Exploring Complexity in Health: An Interdisciplinary Systems Approach
A. Hoerbst et al. (Eds.)

doi:10.3233/978-1-61499-678-1-712

Expectation-Driven Text Extraction from Medical Ultrasound Images

Christian REUL[a,1], Philipp KÖBERLE[b], Nurcan ÜÇEYLER[b] and Frank PUPPE[a]
[a] *Chair for Artificial Intelligence and Applied Computer Science, University of Würzburg, Germany*
[b] *Department of Neurology, University of Würzburg, Germany*

Abstract. In this study an expectation-driven approach is proposed to extract data stored as pixel structures in medical ultrasound images. Prior knowledge about certain properties like the position of the text and its background and foreground grayscale values is utilized. Several open source Java libraries are used to pre-process the image and extract the textual information. The results are presented in an Excel table together with the outcome of several consistency checks. After manually correcting potential errors, the outcome is automatically stored in the main database. The proposed system yielded excellent results, reaching an accuracy of 99.94% and reducing the necessary human effort to a minimum.

Keywords. Optical character recognition, text extraction, image processing

1. Introduction

Although the DICOM format provides an efficient and easy way to keep all necessary information within the header of the image file, sometimes important data is stored only within the picture for various reasons. We propose an expectation-driven method to extract such information with a very high precision in the context of an ongoing study supported by this work:

In an attempt to systematically gather ultrasonography data investigating 26 peripheral nerves 225 measurements are performed per subject. The ultrasound device (Siemens Acuson 1000, Erlangen, Germany) used for this task only allows storing data as pixel structure within the corresponding images. Collecting and typing out all results manually is a time-consuming, monotonous, and tiring process bearing a high risk of erroneous entries. To reduce the necessary human effort, while minimizing the error rate, the system described in this work was developed and implemented. Due to the similarity of the images it is possible to highly adapt to the given task by including certain expectations regarding the position and appearance of the information of interest within the image. By making use of this prior knowledge a very high extraction accuracy of 99.94% was achieved. Additionally, a comfortable way to review and, if necessary, correct these intermediate results is provided by an Excel table which shows the extracted results and the corresponding regions of interest (RoI) within the image. Furthermore, several consistency checks are performed and doubtful values are highlighted. In a test run only five out of 6750 values were considered doubtful after having been checked for

[1] Corresponding Author: christian.reul@uni-wuerzburg.de

consistency; four of those turned out to be actual errors. Not a single mistake was overlooked during the consistency checks, showing the efficiency and accuracy of the system.

The remainder of this paper is organized as follows: In section 2 several approaches for similar problems are introduced. Section 3 describes the task at hand and explains the used algorithm. The results are reviewed in section 4, and section 5 concludes the paper.

2. Related Work

Most publications dealing with automatic text extraction from medical images require a very high level of recognition accuracy. Therefore, most proposed systems make use of certain expectations about the position of the RoI or the target text itself. Another strategy is to apply a semi-automatic approach closely supervised by a human user. During the rest of this section some approaches will be briefly described.

Lee et al. [1] created a pipeline to collect and store the T score values from a quantitative computed tomography image to diagnose osteoporosis. A macro program allowed quick and easy correction of standard optical character recognition (OCR) errors like extra spaces or mixed up characters like “7” and “Z”. They applied a very similar approach in [2] to allow radiologists to extract statistical values about a marked RoI displayed on the screen and copy them into a spreadsheet. Due to the completely supervised character of the task, results of close to 100% were achieved in both applications.

Alte and Werner [3] extracted the zoom factor of different ultrasound units from about 45,000 images in order to analyze the intima media thickness. Open source snipping and OCR tools were used, leading to a recognition rate of 98%.

A modality categorization based on textual annotations in medical images was performed by Florea et al. [4]. The text is extracted by making use of prior knowledge about the color and thickness of characters as well as applying standard morphological operations. The rule based interpretation of extracted annotation yielded excellent precision rates (99%) but only mediocre recall rates (60%).

3. Materials and Methods

3.1. Task Description

The goal of this work was to support a still ongoing study dealing with the ultrasonographic investigation of peripheral nerves in the human body. For each subject 225 measurements are performed and stored as a pixel structure in the corresponding ultrasound image. An example image can be seen in Figure 1. The nerve (cross section tracked by yellow line) is analyzed by measuring its circumference (U), diameter (D), and cross-sectional-area (CSA; F). Furthermore, the name of the nerve, probe position and number, and the body side are stored in a separate line (identifier). In Figure 1 “N. medianus OM 2 LI” means that the median nerve was investigated in the middle of the left arm for the second time on the left side. Additionally, two types of measurements are distinguished: CSA and longitudinal diameter (LD). The latter is indicated by an “L” preceding the measurement number. During the study it is intended to collect ultrasonography data of 26 peripheral nerves per subject, aiming at a total of at least 100

subjects. Thus, the results stored in a minimum of 22,500 images have to be extracted and collected in an Excel file.

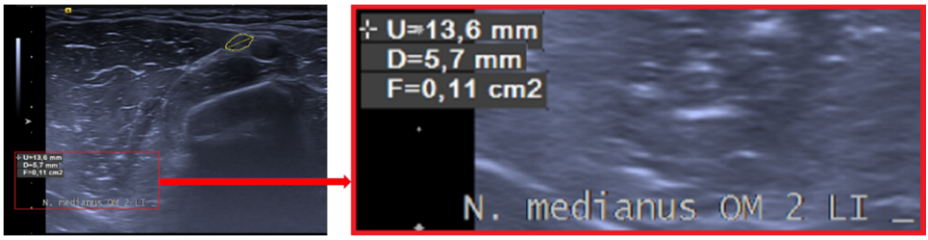

Figure 1. Example of a single measurement.

3.2. Method Overview

The entire workflow is shown in Figure 2. After detecting the RoI the text is separated into lines and converted into a binary image. Next, the OCR takes place; the measurement information is extracted from the recognized text and stored in an Excel sheet. After manual validation the data is transferred into the final Excel workbook.

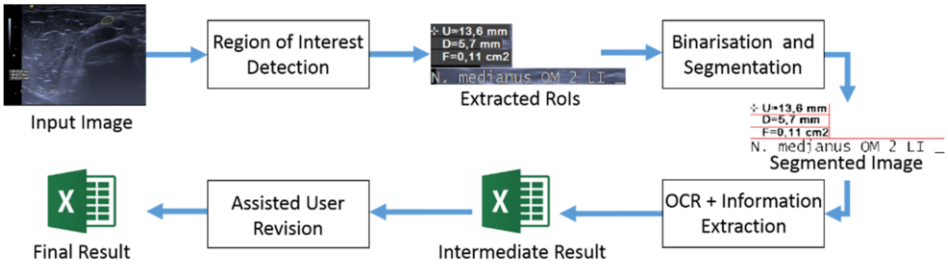

Figure 2. Overview of the entire extraction process.

During the extraction various open source software is used: DCM4Che [5] for extracting the native image and the subject´s name from the DICOM file, the OpenCV library [6] for image processing tasks, TesseractOCR [7] for digitizing the text and Apache POI [8] for creating the Excel documents.

3.3. RoI Detection

Since the position of the identifier is exactly known, the corresponding RoI can be directly extracted. However, the position of the measurement results can vary to a certain degree. As they are always placed on a background patch consisting of pixels of a certain grayscale value, a very accurate RoI detection is still possible. The initial image is converted into a binary image in which “1” is assigned to all the pixels whose grayscale value is equal to the one of the background patches and “0” otherwise. A growing operation is applied to close possible gaps. Finally, the biggest contour is detected in the binary image and considered to be the RoI for the measurement results.

3.4. Binarisation and Segmentation

Analogously to the background patches, the foreground, i.e. the text, always consists of pixels of a certain grayscale value. Therefore, the binarisation process yields accurate

results even when parts of the original image interfere with the text. The detected RoIs are segmented into lines and passed into the OCR engine.

3.5. OCR and Information Extraction

Firstly, the identifier is cleaned by removing special characters like commas and quotation marks which occurred due to non-text pixels randomly having the same grayscale value as the text. Next, the measurement parameters are extracted one by one. For this purpose two lists are used which store all nerve names and all measurement locations which occur in the study. The Levenshtein distance is calculated between the extracted candidates and each item of the corresponding list. The item with the smallest distance is chosen. After determining the measurement type the number and side are assessed as well. If the type is CSA, the area value (F=…) is stored, otherwise (LD) the diameter (D=…). During all steps standard OCR correction techniques are applied. For example, if a number is expected and the detected text is the small version of the letter "L", the number "1" is stored.

3.6. Assisted User Revision

After all extracted parameters had been automatically exported into a structured table (see Figure 3), the user can check for errors. To speed up this process several consistency checks are applied first. For example: if a nerve or location name could not be matched exactly (Levenshtein distance > 0), the corresponding cell is marked yellow. Moreover, a certain sequence of measurement numbers is expected, as there are always three measurements performed per nerve, position, type and side.

Image	Nerve	Location	Type	Side	#	Image	Value
N. suralis IG L3 RE	suralis	IG	LD	RIGHT	3	D=1,1 mm	1,1
N. ischiadicus DOS 1 LI	ischiadicus	DOS	CSA	LEFT	1	F=0,43 cm2	0,43
N. ischiadicus DOS 1 LI	ischiadicus	DOS	CSA	LEFT	1	F=0,58 cm2	0,58
N. ischiadicus DOS 2 LI	ischiadicus	DOS	CSA	LEFT	2	F=0,55 cm2	0,55
N. ischiadicus DOS L1 LI	ischiadicus	DOS	LD	LEFT	1	D=7,5 mm	7,5

Figure 3. Snippet of the reviewing table.

For user convenience the corresponding RoIs from the original images are displayed next to the extracted values. Furthermore, a clickable link to the original image is added to the table, as well as an "Ignore Row" column in which for example the user can mark a duplicate measurement. If the user spots an error, it can be directly edited within the table.

3.7. Final Result

After the reviewing phase the collected data is automatically transferred into the final data table consisting of data regarding nerve, location, type, side, and number.

4. Results and Discussion

For the evaluation five previously unknown healthy subject datasets consisting of 1125 images were used. Six attribute value pairs were extracted as shown in Figure 3. Out of

the 6750 values 6746 were correct. This represents an excellent recognition rate of 99.94%. During the performed consistency checks only five values were highlighted and among these four turned out to be real errors. The last value was a correctly identified nerve name whose Levenshtein distance was greater than 0 due to distortions in the background. Not a single mistake remained undetected during the consistency checks implying a recognition rate of 100% for the 6745 values not highlighted. As it seems highly unlikely that an erroneous value passes the checks undetected, the reviewing process can be sped up significantly by only checking the highlighted values in the intermediate results table. The human effort was dramatically reduced from an estimate of 55 minutes down to below four minutes per subject. It is also viable to assume that the error rate of a human user would be significantly higher.

The four errors are very difficult to avoid due to low quality of the text in the picture. On two occasions the nerve name was divided into two segments. Consequently, the second segment was considered to be the measurement location, therefore causing a total of four errors (2 x nerve name, 2 x location). The consistency check of the system detected these errors. It can be improved to produce two classes instead of currently one class of doubtful results. If the Levenshtein distance is higher than a threshold, the system should show no result at all. As well, checking the number of expected segments can be used to stop further OCR processing in case of an error.

5. Conclusion

A system for the automatic extraction of text from ultrasonography images was developed and validated. Due to making use of prior knowledge and the resulting high degree of adaption to the given task a very low error rate of 0.06% was achieved. No errors occurred when only considering the values that passed the consistency test.

Obviously, the proposed approach currently only allows almost optimal detection rates by making use of very specific constraints. However, it is possible to adapt the main part of the system to other domains. For example, parameters like RoI coordinates or grayscale values of background patches or text can easily be determined by user assisted analysis. Providing an efficient and easy way of doing this will be an interesting and challenging task for the future.

References

[1] Lee, Y., Song, H., Suh, J.: Quantitative Computed Tomography (QCT) as a Radiology Reporting Tool by Using Optical Character Recognition (OCR) and Macro Program. J Digit Imaging; Vol. 25, pp 815-818. 2012.
[2] Lee, Y., Park, E., Suh, J.: Simple and Efficient Method for Region of Interest Value Extraction from Picture Archiving and Communication System Viewer with Optical Character Recognition and Macro Program. Academic Radiology; Vol. 22, Issue 1, pp 113-116. 2015.
[3] Alte, D., Werner, A.: Automatische Texterkennung (OCR) in Ultraschallbildern der A. carotis. 11. Konferenz der SAS-Anwender in Forschung und Entwicklung, Ulm. 2007.
[4] Florea, F. et al.: Modality Categorization by Textual Annotations Interpretation in Medical Imaging. Medical Informatics Europe (MIE 2005); pp 1270-1275. 2005
[5] Open Source Clinical Image and Object Management Library: Homepage. URL: http://dcm4che.org/.
[6] Open Source Computer Vision Library: Homepage. URL: http://opencv.org/.
[7] Tesseract Open Source OCR Engine: Homepage. URL: https://github.com/tesseract-ocr.
[8] Apache POI – Java API for Microsoft Documents: Homepage. URL: https://poi.apache.org/.