

MULTIPLE, ARBITRARY SHAPE ROI CODING WITH ZEROTREE BASED WAVELET CODERS

Aysegul Cuhadar and Sinan Tasdoken

Department of Systems and Computer Engineering
Carleton University, 1125 Colonel By Drive,
Ottawa, ON, K1S 5B6, Canada
{cuhadar,tasdoken}@sce.carleton.ca

ABSTRACT

Region-based video coding schemes employed in MPEG-4 is also promising for still image coding applications where images contain a number of objects that can be encoded at different bit rates, such as compression of medical images for archiving and transmission. Motivated by this fact, in this paper we investigate multi-region and multi-quality (MRMQ) coding based on zerotree wavelet coders. We present a novel scheme which addresses the region size sensitivity problem in region-based coding. The proposed method outperforms the region of interest (ROI) coding unit of JPEG-2000, i.e., it is possible to save 0.3 bits per pixel to attain the same ROI/background rate-distortion performance with the proposed MRMQ coding scheme.

1. INTRODUCTION

Region-based coding is one of the most important features provided by MPEG-4 and JPEG-2000. It allows imposing heterogeneous (or region-dependent) fidelity constraints rather than encoding the entire image with a single fidelity constraint. For image coding applications where images contain a number of objects that can be encoded at different bitrates, such as compression of medical image data for archiving and transmission, it offers better utilization of available bit rate since high fidelity shall be maintained only for relatively small regions rather than for the entire image.

In multiple ROI image coding, it is essential to represent arbitrary shape regions efficiently. MPEG-4 [1] uses a block discrete cosine transform (DCT) based method for region representation due to its computational efficiency, however this representation suffers from annoying blocking artifacts that limits its use for low bit rate coding. As an alternative, a wavelet based method, namely shape adaptive discrete wavelet transform (SA-DWT) is introduced in [2]. This technique offers superior rate-distortion performance and better visual quality than MPEG-4's DCT based mechanism. SA-DWT retains most of the features of conven-

tional DWT, including multiresolution property and locality. Thus, region-based extensions of compression efficient wavelet coefficient coders, EZW and SPIHT, are successfully employed for coding arbitrary-shape regions [2]. In JPEG-2000 [3], heterogeneous fidelity coding problem is examined from a different point of view, that is, instead of using SA-DWT, conventional DWT is used. The transform coefficients required for reconstruction of various regions is differentiated in the transform domain, and heterogeneous fidelity is obtained by appropriately partitioning the available bit rate among transform coefficients, i.e., allocating more bits to transform coefficients that are required for the reconstruction of a region at high fidelity and less bits for the remaining coefficients. This mechanism provides *interactive* encoding utility, however it is highly inferior to the other schemes that employs SA-DWT in terms of rate-distortion (R-D) performance.

In this work, we use SA-DWT for region representation due to its superior representation efficiency. Our aim is to address two important problems observed in encoding of arbitrary shape region representations. 1) For communication over distributed networks, it is beneficial to transmit a scalable bit stream, i.e., the encoding mechanism should provide resolution, distortion and computational scalability. Even though JPEG-2000 encompasses the first two properties, the coding mechanism does not guarantee computational scalability (due to tier 2 of the coding mechanism) for the encoder. Therefore, we adopt SPIHT algorithm for MRMQ coding since it exhibits comparable R-D performance to JPEG-2000 without sacrificing aforementioned properties. 2) Tree based wavelet coders are found to be not very efficient for encoding regions of relatively small size due to use of exponentially growing trees. We address this problem by introducing a novel scheme based on *unbalanced zerotrees*. The proposed techniques outperform the region of interest (ROI) coding unit of JPEG-2000 [3] and the SPIHT-ROI coder [4] in terms of rate-distortion performance. In addition, they provide a more general multi-region multi-quality (MRMQ) coding framework rather than ROI/non-

ROI coding.

The rest of the paper is organized as follows: In Section 2, we present an overview of the MRMQ coding. In Section 3 incorporation of MRMQ coding functionality to zerotree-based wavelet coders is given. Rate-distortion performance of the proposed technique is provided in Section 4. The paper concludes with Section 5.

2. MULTIPLE ROI CODING

The image plane, denoted by Ω , is a set of lattice points over which the image function is defined. Within the transform-based coding framework, the image function is transformed into a more sparse (or energy localized) function defined over the transform coefficient plane Λ . In MRMQ coding, the image plane Ω is partitioned into N disjoint subsets, Ω_i $i = 0, \dots, N - 1$. The union of these subsets gives the entire image plane, Ω . During the encoding process, for each region, a separate distortion criteria is used, i.e.

$$\mathcal{D}^{(\Omega_i)} \leq \mathcal{D}_i, \quad 0 \leq i < N$$

To achieve high compression, one needs a sparse (or highly structured) representation for each region. This can be accomplished by using *region-based transforms*. The region-based transform maps the image samples of each region Ω_i into a set of transform coefficients defined over Λ_i , where $\Lambda_i \subseteq \Lambda$. Note that, although region-based transform provides a maximally-decimated representation of the region samples, it is possible that

$$\Lambda_i \cap \Lambda_j \neq \emptyset, \text{ for } i \neq j$$

In other words, critically sampled representations of completely separated regions does not guarantee complete separation in transform domain. However, by choosing appropriate transforms, it is possible to make subset Λ_i 's mutually disjoint with the property

$$\bigcup_{i=0}^{N-1} \Lambda_i = \Lambda \quad (1)$$

The region-based transforms that satisfy this property have two important features. First, conventional transform coefficient coders are usually defined to encode a rectangular grid of coefficients and their modification to region-based coding usually starts with encapsulating the arbitrarily shaped subset Ω_i to a bounding rectangular frame which is followed by encoding of corresponding region representation, one region at a time. The restriction imposed by equation (1) precludes adaptation of coding mechanism to varying bounding frame sizes and facilitates the use of the conventional transform coefficient plane coding mechanism with small modifications only. Second, it provides flexibility in the efficient

interleaving of bit streams produced for different regions, without the need of a post-processing component. The decomposition strategy proposed in [2] is used in this work, as it exhibits both of these features.

3. ZEROTREE BASED MRMQ CODING

In the following sections, we present MRMQ coding based on the SPIHT algorithm. Application of proposed methods to other zerotree-based wavelet coders is straightforward. Hereafter, the definitions and notations developed in [5] will be used.

3.1. MRMQ-SPIHT with Set of Descendent Labels

The method proposed in this section is inspired by the work of Park et al. [4] where ROI coding capability is incorporated into SPIHT encoder by partitioning the wavelet coefficient plane into two subsets, namely the subset of wavelet coefficients (referred as the set of ROI coefficients) required for the reconstruction of selected ROI samples and the complement of this subset. The problem with this approach is that number of ROI coefficients used to reconstruct the ROI is usually greater than number of samples within the ROI. For example, in [4], one needs to encode 121 wavelet coefficients to describe a 4-by-4 ROI. This non-decimated description of ROI degrades the compression performance substantially. Furthermore, as the ROI representation is over-complete, it is not possible to achieve a precise control on the reconstruction quality of the regions. Thus, it is more beneficial to use critically sampled wavelet transforms to obtain disjoint representations for all regions, which results in better rate-distortion performance. In addition, we propose multi-region coding functionality rather than a ROI/non-ROI coding functionality, which provides a precise control over region fidelities and enables concurrent coding of several regions if required.

In zerotree coding, the tree node with index i has one-to-one correspondence with the wavelet coefficient with index i . Let $v(i)$ be the label of this coefficient (and the corresponding node), i.e., if $i \in \Lambda_l$, then $v(i) = l$. When the wavelet representations of arbitrary shape regions are coded, we encounter different label nodes within the same zerotree structure. Consider encoding of the transform coefficients pertaining to region l in this case. An implementation (such as the one proposed in [6]) can be obtained by partitioning all zerotrees as in conventional case and sending information only about the nodes associated with wavelet coefficients of region l . However, we can save more bits since there is no need to transmit the partitioning information for (sub-)trees which do not have any descendent nodes pointing to wavelet coefficients of region l . To describe the proposed scheme, it is instructive to introduce the notion of

sets of descendent labels, which is defined as:

$$\Upsilon(i) = \bigcup_{k \in \mathcal{O}(i)} \{v(k) \cup \Upsilon(k)\} \quad (2)$$

where $\mathcal{O}(i)$ denotes the direct descendent (offspring or children) nodes spreading from node i , as defined in [5]. Recall from SPIHT algorithm that, list of insignificant pixels (LIP) and list of insignificant sets (LIS) are initialized by using the set of wavelet coefficients within the lowest resolution subband, denoted by \mathcal{H} . However, when region l is to be encoded, these lists should be initialized by using

$$\mathcal{H}_l^{LIP} = \{i \mid i \in (\mathcal{H} \cap \Lambda_l)\} \quad (3)$$

$$\mathcal{H}_l^{LIS} = \{i \mid l \in \Upsilon(i), i \in \mathcal{H}\} \quad (4)$$

where LIP is initialized by using the coordinates, $i \in \mathcal{H}_l^{LIP}$ and LIS is initialized by the zerotree roots with indices, $i \in \mathcal{H}_l^{LIS}$. Another modification to original SPIHT algorithm occurs at the set partitioning stage, i.e., only the zerotrees (represented by root nodes i) for which $l \in \Upsilon(i)$ are partitioned and no bits are transmitted regarding the partitioning information of remaining trees. An important feature of the scheme is that, it is relatively straightforward to extend the algorithm to encode multiple regions at a desired quality altogether, which might be beneficial if complete independence of region bitstreams is not required and a context based arithmetic coder is used to exploit the dependencies among neighboring wavelet coefficients.

MRMQ-SPIHT algorithm described above is not very efficient when relatively small size regions are to be encoded. This problem is discussed in the following section and a low-complexity solution is proposed by introducing the concept of *unbalanced zerotrees*.

3.2. MRMQ-SPIHT with Unbalanced Zerotrees

To further improve the performance of MRMQ-SPIHT algorithm described above, we provide a solution based on the following observation: 1) When a region of relatively small size is encoded, the region's wavelet coefficients are localized over a very small portion of the zerotree. In this case, if we enable the use of nodes from lower levels of zerotrees, i.e., nodes other than roots of zerotrees, upon LIS initialization, we can save the redundant bits transmitted to reach the sub-tree comprising the wavelet coefficients of the region from the root of the zerotree. 2) If we can establish homogeneity in tree structures, i.e., if all nodes of the tree have a common label, then we do not need bookkeeping of sets descendent labels during encoding. Based on these observations, we propose *unbalanced zerotrees* which will be described next.

Assume that the zerotrees are built as quadtrees as described in the original SPIHT algorithm. This construction

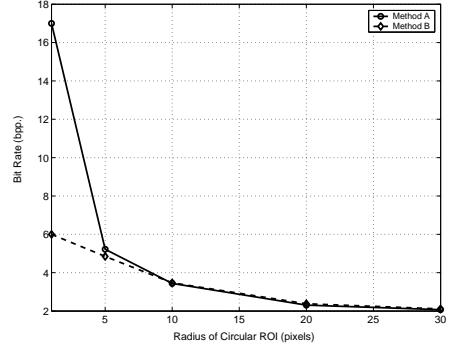


Fig. 1. Comparison of region size sensitivity of proposed methods.

results in mismatch between labels of some nodes and the labels of their parent nodes. To obtain homogeneity in the tree structure, the following tree reconstruction scheme can be used: If the label of a node in a tree does not agree with the label of its parent, the parent node is said to be pseudo-parent, and in this case, the node is disconnected from its pseudo-parent and a new (real) parent is sought for this node. The real parent is defined as one of the eight-connected neighbors of the pseudo-parent and has the same label as the corresponding node. If such a parent node is found, the node is connected to its real parent, otherwise the node itself is used as a zerotree root upon LIS initialization. The proposed tree structure moves the introduced complexity to the initialization stage, no longer requires bookkeeping of the sets of descendent labels and the comparisons about the sets of descendent labels. Furthermore, it performs well when regions of relatively small sizes are to be encoded.

4. EXPERIMENTAL RESULTS

In this section, the performance of MRMQ coders is discussed. For brevity, MRMQ-SPIHT with Set of Descendent Labels will be referred as Method A and RMQ-SPIHT with Unbalanced Zerotrees will be referred as Method B. The 9/7 biorthogonal wavelet basis is used in the experiments.

In the first part of the experiments, compression performance of the two methods is compared against varying region sizes. The 512-by-512 *Angio* image is used in this part. A circular region, centered at the center pixel of the *Angio* image is selected and the objective PSNR value for the region is fixed at 40 dB. The corresponding bit rates for this PSNR value is measured for various radius values. The results are provided in Fig. 1. As can be seen in the figure, Method B outperforms Method A for small region sizes and they perform almost equally as the region size grows. This result favors the use of Method B when the image is to be encoded as multiple regions of relatively small sizes.

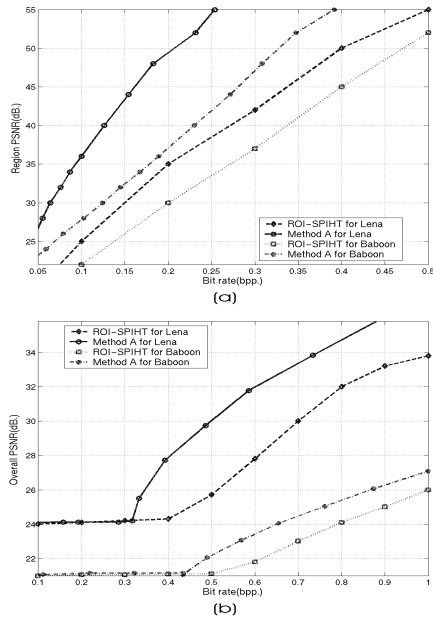


Fig. 2. Compression performance comparison of ROI-SPIHT and Method A.

In the second part of the experiments, the performance of MRMQ-coders is compared with the ROI-SPIHT in [4] and the ROI coding unit of JPEG-2000 [3]. First consider ROI-SPIHT of Park et. al. [4]. This coder is taken as reference since it outperforms all previously proposed SPIHT ROI-coding schemes in terms of R-D performance. The 256-by-256 *Lena* image with 51-by-51 ROI and 512-by-512 *Baboon* image with 101-by-151 ROI are used for comparison. the R-D curves is provided in Figure 2.(a). In terms of ROI-PSNR, Method A performs 10 dB.(/7 dB.) better than ROI-SPIHT when ROI of *Lena*/(*Baboon*) image is encoded. This is due to use of maximally decimated wavelet representation for the ROI rather than use of wavelet coefficients required for reconstruction of the ROI, since the number of wavelet coefficients required for reconstruction of the ROI is usually far greater than number of samples within the ROI. Hence, as can be seen in Figure 2.(b), it is possible to achieve 0.2/(0.15) bpp. gain in the overall bit rate to attain the same ROI/background quality even though the area of ROI is only 4%/(5.8%) of the image area.

The performance of the MRMQ-coder is also compared with the ROI encoder of JPEG-2000. The 720-by-576 *Goldhill* image with a 320-by-200 ROI is used for comparison. JPEG-2000 do not provide a precise control mechanism for the distortions in the ROI/non-ROI and the discrimination of the quality of ROI/non-ROI is achieved by simply down-scaling wavelet coefficients pertaining to the background. The ROI coding performance of JPEG-2000 is taken from [3] for scale = 11. In terms of ROI rate-distortion perfor-

mance, Method A performs, on average 4 dB. better than JPEG-2000, when ROI of *Goldhill* image is encoded. Hence, Method A provides a 0.3 bpp. gain in the overall bit rate to attain the same ROI/non-ROI quality when the ROI area is 15% of the image area.

5. CONCLUSION

In this paper, MRMQ coding based on zerotree-wavelet coding algorithm is presented. The wavelet representations of arbitrary shape regions are encoded using a modified version of the SPIHT algorithm. Region size sensitivity problem of this tree-based coder is addressed by introducing the concept of *unbalanced zerotrees*. Extensive experimental results show that the proposed coders perform significantly better than region coding schemes of JPEG-2000 and ROI-SPIHT, which makes them suitable for region-based image compression applications.

6. REFERENCES

- [1] Andre Kaup, "Object-based texture coding for moving video in MPEG-4," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 5–15, February 1999.
- [2] Shipeng Li and Weiping Li, "Shape-adaptive discrete wavelet transforms for arbitrarily shaped visual object coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 10, no. 5, pp. 725–743, August 2000.
- [3] Raphaël Grosbois, Diego Santa-Cruz, and Touradj Ebrahimi, "New approach to JPEG 2000 compliant region of interest coding," in *SPIE's 45th annual meeting, Applications of Digital Image Processing XXIV*, San Diego, California, Aug. 2001, vol. 4472 of *Proc. of SPIE*, pp. 267–275.
- [4] Keun-Hyeong Park and Hyun Wook Park, "Region-of-interest coding based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 2, pp. 106–113, February 2002.
- [5] Amir Said and William A. Pearlman, "A new, fast, and efficient image codec based on set partitioning in hierarchical trees," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–249, June 1996.
- [6] Eiji Atsumi and Nariman Farvardin, "Lossy/lossless region-of-interest image coding based on set partitioning in hierarchical trees," in *IEEE International Conference on Image Processing*, 1998, vol. 1, pp. 87–91.