Large Scale Image Feature Extraction from Medical Image Analysis

Desai Devanshi Manojbhai[1], R.Rajamenakshi[2]

[1]Gujarat Technical University, India
[2]C-DAC, Pune, India

***Abstract*— *Research in big data is focused on deriving knowledge from multiple data formats using intelligent analytics techniques. Medical analytics is a typical example, which encompasses data in multiple formats available as text, images and the data in the databases. Performing large scale image data analysis for near real time results is a challenging task. The challenge here is to extract features without compromising on the performance. Medical images are derived from multiple devices and are analyzed by health professionals manually which is qualitative. Automated deriving of intelligence and guidance will make the disease diagnosis accurate and faster. Automated analytics will also help to predict the progress of the disease and treatment plans. Data processing on huge image corpus is both storage and compute intensive. This paper presents a comparative survey of the image feature extraction techniques using parallel and high performance computing against non-parallel ones over the medical images.*

Keywords*— *Big data, Medical Image analysis, Feature extraction.

I. INTRODUCTION

Research in medical big data is gaining importance because of the huge corpus of data. This huge corpus of data spans multiple research areas spanning bio- informatics, disease analysis, genomics, drug discovery etc. Image analysis in general, aims at automatic extraction and analysis of spatial context, that characterizes the content from the images for further knowledge engineering. Medical image analytics is focused to derive knowledge from images obtained noninvasively from imaging devices such as X-Ray, CT scan and MRI. Models are designed to extract knowledge in an automated fashion while producing quantitative results. Images thus captured are stored as DICOM images that hold meta data along with the images. Braiamis [31] presses the needs for automatic medical image classification with high accuracy. There is a huge corpus of image data lying in the PACS Systems in the hospitals [24] qualifying it to be termed as big data. Analytics on medical image poses a huge challenge of producing analytics in real time from a huge repository of image data due to its need for huge computational and storage needs. Massively parallel programming techniques and high performance computing infrastructure of the recent times bridges this gap. This paper presents a comparative survey and study of feature extraction from medical images using parallel and high performance computing against serial.

A. Image analysis Pipeline

Figure 1 presents the overview of the image analysis process as defined in [5]. Image analysis comprises of the following steps: (a) image acquisition (b) preprocessing (c) image segmentation (d) feature extraction and feature selection and (e) data analysis. Each of these image analysis phases having own individual processes and their own methods/techniques for medical image analysis.

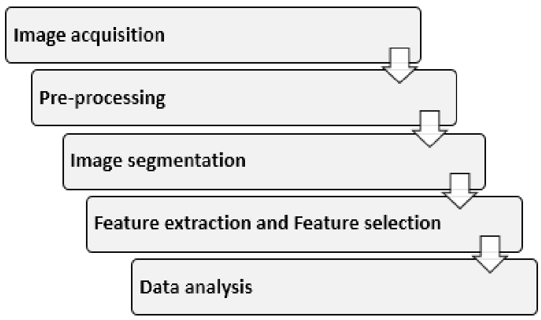

Fig.1: Image Analysis Process

Firstly, high qualities images are collected and stored that are used for diagnosis or for further exploration. Secondly, these images are partitioned into smaller segments, which are referred to as segmentation. Thirdly image features such as color, shape, size, texture are extracted using feature extraction techniques. Feature extraction also relates to dimensionality reduction and feature selection that defines the region for further analysis. The extracted features are expected to contain the relevant information from the input data. The selected features are then analyzed.

B. Overview of Feature extraction

Feature Extraction [14] is the process of creating a representation, or a transformation from the original data. If the features set will extract the relevant information from the input data. Mostly, Feature extraction algorithms used to detect and isolate various segment of a medical image.

Nowadays more research in medical image processing which makes feature extraction and detection of important metadata in large number of image series. However, the process of extracting important features from large image datasets of medical images is extremely demanding in terms of computation time, storage capacity and network bandwidth. The Map Reduce framework is a distributed computing framework, which can be recently been used for large–scale image description and analysis.

This paper mainly presents the state of art study and survey of the image feature extraction techniques. This paper is divided into four sections as follows. Section II describe the related work in image feature extraction, section III presents survey in large scale image processing and feature extraction; Section IV describes analysis and findings and finally section V describes conclusion.

II. IMAGE FEATURE EXTRACTION

A. Classification of Image feature extraction

This section presents the survey of the image feature extraction techniques. Image feature reduction has two aspects feature extraction and feature selection. Feature reduction techniques are used to precisely represent the context of image or it is that portion of image that is under study. For medical images, it is true that not all image feature extraction techniques are suited to extracting features due to the nature of medical image. Mainly the images having organs of different shape and size also complexity and inaccuracy. There is no standard image feature extraction technique but here is some important produce which gives satisfactory results for all imaging applications like brain tumor, brain cancer diagnosis etc.

Fig. 2: Taxonomy of feature extraction/selection methods

Figure 2 describes the taxonomy of Feature extraction and selection methods [8][11][14][18][19].Table 1 presents the set of image features based on shape, intensity, texture and color. Image features at various levels of complexity are extracted from the image data.

Table 1: Image features and their properties

Image Features	Properties
Shape Based Features	Area, Circularity, Irregularity, Perimeter, Roundness[9]
Intensity features	Mean, Median, Intensity, Standard Variance Kurtosis, Skewness[9]
Texture Features	Contrast, Correlation, Entropy, Homogeneity, Sum of square variance[9] Spectral and spatial [1]
Color Based Feature	Impression, Expression and Construction[21] RGB, LUV, HSV and HMMD[1]

Feature extraction of medical images is used to collect effective information from large scale image data. Analyze objects and images to extract the most important features that are represents various medical images. Different methodologies of feature extraction have been used to detect and classification of changes in medical image such as wavelets [2], Feature extraction done using statistical image processing methods. The features that were most promising were color, texture and shape/edge.

B. Algorithms, methods and techniques used in Feature extraction and selection

Table 2: Feature Extraction Models

Category	Models
Text feature Based models	• Gaussian Markow RMF[24] • Fractal Model[24] • Probabilistic model[24] • Simultaneous Autoregrassive Model[24]
Colour Feature Based model	• Steerable Pyramid[12][14] • Contourlet Transform[12][14] • Gabor wavelet Transform[12][14] • K-means based CIS[2]

Table 2 presents the existing models for extracting text and color based features using different models. There exists different tool and packages that progresses images of various types, extract features and much more. [4] Presents different feature selection algorithms such as Correlation based Selection (CBS), Sequential Forward Selection (SFS), Sequential floating forward selection, Goodness based sequential floating forward selection to select best features that brings out the context of image study. [15] has used PCA and LDA analysis for selecting the features for further image analysis. Signal processing techniques [30] such as wavelet transformation and Fourier transformation to extract image features and automatic classification. [25] uses first and second order statistics. Gabor filter and spectral Fourier power spectrum to extract texture based features. [20] used grid color moments, Local binary patterns, Gabor wavelength and edge orientation histogram for extracting color based features from images.

III. LARGE SCALE IMAGE PROCESSIING AND FEATURE EXTRACTION

Kumar et.al [5] presents the current day challenges in medical image analysis and presents a semantic based approach for high throughput image analysis for effective diagnosis. [5] Presents the challenges in image capturing that includes image quality, pixel distortion, noise etc. Krishnan [1] introduced general feature extraction in image processing. Different image feature extraction methods have been used in medicine and science for performing image processing on the images. Singh [2] applied machine learning and pattern recognition techniques on MRI images to brain abnormalities. Current day image segmentations are quantitative and are selected based on visual appearance and shape descriptors. Hence [3] considered spatial domain filtering and frequency domain filtering using spatial interaction for accurate segmentation of medical structures. Serag[4] proposed an optimal feature extraction for classifying FDG-PET in patients with dementia.

Table 3: Parallel Feature Extraction: Methods & Models

Feature Extracted	Methods and Models
Texture Extraction methods	• LoG method [27] • gLoG method [27] • HLoG method [27] • Difference of Gaussian (DoG) method [27] • SVM(Support Vector Machine)
Texture extraction Models	• Gaussian Markov Random Field (GMRF) model [13] • Homogeneous Texture Descriptor (HTD) [13]

In general, Apache Hadoop processes very large scale data files containing large number of sequences of input record, which can be processed parallel. This large data file is split automatically, along with distribution of file system block boundaries, Because of granularity of these input splits in distributed blocks usually contain many data records. Increasingly large and complex algorithms make processing image data more challenging. It requires a lot of computation power, storage and network bandwidth.[22] listed more flexible and scalable infrastructure approaches like (i) Single, powerful machine (ii) Local cluster/ grid (iii) Alternative infrastructures (Graphics cards) and (iv)Cloud computing solutions.[22] computes a decision boundary (hyper plane) using SVM that separates inputs of different classes represented in a given feature space transformed by a given kernel. The main concern is the cost (C) and the Gaussian kernel error σ. Parallel search for optimal SVM parameters has been carried out on Hadoop cluster. There is clear link between runtime map task and resulted classify accuracy can be observed. Most task has long runtime, this interruption of map tasks allowed a reduction of total runtime from 50h to 9h15m, while keeping all values (C, σ) has been considered as best classify performance., But sequential execution would require 990h approximately on desktop computer. [29] Used Map reduce to speed up and make possible large scale medical image processing three paradigm: (i) Parameter optimization for texture classification using Support Vector Machine (SVM), (ii) Content based medical image indexing and (iii) three dimensional wavelet analysis for solid texture classification. [28] Presents a texture pattern classification of tumor using SVM by parallelizing the execution based on parameter values. The execution of the same in a single PC without Hadoop can take weeks to results. [27][13] Demonstrated the use of parallelization methods and models for feature extraction with improved performance and optimization and table 3 presents the summary of the same. Table 4 presents the list of parallel algorithms available for image feature extraction [31] presents an FPGA based architecture for digital image feature extraction. Rao [32] presented a feature extraction from moving images using row-parallel and pixel-parallel architecture based on digital sensor technology for VLSI integration. Here a fully pixel parallel architecture with adaptive binarization of filtered images has been explored to achieve real-time response. [33] Presents FPGA-based architecture for the extraction of texture features using Gray Level Co-occurrence Matrix (GLCM) analysis.

Table 4: Algorithms for parallel features extraction

Features	Algorithms
Texture based	The Marching Squares algorithm
Texture based	The Chain Code algorithm
Color based	JSEG algorithm
Texture based	The Convex Hull algorithms
Texture based	ASGD(Averaging Stochastic Gradient Descent) [13]

IV. ANALYSIS AND FINDINGS

Firstly, there exists a wide class of algorithms and techniques for feature extraction and reconstruction. These methods give a robust performance while retrieving maximum features on medium scale and small scale images. With large scale image analysis, these methods suffer setbacks and are not scalable. With the use of high performance computing techniques and distributed processing platforms for large scale image processing, it is expected to extract maximum features without compromising on performance.

Growing volumes of medical images and the growing need for image analytics for performing large-scale image analysis to understand and analyze hidden patterns and prognosis is on the rise. Secondly, there are no definite models for medical image analysis that performs automated extraction and classification of segments of tumor and associated annotations to derive intelligence. It is very much essential to develop parallel and distributed algorithms to process images in real time.

V. CONCLUSION AND FUTURE WORK

This paper presents a generic survey of the image features, feature selection and extraction methods, models and algorithms. In the context of large scale big image processing, it is necessary to parallelize and utilize high performance computing platform and solutions, to reap the maximum potential. This paper also present the survey of the large-scale image processing techniques available in the literature. Models and algorithms, the challenges present in the literature. We also presented our observations. In this context, we need to develop scalable analytic models that performs image processing that addresses various challenges like image indexing, storage, parallel semantic extraction in multidimensional medical image processing. These models so developed will aid and augment the medical practitioners for effective diagnosis.

REFERENCES

[1] Athira Krishnan, Sreekumar K,A Survey on Image Segmentation and Feature Extraction Methods for Acute Myelogenous Leukemia Detection in Blood Microscopic Images, Athira Krishnan et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014, 7877-7879.

[2] Lavneet Singh, Girija Chetty ,A Comparative Study of MRI Data using Various Machine Learning and Pattern Recognition Algorithms to Detect Brain Abnormalities, in Proceedings of the Tenth Australasian Data Mining Conference (AusDM 2012), Sydney, Australia, 2012.

[3] Ahmed Elnakib, Georgy Gimel'farb, Jasjit Suri, and Ayman El-Baz, Medical Image Segmentation: A Brief Survey, Editors proof,2011.

[4] Ahmed Serag, Optimal feature extraction for the classification of medical images, in Technical Note PR-TN 2008/0033, Issued: 06/2008.

[5] Virendra kumar, Yuhua Gu, Satrajit Basu, Radiomics: the process and challenges, in Elsevier Magnetic Resonance Imaging (2012) 1234-1248.

[6] Ribo Ge 1 and Hailong Shen, Researching on Feature Extraction of Brain CT Image, International Journal of Signal Processing, Image Processing and Pattern Recognition.

[7] Dimitrios Markonis, Roger Schaer, Using MapReduce for Large-scale Medical Image Analysis, 2012 IEEE Second Conference on Helthcare Informatics,Imaging and System Biology.

[8] Matthew Shardlow, A Survey of Stability Analysis of Feature Subset SelectionTechniques, Information Reuse and Integration (IRI), 2013 IEEE 14th International Conference on (2013).

[9] V.P.Gladis Pushpa Rathi and Dr.S.Palani, Brain tumor mri image c lassification with feature selection and extraction using linear discriminant analysis, computer vision and pattern recognition arxiv: 1202.2128 10th august 2012.

[10] Dian Pratiwi, The Use of Self Organizing Map Method and Feature Selection in Image Database Classification System, published in the journal IJENS IJET.

[11] Samina Khalid,Shamila Nasreen, A Survey of Feature Selection and Feature Extraction Techniques in Machine Learning, Science and Information Conference(2014) August 27-29, 2014 | London, UK .

[12] Michele Saad, Low Level Color and Texture Feature Extraction for Content Based Image Retrieval, EE 381K: Multi-Dimensional Digital Signal Processing (2008).

[13] Hong Liu, Ye Wang, Chang-Hai Lei, A Multi-feature Fusion Method for Tongue Image Matching in Traditional Chinese Medicine, Soft Computing and Pattern Recognition (SoCPaR), 2011 International Conference.

[14] Kavya R,Harisha, Feature Extraction Technique for Robust and Fast Visual Tracking: A Typical Review, International Journal of Emerging Engineering Research and Technology Volume 3, Issue 1, January 2015.

[15] Dr. Sanjay Kumar, Er. Ankur Chauhan, A Survey on Image Feature Selection Techniques, International Journal of Computer Science and Information Technologies (IJCSIT), Vol. 5 (5), 2014, 6449-6452.

[16] Shahid Eqbal ,Dr. M. A. Ansari, Medical Image Feature Extraction for Computer Aided Diagnosis of Lung Cancer, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 6, June 2015 ISSN: 2277 128X.

[17] Vadhri Suryanarayana, Dr. M.V.L.N. Raja Rao, Dr. P. Bhaskara Reddy and Dr. G. Ravindra Babu, Image r etrieval system using hybrid feature extraction t echnique, International Journal of Computer Science & Information Technology (IJCSIT) Vol 4, No 1, Feb 2012.

[18] M. Dash, H. Liu, Feature selection for classification, intelligent data analysis, Elsevier, 1997.

[19] Andreas G. K. Janecek, Wilfried N. Gansterer, on the relationship between feature selection and classification accuracy, JMLR 90-105.

[20] Neetesh Gupta, Dr. Vijay anant athavale, Comparative study of different low level feature extraction techniques for content based image retrieval, IJCTEE, August 2011.

[21] Peter L Stanchev, David green Jr, High level color similarity retrieval ICT&P 2003.

[22] Roger Schaer, Using MapReduce for Large-scale Medical Image Ananlysis, HISB 2012.

[23] George Teodoro, Tahsin M. Kurc, Accelerating large scale image analysis on parallel CPU-GPU Equipped system, International parallel and distributed processing symposium, 2012.

[24] Tak-Yee Wong, Dong-Her Shih, a Discrepancy Model Examining the Determinants of User Satisfaction with the PACS System in Radiology Departments of Taiwan Hospitals” ICSAI, 2012.

[25] Prof Vidya Manian, Image Feature Extraction and segmentation, Intel 5046.

[26] Zheng Zhao, Advancing Feature Selection Research, ASU Feature Selection Repository.

[27] Min Zhang, Teresa Wu and Kevin M. Bennett , Small Blob Identification in Medical Images Using Regional Features From Optimum Scale, IEEE transactions on biomedical engineering, vol. 62, no. 4, April 2015.

[28] Yuanquing Lin. Fengjun Lv, Large-scale Image Classification: Fast Feature Extraction and SVN Training, CVPR, 2011.

[29] Andrzej Materka and Michal Strzelecki, Texture Analysis Methods – A Review, Technical University of Lodz, Institute of Electronics, COST B11 report, Brussels 1998.

[30] Maya Nayak and Bhawani Sankar Panigrahi. Article: Advanced Signal Processing Techniques for Feature Extraction in Data Mining. International Journal of Computer Applications 19(9):30-37, April 2011.

[31] D.G. Bariamis, D.K. Iakovidis, D.E. Maroulis and S. A. Karkanis, An FPGA-based Architecture for Real Time Image Feature Extraction.

[32] S.Vishweswara Rao, GLCM architecture for image extraction, International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 3, Issue 1, January, 2014.

[33] A. Veeramuthu and S. Meenakshi, Ph.D, An Efficient Approach for Medical Image Classification using Association Rules, International Journal of Computer Applications (09758887)Volume 90 No.3, March 2014.