

A Content based CT Image Retrieval by Advance Multiple Comparison Technique

J.Bridget Nirmala and S.Gowri

Abstract - In this paper, a content-based scheme to retrieve lung Computed Tomographic images (CT) is presented. It consists of a visual based user interface to allow the query be made by line-drawing the interested (abnormal) regions; and a training scheme to classify the relationship between the images stored in database. The system will output a set of candidate images that are textural-similar to the query image. Using this result system will calculate the average of affected area from these similar images that are once again compared with database images. Now the system will give a set of candidate images. It has only most similar images from the database. This approach is better than existing methods because here the system compared to the affected image with the database image in the first stage. At the second stage the system will calculate the average of affected area of similar images from the output of candidate images. Finally the system produced the images that are similar to the candidate images. The experimental result shows that in simple query, 98% of images can be correctly retrieved.

Index Terms – Lung images MRI, Imaging, Segmentation, Advanced multiple comparison technique.

I. INTRODUCTION

In the past few years, there has been tremendous growth in the database technology to store, retrieve and process large number of images [1],[15],[16]. Initially, images were represented just by textual description and observations. But these methods hardly measured the details of on image.

This led to research on the lines of automatically extracting features of images for the purpose of efficient retrieval and sequencing of images which is referred as Content Based Image Retrieval (CBIR) [3],[4],[14]. Main intention of CBIR is efficient retrieval of images form a huge image database based on some automatically extracted features.

Manuscript received February 20, 2011; revised March14. This work was supported by Anna university Trichy,Tamilnadu,India and St.Michael College of Engg and Technology, Kalyarkoil, Sivagangai.Tamilnadu India.

J.Bridget Nirmala, Dept of Computer Science, St. Michael College of Engg and Technology, Kalayarkoil, TamilNadu, India, Phone Number:9942796744 e-mail:bridget_stalin@yahoo.com.

S. Gowri, Registrar, Anna University of Technology ,Chennai- 600113, Tamil Nadu, India, e-mail: sgowri@annauniv.edu,

These features are extracted from properties such as shape, color and texture of query image and the various images in the repository [7]. Based on some common parameter evaluated from the feature, the relevancy between the query image and the database image are arranged accordingly. A typical CBIR system for retrieving images from database based on their similarity to the input image consists of four main steps. First, extract the features of the image to convert the image from spatial data to feature vector. The feature extraction is the basic process of a CBIR system [8]. The system will select specific features and explore various visual features to uniquely identify an image [9]. Next, construct feature vectors based on selected features, by which images in the database are represented. Later on, compare the feature vectors of a query image with the feature vector of images in database by computing a similarity measure to search for the most relevant images in the database. Here Euclidian distance [6],[10]-[13] is used as similarity measure. The direct Euclidian distance between an image P and query image Q can be given as below.

$$ED = \sqrt{\sum_{i=1}^n (V_{pi} - V_{qi})^2} \quad (1)$$

Where, V_{pi} and V_{qi} are the feature vectors of image P and uery image Q respectively with size 'n'.

II. DISCRETE COSINE TRANSFORM

The discrete cosine transform (DCT) [5],[15] is closely related to the discrete Fourier transform. It is a separable linear transformation; that is, the two-dimensional transform is equivalent to a one-dimensional DCT performed along a single dimension followed by a one-dimensional DCT in the other dimension. The definition of the two-dimensional DCT for an input image A and output image B is

$$B_{pq} = \alpha_p \alpha_q \sum_m A_{mq} \cos \frac{\pi(2m+1)p}{2M} \cos \frac{\pi(2m+1)q}{2N}, \quad 0 \leq p \leq M-1, \quad 0 \leq q \leq N-1 \quad (2)$$

$$\alpha_p = \begin{cases} 1/\sqrt{M} & , p = 0 \\ \sqrt{2/M} & , 1 \leq p \leq M-1 \end{cases} \quad (3)$$

$$\alpha_q = \begin{cases} 1/\sqrt{N} & , q = 0 \\ \sqrt{2/N} & , 1 \leq q \leq N-1 \end{cases} \quad (4)$$

Where M and N are the row and column size of A, respectively if you apply the DCT to real data, the result is also real. The DCT tends to concentrate information, making it useful for image compression applications and also helping in minimizing feature vector size in CBIR. For full 2-Dimensional DCT for an NxN image the number of multiplications required are N² (2N) and number of additions required are N² (2N-2).

III. FEATURE VECTORS EXTRACTION

Fig. 1 shows the block diagram of common content-based image retrieval systems. Three modules constitute the system: the input module, the query module, and the retrieval module [1]. In the input module and the query module, the feature vectors are extracted first. A training process is performed to organize those features. When a query image enters the query module, it extracts the feature vector of the query image. Then in the retrieval module, the extracted feature vector is compared to the feature vectors stored in the database. The target images are the similar images, which retrieved according to their matching scores. Unlike other general content-based image retrieval system, where the shape, size, texture, color and location of an object are usually considered as the feature of the input object in an image. After interviewed with several radiologists, we found that the most important features of lung cancer they are concerned are the Density (texture or the type of distribution of the brighter pixels) and location of the selected ROI area. According to this, we define feature vector as two parts, one is the texture information, and the other is position data. Due to the fact that we use small block to extract texture feature, the abnormal region is usually larger than the block. In order to achieve the best result, we must fully cover the abnormal regions with blocks.

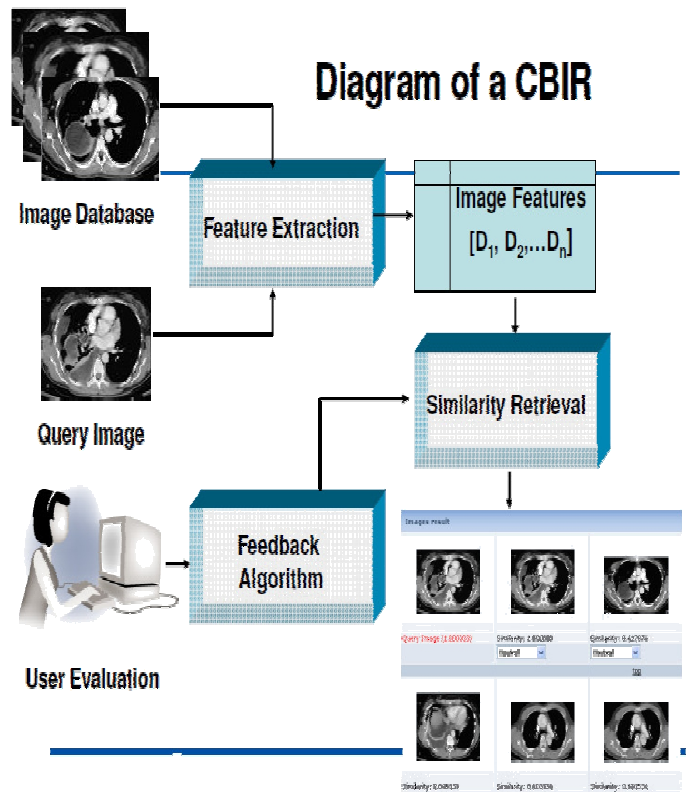


Fig.1 General Diagram of CBIR

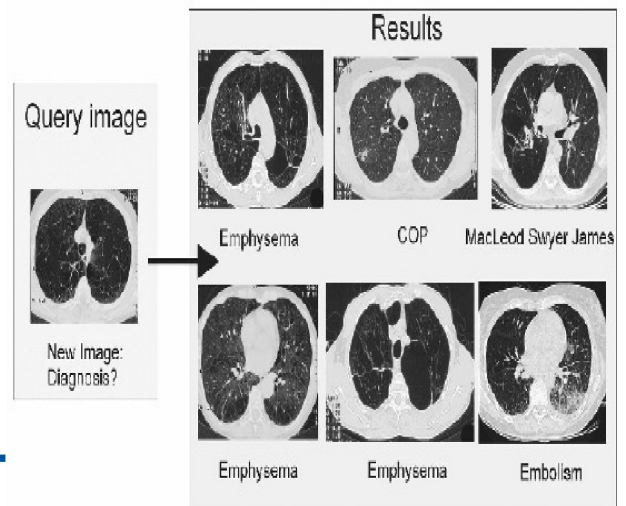
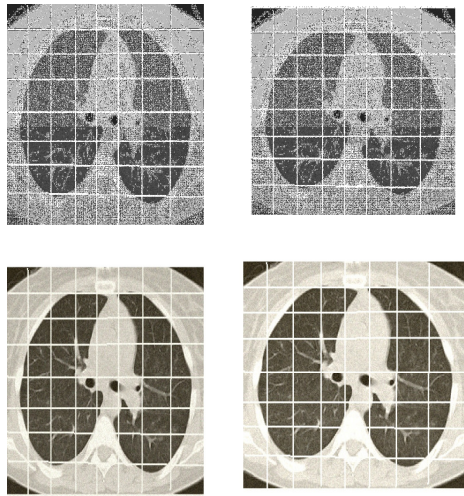


Fig.1 (a) Result images

IV. RESULT OF DATABASE IMAGES



V. IMAGE RETRIEVAL METHOD

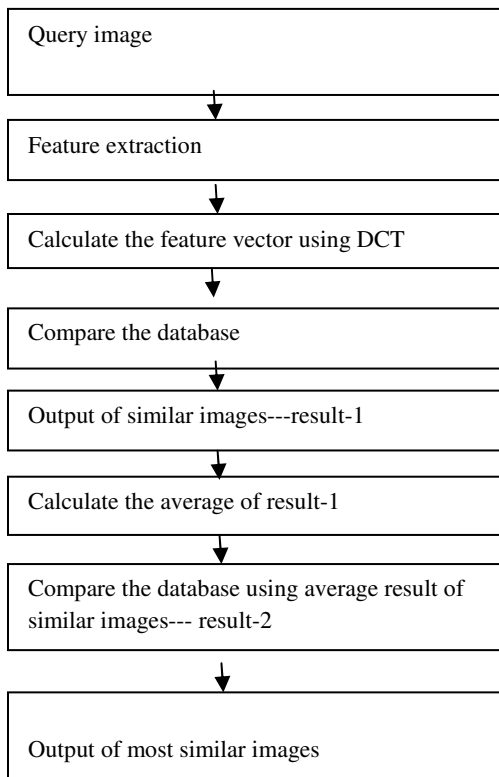


Fig.2 shows the system step by step process

In this system there are five steps to involve retrieving CT lung images.

1. Query image
2. Feature extraction & feature vector calculation (using DCT)
3. Compare the database (result-1)
4. Calculate average of result-1
5. Compare the database (result-2)

In general three modules constitute the CBIR system: the input Module, the query module, and the retrieval module in the input module and the query module, the feature vectors are extracted first. A training process is performed to organize those features. When a query image enters the query module, it extracts the feature vector of the query image. Then in the retrieval module, the extracted feature vector is compared to the feature vectors stored in the database. The target images are the similar images, which retrieved according to their matching scores. In our system contain five modules involve to produce the similar images. First step is a query image, our system input as CT lung image output of candidate images. The second module query image extracts the feature vectors using DCT. The query image of feature vector compared with database feature vector that are stored in the database. Using this result images system will calculate the average of affected area from these similar images. From the average of similar images once again the system will compare the database images. Again the system will gives a set of candidate images. It has most similar images from the database. This approach is better than existing methods because here the system compared to the affected image with the database image in the first stage. At the second stage the system will calculate the average of affected area of similar images on the output of candidate images. Finally the system produced the images that are similar to the candidate images. The experimental result shows that in simple query, 98% of images can be correctly retrieved.

VI. RESULT AND DISCUSSION

Here when a query image divided into 64 sub division by 8 vertical and horizontal lines on the input image, it extracts the feature vector of the query image. This system takes input image feature vector as P and database feature vector as Q. Using DCT can be calculate the feature vector stores in a database table-1 and $n=8$. Thus the system stores the $8*8=64$ feature vector vales store in a table1.

Then the input image feature vector P is compared to Q(Q is database feature vector). Later on, compare the feature vectors of a query image with the feature vector of images in database by computing a similarity measure to search for the most relevant images in the database. Here Euclidian distance [6],[10]-[13] is used as similarity measure. The direct Euclidian distance between an image P and query image Q can be given as below.

$$ED = \sqrt{\sum_{i=1}^n (V_{pi} - V_{qi})^2} \quad (1)$$

Where, V_{pi} and V_{qi} are the feature vectors of image P and Query image Q respectively with size 'n=8'. From the equation -1 the system will measure the most relevant image fig .3 this database result is take result-1.

Table: 1 8 X 8 Feature vector

	Col. 1	...	Col. n
Row 1	35	34	25
Row 2	78	24	68
...
Row n	68	76	45

Table: 2 Average calculations

	Col. 1	...	Col. n
Row 1	35	34	25
Row 2	78	24	68
...
Row n	68	76	45

Avg(Row 1) = $(35+34+...+25)/n$

Avg(Col. n) = $(25+68+...+45)/n$

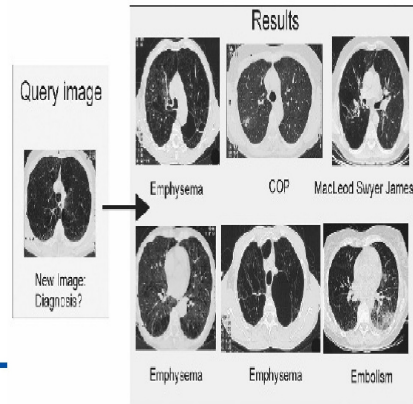


Fig.3 database result-1

Now, the system will output the set of candidate images, from the set of similar image the system calculated the average of distance vector measured by equation- shown in table2.

Table: 3 similar image average vector

10	18	...	22
45	15
....
22	70

After finds the average the system stores the feature vector values into another database table. (Table-3).Now the input image P is average feature vector is given to the database feature vector store in a database .Now the system will produce the most similar image this approach is better than later approaches because in later only one time the image retrieval processes should be done. By our approaches there two times the searching process takes place. Also the system calculates the average of the similar image feature vector store in a database.

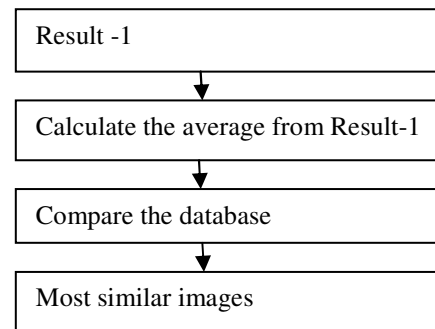


Fig.4 Second process

Average calculation is a simple process from the first result of the similar images feature vector Shown in graph. From the graph system has been calculates the average feature vector values. These values are store in a database table. First each similar image feature vectors were drawn in the graph up to all the similar images feature vectors. From the graph system automatically calculate the average of the feature vector then the result feature vector compare once again in the database.

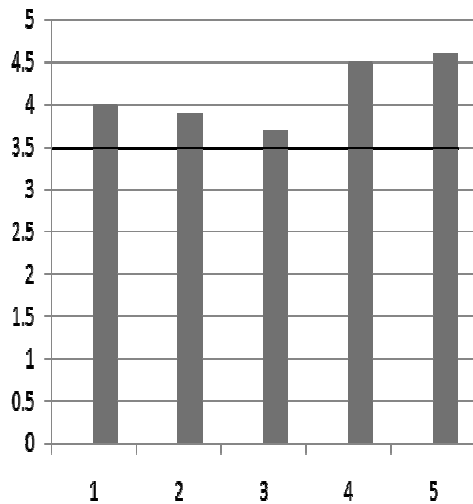


Fig.5 Similar image feature vector graph

From this above graph the average of similar images was lies in 3.5 feature vector values. So the system brought all the similar image affected area (3.5 in vector table) from CT lung image database. Now the system produced the set of similar image but now we can get most similar images because the system compared to average vector values (3.5) with database feature vector values. So the Experimental result will produced the 98% of most similar images.

VII. CONCLUSION

In past year the content based image retrieval is the challenging problem to retrieve the similar image from the database table because in later system produced all similar set candidate images. Now this paper used advance comparison technique to find the most similar image in the database. In this paper, a content-based scheme to retrieve lung Computed Tomographic images (CT) is presented. It consists of a visual based user interface to allow the query be made by line-drawing the interested (abnormal) regions; and a training scheme to classify the relationship between the images stored in database. The system will output a set of candidate images that are textural-similar to the query image. Using this result system will calculate the average of affected area from these similar images that are once again compared with database images. Now the system will give a set of candidate images. It has only most similar images from the database. This approach is better than existing methods because here the

system compared to the affected image with the database image in the first stage. At the second stage the system will calculate the average of affected area of similar images on the output of candidate images. Finally the system produced the images that are similar to the candidate images. The experimental result shows that in simple query, 98% of images can be correctly retrieved.

VIII. REFERENCES

- [1] Yu Xiaohong, Xu Jinhua, "The Related Techniques of Content-based Image Retrieval", In International Symposium on Computer Science and Computational Technology, 2008.
- [2] Pengyu Liu, Kebin Jia, Zhuozheng Wang, Zhuoyi Lv, "A New and Effective Image Retrieval Method Based on Combined Features", In Fourth International Conference on Image and Graphics, 2007.
- [3] H.B.Kekre, Sudeep D. Thepade, "Rendering Futuristic Image Retrieval System", In Proc. of National Conference EC2IT-2009, KJSCOE, Mumbai, 20-21 Mar 2009.
- [4] Dr.N.Krishnan, M.Sheerin Banu, C.Callins Christiyana, "Content Based Image Retrieval using Dominant Color Identification Based on Foreground Objects", In International Conference on Computational Intelligence and Multimedia Applications, 2007
- [5] H.B.Kekre, Tanuja Sarode, Sudeep D. Thepade, "DCT Applied to Row Mean and Column Vectors in Fingerprint Identification", In Proceedings of Int. Conf. on Computer Networks and Security (ICCN), 27-28 Sept. 2008, VIT, Pune.
- [6] H. B.Kekre, Sudeep D. Thepade, "Image Blending in Vista Creation using Kekre's LUV Color Space", SPIT-IEEE Colloquium and Int. Conference, SPIT, Andheri, Mumbai, 04-05 Feb 2008.
- [7] H.B.Kekre, Sudeep D. Thepade, "Boosting Block Truncation Coding using Kekre's LUV Color Space for Image Retrieval", WASET Int.
- [8] H.B.Kekre, Sudeep D. Thepade, "Image Retrieval using Augmented Block Truncation Coding Techniques", ACM Int. Conf. ICAC3-09, 23-24 Jan 2009, FCRCE, Mumbai. Is uploaded on ACM portal.
- [9] H.B.Kekre, Sudeep D. Thepade, "Using YUV Color Space to Hoist the Performance of Block Truncation Coding for Image Retrieval", IEEE International Advanced Computing Conference 2009 (IACC'09), Thapar University, Patiala, INDIA, 6-7 March 2009.
- [10] H.B.Kekre, Sudeep D. Thepade, "Color Traits Transfer to Grayscale Images", In Proc.of IEEE First International Conference on Emerging Trends in Engg. & Technology, (ICETET-08), G.H.Raisoni COE, Nagpur, INDIA. Available on IEEE Xplore.
- [11] H.B.Kekre, Sudeep D. Thepade, "Creating the Color Panoramic View using Medley of Grayscale and ColorPartial Images ", WASET Int.
- [12] H.B.Kekre, Sudeep D. Thepade, "Scaling Invariant Fusion of Image Pieces in Panorama Making and Novel Image Blending Technique", Int.Journal on Ismaging (IJI), Autumn 2008, Volume 1, No. A08, Available online at www.ceser.res.in/iji.html.
- [13] H.B.Kekre, Sudeep D. Thepade, "Rotation Invariant Fusion of Partial

- Images in Vista Creation”, WASET Int. Journal of Electrical, Computer and System Engg.(IJECSE), Vol.2, No.2, Spring 2008. Available online at www.waset.org/ijecse/v2/v2-2-13.pdf
- [14] H.B.Kekre, Sudeep D. Thepade, “Image Orthogonal Kekre’s Retrieval using Non-Involutorial Transform”, International Journal of Engineering Research and Industrial Applications (IJERIA), Ascent Publication House, 2009, Volume 2, No.VII, 2009. Abstract available online at www.ascentjournals.com.
- [15] H.B.Kekre, Sudeep D. Thepade, “Improving the Performance of Image Retrieval using Partial Coefficients of Transformed Image”, International Journal of Information Retrieval, Serials Publications, Volume 2, Issue 1, 2009, pp. 72-79.
- [16] H.B.Kekre, Tanuja K. Sarode, Sudeep D. Thepade, “Image Retrieval using Color-Texture Features from DCT on VQ Codevectors obtained by Kekre’s Fast Codebook Generation”, ICGST International Journal on Graphics, Vision and Image Processing (GVIP), Volume 9, Issue V, Sept. 2009, pp. 1-8. <http://www.icgst.com/gvip>
- [17] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, “Query by image and video content: The QBIC system”. *IEEE Computer*, pp. 23-32, September 1995.
- [18] C. R. Shyu, C. E. Brodley, A. C. Kak, A. Kosaka, A. Aisen, and L. Broderick, “Local versus Global Features for Content-Based Image Retrieval”. *IEEE Workshop on Content-Based Access of Image and Video Libraries*, 1998. Proceedings
- [19] Shyu, C. R., Brodley, C., Kak, A., Kosaka, A., Aisen, A. and Broderick, L., “ASSERT, A physician-in-the-loop content-based image retrieval system for HRCT image databases”, *Computer Vision and Image Understanding*, Vol. 75, Nos. 1/2, pp. 111-132, July/August 1999.
- [20] Teuvo Kohonen. *Self-Organizing Maps*. Springer-Verlag, Heidelberg, 1995.
- [21] Teuvo Kohonen, Jussi Hynninen, Jari Kangas, and JormaLaaksonen, URL:<http://www.cis.hut.fi/research/somresearch/nrc-programs.shtml>
- [22] J.K.Wu and A.D. Narasimhalu and B.M. Mehtre and C.P. Lam and Y.J. Gao, CORE: A Content-Based Retrieval Engine for Multimedia Information Systems, *Multimedia Systems*, 1995, Vol. 3, pp. 25-41.
- [23] S. Berchtold and C. Boehm and B. Braunmueller and D. A. Keim and H. P. Kriegel, Fast Parallel Similarity Search in Multimedia Databases, SIGMOD Conference, AZ, USA, 1997, pp. 1-12.
- [24] A. Yoshitaka and T. Ichikawa, A Survey on Content-Based Retrieval for Multimedia Databases, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 11, No. 1, 1999, pp.81-93.
- [25] V.Oria and M.T. Özsu and L. Liu and X. Li and J.Z. Li and Y. Niu and P.J. Iglinski, Modeling Images for Content-Based Queries: The DISMA Approach, *VIS’97*, San Diago, 1997, pp.339-346.
- [26] Y. Rui and T.S. Huang and S.F. Chang, Image Retrieval: Past, Present, and Future, *Journal of Visual Communication and Image Representation*, 1999, Vol. 10, pp.1-23.
- [27] William I. Grosky, *Managing Multimedia Information in Database Systems*, Communications of the ACM, 1997, Vol. 40, No. 12, pp. 72-80.
- [28] Content-Based Image Retrieval: A Report to the JISC Technology Applications Programme, John P. Eakins and Margaret E. Graham, January, 1999, Inst. for Image Data Research, Univ. of Northumbria at Newcastle.
- [29] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker, “Query by Image and Video Content: The QBIC system,” *IEEE Computer*, vol. 28, no. 9, pp. 23–32, September 1995.
- [30] H. M’uller, A. Rosset, J.-P. Vall’ee, and A. Geissbuhler, “Integrating content-based visual access methods into a medical case database,” in *Proceedings of the Medical Informatics Europe Conference (MIE 2003)*, St. Malo, France, May 2003.
- [31] H. D. Tagare, C. Jaffe, and J. Duncan, “Medical image databases: A content-based retrieval approach,” *Journal of the American Medical Informatics Association*, vol. 4, no. 3, pp. 184–198, 1997.
- [32] H. J. Lowe, I. Antipov, W. Hersh, and C. Arnott Smith, “Towards H. M’uller, N. Michoux, D. Bandon, and A. Geissbuhler, “A review of content-based image retrieval systems in medicine – clinical benefits and future directions,” *International Journal of Medical Informatics*, vol. 73, pp. 1–23, 2004.
- [33] C. Le Bozec, E. Zapletal, M.-C. Jaulent, D. Heudes, and P. Degoulet, “Towards content-based image retrieval in HIS-integrated PACS,” in *Proceedings of the Annual Symposium of the American Society for Medical Informatics (AMIA)*, Los Angeles, CA, USA, November 2000, pp. 477–481.
- [34] H. Abe, H. MacMahon, R. Engelmann, Q. Li, J. Shiraishi, S. Katsuragawa, M. Aoyama, T. Ishida, K. Ashizawa, C. E. Metz, and K. Doi, “Computer-aided diagnosis in chest radiography: Results of large-scale observer tests at the 1996–2001 RSNA scientific assemblies,” *RadioGraphics*, vol. 23, no. 1, pp. 255–265, 2003.
- [35] C.-T. Liu, P.-L. Tai, A. Y.-J. Chen, C.-H. Peng, T. Lee, and J.-S. Wang, “A content-based CT lung retrieval system for assisting differential diagnosis images collection,” in *Proceedings of the second International Conference on Multimedia and Exposition (ICME’2001)*, IEEE Computer Society, Tokyo, Japan: IEEE Computer Society, August 2001, pp. 241–244.
- [36] 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 retrieval system for HRCT image databases,” *Computer Vision and Image Understanding (special issue on content-based access for image and video libraries)*, vol. 75, no. 1/2, pp. 111–132, July/August 1999.
- [37] C. Brodley, A. Kak, C. Shyu, J. Dy, L. Broderick, and A. M. Aisen, “Content-based retrieval from medical image databases: A synergy of human interaction, machine learning and computer vision,” in *Proceedings of the 10th National Conference on Artificial Intelligence*, Orlando, FL, USA, 1999, pp. 760–767.