Volume 5, Issue 5, May 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

# Image Retrieval based on Image Filtering & Similarity fusion using Classification & Relevance Feedback

**Syed Tanzeem*[1], Dr. M.A.PundSurayya Syed[2]**
[1]IT Dept, Theem College of Engg, Boisar, India
[2]Computer Sci & Engg DeptComp Sci & Engg Dept., P.R.M.I.T & RMGM COE,
Badnera-Amravati, India

***Abstract*—In this paper an attempt is made using, the image category information utilized directly to filter out irrelevant images and adjust the feature weights in a linear combination of similarity matching. Using the RF-based technique to update the feature weights based on positive user feedback. Retrieval performance is promising and clearly showing the advantage of searching images based on similarity fusion and filtering in terms of effectiveness and efficiency. Overall, this retrieval framework is useful as a front end for large medical databases where a search can be performed in diverse images for teaching, training and research purposes. Image retrieval also known as query by image content and content-based visual information retrieval 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 image retrieval is opposed to concept based approaches .IR 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. Currently very few methods are available for retrieving images based on similarity and filtering but by implementing this project for biomedical images using SVM classification & relevance feedback we can make searching of similar images quickly & perfectly.***

***Keywords*—Image retrieval,Classification,RF-ranking,SVM,Relevance Feedback.**

## I. INTRODUCTION

AllContent-based image retrieval (CBIR) has attracted much research interest in recent years [1]. In particular, there has been growing interest in indexing biomedical images by content [2, 3, 4, 5]. Manual indexing of images for content-based retrieval is cumbersome, error prone, and prohibitively expensive. Due to the lack of effective automated methods, however, biomedical images are typically annotated manually and retrieved using a text keyword-based search. A common drawback of such systems is that the annotations are imprecise with reference to image feature locations, and text is often insufficient in enabling efficient image retrieval. Even such retrieval is impossible for collections of images that have not been annotated or indexed.

Additionally, the retrieval of interesting cases, especially for medical education or building atlases, is a cumbersome task. CBIR methods developed specifically for biomedical images could offer a solution to such problems, thereby augmenting the clinical, research, and educational aspects of biomedicine. For any class of biomedical images, however, it would be necessary to develop suitable feature representation and similarity algorithms that capture the “content” in the image. The Lister Hill National Center for Biomedical Communications, a research and development division of the U.S. National Library of Medicine (NLM), maintains a digital archive of 17,000 cervical and lumbar spine images collected in the second National Health and Nutrition Examination Survey (NHANES II) conducted by the National Center for Health Statistics (NCHS). Classification of the spine x-ray images for the osteoarthritis research community has been a long-standing goal of researchers at the NLM, and collaborators at NCHS and the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS). Also, the capability to retrieve these images based on geometric characteristics of the vertebral structures is of interest to the vertebral morphometry community. Medical experts have identified visual features of the images specifically related to osteoarthritis, but the images have never been manually indexed for these features which include anterior osteophytes, disc space narrowing for the cervical and lumbar spine, subluxation for the cervical spine, and spondylolisthesis for the lumbar spine. Another archive of 100,000 digitized 35mm color slides of the uterine cervix is being created in collaboration with the National Cancer Institute (NCI).

A classification-driven biomedical image retrieval framework based on image filtering and similarity fusion by employing supervised learning techniques. In this structure, the probabilistic outputs of a multiclass hold up vector machine (SVM) classifier as kind forecast of query and database images are browbeaten at first to filter out irrelevant images, thus reducing the explore space for parallel matching. By means of the introduction of imaging, clinical care could be considerably impacted with enhanced image handling. In current years, rapid advances of software and hardware technology have eased the trouble of maintaining big medical image collections. These images compose an imperative source of anatomical and functional in order for the diagnosis of diseases, medical research, and education. Effectively and efficiently searching in these large image collections poses significant technical challenges as the characteristics of the medical images differ significantly from other general purpose images. The image modality reveals anatomical and/or functional information of different body parts and pathologies. Every imaging modality exposed challenges for acquisition, storage, and retrieval. Presently, the images are retrieved mainly using text based searches. Search results in medical collections might be enhanced by combining text attribute-based search capacity with low-level visual features computed honestly on the image contented, usually identified as the content-based image recovery (CBIR).

\

## II. FEATURE CALCULATION

Image features are plays an important role in image retrieval processing.In oder to perform image retrieval process,the extraction of suitable features from the images are very important and by which both the query image and database images are compared to retrieval of very similar images to query image from the database.There are three levels of feature extraction global ,local and pixel values of the image.Local features are extracted from small subimages from the original image.The global features can be extracted to describe the whole image in an average fashion.The low -level features extracted from images and their local patches are color,texture and shape.For feature extraction CLD and EHD descriptors are used.

### *A.COLOR LAYOUT DESCRIPTOR*

The CLD is a very compact and resolution-invariant representation of color for high-speed image retrieval and it has been designed to efficiently represent the spatial distribution of colors. This feature can be used for a wide variety of similarity-based retrieval, content filtering and visualization. It is especially useful for spatial structure-based retrieval applications. This descriptor is obtained by applying the discrete cosine transform (DCT) transformation on a 2-D array of local representative colors in Y or Cb or Cr color space. The main functions of the CLD are basically Image-to-image matching or video clip-to-video clip matching [7].The extraction process of the CLD is represented as

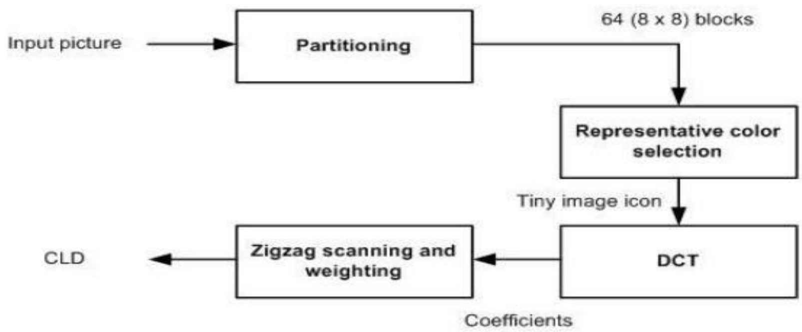


Fig. Extraction process of CLD

### *B.EDGE HISTOGRAM DESCRIPTOR*

The edge histogram descriptor (EHD) represents the local edge distribution by dividing image space into 4×4 sub images and representing the local distribution of each sub image by a histogram. The fact that the EHD consists of the local-edge histograms only, makes it very flexible. In the sense of generating histograms, edges in all sub images are categorized into five types- vertical, horizontal, diagonal and non directional edges (namely edges with no particular directionality), resulting in a total of 5× 16 = 80 histogram bins. Each sub image is further divided into non overlapping square image blocks with particular size which depends on the image resolution. Each of the image blocks is then classified into one of the five mentioned edge categories or as a non edge block [8].

## III. SVM BASED CLASSIFICATION

BeforeMedical imaging is the technique and process used to create images of the hhman body (or parts and function thereof) for clinical purposes (medical procedures seeking to reveal, diagnose or examine disease) or medical science (including the study of normal anatomy and physiology). Although imaging of removed organs and tissues can be performed for medical reasons, such procedures are not usually referred to as medical imaging, but rather are a part of pathology.

As a discipline and in its widest sense, it is part of biological imaging and incorporates radiology (in the wider sense), nuclear medicine, investigative radiological sciences, endoscopy, (medical) thermography, medical photography and microscopy (e.g. for human pathological investigations).

Measurement and recording techniques which are not primarily designed to produce images, such as electroencephalography (EEG), magneto encephalography (MEG), electrocardiography (EKG) and others, but which produce data susceptible to be represented as maps (i.e. containing positional information), can be seen as forms of medical imaging.

Up until 2010, 5 billion medical imaging studies had been conducted worldwide. Radiation exposure from medical imaging in 2006 made up about 50% of total ionizing radiation exposure in the United States. In the clinical context, "invisible light" medical imaging is generally equated to radiology or "clinical imaging" and the medical practitioner responsible for interpreting (and sometimes acquiring) the image is a radiologist. "Visible light" medical imaging involves digital video or still pictures that can be seen without special equipment. As a field of scientific investigation, medical imaging constitutes a sub-discipline of biomedical engineering, medical physics or medicine depending on the context: Research and development in the area of instrumentation, image acquisition (e.g. radiography), modeling and quantification are usually the preserve of biomedical engineering, medical physics and computer science; Research into the application and interpretation of medical images is usually the preserve of radiology and the medical sub-discipline relevant to medical condition or area of medical science (neuroscience, cardiology, psychiatry, psychology, etc.) under investigation. Many of the techniques developed for medical imaging also have scientific and industrial applications.

Medical imaging is often perceived to designate the set of techniques that noninvasively produce images of the internal aspect of the body. In this restricted sense, medical imaging can be seen as the solution of mathematical inverse problems. This means that cause (the properties of living tissue) is inferred from effect (the observed signal). In the case of ultrasonography the probe consists of ultrasonic pressure waves and echoes inside the tissue show the internal structure. In the case of projection radiography, the probe is X-ray radiation which is absorbed at different rates in different tissue types such as bone, muscle and fat.

### *A. Relevance feedback*

Relevance feedback is a feature of some information retrieval systems. The idea behind relevance feedback is to take the results that are initially returned from a given query and to use information about whether or not those results are relevant to perform a new query. We can usefully distinguish between three types of feedback: explicit feedback, implicit feedback, and blind or "pseudo" feedback.

### *B. Explicit feedback*

Explicit feedback is obtained from assessors of relevance indicating the relevance of a document retrieved for a query. This type of feedback is defined as explicit only when the assessors (or other users of a system) know that the feedback provided is interpreted as relevance judgments.

Users may indicate relevance explicitly using a binary or graded relevance system. Binary relevance feedback indicates that a document is either relevant or irrelevant for a given query. The relevance feedback information needs to be interpolated with the original query to improve retrieval performance, such as the well-known Rocchio Algorithm.

- *Implicit feedback (RF):* Implicit feedback is inferred from user behavior, such as noting which documents they do and do not select for viewing, the duration of time spent viewing a document, or page browsing or scrolling actions. The key differences of implicit relevance feedback from that of explicit include. The user is not assessing relevance for the benefit of the IR system, but only satisfying their own needs and The user is not necessarily informed that their behavior (selected documents) will be used as relevance feedback An example of this is the Surf Canyon browser extension which advances search results from later pages of the result set based on both user interaction (clicking an icon) and time spent viewing the page linked to in a search result.
- *Blind Feedback*: Pseudo relevance feedback, also known as blind relevance feedback, provides a method for automatic local analysis. It automates the manual part of relevance feedback, so that the user gets improved retrieval performance without an extended interaction. The method is to do normal retrieval to find an initial set of most relevant documents, to then assume that the top "k" ranked documents are relevant, and finally to do relevance feedback as before under this assumption.

### *C. Support vector machine*

A support vector machine (SVM) is a concept in statistics and computer science for a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. The standard SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the input, making the SVM a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

## IV. FUNDAMENTALS OF CONTENT-BASED IMAGE RETRIEVAL

Content-based image retrieval, a technique which uses visual contents to search images from large scale image databases according to users' interests, has been an active and fast advancing research area since the 1990s.Content-based image retrieval uses the visual contents of an image such as color, shape, texture, and spatial layout to represent and index the image. In typical content-based image retrieval systems (Figure 1-1), 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 retrieveimages, 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 indexing scheme provides an efficient way to search for the image database. Recent retrieval systems have incorporated users' relevance feedback to modify the retrieval process in order to generate perceptually and semantically more meaningful retrieval results. In this chapter, we introduce these fundamental techniques for content-based image retrieval.

 

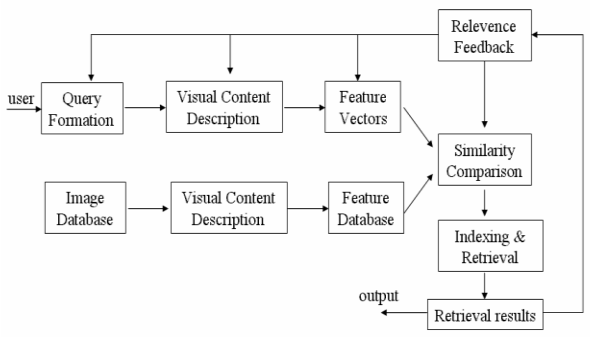


Fig. 1. Content-based image retrieval system.

## V. ALGORITHM

1. Initialize the user authentication
2. Initialize the data base
3. Select the i/p image from data base
4. Create a group of image in data base
5. Store the images in different groups.
6. Perform feature calculation
   i) Divide the relevant group image in to 8x8
   ii) Calculate edge histogram
7. Perform SVM
8. Store the image in to Data base
9. Select the query image
10. Perform feature calculation same as step G
11. Perform initial relevance feedback
12. Perform dynamic fusion
13. Perform RF ranking
14. Store all similar images
15. Compare database image & query image calculate matching score

## VI. RESULTS

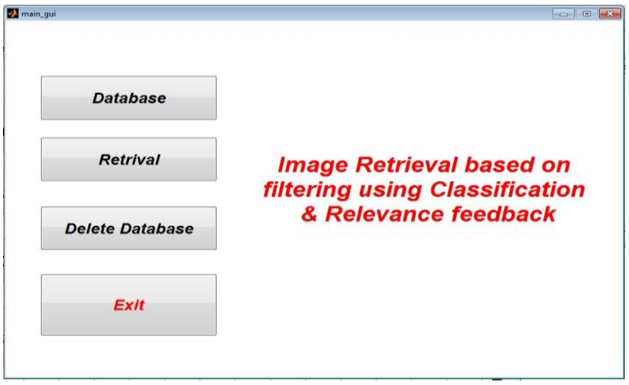


Fig 1. Main Page

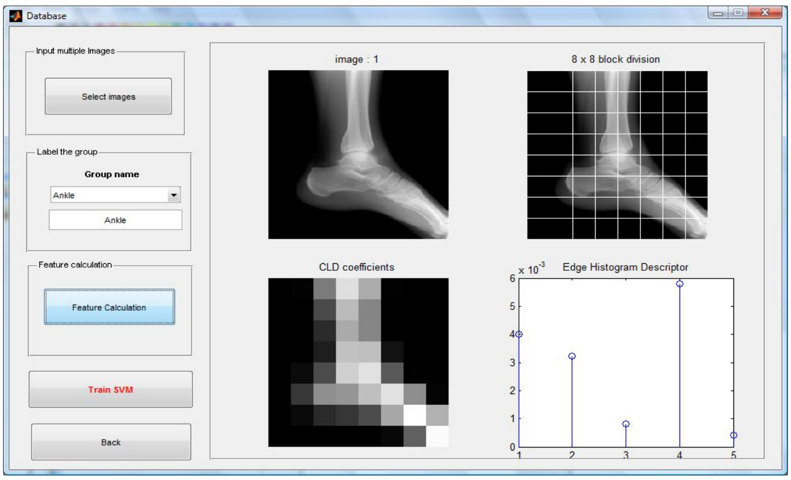


Fig.2 Feature Calculation after click on database

 

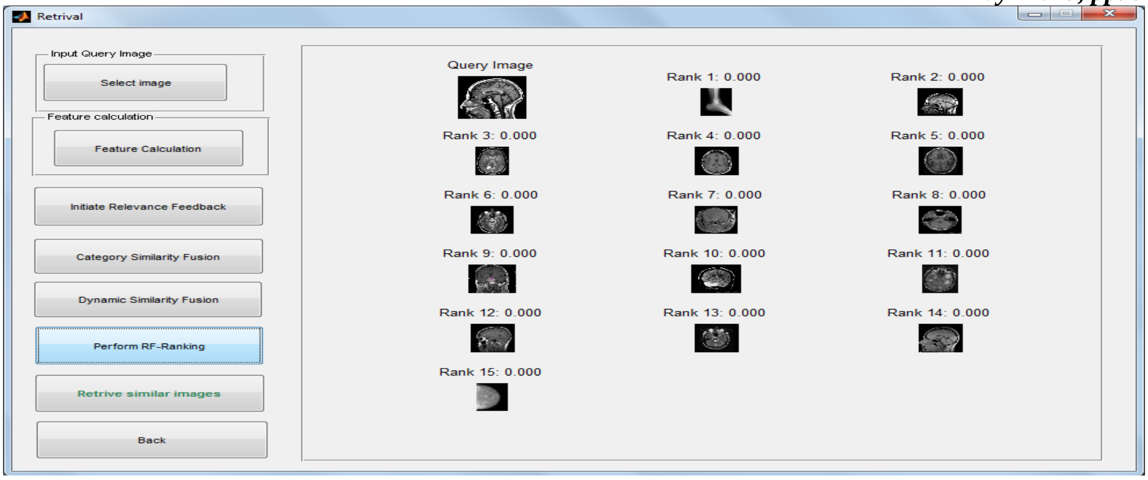


Fig.3 Results after Relevant Feedback

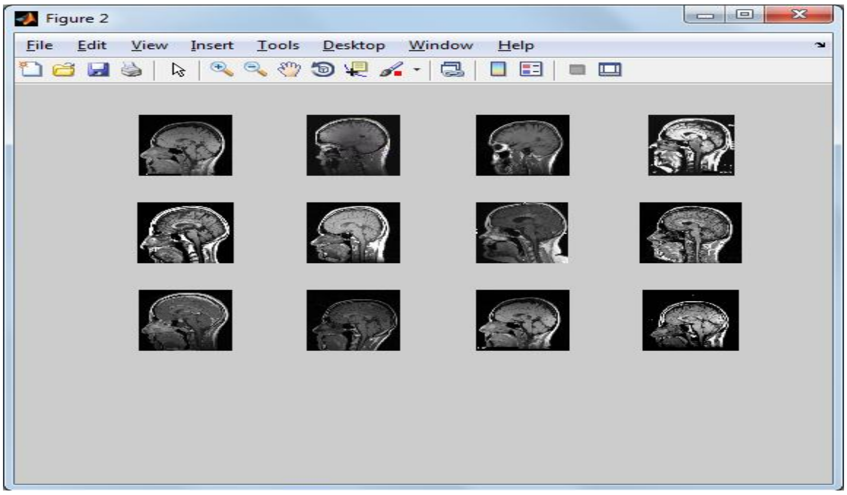


Fig 4. Final Retrived Similar Images

## VII. CONCLUSION

In this Paper, a novel classification-drivenimage retrieval framework is proposed for diverse imagecollections of different modalities. In our approach, wedirectly link classification to retrieval. In this framework, theimage category information is utilized directly to filter out irrelevantimages and adjust the feature weights in a linear combinationof similarity matching. We use the RF-based techniqueto update the feature weights based on positive user feedback.Retrieval performance is promising and clearly shows the advantageof searching images based on similarity fusion and filteringin terms of effectiveness and efficiency. Overall, this retrievalframework is useful as a front end for large medical databaseswhere a search can be performed in diverse images for teaching,training and research purposes.

## REFERENCES


[1] T. C. Wong, Medical Image Databases. New York, LLC: *Springer–Verlag*, 1998.

[2] H. Muller, N. Michoux, D. Bandon, and Geissbuhler, "A review of content-basedimage retrieval applications—Clinical benefits and future directions," *Int. J. Med. Informat*., vol. 73, no. 1, pp. 1–23, 2004.

[3] A. Smeulder, M. Worring, S. Santini, A. Gupta, and R. Jain, "Contentbased image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[4] C. R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka, A. M Aisen, and L. S. Broderick, "ASSERT: A physician-in-the-loop content-based image retrieval system for HRCT image databases," *Comput. Vis. Image Understand*., vol. 75, pp. 111–132, 1999.

[5] W. Hsu, S. Antani, L. R. Long, L. Neve, and G. R. Thoma, "SPIRS: A web-based image retrieval system for large biomedical databases," *Int. J. Med. Informat*., vol. 78, pp. 13–24, 2008.

[6] M. M. Rahman, P. Bhattacharya, and B. C. Desai, "A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback," *IEEE Trans. Inf. Tech. Biomed*., vol. 11, no. 1, pp. 59–69, Jan. 2007.

[7] T. M. Lehmann, B. B. Wein, J. Dahmen, J. Bredno, F. Vogelsang, and M. Kohnen, "Content-based image retrieval in medical applications—A novel multi-step approach," *Proc. SPIE*, vol. 3972, pp. 312–320, 2000.

[8] H. M¨uller, A. Rosset, J. Vallee, and A. Geissbuhler, "Integrating contentbased visual access methods into a medical case database," *in Proc. Med Inf. Eur*., St Malo, France, pp. 480–485.

[9] H. M¨uller, J. Kalpathy-Cramer, I. Eggel, S. Bedrick, S. Radhouani, B. Bakke, C. E. Kahn, Jr., and W. Hersh, "Overview of the CLEF 2009 Medical Image Retrieval Track," 10th Workshop of the Cross-Language Evaluation Forum, CLEF 2009, Corfu, Greece, Sep. 30–Oct. 2, 2009, Proceedings of LNCS, vol. 6242, pp. 72–84, 2010.

[10] A. Mojsilovis and J. Gomes, "Semantic based image categorization, browsing and retrieval in medical image databases," in *Proc. IEEE Int. Conf. Image Process*., 2002, vol. 3, pp. 145–148.

[11] T. M. Lehmann, M. O. G¨uld, T. Deselaers, D. Keysers, H. Schubert, K. Spitzer, H. Ney, and B. B.Wein, "Automatic categorization of medical images for content-based retrieval and data mining," *Comput. Med. Imag. Graph.*, vol. 29, pp. 143–155, 2005.

[12] F. Florea, H.M¨uller, A. Rogozan, A. Geissbuhler, and S. Darmoni, "Medical image categorization with MedIC and MedGIFT," in *Proc. Med. Inf. Eur., Maastricht*, Netherlands, pp. 3–11.

[13] X. S. Zhou and T. S. Huang, "Relevance feedback for image retrieval: A comprehensive review," *Multimedia Syst.*, vol. 8, no. 6, pp. 536–544, 2003.

[14] Y. Rui and T. S. Huang, "Relevance feedback: A power tool for interactive content-based image retrieval," *IEEE Circuits Syst. Video Technol.*, vol. 8, no. 5, pp. 644–655, Sep. 1999.

[15] M. M. Rahman, S. K. Antani, and G. R. Thoma, "A classification-driven similarity matching framework for retrieval of biomedical images," presented at the 11th ACM *Int. Conf. Multimedia Inf. Retrieval*, National Constitution Center, Philadelphia, Pennsylvania, Mar. 29–31.

[16] M. M. Rahman, S.K.Antani, and G.R. Thoma, "A medical image retrieval framework in correlation enhanced visual concept feature space," in Proc. *22nd IEEE Int. Symp. Comput.-BasedMed. Syst*., Albuquerque, NM, Aug.3–4, 2009, pp. 1–4.

[17] T. F. Wu, C. J. Lin, and R. C. Weng, "Probability estimates for multiclass classification by pairwise coupling," *J. Mach. Learning Res*., vol. 5, pp. 975–1005, 2004.

[18] R. Baeza-Yates and B.Ribeiro-Neto, *Modern Information Retrieval*,Reading, MA: Addison-Wesley, Jun. 1999.

[19] M. M. Rahman, S. K. Antani, and G. R. Thoma, "Biomedical image retrieval in a fuzzy feature space with affine region detection and vector quantization of a scale-invariant descriptor," presented at the *6th Int. Symp. on Visual Computing*, Las Vegas, NV, Nov. 29–Dec. 1.

[20] S. F. Chang, T. Sikora, and A. Puri, "Overview of the MPEG-7 standard," *IEEE Trans. Circ. Syst. Video Technol*., vol. 11, no. 6, pp. 688–695, Jun. 2001.

[21] M. Grubinger, P. Clough, A. Hanbury, andH.M¨uller, "Lire: Lucene image retrieval: An extensible java CBIR library," in Proc. *16th ACM Int. Conf. Multimedia, Vancouver, British Columbia*, Canada, 2008, pp. 1085–1088.

[22] V. Vapnik, *Statistical Learning Theory. New York*: Wiley, 1998.

[23] K. Duan and S. S.Keerthi, "Which is the bestmulticlassSVMmethod? An empirical study," presented at the 6th *Int.Workshop on Multiple Classifier Syst*., LNCS 3541, Seaside, CA, Jun. 13–15, 2005.

[24] C.W. Hsu and C. J. Lin, "A comparison of methods for multi-class support vector machines*," IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

[25] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal.Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.

[26] E. A. Fox and J. A. Shaw, "Combination of multiple searches," in *Proc. 2nd Text Retrieval Conf.*, 1994, pp. 243–252 (NIST Special Publication).

[27] J. H. Lee, "Combining multiple evidence from different properties of weighting schemes," in *Proc. 18th Annu. ACM-SIGIR*, 1995, pp. 180– 188.

[28] H. M¨uller, T. Deselaers, E. Kim, C. Kalpathy, D. Jayashree, M. Thomas, P. Clough, andW. Hersh, "Overview of the ImageCLEFmed 2007 medical retrieval and annotation tasks," in *Proc. LNCS 8th Workshop Cross-Lang. Eval. Forum,* 2008, vol. 5152, pp. 472–491.

[29] H. M¨uller, J. Kalpathy-Cramer, I. Eggel, S. Bedrick, C. E. Kahn, Jr., and W. Hersh, "Overview of the CLEF 2010 medical image retrieval track," presented at the *Working Notes for the CLEF Workshop*, Padova, Italy, 2010.

[30] C. C. Chang and C. J. Lin. (2001). "LIBSVM: A library for support vector machines," [Online]. Software available at http://www.csie.ntu.edu.tw/cjlin/libsvm

[31] L. H. Tang, R. Hanka, H. H. S. Ip, and R. Lam, "Extraction of semantic features of histological images for content-based retrieval of images," in Proc. IEEE Symp. *Comput.-Based Med. Syst*., 2000, p. 193.

 