RadLex Terms and Local Texture Features for Multimodal Medical Case Retrieval

Oscar Alfonso Jiménez–del–Toro[1,2], Pol Cirujeda[3], Yashin Dicente Cid[1,2], and Henning Müller[1,2]

[1]University of Applied Sciences Western Switzerland (HES–SO), Switzerland,
[2]University and University Hospitals of Geneva, Switzerland,
[3]Universitat Pompeu Fabra, Spain

Abstract. Clinicians searching through the large data sets of multimodal medical information generated in hospitals currently do not fully exploit previous medical cases to retrieve relevant information for a differential diagnose. The VISCERAL Retrieval benchmark organized a medical case–based retrieval evaluation using a data set composed of patient scans and RadLex term anatomy–pathology lists from the radiologic reports. In this paper a retrieval method for medical cases that uses both textual and visual features is presented. It defines a weighting scheme that combines the RadLex terms anatomical and clinical correlations with the information from local texture features obtained from the region of interest in the query cases. The method implementation, with an innovative 3D Riesz wavelet texture analysis and an approach to generate a common spatial domain to compare medical images is described. The proposed method obtained overall competitive results in the VISCERAL Retrieval benchmark and could be seen as a tool to perform medical case based retrieval in large clinical data sets.

Keywords: Medical Case–Based Retrieval, Content–Based Image Retrieval, 3D Riesz wavelet transform.

1 Introduction

As part of their daily workload, clinicians have to visualize and interpret a large amount of medical images and radiologic reports [11]. In recent years, the volume of images in medical records has increased due to the continuous development of imaging modalities and storage capabilities in hospitals. Going through these large amounts of data is time consuming and not scalable with the current trend of big data analysis [10]. Therefore, the challenge to make efficient use of these large data sets and to provide useful information for the clinicians' diagnostic decisions is of high relevance [12].

The Visual Concept Extraction Challenge in Radiology (VISCERAL) project was developed as a cloud–based infrastructure for the evaluation of medical image analysis techniques on large data sets [10, 3]. Through evaluation campaigns, challenges, benchmarks and competitions, tasks of general interest can be selected to compare the algorithms on a large scale. One of these tasks is the retrieval benchmark, which aims to find cases with similar anomalies based on query cases. A multimodal approach for medical case–based retrieval that uses the RadLex [9] terminology and 3D texture features extracted from the images of the patients is presented in this paper.

2 VISCERAL Retrieval Benchmark

2.1 Data Set

The retrieval data set was composed of patient scans (3D volumes) and RadLex term lists. The 2311 images in the data set were obtained during clinical routine from three different hospitals [1]. The data set had a heterogeneous collection of images including computed tomography (CT) and magnetic resonance (MR) imaging, contrast enhanced and unenhanced images and various fields of view. The RadLex term lists were generated from the radiology reports of the patients corredponding to the images. They included the affected anatomical structures, their RadLex term ID, the pathologies and their RadLex term ID and if the findings were negated or not in the report. The number of findings and anatomical structures involved varied from case to case.

2.2 Evaluation Phase

Additionally to the data set, 10 query topics were distributed to evaluate the retrieval algorithms from the participants. The participants' algorithms had to find and rank relevant cases from the full data set that could aid in the diagnosis of the query cases. Each query topic included a patient's 3D volume, a manually annotated mask from the main affected organ, region of interest (ROI) from the radiologist's perspective and an anatomy–pathology RadLex term list just as those from the cases in the data set. Participants submitted their rankings and medical experts performed relevance judgments on the submitted cases to determine if they were relevant for the diagnosis of each of the query topics.

3 Methods

The proposed approach to retrieve relevant medical cases was based on a weighting score scheme that combined the RadLex terms anatomical and clinical correlations with the information from local texture features. A single main combination of anatomy and pathology was manually selected from each of the query topics RadLex term lists. This decision was based on the region of interest and organ mask provided in the benchmark to the participants.

[1] http://www.visceral.eu/benchmarks/retrieval-benchmark/, as of 1st May 2015

3.1 Text Retrieval

A medical expert provided a list of correlation–based similarities that would be of interest for finding relevant cases in the data set. The different correlations were computed with the RadLex term lists provided in the retrieval challenge from the radiologic reports. Each similarity feature had a different weight in the final decision for the differential diagnosis and retrieval of cases. The textual similarity between two cases was computed according to the following correlations and their correspondent weighting score (in brackets):

1. Same anatomy with same pathology [0.6]
2. Same anatomy with same pathology negated [0.55]
3. Same anatomy present multiple times [0.2]
4. Same anatomy mentioned once [0.1]
5. Same pathology with different anatomy [0.05]
6. Similar anatomies [0.05]
7. Same imaging modality [0.02]

The weights were defined using a heuristic approach after a review of a subset of the RadLex term lists in the data set from the medical expert. The aim of the weightings is to identify and highlight clinical features that could be relevant for a differential diagnosis and incorporate *a priori* knowledge of the types of image scans contained in the data set. An independent score was generated for each case and the ranking was performed with the sum of all the weights from the different similarity features. To define similar anatomies, a list of correlating RadLex terms (e.g. lung, superior lobe, pleura...) was generated from the standard RadLex term hierarchy [2].

3.2 Visual Retrieval

To perform content–based image retrieval, the texture characterization of the region of interest in the query image is computed using 3D Riesz wavelet coefficients. The images in the data set were previously registered to a common spatial domain in a reference image allowing an indirect comparison of the local texture to all the images from the data set. Using a covariance descriptor a similarity score is obtained that will be used for the ranking of the images. Each of these processing steps are further described in the following section.

Texture Features 3D Riesz filterbanks were used to characterize the local texture properties of the regions of interest in the images. 3D Riesz wavelets have been successful in modeling subtle local 3D texture properties with high reproducibility compared to other methods [15, 13]. The N–th order Riesz transform $\boldsymbol{\mathcal{R}}^{(N)}$ of a three–dimensional signal $f(\boldsymbol{x})$ is defined in the Fourier domain as:

$$\widehat{\mathcal{R}^{(n_1,n_2,n_3)}f}(\boldsymbol{\omega}) = \sqrt{\frac{n_1+n_2+n_3}{n_1!n_2!n_3!}}\frac{(-j\omega_1)^{n_1}(-j\omega_2)^{n_2}(-j\omega_3)^{n_3}}{||\boldsymbol{\omega}||^{n_1+n_2+n_3}}\hat{f}(\boldsymbol{\omega}), \quad (1)$$

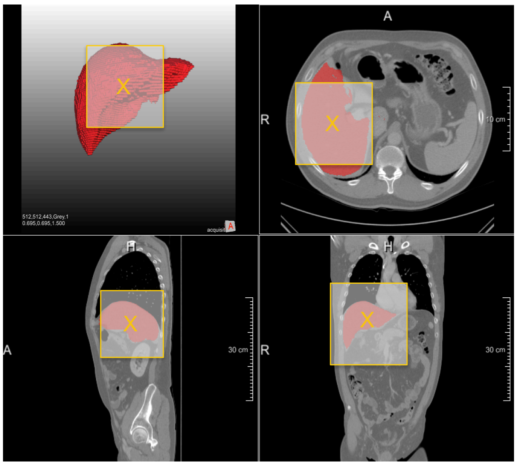

Fig. 1. Sample 3D region of interest selected for the computation of visual features. Shown in red is the binary mask of the main organ affected in the sample query topic. A bounding box is generated containing this organ. The center point of the bounding box is considered as reference to generate a 3D cube around it of $96x96x96$ voxels to compute the Riesz wavelet features. These features will then be compared against the images in the data set

for all combinations of (n_1, n_2, n_3) with $n_1 + n_2 + n_3 = N$ and $n_{1,2,3} \in \mathbb{N}$. Eq. (1) yields $\binom{N+2}{2}$ templates $\mathcal{R}^{(n_1,n_2,n_3)}$ and forms multiscale filterbanks when coupled with a multi–resolution framework.

A single 96x96x96 block was generated using the center of the bounding box surrounding the manually annotated mask of the main organ affected in each the query topics. The obtained texture features were then compared to cases in the data set through the process described in the following sections.

Image Registration An indirect registration from the 3D volume of query topic to the data set images was performed to compute the local texture comparison. A reference image was used to register all the images from the data set and generate a common space domain for visual comparison. Once a new image is provided as a query, it is first registered to the reference image and included in this rough alignment of the data set images. Then, an indirect region of interest was determined in each of the images from the dataset using the same coordinates from the ROI in the query image. The required registrations for this step where computed using the image registration implementation from the Elastix software [7][3]. The quality of the registration is iteratively evaluated in each optimization of a cost function that aims to minimize the normalized cross correlation from the voxel intensities of the transformed moving image to the

[2] www.RadLex.org, as of 1st may 2015

[3] `http://elastix.isi.uu.nl`, as of 1 May 2015.

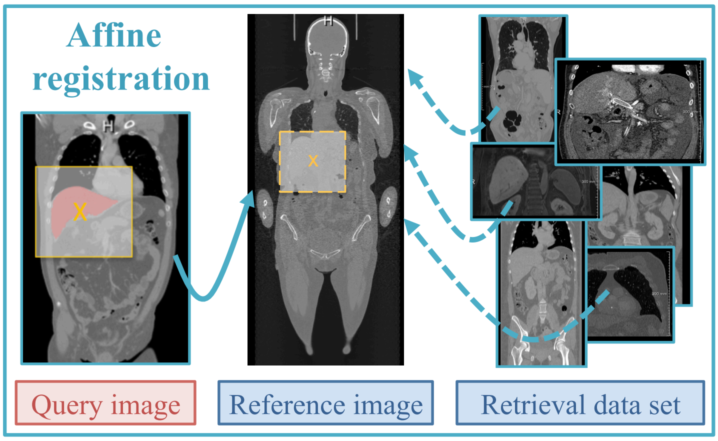

Fig. 2. Finding the region of interest (ROI) from the query image in the data set. The image with the biggest size from the data set, was selected as the reference image. In order to have a common spatial domain to compare the images, all the images from the data set were registered in advance to this reference image using affine registration (dashed blue arrows). With a new query, the query image were also registered to the reference image and the provided binary mask for the ROI (yellow borders) was transformed using the coordinate transformation from the affine registration of the query image. This procedure defined an indirect ROI (dashed yellow borders) in each of the data set images to compare the visual similarities with the query image.

fixed target image, . It uses affine registration that globally aligns the 3D volumes using an iterative stochastic gradient descent optimizer with a multi–resolution approach [6].

Covariance Descriptor In order to represent 3D texture blocks in a compact and accurate notation, we use a 3D Covariance descriptor framework as previously presented in [2]. These descriptors are conceived to translate the different texture patterns found in the aforementioned $96 \times 96 \times 96$ voxels to a common space in which their content can be compared, therefore enabling the evaluation of similarities amongst their inner patterns.

By their construction, Covariance descriptors are suitable for unstructured, abstract texture characterization inside a region, regardless of spatial rigid transformations such as rotation, scale or translations. This is due to a statistical-based representation in which covariance is used as a measure of how several random variables change together – 3D Riesz texture features in our case. Since the distribution of feature variations inside a region is used, rather than the absolute feature values, invariance to region sizes and spatial rigid transformations is achieved. This makes the descriptor robust to rotations or translations, which improves the retrieval accuracy of patient areas in CT images.

In order to formally define the 3D Riesz-Covariance descriptors, we denote a feature selection function $\Phi(ct, v)$ for a given 3D CT volume v as:

$$\Phi(v) = \left\{ \mathcal{R}^{(n_1,n_2,n_3)}_{x,y,z}, \forall x, y, z \in v \right\}, \tag{2}$$

which denotes the set of 6-dimensional Riesz feature vectors, as defined in Eq. 1, obtained at each one of the coordinates $\{x, y, z\}$ contained in the volume v.

Then, for a given region v of the CT image, the associated Covariance descriptor can be obtained as:

$$Cov\left(\Phi(v)\right) = \frac{1}{N-1} \sum_{i=1}^{N} \left(\Phi - \mu\right) \left(\Phi - \mu\right)^T, \tag{3}$$

where μ is the vector mean of the set of feature vectors $\{\Phi_{x,y,z}\}$ within the volumetric neighbourhood made of $N = 96^3$ samples.

The resulting 6×6 matrix Cov is a symmetric matrix where the diagonal entries represent the variance of each Riesz feature, and the non-diagonal elements represent their pairwise covariance, and is used as a discriminative signature of the texture patterns found in the block v. But 3D Riesz-Covariance descriptors do not only provide a representative compactness: they also lie in the Riemannian manifold of symmetric definite positive matrices Sym^+_d. This spatial variety is geometrically meaningful as 3D regions sharing similar texture characteristics will remain under close areas in the descriptor space, and there exist analytical metrics for computing the distance between points of this non Euclidean spatial distribution. Therefore, the similarity between two 3D regions v_1 and v_2 can be computed as the distance of their associated Covariance descriptors C_1 and C_2 via the following *Log-Euclidean* metric [1]:

$$d(C_1, C_2) = \|log(C_1) - log(C_2)\| \tag{4}$$

where $log(C)$ is the matrix logarithm of the symmetric matrix C.

3.3 Multimodal Fusion

It is known from previous medical case–based retrieval benchmarks that the text queries obtain much better results than visual queries [5, 4]. This has been attributed to the currently much more consistent representation of clinical signs in medical images by text labels than by their visual features that are not always very specific. Therefore, it is of high interest to the retrieval information community to find robust visual features that can be combined with semantic terms [8]. To include the information obtained from the visual ranking of the cases into the semantic text weighting scheme, we give an additional weighting if the visual similarity score is high. The additional weight [0.05] is added to the total sum from the textual score of the case if it is in the top 20% of the ranking obtained from the similarity score of the covariance descriptor. These parameters were manually optimized using a small subset of the data set.

4 Results

The proposed method submitted a ranking for each of the query topics (10) in the VISCERAL Retieval benchmark 2015. RadLex–based preliminary results (results computed before obtaining the relevance judgements of participant rankings) were presented at the Multimodal Retrieval in the Medical Domain (MRMD) 2015 workshop, in Vienna, Austria [4]. It obtained the best mean average precision (MAP) result in the benchmark for topics 1,2 and 10, according to these preliminary results presented at MRMD2015.

The final results showed that the method obtained the second best scores from the benchmark in the average over all the query topics. It has also the second best scores from the mixed (textual and visual) retrieval techniques. The mean average precision (MAP), precision after query relevant cases retrieved (Rprec), binary preference (bpref), precision after 10 cases retrieved (P10) and precision after 30 cases retrieved (P30) from our method are shown per query topic in Table 1 and Table 2. The average result of these retrieval metrics when considering all query topics is shown in Table 2. For the complete results and final algorithm comparison with other participants, please refer to [14]. The

Table 1. Runs using the proposed multimodal (text and visual) multimodal retrieval technique.

Metric	01	02	03	04	05
`MAP`	0.2293	0.2227	0.2227	0.2497	0.1949
`Rprec`	0.4576	0.3575	0.3575	0.3106	0.3508
`bpref`	0.5035	0.3466	0.3466	0.4047	0.3542
`P10`	0.2000	0.6000	0.6000	0.8000	0.7000
`P30`	0.5000	0.5333	0.5333	0.8000	0.5667

Table 2. Runs using the proposed multimodal (text and visual) multimodal retrieval technique.

Metric	06	07	08	09	10	**All**
`MAP`	0.3883	0.1780	0.5131	0.1212	0.0467	**0.2367**
`Rprec`	0.4985	0.3483	0.6399	0.1667	0.0851	**0.3572**
`bpref`	0.4912	0.3444	0.6307	0.1580	0.0837	**0.3664**
`P10`	0.9000	0.6000	0.8000	0.3000	0.2000	**0.5700**
`P30`	0.8667	0.7000	0.8000	0.1333	0.1000	**0.5533**

[4] http://www.visceral.eu/workshops/mrmd-2015/, as of 1st May 2015

P10 and P30 results obtained with the method are promising particularly for some of the query topics (e.g. 06 and 08). Almost all query topics have a score in these two metrics above 0.5 and can reach up to 0.9 in P10. There is however, room for improvement in some cases like query topics 09 and 10. An important consideration, is that these results should not be interpreted independently and become meaningful when comparing them against the other retrieval methods.

5 Conclusions

An innovative multimodal (using text+visual information) medical case–based retrieval approach is proposed in this paper. The method is based on a rule–based weighting of the anatomical and clinical RadLex term correlations from radiologic reports. In addition, it also includes state–of–the–art techniques (Riesz wavelets, image registration and covariance descriptors) to compute the similarity between different medical cases through their visual features. The results from the VISCERAL Retrieval benchmark 2015 show that the method can be useful for the retrieval of relevant cases for differential medical diagnosis. Further work is needed to asses the impact of the visual features and obtain more stable retrieval results for different pathologies.

Acknowledgments This work was supported by the EU in FP7 through VISCERAL (318068), Khresmoi (257528) and the Swiss National National Foundation (SNF grant 205320–141300/1).

References

1. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-euclidean metrics for fast and simple calculus on diffusion tensors. Magnetic Resonance in Medicine 56(2), 411–421 (2006)
2. Cirujeda, P., Mateo, X., Dicente, Y., Binefa, X.: MCOV: A covariance descriptor for fusion of texture and shape features in 3d point clouds. In: International Conference on 3D vision (3DV) (2014)
3. Hanbury, A., Müller, H., Langs, G., Weber, M.A., Menze, B.H., Fernandez, T.S.: Bringing the algorithms to the data: cloudbased benchmarking for medical image analysis. In: CLEF conference. Springer Lecture Notes in Computer Science (2012)
4. García Seco de Herrera, A., Foncubierta-Rodríguez, A., Müller, H.: Medical case–based retrieval: Integrating query MeSH terms for query–adaptive multi–modal fusion. In: SPIE Medical Imaging. International Society for Optics and Photonics (2015)
5. García Seco de Herrera, A., Kalpathy-Cramer, J., Demner Fushman, D., Antani, S., Müller, H.: Overview of the ImageCLEF 2013 medical tasks. In: Working Notes of CLEF 2013 (Cross Language Evaluation Forum) (September 2013)
6. Klein, S., Pluim, J.P., Staring, M., Viergever, M.A.: Adaptive stochastic gradient descent optimisation for image registration. International Journal of Computer Vision 81(3), 227–239 (2009)

7. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: Elastix: a toolbox for intensity–based medical image registration. IEEE Transactions on medical imaging 29(1), 196–205 (2010)
8. Kurtz, C., Depeursinge, A., Napel, S., Beaulieu, C.F., Rubin, D.L.: On combining visual and ontological similarities for medical image retrieval applications. Medical Image Analysis (submitted)
9. Langlotz, C.P.: Radlex: A new method for indexing online educational materials. Radiographics 26(6), 1595–1597 (2006)
10. Langs, G., Müller, H., Menze, B.H., Hanbury, A.: Visceral: Towards large data in medical imaging – challenges and directions. Lecture Notes in Computer Science 7723, 92–98 (2013)
11. 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), 1–23 (2004)
12. Rubin, D., Napel, S.: Imaging informatics: toward capturing and processing semantic information in radiology images. Yearbook of medical informatics pp. 34–42 (2010)
13. Jiménez-del Toro, O.A., Foncubierta-Rodríguez, A., Depeursinge, A., Müller, H.: Texture classification of anatomical structures in CT using a context–free machine learning approach. In: SPIE Medical Imaging 2015 (2015)
14. Jiménez-del Toro, O.A., Foncubierta-Rodríguez, A., Müller, H., Langs, G., Hanbury, A.: Overview of the VISCERAL Retrieval Benchmark 2015. In: Multimodal Retrieval in the Medical Domain. Lecture Notes in Computer Science, vol. 9059. Springer (2015)
15. Jiménez-del Toro, O.A., Foncubierta-Rodríguez, A., Vargas Gomez, M.I., Müller, H., Depeursinge, A.: Epileptogenic lesion quantification in MRI using contralateral 3D texture comparisons. In: Medical Image Computing and Computer–Assisted Intervention – MICCAI 2013. Springer Lecture Notes in Computer Science, vol. 8150, pp. 353–360 (2013)