

# Using Self-Authentication-and-Recovery Images for error concealment in wireless environments

Ching-Yung Lin\*, Daby Sow\* and Shih-Fu Chang\*\*

\*: IBM T. J. Watson Research Center, 30 Saw Mill River Rd., Hawthorne, NY 10532  
{cylin, sowdaby}@watson.ibm.com

\*\* : Department of Electrical Engineering, Columbia University, New York, NY 10027  
sfchang@ctr.columbia.edu

## ABSTRACT

Handling packet loss or delay in the mobile and/or Internet environment is usually a challenging problem for multimedia transmission. Using connection-oriented protocol such as TCP may introduce intolerable time delay in re-transmission. Using datagram-oriented protocols such as UDP may cause partial representation in case of packet loss. In this paper, we propose a new method of using our self-authentication-and-recovery images (SARI) to do the error detection and concealment in the UDP environment. The lost information in a SARI image can be approximately recovered based on the embedded watermark, which includes the content-based authentication information and recovery information. Images or video frames are watermarked in a priori such that no additional mechanism is needed in the networking or the encoding process. Because the recovery is not based on adjacent blocks, the proposed method can recover the corrupted area even though the information loss happen in large areas or high variant areas. Our experiments show the advantages of such technique in both transmission time saving and broad application potentials.

## 1. INTRODUCTION

Compressed multimedia data may suffer severe degradation over practical communication channels. Traditionally, in order to guarantee the correct representation in the receiver end, connection-oriented Transmission Control Protocol (TCP) is usually used in applications. However, TCP is a handshaking protocol that it requires a lot of time in exchanging the acknowledgment and re-transmission information. This introduces intolerable time delay in data representation. Also, in the multicasting environment, TCP is not effective because of the lack of an effective congestion control mechanism, which satisfies all receivers. Therefore, using the User Datagram Protocol (UDP) plus a good error control/correction/ concealment scheme in the higher layers may be a promising solution for multimedia transmission in the multicast cases, serious congestion cases, and high bit error rate cases.

Real-time transcoding middlewares, which can convert various types of video sources to versatile hand-held or mobile devices, are becoming more and more popular. Figure 1 shows the system architecture of our video transcoding and streaming project called *Universal Tuner*. In this project, we implemented a prototype of a variable-complexity codec that can transcode MPEG-1/2 video, live TV broadcasting or web-cam video capturing in real-time and stream it to the Palm-OS Personal Digital Assistant (PDA) devices. Our system requires the codec to handle 5 to 10 video signals at the same time and stream them to the clients. Usually, the number of video sources and clients are usually asymmetric. A massive number of clients are usually interested in same contents, such as TV broadcasting. In this multi-casting scenario, a TCP-based server needs multi-threads to transcode video signals for multiple clients, or uses a lot of memory for buffering and retransmission. In the client side, there is usually a time-legacy delay caused by buffering. On the other hand, an UDP-based server needs only single thread to transcode video signals, and there are no delays in the client side. Therefore, applying UDP and error concealment techniques, we can reduce the server's CPU load on transcoding video contents, buffer management and channel resources for retransmission.

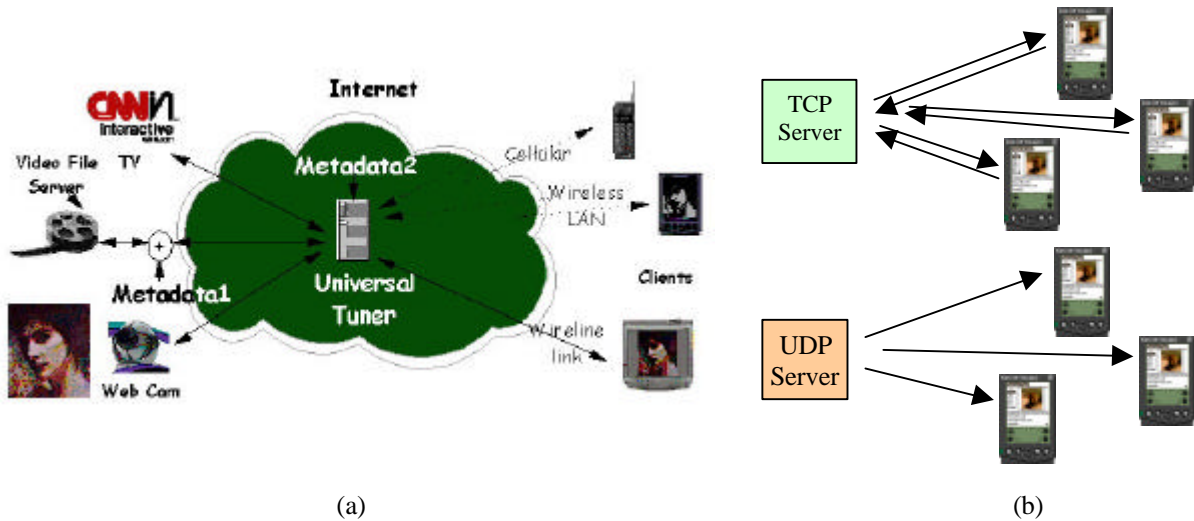


Figure 1: (a) System description of the Universal Tuner video transcoding middleware; (b) Comparison of TCP server and UDP server. TCP-based servers need more resources than UDP-based servers in memory, channel bandwidth, CPU processing, and introduce time-legacy in the client side.

Error concealment is a useful technique in case of the high error bit rate in the mobile environment or packet loss/delay during Internet transmission. Error concealment in videos is generally based on temporal predictive concealment and spatial predictive concealment. Temporal methods utilize the information redundancy in the adjacent frames. This property may not be true in the cases that frames are composed of fast moving objects. For this kind of videos as well as images, spatial predictive concealment is needed.

Error detection in the corrupted bitstream and re-synchronizing the remaining bitstream is an important issue in prior to error concealment. Usually, an error bit in the entropy coded bitstream affects not only a coefficient or a block, but also the following blocks (until the bitstream is resynchronized, if inserted by the encoder). A network layer packet loss may introduce similar effect, which may include several RST intervals, *e.g.*, several rows. Previous researchers dealt this problem in three different ways. The first one is to assume such problems do not exist and the positions of corrupted blocks are provided to the error concealment (EC) algorithm. Although this assumption may be sufficient for evaluating the quality of recovered area, it is not practical in the real world. The second one is to modify the encoder to insert parity bits to isolate the bit errors or inserting the number of bits assigned to each block [5]. The third method is to change transport layer protocol to insert more monitoring functionalities.

In [18], Wang and Zhu made a detailed overview of error control and concealment techniques. These techniques can be characterized as follows: (1) Layered coding with transport prioritization, *i.e.*, a combination of scalability layers and error correction codes [2, 12, 21]; (2) Multiple-description coding [16, 17]; (3) Joint source and channel coding [1]; (4) Robust waveform coding, *i.e.*, redundancy in the waveform coding stage [6]; (5) Robust entropy coding, which places redundancy in the entropy coding stage [11, 13]; (6) Postprocessing at the decoder, which blindly recover the lost pixels based on their surrounding spatial or temporal neighborhood pixels [3, 15]. In [20], Yin *et al.* show that the quality of blind recovery can be enhanced by embedding additional features in the image. For instance, they embed the edge directionality information bit of blocks into images.

In this paper, we propose a new method of using our Self-Authentication-and-Recovery Images (SARI) system for the purpose of error detection and concealment in datagram-oriented image/video transmission. SARI utilizes a semi-fragile watermarking technique. Like a gecko can recover its cut tail, a watermarked SARI image can detect manipulations, resynchronize the remaining bitstream following the corrupted, and recover an approximated original image on the lost area. A SARI image embeds two kinds of information watermarks: authentication bits and recovery bits. The content-based watermark bits generated from a block set, which includes two blocks for extracting authentication bits and four blocks for generating recovery bits, are embedded into other blocks in the image. The

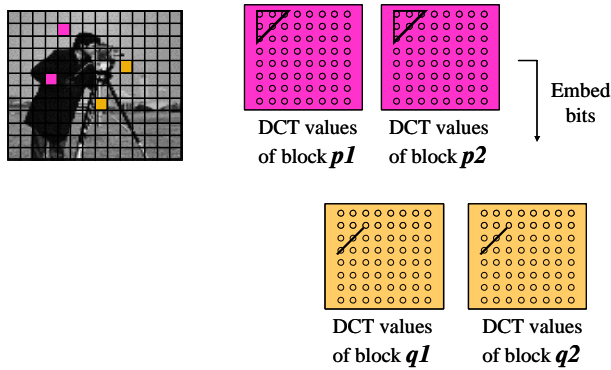


Figure 2: Generate authentication information bits from a block pair and embed them into the other two blocks of image.

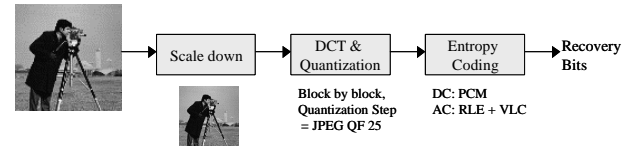


Figure 3: Generate recovery information bits from the scaled down version of the original image.

locations of corrupted blocks are detected by the embedded authentication information, while the lost blocks in a SARI image are approximately recovered based on the recovery information.

SARI images are completely compatible to quantization-based lossy compressions. A watermarked image can be lossy compressed by JPEG/MPEG, and still keep all the embedded information. The lowest acceptable compression quality factor depends on an adjustable watermarking strength controlled by the SARI embedder. The proposed method has two advantages for error concealment. 1.) Users do not need any additional control schemes in network transmission or in the encoding process. 2.) It can recover the corrupted area even though the information loss in large areas or high variant areas, because the recovery is not based on adjacent blocks. The possible drawback of the proposed method is that, comparing a SARI image to the original, there may exist slight visual quality degradation and a possible length increase in the compressed file under the same quantization table.

The rest of this paper is organized as follows. In Section 2, we describe the theorems and design of SARI system. We first show how we generate the compression-resistant authentication information bits and the recovery bits. Thereafter, we show how these information bits are embedded in the image, utilizing the maximum zero-error capacity of the image. In Section 3, we describe the functionality of error detection based on the authentication information in a SARI image. We will discuss how packet loss would affect the image/video under different scenarios. Some example and preliminary experimental results of the proposed method are also shown in this section. Finally, in Section 4, we discuss the advantages and disadvantages of the proposed system and our future work.

## 2. THEOREMS AND SYSTEM DESCRIPTION OF SELF-AUTHENTICATION-AND-RECOVERY IMAGES (SARI)

SARI is based on two invariant properties of quantization. The first one is the invariant relationships between two coefficients in a block pair before and after JPEG compression. The second property shows that if a DCT coefficient is modified to an integral multiple of a quantization step, which is larger than or equal to the steps used in later JPEG compressions, then this coefficient can be exactly reconstructed after later JPEG compression. In SARI, we use the first property to generate authentication bits, and use the second property to embed it, as well as recovery bits, as watermark. These properties provide solutions to two major challenges in developing authentication watermark: how to extract short and invariant information to detect the integrity of the received data, and how to embed information that is guaranteed to survive lossy compression.

Figure 2 shows the procedure of generating the authentication bits. First, we select every two adjacent blocks as a block pair, and then calculate the sign of difference of the DC coefficients and the first five low frequency AC coefficients according to this theorem in [8]:

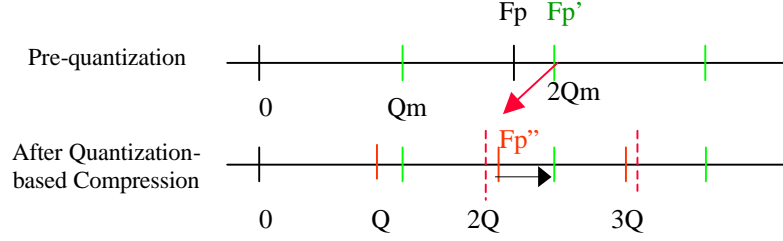


Figure 4: DCT coefficients can be exactly reconstructed using pre-quantization

◆ **Theorem 1:**

- If  $F_{p1}(i,j) - F_{p2}(i,j) > k$  then  $F_{p1}'(i,j) - F_{p2}'(i,j) \geq k'$
- If  $F_{p1}(i,j) - F_{p2}(i,j) < k$  then  $F_{p1}'(i,j) - F_{p2}'(i,j) \leq k'$
- If  $F_{p1}(i,j) - F_{p2}(i,j) = k$  then  $F_{p1}'(i,j) - F_{p2}'(i,j) = k'$

where  $F_{p1}(i,j)$  and  $F_{p2}(i,j)$  are the original  $(i,j)$ th DCT coefficients in the block  $p1$  and  $p2$ , and  $F_{p1}'(i,j)$  and  $F_{p2}'(i,j)$  are their quantized values after lossy compression. If  $k=0$  then  $k'=0$ . In this paper, we only use  $k=0$ , while other values of  $k$  and their resulting  $k'$  are discussed in [8]. To generate an authentication bit,  $Z$ , from a coefficient pair,  $F_{p1}(i,j)$  and  $F_{p2}(i,j)$ , we set  $Z=1$  if  $F_{p1}(i,j) = F_{p2}(i,j)$ , or  $Z=0$  if  $F_{p1}(i,j) < F_{p2}(i,j)$ . Therefore, for each block pair, we can generate 6 authentication bits. In our original design of SARI embedder for the security purpose, the formation block pairs and their corresponding embedding locations are controlled by a user-specific key. Therefore, an attacker cannot manipulate the coefficients to circumvent the authenticator. However, for error concealment, this security design may not be needed, and the mapping functions can be published.

The process of generating recovery information bits is shown in Figure 3. We first scale down the image by half in each dimension. The DCT coefficients are calculated from the scaled image, and then AC values are quantized by the JPEG quantization table of Quality Factor = 25, and DC values are represented by 6 bits. For each block, AC coefficients are run-length encoded (RLE) and variable length coded (VLC) by a defaulted Huffman Table. At the end of each block is a mark of EOB, which is compatible to general JPEG block. These 6 bits of DC coefficient and variable bits of AC coefficients from one block of the scaled image is embedded into 4 other blocks in the original image, using round robin method [9]. In our testing of 9 different types of image, the average length of recovery bits for each block is 10.1 bits, *i.e.*, 40.4 bits/ block in the scaled image [14].

Each block of the original image is divided into three areas. The first area, DC coefficient and 5 low frequency AC coefficients in the zig-zag order are used to generated authentication bits, which will be embedded in other blocks. The second area, midband coefficients are used to embedded both authentication and recovery bits, 1 bit/coefficient. For instance, if 3 authentication bits and 10 recovery bits are embedded per block, then its 6<sup>th</sup>-18<sup>th</sup> AC coefficients are used to embed 13 bits. Consider each coefficient is a communication channel, and it has a pre-defined maximum distortion  $M$ , then we can transmit information based the exact reconstructible coefficient theorem in [9]:

◆ **Theorem 2:** For all  $Q(i,j) = Qm(i,j)$

$$\text{Integer Rounding}[Fp''(i,j) / Qm(i,j)] Qm(i,j) = Fp'(i,j)$$

where  $Fp(i,j)$  is the original DCT coefficient at the position  $(i,j)$  of block  $p$ ,  $Fp'(i,j)$  is a pre-quantized DCT coefficient by  $Qm(i,j)$ , and  $Fp''(i,j)$  is the quantized result of  $Fp'(i,j)$  after lossy compression by Quantization step  $Q(i,j)$ . Figure 4 shows the property of this theorem. Theorem 2 is important to data embedding. In compression operations, any two close real-value coefficients may be quantized to different values if any quantization steps are applied. Therefore, usually the

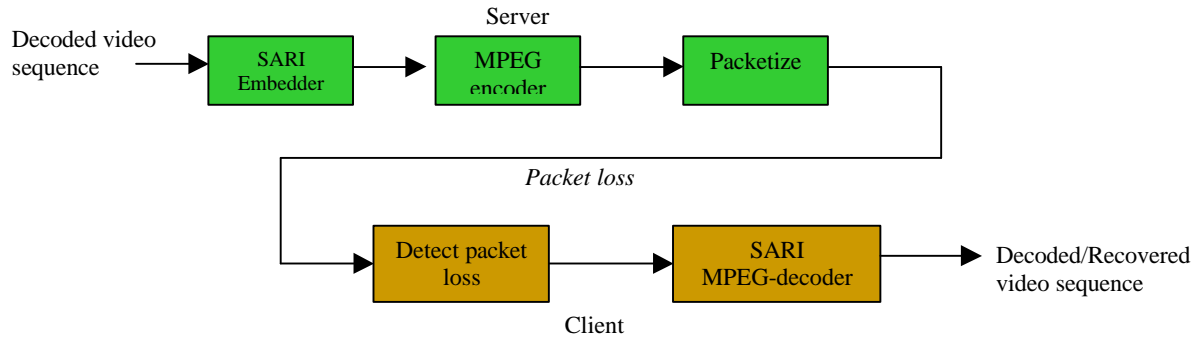


Figure 5: System block diagram of the transcoder utilizing SARI techniques for error concealment.

reconstructed values of the coefficients are unpredictable if we don't know how exactly this coefficient will be quantized. Usually, if the embed information needs to survive future compressions, we do not know what will be the exact quantization steps. Theorem 2 provides a solution to this issue. It indicates that if we pre-quantize a coefficient using a quantization step larger than or equal to the quantization steps in the future operations, then this quantized value can be exactly recovered even after suffering various quantization steps. The positions of possible quantized values work like the only attractors (in the real value lines) that can be exactly recovered after compression. In Theorem 2,  $Q_m = M/2$  is the maximal allowable quantization step, which can be pre-selected in the watermark embedder. Each information bit is inserted to the coefficient by modifying the LSB of  $Fp'$ . Note that embedding one bit only needs modify one original DCT coefficient. We have proved that this is the maximum zero-error channel capacity based on Shannon's graph theory [10]. Distortion of  $M$  can be decided by contrast masking, *a.k.a.* Just Noticeable Distortion, or pre-selected values.

### 3. ERROR DETECTION AND CONCEALMENT

Figure 5 shows the system diagram of the proposed system in the real-time image/video transcoder. SARI watermark embedder is first applied on the decoded image or video I-frames before the re-encoding process. After embedding, all the encoding processes and transmissions processes are all the same as defined in standards. At the client hand, in addition to the standard decoder, we need to detect packet losses and then conceal the errors based on the information embedded in the SARI watermarks.

Typical packet sizes after segmentation are in the order of 1500 bytes for Ethernet and wireless local area networks (IEEE 802.11). This number tends to go down to less than 700 bytes in wireless personal area network (IEEE 802.15) using Bluetooth technology, in order to take into account the buffering capabilities of certain wireless client devices (e.g. cellular phones). Focusing on visual information, it is reasonable to assume that the client device will have enough capabilities to handle this type of traffic and larger packet sizes. Assuming a packet size of 1500 bytes, with a bit error rate (BER) of ranging from  $10^{-4}$  to  $10^{-6}$ , the probability of having a packet error ranges then from 11.3% to 0.12%. Due to the nature of the visual traffic, these errors have serious impact on the overall quality of the delivered content. To see this, recall the representation scheme used in most transform coders. The loss of a packet unit results in the inability to decode all future blocks predicted from the last one included in the lost packet. The situation is even more serious for video traffic as the loss of a packet breaks the temporal prediction chain for several macroblocks. Therefore, in practical cases, we have to consider the damage of lost packets in the IP layers and control its effects in the higher layers.

Errors usually propagate through several blocks in the case of bit error or packet loss. Figure 6 shows the results of an IP packet loss in a JPEG compressed Lena image. If the image is compressed by an encoder without reset (RST) marks, such as in XV, then the bottom part of the image will not be retrieved after packet loss. If the encoder insert RST marks into the image, then the image can be resynchronized, but only after the first RST at the next packets. Usually,



Figure 6: Effects of packet loss in a JPEG bitstream: (a) without reset (RST) marks, (b) with RST marks, after resynchronization.

Figure 7: Recovered image: (a) authentication information can be used to calibrate the position of blocks following the lost packet with or without RST, (b) use authentication information to calibrate and recovery information to recover an approximation of the lost area.

eight RST marks are used cyclically, which provides the location information of the next resynchronized blocks. However, this information may be still falsely reconstructed if there are more than 8 RST marks are included in the lost packets. The reconstructed image is in Figure 6(b).

Using SARI images, the embedded authentication bits can be used to calibrate the location of corrupted error. It does not matter whether RST is available to the detector. From the packet immediately after the lost area, we first find out the possible EOB pattern, which is 1010 in the AC luminance VLC codes, to decide the end of a block. Then, we can reconstruct the AC coefficients and the DPCM of DC of the following blocks. We then shift the block location to find out the match position from other parts of the image. Therefore, the error part can be located. After this process, we then recovery the lost area by using the information hidden in other places. Also, the DC values can be recovered in the blocks right after the lost packet and before the RST. The result of error detection and the error concealment of the previous example is shown in Figure 7. In Figure 7(a), we see that the SARI image can exactly detect the corrupted area caused by lost packets. The system needs not wait until the next RESET flag to regain synchronization. In Figure 7(b), although we can see visible quality degradation in the recovered area if we look at the recovered image closely, the recovered image has shown its superior capabilities of reconstructing lost critical information in the image. Compared to the results in Figure 6, the SARI recovered image can recover the corrupted eye areas that are quite different from their adjacent blocks. In general, for a stand-alone image or MPEG I-frames, recovering such critical areas is a very hard task, if not impossible, for other error concealment techniques based on blind recovery.

For MPEG video streaming, we can apply the technique mentioned above to recover the packet-losses in the I-frames. Similar methods can be applied to DCT residue coefficients in the P and B frames. To recover errors that are corresponding to motion vectors in the compressed bitstream, we use the technique which was proposed by Haskell and Messerschmitt in 1992 and was later improved by Lam *et. al.* in 1993. In their techniques, they use blind recovery methods to estimate lost motion vectors. These motion vectors are estimated by the following methods: (1) setting motion vectors to be zero, (2) using motion vectors of the corresponding block in the previous frame, (3) using the average of the motion vectors from the spatially adjacent macroblocks, and (4) using the median of motion vectors from the spatially adjacent macroblocks. Then, we can apply these four assumptions to recover the macroblocks based on them and their corresponding DCT residue values. The reconstructed macroblock, which has the smallest boundary discontinuity errors, is selected as the recovery result in our system.

#### 4. CONCLUSIONS AND FUTURE WORK

Compare to other error control and concealment techniques, the proposed system has the following advantages:

- There is no need to modify any standards, because the watermark embedder is performed before the stages of compression and transmission standards. In other words, this system is totally compatible with MPEG-1/2, Ethernet, Wireless LAN, and Bluetooth standards.
- The embedded information bits, which include authentication and recovery bits, are evenly distributed inside the image/video frame. Except the bitstream headers that include the format information, no other packets are much important than others. This property can reduce the error propagation damages.
- Using SARI-based error concealment techniques, we can get a better recovery quality compared to the blind recovery techniques. The proposed technique has the unique advantage of recovering critical areas that cannot be estimated by either spatial or temporal neighbors.

On the other hand, although the authentication and recovery information are all embedded in the image/video and then survive compression, the embedded bits do not increase the file size in the raw format. However, they may slightly increase the payload in the compressed domain. Therefore, this technique may have the disadvantage of larger transmission time, compared to the techniques based on blind recovery.

Using the high embedding capacity of SARI, we can hide layered-coding information into waveforms to improve the quality of error concealment without changing the transportation ECC schemes and video codec. Our proposed SARI-based error concealment technique has made UDP-based video multicasting possible and has been integrated in our real-time video transcoding middleware. In the future, we will focus on a large scale testing of this technique, including its performance analysis in its savings on CPU resources, bandwidth, and memory. We will also investigate other techniques for recovering lost motion vectors, such as information hiding techniques.

## 5. REFERENCES

- [1] N. Farvardin and V. Vaishampayan, "Optimal quantizer design for noisy channels: an approach to combined source-channel coding," *IEEE Trans. on Information Theory*, Vol. 38, pp. 827-838, Nov. 1987.
- [2] M. Ghanbari, "Two-layer coding of video signals for VBR networks," *IEEE J. Select. Areas Comm.*, Vol. 7, pp. 801-806, June 1989.
- [3] M. Ghanbari, "Cell-loss concealment in ATM video codes," *IEEE Trans. on Circuits System on Video Technology*, Vol. 3, pp. 238-247, June 1993.
- [4] P. Haskell and D. Meserschmitt, "Resynchronization of motion compensated video affected by ATM cell loss," *IEEE ICASSP*, San Francisco, Mar. 1992.
- [5] C.-S. Kim, R.-C. Kim and S.-U. Lee, "An Error Detection and Recovery Algorithm for Compressed Video Signal Using Source Level Redundancy," *IEEE Trans. on Image Processing*, Vol. 9, No.2, Feb. 2000.
- [6] S. S. Hemami and R. M. Gray, "Image reconstruction using vector quantized linear interpolation," *IEEE ICASSP*, Australia, May 1994.
- [7] W.-M. Lam, A. R. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," *IEEE ICASSP*, Minneapolis, Apr. 1993.
- [8] C.-Y. Lin and S.-F. Chang,, "A Robust Image Authentication Method Distinguishing JPEG Compression from Malicious Manipulation," *IEEE Trans. on Circuits and Systems for Video Technology*, Feb. 2001; also appeared on *CU/CTR Technical Report 486-97-19*, Columbia Univ., Dec. 1997.
- [9] C.-Y. Lin and S.-F. Chang,, "Semi-Fragile Watermarking for Authenticating JPEG Visual Content," *SPIE Security and Watermarking of Multimedia Content II*, San Jose, Jan 2000.
- [10] C.-Y. Lin and S.-F. Chang,, "Zero-Error Information Hiding Capacity of Digital Images," *IEEE Intl. Conf. on Image Processing*, Greece, Oct. 2001.
- [11] J. C. Maxted and J. P. Robinson, "Error recovery for variable length codes," *IEEE Trans. on Information Theory*, Vol. 31, pp. 794-801, 1985.

- [12] K. Ramchandran, A. Ortega, and K. M. Uz, "Multiresolution broadcast for digital HDTV using joint source channel coding," *IEEE J. Select. Areas Comm.*, Vol. 11, pp. 6-23, Jan. 1993.
- [13] D. W. Redmill and N. G. Kingsbury, "Transcoding of MPEG-II for enhanced resilience to transmission errors," *IEEE Intl. Conf. on Image Processing*, Lausanne, Vol. 2, Nov. 1996.
- [14] SARI WebPages: <http://www.ctr.columbia.edu/sari>
- [15] H. Sun and W. Kwok, "Concealment of damaged block transform coded images using projections onto convex sets," *IEEE Trans. on Image Processing*, Vol. 4, pp. 470-477, Apr. 1995.
- [16] V. A. Vaishampayan, "Design of multiple description scalar quantizers," *IEEE Trans. on Information Theory*, Vol. 39, pp. 821-834, May 1993.
- [17] Y. Wang, M. T. Orchard, and A. R. Reibman, "Multiple description image coding for noisy channels by paring transform coefficients," *Proc. IEEE Workshop of Multimedia Signal Processing*, Princeton, June 1997.
- [18] Y. Wang and Q.-F. Zhu, "Error Control and Concealment for Video Communication: A Review," *Proc. IEEE*, Vol. 86, No. 5, May 1998.
- [19] Y. Yang, N. P. Galatsanos and A. K. Katsaggelos, "Projection-Based Spatially Adaptive Reconstruction of Block-Transform Compressed Images," *IEEE Trans. on Image Processing*, Vol. 4, No. 7, July 1995.
- [20] P. Yin, B. Liu and H. Yu, "Error Concealment Using Data Hiding," *IEEE Intl. Conf. on ASSP*, Salt Lake City, May 2001.
- [21] Y.-Q. Zhang, Y. J. Liu and R. L. Pickholtz, "Layered image transmission over cellular radio channels," *IEEE Trans. Veh. Technology*, Vol. 43, pp. 789-796, Aug. 1994.