MULTI-MODAL INFORMATION RETRIEVAL FOR CONTENT-BASED MEDICAL IMAGE AND VIDEO DATA MINING

Peijiang Yuan, Bo Zhang and Jianmin Li
State Key Laboratory of Intelligent Technology and Systems
Tsinghua National Laboratory for Information Science and Technology
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

Keywords: Content-Based Medical Video Retrieval (CBMVR), Artificial Potential Field (APF), Multimodal Information Retrieval.

Abstract: Image based medical diagnosis plays an important role in improving the quality of health-care industry. Content based image retrieval (CBIR) has been successfully implemented in medical fields to help physicians in training and surgery. Many radiological and pathological images and videos are generated by hospitals, universities and medical centers with sophisticated image acquisition devices. Images and Videos that help senior or junior physician to practice medical surgery become more and more popular and easier to access through different ways. To help learn the process of a surgery or even make decisions is one of the main objectives of the content based image and video retrieval system. In this paper, a contented-based multimodal medical video retrieval system (CBMVR) for medical image and video databases is addressed. Some key issues are discussed. A new feature representation method named Artificial Potential Field (APF) is addressed which is specially useful in symmetrical imaging feature extraction. Experimental results show that, with this CBMVR, both the senior and junior physicians can benefit from the mass data of medical images and videos.

1 INTRODUCTION

Content-based Video and Image Retrieval (CBVIR) has been an exciting and fastest growing research area in the last few decades (Smeulders et al., 2000),(Liu et al., 2001). Many approaches for visual representation and vision systems have been addressed (Poggio and Bizzi, 2004),(Taylor et al., 2002). Medical images and videos can provide not only a view of internal structures of patients, but also a direct view for physicians to evaluate the patient's diagnosis and monitor the effects of the treatment, and for researchers to understand the underlying diseases. Currently, many PACS contain terabytes of image data on-line, thereby requiring high quality Content-Based-Medical-Video-Retrieval (CBMVR) technology. However, medicalimages and videos have different characteristics for each imaging system. CBMVR is much different from traditional video retrieval systems. Due to this reason, practical CBMVR applications have not been widely used in medical surgery and education so far. In most cases, professional physicians examine images and videos in conventional ways based on their individual experiences and knowledge. As a result, the increasing demand for efficient management of such data is making CBMVR a very active area of research. Some efforts have already been made in the area of medical image retrieval research. In general, CBMVR is an important technique in medical domain. It can be used in monitoring multiple image modalities, verifying changes or evolution of a certain disease and specially useful in video-guided-surgery.

This paper will focus on some new medical video retrieval techniques to analyze medical imaging and video data from real clinical studies. The main contribution of this paper can be described as follows:

1. In this paper, a new approach for CBMVR system will be presented. Much of current research mainly focus on medical image retrieval.
2. A new method of artificial potential field (APF) based feature extraction will be proposed. This method works well with medical images.
3. A multi-modality keyframe sequences matching algorithm which integrates text, video and audio will be addressed. Late-fusion techniques are used to improve the performance.

Yuan P., Zhang B. and Li J. (2009).
MULTI-MODAL INFORMATION RETRIEVAL FOR CONTENT-BASED MEDICAL IMAGE AND VIDEO DATA MINING.
In *Proceedings of the First International Conference on Computer Imaging Theory and Applications*, pages 83-86
DOI: 10.5220/0001774200830086

2 MOTIVATION

The main purpose of this work is to design a multi-modality CBMVR system by integrating the text, video and audio that realizes the greatest possible benefit from gathering, indexing, communicating, managing, and archiving multimedia data to provide a reliable health-care delivery, medical education, and medical research platform. Medical video retrieval can be useful as a training tool for medical students and physicians in education for detecting the evolution of diseases and for research purposes. For instance. Given a CBMVR system, the senior or junior physicians can benefit from the system both in decision-making and surgery training.

Most current CBMVR methodologies are based on a specific image modality, for example, global features such as color and texture and regional features or local features. To build a practical computer aided diagnostic system, all the relevant technologies, multi-modality technologies especially the most successful text retrieval, audio and video technologies need to be integrated in an inter-operable manner. Fig. 1 gave a lymphoma surgery example where the physicians need to know the current status of the lymphoma disease from thousands of CT and MRI images. From Fig.1, we can see that, medical images share some unified features and quite different from everyday images and videos.

3 RELATED WORK

Some projects for the use of content-based image retrieval methods in the medical domain in general have been addressed from the literature (Orphanoudakis et al., 1994). In (Bucci et al., 1996), CBIR is proposed in the context of a case database containing images and attached case descriptions. A medical reference database was described within a teaching file assistant in (Squire et al., 1999). However, the used visual features are not well defined. An on-line pathology atlas uses the search-by-similarity paradigm in (Cord et al., 2003), (Cai et al., 2001), (Muller et al., 2004), (Kim and Vasudev, 2005), (Tagare et al., 1997).

In general, there are two major approaches of image data description and retrieval in the literature: the metadata oriented and the content-based oriented. However, these known systems mainly focus on retrieval by image and give less or no emphasis to the role of medical video retrieval. CBMVR has not been well addressed.

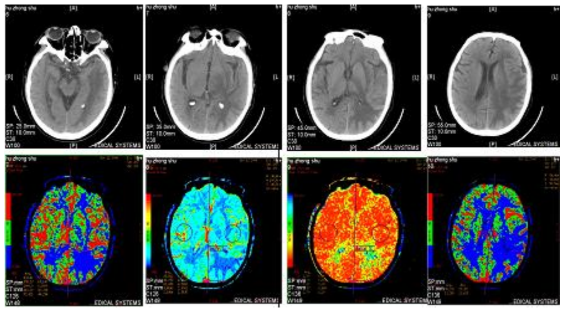


Figure 1: Different evolution status of lymphoma with grayscale and color CT and MRI images.

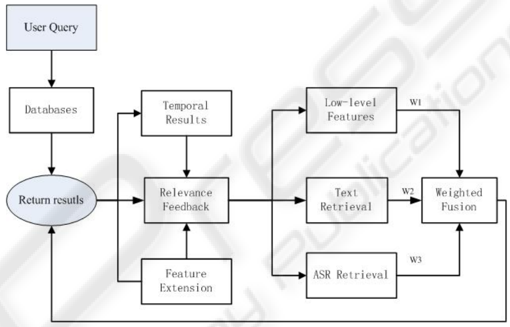


Figure 2: The architecture of the CBMVR system (ASR: Automatic Speech Recognition).

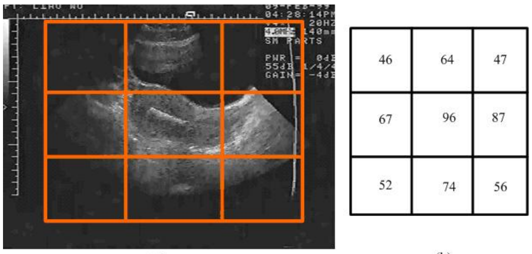
4 OUR APPROACH

We aim at building a multi-modality CBMVR system that can benefit from gathering, indexing, communicating, managing, and archiving multimedia data to provide a reliable healthcare solution, medical education, and medical research platform that can be used for surgery and training. The fundamental features of this CBMVR can be listed as follows:

- The feature description of the image contents should be strong, robust and scalable for different sizes of images and videos.
- The system should be fast or even real-time since some applications need immediate response.

In this case, low-level features like color, texture features are suitable in CBMVR. Use of automatic video retrieval techniques such as text, images, concepts, audio and other modes respectively, the ultimate result of retrieval sequence can be achieved by integration of all the modalities. In addition, by using weighted configuration, automatic classification based on the results of the different modal integration can bring very good results performance. Fig. 2 gives the architecture of the CBMVR system.

The system mainly has three parts: the query, the database and the fusion results. As mentioned above,



(a) (b)

Figure 3: Features of an Ultrasound image (a) the 3*3 matrix representation(b) the average values of each block).

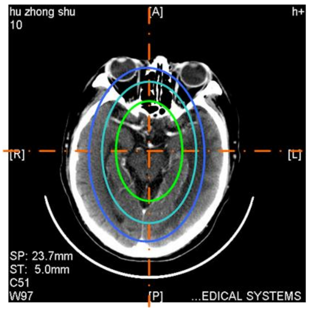


Figure 4: Feature space in artificial potential field.

the features of each keyframe should be robust and not dependent on the image style, for instance, the features can be used in CT, MRI, PET and Ultrasound respectively. Here we choose low-level feature.

The above features are easily to use and robust. The performance, however, sometimes is not very good since the feature space is not salient in many situations. For instance, in an Ultrasound video or MRI video, the grayscale feature space is usually flat. The MAP (minimum average precision) can be low. To solve this problem, we propose a new method which has been successfully implemented in robotics: the artificial potential field.

Potential field methods were addressed for obstacle avoidance in (Khatib and Maitre, 1978). The obstacles were represented as zero level surfaces of scalar valued analytic functions, i.e. $f(x;y;z) = 0$. In this case, we can define an arbitrary cutoff value f_0, which corresponds to the distance where the influence of the potential is no longer important. The potential field can be mathematically described in the following form:

$$P(x,y,z) = \begin{cases} \alpha/f(x,y,z)^2 & f(x,y,z) \leq f_0 \\ 0 & f(x,y,z) > f_0 \end{cases} \quad (1)$$

As depicted in Fig. 4, many medical images are symmetrical and center based. If we adopt potential field as the feature space with the center in the middle of the medical image, the segmentation based Ordinal Measure values might be more precise and robust. We choose the feature vector as:

$$\bar{F} = \{P_1, P_2, ..., P_n\} \quad (2)$$

where P_i is the potential field value of the medical image. This potential based medical image features may impose challenges for efficient processing and indexing. Thus the features, the keyframes, and the video sequences are known. By implementing ANN or LSH method, we can build a CBMVR to find the exact or near-duplicate medical videos and images.

Table 1: The average performance using the average values under Trecvid and medical dataset.

	Precision	Recall
489,655 keyframes	76.9%	85.5%

5 EXPERIMENTAL RESULTS

To test the performance of the given CBMVR, we build two experiments. Our system is based on a Pentium IV PC with 3.4G Hz, 1G memory and 120G hard disk. First, to test the efficiency, TRECVID 2006 database was used.

5.1 Experiment I

First Trecvid videos and some CT and MRI images from Tianjin Medical University, Department of Medical Imaging, are used in this experiment. From 2003 to 2006, about 918 video files, about 530 hours of videos with 485,655 keyframes were used and the medical images are about 4000 keyframes. The Precision and Recall of the experimental results are listed in Table 1: Here we assume that, given a query image, the exact or similar images in the fist m results (Precision @ m) is defined as the percentage of Precision. From the results, we can see that, the precision and the recall is not very good and the time complexity increases dramatically with the growth of the database (with K=30). Therefore, we give another method.

5.2 Experiment II

From experiment I, we can see that, the CBMVR system works fine with the Trecvid data. The performance, however, is not very good (both the precision and recall are not good enough). To solve this problem, while keeping the scalability of the time complexity and the memory usage, we inspired by the text retrieval method and the potential field features. In this experiment, we use potential field features and include the Trecvid data and data from the MRI and

Table 2: The performance using multi-modality integration under both Trecvid and medical dataset.

	Precision	Recall
489,655 keyframes	94.6%	96.4%

CT images. The images are all grayscale, 512*512 pixels. The experiments showed that the accuracy of this algorithm has been greatly improved (Precision 94.6% Recall 96.4%. (See Table 2).

6 CONCLUSIONS AND FUTURE WORK

In this paper, a new approach for medical image and video retrieval system is presented. A new method based on keyframe matching and partial sequence alignment is proposed. An extensive evaluation of different methods for multi-modality automatic categorization of medical images is presented. A new feature space expression named artificial potential field based feature extraction method is discussed. The experimental results show that it is feasible and performs well. The average performance and precision is pretty promising. It is shown that the addressed approaches are promising to offer new possibilities for content-based access to medical images as an accuracy of 94% within the thirty best matches is sufficient for most applications. Content-based image retrieval systems that are no longer limited to a special context are becoming possible. Our future work will focus on the dataset collection and the multi-modality data mining.

ACKNOWLEDGEMENTS

This research was supported in part by the National Natural Science Foundation of China under the grant No. 60621062 and 60605003, and the National Key Foundation R&D Projects under the grant No. 2003CB317007 and 2004CB318108 and China Postdoctoral Science Foundation 20080430422.

REFERENCES

Bucci, G., S., C., and Domicinis, R. D. (1996). Integrating content-based retrieval in a medical image reference databasez. *Computerized Medical Imaging and Graphics*, 20(4):231–241.

Cai, W., Feng, D., and Fulton, R. (2001). Content-based retrieval of dynamic pet functional images. *IEEE Trans. Inform. Tech. Biomed*, 4(2):152–158.

Cord, M., Fournier, J., and Philipp-Foliguet, S. (2003). Exploration and search-by-similarity in cbir. In *Proc. of SIBGRAPI 03*, Sao Carlos, Brsil.

Khatib, O. and Maitre, J. L. (1978). Dynamic control of. manipulators operating in a complex environment. *Proceedings Third International CISM-IFToMM Symposium,September 1978*, pages 267–282.

Kim, C. and Vasudev, B. (2005). Spatiotemporal sequence matching for efficient video copy detection. *IEEE Trans. on Circuits and Systems for Video Technology*, 15(1):127–132.

Liu, Y., Collins, R., and Rothfus, W. (2001). Robust midsagittal plane extraction from normal and pathological 3d neuroradiology images. *IEEE Transactions on Medical Imaging*, 20(3):175–192.

Muller, H., Michoux, N., Bandon, D., and Geissbuhler, A. (2004). A review of content-based image retrieval systems in medical applications–clinical benefits and future directions. *International Journal of Medical Informatics*, 73(1):1–23.

Orphanoudakis, S., Chornaki, C., and Kostomanolakis, S. (1994). I2c-a system for the indexing, storage and retrieval of medical images by content. *Med Informatics*, 19(2):109–122.

Poggio, T. and Bizzi, E. (2004). Generalization in vision and motor control. *NATURE*, 296:768–774.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., and Jain, R. (2000). Contentbased image retrieval at the end of the early years. *IEEE Trans Pattern Anal Machine Intell*, 22(12):1349–1380.

Squire, D. M., Muller, A. W., Muller, H., and Raki, J. (1999). Content-based query of image databases-inspirations from text retrieval-inverted files, frequency-based weights and relevance feedback. *Proceeding Scandinavian Conference on Image Analysis, Kangerlussuaq, Greenland*, pages 143–149.

Tagare, H., Jaffe, C., and Duncan, J. (1997). Medical image databases - a content-based retrieval approach. *Journal of the American Medical Informatics Association*, 4:184–198.

Taylor, D. M., Tillery, S. I. H., and Schwartz1, A. B. (2002). Direct cortical control of 3d neuroprosthetic devices. *SCIENCE*, 296:1829–1832.