A Novel Content Based Image Retrieval Implemented By NSA Of AIS

Monika Daga, Kamlesh Lakhwani

Abstract: Content Based Image Retrieval system was developed long back, a technique using visual content according to the interests of the users', to search images from large scale image databases. Since then various methods and techniques are being applied for generating better results. Growing interest and inspiration from biological immune system i.e. the concept of Artificial Immune System (AIS) immersed, on the other hand, is a new computational paradigm for pattern Recognition. In this paper, a new CBIR system is being implemented using the Negative Selection Algorithm (NSA) of AIS. MATrix LABoratory functionalities is being used to develop a novel CBIR system. It has reduced complexity and an efficiency of retrieval is increasing in percentage depending upon the image type. This is the first ever system to use NSA during image comparison. This new method has been paved in my mind so as it can be helpful in various applications like medical image databases, art collection and World Wide Web.

Index terms: Artificial Immune System (AIS), Content Based Image Retrieval (CBIR), Edge Detection, Euclidean Shape Space, Hamming Shape Space, Negative Selection (NS), Pattern Recognition, RGB Color Histogram.

1. INTRODUCTION

Content-based image retrieval (CBIR) is the application of computer vision techniques to the problem of image retrieval, i.e. the searching for digital images in large databases. Meaning of Content-Based is that the search will analyze the actual contents of the image rather than the metadata associated with the image. It is a technique which uses visual contents to search images from the large scale image database according to users' interests. The term 'content' in this context can lead to colors, textures, or any related other information that can be derived from the image itself. Mostly all researchers are particularly interested in the abilities of this raising AIS system whose complexity is comparable to that of the brain. A new area of research called Artificial Immune Systems has been highlighted by many great researchers, but no formal general framework has been presented yet. It is also known query by image content (QBIC) and content-based visual information retrieval (CBVIR) Most AIS systems aim at solving complex computational problems from mathematics, engineering, and information technology.

2. IMMUNE SYSTEMS AND PATTERN RECOGNITION

As we know, the biological Immune system is identified according to two mechanisms namely: Innate and Adaptive Immunity. Innate immunity is directed against general pathogens that enter the body while adaptive or acquired immunity is the one that enables the immune system to recognize and remember specific pathogens (previously encountered), and to mount stronger attacks beforehand each time the similar pathogen is encountered. In a nutshell, we can say that the human immune system is made up of B Cells and T Cells.

- *Monika Daga is currently pursuing her Master's Degree Program in Computer Science Engineering from Suresh Gyan Vihar University, India, PH-09928107569. E-mail: monikadaga9@gmail.com*

The B Cells are used in the defense against infection. Upon encountering an antigen, B Cells are stimulated by a number of sources and with the help of T Cells undergo cloning and somatic hyper-mutation when the B cell is sufficiently stimulated. The antigens are then attacked by killer T Cells and removed from the system [4]. The immune system maintains a memory of the infection by the antigen so that if ever exposed to the same antigen a quicker response can be elicited against the infection. The biological immune system is an extremely parallel, adaptive, and distributed system. It uses learning, memory, and associative retrieval to solve recognition and classification tasks [5]. As described in [3] from a pattern recognition point of view, the most challenging characteristic of the Immune System is the presence of receptors, on the surface of immune cells which are capable of recognizing an almost limitless range of antigenic patterns. The biological immune system has the ability to learn and recognize applicable patterns, memorize patterns that have been captured previously. It also uses data mining methods to construct pattern detectors effectively. Thus the extraordinarily efficient information processing capabilities of the immune system provide important aspects in the field of computation and pave the way for the emergence of Artificial Immune System as a novel computational paradigm. From the point of view of pattern recognition in the immune system, the most important feature of both B and T cells is that they have receptor molecules on their surfaces that can recognize antigens (either free or bound to an MHC molecule). In the B cells case, the receptor is an immunoglobulin, or antibody, molecule embedded in the membrane of the cell. In the T cells case, the receptor is simply called the T cell receptor (TCR). Recognition in the immune system occurs at the molecular level and is based on shape complementarity between the binding site of the receptor and a portion of the antigen called an epitope. While antibodies posses a single kind of receptor, antigens may have multiple epitopes, meaning that a single antigen can be recognized by different antibody molecules. B and T cell receptors recognize different features of an antigen. The B cell receptor communicates with epitopes present on entire antigen molecules. Antigens may be soluble or confined to a surface. The T cell receptor communicates only with cell surface or boundary molecules. The T cells releases chemical substances which kill other cells or promote their growth and hence playing a very important role in the regulation of the immune responses. By recognizing a cell surface molecule, the

T cell has to identify whether it is interacting with another cell rather than with a soluble molecule. The T cell receptor recognizes antigens bound to a cell surface molecule called a major histocompatibility complex (MHC).

3. METHODOLOGY

As stated earlier, we aim to implement a novel Content based image retrieval system through data mining using Artificial Immune system. The proposed methodology is as follow:
A number of images are stored in the database and two features i.e. RGB color histogram and Edge detection will be extracted using MATLAB. Next, a suitable vector representation (Euclidean & Hamming Shape-space) of these features is stored in the database along with the respective images. Next we will show the steps involved in image retrieval.

1. A query in the form of image is entered by the user using a suitable GUI.

2. This image is fed to MATLAB and features of the image are extracted. These features are again represented in the same vector format as for the stored images.

3. Now the features of the query image are matched with the features of each image in the database. This is done through the pattern recognition feature of data mining. We choose the Negative selection algorithm of Artificial Immune system to match these features. We have modified the algorithm as per our requirements. This algorithm is inspired by the Pattern recognition capabilities of immune system. Working of Negative Selection algorithm is explained in the further sections.

4. The matching between the images is done by using Euclidean shape in case of RGB color histogram since feature vectors of the images are real valued , and by using Hamming distance approach in case of edge detection since the feature vector of images in this case are binary valued.

5. If the features of two images (the query image and the image from the database) match with a certain predefined threshold value, then that image is retrieved and displayed. Otherwise the image is discarded.

6. The above process is repeated with every image of the database and the suitable set of images is displayed using another GUI.

3.1 Edge Detection

Edge detection is a fundamental tool in image processing particularly in the areas of feature detection and feature extraction which focus at identifying points in a digital image at which the image brightness changes sharply or, more formally, has discontinuities. However, in my project the main purpose to apply an edge detection algorithm to an image is that it significantly reduces the amount of data to be processed and may therefore filter out information that may be regarded as less required while preserving the image's important structural properties. If the step of edge detection is successful then the subsequent task of interpreting the information contents in the original image may therefore be substantially simplified.

Similarity Computation using Hamming-shape space distance

The simplest case is to simply calculate the Hamming distance (d) between these two pixels, as given by Equation.

$$d = \sum_{i=1}^{L} \delta, \text{ where } \quad \delta = \begin{cases} 1 & \text{if } p_i = m_i \\ 0 & \text{otherwise} \end{cases} \qquad [2]$$

Here,

p_i = Co-ordinates value obtained from edge detection

m_i = Co-ordinates of binary random matrix

In this method we match the values of query image with the Random matrix; if the value matches and the number of matching co-ordinates is within the already fixed threshold then we will retrieve those images as the similar images to the query images. This approach has an advantage that it is easy to implement and also it consumes less computation time.

3.2 RGB Color Histogram

An RGB image, sometimes referred to as a true color image, is stored in MATLAB as an m-by-n-by-3 where m and n refers to the image size, for each individual pixel data array is defined for red, blue, and green color components . RGB images do not use a palette. Combination of the red, blue and green intensities stored in each color plane at the pixel's location determines the color of each pixel. Red, blue and green components are 8 bits each and Graphics file formats store RGB images as 24-bit images. This gives a potential of 16 million colors (2^{24}). The precision with which a real-life image can be replicated has led to the commonly used term true color image.

Similarity Computation using Euclidean-shape space distance

Using Formula as mentioned below for calculating the Euclidean Distance:

$$d = \sqrt{\frac{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2}{3}} \qquad [2]$$

Here,

p1= Co-ordinates of Red pixel

p2= Co-ordinates of Green pixel

p3= Co-ordinates of Blue pixel

q1, q2 & q3= Co-ordinates of Random matrix

Here, calculation of distance between the query image and Random matrix is performed. If the distance is within the already fixed threshold then we will retrieve those images as the similar images to the query images. This approach has advantage that it is easy in implementation and also has a disadvantage that it consumes more computation time when the number of images in the database increases.

4. PROPOSED WORK

I have modified the existing Negative Selection Algorithm (NSA) according to the suitability of my application. This algorithm is been used to compare the Query image with the images in the database. The block diagram for the modified NSA is shown in figure 4.1. Modified Negative Selection Algorithm is described in the following steps:

1. Generate random matrix (C) using the MATLAB function.

2. Compare the random matrix (C) with the query image matrix (P) using Euclidean shape-space or Hamming shape-space. If a match occurs then discard this coordinate of P; else store this Coordinate of P in the Memory detector set M.

3. Assume a set of different images in the database.

4. Compare the only coordinates saved in memory detector set M for the images in the database using either Hamming-shape space distance or Euclidean-shape space distance. If a match occurs then discard this coordinate of image from database; else presence of non-self coordinate is detected.

5. Set a certain threshold and count the number of coordinates which didn't matched and if the count exceeds the set threshold display the image; else discard the image.

6. Repeat step 4 & 5 for all the images in database.

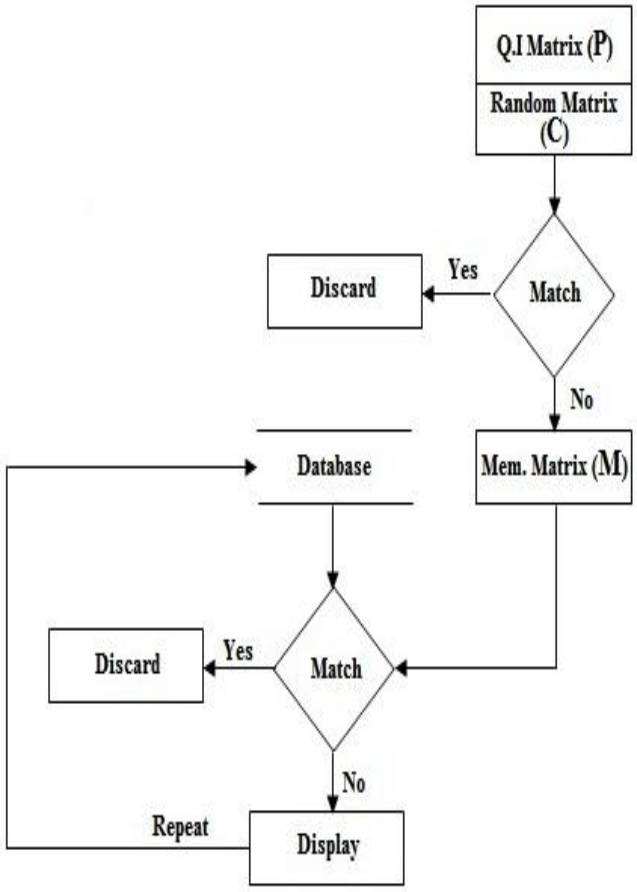


Figure 4.1 Block Diagram for Modified Negative Selection Algorithm

Q.I Matrix: Query Image Matrix

Mem. Matrix: Memory Matrix

Here, there is dealing with large sets of matrix computations. In order to prove the comparison between two images we need to compare their matrices with each other which will require FOR loop iterations which will not only maximize the code complexity but also maximizes the running time of the software. Hence Random matrix comes to the action; we will generate a random matrix of 200X200 of uint8 and uint32 type using inbuilt RAND function in MATLAB.

5. RESULT ANALYSIS

I have experimented with a total of four images. And have calculated the efficiency of each of the CBIR as well as the one developed by me i.e. NSA CBIR System.

Efficiency can be calculated using formula mentioned below-
Efficiency:

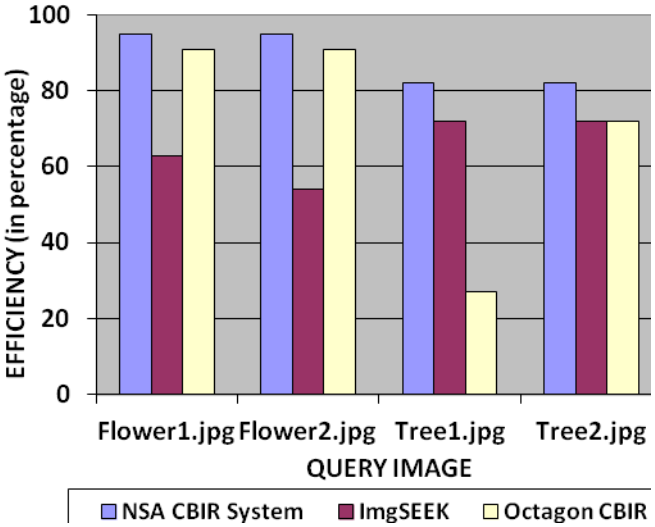
$$\eta) = \frac{\text{Total number of similar (relevant) images displayed}}{\textit{Total number of similar images in the database}} * 100$$

Table 5.1 Comparison of Efficiency of ImgSEEK, Octagon CBIR, and NSA CBIR System

Query Image	*NSA CBIR System*	*ImgSEEK*	*Octagon CBIR*
	95%	63%	91%
	95%	54%	91%
	82%	72%	27%
	82%	72%	72%

The graphical comparison of the three approaches is shown in the following section. This graph below clearly shows the efficiency of NSA CBIR System which is higher compared to other approaches for all the four different images and the same database. The efficiency of the other two systems is nearly equal.

Figure5.2 Efficiency Comparison



6. Conclusion

A novel working system for Content Based Image Retrieval using a new computational paradigm of AIS which in turn is encouraged from the pattern Recognition abilities of biological immune system, is developed successfully and it has been named as NSA CBIR System. We have used two different features and a combined feature, and hence three different approaches for image search. User is also provided with a choice of type of search for his image. A comparative study for efficiency measurement is conducted. As assumed in the beginning of the project, the efficiency of the new CBIR system is higher compared to its counterparts and has been shown above. This is due to the Negative Selection (NSA) approach of AIS for matching used in the implementation of our system. Using the NSA approach also helps reduce the complexity of the code to a great extent. Hence we prove that the Artificial Immune system approach applied in the field of data mining and Pattern Recognition yields unexpectedly efficient results. The most significant part of this project is that this is an absolutely novel approach of constructing CBIR using Artificial Immune System Approach and has not been developed so far. And to add to our success, it shows unexpectedly high efficiency in terms of search of images.

7. Future Prospects

Although we are able to implement a stable CBIR system using AIS approach but there is always a scope for further enhancements. Following are some of the improvements that can be implemented in the current system in order to make it more efficient and provides some new research directions.

1. We have observed that the CBIR system developed by us so far is on a small database of images for the purpose of experimentation. So we can implement the same on a larger database by linking our system to the WEB. This will solve the scalability issue of our system.

2. Another major improvement can be in terms of algorithm developed. We can implement the same CBIR using other approaches i.e. Clonal Algorithm and Immune Network Algorithm of the Artificial Immune System and compare the complexities and efficiency of

our system with that of these two. There may be some scope of improvement.

It has been observed that the NSA CBIR System developed is very efficient from commercial point of view since we have already surveyed that no such system exists online so far. So it has a great potential in terms of marketability. Few minor improvements and experimentations may make this product market ready.

8. REFERENCES

[1] Forrest, S., Somayaji, A & Ackley, D. H. 1997."Building Diversity Computer Systems", in proceedings of the 6th workshop on Hot Topics in Operating Systems, pp. 66-72.

[2] Xia F., Zhu Y., and Gao Y. 2007. "Shape-Space Based Negative Selection Algorithm and Its Application on Power Transformer Fault Diagnosis", in proceedings of the IEEE International Conference on Robotics and Biomimetics, (Dec 2007), pp. 2149 - 2154.

[3] De Castro, L.N, Timmis, J. 2002. Artificial immune systems: a novel paradigm for pattern recognition. In Artificial Neural Networks in Pattern Recognition, Alonso L, Corchado J, Fyfe C (eds), University of Paisley.

[4] Timmis, J, Neal, M. A resource limited artificial immune system for data analysis. Knowledge Based Systems, 2001, pp. 121-130.

[5] S. Sumathi, S. Paneerselvam, Computational Intelligence Paradigms Theory & Applications using MATLAB. CRC Press 1 edition, 2010.

[6] Zeng J., Li T., Liu X., Liu C., Peng L. and Sun F. 2007. "A Feedback Negative Selection Algorithm to Anomaly Detection", Third International Conference on Natural Computation (Aug 2007), Volume 3, pp. 604-608.

[7] Zhengbing H., Ji Z., and Ping M. 2008. "A Novel Anomaly Detection Algorithm Based on Real-Valued Negative Selection System", Workshop on Knowledge Discovery and Data Mining, (Jan 2008), pp. 499-502.

[8] De Castro, L.N, Timmis, J. Artificial Immune Systems: A New Computational Intelligence Approach". Springer-Verlag, 1996.