Content Based Image Retrieval For Histology Image Collection Using Visual Pattern Mining

U.Ravindran,T.Shakila

Abstract - CBIR is trending to an enormous growth in the field of Artificial intelligence based on visual pattern mining in histology images extends the boundaries of CBIR in Genetic Research. This method starts by representing the visual content of the collection using a bag-of-feature strategy. Then, two main visual mining tasks are performed: finding associations between visual-patterns and high-level concepts, and performing automatic image annotation. Associations are found using minimum-redundancy-maximum-relevance feature selection and co-clustering analysis. Additionally includes an interpretation mechanism that associates concept annotations with corresponding image. A neural network as an classifier, It can improve the classification accuracy more than 80%.

Index Terms— Content-based image retrieval, Visual pattern mining,Bag of features (BOF),Visual-codebook, feature selection,Image annotation, Histology and histopathology images, Fundamental tissues

1 INTRODUCTION

BIOMEDICAL images are an important Source of information, and a potential source of knowledge, for both routine clinical decision and biomedical research[2]. Nevertheless, a thorough exploitation of this potential requires techniques able to automatically extract information and knowledge from this vast amount of data done on the area of medical imaging, which is gradually moving from computer assisted image analysis systems, mainly based on image processing techniques , to fully automatic systems based on pattern recognition and machine learning methods . Most of the work on automatic medical image analysis and interpretation has concentrated on individual images rather than on collections of images. Changing this perspective poses new, and potentially useful, questions: What are the relationships between the images? What are the common and distinctive characteristics among them? What are the implicit categories or groups that could be identified in the collection?

The questions discussed in the previous paragraph can be deemed as instances of a more general image understanding problem,in which the focus of the interpretation process is not an individual image, but the image collection as a whole.

- U.Ravindran, M.Tech., A.P/CSE Dept, *ph*-9840096489. Email-ravindranudhya@gmail.com
- Shakila.T is currently pursuing masters degree program in computer science engineering .*ph*-9944994979. E-mail-achunshaki@gmail.com

This introduces new challenges, but also provides new methods to extract hidden knowledge from data. The paper is organized as follows:

Section 1 presents the proposed method content based image retrival [CBIR] for visual pattern mining; Section 2 describes the details of the different stages of the image collection representation strategy based on BOF; Section 3 an 4 describes the proposed automatic annotation,selection and classification strategy; finally, conclusions and future work are discussed in Section 6 and 7.

2 PREVIOUS WORK

The previous approach is composed of three main phases: a preprocessing step, which corrects luminance differences. A segmentation step that uses the normalized RGB color space for classifying pixels either as erythrocyte or background followed by an Inclusion-Tree representation that structures the pixel information into objects, from which erythrocytes are found. Finally, a two step classification process identifies infected erythrocytes and differentiates the infection stage, using a trained bank of classifiers. state-of-the-art nonlinear classifiers were evaluated for these phases: a multilayer perceptron neural network (MLP) [27] and a support vector machine (SVM).

The main characteristic of the works described above, is that the image analysis concentrates on evaluating information in one image to segment cells or regions which is still a very important and fundamental problem in histology image

analysis. However, our approach is different from these works since we follow a image collection analysis strategy to extract meaningful information out of a set of images rather than process or segment tissues in individual slides. Another branch of research in histology image analysis is the automatic image classification, annotation and retrieval. Since large numbers of digital histology slides are being stored more frequently, methods for automatic image organization and access strategies are becoming important. Histology image classification using multiple transformed features was evaluated by Orlov et al. [19], training classifiers that decide what overall category an image belongs to.
The BOF approach is used to learn discriminative models for automatic image annotation, as well as for analyzing relationships between local visual patterns and image categories from a wider perspective, adding an interpretation layer that aims to explain image collection structures and that supports high-level decision making in histology. Classifiers used in this work are support vector machines (SVM), that receives as input a data representation implicitly defined by a kernel function [56]. Kernel functions describe a similarity relationship between the objects to be classified.
In proposed neural network can be used as an classifier for improving accuracy and speed than svm in both testing and training, Using Neural network can solve non-linear problem

3 CONTENT BASED IMAGE RETRIEVAL

Content-based image retrieval (CBIR), also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) is the application of computer vision techniques to the image retrieval problem, that is, the problem of searching for digital images in large databases.
Content-based means that the search will analyze the actual contents of the image rather than the metadata such as keywords, tags, and/or descriptions associated with the image. The term 'content' in this context might refer to colors, shapes, textures, or any other information that can be derived from the image.
CBIR is desirable because most web based image search engines rely purely on metadata and this produces a lot of garbage in the results.
Also having humans manually enter keywords for images in a large database can be inefficient, expensive and may not capture every keyword that describes the image. Thus a system that can filter images based on their content would provide better indexing and return more accurate results.
In CBIR, images are automatically indexed by summarizing their visual contents through automatically extracted quantities, or features, such as color, texture or shape.
Thus, low-level numerical features, extracted by a computer, are substituted for higher-level, text-based, manual annotations or keywords.
In inception of CBIR, many techniques have been developed along this direction and many retrieval systems, both research [1] and commercial, have been built.
Low-level features such as colors, textures and shapes of objects are widely used for CBIR. However, in specific applications, such as medical imaging, low-level features play a substantial role in defining the content of the data[7].
A typical content based image retrieval system is depicted in Figure 1.
There is a growing interest in CBIR because of the limitations inherent in metadata-based systems, as well as the large range of possible uses for efficient image retrieval. Textual information about images can be easily searched using existing technology, but requires humans to personally describe every image in the database.

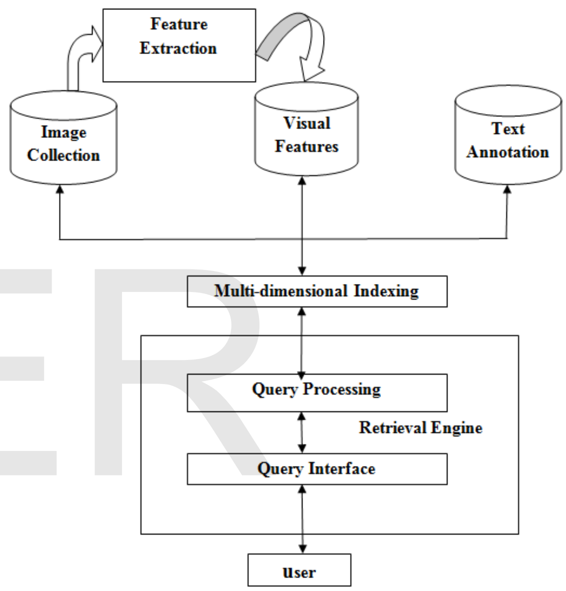

Fig 1 An Image Retrieval System Architecture

4 IMAGE COLLECTION VISUAL CONTENT REPRESENTATION USING BOF

The BOF framework is an adaptation of the bag-of-words scheme used for text categorization and text retrieval.

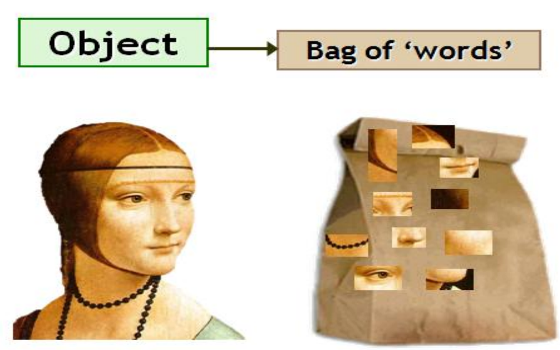

Fig 2 Bag Of Representation

The key idea is the construction of a codebook, i.e., a visual vocabulary in which the most representative patterns are codified as code words or visual words. Then, the image representation as BOF[10] is a histogram generated through a simple frequency analysis of each codeword inside the image.

The four steps to classify images using a BOF representation: (1) feature extraction and representation, (2) codebook construction, (3) the BOF representation of images, and, finally, (4) training of learning algorithms.

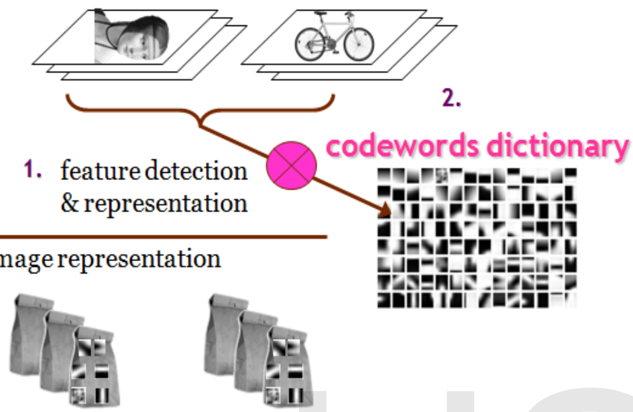

Fig 3 Overview Of The Bof Approach

4.1 Feature Extraction and Representation

The BOF approach starts extracting small blocks (in the present work, 8×8 pixels) from each image in the collection. There are two main alternatives for block extraction, partition of the image by a regular grid or extraction of blocks on salient points.

The regular-grid-based extraction is used; this process take into account a large quantity of blocks, but reduces the probability of missing interesting patterns. Each extracted block must be represented by a set of features.

Three different strategies that have produced good results when used in conjunction with the BOF representation .

The first strategy uses the raw block, i.e., the feature vector has 64 values corresponding to the luminance values of the corresponding pixels .

The second block-representation strategy is based on scale invariant feature transform (SIFT) points . This strategy uses a key-point detector based on the identification of interesting points in the location-scale space. This is implemented efficiently by processing a series of difference-of-Gaussian images.

The final stage of this algorithm calculates a rotation invariant descriptor using predefined orientations over a set of blocks. SIFT points are used with the most common parameter configuration: 8 orientationsand 4×4 blocks of cells, resulting in a descriptor of 128 dimensions. The SIFT algorithm has demonstrated to be a robust key-point descriptor in different image retrieval and matching applications, since it is invariant to common image transformations, illumination changes and noise.

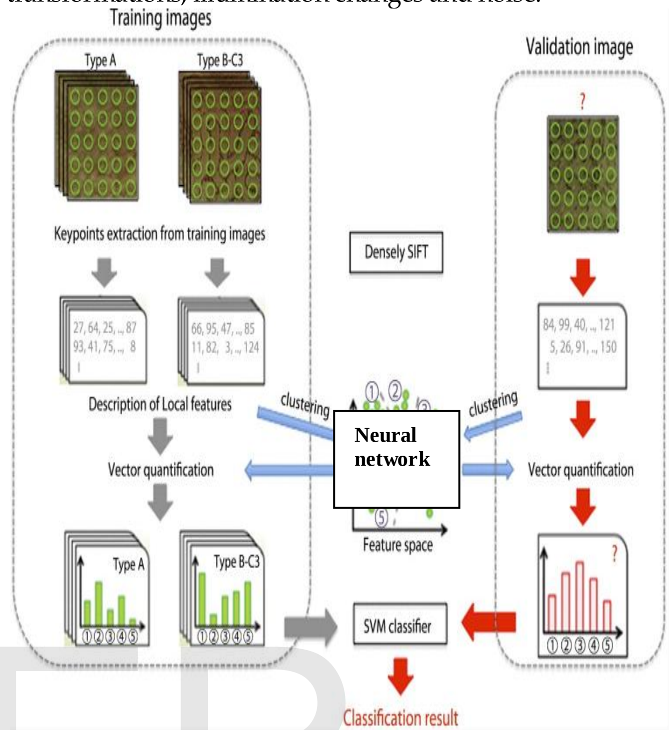

Fig 4 Architecture Of Sift

Finally, the third strategy is the discrete cosine transform (DCT) [41] applied to each channel of the RGB color space by block. The descriptor is built merging the 64 coefficients from each one of the three channels. This strategy generates a visual word that takes into account color and texture information from local features.

4.2 Code Book Construction

The visual dictionary or codebook is built using a clustering or vector quantization algorithm applied to the set of block descriptors extracted from the image collection. All local features, over a training image set, are brought together independently of the source image and are clustered to learn a set of representative visual words from the whole collection.

The k-means algorithm is used in this work to find a set of centroids that correspond to the code words. clustering algorithm has not a big impact in the classification of natural images compared with a random selection of code words. However, this is not necessarily the case for histology images .An important decision in the construction of the codebook is the size selection, that is, how many code words are needed to represent image contents. According to different works on natural image classification, the larger the codebook size, the better. However found that the size of the codebook is not a significant aspect in a medical image classification task..

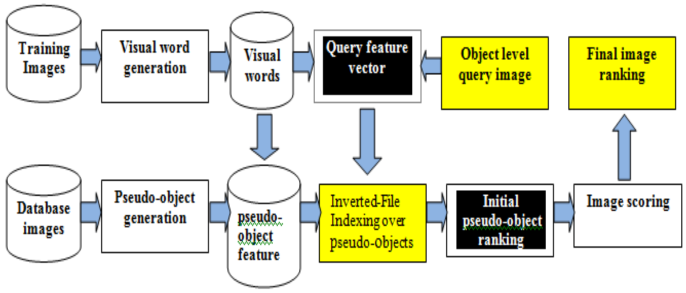

Fig 5 Architecture Of Codebook

4.2.1 VECTOR QUANTIZATION

Vector quantization is a classical quantization technique from signal processing which allows the modeling of probability density functions by the distribution of prototype vectors. It was originally used for data compression. It works by dividing a large set of points (vectors) into groups having approximately the same number of points closest to them. Each group is represented by its centroid point, as in k-means and some other clustering algorithms.

The density matching property of vector quantization is powerful, especially for identifying the density of large and high-dimensioned data. Since data points are represented by the index of their closest centroid, commonly occurring data have low error, and rare data high error. This is why VQ is suitable for lossy data compression. It can also be used for lossy data correction and density estimation.

4.2.2 K-Mean Clustering

In data mining, k-means clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. This results in a partitioning of the data space into Voronoi cells. The problem is computationally difficult (NP-hard), however there are efficient heuristic algorithms that are commonly employed and converge quickly to a local optimum.

1. To minimize sum of squared Euclidean distances between points xi and their nearest cluster centers mk
   Algorithm:
1. Randomly initialize K cluster centers
2. Iterate until convergence
3. Assign each data point to the nearest center
4. Recompute each cluster center as the mean of all points

$$D(X, M) = \sum_{\text{Cluster } k} \sum_{\substack{\text{point } i \\ \text{In cluster } k}} (xi - mk)2$$

These are usually similar to the expectation-maximization algorithm for mixtures of Gaussian distributions via an iterative refinement approach employed by both algorithms. Additionally, they both use cluster centers to model the data, however k-means clustering tends to find clusters of comparable spatial extent, while the expectation-maximization mechanism allows clusters to have different shapes.

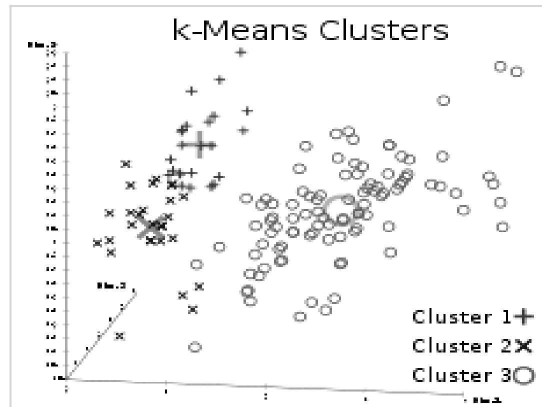

Fig 6 K-Mean Representation

4.3 FEATURE SELECTION AND ANALYSIS

Feature selection, also known as variable selection, feature reduction, attribute selection or variables/Subset selection, is the technique of selecting a subset of relevant features for building robust learning models. Feature selection is a particularly important step in analyzing the data from many experimental techniques in biology, such as DNA microarrays, because they often entail a large number of measured variables (features) but a very low number of samples. By removing most irrelevant and redundant features from the data.

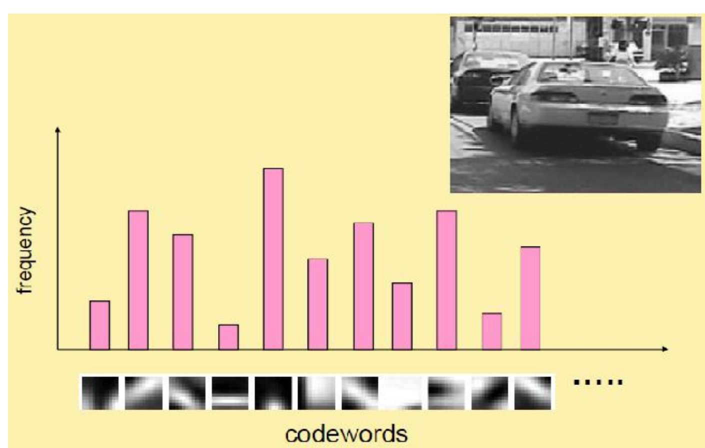

Fig 7 Image Representation

Feature selection helps improve the performance of learning models by:

1. Alleviating the effect of the curse of dimensionality.
2. Enhancing generalization capability.
3. Speeding up learning process.

5 VISUAL PATTERN MINING

The main goal of data mining is to extract useful knowledge from large data bases. This knowledge is usually represented in terms of interesting patterns that uncover hidden, unexpected, and/or interesting relationships among data items. The main goal of the system is to find visual patterns that can be associated with the high-level annotations.
The two analysis can be used in visual mining they are

1. Visual word discrimination analysis
2. Biclustering analysis

5.1 VISUAL WORD DISCRIMINATION ANALYSIS

Visual discrimination analysis used for good representatives of particular classes, i.e., codewords with a high discriminative power. In the general this process is known as feature selection.There are different approaches to perform feature selection, one popular strategy is to choose those features (in this case, codewords) that have a high correlation or dependence with a particular class[50]. This approach is called maximum relevance feature selection. Mutual information (MI) is a popular approximation to measure feature relevance.

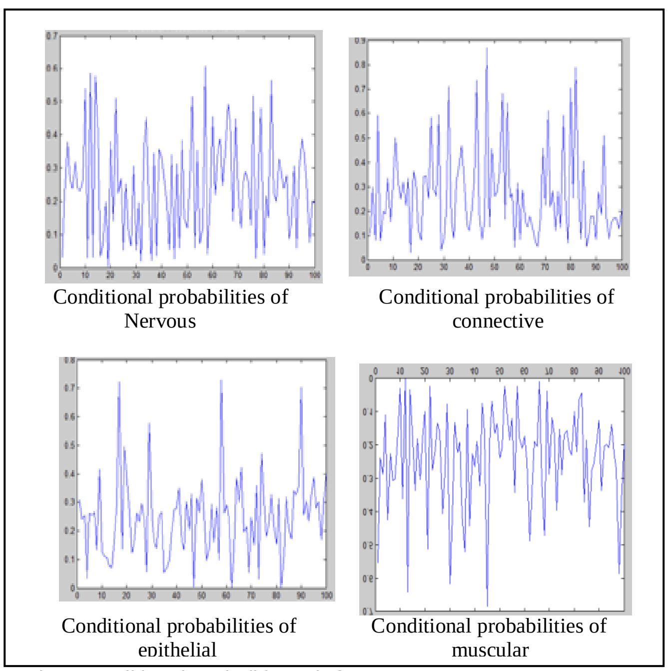
Conditional probabilities of Nervous
Conditional probabilities of connective
Conditional probabilities of epithelial
Conditional probabilities of muscular

Fig 8 Conditional Probalities Of Tissues

6 BICLUSTERING ANALYSIS

Biclustering (or coclustering) analysis is a data mining technique which allows simultaneous clustering by rows and columns of a data matrix. This method, with its respective graphic representation of data, is commonly applied in bioinformatics for gene expression analysis [52]. we propose to apply biclustering to histology image analysis using the following approach: images are analogous to samples (or conditions) and visual words are analogous to genes.
The data matrix is calculated using only the set of most discriminative visual codewords generated by the mRMR feature selection method described in the previous subsection. The main goal is to find biclusters that relate sets of images, which are conceptually connected, with sets of codewords. This goal translates into finding biclusters with high constant values.

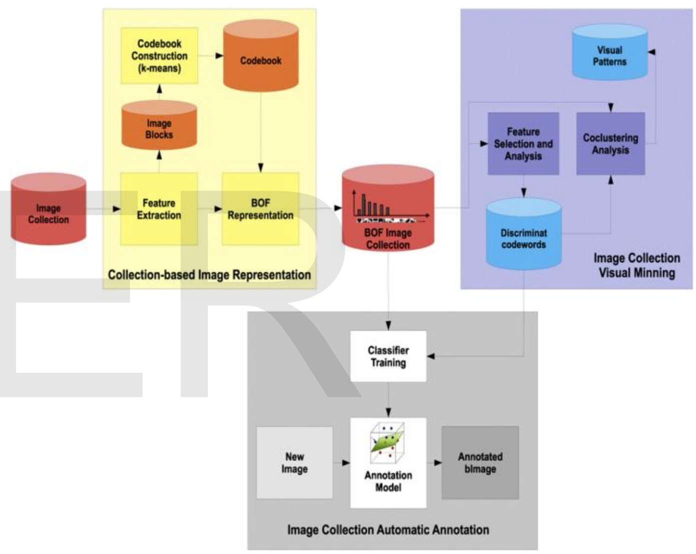

Fig 9 Architecture Diagram

7 CLASSIFIER TRAINING

In image classification, an image is classified according to its visual content. For example, does it contain an airplane or not.An important application is image retrieval - searching through an image dataset to obtain (or retrieve) those images with particular visual content. Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. For example, a classification model could be used to identify loan applicants as low, medium, or high credit risks.
A classification[19] task begins with a data set in which the class assignments are known. For example, a classification model that predicts credit risk could be developed based on observed data for many loan applicants over a period of time.

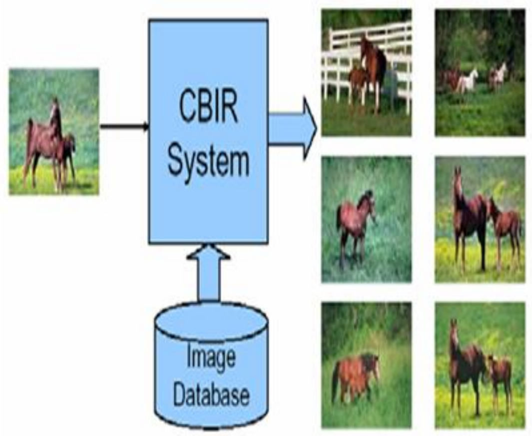

Fig 10 Content Based Retrival

Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. For example, a classification model could be used to identify loan applicants as low, medium, or high credit risks.

A classification task begins with a data set in which the class assignments are known. For example, a classification model that predicts credit risk could be developed based on observed data for many loan applicants over a period of time.

The simplest type of classification problem is binary classification. In binary classification, the target attribute has only two possible values: for example, high credit rating or low credit rating. Multiclass targets have more than two values: for example, low, medium, high, or unknown credit rating.

In the model build (training) process, a classification algorithm finds relationships between the values of the predictors and the values of the target. Different classification algorithms use different techniques for finding relationships. These relationships are summarized in a model, which can then be applied to a different data set in which the class assignments are unknown.

A neural network used as an classifier,neural network consists of units (neurons), arranged in layers, which convert an input vector into some output. Each unit takes an input, applies a (often nonlinear) function to it and then passes the output on to the next layer. Generally the networks are defined to be feed-forward: a unit feeds its output to all the units on the next layer, but there is no feedback to the previous layer. Weightings are applied to the signals passing from one unit to another, and it is these weightings which are tuned in the training phase to adapt a neural network to the particular problem at hand. This is the learning phase.

8 ANNOTATION

Automatic image annotation (also known as automatic image tagging or linguistic indexing) is the process by which a computer system automatically assigns metadata in the form of captioning or keywords to a digital image. This application of computer vision techniques is used in image retrieval systems to organize and locate images of interest from a database.

It can be regarded as a type of multi-class image classification with a very large number of classes - as large as the vocabulary size. Typically, image analysis in the form of extracted feature vectors and the training annotation words are used by machine learning techniques to attempt to automatically apply annotations to new images.

The first methods learned the correlations between image features and training annotations, then techniques were developed using machine translation to try to translate the textual vocabulary with the 'visual vocabulary', or clustered regions known as blobs.

9 HISTOLOGY IMAGE DATA SETS

Histology is a fundamental area of biology that studies theanatomy of cells and tissues at the microscopic level in both plants and animals. The main tool for histology is the microscope (light or electron) that is used to examine thin tissue sections. Histology and histopathology2 images are of great importance for medicine. They are a fundamental asset to determine the normality of a particular biological structure or to diagnose diseases like cancer. Histology courses are designed to train physicians in order to learn different tissue appearances, which vary according to the structure, function and cell organization in different organs of the body. These characteristics are usually highlighted with the help of different types of stains. Histology images are used both for fundamental biological research and for clinical decision making.

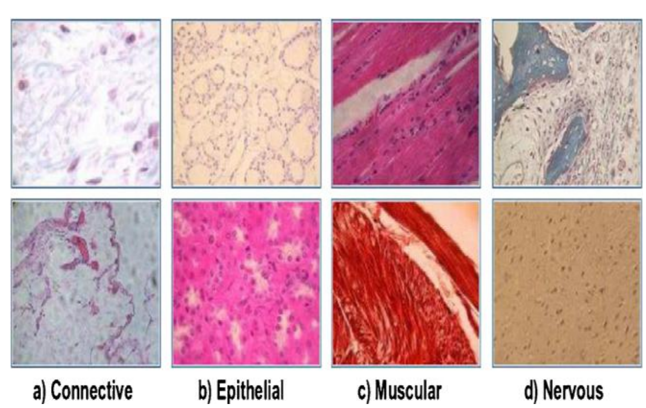

Fig 11 Sample Histology Image

9.1 FUNDAMENTAL TISSUES DATA SET

This data set comprises images from different organs that are representative of the four fundamental tissues. The data set includes 2828 images annotated with a global description of the tissue type. The data set composition is as follows: 484 connective tissue images, 804 epithelial qtissue images, 514 muscular tissue images, and 1026 nervous tissue images. The images show the four tissues in different stains and at different magnifications and cuts

TABLE 1
TISSUE DATA SET

Concept	#Images
Connective	484
Epithelial	804
Muscular	514
Nervous	1026

10 CONCLUSION

A strategy automatically extract visual patterns from a histology image collection. The foundation of the method is a BOF representation that builds a codebook which gathers the building blocks that explain the visual content of the image collection. A state-of-the-art feature selection process is applied to find a set of discriminative codewords. The codewords are related to high-level concepts individually, using conditional probabilities, and collectively, using biclustering.
The method was evaluated using histology image data sets. Histology images are particularly difficult to analyze because of their high variability and complex visual structure. The method was able to successfully find visual patterns that could be related to high-level concepts. The experimental results also showed that the BOF representation is a valuable alternative for histology image representation. The main contribution of the paper does not relies on the individual methods by themselves, but on the overall perspective that focuses on the analysis of the image collection as a whole. This novel perspective allows to use methods, such as biclustering, that traditionally have not been applied to the image analysis problem.
It does not replace traditional biomedical image analysis methods, but complement them. For instance, the method for automatically detecting concept-related regions in images, can extend a conventional annotation method by equipping it with an explanatory capability

11 FUTURE ENHANCEMENT

The exploration of new representation alternatives which take into account structural and multiscale information in order to capture biological and magnification variability, application to other type of biomedical images, use of other data analysis methods as latent semantic analysis, and data fusion from different sources.

REFERENCES

[1] Muller H, Michoux N, Bandon D, Geissbuhler A . A review of content-base Image retrieval systems in medical applications-clinical benefits and future directions International journal of medical informatics 2004;73:1–23.
[2] Gonzalez FA, Romero E. Biomedical Image Analysis and Machine Learning Technologies:Application and Techniques, 1st ed., Information Science Reference - Imprint of: IGI Publishing, Hershey, PA; 2009. Ch. 1. From Biomedical Image Analysis to Biomedical Image Understanding Using Machine Learning.
[3] Bankman IN. Handbook of medical imaging: processing and analysis. 1st ed. San Diego, CA: Academic Press; 2000.
[4] Gonzalez FA, Romero E. Biomedical Image Analysis and Machine Learning Technologies:Applications and Techniques, 1st ed., Information Science Reference - Imprint of: IGI Publishing, Hershey, PA; 2009.
[5] Lewis DD. Naive (bayes) at forty: The independence assumption in information retrieval. In: Nédellec C, Rouveirol C, editors. Proceedings of ECML-98, 10thEuropean conference on machine learning, No. 1398. 1998. p. 4–15.
[6] Tommasi T, Orabona F, Caputo B. Clef2007 image annotation task: an svm based cue integration approach. In: Nardi A, Peters C, editors. Working notes for the cross-language retrieval in Image Collections 2007 Workshop. 2007.
[7] Zheng L, Wetzel AW, Gilbertson J, Becich MJ. Design and analysis of a contentbased pathology image retrieval system. IEEE Transactions on InformationTechnology in Biomedicine 2003;7(4):249–55.
[8] Bonnet N. Some trends in microscope image processing. Micron 2004;35(8):635–53.
[9] Loukas C. A survey on histological image analysis-based assessment of three major biological factors influencing radiotherapy: proliferation,hypoxia and vasculature. Computer Methods and Programs in Biomedicine 2004;74(3):183–99.
[10] Caicedo J, Cruz-Roa A, Gonzalez F. Histopathology image classification using bag of features and kernel functions. In: Combi C, Shahar Y, Abu-Hanna A, editors. Proceedings of the 12th conference on artificial intelligence in medicine: artificialintelligence in medicine, vol. 5651 of Lecture Notes in Computer Science.2009. p. 126–35.

[11] Cruz-Roa A, Caicedo JC, Gonzalez FA. Visual pattern analysis in histopathologyimages using bag of features. In: Bayro-Corrochano E, Eklundh J-O, editors.Proceedings of the 14th Iberoamerican conference on pattern recognition: progress in pattern recognition, image analysis, computer vision, and applications,Vol. 5856 of Lecture Notes in Computer Science. 2009. p. 521–8.

[12] Allalou A, van de Rijke FM, Tafrechi RJ, Raap AK, Wahlby C. Image based measurements of single cell mtdna mutation load. In: Ersboll B, Pedersen K, editors.Image analysis, vol. 4522 of Lecture Notes in Computer Science. 2007. p. 631–40.

[13] Diaz G, Gonzalez FA, Romero E. A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopicimages. Journal of Biomedical Informatics 2009;42(2):296–307.

[14] Diamond J, Anderson NH, Bartels PH, Montironi R, Hamilton PW. The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia. Human Pathology 2004;35(9):1121–31.

[15] Doyle S, Hwang M, Shah K, Madabhushi A, Feldman M, Tomaszeweski J.Automated grading of prostate cancer using architectural and textural imag features. In: 4th IEEE international symposium on biomedical imaging: nano to macro, 2007. ISBI 2007. 2007. p. 1284.–7.

[16] Kong J, Sertel O, Shimada H, Boyer KL, Saltz JH, Gurcan MN. Computer-aided evaluation of neuroblastoma on whole-slide histology images: classifying grade of neuroblastic differentiation. Pattern Recognition 2009;42(6):1080–92.

[17] Mosaliganti K, Janoos F, Irfanoglu O, Ridgway R, Machiraju R, Huang K, et a Tensor classification of n-point correlation function features for histology tissue segmentation. Medical image analysis 2009;13(1):156–66.

[18] Sertel O, Kong J, Catalyurek U, Lozanski G, Saltz J, Gurcan M Histo pathological image analysis using model-based intermediate representations ancolortexture follicular lymphoma grading. Journal of Signal Processing Systems 2009;55(1):169–83.

[19] Orlov N, Shamir L, Macura T, Johnston J, Eckley DM, Goldberg IG. Wnd-charm:multi-purpose image classification using compound image transforms. PatternRecognition Letters 2008;29(11):1684–93.

[20] Tang H, Hanka R, Ip H. Histological image retrieval based on semantic contentanalysis. IEEE Transactions on Information Technology in Biomedicine 2003;7(1):26–36.Naik J, Doyle S, Basavanally A, Ganesan S, Feldman MD, Tomaszewski JE, et al. A boosted distance metric: application to content based image retrieval and classification of digitized histopathology. SPIE Medical Imaging: Computer-AidedDiagnosis 2009;7260, 72603F1-12.

[21] Peng H. Bioimage informatics: a new area of engineering biology. Bioinformatics 2008;24(17):1827–36.

[22] Swedlow JR, Goldberg IG, Eliceiri KW. Bioimage informatics for experimental biology. Annual review of biophysics 2009;38(1):327–46.

[23] Swedlow JR, Eliceiri KW. Open source bioimage informatics for cell biology. Trends in Cell Biology 2009;19(11):656–60.

[24] Tommasi T, Orabona F, Caputo B. Discriminative cue integration for medical image annotation. Pattern Recognition 2008;29(15):1996–2002.

[25] Madabhushi A. Digital pathology image analysis: Opportunities and challenges(editorial).Imaging In Medicine 2009;1(1):7–10.

[26] Madabhushi A, Basavanhally A, Doyle S, Agner S, Lee G. Computer-aided prognosis: predicting patient and disease outcome via multi-modal image analysis.In: Proceedings of the 2010 IEEE international conference on biomedical imaging: from nano to macro, ISBI'10. 2010. p. 1415–8.

[27] Bosch A, Munoz X, Oliver A, Martí J. Modeling and classifying breast tissue density in mammograms. In: Proceedings of the 2006 IEEE Computer Society Conference on computer vision and pattern recognition, vol. 2 of CVPR '06. 2006. p. 1552–8.

[28] Avni U, Greenspan H, Sharon M, Konen E, Goldberger J. X-ray image categorization
and retrieval using patch-based visual words representation. In: ISBI'09: proceedings of the sixth IEEE international conference on symposium on biomedical imaging. 2009. p. 350–3. Bosch A, Munoz X, Martí R. Review: which is the best way to organize/classify images by content? Image and Vision Computing 2007;25:778–91.

[29] Hofmann T. Unsupervised learning by probabilistic latent semantic analysis.Machine Learning 2001;42:177–96.

[30] Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of Machine Learning Research 2003;3:993–1022.

[31] Rogers S, Girolami M, Campbell C, Breitling R. The latent process decomposition of cdna microarray data sets. IEEE/ACM Transactions on Computational Biology
and Bioinformatics 2005;2:143–56.

[32] Bicego M, Lovato P, Ferrarini A, Delledonne M. Biclustering of expression microarray data with topic models. In: Proceedings of the 2010 20th international conference on pattern recognition, vol. 0 of ICPR '10. 2010. p. 2728–31.

[33] Diaz G, Romero E. Histopathological image classification using stain component features on a plsa model. In: Bloch I, Cesar R, editors. Proceedings of the 15th Iberoamerican congress conference on progress in pattern recognition, image analysis, computer vision, and applications, vol. 6419 of CIARP'10. 2010.p. 55–62.

[34] Csurka G, Dance CR, Fan L, Willamowski J, Bray C. Visual categorization with bags of keypoints. In: ECCV international workshop on statistical learning in computer vision. 2004. p. 1–22.

[35] Nowak E, Jurie F, Triggs B. Sampling strategies for bag-of-features image classification.In: Leonardis A, Bischof H, Pinz A, editors. Computer vision – ECCV 2006, vol. 3954 of Lecture Notes in Computer Science. 2006. p. 490–503.

[36] Li J, Allinson NM. A comprehensive review of current local features for computer vision.Neurocomputing 2008;71(10–12):1771–87.

[37] Lowe DG. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 2004;60:91–110.

[38] Kamiya Y, Takahashi T, Ide I, Murase H. A multimodal constellation model for object category recognition. In: Huet B, Smeaton A, Mayer-Patel K, Avrithis Y, editors. Advances in multimedia modeling, vol. 5371 of Lecture Notes in Computer Science. 2009. p. 310–21.

[39] Deselaers T, Ferrari V. Global and efficient self-similarity for object classification and detection. In: IEEE computer society conference on computer vision and pattern recognition, CVPR 2010. 2010. p. 1633–40.

[40] Han J, Kamber M. Data mining: concepts and techniques. Morgan Kaufmann;2000.

[41] Hsu W, Lee ML, Zhang J. Image mining: trends and developments. Journal of Intelligent Information Systems 2002;19:7–23.

[42] Berlage T. Analyzing and mining image databases. Drug Discovery Today 2005;10(11):795802.

[43] Hsu W, Lee ML, Zhang J. Image mining: trends and developments. Journal of Intelligent Information Systems 2002;19:7–23.

[44] Berlage T. Analyzing and mining image databases. Drug Discovery Today 2005;10(11):795–802.
[45] Malik HH, Kender JR. Clustering web images using association rules, interestingness measures and hypergraph partitions. In: ICWE ’06: proceedings of the 6th international conference on Web engineering. 2006. p. 48–55.
[46] Ribeiro M, Balan A, Felipe J, Traina A, Traina C. Mining statistical association rules to select the most relevant medical image features. In: Zighed D, Tsumoto S, Ras Z, Hacid H, editors. Mining complex data, vol 165 of Studies in Computational Intelligence. 2009. p. 113–31.
[47] Smeulders AWM, Worring M, Santini S, Gupta A, Jain R. Content-based image retrieval at the end of the early years. IEEE Transactions on Pattern Analysis and Machine Intelligence 2000;22(12):1349–80.
[48] Zhou XS, Zillner S, Moeller M, Sintek M, Zhan Y, Krishnan A, et al. Semantics and cbir: a medical imaging perspective. In: Proceedings of the 2008 international conference on content-based image and video retrieval, CIVR ’08. 2008.p. 571–80.
[49] Tommasi T, Orabona F, Caputo B. Discriminative cue integration for medical image annotation. Pattern Recognition Letters 2008;29(15):1996–2002.
[50] Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of Machine Learning Research 2003;3:1157–82.
[51] Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transaction on Pattern Analysis and Machine Intelligence 2005;27:1226–38.
[52] Madeira SC, Oliveira AL. Biclustering algorithms for biological data analysis:a survey. IEEE Transactions on Computational Biology and Bioinformatics 2004;1(1):24–45.
[53] Cheng Y, Church GM, Biclustering of expression data. In: Bourne PE, Gribskov M, Altman RB, Jensen N, Hope DA, Lengauer T, et al., editors. Proceedings of the eighth international conference on intelligent systems for molecular biology,vol. 8. AAAI Press; 2000. p. 93–103.
[54] Prelic A, Bleuler S, Zimmermann P, Wille A, Bühlmann P, Gruissem W, et al. A systematic comparison and evaluation of biclustering methods for gene expression data. Bioinformatics 2006;22(9):1122–9.
[55] Caicedo JC, Izquierdo E. Combining low-level features for improved classification and retrieval of histology images. Transactions on Mass-Data Analysis ofImages and Signals 2010;2(1):68–82.
[56] Shawe-Taylor J, Cristianini N. Kernel methods for pattern analysis. Cambridge, UK: Cambridge University Press; 2004.
[57] Fletcher CDM. Diagnostic histopathology of tumors. Amsterdam: Elsevier Science; 2003.
[58] Wong CSM, Strange RC, Lear JT. Basal cell carcinoma. British Medical Journal 2003;327:794–8.
[59] Yang J, Jiang Y-G, Hauptmann AG, Ngo C-W. Evaluating bag-of-visual-words representations in scene classification. In: MIR ’07: proceedings of the international workshop on Workshop on multimedia information retrieval. 2007. p.197–206.
[60] Iakovidis DK, Pelekis N, Kotsifakos EE, Kopanakis I, Karanikas H, Theodoridis Y. A pattern similarity scheme for medical image retrieval. Information Technology in Biomedicine, IEEE Transactions 2009;13(4):442–50.

U.Ravindran received the B.E degree in Computer Science Engineering from Anna University in 2006.and the M.Tech degree in Computer Science Engineering from Bharath University in 2011.Currently, He is an Assistant Professor in the Computer Science Department , Arulmigu Meenakshi Amman College of Engg, Thiruvannamalai Dt, Near Kanchipuram, Anna University, India. Ph- 9840096489.
Email - ravindranudhya@gmail.com

Shakila.T is currently pursuing M.E degree in Computer Science Engineering ,Arulmigu Meenakshi Amman College Of Engg, Thiruvannamalai Dt, Near kanchipuram. Anna University, India, Ph- 9944994979.
E-mail- achunshaki@gmail.com