MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE
IEEE COMMUNICATIONS SOCIETY
http://www.comsoc.org/~mmc

E-LETTER

Vol. 6, No. 7, July 2011

IEEE COMMUNICATIONS SOCIETY

CONTENTS

Call for Participation: MMTC Meeting at ICME 2011 (Barcelona)

Dear MMTC members,

Our next TC meeting has been scheduled on July 12 (Tuesday) at ICME 2011 (Barcelona). If you are attending ICME or in the area by then, please plan to attend our face-to-face meeting. We will report to you the recent development of this TC in the past 6 month, and exchange ideas with you on our plans for the next. Please feel free to bring your friends to our meeting. Snacks, fruits and beverages will be served in the meeting.

IEEE Communications Society MMTC Meeting
Date/Time: July 12, 2011 (Tuesday) 14:00 - 15:30
Location: LaSalle conference venue

Agenda:
Welcome
MMTC sponsored conference reports
Interest Groups reports
Board reports
TMM/ICME special report
Award recognition
General discussion
Adjourn

Thanks. Looking forward to see all of you there!

Haohong Wang
Chair
IEEE MMTC

SPECIAL ISSUE ON IMAGE COMPRESSION TECHNOLOGIES FOR MEDICAL APPLICATIONS

Joan Serra-Sagristà[1] and Ali Bilgin[2]
[1]Universitat Autonoma Barcelona, Spain, [2]University of Arizona, USA
[1]joan.serra@uab.cat, [2]bilgin@email.arizona.edu

The use of imaging in healthcare grows at an astonishing rate. It is estimated that 5 billion medical imaging studies have been conducted worldwide up until 2010 [1] and this number continues to increase rapidly as contemporary medicine relies more heavily on imaging-based investigations and procedures. With such increased use and the ever increasing resolution of medical imaging technologies, the amount of data generated in medical imaging applications can be overwhelming. Efficient storage and transmission of these large data sets necessitates the use of efficient compression techniques. This special issue presents five different articles on compression technologies for medical imaging applications.

The first article titled "Standardized technologies for volumetric image coding" provides an overview of the standardization efforts in volumetric image compression and highlights some of the ongoing research in this area.

The article titled "Scalable Wavelet Compression of Higher Dimension Medical Images" presents an efficient compression technique that can be used to compress three- and four-dimensional volume and functional images that are increasingly used in medical imaging applications. The article also highlights the desirable scalability features of the proposed technique which can be as important as compression efficiency in some applications.

The greater use of mobile devices in medical imaging necessitates transmission of images over wireless channels. The article titled "Compression of 3D Medical Images for Wireless Transmission" introduces a compression system for transmission of 3D medical images over error-prone wireless networks.

While most image compression standards are designed to be used in a variety of applications, their performance in specific applications can sometimes be improved by using preprocessing techniques that utilize application-specific information. One such method is presented in the article titled "Improved Compressibility in JPEG2000 2D and 3D Reversible Compressions of Thin-section Chest CT Images by Increasing the Data Redundancy outside the Body Region". The proposed preprocessing method increases the compressibility of thin-section chest CT images while the compressed data maintains full standards compliance.

Last, but not least, the article titled "What is Different About Medical Image Compression?" discusses attributes unique to medical image compression. Written by a radiologist who has been instrumental in the development of the DICOM (Digital Imaging and Communications in Medicine) standard, the article is a valuable resource for readers interested in developing compression techniques for medical applications.

References

[1] C. A. Roobottom, G. Mitchell, and G. Morgan-Hughes, "Radiation-reduction strategies in cardiac computed tomographic angiography," Clinical Radiology, Vol. 65, No. 11, pp. 859-867, November 2010.

Joan Serra-Sagristà (S'97–M'05) received the B.S., M.S., and Ph.D. degrees in computer science from the Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain, in 1992, 1994, and 1999, respectively. Since 1992, he has been with the Department of Information and Communications Engineering, Universitat Autònoma de Barcelona, where he is currently an Associate Professor and the Director of the

Group on Interactive Coding of Images. From September 1997 to December 1998, he was on a DAAD research grant at the University of Bonn, Bonn, Germany. From June to July 2000, he was a Visiting Researcher with the University of Bonn. His current research interests include image coding, data compression, vector quantization, and wavelet-based techniques, with special attention to remote sensing and telemedicine applications. He has coauthored several papers in these areas. Dr. Serra-Sagristà is a member of the SPIE. He has served on the steering and technical program committees of several international conferences, and he is a reviewer for the major international journals in his research field. He was the recipient of the Intensification Program Young Investigator Award in 2006.

Ali Bilgin received the B.S. degree in electronics and telecommunications engineering from Istanbul Technical University, Istanbul, Turkey, the M.S. degree in electrical engineering from San Diego State University, San Diego, CA, and the Ph.D. degree in electrical engineering from the University of Arizona, Tucson, AZ. He is currently an Assistant Professor with Departments of Biomedical Engineering and Electrical and Computer Engineering at the University of Arizona, Tucson, AZ. His current research interests are in the areas of signal and image processing, and include image and video coding, data compression, and magnetic resonance imaging. He has co-authored over 100 research papers in these areas. He has served as technical committee and organizing committee member for various conferences and currently serves as an Associate Editor for IEEE Transactions on Image Processing and IEEE Signal Processing Letters.

Standardized Technologies for Volumetric Image Coding

Tim Bruylants[1,2], Joeri Barbarien[1,2], Adrian Munteanu[1,2], Jan Cornelis[1] and Peter Schelkens[1,2]

[1]Vrije Universiteit Brussel, Belgium, [2] Interdisciplinary Institute for Broadband Technology (IBBT), Belgium

[1,2]tim.bruylants@vub.ac.be

1. Volumetric image data

Digital medical imaging devices, such as ultrasound, computed tomography (CT) or magnetic resonance imaging (MRI) scanners generate large amounts of volumetric data sets. Tomography used in other sciences such as archeology, biology, oceanography, geophysics, and material science, is another important source of volumetric imagery. Various advancements in digital scanner technology have led to a significant increase of the resolution and sample bit-depth of the individual data sets, causing them to grow accordingly in size. Along with the digitization of these acquisition devices, the industry focus is also shifting towards applications, transport networks and storage systems. This gives rise to new requirements such as efficient storage, transmission, random accessibility, region-of-interest support, long-term archival and interoperability between applications in a multi-vendor environment.

2. DICOM and JPEG 2000

The Digital Imaging and Communications in Medicine (DICOM) Standards Committee is without doubt the most important organization concerned with the standardization of communication for medical imaging. Before DICOM existed, medical devices used proprietary formats to store, transmit and compress medical images. DICOM changed this, by providing a set of standards that allow for interoperability and compatibility in a multi-vendor eco-system. The first DICOM standard was published in 1985, but has since then been updated regularly. Currently, the standard is at version 3.0 and incorporates many changes and improvements since it was created. For the compression of all types of medical images, DICOM 3.0 adopted existing image compression standards from the JPEG working group. Because of the promising results and popularity of wavelet-based compression technologies [3,6], DICOM decided in the late nineties to adopt JPEG 2000 Part 1 as soon as it was finished, in 2001 [1]. In this way the standard is able to offer efficient compression, storage and transmission of two-dimensional medical images. The wavelet-based compression scheme of JPEG 2000 offers many advantages over the more classic compression schemes that were already available in DICOM (e.g. JPEG and RLE), including improved compression efficiency and extended functionality. Most notably is the native resolution scalability that enables lower resolution and thumbnail generation at no extra cost. Additionally, JPEG 2000 with its embedded block coder by optimized truncation (EBCOT) paradigm also delivers excellent quality and bit-rate scalability [3,4,6].

3. Multiple Component Transformations

As technology advanced, medical images with multiple components per slice have become commonplace, raising new compression efficiency concerns. Initially, DICOM only supported JPEG 2000 Part 1, which only includes a color component transformation. However, in order to improve the compression performance for specific types of medical images that consist of multiple spectral components per image slice, DICOM adopted in 2005 the Multiple Component Transformations (MCT) extension of JPEG 2000 Part 2.This extension allows performing arbitrary lossy or lossless transformations on all components of an image. With the rising popularity of volumetric imaging devices, some vendors developed a creative approach to (ab)use this extension for the improved compression of volumetric images. By converting slices of a volumetric data set into virtual components, it is also possible to use JPEG 2000 with the MCT Part 2 extension, which allows for performing a wavelet transformation along the slice axis (i.e. the Z dimension). This significantly enhances compression efficiency. However, this approach also incorporates some serious drawbacks. First of all, it makes the distinction between components and slices for color or multi-spectral volumetric data sets ambiguous. Even the difference between a multi-component two-dimensional image and a volumetric image becomes unclear. Secondly, this approach is unable to treat all dimensions in an isotropic fashion, i.e. every dimension does not have equal configuration functionality. This is an

important drawback as it will negatively affect the rate-distortion performance and functionality of the codec. The fact that the code-blocks used in the entropy coder of JPEG 2000 remain two-dimensional and the fact that the wavelet decomposition structure is more limited with this approach, prevent reaching optimal compression efficiency. Moreover, JPEG 2000 defines a maximum of 16384 components per image, which is rather limited when compared to the maximum width and height of 4 billion pixels. This latter limitation is currently not yet an issue, but with the rapid improvements of scanner technology, it will be in the nearby future.

4. JP3D

In order to address the lack of proper support for the compression of volumetric images, the JPEG committee created Part 10 of the JPEG 2000 standard [2], also referred to as JP3D.Finished in 2008, it defines the volumetric extension for JPEG 2000 that provides isotropic support for handling volumetric images with multiple components and no time-component. It is specifically designed to be compatible with the other existing parts of the standard (e.g. JPIP and JPSEC) and as such offers the exact same functionality as its two-dimensional counterpart. But, because JP3D properly extends the wavelet transformation and the entropy coding to three dimensions, it is able to deliver better compression results than what was previously possible by using only Part 1 or the Part 2 MCT. JP3D also solves the ambiguity between components and slices and handles volumetric data sets in an isotropic fashion, that is, regardless of the orientation of the data. Swapping any of the three dimension axis makes no difference in compression efficiency or coding limitations. This is a huge advantage over other technologies, because it makes JP3D future proof as it is much better suited to handle high-resolution volumetric data sets. Furthermore, those specific functionalities that depend on the actual number of dimensions in a data set, like region-of-interest or random access, are properly supported through JP3D.In combination with JPEG 2000 Part 3 (Motion-JPEG 2000), it is even possible to represent 3D+t data sets.

5. Performance comparisons

As stated before, JP3D improves the compression efficiency for volumetric images. In order to illustrate these performance differences between the different presented compression technologies, this section shows results for three different types of representative medical volumetric images. For the following results, CT represents a typical medical CT-scan with a bit-depth of 12 bits per sample. MRI is an MRI-scan with a sample bit-depth of 12 bits. And finally, US represents a typical UltraSound scan with 8-bit samples. The results were generated by using the official reference implementation of JP3D [5]. Table 1 shows the bit-rates for lossless compression, using the 5x3 wavelet transform, of the respective image data sets, using JP3D, the Part 2 MCT approach and the classical JPEG 2000 2D based compression. Note that even the JPEG 2000 Part 1 2D methodology is still considered state-of-the-art image compression technology.

Lossless (bpp)			
Data set	**JP3D**	**MCT**	**2D**
CT	3.84	3.87	4.08
MRI	4.08	4.13	4.72
US	4.84	4.83	5.05

The results clearly show that applying a wavelet transform in the third dimension (JP3D and MCT) significantly enhances the compression efficiency. JP3D is slightly better, due to the application of 3D entropy coding (3D EBCOT). These improvements are most pronounced for datasets with a high axial correlation (which is less the case for the CT and US datasets in contradiction to the MRI dataset). For near-lossless to lossy compression, our results show improvements of JP3D over 2D based compression of up to 15% at high bit-rates. At lower bit-rates, the compression efficiency improves with up to 50% for JP3D compared to JPEG 2000 Part 1. Also note that JP3D performs slightly worse for the UltraSound data set, but that this difference is negligible when taking into account the added isotropy property and enhanced volumetric functionalities. Worth noticing is the fact that as the redundancies or correlations in a volumetric data set increase – especially along the slice axis – JP3D is in a more favorable position than alternative technologies that are available today. See [3,4] for more in-depth information on the coding.

6. Ongoing research

A new development in medical imaging involves the compression of 3D+t data sets (e.g. a beating heart using CT). Currently, Motion-JPEG 2000 already allows handling this type of data sets, but not in an optimal fashion. However, with the creation of JP3D extra care was taken to prepare the standard for the compression of data sets with more than three dimensions, without breaking compatibility with existing compressed data

streams. As such, JP3D is future proof because it potentially allows for support of up to 255 dimensions. The efficient compression of higher-dimensional data sets will require more in-depth research.

Recent research [7,8] shows that the classical wavelet transformation can still be surpassed in overall compression efficiency by adoption of directional adaptive wavelet transformations. Classical wavelets are excellent in describing pure horizontal, vertical and 45° diagonal edges, but are less efficient when handling discontinuities like smooth changing curves or straight lines at other angles. Therefore, in order to treat these more complex discontinuities, a new class of scalable geometric transforms is proposed as an alternative to the classical wavelet transforms. The directional adaptive wavelet transforms have the ability to better adapt to local geometric structures in the image, without any significant disadvantages over the classical wavelet transforms. These transforms allow for increased compression efficiency when compressing volumetric images, though one should account for the signaling overhead of the directional information.

With the leading chip-vendors now focusing more and more on multiple-core designs for micro-processors, research is also looking to parallelizability of existing and new compression methodologies. JPEG 2000 has, in that respect, a very fitting design because both the wavelet transform and the entropy coder stages allow for easy parallelization. This is a non-negligible advantage of JPEG 2000 over some other competing compression schemes.
Finally, above approaches are still generic in nature. Introducing the exploitation of modality, anatomy and pathology specific knowledge can further boost the coding performance as illustrated by Sanchez et al. [9].

7. Conclusions

JP3D offers a significant improvement for the compression of volumetric imagery. The three-dimensional wavelet decompositions brings superior compression performances for a wide set of medical imaging in the full range of bit-rates. Both lossless and lossy modes benefit from the three-dimensional wavelet transform and entropy coder. Further research shows promising results for even better compression efficiencies by adopting the new directional adaptive wavelet transformations.

References

[1] ISO/IEC 15444-1 and ITU-T Rec. T.800, “Information technology – JPEG 2000 image coding system: Core coding system” (2002).
[2] ISO/IEC 15444-10 and ITU-T Rec. T.809, “Information technology – JPEG 2000 image coding system– Part 10: Extensions for three-dimensional data” (2008).
[3] Taubman, D. S., and Marcellin, M. W., “JPEG2000: image compression fundamentals, standards, and practice”, Kluwer Academic Publishers, ISBN 978-0-792-37519-7 (2001).
[4]Schelkens, P., Skodras, A., and Ebrahimi, T., “The JPEG 2000 Suite,” Wiley Press, ISBN 978-0-470-72147-6 (2009).
[5] Bruylants T. (2008) JPEG 2000 Part 10 Verification Model v1.1.1. ISO/IEC JTC1/SC29/WG1, N4808, Busan, Korea.
[6] Schelkens P., Munteanu A., Barbarien J., Galca M., Giro i Nieto X. and Cornelis J. (2003) Wavelet Coding of Volumetric Medical Datasets. IEEE Transactions on Medical Imaging, vol. 22, no. 3, pp. 441-458.
[7] C. Chang and B. Girod, "Direction-Adaptive Discrete Wavelet Transform for Image Compression", presented at IEEE Transactions on Image Processing, 2007, pp.1289-1302.
[8] A. Munteanu, O.-M. Surdu, J. Cornelis, P. Schelkens, “Segmentation-Driven Direction-Adaptive Discrete Wavelet Transform,” IEEE International Conference on Image Processing, San Antonio, Texas, USA, vol. I, pp. 437-440, September 16-19, 2007.
[9] V. Sanchez, R. Abugharbieh, P. Nasiopoulos, “Symmetry-Based Scalable Lossless Compression of 3D Medical Image Data”, IEEE Transactions on Medical Imaging, vol. 28, no. 7, 2009, pp. 1062-1072.

Tim Bruylants graduated as Master of Science in 2001 at the University of Antwerp. In 2005, he participated as a member of the Forms Working Group (W3C). In 2006, Tim Bruylants became a PhD student at the Vrije Universiteit Brussel. The main topic of his research is the compression of volumetric data sets, using wavelet and geometric transforms. Since 2005, Tim Bruylants is also an active member of the JPEG committee. He is co-editor of the JPEG 2000 Part 10 (JP3D) specification.

Joeri Barbarien obtained the degree of Master in Electrical Engineering in 2000 and the degree of Ph.D. in Engineering Sciences in 2006, both from Vrije Universiteit Brussel, Belgium. Since October 2000, he is a member of the Department of Electronics and Informatics at the same university, where he currently holds a part-time professorship. He is also actively involved as a senior researcher and project-coordinator in the Interdisciplinary Institute for Broadband Technology (IBBT). His research interests include scalable video and still-image coding, distributed source coding, watermarking, and implementation aspects of multimedia processing algorithms. Joeri Barbarien is the author or co-author of more than 50 journal and conference publications, book chapters, patent applications and contributions to standards.

Adrian Munteanu received the M.Sc. degree in Electronics and Telecom-munications from Politehnica University of Bucharest, Romania, in 1994, the MSc degree in Biomedical Engineering from University of Patras, Greece, in 1996, and the PhD degree in Applied Sciences from Vrije Universiteit Brussel, Belgium, in 2003. Adrian Munteanu holds a professorship at the Department of Electronics and Informatics (ETRO) at the Vrije Universiteit Brussel (VUB). His research interests include scalable still image and video coding, scalable mesh coding, error resilient coding, image and video transmission over networks, video segmentation and indexing, multiscale image analysis, and statistical modeling. Adrian Munteanu is the author and co-author of more than 100 journal and conference publications, contributions to books, patents, and contributions to standards.

Jan Cornel (M'80) received the M.Sc. and Ph.D. degrees from the Vrije Universiteit Brussel (VUB), Brussels, Belgium, in 1973 and 1980, respectively. He is a Professor of electronics, medical imaging, and digital image processing and coordinates the research group on Image Processing and Machine Vision (IRIS) at VUB. Current research directions of IRIS include applications in data compression, medical imaging, computer vision, remote sensing, and mapping of algorithms on architectures for real-time applications. He is the author or coauthor of more than 400 scientific publications.

Peter Schelkens obtained an Electrical Engineering degree (MSc) in Applied Physics in1994, a Biomedical Engineering degree (Medical Physics) in 1995, and a PhD degree in Applied Sciences in 2001 from the Vrije Universiteit Brussel (VUB). Peter Schelkens currently holds a professorship at the Department of Electronics and Informatics (ETRO)at the Vrije Universiteit Brussel (VUB) and in addition a postdoctoral fellowship with the Fund for Scientific Research – Flanders (FWO), Belgium. Peter Schelkens is research coordinator at the Interdisciplinary Institute for Broadband Technology (www.ibbt.be), Belgium. He is involved in ISO/IECJTC1/SC29/WG1 (JPEG), WG11 (MPEG), and ITU-T standardization activities.

Peter Schelkens is the Belgian head of delegation for the ISO/IEC JPEG standardization committee, editor/chair of Part 10 of JPEG 2000.

Scalable Wavelet Compression of Higher Dimension Medical Images

William A. Pearlman
Rensselaer Polytechnic Institute, USA
pearlw@ecse.rpi.edu

1. Scalable Image Compression

The number, size, and resolution of medical images acquired for diagnostic purposes have increased tremendously in the last decade. Contributing to this increase in size is the advent of acquisitions of three- and four-dimensional volume and functional (volume versus time) CT (fCT) and MRI (fMRI) images. Furthermore, their delivery is often by wireless transmission over large distances for the purpose of rapid diagnosis at a remote location. Compression is required in order to satisfy the simultaneous demands for storage efficiency and rapid turnaround. Besides efficiency, the compression algorithm may need to deliver features of lossless capability for archiving, progressive lossy-to-lossless decoding capability for progressively increasing display quality, and decoding capabilities that allow progressively increasing resolution and random access to regions of interest.

Our efforts to satisfy these requirements for a single multi-dimensional image have focused on embedding all these features in a single compressed file. From this one file, the full image or a selected region can be decoded directly to reconstructions with the desired resolution and quality without any post processing. The image or image region can be progressively increased in resolution and/or quality by reading and decoding more bits from the compressed file.

2. Wavelet Transform Decomposition

A multi-resolution decomposition of the source image is needed to satisfy the requirement of progressive resolution decoding. The wavelet transform decomposes the source into subbands of various resolutions. After encoding these subbands, one can progressively build the reconstruction by inverting and synthesizing subbands starting from low to high frequency. The idea is best illustrated with a two-dimensional image that is wavelet transformed using three stages of decomposing the low pass subband by half-band wavelet filtering alternately in each dimension. Figure 1 shows the resulting subbands and describes the progressive buildup of the reconstruction in resolution.

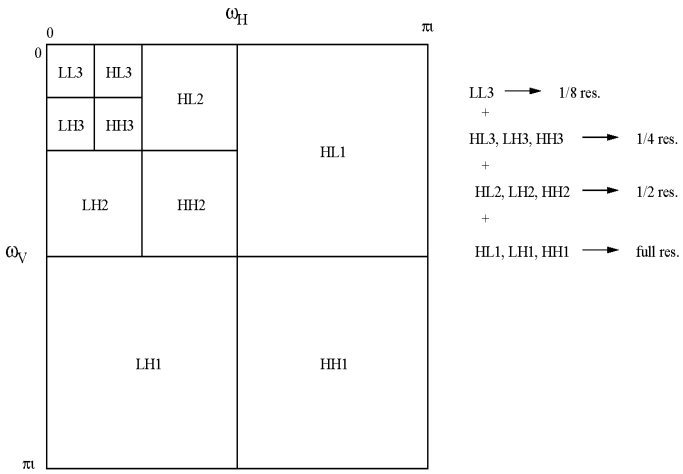


Figure 1. Wavelet subbands from three stages of half-band filtering in each dimension.

The three-dimensional wavelet decomposition in common use today is shown in Fig. 2. Each *xy*-slice is transformed as above followed or preceded by two stages of half-band filtering in the *z*-direction for every *(x,y)* coordinate.

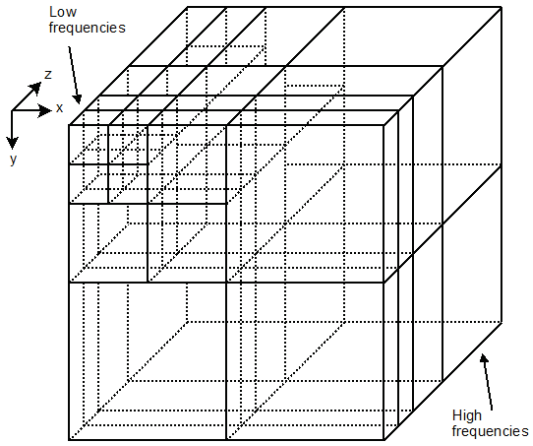


Figure 2. Subbands of a three-dimensional wavelet transform.

In order to preserve progressive resolution, the codes of the individual subbands must be kept separate in the compressed bitstream (hereinafter

called *codestream*) with lower frequency subbands coded before higher frequency ones.

3. Progressive Quality Coding

One way to achieve progressive quality coding is to organize the transform coefficients, so that the coefficients with higher magnitudes always precede those with lower magnitude and encode them in the same order. One can obtain a finer scale of quality progressiveness by sending all the higher order bits of these magnitude-ordered coefficients before the lower order bits. One calls this method progressive bitplane coding. It is *emdedded coding,* since each succeeding bit reduces the distortion less than its predecessor. This ideal is never perfectly realized, because the information needed to convey a perfect magnitude order is excessive and vitiates the compression objective of reducing the number of bits for encoding the image. A partial ordering by highest magnitude bits (containing largest power of 2 values) can be realized and is the objective of the set partitioning techniques of the SPIHT [1], SPECK [2], and SBHP [3]algorithms.

Progressive quality cannot occur simultaneously with progressive resolution, since the subband codestreams cannot be interleaved and still maintain progressiveness in resolution. For SPECK, SPBHP, and JPEG2000 [4] (to be discussed), the subband codestreams are naturally kept separate, but for SPIHT, it is more complicated, but still realizable (see [5]).

The JPEG2000 method [4] encodes blocks within the subbands using bitplane coding. These blocks are typically 64x64 or 32 x 32. This coding algorithm traverses bit-planes of coefficients in a subblock from the highest to lowest bitplane. However, it encodes the bits by context-based, adaptive, arithmetic coding, the purpose of which is to deliver more bits to the first 1's encountered for coefficients and fewer to runs of 0's in a bit-plane. The codestream of any subblock in a subband turns out to be progressive, but clearly the aggregate of the subblock codestreams in a subband are not progressive. These codestreams may then be re-organized to produce a progressive aggregate codestream for the subband.

3. Three-Dimensional Coding Algorithms

The JPEG2000 method operates on the three-dimensional subband structure of Figure 2 by encoding in turn each two-dimensional transform slice as described above. This particular implementation is called JPEG2000 Multi-Component (JPEG200-MC). For lossless coding, the coding proceeds through the bottom bit-plane. For lossy coding, a rate control algorithm is run to set cutting points for each subblock's codestream according to the target bitrate. When decoding to a desired rate, transmission of each subblock's codestream is terminated at the corresponding cutting point.

For both JPEG2000 and SBHP, we can gather together the codestreams of subblocks associated with a specific region of the image, decode and display them. This gives us the capability of random access in the codestream to any desired image region. For further details on three-dimensional SBHP, see [6].

Given this random access capability of SBHP, we also extended it to four dimensions for encoding fMRI and fCT images. (See [7] for further details.) We used subblocks of dimensions 64x64x4x4.

The details of these set partitioning methods are beyond the scope of this article. The aim of these methods is to locate individual pixels whose magnitudes equal or exceed a threshold that is a power of 2 (called a *significant* pixel) and sets of pixels whose magnitudes fall below the threshold (*insignificant* sets). When a cubic set is significant, each dimension is split into 2, to form 8 subcubes. When a tree set has one or more significant pixels, it is split into the immediate offspring of its root (8 for three dimensions) and the set of all descendants of offspring. A significant pixel for threshold 2^n requires no more than *n* raw bits to encode, while the insignificant sets are signified with a single 0 bit. The results of these significance tests (called the *significance map)* are signified with 1 or 0, for significant or insignificant, respectively. This threshold is successively lowered by a factor of 2 to test insignificant pixels or sets found at the previous threshold. In this way, we can economically locate insignificant sets and code significant pixels with a minimum of raw bits.

4. Coding Results for Medical Images

Several coding systems were compared in performance on three- and four-dimensional medical images stored with 8 bits per pixel. To ensure perfectly lossless reconstruction, the wavelet filters reversibly mapped integers at the input to integers at the output without truncation

of precision. The three-dimensional wavelet transforms included a full transform of all the slices in the z-direction, except for SBHP coding, where groups of 16 slices were transformed. All the set partitioning methods produced progressive resolution codestreams. SPIHT and SBHP were enabled also for random access decoding. SPIHT has the lowest granularity of 2x2x2 for random access. The granularity for the other methods is the code-block size, 64x64x4 for SBHP, and 64x64x1 for JPEG000-MC. SPECK is not included in these performance comparisons, because it was not enabled for random access decoding. JPEG2000-MC naturally produces a progressive resolution codestream, but was not enabled for random access decoding, but is included for performance comparisons, as seems mandatory nowadays. JPEG2000 coding of two-dimensional wavelet transforms of every *xy*-slice is included to show the performance gains of three-dimensional transform and coding.

We first show some rate results for lossless coding in Table 1. The numbers in the table are the rates in bits per pixel (b/p). They all show considerable reduction from the uncompressed rate of 8 b/p. SPIHT shows the lowest bit rates in three of the four images, with SBHP not much larger. JPEG2000-MC does considerably better on the MR images than on the CT ones and, in fact, is the winner for MR_sag_head.

Table 1. Lossless Coding Rates (b/p) for 3-D Medical Images

Image Name	SPIHT	SBHP	JP2K-MC	JP2K
CT_Skull	**2.16**	2.27	2.93	3.00
CT_Wrist	**1.27**	1.40	1.78	1.76
MR_sag_head	2.35	2.32	**2.30**	2.91
MR_ped_chest	**1.92**	2.09	2.00	3.11

The competitiveness of SBHP is noteworthy, because it has the least computational and memory complexity among these methods. The SPIHT rates in this table were obtained using arithmetic coding of the significance map only. The SBHP rates were achieved with fixed Huffman codes of 15 symbols for three simple contexts.

In decoding the lossless codestream written by these methods, one can obtain reconstructions at almost any desired rate. In Table 2, we show the PSNR values (PSNR = 10 $\log_{10}$ (255^2/(Mean Squared Error))) resulting when decoding two images at a few different rates.

The PSNR results of JPEG2000-MC are roughly 1.5 dB higher than SPIHT on average. However SPIHT is more flexible and less complex than JPEG2000-MC, especially in this version that uses no arithmetic coding. SBHP is the simplest of the three, and consequently has the worst performance, which is still very respectable.

Table 2. PSNR's of reconstructed images decoded from lossless codestreams.

	Rate (b/p)	1.0	0.5	0.1
CT_Skull (256x256x192)	SPIHT	50.71	45.00	35.47
	SBHP	47.17	42.92	33.36*
	JP2K-MC	52.23	46.31	36.10
MR_sag_head (256x256x56)	SPIHT	51.76	46.45	38.30
	SBHP†	48.29	45.77	39.08*
	JP2K-MC	52.93	47.38	38.89

*Rate = 0.125 b/p †Uses 48 slices

We performed decoding of the CT_Skull volume image compressed at 1.0 b/p at ¼, ½, and full *xy*-plane resolutions. The first slices of the reconstructions are displayed in Fig. 3.

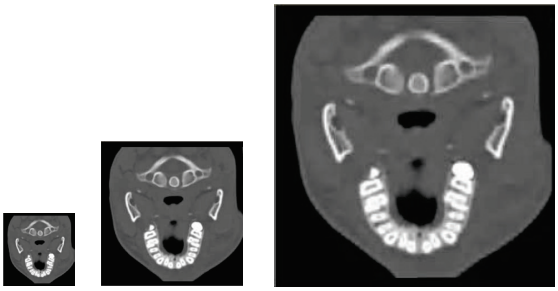


Figure 3. First slice view of CT_skull volume decoded progressively in resolution from 1.0 b/p file. From left to right: 1/4, 1/2 and full resolution.

In Fig. 4, we present a volume rendered display of the entire sequence decoded at these three resolutions from a lossless codestream of CT_Skull. The number of bytes read and decoded is shown below each image. Decoding the ¼ resolution image is 27 times faster and decoding the ½ resolution image is 4.9 times faster than decoding the full resolution image.

Whether decoded from a 1.0 b/p or lossless file, the features and character of the image are clearly recognizable even from the lowest resolution image.

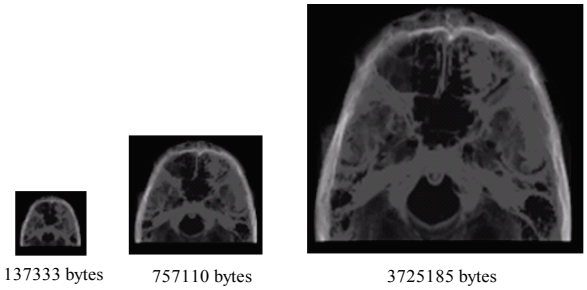


Figure 4. Volume display of CT_Skull image at ¼, ½, and full resolutions. Under each image is the number of compressed file bytes that produced it.

We also selected regions of interest to decode directly from the CT_Skull volume. They are shown in Fig. 5.

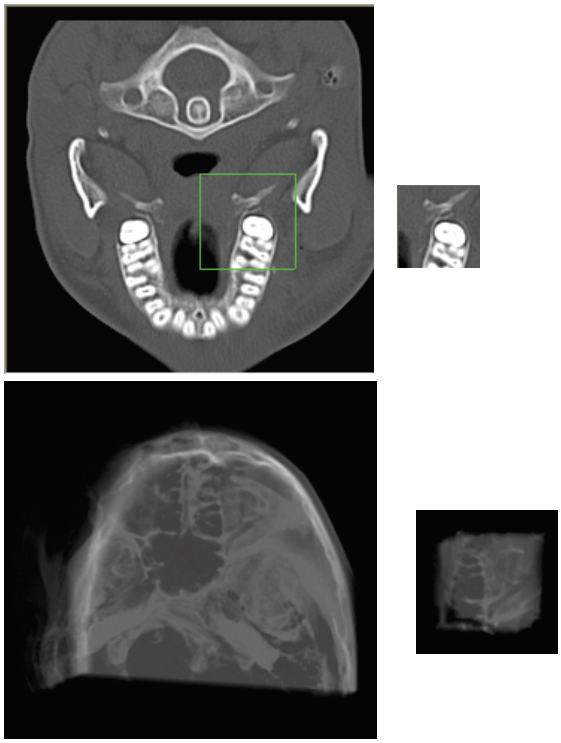


Figure 5. Regions of interest decoded directly from lossless CT_Skull codestream. Top is 2-D view of selection. Bottom is 3-D view of interior region from coordinates (134,117,17) to (198,181,112).

Finally, we extended the SBHP method to four dimensions, the fourth dimension being time *t*, while maintaining all its features. For lossless compression of seven four-dimensional medical images, 4D-SBHP showed an average of 9.7% reduction in bit rate (file size) from encoding each *xyz* cube (or time snapshot) separately and a smaller reduction of 2.0% from encoding each *xyt* cube (time sequence of *xy* slices) separately.

Reconstructions from progressive rate decoding from the lossless codestream are shown in Fig. 6 for the "siem" fMRI image of dimensions 64x64x16x120. In this figure appear volume images at time *t=0* decoded at rates 0.5, 1.0, and 2.0 b/p. In Fig. 7, are shown the full and half resolution volume images at *t=0* decoded from the 0.25 b/p codestream.

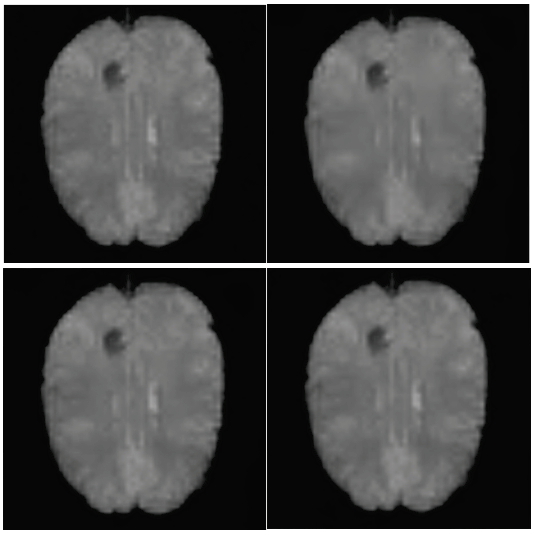
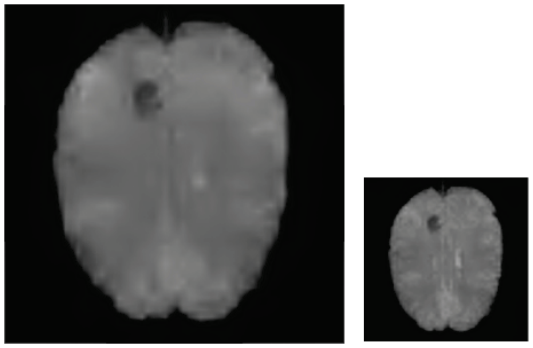


Figure 6. Progressive quality (rate) 4-D SBHP decoding of "siem"fMRI image (64x64x16x120) at time *t=0*. Clockwise from top left: original, 0.5, 2.0, 1.0 b/p.



0.25 b/p full resolution ½ resolution

Figure 7. 4D_SBHP decoding of ½ resolution volume from "siem" fMRI image at *t=0* from 0.25 b/p codestream.

5. Conclusions

Lossless compression of medical images brings not only savings in storage and communication efficiency and speed, it can bring valuable decoding features of progressive decoding in resolution and quality and codestream random access. We specifically demonstrated these features for SBHP, because it is the simplest of the algorithms offered for image compression and delivers the desired features efficiently and effectively. For example, comparisons of 3D SBHP with 3D SPIHT using arithmetic coding of the significance map showed 6 times faster encoding and 6 to 10 times faster in decoding. In other words, the savings in transmission and processing time to get these features from an original or conventionally compressed image are still quite substantial.

Acknowledgements
I wish to acknowledge the contributions of Ying Liu and Emmanuel Christophe, to whom I am indebted for performing much of the work described in this article. Dr. Liu was my doctoral advisee at the time and is now working at ITT Systems in Rochester, NY. Dr. Christophe was a visiting doctoral student from University of Toulouse in France and is now employed at Google in Mountain View, CA. I also gratefully acknowledge the support of the Office of Naval Research under Grant No. N00014-05-1-0507.

References

[1] A. Said and W.A. Pearlman, "A new, fast and efficient image codec based on set-partitioning in hierarchical trees", *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 6, pp. 243-250, June 1996.

[2] W. A. Pearlman, A. Islam, N. Nagaraj, and A. Said, "Efficient, Low-Complexity Image Coding with a Set-Partitioning Embedded Block Coder," *IEEE Trans. Circuits and Systems for Video Technology* , Vol. 14, pp. 1219-1235, Nov. 2004.

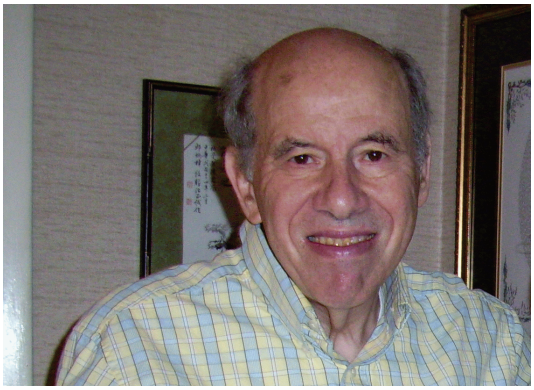
[3] C. Chrysafis, A. Said, A. Drukarev, A. Islam, and W. A. Pearlman, "SBHP- A Low Complexity Wavelet Coder," IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP 2000), Istanbul, Turkey, June 5-9, 2000..

[4] D. S. Taubman and M. W. Marcellin, *JPEG2000:Image Fundamentals, Standards and Practice,*,Boston/DorDrecht/London:Kluwer Academic Publishers, 2002.

[5] E. Christophe and W. A. Pearlman, "Three-dimensional SPIHT Coding of Volume Images with Random Access and Resolution Scalability," *Eurasip J. on Image and Video Processing*, vol. 2008, Article ID 248905, 2008. doi:10.1155/2008/248905

[6] Y. Liu and W. A. Pearlman, "Scalable Three-Dimensional SBHP Algorithm with Region of Interest Access and Low Complexity," *Applications of Digital Image Processing XXIX* , Proc. SPIE Vol. 6312, pp. 631209-1--11, Aug. 2006.
.

[7]] Y. Liu and W. A. Pearlman , "Four-Dimensional Wavelet Compression of 4-D Medical Images Using Scalable 4-D SBHP," Proc. *IEEE 2007 Data Compression Conference,* pp. 233-242, Mar. 2007.



William A. Pearlman is currently Professor Emeritus in the Electrical, Computer and Systems Engineering at Rensselaer Polytechnic Institute (RPI). He obtained his B.S. and M.S. degrees at MIT in 1963 and his Ph.D. at Stanford University in 1974. He joined RPI in 1979 and became Professor in 1988. Prior to joining the RPI, he had been a faculty member at the University of Wisconsin-Madison for five years. .He has authored or co-authored more than 200 publications in the fields of image and video compression, information theory. He is a Fellow of IEEE and of SPIE- The International Society for Optical Engineering. He received the IEEE Circuits and Systems Society 1998 Video Technology Transactions Best Paper Award and the IEEE Signal Processing Society 1998 Best Paper Award in the Area of Multidimensional Signal and Image Processing. He is co-inventor of two celebrated image compression algorithms, SPIHT and SPECK. He is lead author of the textbook, *Digital Signal Compression: Principles and Practice,* by William A. Pearlman and Amir Said, to be issued this year by Cambridge University Press.

Compression of 3D Medical Images for Wireless Transmission

Victor Sanchez[1] and Panos Nasiopoulos[2]
[1]University of California Berkeley, USA, [2]University of British Columbia, Canada
[1]victors@eecs.berkeley.edu ,[2]panos@ece.ubc.ca

1. Introduction

Picture archiving and communication systems (PACS), which contain a collection of specialized networks and software, are commonly used for storage and distribution of 3D medical images. In recent years, the wide pervasiveness use of telemedicine technologies has motivated the integration of mobile devices, such as Personal Digital Assistants, into PACS in order to allow immediate diagnosis by a doctor at any time and place [1]. Consequently, telemedicine applications require that 3D medical images be efficiently transmitted over error-prone wireless networks of various bandwidth capacities.

In this letter, we introduce a compression method with optimal channel protection for transmission of 3D medical images over error-prone wireless networks. Our method, which is based on the 3D integer wavelet transform (3D-IWT) and the embedded block coding with optimized truncation (EBCOT) algorithm, allows compression of 3D medical imaging data into a layered bit-stream that is scalable in quality and resolution, up to lossless reconstruction. The method features optimal channel protection, which is achieved by employing an optimization technique that assigns protection bits to the different sections of the compressed bit-stream based on their mean energy content. The method realizes channel protection by concatenating a cyclic redundancy check (CRC) outer coder and an inner rate-compatible punctured convolutional (RCPC) coder.

We evaluated the robustness of the proposed method over a Rayleigh-fading channel, which effectively models the effect of a propagation environment on radio signals used by wireless devices. Performance comparisons on real magnetic resonance imaging (MRI) volumes are made with the cases of equal channel protection (ECP) and unequal channel protection (UCP). Our results show that the proposed method outperforms the ECP and UCP techniques over a variety of channel conditions and transmission bit-rates.

2. The proposed compression method

The proposed compression method with optimal channel protection is depicted in Fig. 1. At the encoder side, we first apply a 3D-IWT with dyadic decomposition and R levels of decomposition to an input 3D medical image. This type of wavelet transform guarantees perfect invertibility and thus allows for perfect reconstruction of the signal [4]. After the 3D-IWT, we group the wavelet coefficients into code-cubes of $a\times a\times a$ samples. We employ a pyramid approach to define the size and position of code-cubes across the different decomposition levels, so that a code-cube of $a\times a\times a$ samples at position $\{x,y,z\}$ in a particular sub-band at decomposition level r depicts the same spatial information as the code-cube of $a/2\times a/2\times a/2$ samples at position $\{x/2,y/2,z/2\}$ in the equivalent sub-band at decomposition level $r + 1$, where $r = 1$ is the first decomposition level (see Fig. 2). We encode each code-cube independently using a modified EBCOT with 3D contexts to create a separate scalable layered bit-stream for each code-cube [5,6]. We then generate the compressed bit-stream representing the 3D image by collecting the incremental contributions from the various code-cubes into a number of quality layers, so that the code-cube contributions result in an optimal rate-distortion representation of the 3D image, for each quality layer L [5,6]. We use the information about the mean energy of the wavelet coefficients comprising each code-cube in an optimization process to optimally channel-protect the coded code-cubes, so that more protection bits are assigned to those coded code-cubes containing the most energy.

At the decoder side, after transmission over an error-prone wireless network, we first decode the channel-protected data and employ an error concealment technique to minimize the effect of channel errors. We then obtain the wavelet coefficients by applying the EBCOT decoder. Finally, we obtain the re-constructed 3D image by applying an inverse 3D-IWT.

3. Optimal channel protection assignment

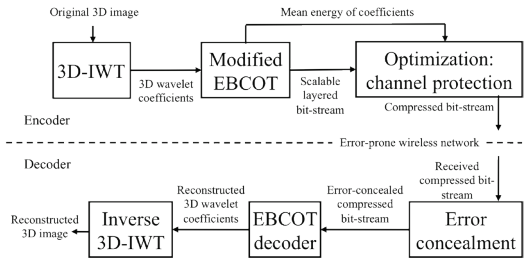


Fig. 1. Proposed compression method. 3D-IWT: three-dimensional integer wavelet transform. EBCOT: embedded block coding with optimized truncation.

We assign channel protection to the compressed bit-stream representing the 3D image based on the effect of bit-errors in each code-cube bit-stream to the overall mean-square error (MSE) of the reconstructed 3D image. Due to the entropy coding process of EBCOT, the distortion in the reconstructed 3D image depends on both the number and position of the bit-errors. A bit-error in the initial few bits of a code-cube bit-stream generally results in higher distortion compared to a bit-error in the later bits, since the initial few bits comprise the most significant bit-planes.

Code-cubes are encoded independently from each other and thus, bit-errors in one code-cube bit-stream do not propagate to others. In order to further limit error-propagation within a single code-cube bit-stream, we employ an error concealment technique at the decoder side. In this error concealment technique, after the occurrence of the first bit-error in a bit-plane, we assign a value of zero to the current and subsequent bit-planes, so that the MSE of a code-cube does not increase any further. Under this scenario, the maximum MSE (MMSE) of a code-cube i at quality layer L [hereafter referred to as code-cube (i, L)] is equal to its mean energy (i.e., errors in all the bit-planes of a code-block):

$$\mathrm{M}_{i,L} = \frac{1}{K}\sum_{k=1}^{K}\left(c_k - \hat{c}_k\right)^2 \tag{1}$$

where c_k is the kth sample of code-cube (i, L), $\hat{c}_k$ is the quantized representation of the kth sample of code-cube (i, L) associated with the bit-stream contribution to quality layer L, and K is the total number of samples in code-cube (i, L). The MMSE of code-cube (i, L) in sub-band s on a per-voxel basis over the entire 3D image may then be calculated as:

$$\overline{\mathrm{M}}_{i,L} = \frac{g_s}{N_s}\frac{q_s}{Q}\mathrm{M}_{i,L} = 2^{2r}\frac{g_s}{N_s}\mathrm{M}_{i,L} \tag{2}$$

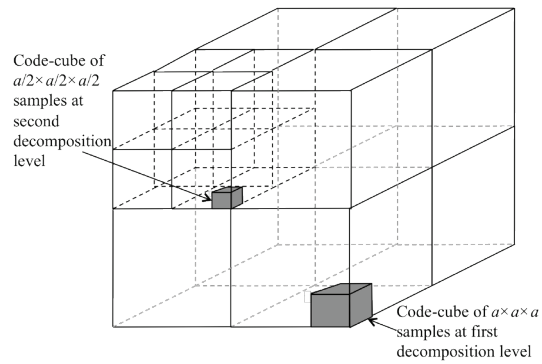


Fig. 2. A code-cube of $a\times a\times a$ samples at the first decomposition level and the equivalent code-cube of $a/2\times a/2\times$a/2 samples at the second decomposition.

where Q is the total number of image voxels, r is the decomposition level to which code-cube (i, L) belongs (r = 1 denotes the first decomposition level), $q_s = Q/2^{2r}$ is the number of wavelet coefficients in s, N_s is the number of code-cubes in s (the code-cubes are of equal size), $\mathrm{M}_{i,L}$ is as defined in (1), and g_s is a factor used to compensate for the non-energy preserving characteristics of the bi-orthogonal Le Gall 5/3 wavelet filter [7].

The overall distortion of the 3D image at quality layer L can be then expressed as the summation of the individual distortions associated to each code-cube (i, L) multiplied by the probability of channel error P_e. The probability of channel error P_e is estimated from the current channel conditions and the RCPC coding rate chosen over a Rayleigh-fading channel [8,9]. For a 3D image coded using a total of I code-cubes, the overall distortion at quality layer L is then:

$$D^L = \sum_{i=1}^{I}\overline{\mathrm{M}}_{i,L}\cdot\mathrm{P}_e \tag{3}$$

where $\overline{\mathrm{M}}_{i,L}$ is as given in (2).

For a fixed target transmission rate, some of the code-cube bit-streams may have to be discarded in order to accommodate for the protection bits. Hence, the distortion in (3) can be expressed as follows:

$$D^L = \sum_{i=i}^{I}\overline{\mathrm{M}}_{i,L}\cdot\mathrm{P}_e\cdot\delta(i) + \sum_{i=i}^{I}\mathrm{m}_{i,L}\cdot\left[1-\delta(i)\right] \tag{4}$$

where $\mathrm{m}_{i,L}$ is the amount of MSE that will be added to the overall distortion if the bit-stream of code-cube (i, L) is discarded, and $\delta(i)$ is 1 if the bit-stream of code-cube (i, L) is included, otherwise it is zero.

We find the optimal channel protection at quality layer L by minimizing D^L in (4) under the following bit-rate constraint:

$$\sum_{i=1}^{I} \frac{S_{i,L}}{R_{i,L}} \cdot \delta(i) \le R_{T,L} \tag{5}$$

where $R_{i,L}$ is the channel code rate for the bit-stream of code-cube (i, L), $S_{i,L}$ is the number of bits in the bit-stream of code-cube (i, L), and $R_{T,L}$ is the available transmission bit-rate at quality layer L.

We solve the optimization problem in Eqs. (4)-(5) by finding the points that lie on the lower convex hull of the rate-distortion plane corresponding to the possible sets of bit-stream assignments.

4. Performance evaluation

We tested the performance of our proposed method over a simulated Rayleigh-fading channel, which effectively models the fading effect on radio signals used by wireless devices in built-up urban areas where buildings and other objects attenuate, reflect, refract and diffract the signals [10]. We employed Jakes' model to simulate a Rayleigh-fading channel, where the channel conditions are specified by the average received signal-to-noise ratio ($\overline{\text{SNR}}$) over the channel [10]. A low $\overline{\text{SNR}}$ value corresponds to poor channel conditions, whereas a high $\overline{\text{SNR}}$ value corresponds to good channel conditions. We used an MRI volume as the test image. The MRI volume comprises 50 slices of a human spinal cord [saggital view, 512×512 pixels per slice (pps), 8 bits per voxel (bpv)]. In order to obtain different channel protection rates, we punctured with a period of eight, the convolutional mother code of rate 1/4 and generator matrix g = [23 35 27 33] (in octal notation) [8]. The decoding process was performed using the Viterbi algorithm [9].

We decomposed the test images with four levels of decomposition in all three dimensions. We employed *32×32×32* samples per code-cube to create a scalable layered bit-stream with 20 quality layers, whose reconstruction quality progressively improves up to lossless reconstruction. We divided the code-cubes bit-streams to be channel-protected into smaller bit-streams of 384 bits. Each of these smaller bit-streams was first protected by an outer 16-bit CRC code defined by the polynomial 210 421 (in octal notation), followed by an inner RCPC code. The information regarding the channel code rates and number of protected code-cube bit-streams is assumed to be common knowledge to both the encoder and decoder and thus, no side information needs to be transmitted. We evaluated the robustness of the proposed method over two different channel conditions ($\overline{\text{SNR}}$ =10dB and $\overline{\text{SNR}}$ =25dB) with frequency-shift keying transmission, a data rate of 15 Kbit/s, a mobile speed of 5 Km/h, and a carrier frequency of 900 MHz, which is one of the operating frequencies for GSM mobile devices [11]. For comparison purposes, we also evaluated an ECP and UCP technique designed for the current channel conditions [12]. Similarly to the proposed method, these techniques employ a 16-bit CRC code (210 421 - in octal notation), followed by an inner RCPC code. The ECP technique assigns protection bits equally across all sections of the compressed bit-stream. The UCP technique assigns protection bits to the different sections according to their mean energy, but unlike the proposed method, it employs no optimization. Both techniques, ECP and UCP, discard code-cube bit-streams to accommodate for the protection bits in a similar manner to the proposed method. In all cases, the decoder performs error concealment on the received data, as explained in section 2

We tested each channel condition with 500 independent trials. Figure 3 shows the average PSNR (in dB) of the received 3D images after transmission at a variety of bit-rates. It can be seen that the proposed method achieves the highest average PSNR values over all channel conditions and transmission rates. This is a consequence of the optimization process employed to assign channel protection, in which the code-cube bit-streams containing the most energy of the image are assigned more protection at the expense of reducing the protection assigned to those code-cube bit-streams with low energy content.

5. Conclusions

We presented a 3D medical image coding method with optimal channel protection for transmission over error-prone wireless networks. The method is based on a 3D integer wavelet transform and the EBCOT algorithm and generates a scalable layered bit-stream. The method optimally assigns channel protection to the different sections of the compressed bit-stream according to their mean energy content. The channel protection is realized by concatenating an outer CRC code and an inner

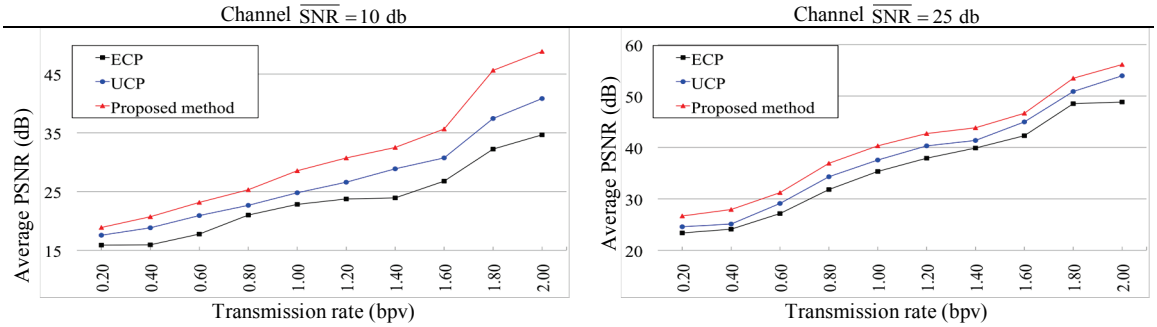


Fig. 3. Average PSNR values (in dB) of the reconstructed MRI slices (saggital view) of a human spinal cord after transmission over a Rayleigh-fading channel with different channel conditions.

RCPC code. We verified the robustness of the proposed coding method over a Rayleigh-fading channel with different channel conditions. Simulation results show that the proposed method outperforms the ECP and UCP techniques over a variety of channel conditions and transmission bit-rates.

References

[1] R. Andrade, A. Wangenheim, M.K. Bortoluzzi, "Wireless and PDA: a novel strategy to access DICOM-compliant medical data on mobile devices," *Int. J. Med. Informatics*, vol. 71, no. 2, pp. 157-163, 2003

[2] S. Lee, T. Lee, G. Jin, J. Hong, "An implementation of wireless medical image transmission system on mobile devices," *J. Med. Syst.*, vol. 32, pp. 471-480, 2008

[3] C. N. Doukas, I. Maglogiannis, G. Kormentzas, "Medical image compression using wavelet transform on mobile devices with ROI coding support," in *Proc. IEEE-EMBS*, pp. 3379-3784, 2005

[4] I. Daubechies and W. Sweldens, "Factoring wavelet transform into lifting steps," *J. Fourier Anal. Appl.*, vol. 41, no. 3, pp. 247-269, 1998

[5] D. Taubman, "High performance scalable image compression with EBCOT," *IEEE Trans. Image Process.,* vol. 9, pp. 1158-1170, 2000

[6] V. Sanchez, R. Abugharbieh, P. Nasiopoulos, "3D scalable medical image compression with optimized volume of interest coding," *IEEE Trans. Med. Imaging.* (in press)

[7] B. Usevitch, "Optimal bit allocation for biorthogonal wavelet coding," in *Proc. Data Comp. Conf.*, Snowbird, UT, pp. 387-395, 1996

[8] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun*, vol. 36, pp. 389-400, 1988

[9] A. Viterbi, "Convolutional codes and their performance in communications systems," *IEEE Trans. Commun. Technol.*, vol.19, pp. 751-772, 1971

[10] W.C. Jakes, *Microwave Mobile Commun.* New York: Wiley, 1974

[11] M. Rahnema, "Overview of the GSM system and protocol architecture," *IEEE Commun. Magazine*, vol. 31, pp. 92-100, 2002

[12] V. Sanchez, M. Mandal, "Efficient channel protection for JPEG2000 bitstream," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, pp. 554-558, 2004

Dr. Victor Sanchez received a Bachelor degree in Electrical Engineering from the Instituto Tecnologico y de Estudios Superiores de Monterrey (199), Mexico; a Master degree in Electrical and Computer Engineering from the Univeristy of Alberta (2002); Canada, and a Ph.D. degree (2010) in Electrical and Computer Engineering from the University of British Columbia, Canada. He is presently a post-doctoral fellow at the University of California, Berkeley. He was been the recipient of research awards from the Consejo Nacional de Ciencia y Tecnologia – CONACYT (2000 and 2006), Mexico; the Canadian Institutes of Health Research – CIHR (2007); and the Natural

Sciences and Engineering Research Council of Canada – NSERC (2010). His research work and interests include signal processing of 3D and 4D data for medical applications, digital communications, information theory and simulation. He has been involved in several projects involving scalable coding of video and imaging data for access and transmission over wired and wireless networks.

Dr. Panos Nasiopoulos received a Bachelor degree in Physics from the Aristotle University of Thessaloniki (1980), Greece, and a Bachelor (1985), Master (1988) and Ph.D. (1994) in Electrical and Computer Engineering from the University of British Columbia, Canada. He is presently the Director of the Institute for Computing, Information and Cognitive Systems (160 faculty members and 800 graduate students) at the University of British Columbia (UBC). He is also a Professor with the UBC department of Electrical and Computer Engineering, the Inaugural Holder of the Dolby Professorship in Digital Multimedia, and the current Director of the Master of Software Systems Program at UBC. Before joining UBC, he was the President of Daikin Comtec US (founder of DVD) and Executive Vice President of Sonic Solutions. Dr. Nasiopoulos is a registered Professional Engineer in British Columbia, the Associate Editor for IEEE Signal Processing Letters, and has been an active member of the Standards Council of Canada, the IEEE and the Association for Computing Machinery (ACM).

Improved Compressibility in JPEG2000 2D and 3D Reversible Compressions of Thin-section Chest CT Images by Increasing the Data Redundancy outside the Body Region

Kil Joong Kim[1], Kyoung Ho Lee[1], Bohyoung Kim[1], Thomas Richter[2, 3], Il Dong Yun[4], Sang Uk Lee[1], Kyongtae Ty Bae[5], Hackjoon Shim[1]

[1]Seoul National University, Korea, [2]University of Stuttgart, Germany, 3Accusoft-Pegasus Imaging, USA, 4Hankuk University of Foreign Studies, Korea, 5University of Pittsburgh School of Medicine, USA

[2, 3]richter@rus.uni-stuttgart.de

1. Introduction

Computed tomography (CT) generates a large amount of image data [1]. Although the cost of storage and network resources has continued to drop, there is increasing demand for data compression of CT images, considering the requirements for long-term preservation and efficient transmission of data, especially between institutions at the regional or national level [2]. While an irreversible compression of data permits a higher degree of compression ratio (CR) than a reversible compression, irreversibly compressed data may be subjected to degradation. Furthermore, a single CR cannot serve as a guideline for the compression of medical images, as the compression artifacts vary considerably with image content, scanning technique, and compression algorithm [3-11]. For example, previous studies advocating the use of irreversible compressions for chest CT images reported a range of variable acceptable thresholds from 4:1 to 10:1 in terms of the CR [8, 12-14]. In contrast, reversible compression would be the safest compression, although it does not provide a very high CR.

In a chest CT image, diagnostically relevant information is typically confined to the body region (*i.e.*, the region of interest, or ROI), whereas the area outside the ROI, including the air, patient clothes, and table, is typically noncontributory to the diagnosis. The compressibility of a chest CT image depends on the degree of data redundancy [15] not only inside the ROI but also outside the ROI, as the image noise, an important factor that decreases the compressibility of an image [3, 9-11], is evenly distributed throughout the entire image, especially in a thin-section image. Therefore, if the data redundancy could be increased selectively outside the ROI in a chest CT image, the overall compressibility of the image would then be improved without affecting the diagnostic information.

This study aimed to propose a preprocessing technique that increases the compressibility in reversible compressions of thin-section chest CT images, and to measure the increase in CR in Joint Photographic Experts Group (JPEG) 2000 two-dimensional (2D) and three-dimensional (3D) compressions.

2. Materials and Methods

Our institutional review board approved this study and waived informed patient consent. We developed a preprocessing technique which automatically segments the body region and replaces pixel values outside the body region with a constant value to maximize data redundancy. One hundred thin-section chest CT scans (50 standard- and 50 low-radiation dose scans) were preprocessed by using the technique. We measured the increase in CR via the preprocessing technique in both JPEG2000 2D and 3D reversible compressions (Fig. 1).

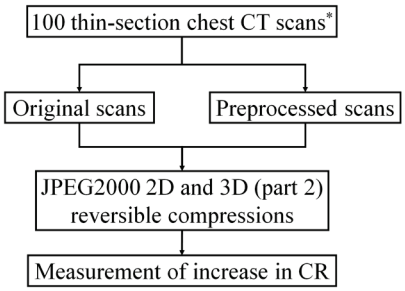


Figure 1: Flow chart of the design of the study. *=studies were performed with standard or low radiational noise (50 samples each)

Only authors who are not employees of or

consultants for Accusoft-Pegasus Imaging Co. (Florida, USA) had control of inclusion of any data and information that might present a conflict of interest for the author (T.R.) who is a consultant of that company.

Study samples

This study consisted of thin-section chest CT scans of 100 adult patients (age: 57.5 ± 12.5 [mean ± SD] years; 57.1 ± 12.1 years for 69 males; and 57.6 ± 12.8 years for 31 females): 50 consecutive scans obtained with the standard radiation dose and 50 consecutive scans obtained with low radiation dose at Seoul National University Bundang Hospital in February of 2009. We included the two scan protocols of different radiation doses which are enumerated in detail in Table 1; both are commonly used in clinical practice, and because we considered it important to test the proposed preprocessing technique at different levels of noise, which is known to affect the compressibility of an image [3, 9-11].

Parameter	Standard-Dose Chest CT	Low-Dose Chest CT
No. of detector rows 3 section thickness (mm)/gantry rotation time (sec)	16×1.5/0.75 for 16-channel multi–detector row CT, 64×0.625/0.42 for 64-channel multi–detector row CT	16×1.5/0.75 for 16-channel multi–detector row CT, 64×0.625/0.42 for 64-channel multi–detector row CT
Tube potential (kVp)	120	120
Pitch	1.071–1.174	1.071–1.174
Effective tube current–time product (mAs) * †	149.3±13.4 (117–172)	23.1±3.9 (16–32)
Effective dose (mSv) * ‡	7.5±0.8 (4.8–9.4)	1.4±0.1 (1.2–1.5)
Scanning range	From vocal cord to adrenal gland.	From lung apex to lung base.
Field of view (mm) *	310.2±29.4 (255–405)	301.6±17.8 (254–342)
Reconstruction thickness (mm)/ interval (mm)	2/1	2/1
Image size	512×512	512×512
Reconstruction fi lter	Medium sharp (filter type C)	Medium sharp (filter type C)

Table 1: CT Scanning Parameters

*: Data are means 6 ± standard deviations, with ranges in parentheses.
†: Automatic tube current modulation was used
‡: Estimated by multiplying the dose-length product measured on the CT console by a conversion factor (0.019 mSv · mGy^{-1} · cm^{-1}) [16].

All scan parameters followed clinical scan protocols of our hospital. Either a 16-channel (n = 44) or a 64-channel multidetector row CT (n = 56) scanner was used (Brilliance; Philips Medical Systems, Cleveland, OH). Scans were acquired during inspiratory breath-hold while patients raised their arms above the shoulders. The field of view (FOV) in each patient was determined as small as possible, while covering the entire thorax for the standard-dose scans and focusing on the lungs for the low-dose scans. Other scan parameters are tabulated in Table 1. The number of images was 372.8 ± 38.6 (mean ± SD) for the standard-dose and 358.3 ± 23.2 for the low-dose scan. Since this study was not intended to evaluate diagnostic performance, potential abnormalities contained within the scans were not considered to be important.

Segmentation method

To identify the ROI in each scan, we developed an automatic segmentation method. It consisted of four steps: rough extraction of the body region, removal of superfluously extracted regions, inclusion of inner holes, and expansion of the ROI. The last step (addition of several pixels beyond the body contour to the ROI) was carried out to ensure the preservation of data inside the body region. The degree of expansion was empirically determined as six pixels by a radiologist (K.H.L., with seven years of clinical experience) in a separate experiment. The first three steps were implemented and processed in a slice-to-slice operation, while the last step was applied to the entire scan. The details of the segmentation method are described in the Appendix. The source code of the segmentation algorithm is available as a supporting document.

Preprocessing

Each of the 100 scans was preprocessed. After the ROI of each scan was determined by using the aforementioned segmentation method, CT number of pixels outside the ROI were replaced by a constant value which corresponded to the median CT number of all the pixels outside the ROI throughout the scan. The rationale for using the median CT number was to maintain the

global luminance level of the preprocessed images close to that of the original images and thus to minimize potential change in luminance adaptation by radiologists, since different luminal adaptation can differently affect the visual perception (16).

Two radiologists independently evaluated the accuracy of the segmentation results of the 100 scans. One of the radiologists (K.H.L.) had seven years of clinical experience while the other, not included as an author, had 10 years of clinical experience. Individual images of each scan were scrolled through and reviewed for the segmentation boundaries superimposed on the original images. If the segmented ROI was determined by both radiologists to completely cover the body region including the skin, the segmentation ROI was considered correct; otherwise, it was considered incorrect.

The computing time of the preprocessing was measured for each scan and then divided by the number of images in the scan. In this calculation, we included the time necessary for loading the input image files into computer memory as well as saving the output image files to storage. We implemented the preprocessing technique using Visual C++ (version 6.0; Microsoft, Redmond, WA) and used a PC platform running Windows XP (Microsoft) with a 2 GHz dual-core processor (Xeon 5130; Intel Co., Santa Clara, CA) and 3 Gbyte of main memory.

In each scan, the *percentage volume outside ROI* was calculated as follows:

$$\frac{\sum_{z=1}^{N}\sum_{y=1}^{512}\sum_{x=1}^{512} M(x,y,z)}{512 \times 512 \times N} \times 100\%$$

where N was the number of images in the scan and $M(x, y, z)$ was 1 outside the ROI and 0 inside the ROI.

Image compression

The 100 original and 100 preprocessed scans were reversibly compressed using both the JPEG2000 2D and 3D (part 2, Multi-Component Transformation extensions) algorithms (Accusoft-Pegasus Imaging Co., Tampa, FL). The Multi-Component Transformation was here a wavelet transformation in Z-direction, implementing together with the regular 2D wavelet decomposition of JPEG 2000 a volumetric decomposition.

Both compression algorithms have been adopted in Digital Imaging and Communications in the DICOM standard. JPEG2000 2D compresses an image by exploiting the data redundancy within the image (i.e., intra-slice correlation) in horizontal and vertical directions only, whereas JPEG2000 3D compresses the datasets additionally in the third dimension through exploitation of the data redundancy between adjacent images (i.e., inter-slice correlation). Both encoders were set to their default settings [6].

The CT images had a pixel bit depth of 12 bits (1.5 bytes) of information (attenuation number raging from -1024 to 3072 HU) per pixel. However, because of practical reasons, the CT images were saved in 16 bits (2 bytes) per pixel with four padding bits. That is, images were encoded as 16 bits/pixel images with the four most significant bits set to zero. This type of padding is typical in the medical world and only justified by the existing toolchain.

For each scan, the CR was defined as the original data size (16 bits/pixel) divided by the compressed data size (bits/pixel) (17). Henceforth, the CRs for the original and preprocessed scans were denoted as $CR_{original}$ and $CR_{preprocessed}$, respectively. The increase in the CR via the preprocessing technique was measured for each scan as the *percentage increase in CR*:

$$\frac{\left(CR_{preprocessed} - CR_{original}\right)}{CR_{original}} \times 100\%$$

Statistical analysis

The accuracy of the segmentation and corresponding Wilson 95% confidence intervals (CIs) [18] were calculated. The $CR_{original}$ and $CR_{preprocessed}$ results were compared using the paired t tests. The linear regression coefficient was calculated between the *percentage increase in CR* (dependent variable) and the *percentage volume outside ROI* (independent variable). Since the compressibility of CT images are affected by compression algorithms and scanning protocols [3, 6, 9-11], the study results were reported at each of the four combinations of the two compression algorithms (JPEG2000 2D and 3D compressions) and the two scan protocols (standard-dose and low-dose scans). A p-value less than .05 was considered to indicate a statistically significant difference. Statistical

software (StatsDirect; StatsDirect, Altrincham, Cheshire, United Kingdom) was used.

3. Results

Segmentation accuracy

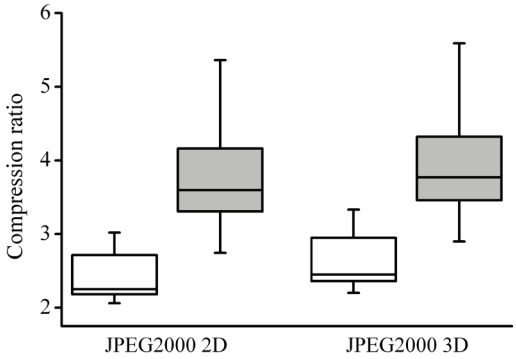
All the ROIs, segmented automatically from the 100 scans using the proposed preprocessing technique, were deemed correct by the two radiologists (100%; 95% CI, 96.3-100.0%). Interobserver agreement could not be calculated because there was no disagreement between the two radiologists.

Computing time of the preprocessing technique

The preprocessing technique required 4.1 ± 0.4 (mean ± SD) and 3.8 ± 0.3 minutes per scan for the standard-dose and low-dose scans, respectively. The computation time per image was 655.3 ± 27.5 and 641.9 ± 31.6 milliseconds per image for the standard-dose and low-dose scans, respectively.

Increase in CR using the preprocessing technique

The CR for the standard-dose scan increased significantly with the preprocessing technique in the JPEG2000 2D (without vs. with the preprocessing, 2.40 ± 0.30 [mean ± SD] vs. 3.80 ± 0.70, $p < .001$) and 3D (2.61 ± 0.34 vs. 3.99 ± 0.73, $p < .001$) compressions. Similarly, the CR for the low-dose scans increased significantly with the preprocessing technique in the JPEG2000 2D (2.38 ± 0.12 vs. 3.36 ± 0.28, $p < .001$) and 3D (2.54 ± 0.13 vs. 3.55 ± 0.31, $p < .001$) compressions (Fig 2).



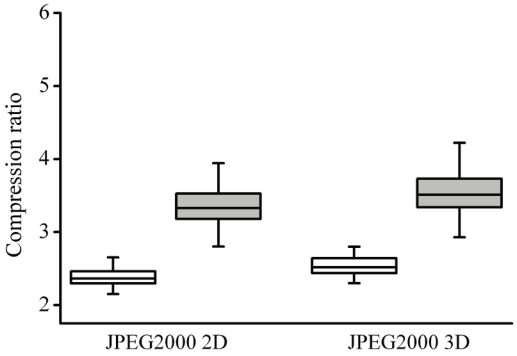


Figure 2: Box-and-whisker plots of reversible CRs in JPEG2000 2D and 3D compressions for original and preprocessed (a) standard-dose and (b) low-dose chest CT studies. For each compression, white and gray boxes = original and preprocessed studies, respect-tively. Middle lines in boxes = medians. Upper and lower margins of boxes = upper and lower quartiles. Ends of the vertical lines = 5th and 95th percentiles.

The mean *percentage increases in CR* per scan were 58.2% (95% CI, 53.1-63.4%) and 52.4% (47.5-57.2%) in the JPEG2000 2D and 3D compressions, respectively, for the standard-dose scans; and 41.1% (38.8-43.4%) and 39.4% (37.4-41.7%) in the JPEG2000 2D and 3D compressions, respectively, for the low-dose scans.

Association between increase in CR and volume outside ROI

The *percentage volumes outside ROI* were 41.5% ± 7.3% (mean ± SD) for the standard-dose and 34.2% ± 4.3% for the low-dose scans. The linear regression coefficients between the *percentage increase in CR* and *percentage volume outside ROI* were 0.95 (95% CI, 0.91-0.97, $p < .001$) and 0.93 (0.89-0.96, $p < .001$) in the JPEG2000 2D and 3D compressions, respectively, for the standard-dose scans; and 0.97 (95% CI, 0.95-0.98, $p < .001$) and 0.99 (0.97-1.0, $p < .001$) for the JPEG2000 2D and 3D compressions, respectively, for the low-dose scans (Fig 3).

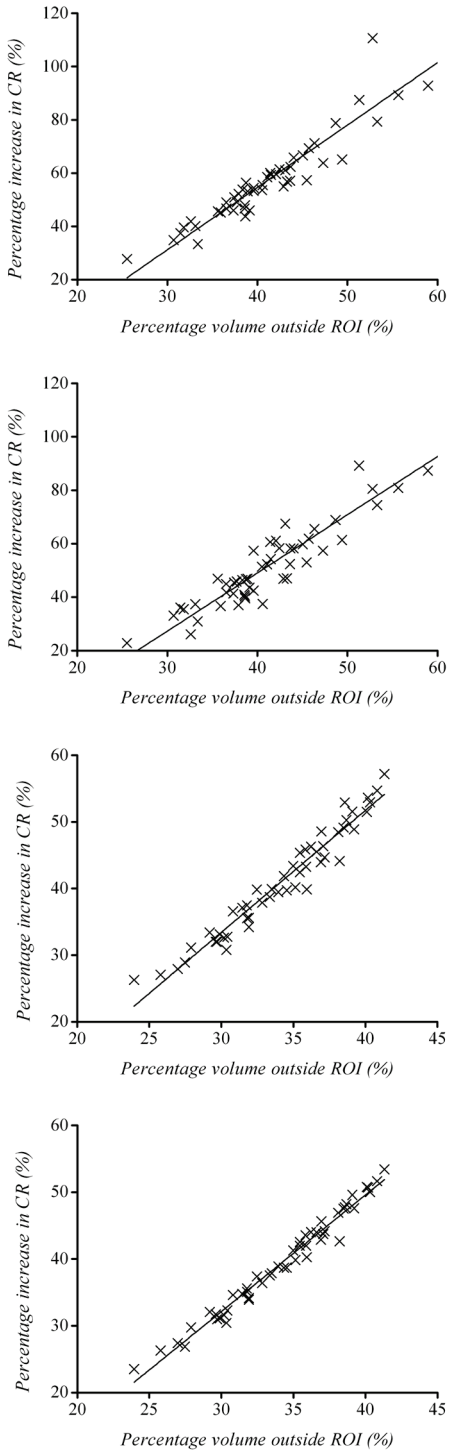


Figure 3: Scatter-plots show the association between percentage increase in CR with the preprocessing technique and percentage volume outside ROI for
(a) JPEG2000 2D (linear regression coeffi cient, r = 0.95) and
(b) 3D compressions (r = 0.93) for standard-dose studies and
(c) JPEG2000 2D (r = 0.97) and
(d) 3D compressions (r = 0.99) for low-dose studies.

4. Discussion

We proposed a preprocessing technique that increases the compressibility in reversible compressions of thin-section chest CT images by increasing data redundancy outside the body region. The results of our study demonstrated that the CR increases considerably with use of our preprocessing technique for the JPEG2000 2D and 3D reversible compressions of thin-section chest CT scans, suggesting considerable saving in system resources required for data storage and transmission without concerns about image degradation inside the body region.

The mean *percentage increases in CR* were 58.2% and 52.4% in the JPEG2000 2D and 3D compressions, respectively, for the standard-dose scans; and 41.1% and 39.4% in the JPEG2000 2D and 3D compressions, respectively, for the low-dose scans.

The increase in the reversible CR with the preprocessing technique was evident across the two compression algorithms and the two scan protocols. Nevertheless, as expected, the degree of increases in CR varied with the compression algorithms and the scan protocols, both of which are known to affect the compressibility of a CT image [3, 6, 9-11]. Had we used a different image sample in terms of image reconstruction algorithm, radiation dose level, or type of scanner, the increase in CR would have been different.

Interestingly, the increase in CR with the preprocessing for the standard-dose scans was greater than that of the low-dose scans. This result seems counterintuitive because the low-dose scans which contain more random noise were expected to gain a greater increase in the data redundancy outside the ROI with the preprocessing technique. Our finding may be explained by differential effects of the volume ratio outside the ROI (*percentage volume outside ROI*) versus the image noise on CR. In our scan protocols, the FOV was set to cover the entire thorax for standard-dose scans and to focus on the lungs for the low-dose scans. Therefore, the *percentage volume outside ROI*, which showed a high linear regression coefficient for the

association with *percentage increase in CR*, was greater in the standard-dose scans than in the low-dose scans. This factor might have played a greater role in determining the increase in CR than the image-noise level. Further investigation is required to systematically understand the increase in CR with the preprocessing technique at different scan protocols.

Earlier studies [19-27] suggested ROI-based compression techniques for medical images. These techniques, similar to ours, aimed to compress the data inside and outside an ROI to different levels. However, our technique has two distinct advantages over the previous techniques in regard to practicability for use in medical image compression. First, previous techniques [20-27] required certain modifications of the compression algorithm in their implementation. This modification necessarily violates, at least in part, the standard of compression algorithms adopted in the DICOM standard. Conformance to the standard is a requisite for medical image compressions that are to be incorporated into a pre-existing PACS. In contrast, the proposed *preprocessing* technique is independent of the compression algorithm. Therefore, it can be incorporated immediately into any PACS. Second, the segmentation methods used in the previous compression techniques were not fully automated [19-22], or the accuracy of the segmentation process was not validated [23-26].

It should be also noted that the proposed technique differs from two ROI coding techniques already adopted in the JPEG2000 standard: the Maxshift method in JPEG2000 part 1 [28] and the General Scaling method in JPEG2000 part 2 [29]. The former method does not allow accurate control of both the image quality *within* and *outside* the ROI[19] because it requires all coefficients within the ROI to be coded before any of the coefficients outside of it to be included in the codestream[28]. A suitably designed encoder could, however, at least encode coefficients within the ROI without any loss and might provide a limited amount of control for the quality of coefficients outside it, while regaining full standard compliance. At the time of writing this article, such an encoder was not yet available and we leave the study of this method to a future work.

With the latter approach, i.e. part 2 ROIs, the ROI can be only a composition of regular shapes such as rectangles or an ellipses [29], and the method has not been widely adopted in industry. Specifically, interoperability within DICOM would be hard to reach.

As shown in our results, the proposed preprocessing technique required approximately four minutes per scan or less than one second per image. This computing time could be reduced by optimizing the source code or through the use of state-of-the-art workstations. Alternatively, the overall preprocessing time for a scan is likely reducible by incorporating the preprocessing method in an on-the-fly manner into the PACS. This is described in the Appendix.
The segmentation method proposed in our study showed segmentation accuracy of 100% with a narrow 95% CI (96.3-100.0%). This result can be attributed to the fact that the body region in chest CT images has well defined boundary facilitating a reliable segmentation. In addition, the segmented body region was further expanded in the final step of the segmentation to improve the fidelity of the segmentation. Nevertheless, it is uncertain if our segmentation method would work for images obtained with different scanners and for images of body regions other then the chest. Further investigation with a greater sample of more heterogeneous nature is required to verify the robustness of the segmentation.

This study has limitations. First, as the test dataset did not include cases with post-surgical or post-traumatic body contour (open thoracotomy or severe subcutaneous emphysema), our segmentation method was not tested in such cases. However, this issue may be less critical for low-dose chest CT scans, as these types of disrupted body contour cases would be very rare in patients undergoing lung cancer screening examinations. Second, we did not examine whether pixel value alteration outside ROIs may hinder clinical interpretation of images inside ROIs. The preprocessing removes the clothes of the patient as well as the table, smoothens the background area by removing image noise, and creates a sharp ROI boundary. We believe that these external changes will not negatively affect the diagnostic performance of radiologists interpreting the anatomy inside the ROI.

In conclusion, the proposed preprocessing technique, which automatically replaces pixel values outside the body region with a constant value, considerably increases CRs for JPEG2000 2D and 3D reversible compressions of thin-

section chest CT scans.

5. Acknowledgements
We would like to thank Prof. Young Hoon Kim in department of radiology in Seoul National University Bundang Hospital who participated as a reader.

T.R. acknowledges the long and fruitful cooperation with Accusoft-Pegasus.

6. Appendices
Segmentation method
We developed an automatic segmentation method which identifies pixels inside the body region (i.e., ROI). The segmentation method consists of four steps: rough extraction of the body region (Fig. 4a), removal of superfluously extracted regions (Fig. 4c), inclusion of inner holes (Fig. 4d), and expansion of the ROI (Fig. 4e).

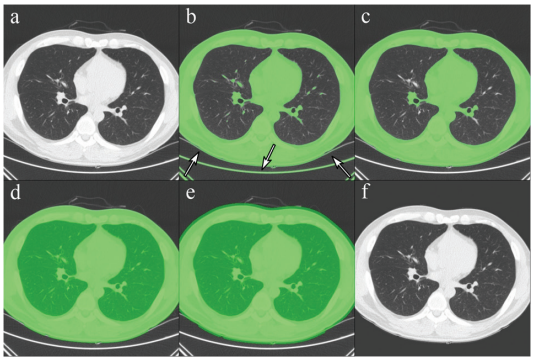


Figure 4: Steps of the segmentation method (see text).

The segmentation process is sequentially operated: the output of one step serves as the input for the next step. The first three steps are applied in a slice-to-slice mode and the last step is carried out over the entire scan. This operational approach was designed to reduce the overall computing time required for the preprocessing of a scan when the preprocessing is incorporated into a picture archiving and communication system (PACS) in an on-the-fly manner. In this way, images can be processed sequentially by the first three steps as they are acquired from a CT scanner. When the last image of a scan is available and processed via the third step, the last segmentation step is applied to the entire dataset of the scan. All the adjustable parameters in our algorithm were determined in a preliminary test with a separate dataset (n = 30), based on judgment by a radiologist (K.H.L., with seven years of clinical experience).

Rough extraction of body region
This step extracts pixels with CT number greater than -400 Hounsfield unit (HU) using a thresholding technique. The threshold value was empirically determined. As the body region in a chest CT scan has higher HU values than the background air (Fig. 4a), the thresholding excludes the background, roughly extracting the body region including the patient table and clothes (note the transparent color overlay in Fig. 4b).

Removal of superfluously extracted region
This step removes superfluously extracted regions from the previous step, such as the patient table and clothes (arrows in Fig. 4b). First, a morphological operator of opening (30) is applied to the roughly extracted body region. The opening operator smoothens out the boundary of the input region by scanning the interior of the region with a structuring element while removing the areas that are not covered by the structuring element. As the structuring element, a circle with a radius of one pixel is used. Because the opening operation isolates regions linked to others with narrow bridges, it detaches the patient table or clothes from the body region. Second, the connected component analysis (30) is applied, resulting in groups of connected pixels. The body region is easily identified as the pixel group with the largest number of pixels (Fig. 4c). As some part of the body region such as the upper shoulders may not be connected to the main body region in some image sections, we additionally identify the pixel groups with more than 100 pixels. The number 100 was also empirically chosen.

Inclusion of inner holes
The result of the previous step does not contain low-attenuating structures such as the lung cavity, trachea, and bronchus in the body region. To include such inner holes in the ROI, the result of the previous step is inverted and the connected component analysis is then applied, yielding the pixel groups of the inner holes and of the background air. With consideration that the background may contain boundary pixels in a section image, the pixel groups which do not contain the boundary pixels are included in the ROI (Fig. 4d).

Expansion of ROI

In order to improve the fidelity of keeping the data inside the body region undisturbed, the ROI result of the previous step is expanded using a morphological operation of dilation (30) (Fig. 4e). The dilation expands the input ROI by driving out the boundary with a structuring element. As the structuring element, a circle with a radius of six pixels is used. The radius (i.e., the degree of dilation) was empirically determined by the radiologist involved in a preliminary experiment.

Replacing CT number of pixels outside ROI with a constant value

Finally, the CT number of every pixel outside the expanded ROI is replaced with a constant value which corresponds to the median CT number of pixels outside ROI throughout the scan (Fig. 4f).

References

[1] Lee KH, Lee HJ, Kim JH, et al. Managing the CT data explosion: initial experiences of archiving volumetric datasets in a mini-PACS. J Digit Imaging 2005; 18:188-195.

[2] Koff D, Bak P, Brownrigg P, et al. Pan-Canadian evaluation of irreversible compression ratios ("Lossy" compression) for development of national guidelines. J Digit Imaging 2009; 22:569-578.

[3] Kim B, Lee KH, Kim KJ, et al. Prediction of perceptible artifacts in JPEG2000 compressed abdomen CT images using a perceptual image quality metric. Acad Radiol 2008; 15:314-325.

[4] Kim B, Lee KH, Kim KJ, et al. Prediction of perceptible artifacts in JPEG2000 compressed chest CT images using mathematical and perceptual quality metrics. Am J Roentgenol 2008; 190:328-334.

[5] Kim B, Lee KH, Kim KJ, Mantiuk R, Kim HR, Kim YH. Artifacts in slab average-intensity-projection images reformatted from JPEG 2000 compressed thin-section abdominal CT data sets. Am J Roentgenol 2008; 190:342-350.

[6] Kim B, Lee KH, Kim KJ, et al. JPEG2000 3D compression vs. 2D compression: an assessment of artifact amount and computing time in compressing thin-section abdomen CT images. Med Phys 2009; 36:835-844.

[7] Kim KJ, Kim B, Lee KH, et al. Regional difference in compression artifacts in low-dose chest CT Images: effects of mathematical and perceptual factors. In:Am J Roentgenol, 2008; 30-37.

[8] Kim TJ, Lee KW, Kim B, et al. Regional variance of visually lossless threshold in compressed chest CT Images: lung versus mediastinum and chest wall. Eur J Radiol 2009; 69:483-488.

[9] Kim KJ, Kim B, Lee KH, et al. Objective index of image fidelity for JPEG2000 compressed body CT images. Med Phys 2009; 36:3218-3226.

[10] Woo HS, Kim KJ, Kim TJ, et al. JPEG 2000 compression of abdominal CT: difference in compression tolerance between thin- and thick-section images. Am J Roentgenol 2007; 189:535-541.

[11] Bajpai V, Lee KH, Kim B, et al. The difference of compression artifacts between thin- and thick-section lung CT lmages. Am J Roentgenol 2008; 191:38-43.

[12] Ko JP, Rusinek H, Naidich DP, et al. Wavelet compression of low-dose chest CT data: effect on lung nodule detection. Radiology 2003; 228:70-75.

[13] Ringl H, Schernthaner RE, Bankier AA, et al. JPEG2000 compression of thin-section CT images of the lung: effect of compression ratio on image quality. Radiology 2006; 240:869-877.

[14] Li F, Sone S, Takashima S, et al. Effects of JPEG and wavelet compression of spiral low-dose CT images on detection of small lung cancers. Acta Radiol 2001; 42:156-160.

[15] Gonzalez RC. Image compression. In: Woods RE, Gonzalez RC, eds. Digital image processing. 2 ed. Upper Saddle River, NJ: Prentice Hall, 2002; 417-419.

[16] Samei E. AAPM/RSNA physics tutorial for residents: technological and psychophysical considerations for digital mammographic displays. Radiographics 2005; 25:491-501.

[17] Kim KJ, Kim B, Choi SW, et al. Definition of compression ratio: difference between two commercial JPEG2000 program libraries. Telemed J E Health 2008; 14:350-354.

[18] Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med 1998; 17:857-872.

[19] Tahoces Pablo G, Ramon VJ, Lado Maria J, Miguel S. Image compression: Maxshift ROI encoding options in JPEG2000. Comput Vis Image Und 2008; 109:139-145.

[20] Strom J, Cosman PC. Medical image compression with lossless regions of interest. Signal Process 1997; 59:155-171.

[21] Menegaz G, Thiran JP. Lossy to lossless object-based coding of 3-D MRI data. IEEE Trans Image Process 2002; 11:1053-1061.

[22] Cuhadar A, Wang K, Tasdoken S. Region of interest image coding for digital mammography. Conf Proc IEEE Eng Med Biol

Soc 2005; 4:3785-3788.
[23] Yin FF, Gao Q. Oncologic image compression using both wavelet and masking techniques. Med Phys 1997; 24:2038-2042.
[24] Gokturk SB, Tomasi C, Girod B, Beaulieu C. Medical image compression based on region of interest, with application to colon CT images. Conf Proc IEEE Eng Med Biol Soc 2001:2453-2456.
[25] Liu Z, Xiong Z, Wu Q, Wang YP, Castleman K. Cascaded differential and wavelet compression of chromosome images. IEEE Trans Biomed Eng 2002; 49:372-383.
[26] Penedo M, Pearlman WA, Tahoces PG, Souto M, Vidal JJ. Region-based wavelet coding methods for digital mammography. IEEE Trans Med Imaging 2003; 22:1288-1296.
[27] Doukas C, Maglogiannis I. Region of interest coding techniques for medical image compression. IEEE Eng Med Biol Mag 2007; 26:29-35.
[28] Information technology - JPEG 2000 image coding system. www.iso.org. Accessed June 1, 2006.
[29] Information technology—JPEG 2000 image coding system: extensions. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=33160. Accessed June 1, 2009.
[30] Gonzalez RC, Woods RE. Morphological image processing. In: Gonzalez RC, ed. Digital image processing. 2 ed. Upper Saddle River, NJ: Prentice Hall, 2002; 519-566.
[31] Bongartz G, Golding SJ, Jurik AG, et al. Bongartz, G, Golding, SJ, Jurik, AG, et al. European guidelines for multislice computed tomography. http://msct.eu/CT_Quality_Criteria.htm. Accessed June 1, 2009.

What is Different About Medical Image Compression?

David A. Clunie
CoreLab Partners, Inc., USA
dclunie@dclunie.com

1. Context

Periodically an article appears in the scientific or medical literature describing some new reversible (lossless) or irreversible (lossy) image compression scheme, and either explicitly or buried within it, is a reference to the scheme's applicability to medical image compression.

It is no secret that the medical community makes extensive use of digital images; indeed greater, if not excessive, use of imaging is one of the major contributors to the burgeoning cost of healthcare [1]. Yet to an outsider, the medical industry appears incredibly conservative about the deployment of new compression technology. Why is this so? What is different about medical image compression?

2. Types of images

The scope of medical imaging is broad, ranging from a simple multi-megapixel single frame digital image of a chest X-ray, through an entire set of tomographic slices through part or all of the body, possibly acquired using hybrid technologies and with a functional as well as structural component, such as a CT-PET scan, through video applications such as visible light endoscopy, to whole slide imaging, which involves multi-resolution acquisition resulting in enormous datasets.

Some of these applications are very similar to consumer applications, particularly single frame digital photography and video recording. In such cases, ordinary consumer compression schemes are routinely applied, albeit often embedded in medically specific storage, management and distribution standards and solutions.

X-ray applications, including single and multi-frame projection images as well as "reconstructed" tomographic slices through the body, are gray scale, but typically require encoding a dynamic range with a resolution beyond 8 bits. The full dynamic range is important to encode information about the structures of different densities. It is usually viewed by the user through a series of "windows" with a narrower range, for example to better see soft tissues, or bone, or air-filled structures. The user needs to be able to adjust this window continuously, hence the full dynamic range data needs to be distributed, not pre-windowed images, despite the fact that the display device is often limited to an 8 bit range.

Further, the numeric pixel data values may have a real world meaning with a well-defined physical unit. For example, X-ray CT images are encoded in Hounsfield Units (HU), a water-relative measurement of linear attenuation and hence related to density. These numeric values have diagnostic significance, so it is important that the values encoded in an image, not just their appearance, be preserved.

3. Uses of images

The most critical use of a medical image is to make a "diagnosis" that affects the patient's management, perhaps better described as "image-based decision making". This may involve the task of detection of one or more abnormalities, as well as their characterization and classification. Usually such interpretations are performed by physicians (such as radiologists) highly trained in the range of diseases they may possibly encounter, as well as highly experienced in the appearance of such diseases and the normal anatomy. Some findings are very obvious, and others are extremely subtle. Some tasks require preservation of high frequency detail, others involve low contrast, and yet others require recognition of subtle changes in particular texture patterns.

What exactly the combination of a human's brain and visual system is doing when making a diagnosis is poorly understood. This is an area of intensive research to the point that a professional society exists to study the matter (Medical Image Perception Society – "http://www.mips.ws/"). A few very specific tasks are amenable to Computer Assisted Detection (CAD) or Diagnosis (CADx), usually to augment rather than replace the human, but as yet, given the broad range of diseases and their appearance, there is no substitute for the human expert.

Calibrated display devices with high luminance, contrast range, and spatial resolution are typically used for diagnostic image interpretation, in an environment controlled for ambient light and specular reflections.

Often distinguished from the critical task of "diagnosis", is the less demanding task of "review". This may involve medical staff with less imaging expertise viewing the image in combination with a radiologist's report, or performing a less demanding task, or one for which they have been specifically trained, such as confirming the appropriate position of a tube or catheter. Arguably such tasks may be performed using images of lesser quality, subject to greater degradation (e.g., more irreversible compression), or displayed on less capable devices or in an ambient lighting environment that is less well controlled.

However, there is a continuum between "review" and "diagnosis" and often the two are not so easily distinguished. In particular, the fallacious assumption is often made that remote users with access over lower bandwidth channels, performing more urgent decision making tasks, perhaps off-hours, can make do with lesser quality. It isn't good enough to make a diagnosis on a perfect image in the morning if the patient died in the middle of the night as a consequence of reviewing a degraded image.

Another distinct "review" task is that of comparison with previous images when interpreting new studies. This is most important, since the ability to detect change significantly affects performance, and because the change itself may be clinically significant (e.g., the progression of the disease or the response to therapy). Arguably, for some tasks, the priors may not need to be of the same quality as the current image.

4. Risk Aversion and Cost Tradeoffs

It is no secret that physicians, particularly in the United States, have become incredibly risk averse, partly as a consequence of the real or perceived threat of malpractice litigation. Despite the fact that irreversible compression has not yet been considered in a court of law [2], the concept of "throwing away" information that was acquired, at first thought, is anathema to radiologists.

However, physicians (or their infra-structure providers) are also sensitive to cost, particularly costs that are not reimbursed, and neither bandwidth nor storage is free. We like to say nowadays, that "disk is cheap", and sure enough it is, but the energy required for power and cooling, is not. Further any decline in storage hardware (if not operational) costs, may be offset by the greater usage of imaging (more studies), and the greater size of images for each study as acquisition technology advances (more, thinner slices acquired faster, for example).

Physicians are also intolerant of inconvenience, and any significant, or even perceptible, delays in the display or navigation of images are unacceptable, and in urgent cases, delay might lead to patient harm [3]. The greater size of each dataset per study exacerbates this effect.

So, even in a local network environment where the facilities are optimized to maximize performance of distribution, perhaps using a single vendor solution, there are still drivers towards compromise that suggest the use of irreversible compression for some less demanding tasks.

If one extends the task to include use of images on a broader geographic scale, the costs rise dramatically, leading to greater pressure to compromise. Yet there is no reason why a patient should have a repeat scan exposing them to cost, discomfort and radiation if an "outside" scan of adequate quality is remotely available. Nor should a geographically remote expert's opinion be unavailable. Nor should opportunities to more efficiently geographically deploy staff and resources be precluded. Sometimes radiologists simply prefer to work at home.

The greater use of mobile devices with less capable display characteristics is another increasingly important factor; there is little point in delivering images of a significantly greater quality than can be displayed, nor should legitimate "closer to the bedside" use cases, which offer the potential to improve patient care, be precluded by limited wireless bandwidth. Mobile devices expedite the timeliness, not just the convenience, of the interaction, and have the potential to improve the efficiency of staff through greater multi-tasking independent of physical location.

Ideally, a compromise could be reached, in which an optimal level of "irreversible"

compression could be defined for any particular class of tasks that is "diagnostically acceptable" [4], not just to contain costs, but to enable implementation of use cases that would otherwise be unachievable.

5. Diagnostically Acceptable Irreversible Compression

Whilst the mathematical literature introduces new compression schemes with a small number of test images demonstrating potential, it remains for the medical literature to evaluate such schemes on a sufficient number of real clinical images to demonstrate safety and efficacy. And therein lies the rub. To conduct such studies "properly" turns out to be extraordinarily expensive. When comparing compressed images to uncompressed images (or images compressed with different schemes or parameters), significant statistical power is required to demonstrate that the absence of any difference in performance found is not just due to an inadequate sample size (type II error). Further, it is necessary to define how much of a difference in performance actually "matters" (i.e., to make the distinction between clinical significance and statistical significance).

It is necessary to conduct studies that are modality-specific, since CT images may tolerate a greater or lesser amount of a specific type of compression than, say, PET or projection X-ray images. Even for a single modality like CT, the characteristics of slices through the chest are very different from those through the brain or the liver, and hence body-region specific trials are required. Further, for one modality and body region, the diagnostic task has a big impact. For example, no reasonable amount of irreversible compression will prevent a large brain tumor from being detected, yet the subtle boundary between grey and white matter might be lost preventing detection of an infarct. Even the acquisition device itself makes a difference, since there is significant variation in reconstructed image characteristics between detectors and algorithms from different manufacturers. Human readers also vary in their ability and may be affected by all of the foregoing factors. The combinatorial explosion of modality, region, task, device and human factors makes performing every experiment impossible.

In attempting to perform a reasonable subset of experiments and reach expert consensus, several approaches have been taken, including performing large scale, well-controlled task specific measurements of performance for narrowly defined tasks that are known or expected to be difficult. The results are then used to set conservative bounds on what is acceptable, recognizing that greater compression could perhaps be used for less demanding tasks, but sacrificing it.

Given the cost of conducting experiments that truly measure an observer's performance, short cuts are often taken, such as assuming that if the reader cannot perceive a difference between displayed images, then their actual performance would not have been affected. Such studies are an unsatisfactory substitute, given the lack of evidence supporting the (widely held) assumption. The removal of noise caused by compression actually causes some observers to express a preference for compressed images, but the effect of this factor on their actual detection or characterization performance is unknown. It is known that the converse is often true, that expert readers can successfully interpret what appear to be grossly degraded images that they would normally reject.

Another option is to compute some numerical metric derived from the pixel data to use as a surrogate for human performance. Simple metrics such as peak signal-to-noise ratio (pSNR) are known to correlate poorly with both subjective assessment of quality and observer performance, but more ambitious metrics based on models of the human visual system (HVS) have shown good results [5] [6]. At the very worse, they help the researcher explore the feasibility of and parameters for conducting human observer experiments.

Several national professional societies have attempted to review the literature, in some cases conduct their own experiments to fill gaps in knowledge [7], and define reasonably conservative limits for what they deem to be acceptable (Canada [8], UK [9], Germany [10]). Other countries (notably the ACR in the US) have shied away from such an exercise and leave it purely to the discretion of each physician. In some scenarios, irreversible compression is expressly forbidden by regulation, for example for digital mammography in the US [11].

6. Quantitatively Acceptable Irreversible Compression

Physicians rarely make measurements on images, though there are some very notable clinical exceptions, such as measuring narrowing in blood vessels. Even if visual interpretation of an image is unaffected by irreversible compression, quantitative performance may be. A classic example was the unexpected failure of an early ACC-DICOM experiment designed to show that JPEG irreversible compression of cardiac angiograms did not affect measurement at compression bit rates that were sufficient to achieve real time playback from single speed CD drives [12]. Improvements in CD drive technology made the use case irrelevant before the study was even completed, but the point was made. Compression may affect both the measurement of the size of a structure, or the values of the pixels contained within the structure. In the latter case, this may affect the classification of the tissue type (in the case of CT density in HU), or the quantitative assessment of function or activity (in the case of a modality like Nuclear Medicine or PET).

Also of concern is the impact of irreversible compression of the performance of machine algorithms designed for detection or classification, such as the CAD or CADx devices previously mentioned. Such algorithms may make use of different information present in the images than that used by the human visual system, and accordingly compression may be inappropriate, or different algorithms or parameters may be needed.

7. Standards and Interoperability

An overriding practical concern for the use of both reversible and irreversible compression is the matter of interoperability, both short and long term.

Whilst some consumers may tolerate the transient inconvenience of having to download a new codec or software version to view a picture or video on the Internet, or even forgo viewing an incompatible image at all, such unreliable performance is completely unacceptable in the practice of medicine.

Absolute reliability is expected, and both equipment manufacturers and institutional staff exercise strict control of the configuration of their systems and make extensive use of standards to assure interoperability. The Digital Imaging and Communications in Medicine (DICOM) standard has been used exclusively for this purpose since it was published in 1993. Amongst other things, it defines a limited set of compression schemes that may be used both on the network and on offline interchange media such as CDs and DVDs.

To the extent that consumer industry standards support the characteristics of medical images, these standards are adopted without modification. DICOM currently supports the use of conventional JPEG for color images, the more exotic JPEG processes for reversible and irreversible encoding of greater than 8 bit images, as well as JPEG-LS and JPEG 2000. For video images, additional schemes supported include various levels and profiles of MPEG2 and H.264. This is not to say that every storage or viewing device supports every scheme, but the problem is constrained by a combination of negotiation of capabilities on the network, strict limits on media content and effective use by purchasers of manufacturer-published conformance statements.

DICOM is extremely conservative about adding new schemes, since the incremental benefit of slightly improved performance, or an additional feature, rarely justifies the risk of compromising interoperability. Medical devices are also highly regulated (e.g., by the FDA in the US), so they are also expensive to modify given the burden of testing and documentation required. Irreversible compression is specifically highlighted as a risk factor in the FDA guidance for approvals [13].

The short-term distribution interoperability issues are readily addressed for both local and remote network viewing and the interchange of CDs, DVDs and USB devices, through the use of the DICOM standard. On the network, proprietary schemes can be negotiated using the DICOM transport protocol, and used if both ends support the scheme. Web-based proprietary thin-client or Web-deployable thick-client solutions are relatively immune to compression-related interoperability problems, as long as the demands for a specific browser version are not too restrictive.

However, in the long term, interoperability is significantly impaired if the images are not archived in a standard compressed format. For serving up images on demand in the short term, proprietary schemes arguably may have performance advantages, but the long term archive must contain images in a standard format,

or at the very worst provide an efficient bulk means to convert them to a standard format. Most institutions can expect to completely replace their Picture Archiving and Communications (PACS) storage and distribution infrastructure every three to five years, very often switching to a different vendor. In such cases, the use of proprietary rather than DICOM standard reversible or irreversible compression schemes to archive the images greatly increases the cost and time required to migrate them.

8. Fundamental Compression Research

The foregoing may seem relentlessly depressing to the researcher interested in developing novel compression schemes or incremental improvements to existing approaches, given the conservatism that the user, manufacturer and standards community manifests.

However, there is still great opportunity to take advantage of increased computing and memory capacity to implement hitherto impractical methods. For example, many of the larger medical image datasets are three or four, not two dimensional, and use of redundancy in these additional dimensions is rarely exploited. Specific features such as the progressive embedded nature of a JPEG 2000 bit stream (achieving progressive transmission without sending more data overall) are also superficially attractive, as is the multi-resolution encoding that offers complete lower resolution images earlier.

Yet these benefits or features are not necessarily easily deployed. Indeed the increased complexity of JPEG 2000 may have significantly slowed its adoption, particularly when many of the practical use cases remain relatively unsophisticated. Neither JPEG-LS, nor JPIP, though included in DICOM, has seen much practical use. For an advanced application like multi-resolution Whole Slide Imaging (WSI) for pathology, which involves interactive remote navigation of enormous datasets (virtual microscopy), a combination of JPEG 2000 and JPIP would seem ideal Yet equipment manufacturers have largely eschewed these standards, ostensibly on the basis of complexity and poor performance, in favor of much simpler tiled JPEG pyramidal decompositions (now supported in both BigTIFF and DICOM).

It is a given that any new scheme will require adoption by a compression standards body like ISO/IEC JTC1/SC29/WG1 (JPEG) before the medical community or DICOM would adopt it. Individual vendors might adopt or develop their own schemes, but the opportunity is limited. Another significant barrier to adoption is the matter of intellectual property restrictions, with which image compression schemes are fraught. The last thing medical software and hardware manufacturers are interested in is paying license fees for compression schemes of dubious benefit. In the best case, they may seek to enhance their own patent portfolio for defensive reasons, but generally rely for that on in-house developed schemes. Few compression schemes succeed without at least one commercially usable open source reference implementation to spur early deployment and testing, and any intellectual property restrictions have a dramatic stifling effect. The extent to which the JPEG and MPEG subcommittees struggle with these issues is no secret.

On a positive note, many image compression researchers have in the past had very limited access to large quantities of realistic medical images with which to experiment. There has been dramatic improvement, with the advent of large Internet-based publicly accessible open image archives. These mostly contain de-identified images collected during the conduct of large-scale clinical trials that are made available for secondary re-use, either through the generosity of the investigators or as a matter of funding agency policy. The National Biomedical Imaging Archive (NBIA) [14], the Osteoarthritis Initiative (OAI) [15] and the Alzheimer's Disease Neuroimaging Initiative (ADNI) [16] are just a few examples. Although there is currently no formally defined collection or “corpus” of specific images by which different studies can be compared, it would be quite reasonable to establish now given the availability of these readily accessible archives. Most such archives contain images in original DICOM format (but de-identified), and hence are compatible with conventional DICOM software toolkits and viewing software. Experimental compression codecs can be readily connected to most such toolkits, for example using standard APIs such as the Java Imaging IO interface. Proprietary or home-grown experimental formats should be studiously avoided, to allow for re-use.

Though perhaps less exciting than fundamental compression research, the need to perform

observer-performance experiments is essentially unsatisfied and unbounded. However, justifying and raising funding to perform sufficiently rigorous studies as to be useful remains extremely challenging.

References

[1] Dinan MA et al. Changes in the Use and Costs of Diagnostic Imaging Among Medicare Beneficiaries With Cancer, 1999-2006. JAMA.2010;303(16):1625-1631. "doi: 10.1001/jama.2010.460" "http://jama.ama-assn.org/content/303/16/1625"

[2] Bull, Housser & Tupper. Lossy Compression Study: Legal Assessment. Report prepared for Fraser Health Authority. December 2004.

[3] American Medical Association. Resolution 539 (A-06). Development of Standards for MRI Equipment and Interpretation to Improve Patient Safety. "http://www.amaassn.org/ama1/pub/upload/mm/471/539a06.doc"

[4] European Society of Radiology. Usability of irreversible image compression in radiological imaging. Position Paper. Insights Imaging. 14 Feb 2011. "DOI 10.1007/s13244-011-0071-x" "http://www.springerlink.com/content/f13u18tv32566005/"

[5] Johnson JP. Nafziger JS. Visual Discrimination Model for Single Image Applications. US Patent Application 12/039,086. Feb 28, 2008. ""

[6] Siddiqui K et al. Correlation of radiologists' image quality perception with quantitative assessment parameters: just-noticeable difference vs. peak signal-to-noise ratios. In SPIE Medical Imaging 2005: PACS and Imaging Informatics. Proc. SPIE 5748, 58 (2005). "doi:10.1117/12.596140" "http://spiedigitallibrary.org/proceedings/resource/2/psisdg/5748/1/58_1"

[7] Koff D et al. Pan-Canadian Evaluation of Irreversible Compression Ratios ("Lossy" Compression) for Development of National Guidelines. Journal of Digital Imaging. 2008. 22(6): 569-578, "doi:10.1007/s10278-008-9139-7" "http://springerlink.com/content/u5r85r66341318w1"

[8] Canadian Association of Radiologists. Standards for Irreversible Compression in Digital Diagnostic Imaging within Radiology. "http://www.car.ca/uploads/standards%20guidelines/Standard_Lossy_Compression_EN.pdf"

[9] Royal College of Radiologists. The adoption of lossy data compression for the purpose of clinical interpretation. 2008. "https://www.rcr.ac.uk/docs/radiology/pdf/IT_guidance_LossyApr08.pdf2".

[10] Loose R et al. Compression of digital images in radiology - results of a consensus conference. Fortschr Röntgenstr 2009; 181(1): 32-37. "DOI: 10.1055/s-2008-1027847" "http://www.thieme-connect.com/DOI/DOI?10.1055/s-2008-1027847"

[11] US Food and Drug Administration. The Mammography Quality Standards Act Final Regulations: Modifications and Additions to Policy Guidance Help System #9. April 19, 2006. "http://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/DocumentArchives/ucm114207.htm"

[12] Nissen SE. Hirshfeld JW. Simon R. Introduction and background: The International Angiographic Compression Study. J Am Coll Cardiol, 2000; 35:1367-1369. "http://content.onlinejacc.org/cgi/content/full/35/5/1367"

[13] US Food and Drug Administration. Guidance for the Submission Of Premarket Notifications for Medical Image Management Devices. July 27, 2000. "http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm073720.htm"

[14] National Biomedical Imaging Archive (NBIA). "http://imaging.nci.nih.gov/ncia/".

[15] Osteoarthritis Initiative (OAI). "http://oai.epi-ucsf.org/"

[16] Alzheimer's Disease Neuroimaging Initiative (ADNI). http://adni.loni.ucla.edu/

David Clunie is a radiologist and the Chief Technology Officer at CoreLab Partners, an imaging contract research organization supporting the cancer clinical trials industry, as well as editor of the DICOM Standard and co-chairman of the Integrating the Healthcare

Enterprise Radiology Technical Committee.

He is a Fellow of the Royal Australian and New Zealand College of Radiologists, and a member of the Society for Imaging Informatics in Medicine, International Society for Optical Engineering, Radiological Society of North America, American College of Radiology, American Association of Physicists in Medicine and several other professsional societies. In 2010, he was an invited participant in the European Society of Radiology's (ESR) Expert Panel on Image Compression, which resulted in the ESR's position paper on the subject. He participated in the Sun Java Image I/O API Experts Group (JSR-15) defining an API for compressed image formats. He is the author of a textbook on DICOM Structured Reporting, published by PixelMed in 2001, as well as several open source software projects for implementing DICOM, and an experimental implementation of JPEG-LS. He is a recipient of the National Electrical Manufacturers Association (NEMA) Kite and Key Award for advancement of the interests of the electrical industry.

SPECIAL ISSUE ON MEDIA CODING IN SOCIAL/P2P NETWORK PARADIGM

Naeem Ramzan
Queen Mary University of London, UK
naeem.ramzan@eecs.qmul.ac.uk

The emerging Social/P2P networking techniques have introduced enormous impact on nowadays network interactions. At the same time, promising methods have been employed to effectively integrate context and content for multimedia mining, management, indexing, retrieval and video coding in Social/P2P networks. In addition, increasing service requirements by multimedia consumers inspire rapid developments on interactive interface design and adaptive visualisation. With these developments, extensive research has been carried out on multimedia content processing and sharing in Social/P2P networks.

Advanced multimedia processing often attempts to optimize the coding efficiency and introduces higher dependency between the processed data, resulting in a fundamental challenge to robust data communication. On the other hand, Social/P2P systems due to limited capacity and unreliability of peers (in P2P), mechanisms are needed to efficiently manage the resources contributed by peers and to adapt to the dynamic nature of the network. Therefore, we are facing unprecedented challenges in multimedia processing and communications in Social/P2P networks. The aim of this special issue is to present the most recent multimedia processing technological and scientific achievements in this scope, in order to improve the end-to-end Quality-of-experience (QoE) in P2P/Social networks.

In the article titled "*Building Incentives in Peer-to-Peer Networks using Social Reciprocation*", discusses incentive protocols based on social norms to persuade cooperation and tone down free-riding to maintain the performance of P2P multimedia sharing applications. Social norms discussed in this paper consist of social strategy "reputation-based behavioral strategy" and peer's reputation scheme. The article concludes that incentive protocol based on social norm outperforms traditional protocols such as Tit-for-tat.

Another important paradigm in networks is considered in the article titled "*Fast Content-Aware Delivery in Overlay Networks*". In this article, the authors describe the approach that they are following in the EU project COAST in order to face the increasing need for a content centric internet. They aim to build a Future Content Centric Network (FCN) overlay architecture able to find the desired data in the closest networking cache and forward it to the users in an efficient, timely and network-friendly way.

The next two articles present the two most advanced coding techniques for social/P2P networks. The article titled "Multiple Description Coding Based Video Streaming in Peer-to-Peer Networks" presents how Multiple Description Coding (MDC) can be applied to achieve robust and adaptive video streaming over P2P networks. This paper briefly revisits the advantages and challenges for video streaming in P2P networks with respect to MDC. They describe a packet scheduling framework for MDC video streaming. The next article "*Adaptable Media Coding Techniques for Social/P2P Networks*" presents an overview of Scalable Video Coding (SVC) with the perspective of content distribution over Social/P2P networks. These coding schemes provide natural robustness and scalability to media streaming over heterogeneous networks. The amalgamation of SVC and P2P are likely to accomplish some of the Future Media Internet challenges. A new piece picking policy and neighbour selection policy is also described in this article to achieve high QoE.

The last article of this issue is "*My Own, Personal Video Broadcast*". This paper explains the way in which personalized media distribution can be achieved in a cost effective way by exploiting the state-of-the-art technologies. The Two EU projects "My e-Director 2012" and "SARACEN" are explained as case study to get personalized media distribution in networking environment.

I believe that this special issue, although ensuing

in presenting a preface to some important research areas in media distribution in social/P2P networks however it is a tip of an iceberg that presents a broad array of demanding and appealing problems.

Finally, I would like to thank all the authors for their original contributions and anticipate these articles can encourage further research on the area and promote emerging technologies in this field.

Naeem Ramzan, M.S in Telecom from Brest France and PhD in Electronics Engineering form Queen Mary University of London in 2004, and 2007 respectively. From 2004-2007 he worked on EU project aceMedia in the Multimedia & Vision Group in Queen Mary University of London. Currently, he is a senior researcher and core member of technical coordination team in the EU funded project PetaMedia, and SARACEN. His research interest includes image and video coding, scalable video coding, multimedia search and retrieval, multimedia transmission over wireless and P2P networks. He has been a co-chair of the IEEE Student Branch Congress, UK, 2008. He has been a co-chair of Special session in International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), UK, 2009; ACM Multimedia Information Retrieval Conference, USA, 2010; and WIAMIS, Italy, 2010. He is also the co-organizer and co-chair of ACM Multimedia workshop SAPMIA, Italy, 2010 and ACM workshop on Social and Behavioural Networked Media Access (SBNMA), Scottsdale, Arizona, USA, 2011. He has served as Guest Editor of a special issue of the Elsevier Journal Signal Processing: Image Communication. He is the author or co-author of more than 50 research publications. He also serves as a reviewer of several international journal and conferences. He is the chair of the Summer School on Social Media Retrieval (S3MR), June 2011.

Building Incentives in Peer-to-Peer Networks using Social Reciprocation

Yu Zhang[1] and Mihaela van der Schaar[2]
UCLA, USA
[1]yuzhang@ucla.edu, [2]mihaela@ee.ucla.edu

With the explosion of communication technologies and multimedia signal processing, the sharing of multimedia content is becoming increasingly popular over the Internet. In particular, Peer-to-Peer (P2P) multimedia applications represent a large majority of the traffic currently exchanged over the Internet. By pooling together the resources of many autonomous devices, P2P networks are able to provide a scalable and low-cost platform for disseminating large files without relying on a centralized infrastructure **Error! Reference source not found.**0. Multimedia sharing systems that have been successfully developed for P2P networks are usually based on data-driven approaches 0,with different types of files beingdivided into chunks and then disseminated over the P2P network. Each peer possesses several chunks, which are shared among interested peers, and information about the availability of the chunks is periodically exchanged among peers through intermediate trackers. Using this information, peers continuously associate themselves with other peers and exchange chunks.

While this approach has been successfully deployed in various applications over P2P networks, it is vulnerable to intrinsic incentive problems since the upload service incurs costs to both the uploader and the downloader, but benefits only the downloader 0. As contributing their contentdoes not generate direct benefit, peers tend to avoid uploading while trying to download content from other peers, a behavior commonly known as *free-riding*.

Such studies demonstrate that designing incentive protocols to encourage cooperation and mitigate free-riding is crucial to maintain the performance of P2P multimedia sharing applications. To achieve this goal, a large body of research was dedicated to this area 00. Many of these existing mechanisms relyon game-theoretical approaches and can be classified into three categories: pricing, reciprocity and intervention000, as shown in Figure 1. As intervention requires centralized control of the system, it is not widely adopted in P2P networks and thus, we focus our discussion on pricing and reciprocity.

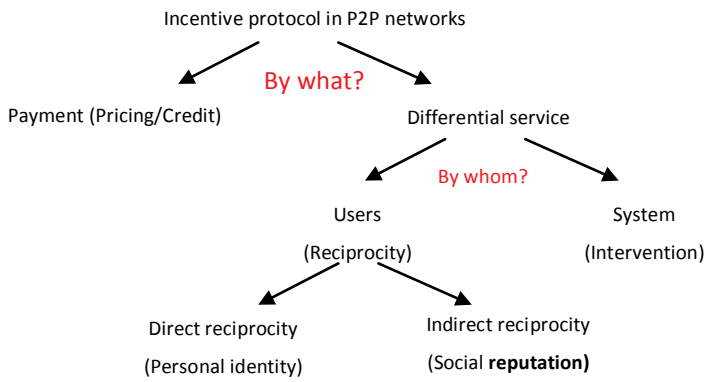


Figure 1. Overview of incentives for P2P protocols

Pricing mechanisms rely on implementing a currency-based system that is resistant to forgery and double-spending0. Peers are incentivized to share their content by rewarding them with virtual currency for uploading and charging them for downloading. However, such solutions are often very cumbersome to deploy because they require an accounting infrastructure to track the transactions of peers, which further necessitates the usage of public keys, a web of trust, or threshold cryptography techniques 0. Furthermore, these systems often deploy auctions to set the price, which may result in high delay and complexity in order to implement a desirable allocation.

Another method for providing incentives is based on reciprocity, where the peers' past reciprocative behavior (e.g. contributing content to other peers or not) is rewarded or punished in future interactions with the same or other peers. Differential service schemes are deployed in reciprocity-based protocols to determine how peers should make their upload decisions 0. Depending on how a peer's rating is generated, reciprocity-based protocols can be classified as direct reciprocity (also known as personal reciprocation) and indirect reciprocity (also referred to as societal reciprocation).

In direct reciprocity, each peer rates a specific peer

individually 0,which requires frequent interactions between two peers in order to establish accurate mutual ratings.This is restrictivein P2P networks characterized by high churn or asymmetry of interests, where a peer normally interacts with a stranger (i.e. with whom it was randomly matched) about whom it has no prior history and with whom it has no expectation to meet again in the future. For example, the investigation in 0shows that over 70% of P2P traffic is exchanged in networks with more than 1000 peers,which implies that a peer normally interacts with a stranger (i.e. with whom it was randomly matched) about whom it has no prior history and with whom it has no expectation to meet again in the future. Hence, protocols based on direct reciprocity such as Tit-for-Tat perform well only in networks dominated by long-lived relationships, where peers have ample opportunities to mutually reciprocate, and where peers are interested in similar content.

Due to the random matching feature of large P2P networks, indirect reciprocity becomes a more appropriate mechanism in designing incentive protocols. Most protocols based on indirect reciprocity use reputation mechanisms 000. A peer is globally rated with a reputation calculated by its past behavior in the network. In order to make a decision, a peer does not need to know the entire action history but the reputation of its opponent. However, the majority of existing works on P2P reputation mechanisms are concerned with system design issues and focus on effective information gathering techniques which only differ in how the global reputation is calculated and propagated, e.g. efficient information aggregation 0, secure peer identification 0, etc. An analytical framework that is able to rigorously study how peers can be incentivized to cooperate in P2P networks and what is the resulting impact on the network performance when various reputation mechanisms are deployed, is still missing.

In this respect, we introducean analytical framework based on social norms that is able to rigorously study how peers can be incentivized to cooperate in P2P multimedia sharing services with a large population of anonymous peers **Error! Reference source not found.**.In an incentive scheme based on a social norm, each individual is attached a label indicating its reputation, which contains information about its past behavior, and individuals with different reputations are treated differently by other individuals they interact with. Hence, a social norm can be easily adopted in social communities with an infrastructure that collects, processes, and delivers information about individuals' behaviour00. Meanwhile, we also consider the following unique features and constraints of P2P multimedia sharing services which will be described in detail later.

- *Asymmetry of interests among peers*.
- *Serviceerrors*.
- *Altruistic peers and malicious peers*.

We consider multimedia sharing applications in which peers would like to associate themselves with other peers that possess media content in which they are interested. The shared media content is coded and divided into media chunks by the content creator. Each peer would like to maximize its long-term benefit from downloading content from other peers while minimizing its consumption on service costs by uploading its own content to others. In multimedia sharing applications with large populations, peers are interested in very diverse content. The traditional analysis on P2P incentive protocols does not consider such diversity in peers' interests by modelling the interests of a pair of matched peers as mutual and homogeneous. Hence, traditional P2P protocols fail easily in providing incentive to a peer to uploadcontent to a requesting peer if itpossesses no content which is interested by this peer. Different from this,we explicitly accommodate in our framework the fact that the peers' interests are asymmetric, andmodel the interaction between a pair of matched peers as a gift-giving game.An example of a gift-giving game is shown in Table 1. When the peer being requested choose to provide the uploading service to the requesting peer, it consumes a service cost of c while the requesting peer receives a benefit of r ; when the peer being requested refuse to provide the uploading service, both peers receive a utility of 0.

	Requested peer	
	Serve	*Not Serve*
Requesting peer	r , $-c$	0, 0

Table 1. The utility matrix of a gift-giving game

Formally, a social norm consists of a social strategy and a reputation scheme. The social strategy is a *reputation-based behavioral strategy*, which regulates the service behaviour of a peer to other peers requesting its content. The reputation

scheme specifies how a peer's reputation will be updated depending on its past behavior. A peer's reputation will be increased if it complied with the social strategy in the past, and will be decreased if it deviated from the social strategy. Different from traditional P2P incentive protocols such as Tit-for-Tat,the social norm framework provides different levels of rewards and punishments to peers with different contributions to the network.This enables the protocol designer to design more sophisticated incentive protocols which can significantly improve peers' incentives to voluntarily contributing their content and hence the sharing efficiency of the entire network**Error! Reference source not found.**. This is especially important in multimedia applications since the shared data is usually heterogeneous (e.g. different chunks have different video distortion impactwhen using a scalable video coder and different delay constraints).

The existing P2P incentive protocol alsorarely considers thatnetwork errors may affect the interactions between peers. This is an idealized assumption which is hard to realize in practical networks. The prevailing protocols such as Tit-for-Tat fails when errors take place during the transmissions between peers and the sharing efficiency of the network will be severely degraded. In contrast, we explicitly takes into consideration that the exchange of chunks between peersmay be subject to service errors and designs protocols which enables peers to quickly re-coordinated to the cooperation phase, thereby increasing the resilience of such sharing applications00.

We also investigate the impact of the presence ofhelpers, which are altruistic peers whichalways provide upload services to other peers, such as seeds, as well as malicious peers who upload corrupted data to others. By investigating the impact of such non-reciprocative peers and adjusting the design of the social strategy as well as the reputation scheme, the resulting protocol could retain peers' incentives for cooperation and the sharing efficiency from being degraded.
When applied to P2P multimedia sharing applications, our experiments showthat the social norm based incentive protocols exhibit significantly higher PSNR than traditional protocols such as Tit-for-Tat **Error! Reference source not found.**. In Figure 2, we explicitly compare the average PSNR of the decoded video among all peers using different protocols. The exchanged video content is the well-known "Foreman" sequenceencoded using H.264/AVC codec and divided into chunks of 0.1s.The following four protocols are considered:

- Optimal cooperation: all peers cooperate unconditionally without considering the incentive constraints. Since all peers provide full services, the performance it delivers remains to be constant and serves as the Pareto boundary of the performance that an incentive protocol can possibly achieve.
- Optimal social norm equilibrium: the social strategy and the reputation scheme are optimized to maximize the social welfare while all peers have the incentive to follow the resulting social norm.
- A fixed social norm with the social strategy being threshold-based **Error! Reference source not found.**.
- Tit-for-Tat

When the service cost c becomes large compared to the service benefit r, peers lose their incentives to follow Tit-for-Tat and do not mutually provide upload services at all. However, the social norm based protocols can still provide sufficient incentives for peers to mutually cooperate, whichleads to significant improvements in terms of PSNR.

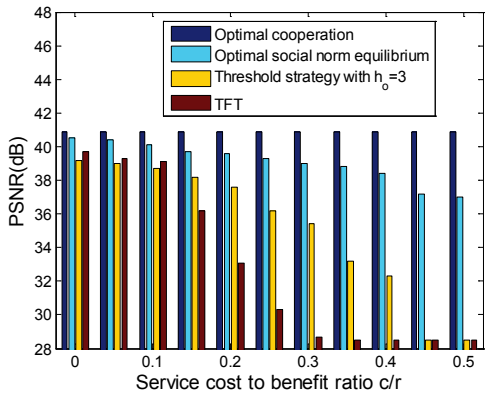


Figure 2. The PSNR of different protocols

References

Y. Zhang, and M. van der Schaar, "Peer-to-Peer Multimedia Sharing based on Social Norms,"*UCLA Technical Report*, http://arxiv.org/abs/1102.1503.

M. Ripeanu, I. Foster, and A. Iamnitchi, "Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design,"*IEEE Internet Computing Journal, Special Issue on Peer-to-Peer Networking*, 2002.

Y. Zhang, J. Park, and M. van der Schaar, "Social Norms for Networked Communities,"*UCLA Technical Report*, http://arxiv.org/abs/1101.0272

X. Zhang, J. Liu, B. Li, and T. Yum, "CoolStreaming/DONet: A Data-driven Overlay Network for Efficient Live Media Streaming,"*PROC. INFOCOM*, pp. 2102 – 2111, 2005.

B. Cohen, "IncentivesBuilding Robustness in BitTorrent,"*Proc. P2P Econ. Workshop*, Berkeley, CA, 2003.

J. Crowcroft, R. Gibbens, F. Kelly, and S. Ostring, "Modeling Incentives for Collaboration in Mobile Ad Hoc Networks,"*Proc. Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, 2003.

P. Golle, K. Leyton-Brown, I. Mironov, and M. Lillibridge, "Incentives for Sharing in Peer-to-Peer Networks,"*Lecture Notes in Computer Science*, vol. 2232/2001, pp. 75-87, Springer, Jan. 2001.

R. Landa, D. Griffin, R. Clegg, E. Mykoniati, and M. Rio, "A Sybilproof Indirect Reciprocity Mechanism for Peer-to-Peer Networks", *PROC. INFOCOM*, pp. 343 – 351, 2009.

Y. Zhang, J. Park, and M. van der Schaar, "Social Norm and Long-run Learning in Peer-to-Peer Networks,"*2011 Int'l Conf. on Acoustics, Speech and Signal Process. (ICASSP 2011)*, 2011.

K. Ranganathan, M. Ripeanu, A. Sarin, and I. Foster, "Incentive Mechanism for Large Collaborative Resource Sharing,"*Proc. IEEE Int'l Symposium on Cluster Computing and the Grid*, pp. 1-8, 2004.

J. Park, and M. van der Schaar, "A Game Theoretic Analysis of Incentives in Content Production and Sharing over Peer-to-Peer Networks,"*IEEE J. Sel. Topics on Signal Process.*, vol. 4, no. 4, pp. 704 – 717, 2010.

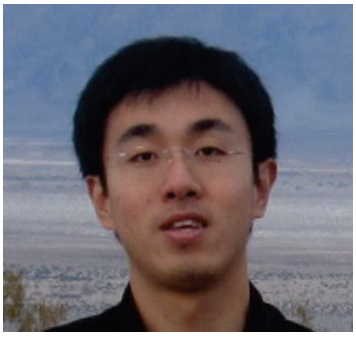
M. Feldman, K. Lai, I. Stoica, and J. Chuang, "Robust Incentive Techniques for Peer-to-Peer Networks,"*Proc. of the 5th ACM Conf. on Elec. Commerce*, Session 4, pp. 102-111, 2004.

J. Park, and M. van der Schaar, "Incentive Provision Using Intervention,"*IEEE INFOCOM* 2011.

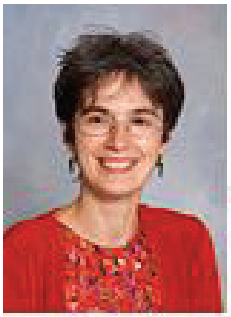
Y. Zhang, J. Park, and M. van der Schaar, "Designing Social Norm Based Incentive Schemes to Sustain Cooperation in Large Communities,"*2nd Int'l ICST Conf. on Game Theory for Networks (Gamenets)*, 2011.

S. Kamvar, M. T. Schlosser, and H. G. Molina, "The Eigentrust Algorithm for Reputation Management in P2P Networks,"*Proc. 12th Int'l Conf. on World Wide Web*, pp. 640 – 651, 2003.

A. Ravoaja, and E. Anceaume, "STORM: A Secure Overlay for P2P Reputation Management,"*Proc. 1st Int'l Conf. on Self-Adaptive and Self-Organizing Systems*, pp. 247 – 256, 2007.



Yu Zhang received the B.S, M.S degrees in Electrical Engineering from Tsinghua University, China, in 2002 and 2006, respectively. He is now pursuing his Ph.D. degree in Electrical Engineering Department, University of California, Los Angeles. His research interests include cross-layer optimization, online learning, and game theoretic network analysis.



Mihaela van der Schaar is Professor in the Electrical EngineeringDepartment at the University ofCalifornia, Los Angeles. Her researchinterests include multimedia networking, communication, processing,and systems, multimedia stream mining,dynamic multi-user networks and system designs, online learning, networkeconomics and game theory. She is an IEEE Fellow, a DistinguishedLecturer of the Communications Society for 2011-2012, the Editor in Chief of IEEETransactions on Multimedia and a member of the Editorial Board of theIEEE Journal on Selected Topics in Signal Processing. She received anNSF CAREER Award (2004), the Best Paper Award

from IEEE Transactionson Circuits and Systems for Video Technology (2005), the OkawaFoundation Award (2006), the IBM Faculty Award (2005, 2007, 2008),the Most Cited Paper Award from EURASIP: Image Communications Journal(2006), theGamenets Conference Best Paper Award (2011), and the 2011IEEE Circuits and SystemsSociety Darlington Award Best Paper Award.She received three ISO awards for her contributions to the MPEG videocompression andstreaming international standardization activities, and holds 33granted USpatents. For more information about her research visit:http://medianetlab.ee.ucla.edu/

Fast Content-Aware Delivery in Overlay Networks

Theodore Zahariadis[1] and Emanuele Quacchio[2]
[1]Synelixis Solutions, Greece, [2]STMicroelectronics, Italy
[1]zahariad@synelixis.com, [2]emanuele.quacchio@st.com

1. Motivation and goals

The explosion of content published on the web poses two types of problems. The first is at the search level: how can a user find the content that s/he is looking for? The second is at the networking level: how can this content be accessed/ streamed quickly?

On searching, the huge number of content requires rethinking the architecture of the search engines. Today, for performance reasons, search engines are centralized and new data centers are full replicas of existing ones. This is obviously not scalable, especially as the overall number of items to be crawled and indexed grows. Distributed architectures for search engines are studied in [1] and further considered in [2].

On networking, it must be noted that the Internet architecture is designed considering an endpoint-based communication model. Every packet should have the addresses of two endpoints (source and destination) to support host-to-host applications. However, recent traffic measurements reveal that more and more applications are oblivious to the addresses of the servers who deliver the contents. This trend has motivated content-oriented networking studies (e.g. DONA [3], CCNx[4])

The proposed content-aware delivery network (CADN) can be envisioned as an overlay over one or multiple underlying access or core networks. One of our major requirements is to implement network architecture and the relevant functional blocks that will be able to take advantage of the Content Centric Networks (CCN) developments, but also keep backwards compatibility and key issues that made IP the greatest networking success for the last 25 years.

2. CADN Architecture

The proposed CADN architecture consists of three logical/abstract overlays (Figure 1). The main components/modules at each overlay are:

At the ***Information Overlay***:

- **Search Engine**: It discovers and indexes the content and the services, processes the queries from the users and returns relevant results ordered according to several criteria.
- **Publishing Front-end**: Besides automatic crawling, this module enables manual publication of content. This functionality may be co-located at the Content overlay Entry Point (CEP).
- **Cache Locator:** It is contacted in order to retrieve an object and redirects the request to the "best" [1] cache node containing a cached replica of the object. In order to make that decision, it may (periodically) communicate with the network monitor entity. It may be co-located at CEPs.
- **Cache Optimizer (CO)**: It supports caches in deciding which object they should store or evict. Coordinating optimizers may offer a distributed cache replication schema. It may be a distributed functionality, co-located at CEPs.

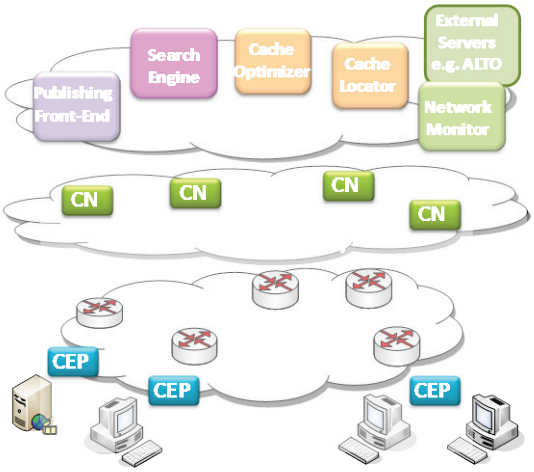


Figure 1: CADN functional architecture

- **Network/Traffic Monitor**: It is responsible for gathering all network related information: topology, traffic, characteristics of the user Internet access and optionally user location. It

[1] "Best" is defined based on the perceived Quality of Service (PQoS) of the user.

may be a variation of an IETF ALTO [5] server or communicate with/supported by external traffic and network optimizer servers. It may also supported by Deep Packet Inspection (DPI).

At the **Distributed Content/Services Aware Overlay,** we define **Content Aware Nodes (CN)**, which may include subsets of the following functionality:

- **DPI**: It contributes to the discovery of content and services. Moreover, it generates information on content popularity and reports that to the Search Engine.
- **Distributed Caching**: It is responsible for content caching and caching replacement in collaboration with the CO and access control.
- **Adaptation & Enrichment Engine**: It is responsible for content adaptation, enrichment and efficient delivery.
- **Hybrid Router**: It is responsible for the content centric delivery, IP acceleration and efficient content streaming (including P2P overlays creation).

Finally, the entry points to the proposed CADN overlay network are the ***Content Overlay Entry Points (CEP***). A CEP may be physically hosted at a local router or a Residential Gateway. CEPs are also responsible for seamless operation and optimal content fetching and streaming; information overlay functionality may be co-located/ hosted by the CEPs. Collaborating CEPs (e.g. linked via DHTs) may offer a distributed Tier 0 caching functionality, while collaborating CNs a Tier 1 caching overlay.

3. Content ULR (CURL)

In the above CADN, we assign a unique ID (UID) to each content object. Based on that UID, fast retrieval may be achieved by a direct index to the content (or a content object replica cached in the CADN caching overlays). However, the proposed schema should:

a) Avoid extended data replications at the network caches and minimize the Content Optimizer load.

b) Detect and retrieve very fast the UID at CEP level. This should be fast enough to allow even seamless real-time video streaming

c) Be backwards compatible with todays' URLs.

In order to meet the (a), UID should be always associated with the content object itself (e.g. encapsulated in the object) or be based on unique characteristics of the content object (e.g. a set of low level descriptors). However, (b), poses that this could be calculated once (or sometimes), but should not be calculated or generated each time the object is requested; instead it should be "carried" and "extracted" in most cases. On the other hand, due to backwards compatibility needs (c), we should not change the standard file format (e.g. we could not encapsulate UID or low level descriptor in the content object). One solution could be to create a wrapper that would encapsulate the UID whenever the content object enters the CADN and extracts that at the time that the content object leaves the CADN, but this would increase the complexity and processing time.

Instead, we propose to use as UID a formal file name format. UID will be a string concatenation in the format:

CM-CID-filename.ext

where:

- ***CM*** is a "Content Marker" e.g. COAST-248036[2], which guarantees an easy and fast detection that this name is a UID
- ***CID*** is a content signature, which could be a self-certifying identifier e.g. MD5 or SHA-1 based-hash function on the file's content or even a combination of searchable low level descriptors.
- the original filename, which is used for making UID easily recognized by humans and avoid complex self-certifying names [6].

Whenever a user publishes new content, the Publishing front end will convert the object filename to a UID and the URL to a Content URL (CURL) with the following format:

http://www. website.com/…/CM-CID-filename.ext

The CURL is a URI, which is designed to enable caching mechanisms and trigger CADN-related functionality (e.g. accessing the CADN overlay network and querying for locally or "nearby" cached content). It should also be emphasized

[2] The idea is based on project COAST. It may be replace by any "magic word".

that from the CURL, the original URL and the UID may be easily extracted, while on the other hand it is fully backwards compatible with existing browsers and no modifications are needed. In this way, we may seamlessly support any kind of data available in the Internet (text, images, various types of audio or video), while in case the content object is not cached in the CADN, we can always go back to the original source URL.

The CID is based on physical characteristics of the content object. Thus, it may guarantee that content caching replications in the CADN will be minimal (even if the names are different). Moreover, it may always be verified that the requested content objects is the one that was actually retrieved. Last but not least, it offers a simple content protection mechanism, even if the content is renamed.

One could argue that CURLs are still not very user friendly due to the CID contained in the UID. Yet, this can be easily overcome by using *both* legacy URLs when the users refer to a Web page (e.g. http://www.coast.eu/my_picture.jpg) and CURLs for a given web page or resource. For direct retrieval of content, the search engine results may be displayed as normal URL, while hiding the relative CURL as hyperlink.

4. CURLs for fast video streaming

Besides fast retrieval of individual content objects (e.g. a single picture) or sets of content objects with loose synchronization constrains (e.g. pictures on a web page), the efficiency of the CURLs may be especially illustrated considering the case of video streaming. Recent technologies for real time video delivery leverage on the possibility to split the content in multiple segments that can be downloaded independently and played in sequence order by a client application; plain HTTP protocol is used to access and retrieve data segments. In case the video source is available at different bitrates, or encoded in a layered format (using advanced video coding standard such as Scalable Video Coding-SVC or MultiView Video Coding-MVC), the client application can dynamically decide to switch the bit-rate in order to guarantee continuity of service in case of network congestion. Such technologies, generally known as *Adaptive HTTP Streaming*, have been widely considered in both commercial solutions and standardization bodies.Within CADN architecture data segments belonging to different views/ resolutions may be cached and distributed separately through different nodes and caches, accelerating video delivery and improving the perceived quality for the user.

The composition of the content in term of data segments as well as the length and the location of each segment are usually described in a manifest file transmitted to users beforehand, and updated during the streaming session. In particular, 3rd Generation Partnership Project (3GPP)-Adaptive HTTP streaming (AHS) solution [7], and emerging MPEG-Dynamic Adaptive Streaming over HTTP (DASH) [8] standard proposal, define an xml basedMedia Presentation Description (MPD) file, in which each segment is identified by a different URL. A streaming client application compliant with 3GPP/DASH specification can interpret MPD file, download and play the content. When streaming services are offered through CADN architecture, URLs in MPD are substituted by CURLs, as depicted in figure 2, where "COAST248036" is used as CM. In this way segments are efficiently retrieved through Content/Service Distributed Overlay, leveraging of caching optimizations provided by content aware nodes. In case of a standard web browser or if a CEP does not exist, the segments will be retrieved directly from the original web server using MPD version containing normal URLs

```
<?xml version="1.0" encoding="UTF-8"?>
<MPD type="OnDemand" minBufferTime="0.96S">
        <Title>MPD Example</Title>
        <Source>COAST</Source>
        <Copyright>COAST, All rights reserved</Copyright>
    </ProgramInformation>

    <Period segmentAlignmentFlag="True">
        <Representation bandwidth="642348" width="640" height="480" mimeType="video/mp4" >

          <SegmentInfo duration="0.96S">
            <!--InitialisationSegmentURL sourceURL="MPD_Example.3gp"/-->
            <Url sourceURL="http://coast.media.net/COAST248036_hashvalue_video_seg_000.mp4" />
            <Url sourceURL="http://coast.media.net/COAST248036_hashvalue_video_seg_001.mp4" />
            <Url sourceURL="http://coast.media.net/COAST248036_hashvalue_video_seg_002.mp4" />
            <Url sourceURL="http://coast.media.net/COAST248036_hashvalue_video_seg_003.mp4" />
            <Url sourceURL="http://coast.media.net/COAST248036_hashvalue_video_seg_004.mp4" />
            ...
          </SegmentInfo>

        </Representation>
     </Period>

</MPD>
```

Figure 2: MPD file embedding CURLS

5. Acknowledgement

The work presented in this paper is based on the project COAST (COntent Aware Searching, retrieval and sTreaming), which is partially funded by the European Commission under contract ICT-248036.

References

[1] Theodore Zahariadis, Flavio Junqueira, Luca

Celetto, Emanuele Quacchio, Saverio Niccolini, Pierre Plaza, "Content aware searching, caching and streaming", 2nd International Conference on Telecommunications and Multimedia, (TEMU), Chania, Greece, 14-16 July 2010, pp. 263-270

[2] Barla Cambazoglu, Emre Varol, Enver Kayaaslan, Cevdet Aykanat, Ricardo Baeza-Yates, "Query Forwarding in Geographically Distributed Search Engines", 33rd Annual ACM SIGIR Conference, July 2010.

[3] Teemu Koponen, Mohit Chawla, Byung-Gon Chun, Andrey Ermolinskiy, Kye Hyun Kim, Scott Shenker, Ion Stoica, "A Data-Oriented (and Beyond) Network Architecture," SIGCOMM'07, August 27–31, 2007, Kyoto, Japan

[4] Van Jacobson, D. Smetters, James Thornton, Michael Plass, Nick Briggs, Rebecca Braynard, "Networking Named Content," ACM CoNEXT 2009, December 1–4, 2009, Rome, Italy.

[5] The Internet Engineering Task Force (IETF), ALTO Protocol, draft-ietf-alto-protocol-07

[6] D. Mazieres, M. Kaminsky, M. F. Kaashoek, and E. Witchel, "Separating Key Management from File System Security" In Proc. of SOSP '99, pages 124–139, Charleston, SC, USA, Dec. 1999

[7] 3GPP mobile broadband standard, www.3gpp.org

[8] T. Stockhammer, "Dynamic Adaptive Streaming over HTTP – Standard and Design Principles", in Proc. of ACM Multimedia Systems Conference, San Jose, CA, USA, Feb. 2011

Theodore Zahariadis is Chief Technical Officer at Synelixis Solutions Ltd and Ass. Professor at the Technological Education Institution of Chalkida, Greece. He is the Technical Coordinator of project COAST and BeyWatch. He is chairing the EC Future Media Internet Architecture Think Tank (FMIA-TT) and the EC Future Internet Architecture (FIArch) Group. His current research interests are in the fields of broadband communications, content aware networks, and sensor networking. He has published more than 90 papers in international Magazines, Journals and Conferences.

Emanuele Quacchio received the Diploma and Master Degree in Electronic Engineering from the Politecnico di Torino, Italy in 1999 and 2003. He worked two years as a researcher under grant in the Dept. of Electronics of the same university and joined ST Microelectronics at the AST-system R&D group in 2006. Since 2002 he has been working on multimedia related activities, in particular in image and video compression standard (JPEG, MPEG2/4, H.264/AVC, SVC, MVC) and multimedia streaming technologies and protocols. He published and co-authored several papers on the principal journals of engineering and conferences. He gave lectures at Politecnico di Torino on video coding and video streaming technologies.

Multiple Description Coding Based Video Streaming in Peer-to-Peer Networks

Yuanyuan Xu[1], Ce Zhu[1], Xue Jun Li[1], Wenjun Zeng[2]
[1]Nanyang Technological University, Singapore, [2]University of Missouri-Columbia, USA
[1]{xuyu0004, eczhu, lixuejun}@ntu.edu.sg, [2]zengw@missouri.edu

1. Video streaming in P2P networks

The majority of video streaming applications involve distributing video from a source server to many clients. Lack of widespread support of IP multicast limits the basic solution of video streaming to the traditional client-server model [1], in which each client sets up a connection with the server to receive a specific video. A server unicasts the video to individual clients, even if all these clients are requesting the same video. The client-server model works well with a reasonable number of users. However, during some special events, flash crowds easily overwhelm or even crash the video server, which results in a scalability problem [2].

To solve such a problem, peer-to-peer (P2P) technology is becoming a popular solution. In P2P networks, a client not only serves as a receiver downloading packets from the server or peer nodes, but also plays the role of a supplier uploading the downloaded packets to other peer nodes. In such systems, a peer node downloads the entire video content using sub-bitstreams transmitted by the server and multiple peers, instead of relying on a dedicated server. Consequently, the resources of individual peer nodes are efficiently utilized and the burden of server node is alleviated. The system capacity grows as more peers join the network, which solves the scalability problem of traditional client-server model in a cost-effective way.

Although P2P technology provides a favorable solution for video streaming, some inherent characteristics of P2P networks also pose considerable technical challenges. Video streaming services are delay sensitive, and they demand huge network bandwidth. Meanwhile, P2P networks are dynamic, and peer nodes may leave or join the network at any time. In addition, peer nodes may be heterogeneous with various resources. For example, some may not have sufficient access bandwidth to support video streaming. To ensure good visual quality, P2P video streaming systems have to be error resilient and supportive to the dynamic and heterogeneous nature of P2P environments. Thus, it remains a challenging issue to efficiently utilize and manage the resources of peer nodes.

2. MD coded P2P video streaming

Among error resilient encoding techniques, multiple description coding (MD coding or MDC) [3] is a promising one to combat transmission errors by generating different encoded versions for the same video source. Each version is referred to as a description and transmitted separately over unreliable networks. Individual description is decodable, and reconstructed quality can be refined as the number of received descriptions increases. Thus MDC offers robustness to description/packet loss over unreliable P2P networks.

Tree-based P2P system CoopNet in [1] was the first to use MDC to introduce redundancy in the transmitted video content and strip them over multiple distribution trees like that in the SplitStream system [4]. Each peer receives the substreams over a diverse set of paths. In case of partial node failures, it is highly likely that the normal peer nodes will continue to receive the majority of the descriptions and hence be able to decode a video of reasonable quality. Similar to CoopNet, many other researchers advocate the use of transmitting MD coded videos over multiple application-layer multicast trees [5-7] and delivering MD coded videos from multiple peers in mesh-based systems [8, 9].

The problem of heterogeneity of peer nodes can also be addressed by MDC. For peer nodes with varying downloading bandwidth, MDC offers a great flexibility for P2P video streaming by utilizing unbalanced MDC [10] and designing an MD coding framework [8] that can adjust MDC design parameters including the number of descriptions, the encoding rate and the redundancy level of each description on the fly. To support dynamic bandwidth heterogeneity due to a time-varying traffic congestion, a bandwidth adaptation protocol is designed in [11] for MD coded video streaming system CoopNet. To accommodate heterogeneous users with different channel conditions, a scalable FEC-based MDC packetization scheme is proposed in [12], where a peer node repacks according to the estimated network condition of its children nodes.

Furthermore, MDC can be utilized to address the incentive problem for peers' contribution to the whole P2P network as well [9], [13].

3. MDC packet scheduling for P2P video streaming

As aforementioned, MD coded video streaming offers a good alternative for robust video streaming over unreliable P2P networks. However, the varying downloading bandwidth and uploading bandwidth impose additional challenges. On one hand, the limited downloading bandwidth of a peer node may only enable the downloading of a portion of the transmitted packets. On the other hand, as the uploading bandwidth of a single peer node is generally insufficient to support delivery of all the packets, the receiver peer has to request video packets from multiple peer nodes. Therefore, mechanisms are needed to efficiently manage and coordinate limited resources in the dynamic and unreliable P2P environments.

Different from those aforementioned works, our recent work focuses on MDC packet scheduling in P2P networks. In our work, we assume the bottleneck of the system is the limited downloading bandwidth of peer nodes. A new packet scheduling framework is formulated for receiver-driven MD coded video streaming. The framework generates a schedule for fetching the expected packets from its supplier peer nodes, based on the information collected from these nodes, in two steps: (1) packet selection and (2) peer node selection. The proposed packet scheduling framework includes: (i) a rate-distortion optimized packet selection scheme that minimizes the expected distortion subject to limited downloading bandwidth; (ii) a rate-distortion based prioritized peer selection scheme that chooses an appropriate peer node for each of the selected packets.

MDC packet skipping or selection is to minimize the distortion caused by the skipped packets while meeting the constraint of limited bandwidth. This can be cast as a type of classical knapsack problem in combinational optimization. Accordingly a prioritized peer selection scheme is employed to choose an appropriate peer node for each of the selected packets. We sort these MDC packets in decreasing order using the ratio of distortion reduction to packet size, and then take turn to select peer node for these packets, taking into account varying uploading bandwidth, data availability and heterogeneous link conditions of peer nodes.

In Fig. 1, the proposed MDC scheduling scheme is compared with other existing packet selection and peer selection approaches as well as their different combinations in downloading the Foreman CIF video sequence. Specifically, we consider the following testing schemes: (i) random packet skipping/selection plus the proposed peer selection; (ii) the proposed packet selection plus random peer selection; (iii) the rarest-first packet selection in [15] plus the proposed peer selection. More simulation results are provided in [16].

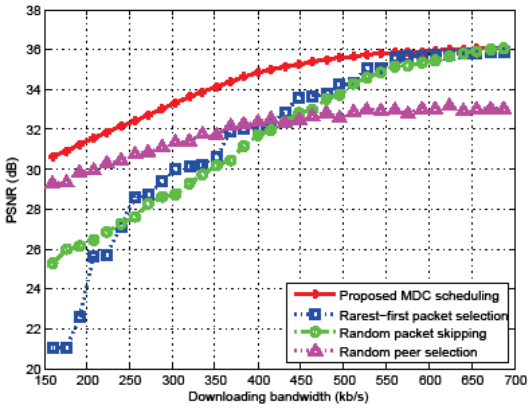


Figure 1: Performance comparison of the proposed scheduling scheme against other packet selection and peer selection counterparts.

4. Conclusion

This letter briefly revisits the advantages and challenges for video streaming in P2P networks, and discusses how MDC can be applied to achieve robust and adaptive video streaming over P2P networks. We also describe a packet scheduling framework for MDC video streaming. In the proposed scheduling framework, a rate-distortion optimized packet selection scheme is developed, followed by a rate-distortion based prioritized peer selection scheme.

References

[1] V. Padmanabhan, H. Wang, and P. Chou, “Resilient Peer-to-Peer Streaming,” in Proceedings of the 11th IEEE International Conference on Network Protocols, pp. 16- 27, 4-7 Nov. 2003.

[2] Y. Liu, Y. Guo, and C. Liang, “A survey on peer-to-peer video streaming systems,” Peer-to-peer Networking and Applications, vol. 1, no. 1, pp. 18–28, 2008.

[3] V. Goyal, "Multiple description coding: Compression meets the network," IEEE Signal Processing Magazine, vol. 18, no. 5, pp. 74–93, 2001.

[4] M. Castro, P. Druschel, A. Kermarrec, A. Nandi, A. Rowstron, and A. Singh, "SplitStream: high-bandwidth multicast in cooperative environments," in Proceedings of the nineteenth ACM symposium on Operating systems principles, 2003, pp. 298–313.

[5] E. Setton, P. Baccichet, and B. Girod, "Peer-to-peer live multicast: A video perspective," Proceedings of the IEEE, vol. 96, no. 1, pp. 25–38, 2007.

[6] G. Dan, V. Fodor, and I. Chatzidrossos, "On the performance of multiple-tree-based peer-to-peer live streaming," in INFOCOM 2007. 26th IEEE International Conference on Computer Communications. IEEE, 2007, pp. 2556–2560.

[7] N. Magharei, R. Rejaie, and Y. Guo, "Mesh or multipletree: A comparative study of live p2p streaming approaches," in INFOCOM 2007, 26th IEEE International Conference on Computer Communications. IEEE, 2007, pp. 1424–1432.

[8] M. Lu, J. Wu, K. Peng, P. Huang, J. Yao, and H. Chen, "Design and evaluation of a P2P IPTV system for heterogeneous networks," Multimedia, IEEE Transactions on, vol. 9, no. 8, pp. 1568–1579, 2007.

[9] Z. Liu, Y. Shen, S. Panwar, K. Ross, and Y. Wang, "P2P video live streaming with MDC: Providing incentives for redistribution," in 2007 IEEE International Conference on Multimedia and Expo. IEEE, 2007, pp. 48–51.

[10] E. Akyol, A. Tekalp, and M. Civanlar, "A flexible multiple description coding framework for adaptive peer-to-peer video streaming," IEEE Journal of Selected Topics in Signal Processing, vol. 1, no. 2, pp. 231–245, 2007.

[11] V. Padmanabhan, H. Wang, and P. Chou, "Supporting heterogeneity and congestion control in peer-to-peer multicast streaming," Peer-to-Peer Systems III, pp. 54–63, 2005.

[12] G. Wang, S. Futemma, and E. Itakura, "Multiple description coding for overlay network streaming," IEEE Multimedia, vol. 14, no. 1, pp. 74–82, 2007.

[13] J. Mol, D. Epema, and H. Sips, "The orchard algorithm: Building multicast trees for p2p video multicasting without free-riding," IEEE Transactions on Multimedia, vol. 9, no. 8, pp. 1593–1604, 2007.

[14] R. Bellman, S. Dreyfus, and R. C. S. M. CALIF., Functional approximations and dynamic programming. Defense Technical Information Center, 1959.

[15] N. Magharei, R. Rejaie, "Prime: Peer-to-peer receiver-driven mesh-based streaming," in IEEE/ACM Transactions on Networking, vol. 17, no. 4, pp.1052-1065, 2009.

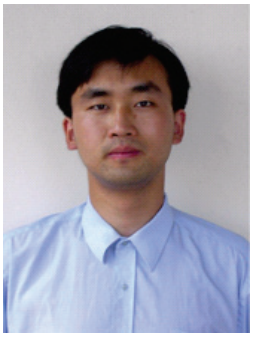
[16] Y. Xu, C. Zhu, W. Zeng, X. J. Li, "Multiple Description Coded Video Streaming in Peer-to-Peer Networks", Signal Processing: Image Communication, Elsevier, submitted.

Yuanyuan Xu received her B.Eng. and M.Eng. degrees from Beijing Jiaotong University, Beijing, China, in 2005 and 2007, respectively. She is currently pursuing her Ph. D degree in Media Technology Lab, School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. Her research interests include image and video coding, joint source network coding.

Ce Zhu (M'03–SM'04) is an Associate Professor with the School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore. He received the B.S. degree from Sichuan University, Chengdu, China, and the M.Eng and Ph.D. degrees from Southeast University, Nanjing, China, in 1989, 1992, and 1994, respectively, all in electronic and information engineering. He pursued postdoctoral research at Chinese University of

Hong Kong, City University of Hong Kong, and University of Melbourne, Australia. His research interests include image/video coding, streaming and processing, 3D video, joint source-channel coding, multimedia systems and applications.

He serves as an Associate Editor of *IEEE Transactions on Broadcasting*, *IEEE Signal Processing Letters*, *Multidimensional Systems and Signal Processing* (Springer), and as an Editorial Board Member of *Multimedia Tools and Applications* (Springer). He has served on technical/program committees, organizing committees and as track/session chairs for over 40 international conferences. He is a member of Technical Committee on Multimedia Systems and Applications (MSA-TC) of IEEE CAS Society, a voting member of Multimedia Communications Technical Committee (MMTC) of IEEE Communications Society.



Xue Jun Li received the B.Eng. (with First Class Honors) and Ph.D from Nanyang Technological University, Singapore, in 2004 and 2008, respectively. From November 2007 to August 2008, he worked as a Research Engineer and later a Research Fellow at Network Technology Research Centre. From August 2008 to September 2008, he worked in Temasek Laboratories @ NTU as a Research Scientist. Currently, he is an academic staff with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His research interests include design / analysis of wireless networking protocols, modeling / design of radio frequency integrated circuits, computer network design and implementation, system optimization and genetic algorithm analysis / development.

Wenjun (Kevin) Zeng (S'94-M'97-SM'03) is an Associate Professor with the Computer Science Department of University of Missouri, Columbia, MO. He received his B.E., M.S., and Ph.D. degrees from Tsinghua University, the University of Notre Dame, and Princeton University, respectively. His research interest is in the general areas of multimedia processing, communications, and networking.

Prior to joining Univ. of Missouri in 2003, he had worked for PacketVideo Corp., Sharp Labs of America, Bell Labs, and Panasonic Technology. From 1998 to 2002, He was an active contributor to the MPEG4 IPMP standard and the JPEG 2000 image coding standard, where four of his proposals were adopted. He is an Associate Editor of the *IEEE Trans. on Info. Forensics & Security, IEEE Trans. on Circuits & Systems for Video Technology,* and *IEEE Multimedia Magazine,* and is on the Steering Committee of *IEEE Trans. on Multimedia,* of which he also served as an AE from 2005 to 2008. He is serving as the Steering Committee Chair of *IEEE Inter. Conf. Multimedia and Expo,* and has served as the TPC Chair or Vice Chair for a number of IEEE conferences. He was a Guest Editor of *the Proceedings of the IEEE's Special Issue on Recent Advances in Distributed Multimedia Communications* published in January 2008 and the Lead Guest Editor of *IEEE Trans. on Multimedia's Special Issue on Streaming Media* published in April 2004.

Adaptable Media Coding Techniques for Social/P2P Networks

Naeem Ramzan
Queen Mary University of London, UK
naeem.ramzan@eecs.qmul.ac.uk

1. Introduction

Future Media Internet will necessitate dispensing high quality multimedia contents in an efficient, supple and personalized way through dynamic and heterogeneous environments. Multimedia content over internet are becoming a well-liked application due to users' growing demand of multimedia content and extraordinary growth of network technologies. A broad assortment of such applications can be found in these days, e.g. as video streaming, video conferencing, surveillance, broadcast, e-learning and storage. In particular for video streaming, over the Internet are becoming popular due to the widespread deployment of broadband access. In customary video streaming techniques the client-server model and the usage of Content Distribution Networks (CDN) along with IP multicast were the most desirable solutions to support media streaming over internet. However, the conventional client/server architecture severely limits the number of simultaneous users for bandwidth intensive video streaming, due to a bandwidth bottleneck at the server side from which all users request the content. In contrast, Peer-to-Peer (P2P) media streaming protocols, motivated by the great success of file sharing applications, have attracted a lot of interest in academic and industrial environments. With respect to conventional approaches, a major advantage in using P2P is that each peer involved in a content delivery contributes with its own resources to the streaming session. However, to provide high quality of service, the video coding/transmission technology needs to be able to cope with varying bandwidth capacities inherent to P2P systems and end-user characteristics such as decoding and display capabilities usually tend to be non-homogeneous and dynamic. This means that the content needs to be delivered in different formats simultaneously to different users according to their capabilities and limitations.

In order to handle such obscurity, scalability emerged in the field of video coding in the form of Scalable Video Coding (SVC) [1–4] and Multiple Description Coding (MDC) [5-6]. Both SVC and MDC offers an efficient encoding for applications where content needs to be transmitted to many non-homogeneous clients with different decoding and display capabilities.

Moreover, the bit-rate adaptability inherent in the scalable codec designs provides a natural and efficient way of adaptive content distribution according to changes in network conditions.

In general, a SVC sequence can be adapted in three dimensions, namely, temporal, spatial and quality dimensions, by leaving out parts of the encoded bit-stream, thus reducing the bit-rate and video quality during transmission. By adjusting one or more of the scalability options, the SVC scheme allows flexibility and adaptability of video transmission over resource-constrained networks.

The eventual objective of employing SVC/MDC in social/P2P is to maximize the end-users' quality of experience (QoE) for the delivered multimedia content by selecting an appropriate combination of the temporal, spatial and quality parameters for each client according to the limitation of network and end user devices .

2. Scalable Video Coding

During the last decade a noteworthy amount of research has been devoted to scalable video coding with the aspire of developing the technology that would offer a low-complexity video adaptation, but preserve the analogous compression efficiency and decoding complexity to those of conventional (non-scalable) video coding systems. This research evolved from two main branches of conventional video coding: 3D wavelet [1] and hybrid video coding [2] techniques. Although some of the earlier video standards, such as H.262 / MPEG-2 [3], H.263+ and MPEG-4 Part 2 included limited support for scalability, the use of scalability in these solutions came at the significant increase in the decoder complexity and / or loss in coding efficiency. The latest video coding standard, H.264 / MPEG-4 AVC [2] provides a fully scalable extension, SVC, which achieves significant compression gain and complexity reduction when scalability is sought, compared to the previous video coding standards.

The scalability is usually required in three different directions (and their combinations). We define

these directions of scalability as follows:

1. Temporal scalability refers to the possibility of reducing the temporal resolution of encoded video directly from the compressed bit-stream, i.e. number of frames contained in one second of the video.
2. Spatial scalability refers to the possibility of reducing the spatial resolution of the encoded video directly from the compressed bit-stream, i.e. number of pixels per spatial region in a video frame.
3. Quality scalability, or commonly called SNR (Signal-to-Noise-Ratio) scalability, or fidelity scalability, refers to the possibility of reducing the quality of the encoded video. This is achieved by extraction and decoding of coarsely quantised pixels from the compressed bit-stream.

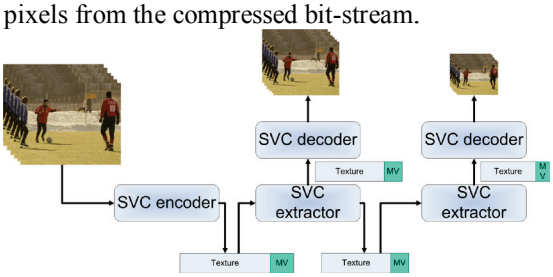


Figure 5: A typical scalable video coding chain and types of scalabilities by going to lower-rate decoding.

An example of basic scalabilities is illustrated in Figure 5, which shows a typical SVC encoding, extraction and decoding chain. The video is encoded at the highest spatio-temporal resolution and quality. After encoding, the video is organised into a scalable bit-stream and the associated bit-stream description is created. This description indicates positions of bit-stream portions that represent various spatio-temporal resolutions and qualities. The encoder is the most complex between the three modules. The compressed video is adapted to a lower spatio-temporal resolution and / or quality by the extractor. The extractor simply parses the bit-stream and decides which portions of the bit-stream to keep and which to discard, according to the input adaptation parameters. An adapted bit-stream is also scalable and thus it can be fed into the extractor again, if further adaptation is required. The extractor represents the least complex part of the chain, as its only role is to provide low-complexity content adaptation without transcoding. Finally, an adapted bit-stream is sent to the decoder, which is capable of decoding any adapted scalable video bit-stream.

3. Scalable Video over P2P network

The proposed system is based on two main modules: scalable video coding and the P2P architecture. In this system, we assume that each peer contains the scalable video coder and the proposed policy of receiving chunk is to make sure that each peer at least receives the base layer of the scalable bit-stream for each group of picture (GOP). Under these circumstances, peers could download different layers from different users, as shown in Figure 2.

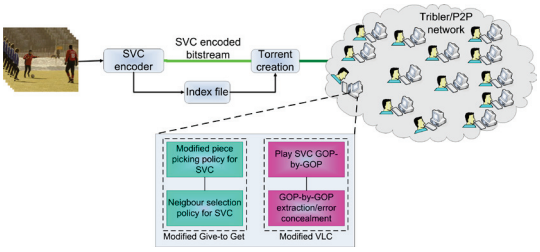


Figure 2: An example of the proposed system for scalable video coding in P2P network.

In this section, we formulate how the scalable layers are prioritized in our proposed system. First we explain how the video segments or chunks are arranged and prioritized in our proposed system

3.1 Piece picking policy

The proposed solution is a variation of the "Give-To-Get" algorithm [8], already implemented in Tribler. Modifications concern the piece picking and neighbour selection policies. Scalable video sequences can be split into GOPs and layers [7] while BitTorrent splits files into pieces. Since there is no correlation between these two divisions, some information is required to map GOPs and layers into pieces and vice versa. This information can be stored inside an index file, which should be transmitted together with the video sequence. Therefore, the first step consists of creating a new torrent that contains both files. It is clear that the index file should have the highest priority and therefore should be downloaded first. Once the index file is completed, it is opened and information about offsets of different GOPs and layers in the video sequence is extracted.

The detail of modified piece picking policy can be found in [7]. Another issue is the wise choice of the neighbours.

3.2 Neighbour selection policy

It is extremely important that at least the base layer of each GOP is received before the window shifts. Occasionally, slow peers in the swarm (or slow neighbours) might delay the receiving of a BT

piece, even if the overall download bandwidth is high. This problem is critical if the requested piece belongs to the base layer, as it might force the playback to pause. Therefore, these pieces should be requested from good neighbours. Good neighbours are those peers that own the piece with the highest download rates, which alone could provide the current peer with a transfer rate that is above a certain threshold [9].

4. Conclusions

This E-letter has presented an overview of SVC with the perspective of content distribution over Social/P2P networks. These coding schemes provide natural robustness and scalability to media streaming over heterogeneous networks. The amalgamation of SVC and P2P are likely to accomplish some of the Future Media Internet challenges. Tangibly, SVC over P2P presumes an excellent approach to facilitate future media applications and services, functioning under assorted and vibrant environments while maximizing not only Quality of Service (QoS) but also Quality of Experience (QoE) of the users.

Acknowledgments

This research has been partially funded by the European Commission under contract FP7-248474 SARACEN.

References

[1] M. Mrak, N. Sprljan, T. Zgaljic, N. Ramzan, S. Wan and E. Izquierdo, Performance evidence of software proposal for Wavelet Video Coding Exploration group, ISO/IEC JTC1/SC29/WG11/ MPEG2006/M13146, 76th MPEG Meeting, Montreux, Switzerland, April 2006.

[2] ITU-T and ISO/IEC JTC 1, Advanced video coding for generic audiovisual services, ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4 AVC).

[3] ITU-T and ISO/IEC JTC 1, Generic coding of moving pictures and associated audio information – Part 2: Video, ITU-T Recommendation H.262 and ISO/IEC 13818-2 (MPEG-2 Video).

[4] H. Schwarz, D. Marpe and T. Wiegand, "Overview of the scalable video coding extension of the H.264 / AVC standard," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, Iss. 9, pp. 1103 - 1120, September 2007.

[5] T. Berkin Abanoz, A. Murat Tekalp, "SVC-based scalable multiple description video coding and optimization of encoding configuration", Signal Processing: Image Communication, Volume 24, Issue 9, October 2009, Pages 691-701.

[6] T. Tillo, M. Grangetto, G. Olmo, "Redundant slice optimal allocation for H.264 multiple description coding",, IEEE Trans. Circuits Syst. Video Technol. 18 (1) (January 2008) 59–70.

[7] S. Asioli, N. Ramzan, E. Izquierdo, "A Novel Technique for Efficient Peer-To-Peer Scalable Video Transmission", in Proc. of European Signal Processing Conference (EUSIPCO-2010), Aalborg, Denmark , August 23-27 2010.

[8] J. A. Pouwelse, P. Garbacki, J.Wang, A. Bakker, J. Yang, A. Iosup, D. H. J. Epema, M. Reinders, M. R. van Steen, and H. J. Sips, "Tribler: A social-based based peer to peer system," in 5th Int'l Workshop on Peer-to-Peer Systems (IPTPS), Feb. 2006.

[9] Naeem Ramzan, Emanuele Quacchio, Toni Zgaljic, Stefano Asioli, Luca Celetto, Ebroul Izquierdo, Fabrizio Rovati, "Peer-to-peer Streaming of Scalable Video in Future Internet Applications", IEEE Communication Magazine, vol. 49, Issue 3, pages 128–135, March 2011.

Naeem Ramzan, M.S in Telecom from Brest France and PhD in Electronics Engineering form Queen Mary University of London in 2004, and 2007 respectively. From 2004-2007 he worked on EU project aceMedia in the Multimedia & Vision Group in Queen Mary University of London. Currently, he is a senior researcher and core member of technical coordination team in the EU funded project PetaMedia, and SARACEN. His research interest includes image and video coding, scalable video coding, multimedia search and retrieval, multimedia transmission over wireless and P2P networks. He has been a co-chair of the IEEE Student Branch Congress, UK, 2008. He has been a co-chair of Special session in International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), UK, 2009; ACM Multimedia Information Retrieval Conference, USA, 2010; and WIAMIS, Italy, 2010. He is also the co-organizer and co-chair of ACM Multimedia workshop SAPMIA, Italy, 2010 and ACM workshop on Social and Behavioural Networked Media Access (SBNMA), Scottsdale, Arizona,

USA, 2011. He has served as Guest Editor of a special issue of the Elsevier Journal Signal Processing: Image Communication. He is the author or co-author of more than 50 research publications. He also serves as a reviewer of several international journal and conferences. He is the chair of the Summer School on Social Media Retrieval (S3MR), June 2011.

My Own, Personal Video Broadcast

Charalampos Z. Patrikakis[1] and Lemonia Argyriou[2]
[1]Technological Educational Institute of Piraeus, Greece, [2] National Technical University of Athens, Greece
[1]bpatr@teipir.gr, [2]lemi_arg21@hotmail.com

1. Introduction

In the last years, multimedia consumption and distribution between internet users is highly increasing from day to day. Moreover, Internet traffic is dominated by peer to peer systems and the exchange of multimedia and all kinds of information over a variety of P2P platforms and protocols (BitTorrent, e.a.). On the other edge platforms such as Youtubeleading to take over the multimedia industry of uploading and sharing video in contradiction to peer to peer systems. Social Networking is also taking from time to time an importantly increasing place in multimedia distribution in everyday usage.

According to Nielsen's latest Quarterly Report [1], today's consumers watch more video, across time and place than ever before. Indicatively, the mobile video audience increases by 51,2% year over year and has reached the 20 million users for the first time. Moreover, the proliferation of broadband internet access, with high-speed connections that improve online video, has reached the 63.5% of homes in USA, while 52.7% of them now are equipped with HDTVs.

All the above, along with the bolstering Video-on-Demand consumption (timeshifting audience increased by 14%) has lead to a strong research interest on efforts to provide higher levels of personalized video viewing. Simultaneously, mobile tablet devices usage in accessing the internet continues to proliferate (versus PCs, laptops and netbooks). Smartphones penetration is detected to the 38% of households this year while new multimedia devices like Apple's iPad,create additional expectations in media consumption demand increment.Therefore, consumers don't end up the usage of any multimedia device. In contrast, what is detected is a strong demand of watching and sharing non-stop video simply by shifting devices according to where they are.

Personalisation owns also a highly important place is this increasing consumption. Users are demanding of sophisticated and added value multimedia services that will offer them new personalised viewing experiences where they will be strongly involved.

2. Supporting InteractivityAndPersonalisation

Following the issue of highly demand on personalisation, sophisticated mechanisms of tagging and annotation [2] have been created in order to provide enhanced personalised services of high Quality of Experience (QoE) [3]. Moreover, new broadcasting network technologies have been used to enhance the viewing experience. Seamless bitrate switching, transparent transitions and mobility between access network technologies during media streaming and automatic, context-aware adaptation of transmission parameters are characteristic examples of those techniques.

In the scope to also provide a better experience strong effort was spent on managing the multimedia transmitted content. Advanced image/video annotation and characterization mechanisms were applied to facilitate the automatic (even for live content) annotation of multimedia content so that truly personalized media streaming can be offered, based on the matching of user selections to content characterisation. In the following sections, we will describe the way in which personalised media distribution in a cost effective way can be achieved, through the deployment of state of the art technologies, as these are used in the context of two EU research projects: My e-Director 2012 and SARACEN ("Socially Aware, collaboRative, scalable Coding mEdiadistributioN.").

2.1 My e-Director 2012approach

The My e-Director 2012[4] platform is a web-based platform created to provide high quality personalised viewing experience of athletic events.My e-Director 2012 platform aims to act as a 'personal director' tailoring itself to each user, based on stipulated preferences. It is focusing on recording and using user preferences to offer recommendations to users about better camera views that should interest them, whilst permitting them to interactively tailor their viewing experience, offering them in addition to DVR-like

controls, extra features such as smart-zooming to focus on areas of interest.

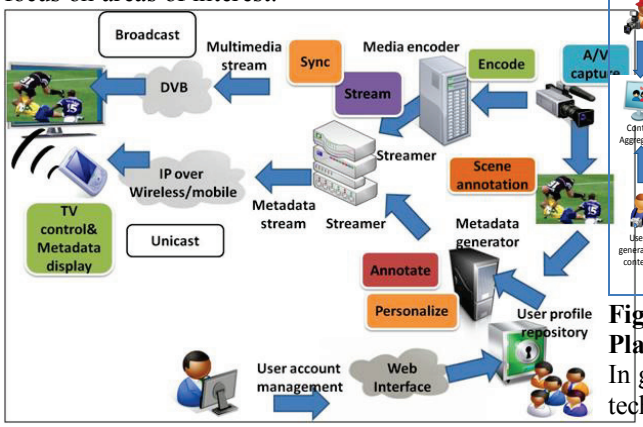


Figure 1.General platform architecture overview

The major challenge posed is the provision of such features utilising automatically extracted metadata for live coverage of parallel events.

My e-Director 2012 is using techniques of automatic visual scene understanding and annotation. Specific sub modules within the Content Distribution Subsystem receive the raw video streams, the generated metadata stream and the recommendation information in order to generate final 'edited' video streams that are delivered to the user (together with recommendations).

2.2. SARACENapproach

SARACEN[5] project addresses two issues that have strongly affected the recent research efforts: the distribution of scalable coded media and the use of peer to peer architectures in order to deliver streaming media

The SARACEN P2P System Architecture concept follows a layered approach with two fundamental interfaces:

-Interface between the bearer Network and the P2P system: derived from the OSI reference model, providing bidirectional transport and measurements of connection to peers in order to optimize the transport.

-Interface between the P2P system and the P2P-enabled device: providing the functions to render the contents and process metadata in metadata repository.

In particular, the SARACEN system creates an enhanced protocol suite that will enable peers to request/exchange different layers of a scalable video stream, or different descriptions.

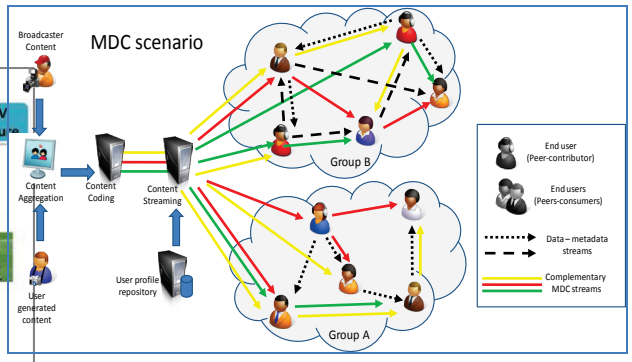


Figure 2. Architecture of the SARACEN P2P Platform for an MDC scenario

In general, the platform is taking advantage of P2P technologies for the efficient distribution of multimedia content, scalable coding techniques for addressing QoE issues and social networking for providing context aware content access and distribution.

3. Added Value through the Combination of P2P And Personalization P2P

Following the concept and architecture of My e-Director 2012 and SARACEN, we will try to combine the ideas and functionalities offered by the two platforms in a unified architecture that is capable of providing interactivity, personalisation and efficiency through the use of P2P. This would be particularly interesting for the coverage of multi-camera, multi-event broadcasts on a world wide range, such as Olympic games in an efficient and cost effective manner. In such cases, there are always cases of events that are of limited interest to the global community but of particular interest to specific groups of viewers: take for example the case of a heavyweight final in boxing between Greece and Turkey, taking place in parallel with the final of 800 meters of men. Much as the second event would be of interest to the general viewers of the games, the first one would interest Greek and Turkish viewers, especially those that are interest in the sport. Taking into account that there are large Greek and Turkish communities in several countries (USA, Germany), there would be interest for added value services to be provided to these users. Internet and the technologies presented earlier as regards personalisation, adaptation and interactivity provide an excellent framework over which coverage of the event could be based.

In this, the innovative techniques used on the two other projects could be combined. More specifically, the coverage of the event could be supported in a cost efficient way, using P2P technologies, thus eliminating the need for

reserving excessive bandwidth. Taking into account that interest to the particular event is higher in particular geographic areas (countries of thecontestants and places with large national communities from the two countries), a P2P distribution could also benefit from this, while bandwidth consumption from the origin of the transmission to these countries could be minimized. Another advantage could come out of the use of the power of social networking, that could assist even further the organisation of P2P distribution network, through the introduction of user profiles in the P2P distribution scheme, so as to be used in the selection of peers, but more importantly, in the personalisation of the coverage of the event. Here, the profiles of the users could be used in order to group them into clusters of similar profiles. For each group, a virtual profile corresponding to an average or characteristic user can be extracted, based on the use of the appropriate clustering method, selected out of the techniques proposed in the global bibliography [6].

A separate channel could be set up for each profile, following different coverage of the event in terms of selected camera, statistics and commentary. The selection of the number of different groups could be based on the availability of resources, while the formation of the virtual profiles can be based on the maximization of interests' commonalities between users [7].

The proposed architecture, makes use of the existing camera feeds, that can be directly formatted into streams (the number of which can be defined through the availability of resources in terms of bandwidth and processing power in the servers), while there is no need for director, as the recommendations coming from the annotation (in real time) of the broadcasted content, combined with the virtual profiles for each group can be combined in order to select the most appropriate camera and accompanying information (i.e. statistics) to be streamed to the users belonging in each profile group. Streaming is based on P2P delivery of the stream to the participants of each group, while the content of the stream is dynamically formed so as to best match the preferences of the Virtual profile corresponding to each group.

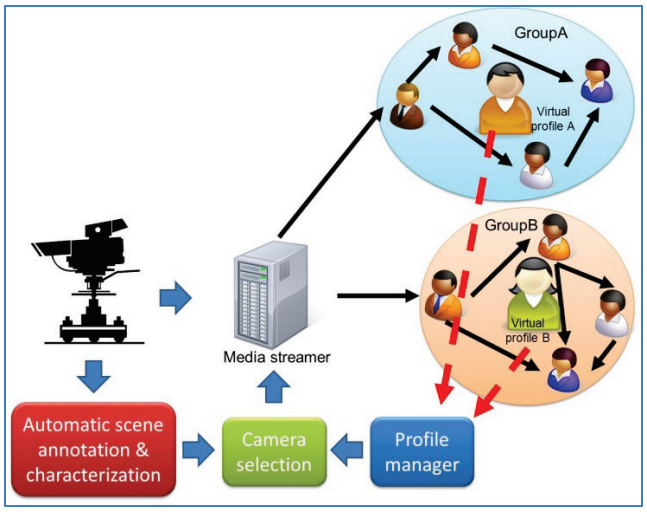


Figure 3.Combination of personalization mechanisms through user profiles and P2P media distribution.

The architecture follows a minimalistic approach, in which the use of technologies for personalization and P2P distribution has been made having in mind the ability to support the corresponding service over a wide set of platforms and terminal equipment. Added value could come through the use of social networking APIs, in order to introduce the ability to support real time communication over the participant peers in the video distribution (i.e. use of tweet feeds or comments in real time), while adaptation of the streamed media to the network and terminal conditions can be supported through the adoption of scalable coding techniques for encoding video such as Multiple Description Coding, so that the quality of the distributed stream can be adapted to best serve the needs and capabilities of each user.

4. Acknowledgments

The research leading to these results has received funding from the European Union's Seventh Framework Programme ([FP7/2007-2013]) under grant agreement n° ICT-215248.

References

[1] What Consumers Watch: Nielsen's Q1 2010 Three Screen Report, published on June 11 2010, Available at: http://blog.nielsen.com/nielsenwire/online_mobile/what-consumers-watch-nielsens-q1-2010-three-screen-report/, last accessed on 09/03/2011.

[2] N. Katsarakis and A. Pnevmatikakis, 'Event Detection in Athletics for Personalized Sports Content Delivery', WIAMIS 2009, London, UK, May 2009.

[3]R. S. Cruz, J. Espadanal, and M. S. Nunes, "A Personalized HTTP Adaptive Streaming WebTV". Proceedings of the First International Conference on User Centric Media Workshop, UCMedia '09. Venice, Italy.

[4] My e-Director 2012 IST project, Available at: http:// www.myedirector2012.eu, last accessed on 10/03/2011.

[5] SARACEN "Socially Aware, collaboRative, scalable Coding mEdiadistributioN." IST project, Available at: http://www.saracen-p2p.eu, last accessed on 10/03/2011.

[6] Sotiropoulos D.N, Tzihrintzis G.A., Savvopoulos A., Virvou M., A comparison of customer data clustering techniques in an e-shopping Application, in: Proceedings of 2nd International Workshop on Web Personalization, Recommender Systems and Intelligent User Interfaces, Ireland, 2006.

[7] Karamolegkos P.N., Patrikakis C.Z., Doulamis N.D., Vlacheas P.T., Nikolakopoulos I.G., An evaluation study of clustering algorithms in the scope of user communities assessment, Computers and Mathematics with Applications, 58 (8), pp. 1498-1519, 2009.

Charalampos Z. Patrikakis received his Dipl.-Ing. and Ph.D from the Electrical Engineering and Computer Science Department of NTUA, Greece. He has participated in more than 25 European and International research projects including ESPRIT, RACE, ACTS, IST, ICT, eContent andeContentPlus. He has published over 90 papers in international journals, books and conference proceedings. He is an Assist. Professor at the Electronics Engineering Dept. of the Technological Educational Institute of Piraeus and a senior member of IEEE.

Lemonia Argyriou received her Dipl. from the Department of Informatics at the University of Piraeus, Greece. She is now a postgraduate student at the Engineering – Economical Systems MBA program of NTUA. She has been working as a research associateof ICCS in EU research projects in multimedia networking.