

Content-Based Lung Image Categorization by Metric Learning for Interstitial Lung Diseases

Geethu Gopal . G Department of Computer Science and Engineering Rajiv Gandhi Institute of Technology Kottayam

*Abstract:--*Content Based Image Retrieval (CBIR) systems retrieve lung images from that database which are similar to the query image. CBIR is the application of computer vision. That has been one on the most vivid research areas in the field of computer vision over the last 10 years. Instead of text based searching, CBIR efficiently retrieves images that are visually similar to query image. In CBIR query is given in the form of image. This paper aims to provide an efficient medical image data Retrieval in Lung Diseases.

Finding similar images or reference is one way to assist radiologist for differential diagnosis of Interstitial Lung Diseases (ILDs). Content Based Image Retrieval (CBIR) has been identified as an important research topic in this direction. This motivated us to design a special purpose CBIR system (Med-IR) for Interstitial Lung Diseases (ILDs), where the user can provide one interstitial disease pattern as input and the system will retrieve few most similar patterns available in the database. CBIR is an effective technique, which is appropriate for large-scale indexing, is adopted, extended and integrated to the proposed framework so as to achieve optimized search and retrieval of rich media content even from large database.

Keywords— computer-aided diagnosis (CAD) · content-based image retrieval (CBIR) · interstitial lung diseases · texture analysis · clinical workflows

I. INTRODUCTION

As the rapid advance of digital imaging technologies, the content-based image retrieval (CBIR) has became one of the most vivid research areas in computer vision. In the last several years, developing computer-aided detection and/or diagnosis (CAD) schemes that use CBIR to search for the clinically relevant and visually similar medical images (or regions) depicting suspicious lesions has also been attracting research interest. CBIR-based CAD schemes have potential to provide radiologists with "visual aid" and increase their confidence in accepting CAD-cued results in the decision making. The CAD performance and reliability depends on a number of factors including the optimization of lesion segmentation, feature selection. reference database size, computational efficiency, and relationship between the clinical relevance and visual similarity of the CAD results. By presenting and comparing a number of approaches commonly used in previous studies, this article identifies and discusses the optimal approaches in developing CBIR-based CAD schemes and assessing their performance.

Although preliminary studies have suggested that using CBIR based CAD schemes might improve radiologists performance and/or increase their confidence in the decision making, this technology is still in the early development stage. Much research work is needed before the CBIR-based CAD schemes can be accepted in the clinical practice. Interstitial lung diseases (ILDs) are a relatively heterogeneous group of around 150 illnesses with often very unspecific symptoms. The most complete imaging method for the characterization of ILDs is the high-resolution computed tomography (HRCT) of the chest but a correct interpretation of these images is difficult even for specialists as many diseases are rare and thus little experience exists. Moreover, interpreting HRCT images requires knowledge of the context defined by clinical data of the studied case. A computerized diagnostic aid tool based on HRCT images with associated medical data to retrieve similar cases of ILDs from a dedicated database can bring quick and precious information for example for emergency radiologists. The experience from a pilot project highlighted the need for detailed database containing high quality annotations in addition to clinical data.



A medical image plays a central role in patient diagnosis, therapy, surgical planning, medical reference, and training. With the recent boom in the availability of filmless radiology equipment, the management of digital medical images is receiving more and more attention. Picture Archiving and Communication Systems (PACS) have been successfully introduced in many hospitals and specialized clinics, providing quick access to screening exams and integrating the actors involved in the workflow. The radiological databases enterprise's originally built for storing digital images have evolved from simple storage servers of past exams, kept for legal reasons, to active and easily accessible repositories for research and decision support.

The two approaches commonly used for image retrieval are referred to simply as global-based image searches and region (or sub-image)-based image searches. An important distinction between these approaches is that global-based methods enables the whole image matching and consider the relevant image, while region based methods focus primarily on specifying a region and on retrieving a large number of images with similar objects. Both methods are useful for image retrieval, but are best suited to queries of different types. Searching by global distinction is the preferred approach in cases where the user provides a whole image for query (i.e. query-byexample). Content-based image retrieval (CBIR) is the application of computer vision techniques to the problem of digital image search in large databases. CBIR enables to retrieve the images from the databases [1, 2]. Medical images are usually fused, subject to high inconsistency and composed of different minor structures. So there is a necessity for feature extraction and classification of images for easy and efficient retrieval [3].

CBIR is an automatic retrieval of images generally based on some particular properties such as color composition, shape and texture [4, 5].Proposed method used Metric learning is a popular supervised learning method for CBIR, where the objective is to learn a metric, which organizes medical exams as closely as possible to subjective ratings on similarity of exams that were previously collected from medical doctors. Current work in medical image retrieval provides an effective result for medical diagnosis with the increased performance of CBIR to radiology practice in future. Feature database creation has often been treated as a preprocessing step for removing the unwanted distortions of X-ray images in large\ databases. The classification step aims to classify the X-ray images with the help of SVM classifier. The number of images produced per day in most modern hospitals followed an exponential growth during the last decade. The mature field of imaging physics brought a large variety of essential diagnosis tools to the clinicians [6]. As a consequence, the explosion of the quantity and variety of medical visual information has the undesirable effect to overwhelm the radiologists with image interpretation tasks. About 15 years ago, the digital form of medical images along with their standardized storage gave birth to a new domain at the crossing of computer vision and medicine: image-based computer-aided diagnosis (CAD). As a response to the imaging physics breakaway, image-based CAD systems have the potential to improve both the effectiveness and effciency of the radiologists [7,8]. Goal of CAD is to use computer vision to assist radiologists in focusing their attention on diagnostically interesting events [9]. The CAD system can be used as first reader in order to improve the radiologists' productivity and reduce reading fatigue [10,11]. Whereas the radiologists' ability to interpret visual information is likely to change based on the domain-specific experience, human factors and time of the day, computerized classification of lung tissue patterns is 100% reproducible and can be quantitative and not only qualitative. Primarily research in Content Based Image Retrieval has always focused on systems utilizing color and texture features .generic medical images, which existing the visual healthful or diseased features with concrete samples, are very important to help the training and diagnosis of doctors. Many physicians construct their medical image libraries freely, which store agent case samples collected in a long time with detailed disease background and evolvement information. In this paper we propose an medical image store and retrieval method based on different extracted features like image histogram analysis, extraction of color values from segmented image and logical shape detection of an medical image.

II. RELATED WORKS

In picture archiving and communication system (PACS), image information is retrieved by using limited text keyword in special fields in the image header (e.g. patient identifier). Content-based image retrieval (CBIR) has received significant attention in the literature as a promising technique to facilitate improved image management in PACS system [12, 13]. The Image Retrieval for Medical Applications (IRMA) project [12] aims to provide visually rich image management through CBIR techniques applied to medical images using intensity



distribution and texture measures taken globally over the entire image. This approach permits queries on a heterogeneous image collection and helps in identifying images that are similar with respect to global features e.g. all chest x-rays in the AP (Anterior-Posterior) view. The IRMA system lacks the ability for finding particular pathology that may be localized in particular regions within the image. In contrast, the Spine Pathology and Image Retrieval System (SPIRS) [14, 15, 16] provides localized vertebral shape-based CBIR methods for pathologically sensitive retrieval of digitized spine x-rays and associated person metadata. Image Map [17] is so far, the only existing medical image retrieval that considers how to handle multiple organs of interest and it is based on spatial similarity. Consequently, a problem caused by user subjectivity is likely to occur, and therefore, the retrieved image will represent an unexpected organ. ASSERT [18] (Automatic Search and Selection Engine with Retrieval Tools) is a content-based retrieval system focusing on the analysis of textures in high resolution Computed Tomography (CT) scan of the lung. In WebMIRS [19] system, the user manipulates GUI tools to create a query such as, "Search for all records for people over the age of 65 who reported chronic back pain. Return the age, race, sex and age at pain onset for these people." In response. the system return values for these four fields of all matching records along with a display of the associated xray images. Cervigram Finder system [20] operates on a subset of the cervigram database. To use this system, the user defines a query by marking a region of interest on an image through GUI.

SPIRS-IRMA [21] is a CBIR system is based on the merits of two already existing systems (SPIRS & IRMA). So there is a need of absolute error free, efficient and automatic CBMIR system which can really helpful in medical stream. In [22], Chandan Singh et all discussed that for An Effective Image Retrieval using the fusion of global and local transforms. based on Angular Radial Transform (ART) and Polar Hough Transform (PHT) techniques. In paper [23], Bikesh kr. Singh et all presents that for The Content Based Retrieval of X-ray images using fusion of Spectral Texture and Shape Descriptors with the efficiency rate of 94.28%. In [24], Hai Jin proposed ET all that for Content and Semantic Context Based Image Retrieval for Medical Image Grid based on CGSP (China Grid Supporting Platform) technique. In [25], Thomas M. Lehman ET all proposed a novel framework for Content-Based Image Retrieval in Medical Applications for Picture Archiving and Communication Systems (PACS) which provides the suitable quality for medical diagnosis. In [26], Soumya Dutta ET all discussed the work that for A Change Detection Algorithm for Medical Cell Images by using Adaptive Median Filter, Standard Deviation calculation for discontinuity measurement Threshold technique and Change Detection methods. In [27], Hamed shahbazi et all proposed a framework that for Content Based Multispectral Image Retrieval using PCA (Principal Component Analysis) with the combination of Local Histogram and GLCM techniques. In [28], Ch. Kavitha et all presents a work for Image Retrieval Based on Combined features of Image Sub-blocks by using Radiance Histogram(RH) and Co-Occurrence Matrix(MCM) Multispectral with Precision=79.5% and Recall=82.5%. In Paper [29], Vimina Е

R ET all proposed a novel work for an Image Retrieval System using Local and Global properties of Image Regions with Relevance Feedback by using HSV (Hue Saturation Value) histograms for colour space, GLCM for Texture feature and Edge Histogram Descriptors (EHD) for Shape features.

III. MATERIALS AND METHODS

To build a CBIR system for ILDs for differential diagnosis and self learning tool for budding radiologists. Query formation, visual content extraction (in form of a feature vector), and checking similarity between feature vector for the query image and those of the images stored in the database (for retrieving images similar to the query image by content) are the different steps of retrieval process. A software platform is designed to store the images and retrieve similar images against a query image on demand. A graphical user interface (GUI) has been developed to facilitate query formation, retrieval and display the retrieved images.

The methods for building the image–based diagnostic aid tool are constituted by several connected subtasks. First, requirements were defined in collaboration with a lung specialist and the radiologists. The state of the art on image– based diagnostic aid was studied to identify challenges and solutions proposed. Clinical parameters associated with interstitial lung diseases were selected by a medical doctor of the project, a lung specialist, and an emergency radiologists



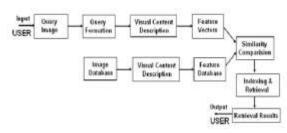


Fig.1.Block Diagram of CBIR System

In typical Content-based image retrieval systems (Fig.1), the visual contents of the images in the database are extracted and described by multi-dimensional feature vectors. The feature vectors of the images in the database form a feature database. To retrieve images, users provide the retrieval system with example images or sketched figures. The system then changes these examples into its internal representation of feature vectors. The similarities /distances between the feature vectors of the query example or sketch and those of the images in the database are then calculated and retrieval is performed with the aid of an indexing scheme. The image database.

A. State of Art in Reviewing System

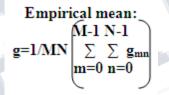
Reviewing medical CBIR systems is an often discussed issue with the first paper appeared in 1997 [30]. Thereafter, reviews usually are specialized on a certain medical or technological application domain such as forensics [31] or Web 2.0 [13], respectively. Rather general reviews have been published by Müller et al. [10] and Akgul et al. [23] referring to CBIR in radiology by current status, clinical benefits, and future directions. Ending up with 187 or 77 references from the two reviews, respectively, inclusion or exclusion criteria are not fully clear, which limits the impact of such work. However, a somewhat more systematic methodology for classifying CBIR systems has also been proposed.

B. Defining the characteristics of Medical CBIR Systems

In [20], for instance, a set of 14 so called gaps are defined to classify medical CBIR systems, which are enriched with additional 7 characteristics. The gaps were identified as being responsible for potential pitfalls and inadequacy of current medical CBIR systems. For instance, the "semantic gap" describes the discrepancies between a high-level of semantic in human image perception and understanding and the simple numerical signature that is extracted by a machine in terms of color, texture and shape.

C. Gray Scale Image Analysis

a) Columnar Mean: Columnar is one of the methodologies that used in CBIR to retrieve gray map images. In this method we take only gray scale image for image analysis. In this approach we calculate the average (Empirical Mean or Simply Mean) value of each column of the image (Because image is stored as a matrix using standard Matlab matrix convensions) and make those values as the index for that image and are stored in the database. While retrieving the image from the database based on the input image, we calculate mean value of each column of the input image and will compare these values with that stored in the database, if there is a match then we will retrieve those images.



b) Diagonal Mean: In this approach we calculate the Empirical mean value of the pixels that lies on the principle diagonal of the image and make that value as the index for that image and is stored in the databases. While retrieving the image from the database based on the input image, calculate mean value of diagonal elements of the input image and will compare these values with that stored in the database. If there is a match then those images are retrieved

c) Histogram Analysis: The Histogram is a summary graph showing a count of the data points falling in various ranges. The effect is a rough approximation of the frequent distribution of the data. The groups of data are called classes, and in the context of a histogram they are known as bins, because one can think of them as containers that accumulate data and fill up at a rate equal to the frequency of that data class.

d) Retrieving similar images using Euclidean

Distance: We use average RGB to calculate color similarity Average RGB is to compute the average value in R, G, and B channel of each pixel in an image, and use this descriptor of an image for comparison purpose. Formula used for calculating Euclidean Distance is as follows:



$$d(I_{a},I_{b}) = \left(\frac{(r_{a}-r_{b})^{2}+(g_{a}-g_{b})^{2}+(b_{a}-b_{b})^{2}}{3}\right)^{1/2}$$
(2)

In this method we calculate the distance between the query image and candidate set images stored in the database, if the distance is within the already fixed threshold then we will retrieve those images as the similar images to the query images.

IV. EXPERIMENTAL EVALUATION

This section compares our approach with the traditional CBIR approaches, by describing and presenting the results for two experiments. Medical CBIR evaluation involves the definition of the following requirements: 1) user needs (or topics) expressed as undiagnosed exams; 2) a collection of documents, in this case PACS exams, from which the CBIR system retrieves the relevant examples; 3) a relevance assessment, which maps the relevance of each of the collection samples to each of the query exams. Performance is evaluated by analyzing the retrieved documents to each topic according to the relevance assessment. Systems can then be ranked by a performance measure that reflects how appropriate the retrieval system was for all topics.

A comparative assessment of the performances of the reviewed medical CBIR systems is not possible as not only their application domains (imaging modalities that the systems are built for) differ, but also there is a lack of common database to evaluate different systems. The Image CLEF med, is one of the few (if not the single) platform to evaluate and compare different systems. The IRMA [29] the med GIFT [30] and the Vis Me [28] projects are participants of Image CLEF med. The evaluation protocols are largely influenced by the tasks addressed by the IRMA system that mainly targets modality and body part similarity-based retrieval(<u>http://www.imageclef</u>. org/: http://ir.ohsu.edu/image/; http://www.clefcampaign.org/). Objective evaluation is even more challenging for the retrieval systems based on patient/case similarity as there is currently no consensus on how to rate the case similarity even manually. Consequently, task specific evaluation platforms are required rather than a generic approach.

Performance of the system was analyzed using precision and recall method. The system achieved high performance compared to existing system. So that the performance of the retrieval process has been improved by comparing the classified images with user's query image medical images Standard formulas have been computed for determining the precision and recall measures. Table I which shows Precision and Recall values in %. And also Table II shows some Parameters to evaluate the performance of the system.

Query Image	Precision	Recall
Imagel	90	10
Image2	85	20
Image3	82	25
Image4	80	28
Image5	75	42
Image6	70	58
Image7	68	62
Image8	64	60
Imageo	04	00

Precision (P) is the ratio of the relevant images to the total number of images retrieved

P = r/n1 (3)

Where,

r - number of relevant images retrieved.

n1 - total number of images retrieved Recall(R) is the percentage of relevant images among all possible relevant images

R = r/n2(4)

Where,

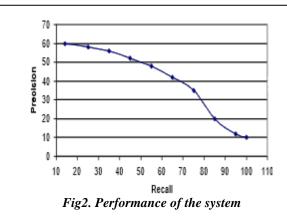
r - number of relevant images retrieved n2 - total number of relevant images in the database.

system		
Parameters	Formulae	
Accuracy	(TP + TN)/N	
Sensitivity	TP / P	
Specificity	TN / N	
Precision	TP/(TP+FP)	
F-measure	2*((p*r)/(p+r))	

Table II. Parameters to Evaluate performance of the system

According to this the performance of our system has demonstrated in Fig 2.





V. CONCLUSION

In this work, the clinical workflows and GUIs of a web-based hybrid detection-CBIR-based CAD system to assist the diagnosis of ILDs are proposed. Most of the systems found in the literature are proposing computeraided detection of the abnormal lung tissue, letting the radiologists down with the problem of the final histological diagnosis. The implemented clinical workflows show that detection-based and CBIR-based CAD are complementary both on the user's side and on the algorithmic side. The next steps are the evaluation of the system in clinical routine to assess the performance of the radiologists with and without the system. In an effort to improve CBIR medical representations, this paper studied the use of radiology reports to supervise CBIR systems. The presented method uses text distances between exam reports to supervise a metric learning algorithm in the exam image space. We compared our approach with traditional CBIR systems, based on visual representations and on expert annotations using a database of ILD CT.

In both cases, our approach consistently increased CBIR performance for the tested image descriptions. Since radiology reports are normally available in all hospital PACS, and since results suggest it is beneficial for a broad range of CBIR configurations and image descriptions, our approach can be applied to a variety of image retrieval applications and, hence, contribute to the introduction of CBIR technology into a clinical context. In the future enhancements we can implement Shape and Texture analysis, Color Image Histogram, Image ranking in Euclidean Distance Method.

REFERENCES

[1] C.B. Akgul, et al., "Content-Based Image Retrieval in Radiology: Current Status and Future Directions", Journal of Digital Imaging, Vol. 24, No. 2, pp. 208-222, 2011.

[2] V.S. Murthy, E.Vamsidhar, J.N.V.R. Swarup Kumar, and P. Sankara Rao, "Content based Image Retrieval using Hierarchical and Kmeans Clustering Techniques", International Journal of Engineering Science and Technology, Vol. 2, No. 3, 2010, pp. 209-212.

[3] B. Ramamurthy, and K.R. Chandran, "CBMIR:Shapebased Image Retrieval using Canny Edge Detection and Kmeans Clustering Algorithms for Medical Images", International Journal of Engineering Science and Technology, Vol. 3, No. 3, 2011, pp. 209-212.

[4] Roberto Parades, Daniel Keysers, Thomas M. Lehman, Berthold Wein, Herman Ney, and Enrique Vidal, "Classification of Medical Images Using Local Representation", Workshop Bildverarbeitung fur die Medizin, 2002, pp.171-174.

[5] Wei Zhang, Sven Dickinson, Stanley Sclaroff, Jacob Feldman, and Stanley Dunn, "Shape – Based Indexing in a Medical Image Database", Biomedical Image Analysis, 1998, pp. 221230.

[6] Simel, D., Drummond, R.: The rational clinical examination: evidence– based clinical diagnosis. McGraw–Hill (2008)

[7] Doi, K.: Computer–aided diagnosis in medical imaging: Historical review, current status and future potential. Computerized Medical Imaging and Graphics 31(4–5), 198–211 (2007)

[8] Duncan, J.S., Ayache, N.: Medical image analysis: Progress over two decades and the challenges ahead. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(1), 85–106 (2000)

[9] Engle, R.L.: Attempts to use computers as diagnostic aids in medical decision making: a thirty-year experience. Perspectives in biology and medicine 35(2), 207–219 (1992)

[10] Mu[°]ller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content–based image retrieval systems in medicine–clinical benefits and future directions.



Vol 3, Issue 9, September 2016

Internation Journal of Medical Informatics 73(1), 1–23 (2004)

[11] Nishikawa, R.M.: Current status and future directions of computer–aided diagnosis in mammography. Computerized Medical Imaging and Graphics 31(4–5), 224–235 (2007)

[12] Engle, R.L.: Attempts to use computers as diagnostic aids in medical decision making: a thirty-year experience. Perspectives in biology and medicine 35(2), 207–219 (1992).

[13] Y. Liu, D. Zhang, G. Lu, W.Y. Ma, "A Survey of Content Based Image Retrieval with High Level Semantics", Pattern Recogn., Vol. 40, pp. 262-82, 2007.

[14] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler," A review of content-based image retrieval systems in medical applications- Clinical benefits and future directions", International Journal of Medical Informatics, Vol. 73, No. 1, 2004, pp. 1-23.

[15] T.M. Lehmann, M.O. Guld, C Thies, B Fischer, K. Spitzer, and D. Keysers," Content-based image retrieval in medical applications", Methods of Info in Med, IOS Press, Vol. 43, No. 4, 2004, pp. 354–361.

[16] S. Antani, L.R. Long, and G.R. Thoma, "Contentbased image retrieval for large biomedical image Archives", Proceedings of 11th World Congress Medical Informatics, 2004, pp. 829–833

[17] L.R. Long, S.K. Antani, and G.R. Thoma, "Image informatics at a national research center", Computer Medical Imaging & Graphics (ELSEVIER), Vol. 29, 2005, pp. 171–193

[18] G.R. Thoma, L.R. Long, and S.K. Antani, "Biomedical imaging research and development: knowledge from images in the medical enterprise", Technical Report Lister Hill National Centre for Biomedical Communications, 2006.

[19] E.G.M. Petrakis, and C. Faloutsos, "ImageMap: An Image Indexing Method Based on Spatial Similarity", IEEE Transaction on Knowledge and Data Engineering, 2002, pp. 979–987 [20] Chi-Ren Shyu, Carla E. Brodley, Avinash C. Kak, and Akio Kosaka," ASSERT:A Physician-in-the-Loop Content-Based Retrieval System for HRCT Image Databases", Computer Vision and Image Understanding, Vol. 75, No. 1, 1999, pp. 111–132.

[21] L.R. Long, S.R. Pillemer, R.C. Lawrence, G- H Goh, L. Neve, and G.R. Thoma, "WebMIRS: Web-based Medical Information Retrieval System", Proceedings of SPIE Storage and Retrieval for Image and Video Databases VI, SPIE, Vol. 3312, 1998, pp. 392-403.

[22] Z. Xue, L.R. Long, S. Antani, J. Jeronimo, and G.R. Thoma, "A Webaccessible content-based cervicographic image retrieval system", Proceedings of SPIE medical imaging, Vol. 6919, 2008, pp. 1-9.

[23] S.K. Antani, T.M. Deserno, L.R. Long, M.O. Guld, L. Neve, and G.R. Thoma, "Interfacing global and local CBIR systems for medical image retrieval", Proceedings of the workshop on Medical Imaging Research, 2007, pp.166171.

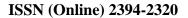
[24] Chandan Singh*, and Pooja, "An effective image retrieval using the fusion of global and local transforms based features", Optics & Laser Technology 44 (2012) 2249-2259.

[25] Bikesh Kr. Singh, G. R Sinha, Bidyut Mazumar, and Md. Imrose Khan, "Content Based Retrieval of X-ray Images Using Fusion of Spectral Texture and Shape Descriptors", 2010 International Conference on Advances in Recent Technologies in Communication and Computing 9789-0-7695-4201-0/10DOI 10.1.1109/ARTCom.2010.51

[26] Hai Jin, Aobing Sun, Ran Zheng, Ruhan He, Qin Zhang, Yingjie Shi, and Wen Yang, "Content and Semantic Context Based Image Retrieval for Medical Image Grid", National High Technology Research and Development Program of China, No.2006AA02Z347 and No.2006AA01A115

[27] Soumya Dutta, Dr. Madhurima Chattopadhyay, "A Change Detection Algorithm for Medical Cell Images", International Journal on Computer Science and Engineering (IJCSE), Feb 2011.

[28] Ch. Kavitha, Dr. B. Prabhakara Roa, and Dr. A. Govandhan, "Image Retrieval based on combined features of image subblocks", Ch. Kavitha et al. / International Journal on Computer Science and Engineering (IJCSE),





connecting engineers...developing research

ISSN: 0975-3397, Vol. 3 No. 4 Apr 2011

[29] Keysers D, et al: Statistical framework for modelbased image retrieval in medical applications. J Electron Imaging 12 (1):59–68, 2003.

[30] Müller H, et al: The Use of MedGIFT and EasyIR for ImageCLEF 2005. in Accessing Multilingual Information Repositories. 2005: Springer LNCS 4022.

[31] Cauvin JM, et al: Computer-assisted diagnosis system in digestive endoscopy. IEEE Trans Inf Technol Biomed 7 (4):256–262, 2003.