Modality Categorization by Textual Annotations Interpretation in Medical Imaging

Filip I Florea, Alexandrina Rogozan, Abdelaziz Bensrhair, Jean-Nicolas Dacher, Stefan J Darmoni

a Perception, Systems and Information Laboratory, CNRS FRE-2645, INSA & University of Rouen, France
b CISMeF Team, Rouen University Hospital & L@STICS, Medical School of Rouen, France
c QUANTIF Team, Rouen University Hospital & University of Rouen, France

Abstract

Our work is concerned with the automatic indexing of medical images for image retrieval purposes inside a large on-line health-catalogue. We present, in this paper, a rule-based medical-image modality categorization approach. The modality information is important for the indexing of medical images present in on-line health documents (thus, mainly in JPEG format). In fact, contrary to the case of the DICOM images, extensively used in PACS systems, JPEG images have any attached metadata (like those present in DICOM headers). Our system is based on the automatic interpretation of textual annotations, once they are extracted from the medical image by image processing and optical character recognition techniques. The system performances were tested on a medical image database, containing six medical modalities: angiography, ultrasonography, magnetic resonance imaging, standard radiography, computer tomography, and scintigraphy. The extraction of this database from a live health environment ensures the diversity of content from the modality, anatomical region and pathological point of view. In order to determinate the medical image modality, the textual annotations have been interpreted using a set of 96 production rules defined by our expert radiologist. The high precision rate of our categorization system (~90%), proves that textual annotations present in medical images are very reliable indicators of the medical modality. Further work will concern the fusion of our rule-based system with an already implemented visual-content based system, for a robust image modality categorization.

Keywords:
Medical Imaging; Cataloguing; Abstracting and Indexing; Internet; Computer-Assisted Image Processing.

1. Introduction

The CISMeF1 health-catalogue (French acronym of Catalogue and Index of On-Line Health Resources) provides on-line searching capabilities for health-resources. The CISMeF project [1] was initiated in 1995 in order to meet the users’ need to find precisely what they are looking for among the numerous health documents available on-line. CISMeF describes

1 www.chu-rouen.fr/cismef
and indexes the most important documents of institutional health-information in French. We are currently focused on the indexing and retrieval of the medical images present in those documents because the content of medical images attached to the health documents are of crucial significance for information retrieval. Thus, adding medical image retrieval functionality to the Doc’CISMeF research-engine will allow the users (i.e. health professionals, students or general public) to perform image-oriented queries depending on what they are searching for (e.g. “find me all the resources/documents containing CT images” or “find me the resource containing medical images closest to a query-image”).

Indexing medical images present in on-line health documents (e.g. guidelines, teaching material, patient information, and so on) is a complex task. First of all, the images present in these on-line documents are in bitmap format, usually compressed to JPEG, and do not contain any additional metadata, contrary to the DICOM format extensively used in PACS systems. Therefore, it is impossible to benefit of the important amount of information regarding the modality, anatomy, pathology, patient ID, acquisition parameters, usually included in the DICOM image headers. Even if the DICOM images were to be available in on-line medical resources (e.g. pdf, doc, html, xml), the DICOM headers have proven to contain errors, for example for the anatomical region field, error rates of 16% have been reported [2]. This way the correct retrieval of all relevant images, based only on the information extracted from the DICOM headers could be biased.

In our case, the health-documents themselves could contain additional information about the images/figures (e.g. modality, anatomical region or pathology). This information could be extracted from 1). the figure caption (i.e. usually placed near the image) or from 2). the paragraphs citing that figure/image. Most of documents in the CISMeF health catalogue are in pdf format, which is the common format for exchanging printable documents on Internet. Give that the pdf permits only displaying and printing, document structure understanding and image extraction from pdf documents become necessary. However, this is a task of significant complexity, treated recently by several papers, like [3]. Preliminary studies we conducted pointed out considerable difficulties in correctly map the figure with the corresponding caption (e.g. several images could often be find in proximity of only one caption) or the corresponding paragraph (e.g. instead of clear citing like “see Fig.2”, form s like “the above figure” or “the figure seen at page 4” are sometimes used). Furthermore, the correspondence between the figures and additional image information contained in the document was often impossible because of the absence of either the figure caption, the figure numbering (e.g. Fig. 2) or the figure citing by its number (e.g. see Fig. 2).

Consequently, we have concentrated our efforts on image processing and optical character recognition, for extracting additional information from the image content itself, for image retrieval purposes. Due to the fact that the visual content of medical images related to specific anatomical region or pathology is highly dependent on the acquisition modality, we have to extract this modality information before even trying to construct pertinent image descriptors for anatomical regions or pathologies.

The modality categorization step is important for content-based medical image retrieval systems. However, very few research works were concerned with, because the medical image systems, to this day, where mostly concerned with image retrieval inside a certain modality. In [4] the modality categorization problem was first raised, the authors presenting a frame that decides the modality by analysis of so called required and frequently occurring image features. A general structure for semantic medical image analysis [5], and recently, body-region classification results are presented, taking in consideration multiple modalities, but focusing on radiographs [6].

In [7], we presented a medical image modality categorization method based on texture and statistical image descriptors. The performances achieved on a 1332 medical-image database.
representing six main imaging modalities, were significant (more than 90% of classification accuracy). We present, in this paper, another medical image modality categorization approach by rule-base automatic interpretation of textual annotations in medical images. Further work will concern the fusion of this rule-based approach with the previous visual-content based system, for a robust image modality categorization.

2. Material and methods

2.1. The medical image database

The implementation of medical image indexing and retrieval methods requires the constitution of a representative medical image database. A list of medical image modalities used in daily practice was constituted by a medical expert from the Rouen University Hospital (RUH), and implemented in the CISMeF terminology as resource type. Since the beginning of this study, the CISMeF team has developed an exhaustive taxonomy of medical image types (N=65) derived from the MeSH tree of diagnosis imaging.

For the experiments presented in this paper, we consider the six main categories of medical-image modalities: angiography, ultrasonography, magnetic resonance imaging (MRI), standard radiography (RX), CT scan (Computer Tomography), and scintigraphy. Our medical image database contains 1332 images extracted from Radiology, Radio Pediatrie and Nuclear Imaging departments of RUH and of the “Henri Becquerel - Fight against Cancer Center of Rouen”. These images are digitally acquired or secondary digitized, to JPEG format, as they are mostly found in health documents published on the Internet. Thus, when using medical images in educational image repositories, medical image atlases (e.g. for teaching purposes) or in PACS systems (for clinical use), the JPEG compression offers a means to reduce the cost of storage and increase speed of transmission. When JPEG lossy scheme is used, the rate of compression has usually been validated by a medical specialist prior to publication to avoid losing valuable details.

The images do not have the same dimension and quality, being acquired with different digital or analogical equipments, in different hospital services, with different parameters, in a time-period of several years. Moreover, only a fraction of the images presents textual annotations. All this variability increases the difficulty of the modality categorization process. Consequently, efficient modality categorization process requires the extraction of robust and relevant modality information carried not only by the textual annotations, but also by visual-content descriptors.

Figure 1 - The rule-based modality categorization system architecture.

In the next section, we present the medical image modality categorization approach based on automatic interpretation of textual annotations in medical images.
2.2. The rule-based modality categorization system

The architecture of our medical image categorization system (see Figure 1) is reposing on three steps: i) textual annotation extraction of all the text regions from the original medical image; ii) optical character recognition of textual annotations; iii) rule-based automatic interpretation of the recognized annotations. Those annotations are cross-referenced with a list of the most pertinent and discriminative modality annotations (i.e. a set of medical modality production rules, created a priori for each modality, by an expert radiologist).

2.3. The textual annotation extraction

Nowadays, most of digital acquisition equipments used in PACS systems are DICOM-compliant. The DICOM header provides textual meta-information about the acquisition process. However, when the images produced by these equipments are inserted in health documents or published on online databases (and, thus, saved to JPEG format), the DICOM text layers concern with personal information have to be censored (i.e. anonymised) due to legal constraints. The manual anonymization of DICOM text layers is time-consuming, and some PACS are even still lacking automatic anonymization tools. Therefore, medical image professionals use to erase all DICOM text layers before image publication. That is why many images found on online digital libraries or attached to resources in online health-catalogues (e.g. CISMeF) do not have textual annotations directly on the image.

Still, there is an important volume of medical images that are automatically anonymised (striped of the personal information). This way the others information (i.e. important from the radiologist point of view), regarding acquisition parameters, contrast agents and body region specific to the imaging modality acquisition, are still available and visible on the images. The goal of this paper is to evaluate if these textual annotations are representing relevant descriptors of medical imaging modality.

Even though we could expect important text variability because of the different modality acquisition equipments, our experiments showed that the text regions have relatively resembling characteristics (i.e. color, font, thickness) throughout the imaging modalities and acquisition systems. Thus, the characters are white (light grey) or black, have approximately the same size (reported to the image size), and resembling fonts. The characters always form horizontal lines, which are close to the image borders (see Figure 2a).

Therefore, we could extract with high precision the textual annotations from medical images with a TopHat filter [8] set on the character’s thickness (see Figure 2b). The Top Hat Transform is a method that can isolate objects lighter than the neighbourhood and smaller than a structural element conveniently defined. It is computed by subtracting the result of the morphologically open-ing of the image with the structural element \((A \circ B)\) from the original image \(A\), as shown in Equation (1). The structural element \(B\) is dependent of the text size (i.e. the structuring element is a circle with the diameter equal to the character’s thickness).

\[
TopHat \ (A, B) = A - (A \circ B) \tag{1}
\]

Note that for an image with black textual annotations, either the image may be inverted prior to the Top Hat Transform, or the complement of the Top Hat Transform may be used. Also known as the Bottom Hat Transform, this transform is based on a morphological close function \((A \bullet B)\), as shown in Equation (2).

\[
BottomHat \ (A, B) = (A \bullet B) - A \tag{2}
\]

The text color detection is done before applying the transform, by interpretation of the grey-level distribution (taking into consideration that the background corresponds to the extremities of the grey-level histogram and that the background is of opposed color than the
text. Finally the text is obtained by thresholding the Top Hat filtered image (as in Figure 2c).

![Original Image](image1.png) ![TopHat Transform](image2.png) ![Thresholding](image3.png)

**Figure 2 - Extraction of the textual annotations.** a). original image, b). TopHat transform (the filtered text is highlighted in white), c). thresholding the TopHat filtered image

The false detections were removed with additional morphological transformations (i.e. considering the horizontal disposition of text in lines and the connectivity to the image edges). Managing to extract the textual annotations with considerable accuracy, we can now consider the stages of recognition and interpretation.

### 2.4. The optical character recognition and the rule-based interpretation

To rapidly evaluate the efficiency of our approach, a commercial OCR software was used (Abbyy FineReader 7.0 Professional Edition Try&Buy Edition), for the optical recognition of the textual annotations. No specific dictionary was used to help the character recognition process.

For the interpretation of the recognized textual annotations, a set of production rules was defined by a medical specialist. A sample of the full set of 96 rules is presented in Table 1 (e.g. “TR”, “TE” and “TA” are typical annotations for MRIs, and stand for the French “Repetition Time”, “Exposure Time” and respectively “Acquisition Time”). A majority vote decision scheme was used for a precise and robust rule-based modality categorization process.

**Table 1 - Set of medical rules with respect of image modalities.**

<table>
<thead>
<tr>
<th>Annotation</th>
<th>Modality</th>
<th>Annotation</th>
<th>Modality</th>
<th>Annotation</th>
<th>Modality</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEX</td>
<td>MRI</td>
<td>GAIN</td>
<td>ultrasono</td>
<td>ARM</td>
<td>MRI</td>
</tr>
<tr>
<td>TR</td>
<td>MRI</td>
<td>kV</td>
<td>CT</td>
<td>dB</td>
<td>ultrasono</td>
</tr>
<tr>
<td>TE</td>
<td>MRI</td>
<td>mA</td>
<td>CT</td>
<td>FLTR</td>
<td>angiography</td>
</tr>
<tr>
<td>TA</td>
<td>MRI</td>
<td>POST CM</td>
<td>MRI or CT</td>
<td>THALLIUM</td>
<td>scintigraphy</td>
</tr>
<tr>
<td>Tilt</td>
<td>CT</td>
<td>DMSA</td>
<td>scintigraphy</td>
<td>COLLIMATEUR</td>
<td>scintigraphy</td>
</tr>
</tbody>
</table>

### 3. Results

From the 1332 medical images of our database, only a third are containing textual information (see Table 2). Among the 456 images with textual annotations, only 376 were relevant, meaning that only 82% have textual annotations belonging with at least one entry in the set of production rules. Unfortunately, from those 376 images with relevant (and easily readable) annotations, only 221 were correctly recognized by the OCR software.

**Table 2 - Database composition and successful modality recognition**

| No. of images in the database | 1332 |
| No. of images with textual annotations | 456 |
| No. of images with relevant textual annotations | 376 |
| No. of images where the modality was successfully recognized | 221 |

The rule-based modality categorization performances are summarized in Table 3. The precision rates are very high (~99%) implying that when relevant textual annotations are
present and correctly recognized by the OCR, they are very reliable indicators of medical modality. The fact that we do not adapt the OCR software to our application explains the weak categorization recall. Thus, the global recall rate when textual pertinent annotations are present is only of 60%. Failing to recognize some image annotations make impossible the rule-based interpretation of the image modality. Still, the F-measures, computed to show the combined precision/recall performances, are rather good, except for the angiography category. The angiography weak recall caused a significant F-measure drop.

<table>
<thead>
<tr>
<th>Table 3 - Categorization results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Precision</strong></td>
</tr>
<tr>
<td>Angiography</td>
</tr>
<tr>
<td>Ultrasonography</td>
</tr>
<tr>
<td>MRI</td>
</tr>
<tr>
<td>RX</td>
</tr>
<tr>
<td>CT</td>
</tr>
<tr>
<td>Scintigraphy</td>
</tr>
<tr>
<td>Global</td>
</tr>
</tbody>
</table>

Note that in our database, the RX and scintigraphy images do have any textual annotations. The lack of text for the RX is normal, because all the RX images in our database are secondary digitized from film imprints (who usually do not have any annotations; exceptions are represented by L-left R-right orientation markings).

4. Conclusion

Our experiments show that the textual annotations are very reliable indicators of medical imaging modality. However, the fact that these textual annotations are not always available, implies that in real content-based access technologies (i.e. online health catalogues), we can not rely only on a rule-based automatic interpretation approach for modality categorization. The visual content of medical images has also to be taken into account (as we presented in [7]). Furthermore, the rule-base and visual-content modality categorization approaches have to be combined, for a more accurate and reliable medical modality categorization architecture.

5 References


Address for correspondence

Filip-Ionut FLOREA, Perception, Systems and Information Laboratory, FRE CNRS 2645, INSA de Rouen, BP8 - Avenue de l'Université, 76801 Saint-Etienne-du-Rouvray Cedex, Tel: +33 (0) 2 32 95 98 81, mailto:filip.florea@insa-rouen.fr