A COMPREHENSIVE QUALITY EVALUATION SYSTEM FOR PACS

Dinu Dragan, Dragan Ivetic
Department for Computing and Automatics, Republic of Serbia
dinud@uns.as.rs, ivetic@uns.ac.rs

ABSTRACT
An imposing number of lossy compression techniques used in medicine, represents a challenge for the developers of a Picture Archiving and Communication System (PACS). How to choose an appropriate lossy medical image compression technique for PACS? The question is not anymore whether to compress medical images in lossless or lossy way, but rather which type of lossy compression to use. The number of quality evaluations and criteria used for evaluation of a lossy compression technique is enormous. The mainstream quality evaluations and criteria can be broadly divided in two categories: objective and subjective. They evaluate the presentation (display) quality of a lossy compressed medical image. Also, there are few quality evaluations which measure technical characteristics of a lossy compression technique. In our opinion, technical evaluations represent an independent and invaluable category of quality evaluations. The conclusion is that quality evaluations from each category measure only one quality aspect of a medical image compression technique. Therefore, it is necessary to apply a representative(s) of each group to acquire the complete evaluation of lossy medical image compression technique for a PACS. Furthermore, a correlation function between the quality evaluation categories would simplify the overall evaluation of compression techniques. This would enable the use of medical images of highest quality while engaging the optimal processing, storage, and presentation resources. The paper represents a preliminary work, an introduction to future research and work aiming at developing a comprehensive quality evaluation system.

Keywords: medical image quality metrics, medical image compression, PACS

1 INTRODUCTION

Picture Archiving and Communication System (PACS) represents an integral part of modern hospitals. It enables communication, storage, processing, and presentation of digital medical images and corresponding data [1]. Digital medical images tend to occupy enormous amount of storage space [2, 3]. The complete annual volume of medical images in a modern hospital easily reaches hundreds of petabytes and is still on the rise [4]. The increased demand for digital medical images introduced still image compression for medical imaging [5], which relaxes storage and network requirements of a PACS, and reduces the overall cost of the system [3].

In general, all compressed medical images can be placed in two groups: lossless and lossy. The first group is more appealing to physicians, because decompression restores the image completely, without data loss. It achieves modest results and maximum compression ratio of 3:1 [6, 7, 8]. Several studies [9, 10] showed that this is not suitable for PACS, and that at least 10:1 compression ratio has to be achieved.

The second group of compression techniques achieves greater compression ratios, but with data distortion in restored image [6, 7, 8]. Lossy compression provoked serious doubts and opposition from medical staff. The opposition rose from the fact that the loss of data can influence medical image interpretation and can lead to serious errors in treatment of a patient. Therefore, the main research area for lossy compression of medical images is finding of the greatest compression ratio that still maintains diagnostically important information. The degree of lossy compression of medical images which maintains no visual distortion under normal medical viewing conditions is called “visually lossless” compression [10]. Several studies [8, 11, 12] and standards [13] proved clinical acceptability to use lossy compression of medical images as long as the modality of the image, the nature of the imaged pathology, and image anatomy are taken into account during lossy compression. The medical organization involved has to approve and adopt a lossy compression of medical images applied in PACS. Therefore, it is necessary to provide a
quality evaluation of different compression techniques from PACS point of view.

During our work on a PACS for a lung hospital, we tried to adopt image compression for medical images which achieves highest compression ratio with minimal distortion within decompressed image. Also, we needed image compression suitable for telemedicine purposes. We consulted the technical studies in search for quality evaluation of image compression technique. The sheer amount of studies is overwhelming [14, 15]. There is no unique quality evaluation which is suitable for various compression techniques and different applications of image compression [16, 17]. In most cases the studies are focused only on presentation (display) quality of the lossy compressed medical image. Technical features of compression technique are usually ignored.

This paper represents a preliminary research. Its purpose is to identify all the elements needed to evaluate the quality of a compression technique for PACS. We identified three categories of quality evaluations and criteria: presentation-objective, presentation-subjective, and technical-objective. Overview of technical studies led us to conclusion that quality evaluations from each category measure only one quality aspect of an image compression technique. To perform the complete evaluation of medical image compression technique for PACS, it is necessary to apply a representative of each category. A correlation function between the representatives of each category would simplify the overall evaluation of compression techniques. A 3D evaluation space introduced by the paper is a 3D space defined by this correlation function and quality evaluations used. Our goal is to develop an evaluation tool based on the 3D evaluation space which is expected for 2011. All the elements of the quality evaluation system are identified in the paper.

The organization of the paper is as follows: section 2 gives the short overview of the lossy compression techniques used in medical domain; section 3 describes the quality evaluations used to measure the quality of compression techniques; 3D evaluation space is discussed in section 4; section 5 concludes the paper.

2 LOSSY COMPRESSION OF MEDICAL IMAGES

Over the past decades an imposing number of lossy compression techniques have been tested and used in medical domain. Industry approved standards have been used as often as the proprietary compressions. On the part of the image affected, they can be categorized in two groups:

1. medical image regions of interest (ROI) are compressed losslessly while the rest of the image background is compressed lossy,

2. the entire medical image is compressed lossy targeting the “visually lossless” threshold.

The first group offers selective lossy compression of medical images. Parts of the image containing diagnostically crucial information (ROI) are compressed in a lossless way, whereas the rest of the image containing unimportant data is compressed lossy. This approach enables considerable higher compression ratio than ordinary lossy compression [18, 19]. Larger regions of the medical image contain unimportant data which can be compressed at higher rates [19]. Downfall of this approach is computational complexity (an element of technical-objective evaluation). Each ROI has to be marked before compression. Even for images of the same modality, ROIs are rarely in the same place. ROIs are identified either manually by qualified medical specialist or automated based on a region-detection algorithm [20]. The goal is to find a perfect combination of automated ROI detection algorithms and selective compression technique.

Over the years various solutions for ROI compression of medical images emerged which differ in image modalities used, ROI definitions, coding shames and compression goals [20]. Some of them are: a ROI-based compression technique with two multi-resolution coding schemes reported by Strom [19], a block based JPEG ROI compression and a importance schema coding based on wavelets reported by Bruckmann [18], a motion compensated ROI coding for colon CT images reported by Bokturk [21], a region based discrete wavelet transform reported by Penedo [22], a JPEG2000 ROI coding reported by Anastassopoulos [23].

The second group of lossy compression techniques applies lossy compression over entire medical image. Considerable efforts have been made in finding and applying the visual lossless threshold. Over the years various solutions emerged which differ in goals imposed on a compression technique (for particular medical modality or for a group of modalities), and in compression techniques used (industry standards or propriety compression techniques).

Some of the solutions presented over the years are: a compression using predictive pruned tree-structured vector quantization reported by Cosman [17], a wavelet coder based on Set Partitioning in Hierarchical Trees (SPIHT) reported by Lu [24], a wavelet coder exploiting Human Visual System reported by Kai [25], a JPEG coder and wavelet-based trellis-coded quantization (WTCQ) reported by Slone [10], a JPEG2000 coder reported by Bilgin [26].

Although the substantial effort has been made to develop a selective lossy compression of medical images, the industry standards that apply lossy compression on the entire medical image are commonly used in PACS.
3 QUALITY EVALUATIONS

The significant effort has been made to solve the problem of measuring digital image quality with limited amount of success [13, 14]. Various studies tried to develop new metrics or to adopt existing ones for medical imaging [5, 6, 7, 8, 17]. The quality evaluations used can be broadly categorized as [5, 17]:

- objective quality evaluations – based on a mathematical or a statistical model, which is easy to compute and rate,
- subjective quality evaluations – based on a subjective observer evaluation of restored image, or questionnaires with numerical ratings.

These categories can be further sub-categorized, but this falls out of the scope of the paper [5, 17].

The quality evaluations proposed measure presentation (display) quality of the lossy compressed medical image. Therefore, they can be categorized as presentation-objective and presentation-subjective quality evaluation. Although, these quality evaluations have been devised for image quality measurement, they can be also used for evaluation of lossy compression techniques. The quality of the reconstructed image should not be the only criteria for adoption of a compression technique for PACS. The quality evaluation of medical image compressions for PACS is inseparable from technical aspects of the system. The lossy compression can uphold remarkable presentational quality (objective and subjective) of medical images but with high technical demands. In some cases these technical demands are not achievable and in most cases they are too expensive. In many countries this will impose too high price for PACS. Evaluations measuring image compression quality from technical point of view can be categorized as technical-objective quality evaluations.

3.1 Presentation-objective evaluations

Presentation-objective evaluations represent the most desirable way to measure image quality. They are based on a mathematical model, and are usually easy to compute. Their main advantage is objectivity [27]. The numerical distortion evaluations like mean squared error (MSE), Eq. (1), signal-to-noise-ratio (SNR), Eq. (2), or peak-signal-to-noise-ratio (PSNR), Eq. (3), are commonly used [6].

\[
\sum_{i=1}^{m} \sum_{j=1}^{n} [f(i, j) - f'(i, j)]^2 / m \cdot n \tag{1}
\]

\[
\frac{\sigma^2}{MSE} = \frac{1}{m \cdot n} \sum_{i=1}^{m} \sum_{j=1}^{n} (f(i, j) - \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} f(i, j)}{m \cdot n}) \tag{2}
\]

These measures fail to measure local degradations and do not provide precise descriptions of image degradations [5, 27]. Still, many studies use this quality evaluations to rate their implementations of lossy medical image compression techniques. Quality of the lossy compressions studied in [9, 24, 25, 26, 28] was measured by these numerical distortion evaluations. For example, Chen [9] used PSNR to evaluate propriety DCT based SPIHT compression, original SPIHT and JPEG2000. The DCT based compression achieved highest PSNR values for the tested medical images, which indicated that it is more suitable for medical imaging then the other two compression techniques.

Beside scalar numerical evaluations, graphical evaluations such as Hosaka plots and Eskicioglu charts, and evaluations based on HVS model have been used [14, 15, 29]. Their applicability in medical domain has been reported in [6, 27].

Also, a hybrid presentation-objective metrics have been studied for medical domain. Przelaskowski [27] proposed a vector quality measure reflecting diagnostic accuracy, Eq. (4).

\[
HVM = \sum_{i=4}^{6} \alpha_i V_i \tag{4}
\]

The values \( V_i \) represents one presentation-objective measure. The vector measure was designed to include the formation of a diagnostic quality pattern based on the subjective ratings of local image features. This quality measure represents a way of combining presentation-objective and presentation-subjective evaluations. Evaluation of lossy JPEG2000 compressed medical images found that compression ration of 20:1 is diagnostically acceptable.

3.2 Presentation-subjective evaluations

Presentation-subjective evaluations have been used to evaluate lossy compressed medical images more often than presentation-objective [30]. Presentation-subjective evaluations are based on observer’s subjective perception of reconstructed image quality [5]. The subjective quality of a reconstructed medical image can be rated in many ways [5]. In some studies, observer is presented with several reconstructed versions of the same image. The observer has to guess the image compression level and to order the sample images in order from the least compressed to the most compressed [5, 31]. If the difference between original image and reconstructed image at some level of compression is not distinguishable, then that level of compression is
diagnostically acceptable [32]. Other studies used qualified observers to interpret reconstructed medical images compressed at various levels. The compression levels on which results were the same as for the original image have been rated as acceptable [5]. Also, some studies used qualified technicians to define a “just noticeable” difference used to select the point at which compression level is not diagnostically usable. The observers have been presented with series of images, each compressed at higher level. They simply had to define the point at which changes became obvious. The studies were based on presumption that one can perceive “changes” in the image long before an image is degraded enough to lose its diagnostic value [5].

When subjectively evaluating medical images, it is not sufficient to say that image looks good. It should be proved that image did not lose the essential information and that it has at least the same diagnostic values as the original medical image [6]. Therefore, beside pure subjective evaluations, semi-subjective evaluations of a reconstructed medical image which measure diagnostic accuracy have been used. Observers often rated the presented images on a scale of 1 to 5 [10, 17]. Collected data have been further statistically analyzed highlighting averages and other trends in collected data. Quality of reconstructed medical images is most often measured by semi-subjective evaluation based on Receiver Operating Characteristic (ROC) analysis, which has its origins in theory of signal detection [6, 7, 27, 33]. A filtered version of the signal plus Gaussian noise is sampled and compared to a threshold. If it exceeds the threshold then the signal is declared to be there. As the threshold varies, so does the probability to erroneously declare the signal present or absent. The ROC analyses are based on ROC curves (Fig. 1), which are a simple complete empirical description of this decision threshold effect, indicating all possible combinations of the relative frequencies of the various kinds of correct and incorrect decisions [6]. The plot is a summary of the trade-off between true positive rate (sensitivity) and false positive rate (the complement of specificity). The area under the curve can be used to summarize overall quality or the efficiency of the detection process [6, 7].

The ROC curves are not applied directly to medical imaging. The decision threshold is based on diagnostic accuracy and physician’s judgment. Reconstructed medical images, which either possessed or not an abnormality, were presented to qualified specialists. Observers had to provide a binary decision if abnormality is present or not, along with a quantitative value for their degree of certainty (a number from 1 to 5). A subjective confidence rating of diagnoses is then used as if it were a threshold to adjust for detection accuracy [6]. A resulting diagnostic accuracy is compared with original image and used to define an acceptable compression level. The results for the different compressions or compression levels could be used for quality evaluation of compression techniques.

The success of ROC analysis depends on the number of test images and observers included in the study. Therefore, the ROC analyses tend to be expensive and time consuming. For example, a typical ROC study would require over 300 images to obtain a reasonable statistical confidence level, five or more radiologists to view these images, and a full-time statistician to coordinate and analyze the data [6].

Various results were obtained for image quality by presentation-subjective evaluations. Smith [31] reported that lossy JPEG compression of chest radiographs can be set at levels as high as 30:1. Perlmutter [7] reported that lossy wavelet compression of digital mammograms can achieve compression ratio of 80:1 with no influence on diagnostic accuracy. Przelaskowski [33] reported even better results for JPEG2000 compression of digital mammograms of 140:1 compression ratio. The study [33] was based ROC analysis.

3.3 Technical-objective evaluations

PACS as a part of a modern hospital becomes a highly interactive environment that is forming a ubiquitous computing environment for medical work [34, 35]. It is not limited to only one medical facility or to a group of closely spaced facilities. PACS often spreads over vast areas including not only the most prominent and richest of medical facilities, but also the facilities in rural and less developed areas [3]. The best, and expensive devices, are not available for this facilities. Also, it is unreal to expect a 100 Mbit connection (minimum for efficient PACS communication [3]) to all sides of such a sparse system.

As a part of the mobile health, devices with less storage, processing, and display capabilities are also a common part of a PACS. These devices can
process only the limited number of medical images and images of limited size [2]. Also, these devices usually use wireless networks which have capabilities far beneath connected ones [2]. Therefore, to view a medical image on these devices it is necessary to have images scaled for the display size of the mobile device. This could have a negative impact on PACS storage space [36], but it is minimized when the scaled medical images are acquired from the same image codestream as the original sized image i.e. when streaming of medical images is used [37, 38]. Image streaming is a process of gradual buildup of an image by resolution or by pixel accuracy [28]. It enables extraction of a lower-resolution image from the codestream.

The architecture of modern PACS is described by Fig. 2. Beside high class hospitals, the system contains less equipped hospital in rural areas and medical mobile devices.

These are all reasons for adopting lossy compression of medical images for a PACS, but they are also restrictions which one developing a PACS system should consider. They represent technical-objective criteria for evaluating a medical image compression technique for a PACS. The parameters of the criteria are overall cost of the system equipment, storage and network requirements, the cost for implementation of the compression technique, compression/decompression speed, streaming possibility of the compression technique, image modalities suitable for the compression, and compression ratio achieved under certain quality assumption. Technical studies comparing different compression techniques evaluated several things [39, 40, 41, 42]:

- **Compression speed** [39, 40, 41, 42]. The studies measured the time elapsed while the sample image was compressed to target compression ratio, and the time elapsed during decompression.

This time has impact on overall performance of the system because it can cause data transmission delay. Better PACS performance is achieved if decompression time is minimized, because decompression occurs more often than compression. Therefore, the retrieval oriented compression techniques are common for medical imaging.

- **Memory and processor power used** [40]. The study measured the amount of memory and processor power used during compression and decompression process. The values measured inform about the overall complexity of compression technique which influence overall cost of the system. High requirements influence higher cost.

- **Compression ratio** [40]. The influence of compression technique on storage requirements is expressed as achievable compression ratio. It is measured in respect to image presentation quality, like numerical distortion measures, section 3.A. Storage requirements influence the overall cost of PACS.

- **Functionalities of a compression technique** [41]. Most applications require other features beside quality and coding efficiency of the compression technique. The technical-objective quality evaluations in consulted studies did not evaluate functionalities of a compression technique numerically. Rather they use some method of description. Santa-Cruz [41] provided a functionality matrix that indicated the supported features in each compression technique and an appreciation of how well they are fulfilled, Table 1. They compared JPEG2000, JPEG-LS, JPEG, MEPG-4 VTC, and PNG compression techniques. A set of features (functionality) is included in Table 1. A “+” mark indicates whether the functionality is supported. The more “+” marks,
Table 1: Functionality matrix – various functionalities of different compression techniques are compared [41].

<table>
<thead>
<tr>
<th>Lossless Compression Performance</th>
<th>JPEG2000</th>
<th>JPEG-LS</th>
<th>JPEG</th>
<th>MPEG-4 VTC</th>
<th>PNG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lossy Compression Performance</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>-</td>
</tr>
<tr>
<td>Progressive Bitstreams</td>
<td></td>
<td></td>
<td>++</td>
<td>++</td>
<td>-</td>
</tr>
<tr>
<td>Region of Interest (ROI) Coding</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>Arbitrary Shaped Object</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Random Access</td>
<td>++</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Low Complexity</td>
<td>++</td>
<td>++++</td>
<td>++++</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>Error Resilience</td>
<td>+++</td>
<td>++</td>
<td>++</td>
<td>+++</td>
<td>+</td>
</tr>
<tr>
<td>Non-Iterative Rate Control</td>
<td>+++</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Generality</td>
<td>+++</td>
<td>+++</td>
<td>++</td>
<td>++</td>
<td>+++</td>
</tr>
</tbody>
</table>

the more efficiently or better is the functionality supported by compression technique.

- **Error Resilience** [40, 41]. It is important to measure the error resilience of compressed images sent over network transmission channels. This is tested by transmitting the compressed data over simulated noise channel.

The quality evaluation of medical image streaming has not been studied in the consulted literature. Streaming of medical images is an important issue for PACS trying to achieve mobile health (and ubiquitous healthcare, also) and it should be considered during quality evaluation. Because it is supported by limited number of compression techniques, quality evaluation should indicate whether the streaming is supported or not. If compression techniques support image streaming, the quality of extracted low-resolution images should be evaluated.

An important issue considering technical aspects of medical image compression techniques is weather to use industry wide standards or to develop a proprietary compression technique [43]. The second approach could lead to more efficient compression techniques, but in long term, it would show more costly. It could compromise PACS communication with equipment and networks not supporting the proprietary compression technique [43]. The long term archives of medical images could be compromised if the system transgresses to another compression technique. The use of industry approved standards can reduce the cost and risk of using compression.

4 3D QUALITY EVALUATION SPACE

Quality evaluations of lossy compression techniques differ in many ways. They differ in way whether they consider the application for which the compressed image has been used. Some quality evaluations measure only the performance of the compression technique while other measure only the presentation quality of the restored image. Overall, there is no quality evaluation which measures all the elements of a medical image compression technique.

When evaluating medical image compression techniques it is important to measure the quality of the restored images. Presentation evaluations (objective and subjective) measure the presentation (perceptual) quality of a restored image. The values obtained are used to compare the quality of the compressed images and to observe which compression technique achieves higher compression ratio under the same quality assumption. It is easier to compute the presentation-objective measure which is usually presented as a scalar or a vector. These values are comparable and it is easy to obtain which compression technique is better – the one heaving bigger value. They fail to measure precise (local) characteristics of the restored image i.e. they do not consider the medical application of the compression technique. On the contrary, the presentation-subjective evaluations consider the medical application of the compression technique, but they are harder to obtain and cost more than presentation-objective evaluations. The presentation-subjective evaluations are harder to interpret and compare, and they are dependable of observer’s knowledge, experience and perception. The advantage of presentation-subjective evaluations is that they are recommended by official medical organizations which consider compression of medical images (like CAR).

The presentation quality evaluations fail to measure technical aspects of a compression technique. Beside restored image quality, it is necessary to obtain technical information about compression technique, like: efficiency (compression speed, achievable ratio, transmission possibilities), error resilience, features (image streaming), and implementation cost and maintenance. The technical-objective quality evaluations measure technical elements of a compression technique. There are several important technical-objective evaluations measuring different features of a compression technique. The issue is how to correlate them to one value.
To obtain the complete quality evaluation of a medical image compression technique, it is necessary to use all three previously described quality evaluations: presentation-objective, presentation-subjective, and technical-objective. Only then will the observers adopting a medical image compression for PACS have a complete insight of a given compression technique. This will present the medical staff with highest quality medical images while engaging the optimal processing, storage, and presentational resources.

The complete quality evaluation could be improved if there is a correlation function between the quality evaluations used, such as the one described by Eq. (5).

$$ev = f(a \cdot po, b \cdot ps, c \cdot to)$$

(5)

Variables $po$, $ps$, and, $to$ represent values obtained by applying presentation-objective, presentation-subjective, and technical-objective quality evaluations. Factors $a$, $b$, and $c$ are weighting factors ranging from 0 to 1 used to define the influence of a particular quality evaluation. Value of 0 cancels the influence of a particular evaluation.

Ideally, the result of the correlation function should be a scalar which should define the quality of a compression technique in a simple and a comparable way. A higher value indicates a better quality. Unfortunately, it is more realistic to expect that the result of the correlation function would be a vector which defines the quality of a compression technique in a space defined by presentation-objective, presentation-subjective, and technical-objective evaluations, Fig. 3. Higher vector intensity indicates a better quality.

One possible combination for 3D quality evaluation space would be the use of PSNR [6] (as it is the one most often used), Eq. (3), or Przelaskowski [27] vector measure (because combines it several objective measures), Eq. (4), for presentation-objective evaluation, ROC analysis [6] (as it is subjective measure used most often) for presentation-subjective evaluation, and functionality matrix [41] (being the most comprehensive technical evaluation) for technical-objective evaluation. The proposed combination is still under review.

5 CONCLUSION

This paper represents a preliminary research. We identified three categories of quality evaluations and criteria: presentation-objective, presentation-subjective and technical-objective. To obtain the complete quality evaluation of a medical image compression technique, it is necessary to use all three categories of quality evaluations.

The development of a comprehensive evaluation of all the aspects of a compression technique would ease the task of adopting a medical image compression for a PACS. Our future research will include devising a technical-objective correlation function which will uniformly present the results of technical-objective quality evaluations. The major focus of our future research will be devising a correlation function between all the groups of quality evaluations. We strive to achieve quality evaluation space like the one described by the Fig. 3 which would represent an environment for simple and comprehensive evaluation of medical image compression techniques for PACS.

In the case of PACS for the lung hospital we did not have time to wait for development of 3D quality evaluation space. Therefore, we adopted the compression technique that in our opinion (which was drawn from numerous technical studies) offered the most - JPEG2000 compression [36, 37, 38, 39, 40, 41, 42]. It would be interesting to see if this decision correlates with the results in the 3D quality evaluation space for compression of medical images.
ACKNOWLEDGEMENT

This research has been conducted under IT project: “WEB portals for data analysis and consulting”, No.13013, financed by the Government of Republic of Serbia.

6 REFERENCES


