Usability of irreversible image compression in radiological imaging. A position paper by the European Society of Radiology (ESR)

European Society of Radiology (ESR)

Preface

The usability of image compression has been a relevant topic in radiological image management for a long time. Despite some well-prepared recommendations by the national radiological societies in Canada, Germany and the United Kingdom, there are still different concerns by users and vendors about implementing such tools.

Based on these experiences, the ESR has initiated an international expert discussion on open issues using image compression. This paper summarizes the results of this process. It is focused on the use of "diagnostically acceptable irreversible compression" (DAIC). The so-called "lossless compression", meaning that the information is fully preserved, is out of scope of this paper.

The paper is of interest for radiologists, picture archiving and communication systems (PACS) administrators, researchers, vendors and imaging management service providers. Therefore, special background information and detailed technical information are also part of this paper, to present the best overview.

Introduction

Radiological imaging methods have become digital over the last decade. Furthermore, new advancements—including multi-detector computed tomography (CT) and functional imaging in CT and magnetic resonance imaging (MRI)—have substantially increased the size and number of digital images in radiological imaging departments.

European Society of Radiology (ESR)
Neutorgasse 9/2,
1010 Vienna, Austria
e-mail: comminucations@myESR.org

Decreasing costs of storage solutions have been equalized or surpassed by the increasing amount of digital images. In telemedicine, limited bandwidth and resulting duration of image transmission for huge studies is sometimes critical.

Therefore, the interest in using image compression is high. Picture archiving and communication system (PACS) vendors offer the use of different image compression methods, including reversible ("lossless") and irreversible ("lossy") compression algorithms. This paper discusses irreversible image compression only. The goal is to review the scientific, technical and organizational issues that impact upon the use of irreversible image compression in a manner that is visually imperceptible and/or without loss of diagnostic performance.

Recommendations on the use of irreversible compression have so far been published by three radiological societies over the last two years:

- Royal College of Radiologists (RCR, UK) "The adoption of lossy data compression for the purpose of clinical interpretation" (April 2008) https://www.rcr.ac.uk/docs/radiology/pdf/IT_guidance_LossyApr08.pdf [1]
- German Röntgen Society (DRG, Germany) "Compression of digital images in radiology—results of a consensus conference" (Rofo 2009) [2]
- Canadian Association of Radiologists (CAR, Canada) "Pan-Canadian evaluation of irreversible compression ratios ("lossy" compression) for the development of national guidelines" (J Digit Imaging, 2009) [3]

These recommendations all assert that the use of irreversible image compression is possible without losing relevant clinical information. There are slight differences in the amount of compression accepted (expressed as "compression ratios"). Based on these guidelines, users should be able to implement workflows using irreversible image compression. There are still some concerns that remain,

including responsibility, choice of algorithms, and effect on post-processing. This is due to different reasons. These national recommendations also reflect results from many scientific studies, but parts of these studies have different weaknesses due to non-standardized tools, compression algorithms, characteristics of cases (e.g. subtle findings, image noise).

Therefore, the European Society of Radiology (ESR) has initiated a process to discuss these open issues, including well recognized international experts in this field. This activity has been coordinated by the ESR Subcommittee for Information and Communication Technology (ICT).

Overview on image compression

There are two types of image (or data) compression.

One is "reversible", or "lossless" compression, which preserves mathematically the full information in the original. That is, after compression and decompression, the pixel values are identical to those in the original image. Normally, a compression ratio of 1:2 up to 1:3 could be achieved by this technique [4]. This "lossless compression" can be used without any concerns, due to the fact that the full information is preserved "bit by bit" and will be available following the decompression. Therefore, this kind of image compression is out of the scope of this paper.

The other type is "irreversible" or "lossy" compression, which, although it does not exactly preserve the pixel values on decompression, recovers values that may be fit for a particular purpose yet achieve greater compression. Irreversible compression is used in many different applications in real life every day, e.g. digital photography. Most known examples for such algorithms are JPEG and JPEG 2000 (syn. J2K), which can be used with both methods as reversible and irreversible compression. Video images may also be compressed, usually in an irreversible manner, in order to achieve a particular frame rate in a fixed bandwidth, e.g. using MPEG.

In medical applications, it is possible that irreversible compression may be used without producing perceptually visible differences ("visually lossless"), and/or without producing differences that affect diagnostic performance ("diagnostically lossless"). Depending on the amount of information in the image, and the diagnostic task, the degree of compression that may be used safely with a particular algorithm depends on many different factors, including investigation and body part and disease. For example, CT of the brain may tolerate far less irreversible compression than projection radiographs, such as CR images of the breast, which can be compressed with surprisingly high ratios without producing visually recognizable differences or affecting diagnostic performance.

Many studies have been published regarding medically acceptable compression ratios. Objective methods to determine the best compression ratio are studies based on comparing diagnostic performance (accuracy) using compressed and original images. The ability of a human observer to detect the difference between compressed and original images can also be tested objectively, or their subjective preference or opinion as to the quality can be determined. There are also software tools available to determine "just noticeable difference" (JND) values based on mathematical models that simulate the human visual system (HVS) [5, 6].

As already mentioned, recommendations on the achievable compression ratios have been published by CAR, DRG and RCR. There will be no new or modified recommendations based on this ESR initiative (Table 1).

It is obvious, that the relationships for different modalities and regions correlate very well between these recommendations, even that there are more progressive or conservative absolute values. One should be aware that these compression ratios should be correlated with reversible ("lossless") compression, which means that an irreversible compression ratio of 1:10 is about 1:2.3 (based on bits) up to 3.1 (based on bytes).

The use of irreversible compression tools should be adjusted to the diagnostic performance, which should be the relevant criteria for decision-making. Therefore, it is recommended to use the term "diagnostically acceptable irreversible compression" (DAIC).

Use cases

There are different meaningful reasons for using visually or diagnostically lossless compression in radiology; some examples are:

- *Teleradiology:* cooperative work between different healthcare providers is usual. Digital images could and should be transmitted electronically to provide full information at the point of care. Limited or expensive access to bandwidth, especially in remote regions, increases transmission time. Irreversibly compressed studies could be transferred in less time than reversibly compressed images (perhaps by a factor of 3 or more, depending on the investigation, body part and task).
- *Functional or volumetric imaging:* new imaging techniques with high-resolution and isotropic voxel or time-resolving acquisition protocols produce studies with a huge number of images, sometimes several gigabytes. Storage of these images is required for further work-up in clinical studies or follow-up examinations.

Table 1 Medically acceptable compression ratios

	UK	Canada JPEG	Canada J2K	Germany
Radiography - chest	10	30	30	10
Radiography - skeletal	10	30	20	10
Radiography - body	10	30	30	10
Radiography - pediatric		30	30	
Radiography - mammo	20	25	25	15
CT - head	5	12	8	5
CT – skeleton/chest/lung	5	15	15	8
CT - body		15	10	10
CT - angio		15	15	
CT - pediatric		15	15	
MR	5	24	24	7
NM		11	11	
US	10	12	12	
XA	10			6
XRF				6
Images for RT planning	None			

- *Long-term storage and PACS migration:* even with declining costs for storage, the growth of storage requirements in enterprises, regional or national, for PACS implementations is evident. Therefore, it is an economical aspect if less capacity is required, and especially regarding the migration of data over years—which will happen regularly every 3–6 years—it is also a matter of time how much data have to be copied to newer storage systems.

Methods to evaluate compression results

During the evaluation of compression results, an essential step is to monitor the exact compression methods and parameters being used, the image encoding characteristics and (to understand their impact on the compression algorithms) the image complexity, its radiological "entropy".

A desirable goal would be to store the output values described by (mathematical) methods automatically computed either on the uncompressed input or "difference" images produced by a pixel-wise subtraction of the compressed image from the original.

Having this kind of statistic is essential for designing any quality assurance strategy to evaluate compression results. Further research should go into a definition of the algorithms to be used, since their output should preferably correlate with the accepted irreversible compression ratios agreed on in the consensus decisions. Most likely there will be no single test to describe the quality of a compression workflow, but rather a matrix of different tests and conditions.

There are different ways to evaluate the quality, results and effects of irreversible compression. One is simply subjective assessment, where the assumption is that one cannot tolerate any visually perceptible difference. This can be done with a side-by-side (or "flicker") comparison of studies in blinded presentations to different readers.

Another approach is to measure the diagnostic accuracy for defined tasks (such as detection of a particular disease with a particular investigation) in a standardized setting and to compare performance using original images to performance using images compressed on different levels. Receiver operating characteristic (ROC) methods may be used.

The relationship between "visually lossless" and "diagnostically lossless" is complex. There is evidence that despite obvious visual degradation from irreversible compression, high performance equivalent to that on uncompressed images can be achieved for some modalities, body parts and diagnostic tasks. This is not dissimilar to the ability of a radiologist to interpret a poor quality image successfully (albeit less confidently, perhaps). That said, many radiologists are reluctant to interpret images that have been degraded by compression, so the "visually lossless" threshold may be the limiting factor rather than the "diagnostically lossless" threshold in practice, assuming that the former entails less compression than the latter. Conversely, though it is often assumed that if there is no visual loss, then there can be no diagnostic loss; this has been less well investigated, and it is possible that the means by which visually lossless thresholds are determined experimentally are insufficient to guarantee diagnostic performance, particularly for difficult tasks that involve low-contrast detection, require high frequency information to be preserved, require texture to be preserved or are vulnerable to compression artifact being misinterpreted as false positive findings.

Simple mathematical metrics that quantify the difference between the original and the decompressed image, such as peak signal-to-noise ratio (PSNR) and mean squared error (MSE), have been shown to be not well correlated with visual or diagnostic performance, so more sophisticated metrics have been developed. Another method based on mathematical models is simulating the human physiology. Using such software tools could help to measure the similarity or differences of images and to determine JND values, signal-to-noise ratios (SNRs) or probability levels for detecting differences in a number of pixels. Here, the structural similarity (SSIM) method has shown to be an improvement on traditional methods like PSNR and MSE because it seems to be more consistent with the performance of the HVS.

The amount of "information" contained in an image is described as its "entropy". This may be estimated mathematically, with varying degrees of sophistication. A simple measure is the zero order entropy (sum of the context-independent probabilities of each pixel value occurring). The extent to which an image can be reversibly compressed can also be used as a measure of entropy. Note that "information" in this context does not mean "useful" information, i.e. noise is included. The entropy of an image affects how much it can be irreversibly compressed before the difference is visually or diagnostically detectable. A large factor in the entropy in a medical image is the amount of the rectangular pixel matrix occupied by the body part (e.g. consider a small versus large breast on a fixed sized mammography detector); also important is the amount of noise in any unmasked background air or collimated area.

Images with high entropy should presumably be treated with lower compression ratios in irreversible compression than those with more uniform content. A simple approach is to measure the file output size of a reversible image compression method (JPEG lossless or JPEG 2000 reversible), which should be greater for images with higher entropy. Other computable methods like image compositional complexity (ICC), fractal dimension (FD), or region of interest (ROI) scores might be more efficient to calculate and more robust against noise in the original images.

In reviewing the literature on irreversible compression evaluation, and attempting to generalize from the results and conclusions of a study, a number of important factors should be considered:

- Whether or not the end point in the study was an objective measure of diagnostic performance, an objective attempt to determine a visually lossless threshold without a diagnostic task, or a subjective comparison
- For a performance based study, the generalizability of the body part, investigation and task (e.g. a study of the detectability of microcalcifications from digitized mammography film is not necessarily generalizable to detection of masses on direct digital or computed radiography mammograms)
- The selection of images considered with respect to the difficulty of the task, i.e. is it a "stress" set, in which only hard cases likely to be vulnerable to compression are included, or is it a "field" set, in which a representative prevalence of abnormal and normal findings is considered, or an "enriched" set, which is somewhere in between
- The overall performance of the reader with respect to the task, and the intra- and inter-reader variability, since if readers generally perform poorly (i.e. low sensitivity and specificity) in general (because the task is particularly hard or poorly defined, or they are lacking expertise or training), or vary of time or between each other, then the amount of variance may overwhelm the ability to detect differences caused by compression
- Whether or not the methodology accounts for localization information (e.g. many ROC studies do not)
- How reliable is the truth information that is used in the performance evaluation (e.g. in ROC studies), or whether a truth-independent methodology is used (e.g. agreement between readers of uncompressed and compressed images)
- Whether or not the underlying assumptions of the statistical methodology have been violated [e.g. pooling readers in ROC studies as opposed to using a formal multi-reader multi-case (MRMC) analysis], and how generalizable the results are (e.g. to a population of readers as opposed to just those involved in the study)
- The impact of side effects of compression on improving visual acceptance and diagnostic performance (or speed of diagnosis), e.g. the smoothing of the noise in irreversibly compressed images can improve performance (and ideally should be factored out as a separate processing step from compression and utilized routinely if so)
- Whether or not the algorithm and parameters used are generalizable to those that will be used in practice (e.g. much of the literature evaluates novel or proprietary rather than standard algorithms, and even if of the same family (e.g. wavelet) results in terms of compression ratio or bit rate may not be directly comparable with algorithms available for operational use (e.g. different basis function, different entropy coder, etc.)
- The failure to detect a statistical difference between uncompressed and compressed images does not necessarily mean that there is no clinically significant difference; the size of the study (in terms of cases and readers) may have insufficient statistical power to detect such a difference; the statistical power of the study to detect a predetermined "effect size" (clinically signifi-

cant difference) should be computed a priori (using variance estimates from the literature and/or pilot studies) and stated, and the confidence interval of the difference in performance between the uncompressed and compressed figure of merit (e.g. total area under the ROC curve) should be reported

- It can be difficult to establish what effect size (negative impact of compression) is acceptable, since this may be a matter of the nature of the resulting hazard, health policy with respect to quality of care expectations, and the balance of operational savings and the cost of unfortunate outcomes, and such information is often not available

- How the choice of case mix and allowable effect size are generalizable to individual patient care decision algorithms (e.g. a level of irreversible compression may have negligible negative effect on a mixed population of easy and difficult cases and readers of varying skills, but reduce the probability of detection by inexpert readers of a difficult finding, with potentially devastating consequences, consequently irreversible compression may need to be avoided or less aggressively applied for specific high risk tasks)

Compression technology

Medical images are usually stored in DICOM format, which contains image as well as non-image ("header") content. The DICOM standard defines how to encode the image content ("pixel data") either without compression, or by using various irreversible or reversible, non-proprietary image compression schemes. Of the compression schemes for still images currently defined in the DICOM standard, three are ISO standard schemes that can support reversible, irreversible and visually lossless irreversible image compression, JPEG and JPEG-lossless (defined in ITU-T.81 or ISO 10918-1 from 1992), JPEG-LS (ITU-T.87 or ISO 14495) and JPEG 2000 (Part 1 ITU-T.800 or ISO/IEC 15444-1 from 2002, and Part 2 Annex J). All these schemes are intended for the compression of continuous-tone still images and not designed to be applied to image types like scanned text documents, drawings or computer graphics. They may be applied to individual frames of a multi-frame image, but are not specifically intended for video compression (for example, they do not take advantage of intra-frame motion prediction). JPEG 2000 Part 2 Annex J supports a limited form of three-dimensional (3D) compression by providing a multi-component transform that can be used to exploit redundancy between slices.

Basically all compression schemes are based on the same technical principles using entropy coding, de-correlation and quantization steps.

Entropy encoding exploits a certain, non-uniform statistical distribution of a set of input symbols in a way that individual, frequent input symbols are mapped to codes allocating lesser bits and less frequent symbols are mapped to codes allocating more bits than the original input symbols, resulting in an overall data-size reduction compared with the set of input data. Depending on the chosen entropy coding scheme, there is not necessarily a one-to-one mapping of input symbols to codes. Arithmetic entropy coding schemes can map a number of input symbols to a single code-bit or to fractions of code-bits resulting in improved overall compression efficiency. JPEG, JPEG-lossless and JPEG-LS use a Huffman entropy coder which applies a one-to-one mapping of uniquely sized input symbols to output symbols of different sizes. JPEG 2000 is using a context adaptive binary arithmetic coder (CABAC), which encodes binary symbols using statistical distributions corresponding to the type of binary symbol to be encoded. A binary symbol representing, e.g. the sign of a numeric value, is encoded using the statistical distribution of all already encoded sign symbols, which is different to the statistical distribution used to encode numeric value bits. The encoder switches its contexts depending on the type of binary symbol to be encoded.

Since entropy coding performs efficiently on statistical non-uniform distributed data additional processing steps modify the distribution of the data prior to encoding. Those steps are decorrelation and quantization steps.

Spatially organized sample values, reflecting Hounsfield units, radiation absorption (X-ray) or dose accumulation (PT) are spatially correlated in medical image data. Decorrelation steps are converting those spatial sample values into a set of coefficients. This can be achieved either by applying a decorrelating transform or a prediction scheme.

Prediction schemes predict a certain sample value by a linear or non-linear combination of its spatial neighbours and encode the difference of the sample value and its prediction (prediction error). The more spatially correlated the sample values are, the smaller the resulting prediction error becomes. Highly correlated input sample values result in a high number of small prediction coefficients following a statistical distribution suitable to be efficiently encoded by entropy coders described before. Prediction schemes are fully reversible and are used by JPEG lossless and JPEG-LS.

A decorrelating transformation describes a spatial rectangular area of $N \times M$ sample values by a weighted superposition of $N \times M$ two-dimensional so-called base-functions. Spatially correlated input samples can be described by just a few prominent base functions, resulting in a few corresponding weighting coefficients, while weighting coefficients of the remaining base-functions are smaller or almost zero. The statistical distribution of those weighting coefficients—so-called transformation coefficients—is non-

uniform and suitable to be efficiently encoded by entropy coders described before. Decorrelating transformations are theoretically fully reversible.

JPEG applyies an 8×8 decorrelating discrete cosine transform (DCT). The spatial sample values are represented by a set of weighted superposition of 8×8 cosine functions. Due to the lack of integer implementations of the DCT in JPEG, rounding effects result in an irreversible DCT.

JPEG 2000 is using a full-size hierarchical Discrete Wavelet Transform (DWT) using easy to compute so-called lifting implementations that show some similarities to prediction schemes. There exist reversible integer implementation and irreversible floating point implementations of individual DWTs.

After the decorrelating transformation and prior to entropy coding, irreversible compression schemes perform a mapping of almost continuous range transformation coefficients to a significantly smaller number of "representative" coefficient values. That process is called quantization. Since the diversity of "representatives" is smaller than the diversity of transform coefficients, this quantization process is non-reversible. The difference between the quantized "representative" value and the original value is called the quantization error. The smaller the number of allowed quantized values, the higher the frequency of each quantized value becomes, which results in a more efficient entropy coding, resulting in a higher compression ratio. On the other hand, this reduction of the diversity of quantized values results in less accuracy of the coefficient to "representative" mapping, which results in an overall loss of image quality.

The quantization step is unique to irreversible encoding schemes. Reversible encoding schemes do not perform any quantization of coefficients; without quantization, and without loss of precision in intermediate steps (including decorrelation and colour space transformation), all values are encoded without loss.

Additional differences

JPEG ITU-T.81/ISO 10918-1

JPEG supports 8 or 12 bits of contrast resolution in irreversible modes, and 1–16 bits in reversible mode.

JPEG 2000 ITU-T.800/ISO 15444-1

JPEG 2000 supports any arbitrary contrast resolution between 1 and 16 bits.

Irreversible and reversible modes are supported. JPEG 2000 supports a so-called "progressive mode", where the JPEG 2000 bitstream is self-contained and can be truncated at any arbitrary place and still represents a valid bitstream.

JPEG-LS is rarely used in DICOM applications, perhaps because fewer codecs are commercially available, because its reversible compression performance, although excellent, is approximately the same as reversible JPEG 2000 Part 1,

and because its "near-lossless" mode (intended to be used to achieve a "visually lossless" result) has not been well studied.

JPEG XR is a relatively new development, accepted by ITU-T Recommendation T.832 and im Juni SO/IEC 29199-2 since 2009. Actually, there are no relevant experiences in radiological imaging. Nevertheless, a work item proposal has been accepted by the DICOM Standard Committee in December 2010 to include this compression algorithm into the DICOM standard.

Special aspects on compression algorithms

For the JPEG 2000 irreversible compression, it is primarily required to define how the compression ratio to be calculated should effect the size of the "data output stream" produced by the compressor. Most (all) JPEG 2000 implementations expect a bit ratio parameter as input that basically defines "how many bits of information of the original data should be packed into one bit of the output". Thus, the calculation of the value should be done on the size of the pixel data (DICOM tag) in the original image, not on the overall size of the DICOM image. We further need to look at the implementation latitudes of the DICOM image producers. The standard defines the properties of bits allocated (e.g. typically 16 in CT) versus bits stored (e.g. typically 12 in CT). Here, the bit ratio calculation has to be done on the bits stored setting.

If, however, the maximum pixel grey value in the image effectively is not in the bit layer defined through the DICOM bits stored but some lower value, the input for the compression algorithm has (one or more) "empty bit layers" in the high frequency range, which should be taken in to account while running the compression. This kind of encoding is typical for, e.g., CT brain or many MR cases, currently reflected by lower compression thresholds in the trails that have been performed.

Thus, a more secure approach would be to address the bit layer defined though the maximum pixel gray value for encoding, with the exception of "burned in" annotations/ markers, see below.

The image content complexity will definitely impact on the results produced when compressing images with a static ratio. From the point of quality assurance, it would be highly desirable to define generic rules that correlate "how much less a compression ratio" has to be selected for images with higher entropy.

Images where the pixel data are represented through a look-up table (LUT) may be a problem to the compression algorithms. There are two cases, one in which the LUT encodes a quantized continuous function (such as a sigmoid curve) intended to be applied to a continuous range of input

pixel values, and the other when the pixel values are not continuous, but arbitrary (unordered) indexes. Irreversible compression is harmless for the first (continuous) case. If irreversible compression is applied to arbitrary index values, the decompressed output will be garbage. For the indexed case, a recoding is essential to eliminate the LUT and replace all pixel values with real gray or colour value representations before compressing the image. In practice, DICOM images with a photometric interpretation of PALETTE COLOUR are vulnerable to this problem, and should be recoded as true colour images before irreversible compression

The absence of high frequencies in original images will also impact on the compression output. This, in particular, applied to thick-slice CT images computed from originally thinner raw or thin-slice data. It is in the nature of MIP creation to smoothen the content, with a typical reduction of the high frequencies in the resulting image. This might be a pitfall for the output quality produced by compression irreversible algorithms, since the main strategy of removing data in an image is to eliminate/reduce the high frequency parts which typically encode less visual information than the low frequency parts.

Images with "burned in" annotations/markers should also retrieve exceptional handling. These images often show a histogram abnormality with a very few pixels used for rendering the markers or annotations (e.g. "L" and "R" in CR or arrows/measurements in CT) at a very high grey value, followed by a probably huge gap to the maximum gray value representing the real pixel data information of the image. As discussed with possibly empty bit layers and the absence of high frequencies above, we need to reduce the ratios for the irreversible encoding to avoid putting too much compression on the images, losing their diagnostic applicability.

The choice of scheme—standard versus proprietary (or academic)—is relevant for interoperability; which is important due to the widespread implementation by DICOM receivers. Another issue is the flexibility of scheme to support advanced techniques in generic decoder (e.g. region-based coding in J2K requires no special decoder); note that use of proprietary schemes precludes distribution on standard media, importation by other PACS and data migration (to new archive or regional archive) without decompression (and causes problems in repeated cycles of recompression).

Effects of irreversible compression in specialized imaging procedures

As already mentioned, there are good reasons to use irreversible compression for modern imaging methods, like multi-detector CT, functional imaging and so on. At the moment, there are limited experiences on effects of irreversible compression in specialized imaging procedures like 3D imaging, measurements in functional imaging, computer aided detection (CAD), etc. Preliminary results indicate, that there could be significant influence in CAD results performed after irreversible compression.

It is recommended by different groups to use original images (non-compressed) in radiation oncology for therapy planning or in surgical/interventional navigation. Up to now, there is no evidence for this decision, but also there are not enough data on the use of compressed images in these applications. Therefore, it is out of the scope of this paper.

Also, visually differences in thin-slice CT studies have been reported using irreversible compression ratios by 1:8 up to 1:16 [7].

The effect of irreversible compression on measurements of size and signal intensity (HU in CT, SUV in PET) has not been well studied until now; it is recommended not to use irreversible compression in clinical trials that depend on such quantitative analysis.

Note: The Food and Drug Administration (FDA) does not allow irreversible compression of digital mammograms at this time for retention, transmission, or final interpretation (ACR Technical Standard for Electronic Practice of Medical Imaging, 2007) [8].

General description of a framework for acceptable irreversible compression

Diagnostically acceptable irreversible compression does not mean that the compressed image is completely equivalent to the original; instead it means that compression has been done to acceptable levels, as established by an appropriate quality assurance system. The system consists of the following components:

1. An analytical measure of quality that does not need to be perfect but should have the following additional properties:

   (a) The analytical quality metric should generally increase with increasing compressed file size, at least for the same original image and the same compression algorithm.

   (b) The metric should have the ability to be applied at reduced resolutions, so that the quality of a compressed image or data set at one resolution can be assessed at a reduced resolution. This property allows the quality assurance system to provide assurances with regard to the diagnostic acceptability of the irreversibly compressed data at multiple resolutions, allowing high-resolution images to pass the same acceptability tests as

lower resolution images that might be produced by less capable imaging equipment or by the same imaging equipment operating at reduced resolution. Any measure of quality that does not have this property necessarily introduces perverse incentives into the quality assurance system—e.g. an incentive to generate and compress low-resolution images because they can pass the acceptability test with small compressed size, even though high resolution images might pass the acceptability test at the same resolution and similar compressed size while simultaneously providing higher resolution information that could favour further improvements in diagnosis reliability.

More generally, the term "resolution" here can be considered to represent any spatial or volumetric transformations that are used to produce the images actually used in a diagnostic procedure for which acceptability is being assessed. These transformations might include reduction of thin-slice CT volumes to thick-slice data sets along one or more slice orientations.

(c) The metric should have the ability to be measured over spatial ROIs, as determined during diagnostic inspection. This property is important, because a single measure of quality inevitably pools evidence from across the image. This pooling process may allow the entire image to pass an acceptability test even though a specific ROI might not pass the test. Conversely, the pooling process may cause an image to fail the acceptability test globally, even though it passes the test on all regions of interest. For these reasons, the quality assurance system needs to be capable of evaluating the quality metric on prescribed ROIs, as well as the entire image, an report the assessed quality in each case.

(d) The metric should have the ability to be evaluated on the decompressed image after it has been subjected to one or more diagnostic viewing transformations. These transformations would typically be windowing functions that are commonly used with the radiological investigation in question. Note that viewing transformations can subsume other specifics of the investigation, such as bit depth and even sample value histograms. To make this so, viewing transformations should be expressed with respect to a notional display device in which the transformed data are rendered as black when 0 and maximally bright when 1, with a prescribed display gamma function—for that, a gamma value of one-third would seem appropriate since it correlates rather well with human perception. Conversions required to map imagery at some particular bit depth to the

actual diagnostic display monitor can always be folded into the formal viewing transformations.

Note: compressed file size, compressed bit rate, and the reciprocal of the compression ratio are all measures of quality that clearly satisfy the first property, but they clearly do not satisfy the other three properties.

2. An automated process to adjust the compression parameters of the compression algorithm used is required so as to ensure that the analytic quality measure meets defined thresholds. Note that this automated process might be quite straightforward, possibly requiring some iteration if initially selected compression parameters do not lead to a compressed result that satisfies the thresholds. Separate thresholds might potentially be defined for the overall image and for ROIs that might be available. Similarly, separate thresholds may apply at different resolutions and for different viewing conditions. This possibly iterative process is applied separately to each image (or possibly larger data set) that is subjected to irreversible compression. The thresholds, however, are established for all images within a particular diagnostic workflow. These thresholds need to be sufficiently high to guarantee acceptability of the compressed content for diagnostic purposes under the diagnostic conditions for which the thresholds are derived. There is no implication here that determination of suitable thresholds is a simple process or one that can be reliably carried out based on limited clinical trials. However, it should be clear that the ability to generalize limited clinical observations depends at least partly on the choice of a "good" quality metric.

3. A feedback process to set and adapt the thresholds based on diagnostic evidence. The feedback process is of course central to the success of the quality assurance system. Separate feedback processes could be set up for different parts of the workflow, depending on how compression is used and what the associated level of risk is. For example, the compression of content for primary view would be expected to have higher thresholds of acceptability than the compression of content for longevity studies or for remote viewing by third parties.

One very simple quality metric that could be used within the framework

One measure of image quality that is extensively used is PSNR. PSNR is equivalent to the MSE between the original and compressed images, normalized with respect to the largest dynamic range that can be represented within

the sample data bit-depth and converted to the log domain. There is nothing special about PSNR and it may not correlate well with visual quality. It is obvious, for example, that adding an offset to all samples in the image will radically alter the MSE and hence PSNR, while having little impact on perceived visual quality. That said, adding an offset to all values in an image is not characteristic of the distortions typically created by irreversible compression; moreover, adding offsets to all values in certain types of radiological images (e.g. CT) might well be considered a serious error, even if not visually noticeable.

Despite its short-comings, PSNR is superior to compression ratio as an image quality metric. PSNR can be made to satisfy all of the properties identified in the previous section. Moreover, it is possible to directly target PSNR as a quality metric during JPEG 2000 compression, without the need for iterative recompression of the data.

As a starting point, it seems that a reasonable workflow for distortion metric computation should be as follows:

1. Identify the image resolutions for which diagnostic acceptability is to be established, either through clinical trials or by means of an incremental quality assurance (QA) system. This set of image resolutions may grow as more sophisticated equipment becomes available, but it is important to be able to assess the acceptability of irreversibly compressed high-resolution content on lower resolutions for which measures of diagnostic acceptability have been (or are being established).
2. Identify the diagnostic viewing transformations for which diagnostic acceptability is to be established. Again, this set of transformations may be expanded as new radiological investigations become better understood.
3. Identify any spatial ROIs that may apply to a specific image—as potentially identified by a radiologist or other expert (typically during primary read).
4. For each combination of resolution R, viewing transformation X, and ROI G:

    (a) Map the uncompressed and compressed images separately to the desired resolution R using appropriate anti-aliasing filters
    (b) Apply viewing transformation X to each image—note that this leaves the image samples with floating-point values with a nominal range from 0 to 1
    (c) Evaluate the MSE between the transformed compressed and uncompressed images over each ROI G and also over the whole image
    (d) Compute the PSNR from $10 \times \log10(1/MSE)$—note that this is the correct PSNR expression for sample values with a nominal range from 0 to 1

    (e) Compare each calculated PSNR at resolution R against the acceptability threshold T_R established for that resolution

Before concluding this section, two remarks are in order. First, one could almost certainly do better by replacing PSNR with a structured similarity image metric (SSIM), but PSNR is both simpler and more well-defined. Second, it may appear that things are becoming very complicated indeed. However, this impression should tempered by the following observations:

1. One can start with just one resolution R and one viewing transformation for each investigation and it would still be better off than using compression ratios as a quality metric.
2. Where the original datasets from clinical trials of diagnostic acceptability for irreversible compression are already available, the compressed and uncompressed images from those trials can be passed through the above workflow to determine initial thresholds, T_R, that should already be more useful than adopting the compression ratios recommended as a result of such trials.
3. Thresholds for diagnostic acceptability must be tied to resolution if creating perverse incentives to store only low-resolution compressed content should be avoided. The PSNR method outlined above is about as simple a scheme as one could invent that would avoid creating such perverse incentives.

Workflow aspects using irreversible compression

Implementing and using irreversible compression in a radiological workflow should consider the entire image lifecycle in complex collaborative workflows, including enterprise wide image distribution or eHealth applications. Therefore, there is responsibility on both sides, the vendor of imaging modalities or PACS and also the users themselves.

Figure 1 puts irreversible compression in the perspective of the overall "image lifecycle pipeline", and also in perspective of other activities that affect the amount of information in the medical images (Fig. 1).

This document specifically deals with irreversible compression parts of Fig. 1. Nevertheless, it is important that the principles used to determine irreversible compression strategies are generic and can be applied to other types of reduction of medical image information. Otherwise, the end result may be contradictory usage in different parts of pipeline, now or in the future. As one example, the not uncommon practice of obtaining CT images as thin slices, but only sending thick slices to the PACS is arguably a much more severe data reduction than an irreversible JPEG
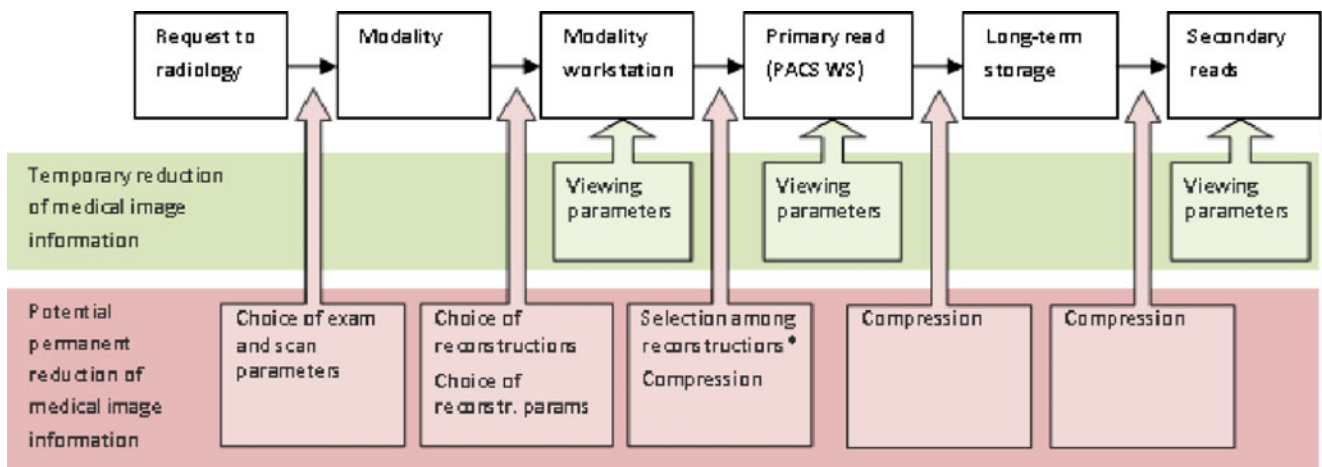
Fig. 1 Workflow aspects using irreversible compression

2000 compression (though the impact of this practice on diagnostic performance has not been well documented; neither have the legal implications of discarding data that were used in making an interpretation).

The *asterisk* in Fig. 1 highlights a risk for unnecessary data expansion. Creating multiple reconstructions of an image volume (for instance, sagittal and coronal versions where axial slices already exist, or thick slices where thin slices already exist) constitute a redundancy that can be removed without loss of information. This removal requires, however, that there is fast and effective on-the-fly multi-planar reconstruction (MPR) in the PACS. Further, the reconstructions of thick slices from thin by average intensity projection (AIP) or similar means does not produce exactly the same image as reconstruction from the CT-acquired raw data (which is rarely saved for this purpose).

In principle, there are two general modes of evaluation of images including different workflows: one is reading images with medico-legal responsibility, e.g. primary read or secondary read for an addendum or supplementary report; the other one is reviewing for comparison while follow-up studies or reviewing by other physicians.

Issue with multiple compression cycles

A workflow issue of high importance is that great caution must be exercised when applying repeated compression because of unexpected and potentially dangerous side effects due to repeated compression using different algorithms (or different implementations of the same algorithm) could result. One example is quantization in JPEG 2000. Consider the following setting: an image is compressed to a 10:1 ratio using quantization method A, and then decompressed. A subsequent compression to

20:1 using quantization method B may yield far worse quality than a 20:1 compression on the original image. However, compressing to 20:1 using method A may not cause additional distortion (if exactly the same implementation and parameters are used, which would be unlikely in real life—not enough evidence based on scientific studies is available at the moment). Vendors must make sure that when applying compression to an image that previously has undergone irreversible compression, the exact same compression parameters, such as quantization factors and tile sizes, must be used again. In a compressed image, these parameters are embedded in the image data, but if a decompressed version of the irreversible compressed image is to be distributed, all parameters must be provided alongside the image.

For each of JPEG and JPEG 2000, it could be rather precisely predicted what is required to avoid unnecessarily large degradation through multiple compression cycles. It is worth pointing out that if multiple compression cycles are envisaged as a possibility, then one would do best to provide a comprehensive set of quality layerings (costs almost nothing) within a single original JPEG 2000 image (perhaps running up to reversible) so that the quality can be progressively degraded by discarding quality layers rather than recompressing—little is lost by doing things this way (except that a lower performance integer wavelet must be used to avoid loss through floating point imprecision if reversible compression is required) and a great deal is gained, so it might be a matter of recommendation that where JPEG 2000 content is used, a reasonable collection of quality layers is provided from the outset; alternatively, smart algorithms can mostly recreate these after the fact, but with lower reliability.

Note: it is harmless to reversibly recompress an image that has previously been irreversibly compressed, since by definition reversible compression is always harmless (same bits out as in).

Users should implement QA systems to check the quality of irreversible compression over time. There is a need to control the compression application.

Different settings using irreversible compression are known. A conventional approach is using original images for primary reporting and to do irreversible compression afterwards, e.g. following a time stamp and compress images in the PACS archive after a specific time, e.g. 6 months. In this case, there is discussion on the question of which image—the original one or the later compressed one—is legally relevant or should be archived.

Another approach is the direct use of irreversibly compressed images, even for primary reading. This is done on the assumption that only compression solutions (ratios, algorithm, etc.) ensuring visually and diagnostically lossless compression are used. This setting is relevant in Canada [9, 10].

Following this second statement, some experts recommend to implement irreversible compression in the different modalities itself, this would be in analogy to other image manipulation like filtering, enhancement, etc. The risk of this approach would be that potentially many different compression solutions, based on different algorithms and technical solutions, would be implemented in one department. This would lead to an enormous challenge for QA processes, not to mention interoperability concerns. While it can be argued that irreversible compression is in many aspects very similar to image processing usually performed within the investigation, one big difference is that irreversible compression must be followed by decompression, while all other image processing steps are one way.

What is relevant in a QA setting?

Any facility using a digital image data management system must have documented policies and procedures for monitoring and evaluating the effective management, safety, and proper performance of acquisition, digitization, processing, compression, transmission, display, archiving, and retrieval functions of the system.

Display capabilities: displaying prior application of irreversible compression ratio, processing, or cropping (ACR Technical Standard for Electronic Practice of Medical Imaging, 2007) [8].

Issue with limited bandwidth

It might be worth pointing out that JPEG 2000 provides a particularly attractive way of addressing the teleradiology application with limited bandwidth, since standardized tools (especially JPIP) already exist to view the content remotely while incrementally retrieving content (even up to reversible) on an ROI basis. This would seem to allow the best of both worlds (reversible compression, with rapid availability over slow communication channels), especially if backed up by metadata identifying the most interesting regions identified in a first screening.

Issue on harmonizing compression tools

An open issue in the discussion between vendors and users is if there could be a consensus to use only one compression type (JPEG or JPEG 2000) to avoid interoperability problems in PACS implementations.

Also, it should be discussed, if—over time—there is a need for an IHE profile regarding this issue.

Legal issues

In principal, legal issues are out of the scope of these guidelines, because these are generally based on national law.

For example, in Canada there is a governmental supported/driven process in evaluating and implementing irreversible compression for the national Canada Health Infoway Project. In Germany, irreversible compression is allowed by the relevant law, if "there is no loss in medical information".

General guidelines

1. Image compression using irreversible compression algorithms may be acceptable in diagnostic radiological imaging, if used according to accepted guidelines.
2. To ensure visually and diagnostically lossless irreversible compression, radiologists should follow the recommendations of CAR, DRG or RCR. The amount of compression achievable depends on the investigation, the body part and the compression algorithm used. Therefore, it is recommended to use the term DAIC.
3. Vendors and users should know about principles of different irreversible compression algorithms like JPEG and JPEG 2000, and their specific advantages and risks.
4. Irreversible compression may be helpful to provide high-quality imaging transmission in teleradiology using limited bandwidth.
5. Irreversible compression may lower the cost of maintaining long-term archives and reduce the time and costs of data migration.

6. Vendors should respect their responsibilities in offering reliable and confident compression solutions, e.g. to implement all DICOM-related requirements, like indicating irreversible compression in the DICOM header. Furthermore, vendors must make sure that when applying compression to an image that previously has undergone irreversible compression, the exact same compression parameters must be used again. If a decompressed version of the irreversible compressed image is to be distributed, all compression parameters must be provided alongside the image, for instance in the DICOM header.

7. Users should respect their responsibilities in using irreversible compression in choosing the optimal compression tools and ratios, and also to implement appropriate QA processes.

8. It is important that repeated irreversible image compression be performed with care to avoid unexpected further loss.

9. At the moment there are limited data on the effects of irreversible compression in specialized imaging, like 3D, measurement or CAD. Evaluation of implementations before clinical implementation is strongly recommended.

10. Systems in radiation oncology or surgical navigation are likely not to accept compressed images, workflow concepts should reflect this. It seems that there are no specific technical reasons, therefore, but it might be a "maturity issue" of those systems at the moment.

11. Different workflows using irreversible compression are known, e.g. using compressed images already for primary reading up to do the compression for "old images" in the PACS only, or a mix between. In principle, there are different advantages and risks of these different approaches, but each of these concepts is acceptable in principle.

12. Users should be aware that the storage of multiple reconstructions of an image volume constitute a redundancy that can be removed without loss of information.

References

1. Royal College of Radiologists (RCR, UK) (2008) The adoption of lossy data compression for the purpose of clinical interpretation. https://www.rcr.ac.uk/docs/radiology/pdf/IT_guidance_LossyApr08.pdf
2. Loose R, Braunschweig R, Kotter E, Mildenberger P, Simmler R, Wucherer M (2009) Compression of digital images in radiology—results of a consensus conference. Rofo 181:32–37
3. Koff D, Bak P, Brownrigg P, Hosseinzadeh D, Khademi A, Kiss A, Lepanto L, Michalak T, Shulman H, Volkening A (2009) Pan-Canadian evaluation of irreversible compression ratios ("lossy" compression) for development of national guidelines. J Digit Imaging 22(6):569–578, Epub 2008 Oct 18
4. Huang HK (2010) PACS and imaging informatics: basic principles and applications, 2nd edn. Wiley-Blackwell, Hoboken
5. Krupinski EA, Johnson J, Roehrig H, Nafziger J, Fan J, Lubin J (2004) Use of a human visual system model to predict observer performance with CRT vs LCD display of images. J Digit Imaging 17(4):258–263
6. Shiao Y-H, Chen T-J, Chuang KH et al (2007) Quality of compressed medical images. J Digit Imaging 20:149–159
7. Erickson BJ, Krupinski E, Andriole KP (2010) A multicenter observer performance study of 3D JPEG 2000 compression of thin-slice CT. J Digit Imaging 23(5):639–643, Epub 2009 Jul 15
8. American College of Radiology (2007) ACR technical standard for electronic practice of medical imaging. http://www.acr.org/secondarymainmenucategories/quality_safety/guidelines/med_phys/electronic_practice.aspx
9. Koff DA, Shulman H (2006) An overview of digital compression of medical images: can we use lossy image compression in radiology? Can Assoc Radiol J 57(4):211–217
10. Norman G, Bak P, Matos A, Koff DA (2010) Re: PanCanadian evaluation of irreversible compression ratios ("lossy" compression) for development of national guidelines. J Digit Imaging 23:517–518

List of Experts involved preparing this paper (in alphabetical order)

Peter Bak, PhD, CMC Assistant Professor Medical Imaging Informatics Research Centre at McMaster University Department of Radiology, Faculty of Health Sciences McMaster Innovation Park, Canada.

Michel Claudon, Service de Radiologie, Hôpital d'Enfants, Vandoeuvre les Nancy, France.

David Clunie, MD, MB,BS, FRACR, Chief Technology Officer RadPharm, Inc., Chair DICOM WG 5, USA.

Remy Demuth, MD, Radiologist, Président Société luxembourgeoise de radiologie, Luxembourg.

Neelam Dugar, MD, Consultant Radiologist Clinical PACS Lead Doncaster and Bassetlaw Hospitals NHS Trust Chairman of The UK PACS and Teleradiology Group, Great Britain.

Dave Harvey, MD, Radiologist, Director Medical Connections, Great Britain.

Peter Heiles, Agfa HealthCare GmbH, Bonn, Germany.

Marc Kämmerer, MD, PhD, Radiologist, VISUS Technology Transfer GmbH, Bochum, Germany.

David Koff, MD FRCPC Chief/Chair Dept. of Radiology, McMaster University, Hamilton, Ontario, Canada, Director MIIR-CAM, co-founder, treasurer IHE Canada.

Elmar Kotter, MD, PhD, Radiologist, Vice-Chair Dept. of Radiology, University Freiburg, Germany; Chairman of the DRG IT Working group.

Torbjörn Kronander, PhD, CEO Sectra Imtec AB, Sweden.

Reinhard Loose, MD, PhD, Chair Dept. of Radiology, Hospital Nuremberg, Germany.

Claes Lundström, PhD, Research Director Medical ITSectra Imtec AB, Sweden.

Jens Martin, VISUS, Bochum, Germany.

Uwe-Erik Martin, Siemens AG, Healthcare Sector Imaging & IT Division, Erlangen, Germany.

The Expert Panel on Image Compression in Palma de Mallorca in October 13, 2010 was kindly supported by