

# Quantitative Computed Tomography (QCT) as a Radiology Reporting Tool by Using Optical Character Recognition (OCR) and Macro Program

Young Han Lee · Ho-Taek Song · Jin-Suck Suh

Published online: 8 March 2012

© Society for Imaging Informatics in Medicine 2012

**Abstract** The objectives are (1) to introduce a new concept of making a quantitative computed tomography (QCT) reporting system by using optical character recognition (OCR) and macro program and (2) to illustrate the practical usages of the QCT reporting system in radiology reading environment. This reporting system was created as a development tool by using an open-source OCR software and an open-source macro program. The main module was designed for OCR to report QCT images in radiology reading process. The principal processes are as follows: (1) to save a QCT report as a graphic file, (2) to recognize the characters from an image as a text, (3) to extract the *T* scores from the text, (4) to perform error correction, (5) to reformat the values into QCT radiology reporting template, and (6) to paste the reports into the electronic medical record (EMR) or picture archiving and communicating system (PACS). The accuracy test of OCR was performed on randomly selected QCTs. QCT as a radiology reporting tool successfully acted as OCR of QCT. The diagnosis of normal, osteopenia, or osteoporosis is also determined. Error correction of OCR is done with AutoHotkey-coded module. The results of *T* scores of femoral neck and lumbar vertebrae had an accuracy of 100 and 95.4 %, respectively. A convenient QCT reporting system could be established by utilizing open-source OCR software and open-source macro program. This method can be easily adapted for other QCT applications and PACS/EMR.

**Keywords** Computer in medicine · PACS · OCR · QCT · Reading room

## Introduction

Quantitative computed tomography (QCT) has become a diagnostic tool to evaluate the bone density. QCT has some important advantages over dual-energy X-ray absorptiometry (DXA) and is being increasingly utilized in diagnosis and screening of osteoporosis [1]. Currently, QCT usually outputs the QCT report as a graphic image of the *T* score values, which is not recognizable as a text (Fig. 1). Physicians or radiologists had to read the values on the picture archiving and communicating system (PACS) and typewrite the values on the electronic medical record (EMR) or radiologic report. This situation might be a tiresome task and has some probability of typographical error.

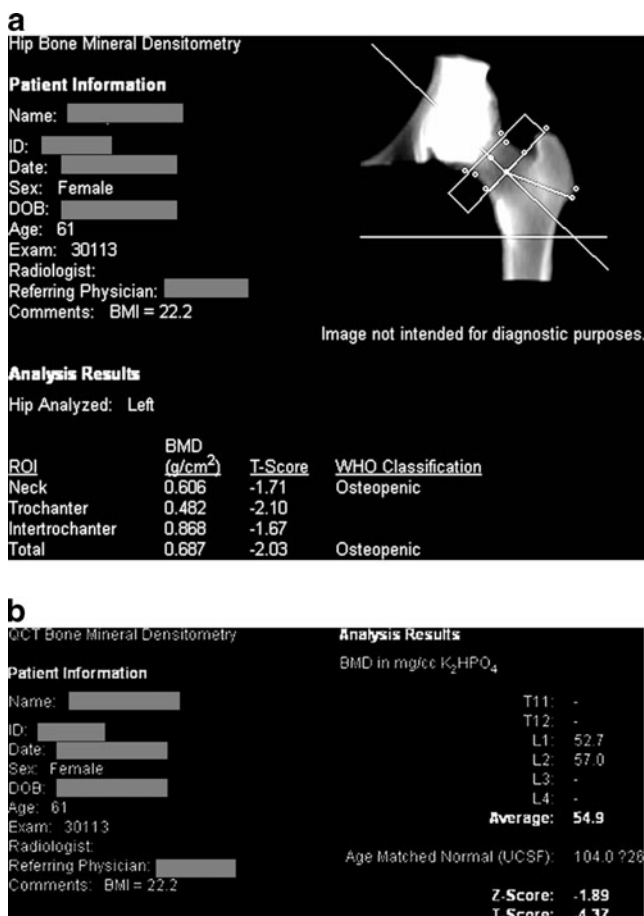
Optical character recognition, usually abbreviated as OCR, is a mechanical or electronic translation of scanned images from typewritten or printed text into machine-encoded text [2]. The OCR can be used for converting imaged values to text data. The OCR-powered radiology reporting system would be useful. In this article, we introduce the development of an automated QCT radiology reporting system by using OCR and macro program and illustrate the practical usages of automating QCT reports in radiology reading environment.

## Materials and Methods

### Hardware and Software

The software we used was Centricity® PACS RA1000 (GE Healthcare, Barrington, IL). The PACS workstation

Y. H. Lee · H.-T. Song · J.-S. Suh (✉)  
Department of Radiology, Research Institute of Radiological Science, Medical Convergence Research Institute, and Severance Biomedical Science Institute, Yonsei University College of Medicine, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-752, Republic of Korea  
e-mail: jss@yumc.yonsei.ac.kr



**Fig. 1** The comparison of two report forms using a QCT application. **a** The image of femoral neck showed the *T* score values with a clear font. **b** The image of vertebrae showed blurry *T* score values

was used (XW6200, Hewlett-Packard, Palo Alto, CA) with 2.8-GHz Xeon processors (Intel, Santa Clara, CA), 3,072 MB RAM, Microsoft Windows XP SP3 with a video graphic array (NVIDIA Quadro FX 1400), and two flat panel LCD devices of 5 megapixels (Totoku Electric, Japan).

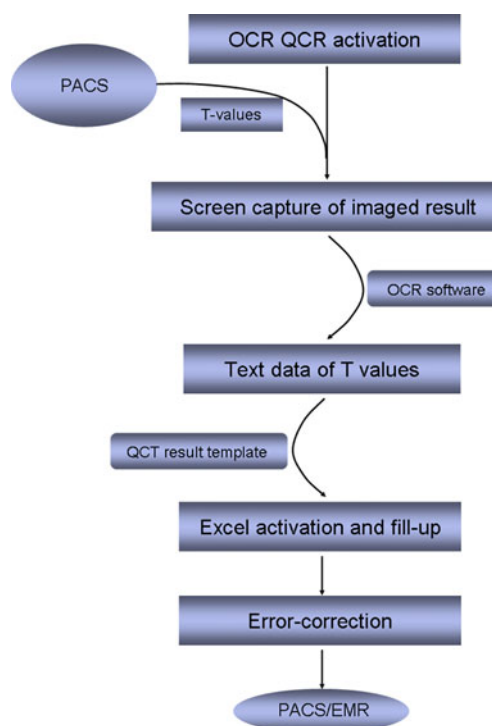
The macro program (AutoHotkey version 1.1.0) was used to design the main module. This open-source macro program is downloadable from the official webpage (<http://www.autohotkey.com>) [3]. GOCR version 0.49 was used for OCR, and this open-source OCR software is downloadable from the official webpage (<http://jocr.sourceforge.net/api>) [4].

The QCT images were acquired by means of Light-Speed VCT64 (GE Healthcare, Milwaukee, WI). A QCT application (QCT PRO version 4.2.3., Mindways Software) generated the QCT result image of the *T* scores of the lumbar vertebrae and femoral neck. The result images were transferred to the PACS server as DICOM files. These result images were reviewed at the PACS workstation.

## Main Module of Radiological Reporting Tools

This reporting system was developed by using OCR software and macro program. The main module was designed to acquire OCR of QCT images in radiology reading process (Fig. 2). All things except OCR were written by macro program AutoHotkey: screen capture, diagnosis according to given *T* score, switching to EMR/PACS, and transferring the values. By one predefined keystroke event, we developed a tool to automate the entire process for converting the QCT images into text reports. The principal processes are described below:

1. Capture of QCT report image. The first step is to save the QCT image (Fig. 1) as a graphic file (JPG format) in a temporary folder of a hard disk drive. For efficiency, we designed the OCR to work within the certain area (i.e., localized box). The location of the *T* score can be fixed by the dedicated template in the PACS.
2. OCR of the image. The OCR engine loads the QCT report saved as an image file, and the characters in the image is converted into a text file (Fig. 3). The text file is saved in a temporary folder.
3. Extraction of *T* scores. The critical values are extracted from the temporary text file. The *T* scores are retrieved by using a specific format pattern and certain position which can be obtained from a string "*T* score."



**Fig. 2** A schematic diagram demonstrates workflow of QCT as a radiology reporting tool by using OCR and macro program

**a**

ROI	(g/cm <sup>3</sup> )	T-Score	WHO Classification
Neck	0.724	-0.64	Normal
Trochanter	0.626	-0.70	
Intertrochanter	1.063	-0.24	
Total	0.843	-0.68	Normal

**b**

ROI	T-Score	WHO Classification
Neck	0.724 -0.64	Normal
Trochanter	0.626 -0.70	
Intertrochanter	1.063 -0.24	
Total	0.843 -0.68	Normal

**Fig. 3** The QCT images and OCR-generated texts QCT report images (a) and OCR-generated texts (b). The text from the primary OCR recognition has some errors such as splitting, segmentation, and alpha-bet vs. numbers

- Correction of the OCR errors. The error correction is automatically performed with the software written by macro using the if-then routine and string operations. Based on the common misrecognition of the OCR engine, the error correction of the common patterns are below:

For corrections of segmentation errors (e.g., “1 .0 2” instead of “1.02”), the macro program deleted the space between the numbers. The extra spaces were replaced with single space. Because the *T* values are the number, when the numbers are recognized by character, the macro program corrects the misrecognition, e.g., “0”/zero/ instead of “O” or “o” /ou/, “8” instead of “B,” and “7” instead of “Z.”

- Making the QCT report with the formal format. The reporting format is generated from the QCT values and the patient information. The diagnosis of osteopenia, osteoporosis, or normal bone mineral density based on the *T* score can be inserted into the QCT radiology reporting template.

Predefined templates of radiologic QCT report are below:

*T score of the femur is [T score1], compatible with [diagnosis1].*  
*T score of the spine is [T score2], compatible with [diagnosis2].*

The [diagnosis1] and [diagnosis2] can be determined by the *T* score according to World Health Organization guideline: *T* score -1.0 or greater is “normal,” *T* score between -1.0 and -2.5 is “low bone mass” (or “osteopenia”), and *T* score -2.5 or below is osteoporosis [5, 6]. The macro program can fill up the [T score1] and [T score2] with the templates, and copy to clipboard of Windows.

- Export of the QCT report. The QCT report is exported and pasted into the EMR or PACS (Fig. 4). This step is achieved by using the macro program: Window switching



**Fig. 4** Screenshot of the QCT reports. This report can be automatically generated with QCT-OCR reporting tools

to EMR or PACS window, focusing the cursor on the input textbox, and sending the key [control V] for paste.

### Accuracy Evaluation

The accuracy of OCR was evaluated from 500 randomly selected QCT images between May 2011 and June 2011. The OCR-recognized *T* scores from the reporting system were compared for accuracy with *T* scores which is eye recognized by two musculoskeletal radiologists. The confidence intervals of accuracies were calculated.

### Results

The QCT by using open-source OCR and macro program acted successfully as a radiology reporting tool that used OCR to process QCT which evaluates the bone mineral density of the femoral neck and lumbar vertebrae. The macro inputs the QCT report into the EMR or PACS by automating the QCT reporting steps of reading the *T* scores from the QCT report (Fig. 4). The diagnosis of normal, osteopenia, or osteoporosis is also determined according to *T* scores. In accuracy evaluation of *T* scores of the OCR and eye recognition, the results of the femoral neck and lumbar vertebrae *T* scores had an accuracy of 100 % (confidence interval 1.0–1.0, confidence interval;  $p < 0.01$ ) and 95.4 % (confidence interval 0.9356–0.9724;  $p < 0.01$ ), respectively.

### Discussion

For the development of this QCT reporting tool, we used the OCR software GOCR and a macro program AutoHotkey.

The GOCR is a widely used OCR engine for personal computer and smart phone. By using an OCR program, a more accurate task can be accomplished with the primary mechanical recognition performed by OCR and the secondary eye recognition verification.

The macro program AutoHotkey is an open-source application for computing environment and is also applied in the field of radiology for more convenient radiology reading environment [7–9]. By using this macro, we made an automation of QCT reporting system. Also the macro program was utilized for the error correction. The accuracy of text recognition depends on the OCR application and the image quality. One should consider the optimal QCT image size because GOCR usually recognize image letters with 20–80 pixel size. In terms of text recognition of the *T* values visualized as numbers, the number figures were well recognized except for the some cases. These misrecognitions can be automatically corrected with macro-coded error correction. Moreover, they might be further improved by applying better algorithm.

Our method looks similar to the method in radiologic dosage report [10]. However, there are some differences: (1) we used the different OCR engine which is an open-source software and commonly used for development, and (2) we also used the open-source macro program AutoHotkey as a main program. (3) While the radiologic dosages are saved in the Excel file, the QCT reports of QCR report system are transferred to EMR or PACS as a text. Our method of open-source software can be applied in other field of the medical record system.

In our study, the accuracies of OCR were 100 % for femoral neck and 95.4 % for lumbar vertebrae. The different results were resulted from the different fonts of the QCT report images. In a QCT application we used, the report was generated with blurry fonts (Fig. 1). In terms of the vendor issue, our QCT software (QCT PRO, Mindways software) produced the different formats of QCT reports: One is clear

font and an easily OCR-recognized font (Fig. 1a) and the other is a blurred font (Fig. 1b). The latter needs to be error corrected. We have asked the vendor the problem. We expect the vendor of the QCT application would update the report form with clearer fonts to resolve the font variation. And this point may be a factor for the future development of the QCT software.

In conclusion, the convenient QCT reporting system could be established by utilizing open-source OCR software and open-source macro program. This method can be easily adapted for other QCT applications and PACS/EMR.

## References

1. Adams JE: Quantitative computed tomography. *Eur J Radiol* 71:415–424, 2009
2. Optical character recognition . Wikipedia. Available at [http://en.wikipedia.org/wiki/Optical\\_character\\_recognition](http://en.wikipedia.org/wiki/Optical_character_recognition). Accessed 24 Jun 2011
3. AutoHotkey Web site. Available at <http://www.autohotkey.com>. Accessed 24 Jun 2011
4. GOCR Web site. Available at <http://jocr.sourceforge.net/api/> Accessed 24 Jun 2011
5. Kanis JA, Melton 3rd, LJ, Christiansen C, Johnston CC, Khaltaev N: The diagnosis of osteoporosis. *J Bone Miner Res* 9:1137–1141, 1994
6. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis. Report of a WHO Study Group. *World Health Organ Tech Rep Ser* 843:1–129, 1994
7. Lee YH: Efficient radiologic reading environment by using an open-source macro program as connection software. *Eur J Radiol* 81:100–103, 2012
8. Siström C, Honeyman-Buck J: A simple method for importing multiple image files into PowerPoint. *AJR Am J Roentgenol* 182:1591–1596, 2004
9. Yam CS: Simple method for inserting Flash movies into PowerPoint presentations. *AJR Am J Roentgenol* 188:W374–W378, 2007
10. Li X, Zhang D, Liu B: Automated extraction of radiation dose information from CT dose report images. *AJR Am J Roentgenol* 196:W781–W783, 2011