Multiple Description Image Coding Using Regions of Interest

Marcus Nyström*, Jerry D. Gibson† and John B. Anderson*
*Dept. of Electrical and Information Technology, Lund University, Sweden
†Dept. of Electrical and Computer Engineering, University of California, Santa Barbara
Email: marcus@eit.lth.se, gibson@ece.ucsb.edu, john_b.anderson@eit.lth.se

***Abstract*— A multiple description (MD) coding scheme encodes an image such that each description alone yields an acceptable reconstruction quality, and the descriptions together render increasingly better approximations of the image. Traditionally, each description contributes equally to the reconstruction quality throughout the image. This is inefficient from a coding perspective since all descriptions largely represent similar source information. We suggest a novel region of interest (ROI) based MD coding scheme with two descriptions and two ROIs, such that each ROI is coded into a separate description. The impact this scheme has on visual reconstruction quality and coding efficiency is investigated. We argue that ROI based MD coding is beneficial at low bit rates since it *(i)* introduces little redundancy between the descriptions with practically no loss in quality at the central decoder, and *(ii)* still produces useful side descriptions. Extensions to video coding are discussed in the concluding section.**

I. Introduction

When transmitting information over packet-switched, heterogeneous networks such as the Internet, a certain quality of service cannot be guaranteed; data packets may be lost, corrupted or delayed and hence degrade or render the transmitted source useless. Errors can occur due to a variety of factors such as noise, congestion caused by a sudden drop in bandwidth, or when retransmissions are infeasible due to real-time constraints, to name a few examples. One of the methods addressing the transmission of coded information over error-prone networks is *multiple description (MD) coding* [1], [2]. MD coding strives to alleviate the effects of packet losses or errors by making best possible use of the (partial) source information received at the decoder.

In MD coding, a source is encoded into a number (say n) of representations, usually such that each description carries equally important information about the source. The goals of MD coding are to achieve an acceptable reconstruction quality from one or a subset of the descriptions, and that the quality should improve as more descriptions are available to decoder. Thus, it is possible to obtain useful representations of the source despite large error rates yielding significant description losses. However, since the resilience against a lost description relies on a certain level of inter-description redundancy, MD coding carries a penalty in terms of reduced coding efficiency. For an more comprehensive introduction to MD coding, we refer to the overview paper by Goyal [2].

In this paper, we propose an approach that efficiently implements MD coding for images; ROIs are utilized to create spatially diverse descriptions such that each description carries substantially different source information. A progressive scheme is used to encode the descriptions such that interesting regions are coded and transmitted a priori. Section II introduces and motivates the proposed method, and also describes how the descriptions are encoded, synthesized and decoded, given the ROIs. Section III presents simulation results and shows a number of images coded with the proposed ROI based method for MD coding. Section IV concludes the paper.

II. MD Coding With Two ROIs

A. Overview of the suggested approach

A fundamental trade-off in MD coding is to generate source descriptions that are not too similar, and yet be able to reconstruct an acceptable estimate of the source from one or a fraction of the descriptions. If the descriptions are too similar, they are inefficient from a coding perspective since bits are wasted in representing essentially the same data. If on the other hand the descriptions carry substantially different information about the source, each description might not suffice to reconstruct the source with acceptable quality. For example, consider the case of two descriptions $n = 2$ with a total rate $R = R_1 + R_2$, where R_1 and R_2 denote the rates of descriptions 1 and 2, respectively. If R_1 and R_2 are sufficient to yield two individually good representations, then probably the bits could have been used more wisely when both descriptions are received. On the other hand, a source acceptably represented with R bits may be difficult to split in two different descriptions of rates R_1 and R_2 such that each description is of acceptable quality. To address this fundamental trade-off, we introduce a novel, region of interest (ROI) based method for MD coding with two ROIs and $n = 2$ descriptions, such that each description accommodates one ROI. We consider the most general case when the total rate R is equally divided between the two descriptions, i.e., $R = R_1 + R_2$ where $R_1 = R_2 = R/2$. This method is

This work was done while the first author was visiting the ViVoNets Lab at the University of California, Santa Barbara.

The second author is supported by the California Micro Program, Applied Signal Technology, Dolby Labs, Inc., Mindspeed, and Qualcomm, Inc., by NSF Grant Nos. CCF-0429884 and CNS-0435527, and by the UC Discovery Grant Program and Nokia, Inc.

(a) (b)

(c) (d)

Fig. 1. *Example of ROI based MD coding.*(a) Original image with ROIs marked by rectangles. (b) Reconstruction at first side decoder ($R_1 = 0.025$ bpp). (c) Reconstruction at second side decoder ($R_2 = 0.025$ bpp). (d) Reconstruction at central decoder ($R = R_1 + R_2$ bpp).

illustrated in Fig. 1; 1(a) shows the original image with the two ROIs (ROI #1 to the left) indicated by boxes. Figs. 1(b), 1(c), 1(d) depict, respectively, side description #1, side description #2, and the central (merged) description after reconstruction. A detailed description of how these images were generated will be given in Section II-B.

Using ROI based MD coding, consider the naive yet illustrative case where R_1 is allocated entirely to the ROI in description 1 and R_2 entirely to the ROI in description 2. Clearly, the descriptions carry no information about each other and are hence comparable to a single description scheme in terms of coding efficiency for the joint rate case. In addition, the descriptions are extremely simple to synthesize at the decoder. However, if only one description, and thus only one ROI is received, the other ROI is completely lost, which is unacceptable.

In a more realistic scenario, the ROIs are coded a priori with a majority of the bit budget, and the non-ROIs are allocated remaining bits. This way, we keep the advantage of having two largely uncorrelated descriptions (thus adding little excess joint rate), and at the same time having acceptable non-ROI quality. If only one description is received in this scenario, one of the ROIs is still available in good quality. Consequently, a single description can render "useful" descriptions in terms of interpretability or recognition where standard MD coding methods would fail to do this at similar (low) bit rates.

Clearly, the inter-description similarity increases as a larger fraction of the bit budget is spent outside of the ROIs. We will in the next section present numerical and visual results on how the trade-off in allocating bits between ROI/non-ROI regions affects the reconstruction quality.

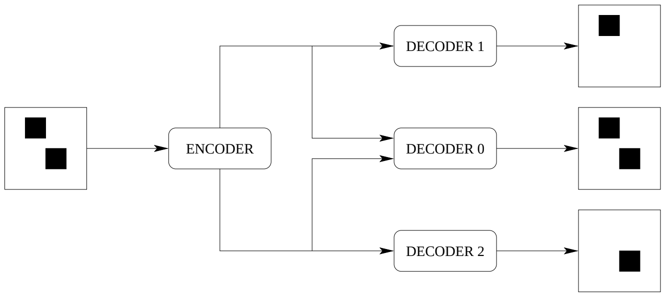


Fig. 2. *Overview of the proposed ROI based scheme for MD coding.* The black boxes represent ROIs. Each ROI is placed in a separate description, which is coded at $R/2$.

B. Implementation based on SPIHT

Due to its low complexity, good rate distortion trade-off, and its inherent capabilities for progressive transmission, we choose to implement the ROI based MD coding scheme with the set partitioning in hierarchical trees (SPIHT) algorithm [3], modified to suit our application. SPIHT is a coding algorithm that exploits structural similarities between different subbands in a wavelet transformed image through efficient quantization and progressive refinement of the wavelet coefficients. The structural similarities are used to derive hierarchical trees in a parent-child oriented structure, such that coefficients in the lower frequency subbands (parents) derive the significance for coefficients in the higher frequency subbands (children and descendants).

Each run n in the SPIHT algorithm contains a *sorting pass* and a *refinement pass*. The sorting pass utilizes three different lists: List of insignificant coefficients (LIC), list of insignificant sets (LIS), and list of significant coefficients (LSC). The coefficients in LIC are labeled as significant or insignificant depending on if it exceeds a threshold value, T_n, that is reduced by two for each run. If a coefficient in LIC is significant it is passed to the LSC, otherwise it remains in LIC. LIS contains coefficients whose descendants are checked for significance. Depending on their significance, they are either moved to LSC or LIC to be processed in the next run. In the refinement pass, significant coefficients in LSC are updated increasingly more fine-grained as the threshold decreases. One of the key features that makes SPIHT attractive is that the generated bit stream is completely embedded; an increasingly improved version of the image can be obtained from a single bit stream by truncating the bit stream at later and later points. This yields an exact rate control, and an image can be retrieved at any resolution until it becomes nearly lossless. The algorithm performs optimally in the sense that it guarantees the best possible reconstruction, measured in mean squared error, at any truncated bit rate. Decoding is straight forward, where the decoder follows the same basic structure as the encoder.

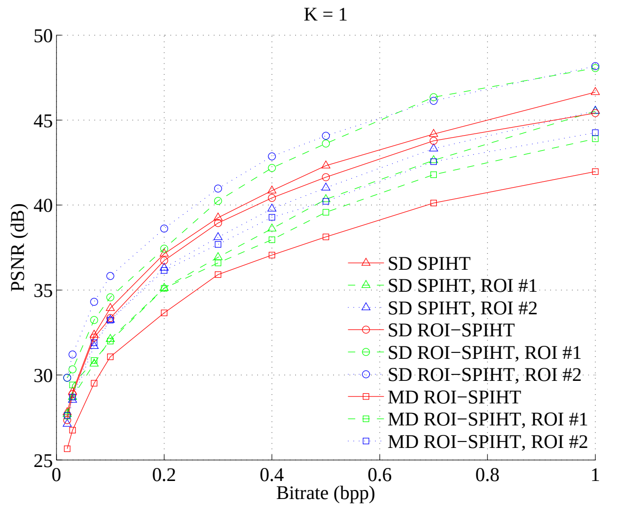
An overview of the ROI based MD coding system is schematically depicted in Fig. 2. Two ROIs, illustrated by black boxes in the figure, are defined for an input image, each encoded into a separate description with a modified SPIHT algorithm. If only one of the descriptions is received, it is directly decoded by the modified SPIHT decoder (DECODER 1 or 2). In case both descriptions are available to the decoder, they are first synthesized and then decoded by the central decoder (DECODER 0). Since it is simple and fast, although not necessarily optimal, we perform the synthesis in the wavelet domain by choosing the coefficients from the ROIs in the first and second descriptions, and otherwise the average coefficient values from descriptions one and two.

The implementation for ROI prioritization presented in this paper largely follows the philosophy of the *general scaling based method*, as defined in the JPEG2000 image coding standard [4]. With this method, after the wavelet transform has been applied to the input image, coefficients residing outside the ROIs are downscaled causing ROI coefficients to be coded a priori. Before inverse transformation, the decoded non-ROI coefficients are upscaled by the same factor. We use a similar strategy by multiplying non-ROI coefficients with factors $2^{-\mathrm{K}}$ and $2^{\mathrm{K}}, \mathrm{K} \in \mathbb{N}$ prior to SPIHT encoding and wavelet inverse transformation, respectively. Consequently, the choice of K controls the degree to which an ROI is prioritized and thus how fast the ROI is reconstructed at the expense of non-ROIs.

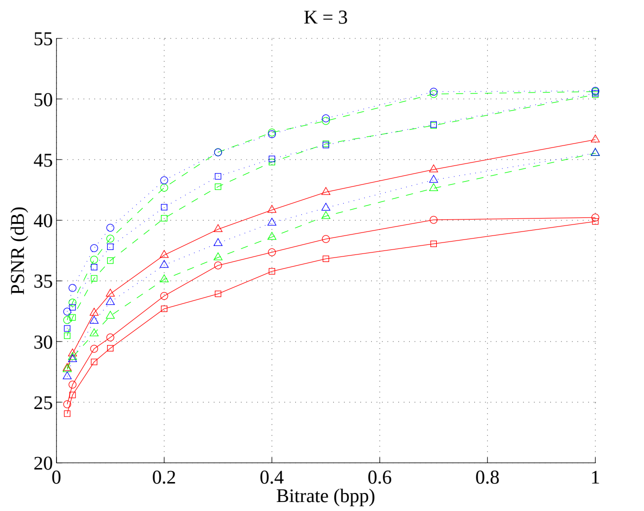
Since SPIHT in its original setting is not designed for ROI coding, we implement two algorithmic modifications motivated by the following. *(i)* SPIHT tests all coefficients in LIC and all descendants of LIS for significance against the threshold T_n. However, since the non-ROI coefficients are downscaled, most of them will with a high probability be insignificant during the initial runs. Since the decoder needs to be informed about the insignificance, several bits are wasted in coding "known" information. To resolve this problem, we do not test non-ROI coefficients for significance during the K first runs, given they have been downscaled by a factor 2^{K}. *(ii)* Since downscaling changes the dynamic range, i.e., the ratio between the largest and smallest coefficient value, the number of runs is expected to increase at any given bit rate. As a result, so will the number of refinement bits. Refinement bits generated when the ROI reconstruction quality is already high have shown to give negligible improvement in visual quality [5]. In this paper, we therefore use a maximum of 10 refinement bits for each significant ROI coefficient. Due to the loss in precision associated with downscaling/upscaling of coefficients outside of the ROIs, the number of refinement bits for non-ROI coefficients can be reduced even further. Given a scaling factor 2^{K}, simulations have shown that using only $10-\mathrm{K}$ refinement bits for each coefficient essentially does not affect the reconstruction quality.

III. RESULTS

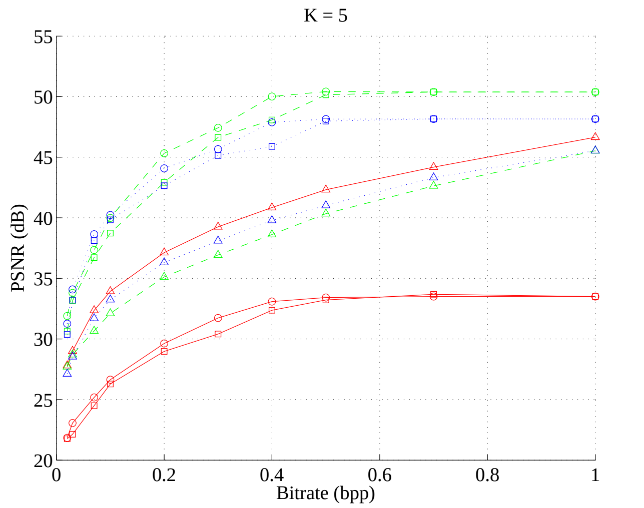
The ROI based MDC scheme is tested on a gray scale, 8-bit image with two obvious ROIs, each containing a facial region as depicted in Fig. 1(a). The test image comprises the first frame of the *Mother and Daughter* test sequence, which for convenience is resized to dimension 512×512 by nearest-neighbor interpolation. The test image is transformed with a



(a)



(b)



(c)

Fig. 3. *Simulation results for three different levels of ROI prioritization K.*(a)K = 1 (b) K = 3(c) K = 5

six level dyadic wavelet transform using the 9/7 bi-orthogonal filter [6] and symmetric border extension. To ensure that a very crude representation of the background is generated even at very low bit rates, coefficients in the lowest frequency subband are not affected by the the ROI scaling.

Simulation results are presented in Fig. 3 for three different levels of ROI prioritization, K $= \{1, 3, 5\}$, and bit rates ranging from 0.02 bpp to 1.00 bpp. Curves annotated single description (SD) SPIHT show rate distortion (RD) results for standard SPIHT coding. Distortions are measured by peak signal-to-noise ratios (PSNR) in decibels (dB), and are given for the whole image as well as within the ROIs (ROI #1 and ROI #2). A similar set of RD curves are given for SD ROI-SPIHT, where coefficients residing in *both* ROIs, along with the downscaled non-ROI coefficients, are coded in a single description with the modified SPIHT algorithm. Thus, SD ROI-SPIHT represents the best possible performance that the proposed ROI based MD coding scheme can achieve for the joint rate case. Finally, MD ROI-SPIHT represents RD curves obtained at the output of the central decoder (DECODER 0).

At low bit rates, the quality in the ROIs are improved due to ROI based SPIHT coding at the expense of the overall reconstruction quality, both when using single and multiple descriptions and for all choices of K. Fig. 3(a) shows RD results for a moderate level of ROI prioritization, K $= 1$. Since the ROI coefficients are given only little priority, and therefore two largely similar descriptions coded at $R/2$ are merged into one description, the excess joint rate is large. Moreover, the ROI quality in the side descriptions is unacceptably poor at very low bit rates. When K is increased to K $= 3$, as in Fig. 3(b), the RD results for ROI based MD coding change drastically; the ROIs are reconstructed with good quality already at low bit rates, and since only little of the background is included in the first part of the bit stream, the inefficiency of using ROI based MD coding compared to ROI based SD coding is small up to a rate of about 0.2 bpp. This tendency is even more clear in Fig. 3(c), where level of prioritization is increased to K $= 5$. Again the ROI based MD coding approach renders high ROI quality at very low bit rates, and the ROI quality increases rapidly up to 0.1 bpp. At this level, the penalty of using ROI based MD coding is small compared to ROI based SD coding even at higher bit rates. This is because at higher bit rates, the $10 - K$ refinement bits may have been spent already or the threshold is decreased to such a low level that refinements are too small to yield noticeable improvements.

In order to assess how ROI based MD coding affects visual quality, Fig. 4 presents the impact different prioritization levels have on the test image decoded at the central decoder for three illustrative bit rates, $R = \{0.02, 0.07, 0.50\}$ bpp. The figure clearly illustrates the benefit of using ROI coding at low bit rates; without ROI coding at $R = 0.02$ bpp, it is virtually impossible to see the facial expressions of the mother and the daughter. Moreover, notice how at a higher bit rate, R $= 0.50$ bpp, and a high level of ROI prioritization, K $= 5$, the background quality is penalized from the precision losses associated with downscaling/upscaling of wavelet coefficients. Considering the trade-offs between the level of ROI prioritization, the loss in precision due to scaling, and the degree of excess joint rate, the choice of K $= 3$ seems to offer a good compromise. Visual representations of the side descriptions can be inferred from the joint descriptions; simply degrade the quality in one of the ROIs to a similar level as in the non-ROI regions in Fig. 4.

IV. Conclusions and Future Work

We have presented a novel coding scheme that exploits region of interests (ROIs) for multiple description (MD) coding by placing each ROI into a separate description. This method enables efficient coding when both descriptions are received, and the ROIs are coded with higher priority than non-ROIs at low bit rates. If only one of the descriptions is received, one ROI is still reconstructed with good quality, and may be sufficient for recognition or interpretation purposes.

There is little doubt that ROI based MD coding would be even more advantageous for coding of videos than still images. First, problems associated with real-time video coding and transmission over heterogeneous networks such as the Internet are inherently captured by the philosophy of MD coding. Second, if one description, and thus one ROI is lost during transmission, the missing ROI can be estimated from the previous frame. Since current state of the art video codecs such as H.264 use multi-frame inter prediction, the prediction accuracy in the current frame would suffer only little if descriptions were dropped randomly (at a sufficiently low error rate) over the past frames.

Since ROI based MD *video* coding would be a natural continuation of the work presented here, we plan to extend our approach by implementing and simulating the performance of ROI based MD coding in a video coding framework.

References

[1] A. El Gamal and T. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. 28, pp. 851–857, Nov. 1982.

[2] V. Goyal, "Multiple description coding: Compression meets the network," *IEEE Signal processing mag.*, vol. 18, pp. 74–93, Sept. 2001.

[3] A. Said and W. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 6, no. 3, pp. 243–250, June 1996.

[4] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The jpeg 2000 still image coding system: An overview," *IEEE Trans. Consumer Electron.*, vol. 46, no. 4, pp. 1103–1127, Nov. 2000.

[5] Z. Wang and A. Bovik, "Embedded foveation image coding," *IEEE Trans. Image Processing*, vol. 10, no. 10, pp. 1397–1410, Oct. 2001.

[6] A. Cohen, I. Daubechies, and J. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Commun. Pure Appl. Math*, vol. 45, no. 5, pp. 485–560, 1992.

Fig. 4. *Visual coding results at different bit rates and levels of ROI prioritization.* Images in the first row are coded with $R = 0.02$ bpp, second row with $R = 0.07$ bpp, and third row with $R = 0.50$ bpp. Columns represent prioritization levels $\mathrm{K} = \{1, 3, 5\}$, respectively. All of the images are decoded with the central decoder (DECODER 0).