System and protocol for collecting subjective data on clinical quality of compressed DICOM images

Luana Gonçalves*, Ronaldo de Freitas Zampolo**, Leonardo Miranda de Brito*,
Valquíria Gusmão Macedo *, Fernando Antonio Silva Rocha***
*Department of Electrical and Biomedical Engineering
**Department of Computer and Telecommunications Engineering
Federal University of Pará, Belém, Brazil
***Vila Rica Medicina Diagnóstica, Santos, Brazil
Email: zampolo@ufpa.br

***Abstract*—Compression in medical images has the potential to reduce costs in data transmission and storage in Picture Archiving and Communications Systems (PACS). In this case, the availability of adaptive schemes for codec tuning would be worthwhile for maximising compression and, at the same time, avoiding quality loss, considering the diagnostic value of such images. However, ordinary visual quality metrics are not usually developed considering this specific context. In this paper, we present a system and a protocol for subjective data acquisition, which aims at the modelling of the quality assessment criteria applied by experts in the preliminary phase of a medical analysis. The system is implemented by using Python/PsychoPy, and the experimental protocol following ITU recommendations and criteria defined in the European Guidelines on Quality Criteria for Computed Tomography. Snapshots of the graphic user interface are exhibited, but emphasis is given on the proposed experimental procedure.**

I. Introduction

Medical imaging technologies are increasingly important for diagnosis and treatment of various pathologies, reinforcing the demand for more effective methods of transmission and storage of medicals images. At present, a huge volume of medical images are produced every day in hospitals and radiological centres, requiring larger storage facilities. In addition, if such images need to be transmitted, the volume of data might be an issue, considering eventual limitations of bandwidth in the communication channel. The use of compression algorithms might reduce both storage requirements and transmission time of medical images, leading to decreasing costs concerning implementation and operation of a Picture Archiving and Communications System (PACS). When compared with the lossless techniques, lossy compression algorithms provide higher compression at the expense of signal fidelity. Note that lossy algorithms can be adopted, since visual relevant information is selected and preserved, in such a way to allow the reconstructed image has the same diagnostic accuracy as the original medical image. Aware of this context, the American College of Radiology and National Equipment Manufacturers Association (ACR/NEMA) establishes that lossy compression schemes, such as JPEG, JPEG-LS, JPEG-2000 or MPEG, can be used with DICOM images, in the condition of keeping the same diagnostic accuracy or clinical quality [1]. The ACR/NEMA coins the term *diagnostically acceptable irreversible compression* (DAIC) as mathematically irreversible (lossy) compression that does not affect a particular diagnostic task [2]. Currently, DAIC can be used by a qualified physician to decrease the size of medical images without reduction in their clinical diagnostic features. Thus, the visual quality of medical images has a decisive role in diagnostic and represents the main constraint to any kind of processing that may be applied to such images. Although the clinical quality assessment given by qualified physicians is the most reliable way to evaluate the quality of medical images, such a method is expensive and time consuming. Visual quality metrics would be an interesting alternative as they are faster and can be embedded in systems. However, automatic parameter determination of compression algorithms to the specific context of medical images is difficult when using non specialised visual quality metrics. This happens because visual quality metrics are commonly developed from experiments whose test images are composed of ordinary scenes (nature, people, faces, animals, etc.), subjects are regular users (no experts in medical images), and the procedure of evaluation is quite different of those in clinical practice. There are few works on the objective characterisation of the clinical or diagnostic quality of medical images, among which we highlight [3]. In this paper, we propose a system and an experimental protocol for collecting subjective data on diagnostic quality of compressed medical images, following ITU recommendations [4] and criteria defined in the European Guidelines on Quality Criteria for Computed Tomography [5]. The aim is to obtain a database that can be used for comparing the performance of visual quality metrics as well as supporting the development of new metrics specific for medical images. The remainder of this text are organized as follows. Section II presents the quality criteria for medical images adopted in this work. The windowing process for medical images are discussed in Section III. Section IV shows the proposed methodology and system. Conclusions are given in Section V.

II. Quality Criteria for Diagnostic Images

In comparison with conventional radiology, the relative complexity of settings in CT may adversely influence the image quality. Optimal use of medical imaging technologies with ionising radiation involves the following three important aspects of the imaging process: diagnostic quality of the image, radiation dose to the patient, and choice of examination technique. In this context, the European Study Group of Radiologists and Physicists conducted a study to investigate overall principles associated with good imaging technique and to define quality criteria for CT, which would result in images with the necessary clinical information, keeping radiation dose to the patient at a minimum for different examination techniques. These quality criteria can be used by radiologists, qualified medical physicists or radiologic technologist as a checklist on routine evaluation of every medical imaging process. Quality criteria are useful for the prompt checking of the imaging quality, while the patient is still in the scanner during acquisition procedure. However, the same quality criteria cannot be applied to all examination techniques because for certain clinical indications a lower level of image quality may be acceptable. Then, for each kind of CT examination, a list of quality criteria is given. Diagnostic requirements are basically a list of visual criteria which specifies important anatomical structures that must be clear and well defined in a medical image to allow accurate diagnosis. Thus, the assessment of a medical image quality takes into account both the anatomy of the area under examination and the contrast between different tissues, being the latter essential for the detection of pathological changes. Such criteria for CT, in which the level of visualization or critical reproduction of essential anatomical features are listed, are defined in the European Guidelines on Quality Criteria for Computed Tomography [5]. In Table I, we reproduce the criteria regarding head CT images for different windowing choices (soft tissue and bone), just to give a practical example.

TABLE I
QUALITY CRITERIA FOR HEAD COMPUTED TOMOGRAPHY EXAMS

	Critical reproduction of Soft Tissue
1	Visually sharp reproduction of the border between white and grey matter
2	Visually sharp reproduction of the basal ganglia
3	Visually sharp reproduction of the ventricular system
4	Visually sharp reproduction of the cerebrospinal fluid space around the mesencephalon
5	Visually sharp reproduction of the cerebrospinal fluid space over the brain
6	Visually sharp reproduction of the great vessels and the choroid plexuses after intravenous contrast media
	Critical reproduction of Bone
1	Visually sharp reproduction of the cortical and trabecular bone structures
2	Visually sharp reproduction of the air filled compartments
3	Visually sharp reproduction of the sella turcica
4	Visually sharp reproduction of the cerebellar contours
5	Reproduction of the border between the white and grey matter (cerebellum)
6	Visually sharp reproduction of the cerebrospinal fluid space around the brain stem
7	Visually sharp reproduction of the great vessels and choroid plexuses after intravenous contrast media

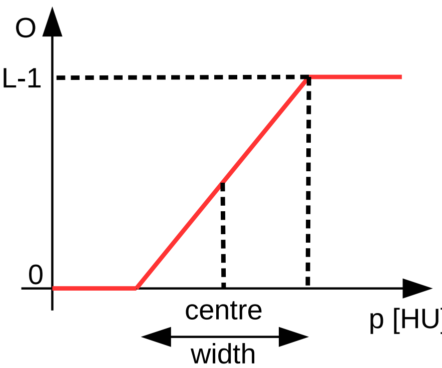


Fig. 1. Example of a windowing function: p [HU], and O denote the values of pixels before (in Hounsfield Units) and after windowing procedure, respectively; the interval [0, L-1] represents the exhibition device range from black to white.

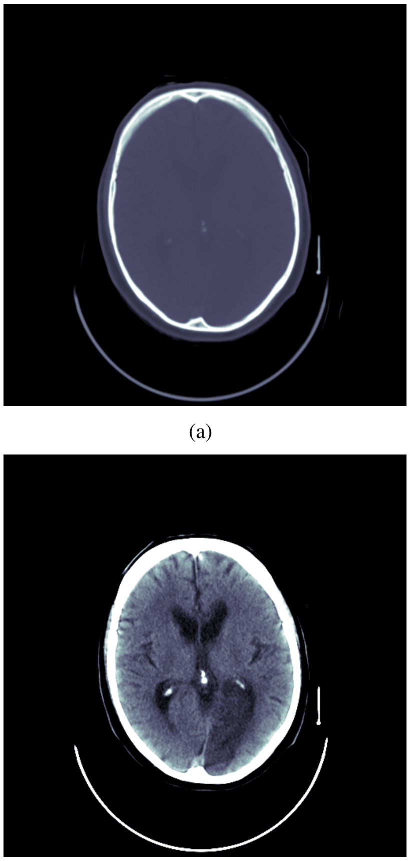
III. WINDOWING

Windowing, in the context of diagnostic images, consists in a pixel intensity transformation, where the range of actual pixel values corresponding to the whole grey scale from white and black is adjusted according to the anatomical structure of interest (Fig. 1).

Two reasons justify the necessity of windowing for viewing DICOM images. First, medical images have pixel depth greater than 8 bits, requiring a mapping function to make pixel values consistent with display device capability. Second, the choice of windows with different values of centre and width allows to emphasize different types of tissues. Air, water, bone and soft tissue, for instance, have windows previously defined according to the scale Hounsfield Unit (HU) [6]. In Fig. 2, one can see two results after applying different windows to the same DICOM image.

IV. PROPOSED SYSTEM AND PROTOCOL

The proposed system for collecting subjective clinical quality of compressed DICOM images was conceived taking into account the criteria pointed out in [5]. A pilot experiment was conceived to evaluate and improve the initial proposed protocol. In this, the subject cannot modify the exhibited test image, and instead of providing all slices produced by a given exam, as it would be the case in a regular quality evaluation, just some slices related to 2 previously chosen criteria (see Table II) are used as original test images. Such criteria were chosen because they were pointed out by the expert in radiology of our research team as the most meaningful among the others for head CT (Table I). In an actual diagnostic centre, immediately after the acquisition step, either a technician or a physician analyses specific parameters of the images from an exam, such as degree of precision, and regularity of boundaries between anatomical structures. Analysis of such



(a)

(b)

Fig. 2. Results after windowing a DICOM image. Window centre and width were set to emphasise (a) bone, and (b) soft tissue.

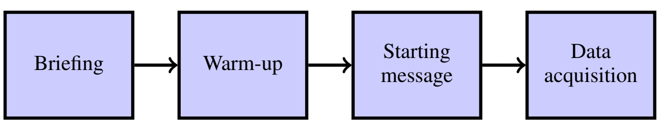


Fig. 3. Proposed experimental protocol.

aspects indicates whether medical images present artefacts, noises or distortions that could prevent correct diagnosis. In this case, the acquisition phase must be repeated. Otherwise, those images are sent to the next stage, where anatomical structures are evaluated and a corresponding clinical report is written. Considering that clinical analysis is performed at the same place or nearby acquisition premisses, no image compression is really necessary. However, if the whole exam needs to be, at some moment, either transmitted or stored, image compression might play an important role. In order to be useful, compression must not result in lost of clinical quality. Python language was used in system implementation as well as the open source application PsychoPy [7], [8], which has several resources for the design and execution of experiments in neuroscience, psychology and psychophysics. The designed experimental protocol can be divided into the steps illustrated in Fig. 3. First of all, some guidelines are presented, regarding

TABLE II
QUALITY CRITERIA USED FOR HEAD COMPUTED TOMOGRAPHY IMAGES

	Critical reproduction of soft tissue
1	Visually sharp reproduction of the border between white and grey matter
2	Visually sharp reproduction of the basal ganglia
	Critical reproduction of bone
1	Visually sharp reproduction of the cortical and trabecular bone structures
2	Visually sharp reproduction of the air filled compartments

TABLE III
RATING CATEGORIES USED TO EVALUATE CLINICAL IMAGES

Rating	Appearance of structure
1	Not visible
2	Poorly reproduced
3	Adequately reproduced
4	Very well reproduced

the experimental procedure and its objectives. Next, a warm-up phase familiarises the subject to the system interface. Then, a starting message is displayed indicating the beginning of the actual data collection phase, where the test images are presented randomly, one at a time. For each displayed test image, the subject must rate the criteria in Table II (depending if the windowing procedure was set to emphasise bones of soft tissue), according to the categories in Table III [9].

The test set for this pilot experiment is defined from 10 original head CT DICOM images, 5 of them with abnormal lesions and the other 5 without any pathological structures, all in axial plane. For each original image, two versions were produced by using windowing procedure to emphasise bone and soft tissue. The choice of bone and soft tissue windows is justified because they are the standard options in the analysis of head CT images. The size of the slices are originally 512×512. Following a common procedure in daily practice, windowed images were resized to 1024×1024 by using bicubic interpolation. Two compressing schemes, among those allowed in DICOM recommendations, were chosen in this preliminary phase: JPEG and JPEG2000. The test set is created by coding every windowed and resized image with both coding schemes (JPEG and JPEG2000) at different quality factors (10, 20, 30, 50 and 100). For the purpose of this clinical quality assessment, subjects must be either medical image experts, radiologic physicists or radiologic technologists. Each trial of the experiment comprises the evaluation of 180 images. Although time limit for analysis is not fixed, tentative tests show that the average time for one single trial is around 20 minutes.

Although the first version of the system for collecting experimental data is already implemented and available, the experimental protocol concerning actual data acquisition (clinical quality assessment from experts) is currently in final preparation to be submitted to the ethics committee of authors' university for analysis and consent. All the images used in this

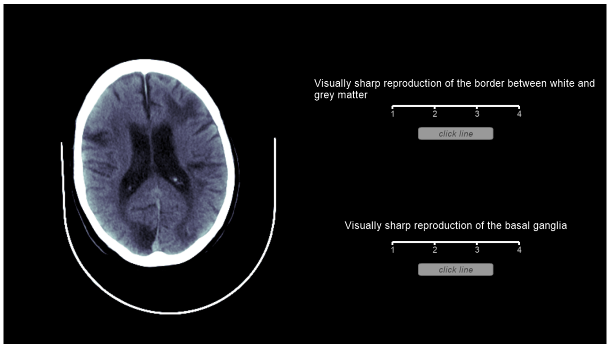


Fig. 4. Graphical user interface of the implemented system for assessing clinical quality of compressed medical images.

text are illustrative and available in public online databases.

V. Conclusion

In this paper, an experimental system and protocol for clinical quality assessment of compressed medical images are presented. The general concept is currently applied just for the case of head tomography and follows the recommendations of the European Guidelines on Quality Criteria for Computed Tomography. This proposition is part of a research in progress intended to allow performance assessment and development of visual quality metrics for medical purposes. Next steps include experimental protocol submission to ethics committee for analysis and consent, collection of clinical quality assessment data, data analysis, and performance evaluation of different visual quality metrics. In the near future, extensions to other types of CT images and modalities of medical imaging systems are intended. The software source code of this project is available at https://github.com/luaffjk/AQVIM, where interested parties can obtain the latest stable version, contribute with suggestions or even collaborate in the development of the system.

Acknowledgment

The authors would like to thank CNPq, FAPESPA and UFPA, which support partially this project by their corresponding Scientific Initiation Programmes (PIBIC).

References

[1] M.-M. Sung, H.-J. Kim, E.-K. Kim, J.-Y. Kwak, J.-K. Yoo, and H.-S. Yoo, "Clinical evaluation of JPEG2000 compression for digital mammography," *IEEE Transactions on Nuclear Science*, vol. 49, no. 3, pp. 827–832, Jun 2002.

[2] P. Bak, M. Claudon, N. Dugar, D. Harvey, P. H. M. Kmmerer, D. Koff, E. Kotter, T. Kronander, R. Loose, C. Lundstrm, J. Martin, U.-E. Martin, P. Mildenberger, D. Mller, J. Reponen, A. Rowberg, L. Sutton, D. Taubman, and A. Tzannes, "Usability of irreversible image compression in radiological imaging. a position paper by the european society of radiology (esr)," *Insights Imaging*, vol. 2, no. 1, p. 103115, Dec 2011.

[3] C. Cavaro-Menard, L. Zhang, and P. L. Callet, "Diagnostic quality assessment of medical images: Challenges and trends," in *2010 2nd European Workshop on Visual Information Processing (EUVIP)*, July 2010, pp. 277–284.

[4] ITU-T, *Tutorial – Objective Perceptual Assessment of Video Quality: Full Reference Television*, International Telecommunication Union Geneva, Switzerland, 2004.

[5] E. Commission, "European Guidelines on Quality Criteria for Computed Tomography ," http://www.drs.dk/guidelines/ct/quality/htmlindex.htm, 1999, [Online; accessed 08-September-2017].

[6] W. Birkfellner, Ed., *Applied Medical Image Processing: a Basic Course*. Taylor and Francis, 2014.

[7] J. W. Peirce, "Psychopypsychophysics software in Python," *Journal of Neuroscience Methods*, vol. 162, no. 12, pp. 8 – 13, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165027006005772

[8] J. Peirce, "Generating stimuli for neuroscience using PsychoPy," *Frontiers in Neuroinformatics*, vol. 2, p. 10, 2009. [Online]. Available: http://journal.frontiersin.org/article/10.3389/neuro.11.010.2008

[9] A. D. Crop, P. T. V. H. Klaus Bacher, P. V. Smeets, B. S. Smet, M. Vergauwen, U. Kiendys, P. Duyck, K. Verstraete, K. DHerde, and H. Thierens, "Correlation of contrast-detail analysis and clinical image quality assessment in chest radiography with a human cadaver study," *Radiology*, vol. 262, no. 1, pp. 98–304, Jan. 2012.