighborhood of the processed pixel has to be undergone in the pre-processing methods. iv. The whole knowledge of the image is acquired for image restoration.

Tier II: SBIR Engine Module: SBIR Engine plays a vital role in taking the input from the user using the web interface;

Image Search: Matching module browses through primary search engines for retrieving the relevant images (i.e. grabbed from the web and refined in the previous step) to the ontology knowledge base.

Image Processing: The images resulted from the search engine may or may not relate the user query; hence, the content of each and every image are believed to be verified. This module is responsible to verify the images about its compliance with the input query [29]. The objects in each and every image are detected using the shape based feature extraction and these objects are moved on to the 1) Color based Feature Extraction technique the prolongs with the MTH algorithm proposed by Guang-Hai Liu et al in [30] to calculate the pixel color and color value of the objects and 2) texture classification technique proposed by Mohsen Zand et al in [31] to recognize the texture / pattern (if any) in the objects. In the very next step, the low level features extracted in the previous step are converted into high level ontology concepts; when the resulted class (es) matches the user search query, the image will be incorporated in the resultant set or else it might be discarded. Thus endorsing with results; that only the related images are stick with it and the non-relevant images will be discarded in this step.

Ontology Manager: Ontology Manager is answerable to insert in the new relevant images features and concepts. [32,33] The concept experimented into our proposal is Bio-Medical domain Ontology.

Tier III: Boosting Module:

Boosting algorithm comes under machine learning and this initiates in testing the effectiveness of the classifier executed in Tier II [34]. The boosting algorithm has been trained for this process execution. Now, though the process has been trained the effectiveness of the classifier is the main crisis for the Classification [35]. Technically conveying, simple classifiers never meet the requirements as they generally shoot-out the results which is slightly convincing when compared with random result. The main assignment of the boosting algorithm is to convert the weaker classifiers into a stronger one through learning.

In order to alleviate the problems found in applying traditional AdaBoost algorithm to multi-class classification problems, unsupervised clustering algorithms are adopted as a multi-class classifier to AdaBoost algorithm in this paper [36]. The proposed multi-class classifier-based Adaboost algorithm (MC) gives a final strong classifier that requires less training time and more stable results by adopting clustering algorithms in each training step (T). When the number of training data is N and the number of classes is M, each of the data has a distribution weight w_{ij} , $1 \leq i \leq N$, $1 \leq j \leq M$ to each of M classes as defined in the original AdaBoost algorithm.

The calculation of classification errors using pseudo-loss error is the same as the case in Adaboost.M2 and the error of each class, k on each training step, t, defined as follows :

$$\epsilon_k^t = \frac{1}{2} \sum_i^N w_{i,k}^t |1 - h_k(x_i)| + w_{i,k}^t \sum_{j \neq k}^J w_{i,j}^t h_j(x_i), j \in J - k \quad (1)$$

where $h_k(x_i)$ is the classifier output for data i.

In Adaboost.M, the pseudo-loss error includes binary classifier outputs that have target values for other classes as in Eq.(1). However, since the proposed multi-class classifier algorithm produces the output, 1, for the closest class and output 0 to all the other classes. Eq. (1) can be changed to Eq. (2):

$$\epsilon_k = \frac{1}{2} \sum_{i=1}^N w_{i,k}^t |1 - h_k(x_i)| \quad (2)$$

By using Eq. (2), the reliability of a weak classifier for class, k, can be also obtained as follows:

$$\beta_k^t = \frac{\epsilon_k^t}{1 - \epsilon_k^t} \quad (3)$$

Therefore, the weight update equation for class, k, can be found as follows:

$$w_k^{t+1} = w_k^t (\beta_k^t)^Q, Q = \frac{1}{2} (1 - h_k^t(x_i)) \quad (4)$$

After finishing all the predetermined weak classifier training steps, ($t = T$), and assigning all weak classifier weights, the strong classifier, $H(\cdot)$, for data (x) can be found as follows [37]:

$$H(x) = \arg \max_k \sum_{t=1}^T \alpha_k^t h_k^t(x) \alpha_k^t = \log \frac{1}{\beta_k^t} \quad (5)$$

V. SIMULATION AND EXPERIMENTAL RESULTS

The Semantic Based Image Retrieval (SBIR) emerged out in developing the Ontology concepts. To accomplish this, it was literally pushed to use up to 2000 images and from which 15 of them are various Bio Medical Image Parts. Figure 8 technically explains the training dataset. In this particular section, we discuss and present about some of the test cases.

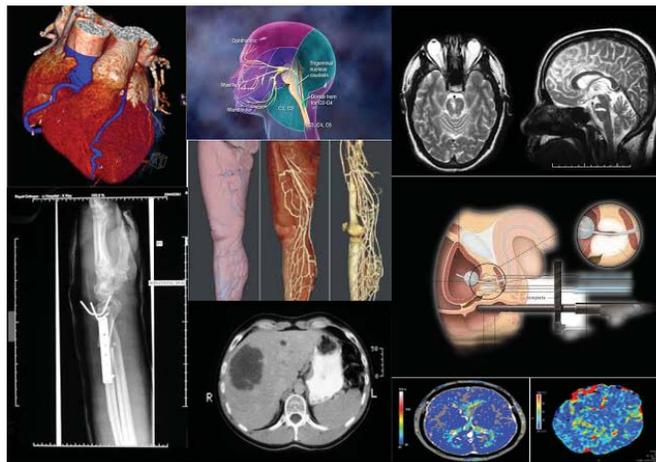


Figure.8 Partial Training Set

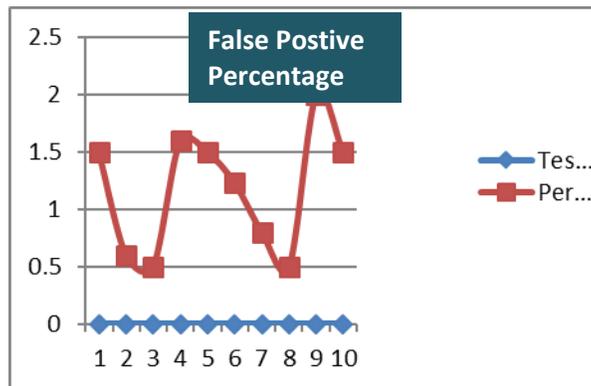


Figure.9 Percentage Improvement Vs Test Case

Figure.9: this illustration gives details about the percentage improved. This improvement in the proposed SBIR technique is illustrated in the fig only after carrying out several numbers of test cases. Considering the test case, it focuses mainly on the content of the images; that plays a vital role either.

Figure 10 shows the false positive percentage that has been done over number of test cases, and as per the proposed solution false positive percentage ranges from 0.60 to 2 percent in the test cases and this shows the result accuracy of the proposed solution.

SBIR Tool:

Our Proposed SBIR Tool is designed with 4 Main areas:

1. User Input Area
2. Visual Analyzer area
3. Refinement Option.
4. Log Status Window

In Input area, user can choose an image from the local drive. In Figure 11. Shows the Home screen of our SBIR Tool.

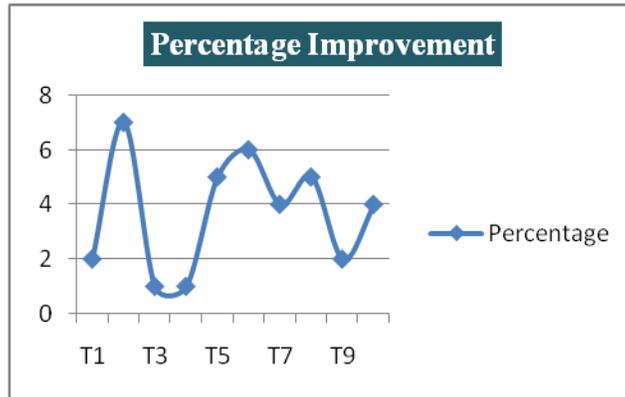


Figure.10 False Positive Percentage Vs Test Case

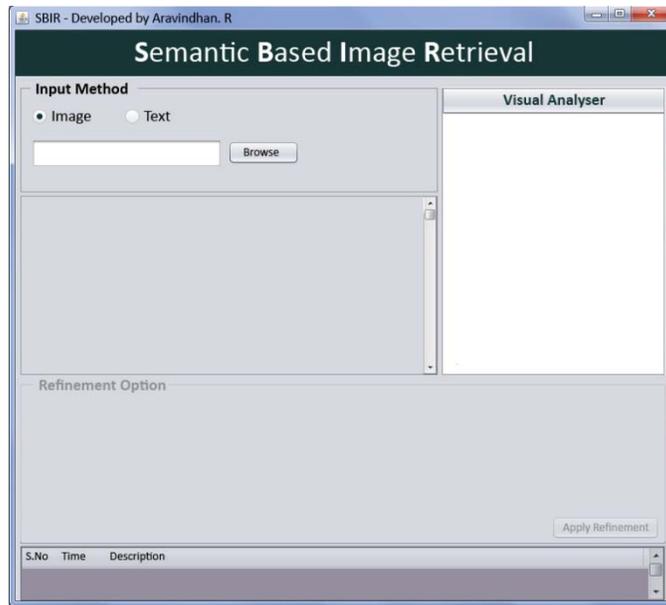


Figure 11: SBIR Tool Home Page

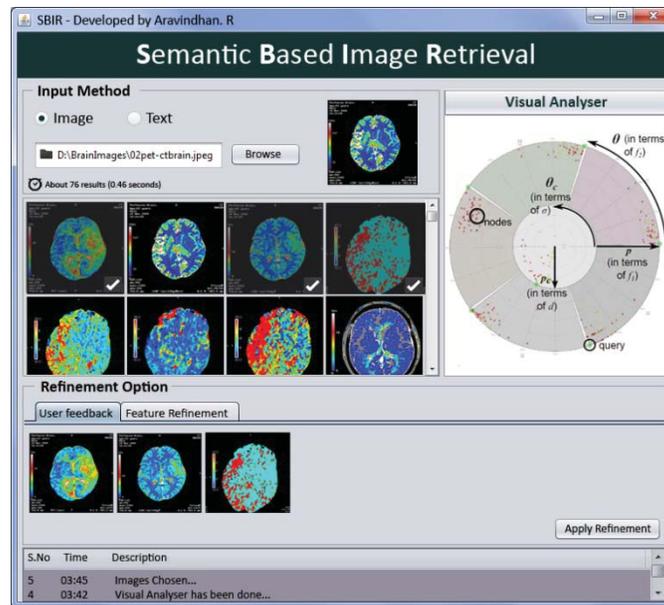


Figure.12 Test Case 1 and Refinement Option (User feedback)

Figure 12, Illustrates a test case that used PET-CT images. The given result is based on the ontology concepts. The images are obtained from the knowledge base. Figure 12, the users have used brain PET-CT image as an input; all queries are generated through the SBIR Engine and are executed based on the ontology knowledge. However, the knowledge base plays the lead role in grabbing the resultant images. Thus attained results will be automatically promoted to the Boosting Module. This is the execution criteria where the results are again ranks/boosts the results and output is shown out to the users

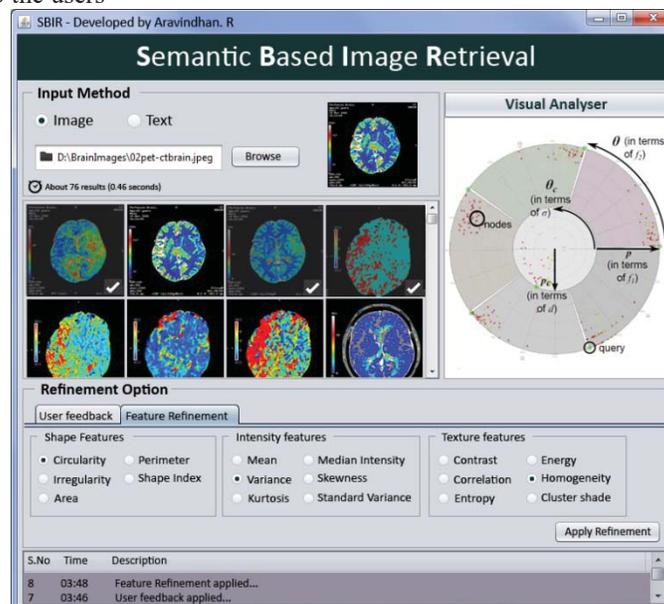


Figure.13 Test Case 1-Refinement Option (feature Refinement)

In Figure 12 and 13, the Test case1's Refinement Option shows the User Feedback tab and Feature Refinement tab. When the users are unable to retrieve the relevant result(s), then the users are facilitated with the feature selection option and they may go for it. After choosing the User Feedback tab, the users can select one or more images, from the results displayed. Then they can press Apply Refinement option for better results. In Feature Refinement tab,

this is another criterion for allowing the refinement process by the users in the SBIR Engine. Figure 12, elaborates the Test Case 1's Feature Refinement tab under the refinement option. The selections available are:

- i) Shape Features
- ii) Intensity Features
- iii) Texture Features

In figure 14, Test Case 2, the user have chosen Heart CT Scan image as an input; The SBIR processes the query using the ontology concepts. The displayed images will be taken from the knowledge base.

Visual Analytics Components:

As a rule, the results that are retrieved based on the CBIR (Content Based Image Retrieval) engine are analyzed to an extent. The VA components accessible in the workflow will primarily involve themselves in structuring the interactive visualization scheme. Here, we include the suggestions of the interfaces in the search criteria [38] and the visualization guidelines that are versatile [39]. They are: i) there are always abundance of dissimilarities between visualized information and the effectiveness of human interpretation and this semantic gap could compete only through complex data and too with multiple observations, ii) compound and complex information are the one that are solicited everywhere and there are notions to summarized such ones involved, iii) subordinate information comes under surplus presentation, iv) users will always be privileged to intermingle with the visualization information and refine them as well.

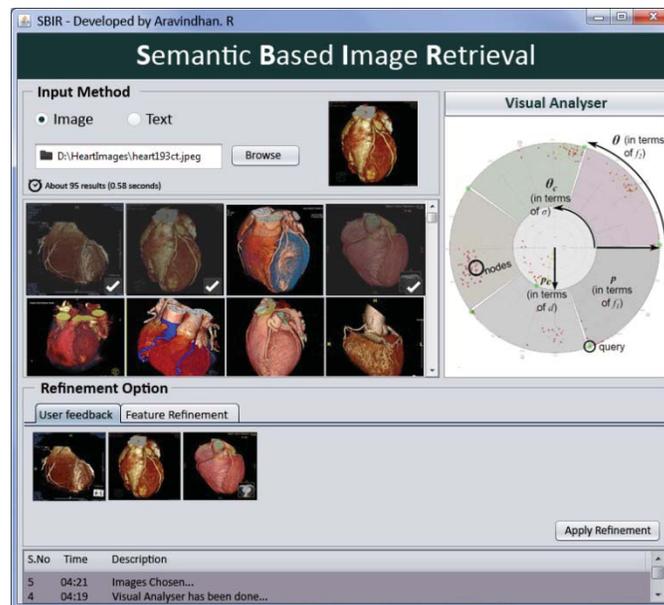


Figure.14 Test Case 2

Justification for Circular Visualization: These are several peaks that were really scrutinized, while we decided to fabricate into a circular layout. i) They are prioritized for the sectors placement in counterpart positions, ii) the logical consistency of the layout will not be affected though after deleting or adding sectors and hence proved that it enables smooth dynamic changes, iii) they always concentrate for better angular resolutions. For instance, always the nodes in each sector are visualized through high radial (p) value for the images that accomplish relevancy, iv) amidst the other sectors the circular layout outstands in positioning the central disk (the super center). Literally speaking, the super center acquires the functionality of both the radial and angular components. Radial component: communication done by the CBIR engine in visualizing the correspondence and Angular component: the image and the standard deviation of its features i.e., the compactness. The above is illustrated is obtained to compare the super center (features used to calculate image similarity) and outer sectors (subset of features used to calculate image similarity)

The three main assignments undertaken by the VAMIR are: i) the relationships between the retrieved images and the queries are visualized, ii) also visualizes the between the retrieved images, here, the chosen feature subset works as a foundation for all, iii) displays the mainframe similarities that occurs between the query and the retrieved images are segmented with reference to all the features. Figure2 illustrates the VAMIR interface;

VI CONCLUSION

This paper describes that the Visual Analytics Tool for Semantic Retrieval (VATSR) is proposed which allows user to visually interact and refine the query and/or search results. Moreover, in this paper, a hierarchical classifier is proposed which classifies at different levels with different set of features, which ensures that a relevant image missed out in one classifier may be retrieved in another classifier. Along with this, a boosting algorithm is used to improve the precision of the search results. Future work is fabricated into three different ideas: The VATSR Tool will be converted into a web service. This web service provides many advantages especially its interoperability including i) various applications to talk to each other, ii) share data and iii) services among themselves. The current scenario elaborates that, the Big Data is shooting out in the industry. In this proposal, we have portrayed only with small Data set; excluding Big Data. However, we will very well implement our Knowledge Base into the Big Data in the future. As a final point we have combined here all the worked out and constructed biomedical domain based focused web crawler and even launched the Search engine for the end user(s).

REFERENCES

- [1] H.W. Hui, D. Mohamad, and N.A. Ismail, "Approaches, challenges and future direction of image retrieval," *Journal of Computing*, vol. 2, 2010, pp. 193–199
- [2] Alaa M. Riad, Hamdy.K. Elminir, SamehAbd-Elghany, "A Literature Review of Image Retrieval based on Semantic Concept," *International Journal of Computer Applications (0975 – 8887) Volume 40– No.11, February 2012*
- [3] Y. Rui, T.S. Huang, and S.-F. Chang, "Image retrieval: Current techniques, promising directions, and open issues," *Journal of visual communication and image representation*, vol. 10, 1999, pp. 39–62.
- [4] H. Müller, J. Kalpathy-Cramer, B. Caputo, T. Syeda-Mahmood, and F. Wang, "Overview of the first workshop on medical content-based retrieval for clinical decision support atMICCAI 2009," in *Medical Content-Based Retrieval for Clinical Decision Support (Lecture Notes Computer Science)*. Berlin, Germany: Springer, 2010, vol.5853, pp. 1–17
- [5] G. MailaivasanParthibanKarthikram, "Tag based image retrieval (TBIR) using automatic image annotation," *International Journal of Research in Engineering and Technology*, vol. 03, 2014
- [6] N. Chauhan and M. Goyani, "Enhanced Multistage Content Based Image Retrieval," *IJCSMC*, vol. 2, Issue. 5, 2013, pp. 175–179..
- [7] M. Singha and K. Hemachandran, "Content based imageretrieval using color and texture," *Signal & Image Processing: An International Journal (SIPIJ)*, vol. 3, 2012, pp. 39–57.
- [8] H. Zhang H. Dagan F. Long and D. Fen, Fundamentals of content based image retrieval," in D. Feng, W. Siu, H.Zhang (Eds.): "Multimedia Information Retrieval and Management. Technological Fundamentals and Applications,.
- [9] N. Chauhan and M. Goyani, "Enhanced Multistage Content Based Image Retrieval," *IJCSMC*, vol. 2, Issue. 5, 2013, pp. 175–179.
- [10] N. Chauhan and M. Goyani, "Enhanced Multistage Content Based Image Retrieval," *IJCSMC*, vol. 2, Issue. 5, 2013, pp. 175–179.
- [11] H. Zhang H. Dagan F. Long and D. Fen, Fundamentals of content based image retrieval," in D. Feng, W. Siu, H.Zhang (Eds.): "Multimedia Information Retrieval and Management. Technological Fundamentals and Applications, Springer-Verlag,
- [12] H.H. Wang, D. Mohamad, and N. Ismail, "Image Retrieval: Techniques, Challenge, and Trend," *International conference on Machine Vision, Imageprocessing and Pattern Analysis, Bangkok, Citeseer, 2009.*
- [13] R. Datta, D. Joshi, J. Li, and J.Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Computing Surveys (CSUR)*, vol. 40, 2008, p.5.
- [14] B. Luo, X. Wang, and X. Tang, "A world wide web based image search engine using text and image content features," in *IS&T/SPIE Electronic Imaging, Internet Imaging IV, 2003.*
- [15] X. Tang, K. Liu, J. Cui, F. Wen, and X. Wang, "Intentsearch: Capturing user intention for one-click internet image search," *T-PAMI*, vol. 34, no. 7, pp. 1342 – 1353, July 2012.
- [16] J. Cui, F. Wen, and X. Tang, "Intentsearch: interactive on-line image search reranking," in *ACM Multimedia*, 2008.
- [17] Mohsen SardariZarchi, Amirhasan Monadjemi, Kamal Jamshidi "A semantic model for general purpose content-based image retrieval systems" *Computers & Electrical Engineering*, Volume 40, Issue 7, October 2014, Pages 2062-2071
- [18] Stefan Poslad, Kraissak Kesorn "A Multi-Modal Incompleteness Ontology model (MMIO) to enhance information fusion for image retrieval", *Information Fusion*, Volume 20, November 2014, Pages 225- 241.
- [19] Camille Kurtz, Adrien Depeursinge, Sandy Napel, Christopher F. Beaulieu, Daniel L. Rubin "On combining image-based and ontological semantic dissimilarities for medical image retrieval applications" *Medical Image Analysis*, Volume 18, Issue 7, October 2014, Pages 1082-1100.
- [20] Gowri Allampalli-Nagaraj, Isabelle Bichindaritz "Automatic semantic indexing of medical images using a web ontology language for case based image retrieval" *Engineering Applications of Artificial Intelligence*, Volume 22, Issue 1, February 2009, Pages 18-25.
- [21] Ying Liu, Dengsheng Zhang, Guojun Lua, Wei-Ying Ma "A survey of content-based image retrieval with high-level semantics" *2006 Pattern Recognition Society.*
- [22] Sohail Sarwara, Zia UI Qayyuma, Saqib Majeed b. "Ontology Based Image Retrieval Framework using Qualitative Semantic Image Descriptions". *17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems - KES2013.*

- [23] Deb S and Zhang. "An Overview of Content-based Image Retrieval techniques". Proceedings of 18th International Conf on Advanced Information Networking and Application (AINA04), 1, issue 2004:59-64, 2004 .
- [24] Veltkamp and M Tanase. "Content-Based Image Retrieval Systems: A Survey". Tech. Rep. UU-CS 2000-34, Univ. Utrecht, Utrecht, The Netherlands, 2000.
- [25] N. Magesh and Dr.P.Thangaraj "Semantic Image Retrieval Based on Ontology and SPARQL Query". International Conference on Advanced Computer Technology (ICACT) 2011.
- [26] Eero Hyvönen¹,Avril Styrman¹,and Samppa Saarela², "Ontology-Based Image Retrieval".
- [27] Oleg Starostenko, and Vera Tyrsa "Shape Indexing and Semantic Image Retrieval Based on Ontological Descriptions".
- [28] Stefan Poslad, KraisakKesorn "A Multi-Modal Incompleteness Ontology model (MMIO) to enhance information fusion for image retrieval", Information Fusion, Volume 20, November 2014, Pages 225-241.
- [29] Mohsen SardariZarchi, AmirhasanMonadjemi, Kamal Jamshidi "A semantic model for general purpose content-based image retrieval systems" Computers & Electrical Engineering, Volume 40, Issue 7, October 2014, Pages 2062-2071.
- [30] Guang-Hai Liu, Lei Zhang, Ying-Kun Hou, Zuo-Yong Li, Jing-Yu Yang, Image retrieval based on multi-texton histogram, Pattern Recognition (2010), Volume 43, Pages 2380-2389.
- [31] Mohsen Zand, ShyamalaDoraisamy, Alfian Abdul Halin, Mas Rina Mustaffa "Texture classification and discrimination for region-based image retrieval" Journal of Visual Communication and Image Representation (2014), doi:10.1016/j.jvcir.2014.10.005.
- [32] Umar Manzoor, SamiaNefti, YacineRezgui, "Categorization of malicious behaviors using ontologybased cognitive agents", Data & Knowledge Engineering (2012), <http://dx.doi.org/10.1016/j.datak.2012.06.006>.
- [33] Awatef AI Azemi, SamiaNefti, Umar Manzoor, YacineRezgui Building a Bilingual Bio-Ontology Platform for Knowledge Discovery, International Journal of Innovative Computing, Information and Control, Volume 7, Number 12, Dec 2011, Pages 7067-7075.
- [34] Liu Y, Rong J et al. (2010). A boosting framework for visuality-preserving distance metric learning and its application to medical image retrieval pattern analysis and machine intelligence, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 32(1), 30–44.
- [35] Y. Freund and R. Schapire, "A short introduction to boosting," Journal of Japanese Society for Artificial Intelligence, 14(5):771-780, Sep., 1999.
- [36] L. Reyzin and R. Schapier, "How Boosting the Margin Can Also Boost Classifier Complexity," Intl. Conf. on Machine Learning, 2006.
- [37] Kaur, Harpreet, and Jyoti, Kiran, "Survey of Techniques of High Level Semantic Based Image Retrieval", International Journal of Research in Computer and Communication technology, IJRCCCT, ISSN22785841, Vol 2, Issue 1, January 2013.
- [38] M. Tory and T. Moller, "Human factors in visualization research," IEEE Trans. Vis. Comput. Graphics, vol. 10, no. 1, pp. 72–84, Jan./Feb. 2004.
- [39] M. L. Wilson, "Search user interface design," Synthesis Lectures Inf. Concepts, Retrieval, Services, vol. 3, no. 3, pp. 1–143, 2011.