INTER-MODAL SELECTIVE 3D CODING OF PET-CT DATASETS

Alberto Signoroni [†], Stefano Masneri [†], Riccardo Leonardi [†], and Isabella Castiglioni [§]

[†] DEA – University of Brescia
via Branze, 38, I25123, Brescia, Italy
phone: + (39) 030 3715432, fax: + (39) 030 380014
email: *firstname.lastname*@ing.unibs.it

[§] IBFM – CNR
via Fratelli Cervi, 93, I20090 Segrate (Mi), Italy
phone: + (39) 02 26432715
email: isabella.castiglioni@hsr.it

ABSTRACT

In this work we introduce a new selective coding approach suitable for co-registered multi-modal medical images and we apply it to large PET-CT volumes. Salience information guiding a space variant reconstruction quality of the anatomical volume (CT) is generated through an automatic analysis of the functional volume (PET). This allows a versatile multiple volume-of-interest coding with arbitrary 3D-shape and scaling-factors and without the need of side-information to be transmitted. The proposed solutions are suitable for critical applications where high and optimized compression ratio, minimization of human intervention and full diagnostic quality preservation are all required.

1. INTRODUCTION

Telemedicine systems are in constant evolution worldwide, especially in the filed of diagnostic imaging, where teleradiology solutions are more and more pervasive. However, there are a number of relevant clinical applications and image modalities which are still asking for new enabling or improving technological solution because of their critical requirements. This is, for example, the case of scenarios where large 3D image datasets (volumes) are to be exchanged frequently and in short time between two hospitals for tele-diagnosis or second opinion purposes. Emblematic cases are those related to multi-modal tomographic systems, which allow to produce, in a single examination session, spatially co-registered image volumes related to different and complementary imaging modalities. In particular, in the past ten years, Positron Emission Tomography (PET) combined with Computed Tomography (CT) has shown a dramatic increase, becoming nowadays the reference methodology for the cancer staging and restaging and affecting the therapeutic decision (surgical, medical, and radiotherapeutic) in the 40-50% of cases. However, the use of these technologies and the interpretation of the images produced by these systems require specific experiences which are not so diffused as the scanners are.

The reduction of the transfer time (by mean of a high compression ratio) and the preservation of the diagnostic quality of medical images are concurrent but colliding primary needs, so that approaching them disjointly would likely lead to suboptimal solutions. Prominent technological solutions are those which involve *selective* (also named *region-of-interest*) coding approaches under, or driven by, diagnostic quality preservation requirements and criteria.

In this paper we present a new approach to the selective coding of multi-modal datasets. We explore and exploit the possibility of a cross-modal derivation of meaningful saliency information and we demonstrate how this can work in the case of CT volume coding with *volumes-of-interest* derived from the spatially co-registered PET volume. In particular we provide an algorithm which is able to recognize in PET data (functional characterization of cancer disease) a diagnostic information suitable to modulate the coding strength of a coding algorithm on the corresponding CT image sets (anatomical characterization of cancer disease), optimizing CT image compression procedure with respect to diagnostic quality.

The system is also designed to guarantee an adequate interplay among what can be reliably done automatically and what can only result from a high level professional reasoning. Combined with an efficient 3D wavelet coding architecture, the overall selective coding system will entail concrete, effective and reliable solutions for teleradiology and/or archiving applications in the diagnostic loop.

2. RATIONALE

2.1 PET-CT scans: description and applications

PET is an imaging diagnostic technique allowing images of the spatial distribution of a radiotracer, administered to a patient, to be obtained in vivo and non-invasively. PET with ^{18}F-Fluorodeoxyglucose (^{18}F-FDG) allows the detection of neoplastic lesions as foci of increased metabolism of glucose, whose solid tumors are extremely avid. Current generation PET-CT systems makes a more accurate lesion assessment possible, thanks to the simultaneous analysis of fused functional PET and anatomical CT images, PET being crucial to detect presence of lesions (also when CT does not show), CT being crucial in univocally localizing PET lesions, providing the size, shape, and anatomical characterization of lesions. This results in having two diagnostic procedures in one session, with great advantages for the patient but at the price of very large amount of digital data, particularly of CT data. Reporting PET-CT studies is performed by nuclear medicine physicians or by radiologists, depending by hospital conventions or country healthy rules, but, in all cases, notwithstanding expertise in interpreting both PET and CT images is mandatory. Therefore, in this scenario, a decisive role can be played by the cooperation between a hospital with a consolidated experience in the field and other hospitals which approach the recent technologies for the first time or with less experience. PET-CT report is a report of a PET study for which CT co-registered study is used to better characterize oncological lesions identified by PET. In such a context, anatomical accuracy of CT image volume outside lesion foci revealed by PET is not a primer interest. This corroborates the use of a selective coding approach for CT data. Moreover, due to its vocation, PET-CT image sets often consist in whole-body scans occupying up to 350Mbytes, CT volume being the dominant bit-demanding modality (80% of the whole data sets). On common ADSL connections, the transfer time for a single lossless coded PET-CT dataset may vary from minutes to hours depending on the link throughput. This can be unacceptable from both quality of (health) service, logistic and economic points of view, and urgently asks for proper coding solutions which could guarantee in all cases efficient services and accurate and reliable remote data interpretation.

2.2 Selective coding for 2D and 3D images

Aiming at coding efficiency, what stated above suggests to seek selective coding solutions, where coding strength and reconstruction quality are modulated in the spatial domain by means of a saliency map derived from some form of a-priori information about the presence of regions-of-interest (RoIs) on the image. Depending on application and requirements, RoI shape information can arrive at the decoder site either explicitly (as side-information) or implicitly (no bit spent for it). Selective coding can be introduced in different ways to work in conjunction with transform coding schemes and standards, either block-based (e.g. DCT-based, JPEG) or full-

frame based (e.g. wavelet-based, JPEG2000). For small block partitions and frequency-domain transform coding (e.g. 8x8 DCT based coding) target distortion can be differentiated block-wise. Here, only block shaped RoI are possible and blocking artifacts affecting this kind of schemes can get accentuated. In whole frame multi-resolution transform coding (e.g. wavelet coding), RoIs defined in the image domain have a quite natural representation in the transformed domain because of the finite spatial support of influence of the transformed coefficients. Two possible exploitation of transformed domain RoI mapping have been proposed which have also been implemented in the wavelet-based JPEG2000 standard:

1. RoI and background (BG) coded in virtually separated bit-streams. Two main approaches can be reported:
   – the one known as "max-shift", which is implemented in JPEG2000 part1 (baseline) [1], allows arbitrary RoI shapes without the need to code and transmit them: RoI coefficients are scaled-up such as $\min(scaled\text{-}up\ RoI\ coeffs) \leq \max(BG\ coeffs)$, therefore, all subband RoI coefficients are coded first. This separation causes some RD inefficiencies, sharp quality transitions between RoI and BG region, difficulty in handling different saliency levels in the same image. Some flexibility for "max-shift" method has been introduced in [2];
   – the use of shape-adaptive wavelet transform (e.g. see [3]). This lead to the same "max-shift" drawbacks and, in addition, the shape info must be explicitly encoded and transmitted.

2. "Scaling-based" solutions: RoI coefficients are scaled-up by a suitable factor and remain immersed in the BG ones. Therefore a unique, possibly scalable or embedded, bitstream is produced where RoI image quality is steady maintained superior with respect to the BG one with a quality disparity depending on the required/selected scaling-factor. This leads to minor performance loss and allows some more degree o freedom in managing multiple RoI saliency levels and/or gradual RoI-BG quality decay [4, 5]. However, RoI shape must be encoded as side information. As a consequence, in order to not impair coding performance, only simple RoI shapes (e.g. rectangles or ellipses) are usually allowed. A "scaling-based" standardized implementation can be found in JPEG2000 part 2 (extensions [6]).

Medical imaging is populated by many modalities which generate 3D datasets (volumes). Volumetric data require maximum bit-saving efforts but are not so easy to handle and to interact with for the definition of Volumes of Interest (VoIs). High-performance 3D extensions of wavelet coding techniques have been proposed [7, 8] which exploit the redundancy in the third dimension. Also RoI coding has been extended to VoI coding. In [9] different VoI approaches are presented and compared, while solutions to the problems of handling VoI-BG quality decay and of concisely defining 3D shapes for scaling-based coding can be found in [8].

In the next section we propose a solution presenting the advantages of both "max-shift" and "scaling-based" approaches and avoiding their respective drawbacks. A peculiarity of the proposed solution for inter-modal PET-CT VoI coding is in fact to allow arbitrary volumetric shapes without side-information to be transmitted.

3. MATERIALS AND METHODS

3.1 PET-CT scans

PET-CT whole-body scans with ^{18}F-FDG were considered as datasets for the application of selective coding algorithms. They were performed on oncological patients by the three-dimensional PET-CT system HD Reveal [10], a multi-modal scanner recently installed at the Hospital São Raphael of Salvador de Bahia, Brasil, and connected by a telemedicine system to the Hospital San Raffaele of Milan, Italy, for second-opinion purposes. Due to the presence of network bottlenecks (mostly on the local Brazilian side), actual time required for the transmission of whole-body PET-CT studies is up to few hours, thus recommending the use of high-compression algorithm for real-time remote reporting.
Brazilian patients were selected considering both staging and restaging diagnostic question, and different lesion foci. This is

Table 1: PET-CT scans considered in this study

Pat.	Diagnostic question	Organs	slices
1	Restaging (recurrence in patient with brain metastasis from lung cancer)	brain	108
2	Staging (prostate adenocarcinoma)	abdomen	113
3	Detection of primary tumour site	whole-body	307
4	Restaging (recurrence in patient with colon adenocarcinoma, evidence of tumoral markers increase)	whole-body	355
5	Staging (pre-chemio therapy in patient with Hodgkin lymphoma)	whole-body	404

reported in Table 1, where the diagnostic question, the imaged organ and the number of PET-CT images to be transmitted are shown. Patients were administered with 1mCi/10kg of ^{18}F-FDG and scanned by the PET-CT system after 45 minutes from the radiotracer injection. All the patients underwent a low voltage CT scan (10mA), a 140keV (from 90 to 110mA) CT study and a 3D PET study (3 min/scan). PET images were reconstructed with CT attenuation correction, and both PET (128x128) and 140Kev-CT (512x512) image sets were represented in DICOM 16 bit/voxel binary format.

3.2 Selective coding

3.2.1 Mask Generation

We consider, as required in any PET-CT application, contextual transmission of PET and CT datasets. Therefore VoI masks must be extracted from decoded PET volumes as they result available at the decoder site. This fact is non critical and a lossless scenario for PET data will be considered thereinafter. Oncological PET datasets show both pathological and physiological uptake and it is not affordable for an automatic analysis to reliably distinguish between them. On the other hand, insofar as relevant uptake involves a minority of voxels, and this is the case of the majority of datasets, inclusion of physiological uptake regions can be done without consequences on the effectiveness of the selective coding solution. This has been also considered appreciable in general as it can be seen as a safeguard criterion. Therefore our action has been concentrated on trying to automatize the exclusion of bigger physiological uptake regions and removing irrelevant uptake noise. In particular we derive inclusive criteria for saliency mask construction by analyzing a two variable histogram derived by a bivariate ranking of binary connected components extracted from the PET volume. For each PET volume we produce a diagram $D(lev, vol)$, like the one in Figure 1, that represents, for each intensity channel (lev on x-axis), a morphometric ranking based on a volume measure (vol on y-axis) of the clusters of voxels above the threshold lev. Representative vol classes are taken on a cubic scale, i.e. delimited by values $vol = i^3, i \in [1, i_{max}]$. This diagram is used to derive exclusion and inclusion criteria for PET voxel clusters in mask generation. As stated, high PET voxel intensities are not necessarily related to pathological tissues, some parts of the body showing in fact high uptake because of their physiological activity (e.g. the brain), or because they simply drain the radiotracer (e.g. the bladder). Therefore, within the possibility of an automatic analysis, such body parts should not be considered during the generation of the mask. In addition, PET volume is highly noisy, so that it is possible to find a certain amount of uptake granularity consisting in small clusters of captation voxels with no diagnostic interest. Following these considerations we developed a mask generation algorithm that acts in two steps:

1. **Cluster exclusion**: clusters are excluded by setting to zero points of the diagram $D(lev, vol)$ according to the following criteria:
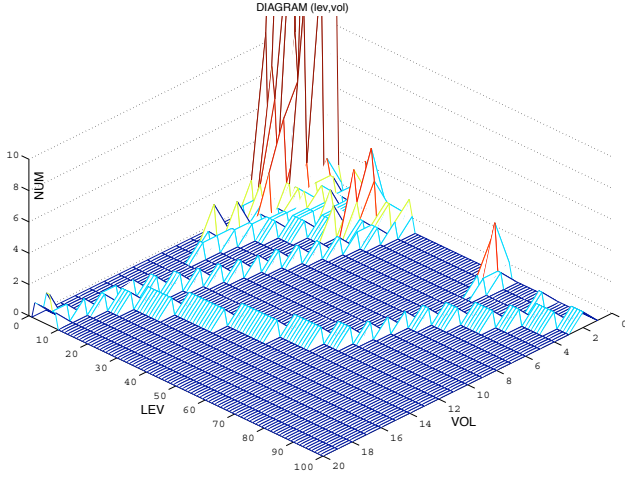
Figure 1: D(*lev,vol*) distribution of clusters in the PET dataset depending on their dimension *vol* and the threshold value *lev*.
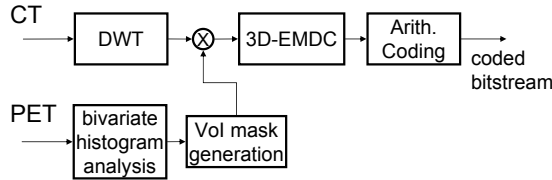

Figure 2: CT coding framework.

E-a) Removal of single-voxel clusters (noise). E-b) Smaller clusters are regarded as a noise component and excluded if their number surpass, at each *lev*, a defined number. E-c) Starting from the highest *lev* and smallest *vol* we follow, by a neighborhood navigation of $D(lev,vol)$, a typical footprint of the bladder consisting in a tail which is clearly visible in Figure 1. We decrement $D(lev,vol)$ on all points of that tail. E-d) Too big cluster are excluded because belonging to the bladder or other big physiological uptake structures (e.g. the heart).

2. **Cluster inclusion**: for each value of *lev* we analyze the corresponding binary thresholded PET dataset. For each cluster found, if its representative point in $D(lev,vol)$ is different from zero we mark the voxels of the cluster as belonging to the mask and decrement the value of this point in the diagram, otherwise the cluster voxels are not considered as belonging to the mask at stage *lev*.

This is formalized in pseudocode in Algorithm 1, where the same reference labels to the above introduced sub-steps has been used and the values of the variables that have been experimentally set for proper mask generation are showed.

Once obtained the mask in spatial domain, we derive the one in the wavelet domain following the subband decimation. Moreover, subband masks are dilated at each wavelet decomposition level, with one or two steps of dilation, depending on the length of the used filters. All solutions to manage a gradual quality decay between VoI and BG are possible (e.g. those described in [4, 8]). As already stated there is also the possibility to complete the VoI mask with explicit indications of nuclear medicine physicians or radiologists. It is worth noting that intermodal information exploitation cannot be applied viceversa, i.e. to empower VoI estimation. In fact, despite the concrete problems in applying this principle in a selective coding setting (which would require closed loop solutions), VoI detection in PET data cannot be reliably steered by an eventual CT data analysis (e.g. segmentation) because organ boundaries do not necessarily correspond with pathologic borderlines.

Algorithm 1 Mask generation algorithm

Require: $D(lev, vol)$
Require: $PETVOL$ {Pet dataset}
Ensure: $MASKVOL$ {Mask dataset}

    Function: *FindNeigh(lev,vol)* Decrements $D(lev, vol)$ then sets *lev* and *vol* according to the position of the nearest neighbor (i, j) of the considered point with $D(i, j)$ different from zero. Returns **False** if no neighbors are found.
    Function: *GetCluster(vox, s)* Returns a cluster object consisting in all voxel connected to *vox* and above threshold *s*.
    Function: *SetMask(cl, mask, s_val)* For each voxel of the cluster *cl*, set the corresponding voxel in *mask* to the saliency value *s_val*.
    Define: $LEV\_CHANNEL = 100$
    Define: $SMALL\_CLUSTER\_LIMIT = 4$
    Define: $TOO\_MANY\_CLUSTERS = 10$
    Define: $MIN\_CHANNEL = 5$
    Define: $TOO\_BIG = 50000$
    Local: *current_lev*
    Local: *current_vol*

```
 1: /*E-a + E-b*/
 2: for i = 0 to LEV_CHANNEL do
 3:     D(i, 0) ←0
 4:     for j = 1 to SMALL_CLUSTER_LIMIT do
 5:         if D(i, j) ≥ TOO_MANY_CLUSTERS then
 6:             D(i, j) ←0
 7:         end if
 8:     end for
 9: end for
10: /*E-c*/
11: current_lev ←LEV_CHANNEL
12: current_vol ←0
13: while D(current_lev, current_vol) == 0 do
14:     current_vol + +
15: end while
16: D(current_lev, current_vol) − −
17: repeat
18:     neigh_found ←FindNeigh(current_lev, current_vol)
19: until neigh_found
20: /*Inclusion + E-d*/
21: for w = MIN_CHANNEL to LEV_CHANNEL do
22:     threshold = w · 32767/LEV_CHANNEL
23:     for all voxel ∈ PETVOL do
24:         if PETVOL(voxel) ≥ threshold then
25:             clust ←GetCluster(voxel, threshold)
26:             if D(w, clust.size) > 0 & clust.size ≤ TOO_BIG then
27:                 SetMask(clust, MASKVOL, s_value)
28:             end if
29:         end if
30:     end for
31: end for
```

3.2.2 Embedded Morphological Dilation Coding

The overall coding framework for CT volumes is depicted in Figure 2. The coding engine we selected is an adapted version of the Embedded Morphological Dilation Coding (EMDC) algorithm (with context based adaptive arithmetic coding) [11], extended for 3D datasets [12]. This algorithm demonstrated comparable or slightly superior performance with respect to JPEG2000 and superior performance with respect to other 3D coding schemes [11, 12]. EMDC generates a progressive bitstream by means of the codification of the action of a multiresolution morphological operator on quantization bit-planes which is able to exploit non-linear intra and inter-subband statistical dependencies of wavelet coefficients. We adapted the 3D-wavelet transform and 3D-EMDC to handle data of any spatial di-
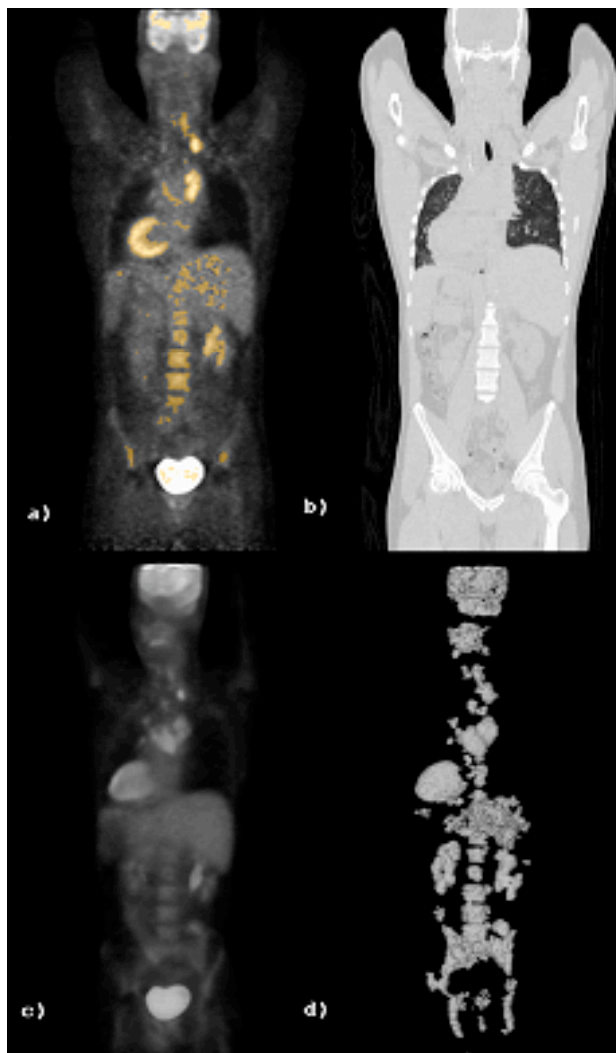
Figure 3: a) PET coronal slice with mask overlay, b) corresponding CT slice, c) PET volume rendering, d) Mask volume rendering

Figure 4: a) PET axial slice with mask overlay, b) CT compressed with CR=30 and mask=8, c) with CR=65 and mask=8

mension. For the transform stage we used symmetric boundary data extension implemented for both odd-length and even-length linear phase wavelet filters and we do not introduce extra-coefficients in the cases of odd length subband splitting [13]. Moreover, EMDC inter-band scanning has been adapted to the uneven subband dimensions generated by the wavelet transform for arbitrary length signals. This gives a high versatility to our 3D transform coding system which can adapt itself to real settings not only in terms of original data dimensions but also with respect to an eventual data partitioning and free choice of wavelet filters in response to computational load issues and/or quality requirements.

4. RESULTS

4.1 Mask evaluation

We considered PET-CT volumes generated in the Brazilian hospital with associated different diagnostic questions (see Table 1). All VoI masks have been generated following Algorithm 1. From an expert point of view the algorithm produced VoI masks presenting a good fit to physiological and most evident pathological uptake regions. Variability in PET volumes in terms of absolute uptake level is high and depends from many factors. Criteria and parameters that guide the automatic VoI extraction have been thought and set to handle this PET variability. Figure 3 is representative of the
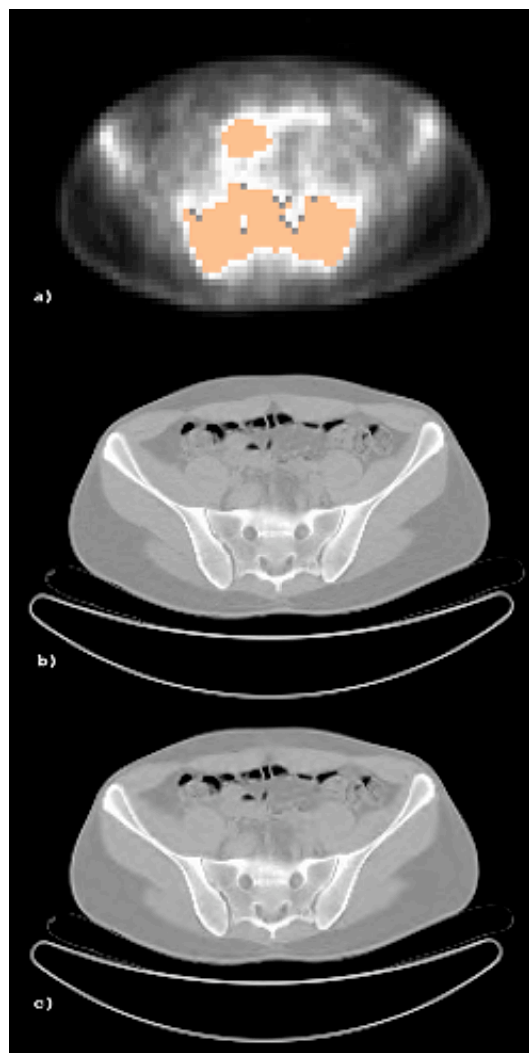
datasets involved and of the mask outcomes (volume rendering has been used to give a global view). It can be observed that uptake noise is not present into the mask. The bladder has been successfully automatically removed in all considered examples. To be suitable for diagnostic quality assurance, the described selective coding method could be completed with a "light" user-interaction (i.e. low user-burden and few bits generated). This may be required in order to clinically validate the generated VoI mask and/or to add regions (e.g. by means of "one-click" interactions) which are "dark" on the PET data (e.g. necrotic tissues) but considered relevant from a clinical point of view and on the base of the a-priori physician knowledge about the diagnostic question. These aspects will be carefully considered in a clinical validation phase of the presented technique.

4.2 Selective coding evaluation

We report representative coding results regarding the compression of a whole-body CT volume (404 slices) at different coding ratio and different saliency mask levels. Figure 5 shows PSNR results at various compression ratio (CR) and various VoI mask values (2,4 and 8). Meaningful quality differences can be seen between the VoI region and the whole volume with marginal variations of global PSNR values at various mask levels (as in this case, a typical VoI contains about 0.5% of the volume voxels). Less familiar for coding experts are the obtained absolute PSNR values because they are
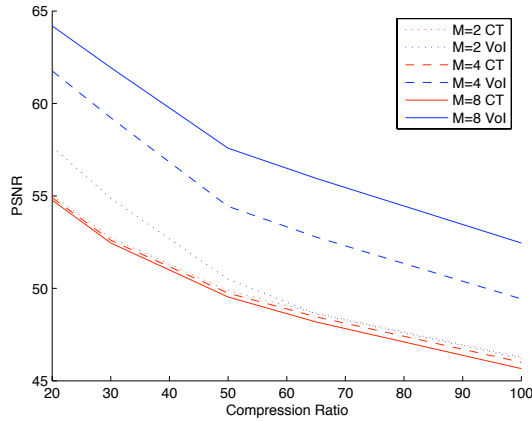
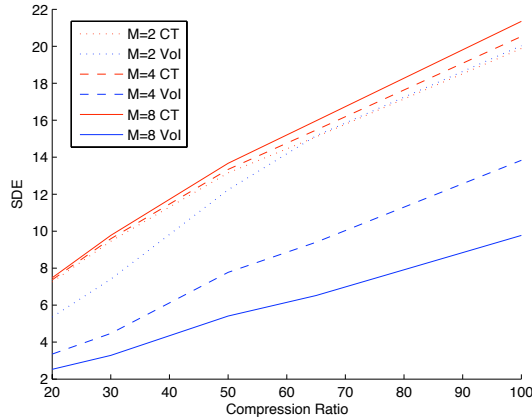Figure 5: CT (blue) and VoI CT (red) PSNR values at various CR

Figure 6: CT (blue) and VoI CT (red) SDE values at various CR

calculated with respect to the nominal CT range of 2^{12} levels, while mean square error (MSE) measures depends on the actual range. For these reasons values are higher than typical ones for generic images. To give a complementary quantitative idea of coding errors we show, in Figure 6, corresponding outcomes in terms standard deviation error (SDE). A SDE value of 10 corresponds approximately to 1/2500 of the actual range of CT values and therefore it can be thought as about 1 level deviation for a 256-graylevel image.

In Figure 4 we give a visual coding example. We show the reference PET slice with the mask overlay and the corresponding CT slice compressed at two different CR with the same VoI mask value. Visual quality difference can be seen only outside the VoI regions. What is important to say is that, as a consequence of the correct application of the proposed method, the two VoI-coded volumes present the same diagnostic quality. A clinical validation will allow to establish, with respect to additional clinical requirements and transmission channel performance measurements, more defined coding settings in terms of suitable mask values and CRs.

5. CONCLUSIONS

In this work we proposed a novel approach to selective PET-CT image coding problems which guarantees highly efficient CT coding/decoding tools preserving image regions on CT volumes to be visually inspected by clinicians during PET-CT oncological re-

mote reporting (second-opinion), according to the specific diagnostic question. The proposed solution, which can be used both in an automatic and in a semiautomatic way, shows potential to be applied in all inter-modal imaging clinical environments where telemedicine or archiving services ask for high coding ratios, diagnostic quality assurance and minimization of user burden.

6. ACKNOWLEDGMENTS

REFERENCES

[1] ISO/ISC JTC 1/SC 29/WG 1 (ITU-T SG8), *JPEG2000 Part I: Core coding system* Number 15444-1, 2002.

[2] P.G. Tahoces *et al.* "Image compression: Maxshift ROI encoding options in JPEG2000," *Computer Vision and Image Understanding* vol. 109, pp. 139-145, Jan.2008.

[3] S.Li and W.Li, "Shape-adaptive discrete wavelet transform for arbitrary shaped visual object coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 10, pp.725–743, 2000.

[4] A. Signoroni and R. Leonardi, "Progressive Medical Image Compression using a Diagnostic Quality Measure on RegionsofInterest " in *EUSIPCO98*, pp.23252328, Rodi (GR), Sept. 1998.

[5] A. Signoroni, F. Lazzaroni and R..Leonardi, "Exploitation and extension of the Region-of-Interest coding functionalities in JPEG2000," *IEEE Transactions on Consumer Electronics*, vol. 49, n. 4, pp.818-823, Nov. 2003.

[6] ISO/ISC JTC 1/SC 29/WG 1 (ITU-T SG8) *JPEG2000 Part II: Extensions*, Number 15444-2, Final Publication Draft, 2002.

[7] P. Schelkens *et al.*, "Wavelet Coding of Volumetric Medical Datasets," *IEEE Trans. on Medical Imaging*, vol. 22, n. 3, Mar. 2003.

[8] A. Signoroni, F. Lazzaroni and R. Leonardi, "Selective coding with controlled quality decay for 2D and 3D images in a JPEG2000 framework," in *Proc. Visual Communication and Image Processing, VCIP 2003*, SPIE vol. 5150, pp.830-841, Lugano, CH, July 2003.

[9] I. Ueno and W. A. Pearlman, "Region of Interest Coding in Volumetric Images with Shape-Adaptive Wavelet Transform," in *Proc. SPIE/IS&T Electronic Imaging 2003*, SPIE vol. 5022, January 2003.

[10] T. Beyer, C. C. Watson *et al*, "The Biograph: A Premium Dual-Modality PET/CT Tomograph for Clinical Oncology," *Electromedica* vol.69, n.2, pp.120–126, 2001.

[11] F. Lazzaroni, R. Leonardi and A. Signoroni, "High-performance embedded morphological wavelet coding," *IEEE Signal Processing Letters*, vol.10, n.10, pp. 293–295, Oct. 2003.

[12] F. Lazzaroni, A. Signoroni and R. Leonardi, "Embedded Morphological Dilation Coding for 2D and 3D Images," in *Visual Communication and Image Processing, VCIP 2002*, SPIE vol. 4671, pp.923934, San José, CA, USA, Jan. 2002.

[13] Barnard, H. J., "Image and video coding using a wavelet decomposition," *Ph.D. Thesis* Technische Univ., Delft (Netherlands), 1994.