USYD/HES-SO in the VISCERAL Retrieval Benchmark

Fan Zhang[1], Yang Song[1], Weidong Cai[1], Adrien Depeursinge[2], and Henning Müller[2]

[1] Biomedical and Multimedia Information Technology (BMIT) Research Group, School of Information Technologies, University of Sydney, NSW, Australia
[2] University of Applied Sciences Western Switzerland (HES-SO), Sierre, Switzerland

Abstract. This report presents the participation of our joint research team in the VISCERAL retrieval task. Given a query case, the cases with highest similarities in the database were retrieved. 5 runs were submitted for the 10 queries provided in the task, of which two were based on the anatomy-pathology terms, two were based on the visual image content, and the last one was based on the fusion of the aforementioned four runs.

1 Introduction

Medical image data produced has been growing rapidly in quantity, content and dimension, due to an enormous increase in the number of diverse clinical exams performed in digital form and to the large range of image modalities and protocols available [1–3]. Retrieving a set of images that are clinically relevant to the query from the large image database has been the focus of medical research and clinical practice [4, 5]. In the VISCERAL retrieval task, we conduct medical image retrieval based on multimodal and multidimensional data. The similarities between medical cases are computed based on extracts of the medical records, radiology images and radiology reports. 5 runs were submitted, of which two were obtained with term retrieval that utilized the anatomy-pathology terms, two were obtained with visual content retrieval that made use of the visual content features, and one was obtained with information fusion that combined the above results.

The structure is as follows: in Section 2.1, we give an overview of the VISCERAL retrieval dataset; in Section 2.2 to 2.4, we introduce the term, visual content and fusion retrieval methods that were used for our submission; in Section 3, we present the results given the 10 topics; and we provide a conclusion in Section 4.

2 Datasets and Methods

2.1 VISCERAL Retrieval Dataset

The VISCERAL retrieval dataset consists of 2311 volumes originated from three different modalities of CT, MRT1 and MRT2. The volumes are from different

human body regions such as the abdomen, thorax and the whole body. Within the whole dataset, 1815 volumes are provided with anatomy-pathology terms extracted from the radiologic reports. A total of 10 topics were used as queries. Each of them was annotated with the 3D bounding box of the region of interest (ROI), binary mask of the main organ affected and the corresponding anatomy-pathology terms. More on the VISCERAL data in general and the evaluation approach can be found in [6].

2.2 Term Retrieval

Medical image retrieval is conventionally performed with text-based approaches, which rely on manual annotation with alpha-numerical keywords. The anatomy-pathology term file provided in the VISCERAL retrieval benchmark lists the pathological terms and affected anatomies that are recorded in the report and were extracted automatically from the German radiology reports and mapped to RadLex.

In our design, we constructed a co-occurrence matrix between the terms and cases. TF-IDF [7] was used to weight the terms in each case and Euclidean distance was applied for similarity computation. This led to the first run of our retrieval result submission. Our second attempt was using the latent topic model to measure the similarity between different cases in terms of semantic description. Probabilistic Latent Semantic Analysis (pLSA) [8] was used to infer the latent topics between the terms and cases and to represent the cases as probability distributions given the extracted latent topics. The similarity between cases was measured by the Euclidean distance between the probability distributions. 20 latent topics were used.

2.3 Visual Content Retrieval

In the literature, there are many methods that can automatically extract the visual features to characterize the images. The bag-of-visual-words (BoVW) [9] method, which is one of the popular methods for visual content-based image retrieval, was applied to obtain the third run of our submission. The scale invariant feature transform (SIFT) [10] descriptors were extracted from each scan of the 3D volume from the axial view. A visual dictionary of size 100 was then computed based on the k-means clustering method. For the retrieval step, given the ROI of a query case, we traversed all sub-regions (of the same size as the ROI) in a candidate volume. The sub-region that has the smallest Euclidean distance from the query ROI in terms of the visual word frequency histograms was regarded as the most similar area of the candidate to the query ROI. The distance between the two regions represented the similarity between the query and candidate images.

Based on the results of the BoVW method, we further conducted a retrieval result refinement process based on our recent work [11] to obtain the fourth run. The method assumed that the similarity relationship between the initial retrieved results and the remaining candidates can be used as relevance feedback for result

refinement. Specifically, the initial results were ranked based on whether the neighboring candidates were related to the query; the candidates were ranked according to the ranking scores of the neighboring initially retrieved items. For our submission, we selected the top 30 volumes based on the BoVW outputs as the initial results. Then, a bipartite graph between the initial results and the remaining candidates, which represented the neighbourhood, was constructed by keeping the top 30 candidates for each initial result. The iterative ranking method [11] was applied to recompute the similarity score of each candidates with an iteration number of 20.

2.4 Fusion Retrieval

It is often suggested that the combination of textual and visual features can improve the retrieval performance [12]. Given the aforementioned four runs from the term and visual content retrievals, we obtained our fifth run by fusing them together. We applied the sum combination method that has been effective for textual and visual feature fusion [13]. To do this, a normalization step was firstly incorporated to normalize the similarity scores of the first four runs, as:

$$S' = \frac{S - S_{min}}{S_{max} - S_{min}} \tag{1}$$

where S_{min} and S_{max} are the lowest and highest similarity scores obtained within a certain run. The sum combination was then adopted to compute a fusion score for each candidate, as:

$$S_F = \sum_{r \in [1,4]} S'_r \tag{2}$$

where $r \in [1, 4]$ represents the first four runs. The ones with the higher scores were for the retrieval results of our fifth run.

3 Results and Discussion

To evaluate the performance of retrieval results, medical experts were invited to perform relevance assessment of the top ranked cases for each run. Difference evaluation measures were used considering the top-ranked X cases, including: the precision for top-ranked 10 and 30 cases (P@10, P@30), mean uninterpolated average precision (MAP), bpref measure, and the R-precision. More on the evaluation measures and result comparisons can be found in [14].

Fig. 1 displays the retrieval result for each of the topics given the aforementioned measures. The performance was diverse across the cases. It can be generally observed that better results were obtained for topics 1 and 7 but the results for topics 9 and 10 were unfavorable. The differences were due to the different affected regions. Our methods computed the similarity between cases using the entire volumes, instead of focusing on the local details. Therefore, for the eases that have a small affected region, e.g., case 10, the similarity tended to

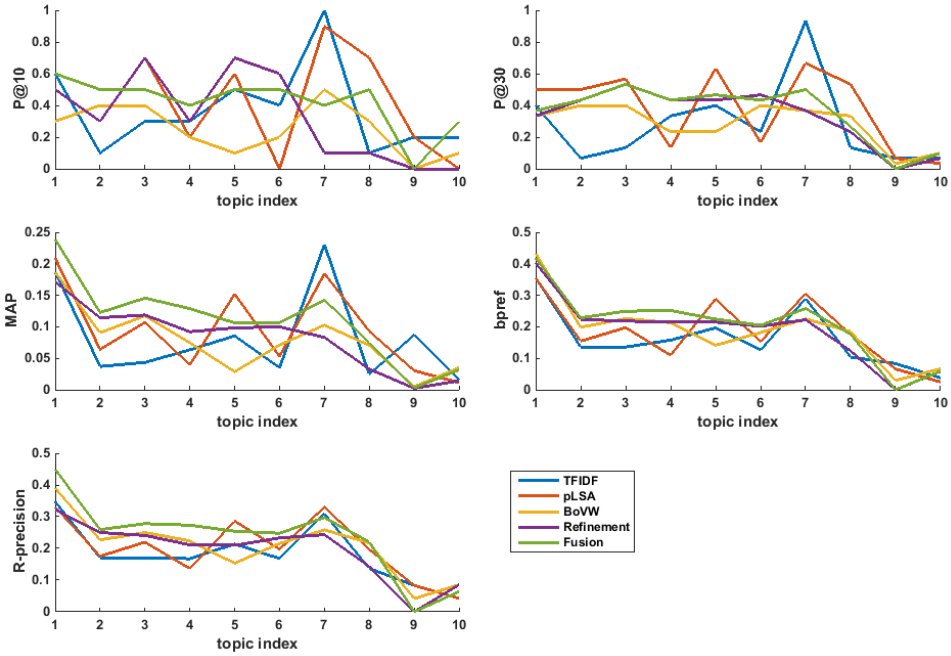

Fig. 1. Retrieval results of the 10 topics given different evaluation measures.

Table 1. Average results of the different measures across the 10 queries.

	P@10	P@30	MAP	bpref	R-precision
TFIDF	0.370	0.277	0.081	0.162	0.181
pLSA	0.410	0.380	0.094	0.183	0.200
BoVW	0.250	0.283	0.078	0.190	0.206
Refinement	0.330	0.330	0.083	0.188	0.194
Fusion	0.420	0.353	0.110	0.207	0.234

be inaccurate. Table 1 shows the average results of the measures across the 10 queries. The better performance was obtained from the term retrieval when compared to the visual content retrieval. The fusion approach achieved the overall best result, which is in accordance with findings in the literature.

4 Conclusions

This report describes our approaches to address the VISCERAL Retrieval task. 5 runs were submitted, based on similarities computed from the anatomy-pathology terms, visual content descriptors, and fusion of these two.

References

1. Doi, K.: Diagnostic imaging over the last 50 years: research and development in medical imaging science and technology. Physics in Medicine and Biology 51, R5–R27 (2006)
2. Müller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content-based image retrieval systems in medicine – clinical benefits and future directions. International Journal of Medical Informatics 73, 1–23 (2004)
3. Cai, W., Feng, D., Fulton, R.: Content-Based Retrieval of Dynamic PET Functional Images. IEEE Transactions on Information Technology in Biomedicine 4(2), 152-158 (2000).
4. Müller, H., Antoine R., Arnaud G., Jean-Paul V., Antoine G.: Benefits of Content-based Visual Data Access in Radiology 1. Radiographics 25(3), 849-858 (2005)
5. Song, Y., Cai, W., Zhou, Y., Wen, L., Feng D.: Pathology-centric Medical Image Retrieval with Hierarchical Contextual Spatial Descriptor. IEEE International Symposium on Biomedical Imaging (ISBI), 202-205 (2013).
6. Hanbury, A., Müller, H., Langs, G., Weber, M. A., Menze, B. H., Fernandez, T.S.: Bringing the algorithms to the data: cloud–based benchmarking for medical image analysis. CLEF conference, Springer Lecture Notes in Computer Science (2012)
7. Jones, K. S.: A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation 28, 11–21 (1972)
8. Hofmann, T.: Unsupervised learning by probabilistic latent semantic analysis. Machine Learning 42, 177–196 (2001)
9. Sivic, J., Zisserman, A.: Video Google: a text retrieval approach to object matching in videos. IEEE International Conference on Computer Vision (ICCV), 1470–1477 (2003)
10. Lowe, D. G.: Object recognition from local scale-invariant features. IEEE International Conference on Computer Vision (ICCV), 1150–1157 (1999)
11. Cai, W., Zhang, F., Song, Y., Liu, S., Wen, L., Eberl, S., Fulham, M., Feng, D.: Automated feedback extraction for medical imaging retrieval. IEEE International Symposium on Biomedical Imaging (ISBI), 907–910 (2014)
12. Müller, H., Kalpathy-Cramer, J.: The ImageCLEF Medical Retrieval Task at ICPR 2010—Information Fusion to Combine Visual and Textual Information. Recognizing Patterns in Signals, Speech, Images and Videos, 99–108 (2010)
13. Zhou, X., Depeursinge, A., Müller, H.: Information Fusion for Combining Visual and Textual Image Retrieval. International Conference on Pattern Recognition (ICPR), 1590–1593 (2010)
14. Jiménez–del–Toro, O. A., Foncubierta–Rodríguez, A., Müller, H., Langs, G., Hanbury, A.: Overview of the VISCERAL Retrieval benchmark 2015. Multimodal Retrieval in the Medical Domain, Lecture Notes in Computer Science 9059 (2015)