ISSN 1746-7659, England, UK
Journal of Information and Computing Science
Vol. 11, No. 2, 2016, *pp*. 146-152

Content Based Image Retrieval for Space Weather Application

Esraa Zeki Mohammed
State Company for Internet Services, Ministry of Communication, Kirkuk, Iraq.
(*Received November 29, 2015, accepted March 13, 2016)*

Abstract. The research presented in this paper was aimed to develop an CBIR system to aid in automatically detect and retrieve solar flares images based on detected edges and color features. In general, the work consists of two phases: (1) enrollment phase, which consist of feature extraction based on Conventional Color Histograms (CCH) and edge detection for sunspots and active regions on the sun using Sobel filter, (2) retrieving phase, which use the similarity measurement to retrieve most similar images from the database for a given query sample. The conducted tests were carried on 250 images and give good precision and recall rates.

Keywords: color histogram, image retrieval, similarity measure, sobel filter, solar flares, sunsopts.

1. Introduction

Digital imagery is getting more popular in many perspectives. Private photo collections, medical imaging, solar phenomenon and geographical information systems are only some to mention. As the computation power is growing and the cost of storage media is decreasing, the size of digital image collections is increasing rapidly. There is a need for techniques that enables us to access and retrieve the huge amount of information embedded in these image collections. Now, simple manual browsing mechanism is getting cumbersome even with private collections. So, automatic image retrieval is inevitable [1].

The field of content-based image retrieval (CBIR) focuses on the analysis of image content and the development of tools to represent the visual content in a way that can be efficiently searched and compared. CBIR techniques extract visual features and perform the indexing, querying and retrieval based on such features. General CBIR approaches are intended for browsing large archives of general (arbitrary) content [2]. More recently, CBIR systems are starting to emerge in different applications, and the solar flares phenomenon studying domain is currently cited as one of the principle application domains for content-access technologies.

The term "space weather" refers to adverse conditions on the Sun that may affect space-borne or ground-based technological systems and can endanger human health or life .The importance of space weather is increasing day after day because of the way solar activities affect life on Earth and it will continue to increase as we rely more and more on different communication and power systems [3]. The established effects of space weather activities on our daily lives and can be summarized as follows:

Ground based systems: induced electric field and currents can disrupt the normal operation of high voltage power transmission grids. Pipeline, telecommunications cables, metallic oil and gas pipelines and railway signling [4].

Communication systems: wireless communication systems suffer from interruption of service like frequency jamming and dropped communications due to radio bursts caused by solar microwave emission. Solar activity can be produce X-ray that disrupt poin-to-point high frequency radio communications and radio noise that interferes with communication and radar systems [5,6].

Space based systems: adverse space weather conditions can be cause anomalies and system failures and increased drag on movement of satellites and spacecraft leadind to slow-down, change in orbits and shorter life-times of missions. Other radiation hazards include direct collision damage and/or electrical effects, caused by charged particles [7].

Astronomical observations have a long history, dating back to ancient times. However the technological progress in recent years opened new areas of research and exploration of previously unknown astronomical phenomena and dependencies among them. Since most of the astronomical data come in the form of images that lead to make image processing techniques natural means of data analysis. Some of these techniques methods have been used in astronomy for many years and led to revolutionary discoveries and observations. One of the areas relatively undeveloped in terms of image processing is solar image feature extraction [8].

With a substantial increase of solar image databases an automated detection and verification of various features of interest is becoming increasingly important for a reliable forecast of the solar activity and space weather [9].

The visual contents of the images are extracted and described by feature vectors that form a feature database. During the retrieval stage, the users provide the retrieval system with a sample image (query image). The system then changes this query into feature vectors. In the matching stage, the system calculates the similarities between the feature vectors of the query sample and those of the images in the database, and then performs the retrieval accordingly [10].

Many studies have applied the concepts of feature extraction and retrieval to analyze and retrieve images such as:

Csillaghy, et al [11], presented a method that summarizes the image information content by partitioning the image in regions with same texture. They call this process texture summarization. Second, indexing features are generated by examining the distribution of parameters describing image regions. Indexing features can be associated with global or local image characteristics. Both kinds of indexing features are evaluated on the retrieval system of the Zurich archive of solar radio spectrograms. The evaluation shows that generating local indexing features using self-organizing maps yields the best effectiveness of all tested methods.

Seo, et al [12] presented a ROI (Region-Of-Interest) based medical image retrieval system that is considered a combination of feature descriptors and initial weights for similarity matching. For semantic ROI segmentation, they create attention window to remove the meaningless regions included in the image such as background and propose a quad-tree based ROI segmentation method.

Banda, et al [13], presented an extensive comparative analysis of several different domain-specific datasets in order to provide some guidance for the solar physics community on the well- researched field of medical image analysis allowing them to transfer knowledge from one applied field to their own. They report on the transfer of image parameters that produce good results for medical images to the domain of solar image analysis. Using the first solar domain-specific benchmark dataset that contains multiple types of solar phenomena they discovered during their work for constructing a content-based image retrieval (CBIR) system for NASA's Solar Dynamics Observatory (SDO) mission that they could take advantage of the research on the analysis of images in the medical field.

Jyothi et al [14] proposed a retrieval system using region based image retrieval system, finding region in the pictures using a new image segmentation method by improved mountain clustering (IMC) technique and features are extracted using a set of orthogonal set of moment functions for describing images. The achieved average precision rates were in the range of 68% to 76% considering a database covers brain tumor, tooth decay, lung cancer, and tuberculosis images.

2. Concepts and Methods

2.1. CBIR

In 1992, the National Science Foundation of the United States organized a workshop on visual information management systems in order to identify new directions in image database management systems. At this conference, Kato [10] introduced the term Content-Based Image Retrieval (CBIR) to describe automatic retrieval of images from a database. He emphasized the use of color and shape as the criteria for the automatic image retrieval process. Since then, the term CBIR has been adopted to describe an image-retrieving process that is used for large collections of images; which is based on features that can be automatically extracted from the images themselves [15].

Content Based Image Retrieval is a set of techniques for retrieving semantically-relevant images from an image database based on automatically-derived image features. The main goal of CBIR is the retrieval efficiency during image indexing and retrieval, thereby reducing the need for human intervention in the indexing process. The computer must be able to retrieve images from a database without any human assumption on specific domain.

One of the main tasks for CBIR systems is the similarity comparison. It is involved with the extraction of feature signatures of every image depending on its pixel values and, then, defining the rules for comparing images. The extracted features are considered as the image representation structure which used for measuring its degree of similarity with the images registered in a database. Images are compared by calculating the difference of its feature components with the corresponding features of other image [16].

2.2. Edge Detection

Edge detection is an important first stage in the determination of existing orientations in images. Edges correspond to local intensity discontinuities of an image. Edge detection can be used for feature extraction and object or boundary description.

Sobel operator is applied in this work to evaluate the strength of existing edges and to produce a new image with edge information. Sobel operator uses the convolution concept, where a set of two 3 x 3 convolution kernels will be used. One of the kernels is used to detect the brightness changes in the horizontal direction (Gx), and the other one used to detect the brightness changes in the vertical (Gy) direction. Figure (1), illustrates to the kernel of Sobel filter [17].

-1	0	+1
-2	0	+2
-1	0	+1

Gx

+1	+2	+1
0	0	0
-1	-2	-1

Gy

Fig. 1: Sobel operators

Then, the gradient magnitude (i.e., edge strength) can be calculated from the following equation:

$$|G| = \sqrt{Gx^2 + Gy^2} \tag{1}$$

The new |G| is considered as the edge strength of the tested pixel. A new edge image will be created once all the |G| values over the entire coordinates are collected. An example of the input image before and after edge detection process is shown in Figure (2).

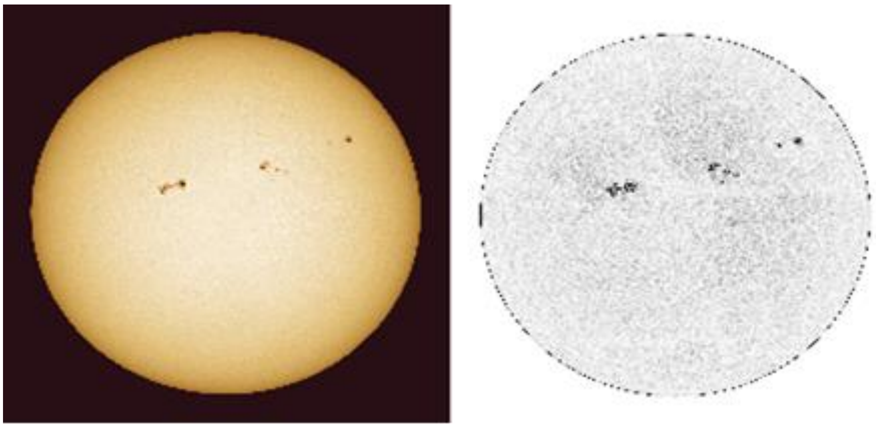

Fig. 2: The input before edge detection and the output after edge detection

2.3. Color Histogram

In this work a conventional color histogram (CCH) is used to indicate the frequency of occurrence of every color in an image, which in turn used as a feature vector represent the statistical behavior of the image color. From a probabilistic perspective, it refers to the probability mass function of the image intensities. It captures the joint probabilities of the intensities of the color channels. Computationally, it is constructed by counting the number of pixels of each color (in the quantized color space). The appealing aspect of the CCH is its simplicity and ease of computation [18].

For an m*n image I, the colors in that image are quantized to C_1, C_2, ……, C_k. The color histogram H(I)={h_1, h_2, ……, h_k}, where hi represents the number of pixels in color Ci. The color histogram also represents the possibility of any pixel, in image I, that in color Ci.

$$p_r(p \in c_i) = \frac{h}{m*n} \tag{2}$$

2.4. Distance Measure

In this work the similarity measure between a pair of images Q and T having feature vectors Qi and Ti is computed using Euclidean distance metric. The uniformity assumption of the feature space implies that the perceptual distances between points in the space correspond to the Euclidean metric. The similarity measure is therefore:

$$d(Q,T) = \sum_{i=0}^{N-1} |Q_i - T_i| \tag{3}$$

Where Q = $\{Q_0, Q_1, \ldots Q_{N-1}\}$ and T = $\{T_0, T_1, \ldots, T_{N-1}\}$ are the query and target feature vectors respectively, d(Q, T) is the Euclidean distance, N values which represent feature vector length.

The distance between two identical images is zero, (i.e., d(Q, T)=0). Smaller values of distance d() indicate more similarity between images and vice-versa. For similarity retrieval of images, the Euclidean distance d(Q, T), can be computed between the query image and all the database images. Then the thresholds value will be used to determine images to be retrieved.

3. Proposed System

Content-Based Image Retrieval (CBIR) systems operate under the query-by-example (QBE) paradigm. As shown in figure (3) an example image is presented to the system and the user makes a query for images that are similar to the given example. Performance of CBIR systems is highly dependent on the properties of the example image.this section we present some lemmas which will be needed in the sequel.

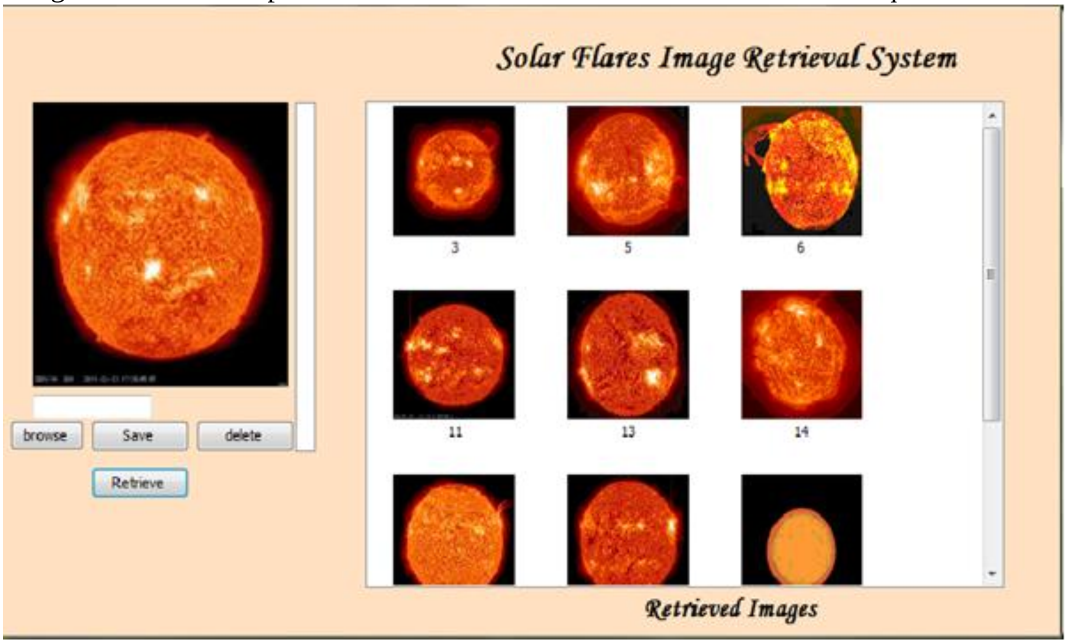

Fig. 3: The Proposed System Interface

Firstly, a user uploads the query image through a user interface, and then its features will be extracted and assembled in a feature vector, which is then compared with archived feature vectors those pre-extracted from the images stored in the image databases. Then a set of images that are very similar to the query image are retrieved and displayed.

The operation of any typical image retrieval system passes through two main phases (enrollment and retrieving similar images). In the first phase the whole set of databases images are passed through the feature extraction module to extract the features vectors and then stored in feature vector databases. The collection of samples consists of 250 images stored in image database, and same number of feature vectors have been registered and saved in a dedicated database called feature vector database. Examples of image together with its corresponding color histogram are shown in Figure (4).

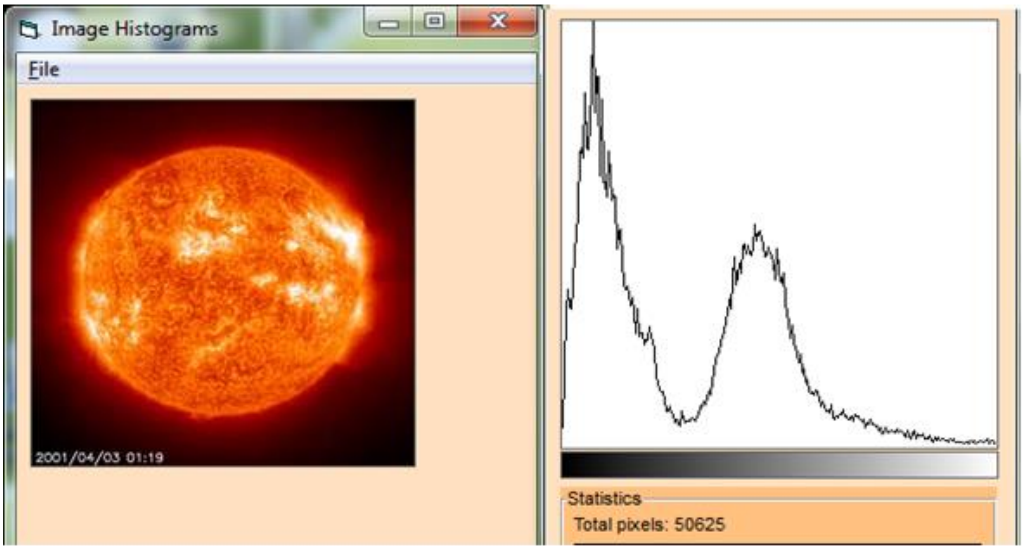

Fig. 4: An example of a solar image and its corresponding histogram

In the second phase of system operation (i.e., retrieving phase), a set of test samples had been used to investigate the efficiency of the established retrieval system. In this phase, the tested image sample is passed through the feature extraction module to extract its feature vector and match this extracted feature vector with stored vectors in feature vector database. Finally, a list of similar images to a given sample is retrieved from image database. For purpose of performance evaluation some of the retrieval results for the conducted tests were used to determine the precision and recall rates of the proposed system. The block diagram for the proposed system is shown in Figure (5).

In both system operation phases (i.e., enrollment and retrieval), the same image loading module is included and also the same feature extraction module is applied.

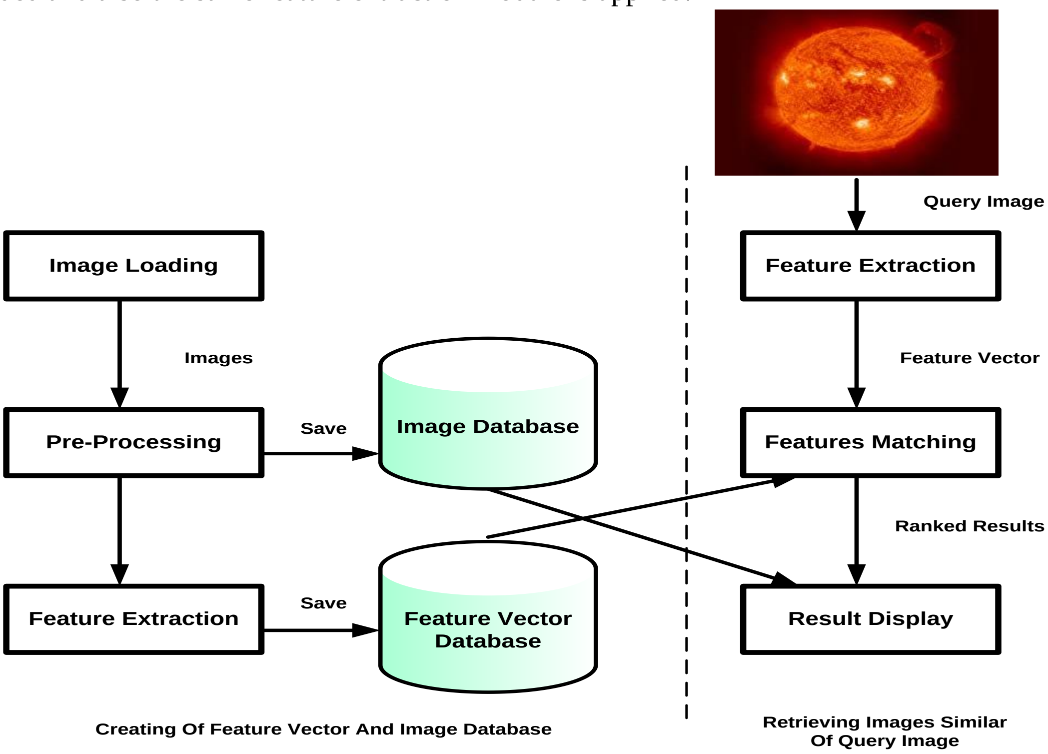

Fig.5: The Diagram of Proposed System

4. Results and Discussion

JIC email for contribution: editor@jic.org.uk

DOI https://doi.org/2024-JICS-22523 | Generated on 2025-04-20 22:03:13

The main stages of the established system are: feature extraction and retrieving using similarity measurement. Two metrics for retrieval effectiveness were used; they are recall and precision. Recall signifies the relevant images in the database that are retrieved in response to a query. Precision is the proportion of the retrieved images that are relevant to the query. They defined as follows [17]:-

$$\text{Precision} = \frac{\text{Retrieved related images}}{\text{Total retrieved images}} * 100\% \tag{4}$$

$$\text{Recall} = \frac{\text{Retrieved related images}}{\text{Total related images}} * 100\% \tag{5}$$

The below test aimed to investigate the effect of using different threshold values on system precision and recall. Table(1) presents the result of this test.

Table1. The effect of threshold value on solar image retrieval precision and recall

Threshold	Precision	Recall
0.3	87%	63%
0.4	85%	67%
0.5	82%	71%
0.6	79%	75%
0.7	75%	78%
0.8	69%	80%

Also, for sunspots image retrieval in preprocessing stage the sobel filer is applied first on the query image and the most similar images retrieved from the database based on detected edges in the images as shown in figure (6).

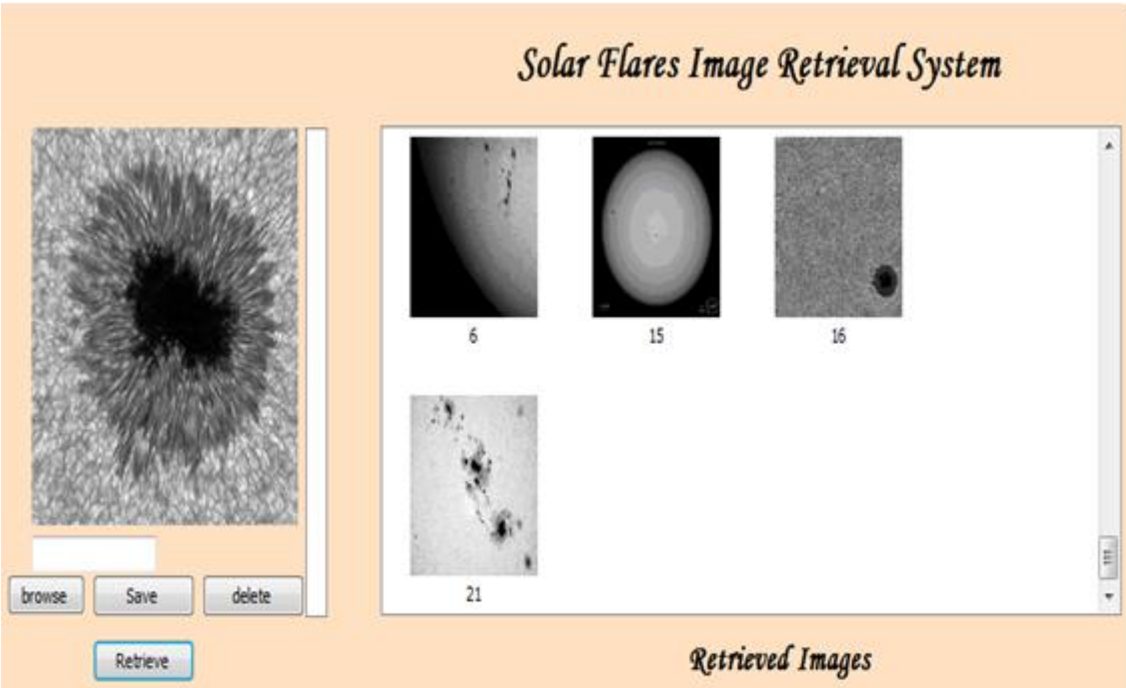

Fig. 6: Sunspots image retrieval

5. Conclusions

The proposed retrieval system facilitates access to images in large archives of astronomical images based on interaction user interface that allow user to quickly obtain an overview of information contained in the images.

The use of color histogram and edge detection using sobel filter can be utilized to describe the color and shape content of solar images.

Testing the effect of using different threshold values helps to find the suitable values which lead to best retrieval results. The established solar images retrieval system gave better precision and recall rate, when the threshold value less than or equal (0.5).

6. References

[1] Halawani A., Teynor A., Setia L., Brunner G., and Burkhardt H., ”Fundamentals and Applications of Image Retrieval: An Overview”, Datenbank-Spektrum, vol. 18, pp. 14-23, 2006.

[2] Greenspan H., and Pinhas A.T., “Medical Image Categorization and Retrieval for PACS Using the GMM-KL Framework”, IEEE Transactions on Information Technology in BioMedicine, vol. 11, pp. 190-202, 2007.

[3] Qahwaji R. and Colak T., “Automated prediction of solar flares using neural networks and sunspots associations,” Solar Physics, vol. 241, pp. 195–211, 2007.

[4] Clark, T.D.G., “A Review of the Effects of Space Weather on Ground Based Technology”, in Space Weather Workshop: Looking towards a European Space Weather Programme: Noordwijk, The Netherlands, 2001.

[5] Gary, D.E., Lanzerotti, L.J., Nita, G.M. and Thomson, D.J., “Effects of Solar Radio Bursts on Wireless Systems”, in NATO ARW on Effects of Space Weather on Technology Infrastructure, 2003.

[6] Shea, M.A. and Smart, D.F., “Space weather: The effects on operations in space”, in Solar-Terrestrial Relations: Predicting The Effects On The Near- Earth Environment, pp. 29-38, 1998.

[7] Guy, F. and Johnson D.C., “Space Environmental Impacts on DoD Operations”, AIR FORCE SPACE COMMAND, 2003.

[8] Artur J. Kowalski, “Image Enhancement Techniques Applied to Solar Featur Deection”, Ph.D. Thesis, Department of Computer Science, Faculty of New Jersey, Institute of Technology, 2003

[9] Zharkova V V., Aboudarham J., Zharkov S., Ipson S.S., Benkhalil A.K., and Fuller N., “Intelligent Approach to Data Analysis with the Solar Feature Catalogues”, Advances in Space Research Journal, ELSEVIER, 2004.

[10] Eakins J., and Graham M., “Content-based Image Retrieval”, University of Northumbria at Newcastle, Report no. 39, 1999.

[11] Csillaghy A., Hinterberger H., and Benz A.O., “Content-Based Image Retrieval in Astronomy”, Kluwer Academic Publishers , pp. 229–241, 2000.

[12] Seo M., Ko B., Chung H., and Nam J., “ROI-Based Medical Image Retrieval Using Human-Perception and MPEG-7 Visual Descriptors”, Springer-Verlag Berlin Heidelberg, pp. 231 – 240, 2006.

[13] Banda J.M., Angryk R.A. , and Martens P.C., “On the surprisingly accurate transfer of image parameters between medical and solar images”, pp. 3669 - 3672 , 2011.

[14] Jyothi B., Madhavee Y., Krishna P., and Reddy V., “Medical Image Retrieval Using Moments”, International Journal of Application or Innovation in Engineering & Management (IJAIEM), vol. 2, issue 1, pp. 195-200, 2013.

[15] Kato T., “Database Architecture for Content-Based Image Retrieval,Proceedings of Image Storage and Retrieval Systems (SPIE), pp. 112-123, 1992.

[16] Aulia E., “Hierarch Indexing for Region Based Image Retrieval”, M.Sc. Thesis, Department of Industrial and Manufacturing Systems Engineering, Louisiana State University, 2001.

[17] Kekre H. B., Thepade S., Mukherjee P., Kakaiya M, Wadhwa S., and Singh S., “Image Retrieval with Shape Features Extracted using Gradient Operators and Slope Magnitude Technique with BTC”, International Journal of Computer Applications, vol. 6, no. 8, 2010.

[18] Gonzalez A., and Reyes E., “Content-Based Image Retrieval Using Topological Descriptors”, ISSN 2072-6287, no. 2142, 2011.