

Parallel Approach for Content Based Medical Image Retrieval System

¹M. Emmanuel, ²D.R. Ramesh Babu, ³Jayashree Jagdale, ³Pravin Game and ³G.P. Potdar

¹Dr. MGR Research Institute Chennai, India

²Department of Computer Engineering, Dayanand Sagar Engineering College, Bangalore, India

³Department of Information Technology, PICT Pune, India

Abstract: Problem statement: Parallel implementation of a CBIR system for medical applications due to the rapid progress in the technologies for obtaining and storing digital images for diagnostic purposes in medicine. **Approach:** For feature extraction Wavelet decomposition was used for texture, global color histograms for color and novel wavelet based approach for shape feature extraction. The system had been implemented in the Matlab, for the parallel processing, using the master slave approach the scheme was tested up to 16 hosts, Here parallelism is inherent in program loops, which focused on performing searching operation in parallel. **Results:** The experimental results showed that parallel CBIR systems were the need of the hour for evidence based medicine; the parallelization approach enabled the utilization of the similarity measurement technique because of the capability of parallel data processing in the several computers connected to the computational grid. **Conclusion:** Medical image retrieval played vital roles for many health-related applications such as medical diagnostics, drug evaluation, medical research, training and teaching. Due to the rapid progress in the technologies for obtaining and storing digital images for diagnostic purposes in medicine and the rapid expansion of computer networks and the Internet, parallel CBIR systems are technologically feasible for Medical Domain and should be used in evidence based medicine.

Key words: CBIR, image retrieval, medical image analysis

INTRODUCTION

A content based image retrieval system is an extension of the text-based information retrieval system. However, the techniques used in retrieval systems are deviated from text-based retrieval systems and CBIR has now matured into a distinct research discipline in its own right (Lew *et al.*, 2006). CBIR is a type of system that retrieves images based on the content of the image. This image content can be described by using either its semantic or its visual information. The retrieval of images based on the semantic content is mostly done via keywords or a text phrase. This is often achieved by applying the traditional text-based retrieval approach to analyze the image content through its file name and description tags. This approach is not suitable for large existing databases, where text annotations in images are often not available and the image filename rarely reflects the true interpretation of the actual content.

The primary goal of study and implementation of parallelization is to enhance the performance and computational time of Sequential program by converting them into parallel to run on more than one processor. Parallelization serves the following purpose.

Any group of Windows machines dedicated to a single purpose can be called a cluster, Master node monitor workstations in its domain, If the users have left their workstations on, till command is executing at the backend, The Master computer has a feature of executing applications at the client Workstations. Content based image retrieval systems are widely used in medical imaging. The aim is to provide relevant support in medical imaging diagnosis. The objective of CBIR in the hospital information systems is to give the right information to the specialist in the appropriate time, to improve the quality and efficiency of the diagnoses. In the process of clinical decisions, CBIR provides benefits, being able to retrieve images in databases of the same form, anatomical region and pathology. The very specific characteristics of the medical image content require a specialized design of CBIR systems. In that manner, a variety of information provided by the different types of medical image acquisition devices and technologies such as Magnetic Resonance Imaging (MRI), Magnetic Resonance Spectroscopy (MRS), Computed Tomography Scanners (CTS), Single Photon Emission Computer Tomography (SPECT), Positron Emission Tomography (PET),

Corresponding Author: M. Emmanuel, Dr. MGR Research Institute Chennai, India

Electrical Impedance Tomography (EIT), Ultrasound Probes (US), need to be adequately analyzed. CBIR systems can be used for searching and retrieving different types of images from large databases on the bases of the visual content of the images. Thus the precise retrieval from the large medical images collections can be accomplished by indexing or feature extraction, namely, by describing the images. So the phase of making appropriate description of the medical image content is very important to achieve accurate, efficient and correct retrieval process.

Related study: Several forms of parallel computing are Instruction level, bit-level, task and data parallelism, all forms are exploited from many years for high-performance computing, due to this reason parallel computing has become the dominant paradigm in computer architecture. Parallel image retrieval has proposed by Abacus group from City University of Hong Kong uses Beowulf cluster is for a distributed image retrieval system, Plaza *et al.* (2006) worked on cluster based retrieval, in (Petcu *et al.*, 2007) proposes Grid based platform for parallel processing, In contrast to that, here we present an approach that uses data parallelism. This approach has the considerable advantage in terms of retrieval speed. In the remainder of this study, we discuss the approaches used for extracting texture, color and shape features in content-based medical image retrieval system, with parallelization and experimental results. Finally, we summarize our findings and their importance.

MATERIALS AND METHODS

Texture describes the content of many real world images such as clouds, bricks and trees. All of which have textual properties, CBIR systems use texture characteristics which are derived from either Gabor Wavelets or the DWT. Ma and Manjunath (1996) have done experimental comparison of features and found that best features were based on the mean and standard deviation of the Gabor Wavelets. Kingsbury (1998) has proposed a new complex wavelet transform which gives fast way of computing Gabor like wavelets, accuracy in this method is similar to the Gabor Wavelets.

Wavelet transformation is used over the years for image analysis, in wavelet transform mother wavelet function $\Psi(t)$, chosen as basis function to represent a signal, while the scaling function $\Phi(t)$ is used to approximate the function at different scales. In the case of the wavelet transform of a signal f of length N , the approximate space V_j of the signal at resolution 2^{-j} is

decomposed into lower resolution space and a detail space. Low pass and high pass filters are used to divide the orthogonal basis into two bases in a two channel filter bank. A more general form of the wavelet basis, adaptively segments the frequency axis. Frequency intervals of varying width are adaptively selected to extract the frequency content present in the given signal. In order to approximate the signal f belonging to an approximation space $V_j = W_j^0$ at a resolution 2^{-j} using wavelet packets, the space W_j^0 is decomposed into 2 orthogonal spaces. The scale, frequency and position indices of the wavelet function are denoted using d , n and l respectively. As given in various text books, continuous WT decomposes a signal f into a set of scaling functions by using a wavelet functions basis:

$$(W_a f)(b) = \int f(x) \Psi_{a,b}^n(x) dx \tag{1}$$

The basis of wavelet functions is obtained by scaling and shifting a single mother wavelet function $\Psi(x)$:

$$\Psi_{a,b}(x) = \frac{1}{\sqrt{a}} \Psi\left(\frac{x-b}{a}\right) \tag{2}$$

The mother wavelet should only satisfy the zero-average condition, i.e., $\int \Psi(x) dx = 0$. The Discrete WT is obtained by taking $a = 2^n$ and $b \in \mathbb{Z}$, The oldest and simplest, in this study pyramid-structured wavelet transform used for texture classification. It recursively decomposes sub signals in the low frequency channels, the texture image is decomposed into four sub images, in low-low, low-high, high-low and high-high sub-bands. At this point, the energy level of each sub-band is calculated using Eq. 3:

$$E = \frac{1}{MN} \sum_{i=1}^m \sum_{j=1}^n |X(i, j)| \tag{3}$$

Distance between the two sets of energies calculated using Eq. 4:

$$D_i = \sum_{k=1}^k (x_k - y_{i,k})^2 \tag{4}$$

Visual features can be classified in Eakins and Graham (1999) into primitive features and logical features. All currently available systems use primitive features such as color or shape. Color has been the most effective feature. Alaoui *et al.* (2009) and Pourghassem and Ghassemian (2008) explore the idea

of color spatial indexing and image classification, in this work color features are extracted using Global color histograms. The equation we used in deriving the distance between two color histograms is the quadratic distance metric, the problem of comparing totally unrelated bins is solved by using below equation:

$$d^2(Q,I) = (H_Q - H_I)^t A (H_Q - H_I) \quad (5)$$

RESULTS AND DISCUSSION

To reduce code redundancy and enhance reusability, our goal when designing the parallel algorithm was to reuse much of the code for the sequential one. The algorithm has been implemented in the Matlab programming language. For the parallel processing, using the master slave approach the scheme is tested up to 16 hosts, the results are compared with that of a sequential program on a single host. In the experiment, for the Master Slave approach each host is already installed CBMIR software and connected by a 100 Mbps Ethernet, for the parallel implementation, the experiments were conducted by using many sample images; in the parallelization approach it has observed that the communication time is more than the total processing time. Recall-precision graphs for texture, color, shape and interface, loading query image and the search result is shown from Fig. 1-6.

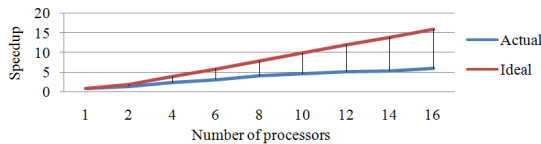


Fig. 1: Speedup graph for color

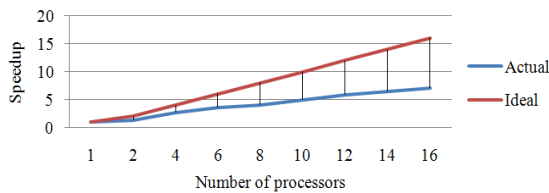


Fig. 2: Speedup graph for texture

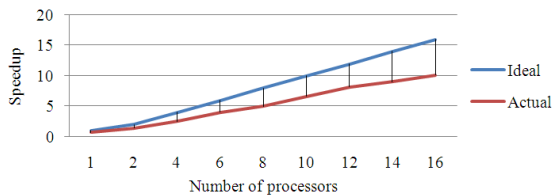


Fig. 3: Speedup comparison graph

Algorithm:

- Step1: Read the query image
- Step 2: Extract the image feature
- Step 3: Divide the whole database into n workers

Step 4: Each worker is assigned N/n work for the evaluation and comparison (where N is the total work), using peval

Example: for n=2 peval ('up = up_sort(last_image, color_feature, d2);', 'down=down_sort(last_image, color_feature,d2);');

Step 5: Compare these values with the values stored in the database

```
d11 = diff(h1)
d1 = Sum(d11)
d2 = Double(1/(1+exp(d1)))
```

Step 6: The computational result is sent to the master again

Step 7: Master chooses the top ten values which are above than the threshold value

Step 8: Use the insertion sort to get the images in the ascending order and display the top three images

Step 9: End

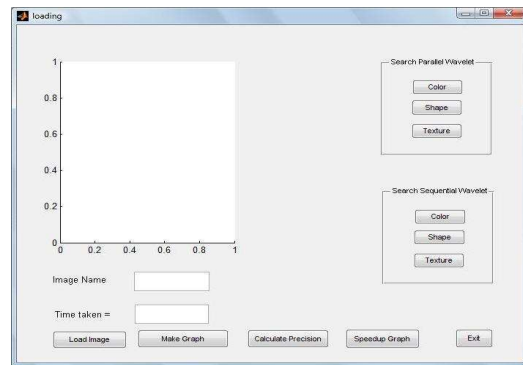


Fig. 4: GUI for loading query image

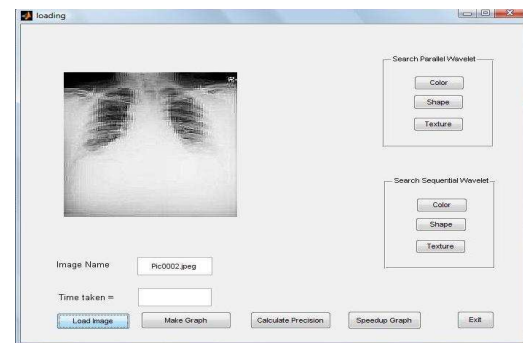


Fig. 5: Loading query image

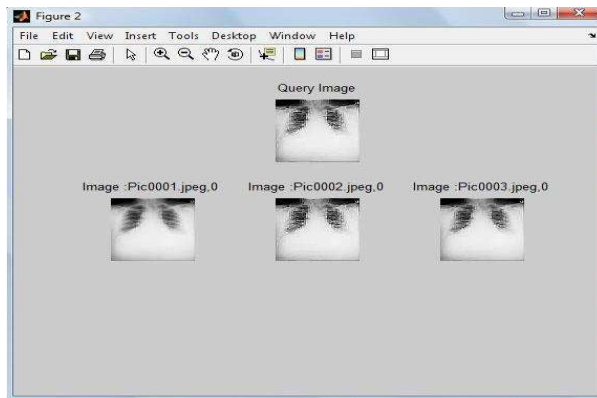


Fig. 6: Parallel search result

Like other information retrieval systems, a CBIR system resolves queries in an approximate way, because users are not aware of the exact results that should be delivered. That means that CBIR systems must be evaluated to determine a degree of precision of the retrieval process, revealing how good the system is at locating relevant images associated to the user query. Since the most important contributions of the proposed architecture is the content-based image retrieval module, it is important to assess its performance. In this work a specific evaluation has been made using an experimental dataset of cells, hand and chest images. to evaluate the performance of the proposed approach, system provides the user with the ability to query the image database with an image sample. Results obtained in this work were evaluated with “precision” and “recall” parameters which are typically utilized for evaluating systems of content-based images retrieval and information retrieval. “Recall” means the ratio of relevant images over the number of images retrieved in the query. On the other hand, “precision” is the ratio of retrieved images that are relevant for reference. The medical staff implied in our evaluations and experiments considered it useful both in the diagnosis process and in the training of the medical students.

CONCLUSION

Medical image retrieval plays vital roles for many health-related applications such as medical diagnostics, drug evaluation, medical research, training and teaching. Due to the rapid progress in the technologies for obtaining and storing digital images for diagnostic purposes in medicine (from photography over digital

radiography to functional MRI and PET) and the rapid expansion of computer networks and the Internet, medical image databases for training and supporting diagnostics have become technologically feasible. Many algorithms, architectures and systems have been studied and developed to help search and browse through large multimedia databases based on content. Because of the importance of medical imaging, recently there is increasing interest by informatics researchers and physicians to develop CBIR algorithms, as well as architectures for medical image applications. In addition to efficient and convenient repositories of medical images, these CBIR systems can be also used as aids for medical diagnostics and training of physicians. Because of their high computational cost, the utilization of similarity measurement algorithms executed in a single computer becomes unfeasible in computer-aided diagnosis. The utilization of parallelization in medical applications is still at its beginning; however this is a promising technology and significant developments in IT applied to the health care field can be expected in the near future. The parallelization approach has shown as a promising tool in the processing and storage of great data.

REFERENCES

- Alaoui, R., S.O. El Alaoui and M. Meknassi, 2009. Spatial color indexing: An efficient and robust technique for content-based image retrieval. *J. Comput. Sci.*, 5: 109-114. <http://www.scipub.org/fulltext/jcs/jcs52109-114.pdf>
- Eakins, J.P. and M.E. Graham, 1999. Content-based image retrieval: A report to the JISC Technology Applications Programme. Springer. <http://www.citeulike.org/user/GJNauta/article/4115427>
- Kingsbury, N., 1998. The dual-tree complex wavelet transform: A new efficient tool for image restoration and enhancement. *Proceeding of the European Signal Processing Conference*, Sept. 1998, Scientific Literature Digital Library and Search Engine, United States, pp: 319-322.
- Lew, M.S., N. Sebe, C. Djerba and R. Jain, 2006. Content-based multimedia information retrieval: State of the art and challenges. *ACM Trans. Multimedia Comput. Commun. Appli.*, 2: 1-19. DOI: 10.1145/1126004.1126005
- Ma, W.Y. and B.S. Manjunath, 1996. Texture based pattern retrieval from image database. *Multimedia Tools Appli.*, 1: 35-51. DOI: 10.1007/BF00717822

- Petcu, D., D. Zaharie, D. Gorgan, F. Pop and D. Tudor, 2007. MedioGrid: A grid-based platform for satellite image processing. Proceeding of the 4th IEEE International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, Sept. 6-8, IEEE Xplore Press, Dortmund, pp: 137-142. DOI: 10.1109/IDAACS.2007.4488392
- Plaza, A., D. Valencia, J. Plaza and P. Martinez, 2006. Commodity cluster-based parallel processing of hyperspectral imagery. *J. Parall. Distribut. Comput.*, 66: 345-358. DOI: 10.1016/j.jpdc.2005.10.001
- Pourghassem, H. and H. Ghassemian, 2008. Content-based medical image classification using a new hierarchical merging scheme. *Comput. Med. Imag. Graph.*, 32: 651-661. PMID: 18789648