COMPLEXITY-QUALITY ANALYSIS OF TRANSCODING ARCHITECTURES FOR REDUCED SPATIAL RESOLUTION

Anthony Vetro, Toshihiko Hata, Naoki Kuwahara, Hari Kalva and Shun-ichi Sekiguchi

***Abstract*—This paper presents a detailed complexity-quality analysis of various transcoding architectures that perform spatial resolution reduction. In addition to the Reference architecture, two additional architectures are considered. The optimization of several processing components, including the down-sampling operation, is presented. We have found that the two additional architectures provide a good trade-off in terms of complexity and quality compared to the Reference.**

***Index Terms*—Transcoding, Down-Conversion, Optimization, Architecture, Complexity-Quality.**

I. INTRODUCTION

GENERALLY speaking, transcoding is an important application area for consumer electronics. With the growing amount of content available in various formats, the ability to transcode audio-visual content can impact the design of receivers and their complexity. More importantly, it is expected that transcoding functionality will be incorporated into a number of consumer electronic devices, e.g., set-top boxes, PC's, still/video camera, etc.

To enable broadcast-quality video streams to be decoded and displayed on mobile devices, transcoding from MPEG-2 MP@ML to MPEG-4 Simple Profile is needed. This conversion implies a reduction in bit-rate from approximately 6Mbps to 384kbps and lower, as well as a reduction in spatial resolution from 720x480 interlace to 352x240 progressive. In [1], several low-complexity video transcoding architectures that meet these requirements have been proposed. In this paper, a detailed analysis of the complexity and quality of select architectures is presented. The optimizations of several processing components, such as the DCT down-conversion, will also be presented.

The rest of this paper is organized as follows. In section II, an overview of the transcoding architectures is given. Section III describes algorithmic and processor-specific optimizations that were considered, including the optimization of the down-sampling processes. In section IV, the architectures are analyzed in terms of their complexity and quality. Finally, conclusions are presented in section V.

Anthony Vetro and Hari Kalva are with Mitsubishi Electric Research Labs, Murray Hill, NJ, USA (contact author: avetro@merl.com).

Toshihiko Hata and Naoki Kuwahara are with Mitsubishi Electric Corporation, Advanced Technology R&D Center, Amagasaki City, Japan.

Shun-ichi Sekiguchi is with Mitsubishi Electric Corporation, Information Technology R&D Center, Kamakura City, Japan.

II. TRANSCODING ARCHITECTURES

The architectures under consideration are shown in Figs. 1-3. Fig. 1 illustrates the Reference architecture, which is simply a cascaded approach that decodes, down-samples and re-encodes the video. Fig. 2 shows the proposed Intra Refresh architecture (Pro1), which compensates for various errors by converting select macroblocks to intra-coded blocks. Fig. 3 shows the proposed Partial Encoder architecture (Pro2), which is similar to the Reference architecture, but simplifies the re-encoding process by not compensating for re-quantization errors. The background regarding the development of these architectures can be found in [1], but a brief description of the two proposed architectures shown in Figs. 2 and 3 is included below for completeness.

In reduced resolution transcoding, drift error is caused by many factors, such as requantization, motion vector truncation and down-sampling. Such errors can only propagate through inter-coded blocks. By converting some percentage of inter-coded blocks to intra-coded blocks, drift propagation can be controlled. In the past, the concept of intra-refresh has successfully been applied to error-resilience coding schemes [2], and we have found that the same principle is also very useful for reducing the drift in a transcoder. The Intra Refresh architecture shown in Figure 2 is based on this concept.

In the Intra Refresh scheme, output macroblocks are subject to a DCT-domain down-conversion, requantization and variable-length coding. Output macro-blocks are either derived directly from the input bitstream, i.e., after variable-length decoding and inverse quantization, or retrieved from the frame store and subject to a DCT operation. Output blocks that originate from the frame store are independent of other data, hence coded as intra blocks; there is no picture drift associated with these blocks.

The decision to code an intra-block from the frame store depends on the macroblock coding modes and picture statistics. In a first case based on the coding mode, an output macroblock corresponds to four input macroblocks for size conversion by a factor of two in each direction. Since all sub-blocks must be coded with the same mode, the transcoder must avoid having *mixed-blocks*, i.e., inter-coded and intra-coded sub-blocks in the same output macroblock. This is detected by the mixed-block processor, which will trigger the output macroblock to be intra-coded. In a second case based on picture statistics, the motion vector and residual data are used

Manuscript received June 24, 2002

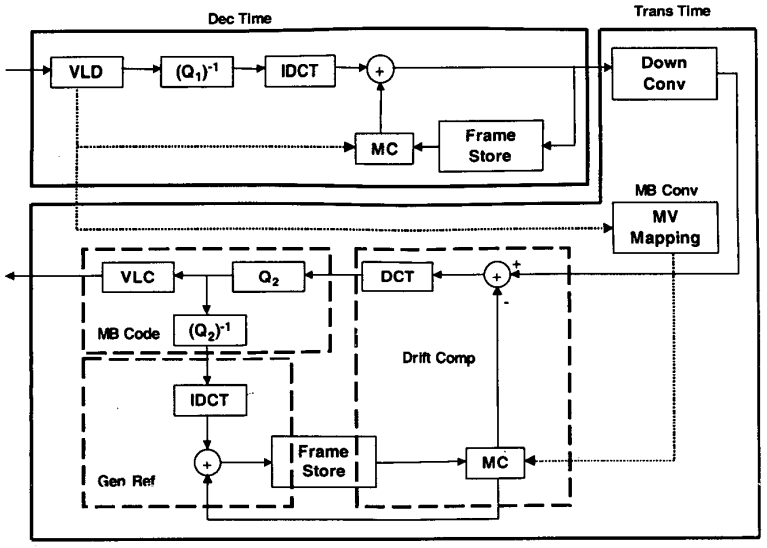


Fig. 1 Reference architecture for reduced spatial-resolution transcoding.

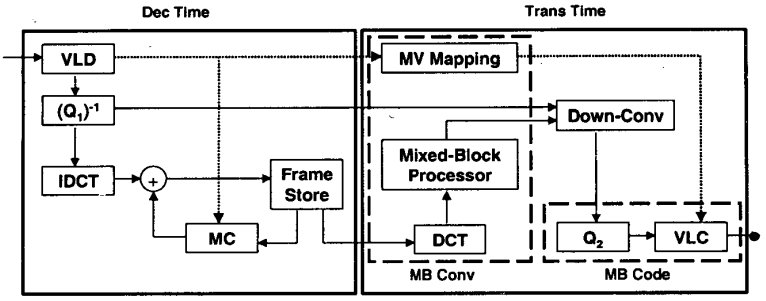


Fig. 2 Intra Refresh architecture for reduced spatial-resolution transcoding (Pro1).

to detect blocks that are likely to contribute to larger drift error. For this case, picture quality can be maintained by employing an intra-coded block in its place. Of course, the increase in the number of intra-blocks must be compensated for by the rate control. Please refer to [1] for details on the operation of the rate control function.

As an alternative to the Reference architecture, the Partial Encode architecture is considered in Fig. 3. This architecture aims to eliminate the drift error due to down-sampling and motion vector scaling; drift error caused by requantization is neglected. It operates under the assumption that the error due to requantization in a reduced resolution transcoder is much less than the error due to down-sampling and motion vector scaling.

In the Reference architecture, the reconstructed reference frame used for re-encoding consists of two parts, the low-resolution motion compensated prediction and the reconstructed low-resolution residual. In contrast to this Reference architecture, the Partial Encode architecture essentially removes the feedback components (inverse quantization and inverse DCT) that contribute the residual component to the reconstructed reference frame.

III. Optimizations of Transcoder Processes

This section describes both algorithmic and processor-specific optimizations that were used to improve the performance of the transcoding architectures.

A. *Optimization of DCT-Domain Down-Conversion*

In our earlier work [3], the concept of frequency synthesis has been proposed to transform an input DCT macroblock consisting of four 8x8 DCT blocks into an single 8x8 DCT block. These filters are used in the transcoder to perform the DCT-domain down-sampling operation. The computations are actually performed on the rows and columns of the macroblock using separable 1D filters. Let $\underline{A}$ and $\underline{B}$ denote the input vectors of size N. Then, the output data, $\underline{E}$, also of size $\underline{N}$, is computed according to,

$$\underline{E} = f_1 \cdot \underline{A} + f_2 \underline{B} \quad (1)$$

where

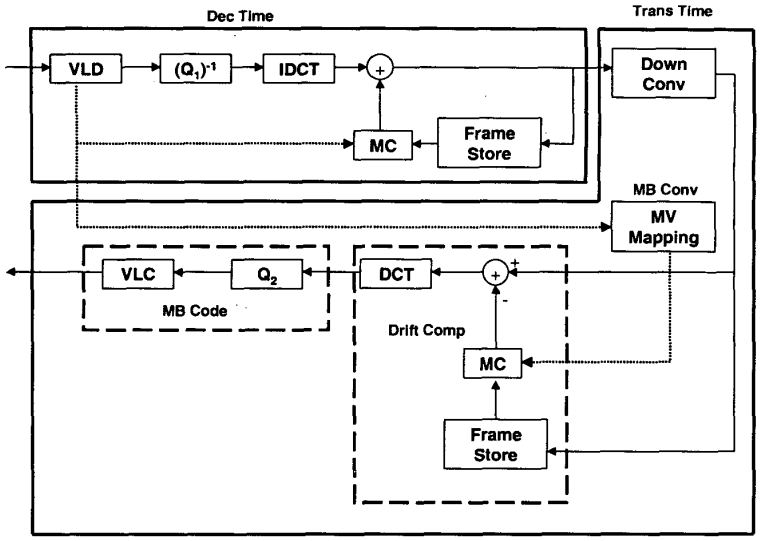


Fig. 3 Partial Encode architecture for reduced spatial-resolution transcoding (Pro2).

$$f_1(k,p)=\sum_{i=0}^{N-1}\psi_p^N(i)\cdot\psi_k^{2N}(i)$$
$$f_2(k,p)=\sum_{i=0}^{N-1}\psi_p^N(i)\cdot\psi_k^{2N}(i+N) \quad (2)$$

and

$$\psi_k^N(i)=\sqrt{\frac{2}{N}}\cdot\alpha(k)\cdot\cos\left(\frac{(2i+1)k\pi}{2N}\right) \quad (3)$$

and $\alpha(k)=1/\sqrt{2}$ for $k=0$, and 1 for $k\neq 0$. The above down-conversion filters can be applied in both the horizontal and vertical directions, and to both frame-DCT and field-DCT blocks. In our experiments, we aim to produce a progressive (MPEG-4) output. Therefore, if the input is a frame-DCT block, we apply the above filters to both fields. On the other hand, if the input is a field-DCT block, a similarly derived set of filters is used to simultaneously perform a field-to-frame conversion and down-conversion in the DCT domain.

We have observed that these filters exhibit favorable properties towards reducing the complexity. A sample set of filters that have been derived by equation (2) are given below, where k denotes the output index and p denotes the input index. For even pixel outputs, the filtering can be reduced from a 16-tap filter to a simple averaging operation. For odd pixel outputs, we can exploit the symmetric properties of the filter taps to significantly reduce the number of multiplications and additions. The relationship between the odd tap filters is given by,

$$f_1(i,j)=(-1)^{j+1}f_2(i,j) \quad (4)$$

Using the above relationship, a significant amount of computation can be saved. The amount of savings will be further analyzed in section IV-A.

```
f1[k][p] =
        { 0.50000, -0.00000, -0.00000, -0.00000,  0.00000,  0.00000,  0.00000,  0.00000}
        { 0.45088,  0.21117, -0.04136,  0.01703, -0.00883,  0.00499, -0.00278,  0.00125}
        { 0.00000,  0.50000,  0.00000,  0.00000,  0.00000,  0.00000,  0.00000,  0.00000}
        {-0.15224,  0.38511,  0.26956, -0.06723,  0.03086, -0.01660,  0.00903, -0.00403}
        { 0.00000,  0.00000,  0.50000,  0.00000,  0.00000,  0.00000,  0.00000,  0.00000}
        { 0.09375, -0.15691,  0.35925,  0.28339, -0.07500,  0.03489, -0.01786,  0.00777}
        { 0.00000,  0.00000,  0.00000,  0.50000,  0.00000,  0.00000,  0.00000,  0.00000}
        { 0.06966,  0.10672, -0.14309,  0.35148,  0.28742, -0.07625,  0.03364, -0.01383}

f2[k][p] =
        { 0.50000, -0.00000,  0.00000,  0.00000,  0.00000,  0.00000,  0.00000,  0.00000}
        {-0.45088,  0.21117,  0.04136,  0.01703,  0.00883,  0.00499,  0.00278,  0.00125}
        { 0.00000, -0.50000,  0.00000,  0.00000,  0.00000,  0.00000,  0.00000,  0.00000}
        { 0.15224,  0.38511, -0.26956, -0.06723, -0.03086, -0.01660, -0.00903, -0.00403}
        { 0.00000,  0.00000,  0.50000,  0.00000,  0.00000,  0.00000,  0.00000,  0.00000}
        {-0.09375, -0.15691, -0.35925,  0.28339,  0.07500,  0.03489,  0.01786,  0.00777}
        { 0.00000,  0.00000,  0.00000, -0.50000,  0.00000,  0.00000,  0.00000,  0.00000}
        { 0.06966,  0.10672,  0.14309,  0.35148, -0.28742, -0.07625, -0.03364, -0.01383}
```

TABLE I
COMPARISON OF COMPLEXITY USING NON-OPTIMIZED CODE.

Arch	Frame Rate	Total Time	Dec Time	Trans Time	MB Conv	Down Conv	MB Code	Rate Control	Drift Comp	Gen Ref	Other
Ref	10fps	43.80	7.43	36.1	0.40 (1.1%)	10.00 (27.7%)	5.40 (14.9%)	0.88 (2.4%)	16.22 (44.9%)	1.88 (5.2%)	1.32 (3.7%)
Pro1	10fps	24.79	7.40	17.11	5.68 (33.2%)	5.38 (31.4%)	5.09 (20.5%)	0.88 (5.14%)	N/A	N/A	0.08 (0.46%)
Pro2	10fps	38.32	7.40	30.63	0.32 (0.01%)	8.90 (29.1%)	4.91 (16.0%)	0.89 (2.9%)	15.49 (50.6%)	N/A	0.11 (0.39%)

TABLE II
COMPARISON OF COMPLEXITY USING OPTIMIZED CODE.

Arch	Frame Rate	Total Time	Dec Time	Trans Time	MB Conv	Down Conv	MB Code	Rate Control	Drift Comp	Gen Ref	Other
Ref	10fps	12.32	3.69	8.34	0.02 (0.24%)	0.55 (6.59%)	2.88 (34.5%)	0.06 (0.72%)	2.64 (31.6%)	2.00 (23.9%)	0.19 (2.28%)
Pro1	10fps	8.39	3.66	4.45	0.47 (10.5%)	1.51 (33.9%)	2.34 (52.5%)	0.05 (1.12%)	N/A	N/A	0.08 (1.8%)
Pro2	10fps	7.78	3.72	3.78	0.02 (0.5%)	0.40 (10.5%)	2.20 (58.2%)	0.06 (1.59%)	1.06 (28.0%)	N/A	0.04 (1.06%)

B. *Processor-Specific Optimizations*

The transcoder optimizations are targeted for the Intel Pentium-4 processors. The Intel Pentium-4 processor has architectural and instruction-level support to speedup compute-intensive processes such as MPEG video processing. The processor's single-instruction multiple-data (SIMD) capability allows operations on packed integer or floating point data contained in 64-bit MMX or 128-bit XMM registers. The MMX, SSE, and SSE2 instruction sets include instructions that operate on packed integer or floating point data to speed up execution.

The MPEG video encoding and decoding algorithms operate on 8x8 blocks of data. For the first round of optimizations, the most common block operations were optimized taking advantage of the SIMD capabilities of the processor. The optimized block operations include FDCT, IDCT, copying, type-converting, zeroing, summing, and clipping elements of a block. These optimized block operations are used in the MPEG-2 decoder as well as the MPEG-4 transcoder resulting in an improved performance in both the MPEG-2 decoder and the MPEG-4 transcoder.

To further speedup the transcoder, the portions of the transcoder that were profiled to be taking up a significant amount of processing power were identified and optimized. The MPEG-2 decoder optimizations included motion compensation and frame reconstruction operations. The motion compensation optimizations took advantage of the processor's PAVGB instruction to average a row of a block in a single instruction.

The MPEG-4 transcoder specific optimizations take advantage of the processor capabilities to speedup compute-intensive processes such as quantization, drift compensation, and down sampling. All these optimizations use the SIMD capability to process multiple elements in a parallel. The memory allocations are aligned on 16-byte boundary to speedup memory access latency in the processor cache. With aligned memory, the load and store instructions that operate on 128-bit XMM registers and aligned memory are be used for faster load/store operations. The quantization optimizations take advantage of the division of packed floating-point data stored in XMM registers. The data range clipping uses min and max operations on packed data in XMM registers to efficiently clip quantized coefficients.

In addition to the optimizations to the source code, the Intel compiler was used to generate optimized, processor specific code. The processor specific optimizations resulted in significant performance improvements making it possible to transcode multiple MPEG-2 video streams simultaneously.

IV. COMPLEXITY-QUALITY ANALYSIS

In this section, we compare the complexity and quality of the proposed transcoding architectures to the Reference architecture.

A. *Complexity Analysis*

In our experiments, we consider a set of MPEG-2 MP@ML bitstreams encoded at a rate of 6Mbps with GOP parameters

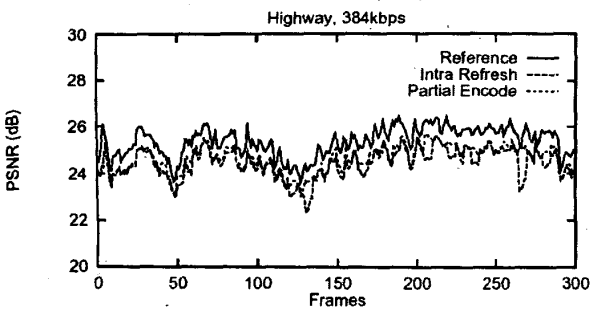


Fig. 4 Quality comparison of transcoding architectures for *Highway*.

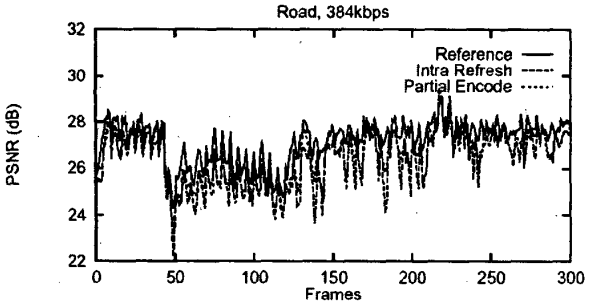


Fig. 5 Quality comparison of transcoding architectures for *Road*.

N=15 and M=3. Each bitstream contains 900 frames, with duration of 30 seconds. The transcoder drops B-frames, hence the transcoded output has a frame-rate of 10 frames/sec.

The platform used in our simulations is a 1.8GHz Pentium-4 Processor with 512MB of RAM. A four-pixel averaging operation was used to perform the spatial down-sampling in the Reference and Partial Encode architectures.

Table I provides a comparison of the execution time using the non-optimized software for the three architectures under consideration. The headings of this table correspond to the blocks outlined in Figs. 1-3. From this analysis, it is clear that the complexity of the Reference is higher than both of the proposed architecture and that the complexity of Pro2 is between that of the Reference and Pro1. Analyzing the breakdown of the transcoding time, we find that the down-conversion process occupies a significant portion of the time for all architectures. For the Pro1 architecture, the processing contributing to MB_Conv are also time-consuming, mainly due to the DCT process. Similarly, in both the Reference and Pro2 architectures, the processes contributing to Drift_Comp consume a significant percentage of time due to both DCT and MC processes.

The results of the optimized transcoding software are shown in Table II. From the data, we can see that a speed-up of approximately 70% in the down conversion process for the Pro1 architecture was achieved. For the overall execution time, an improvement of 26% and 9% were calculated using the MMX-based FDCT and IDCT, respectively. Overall, the total complexity of the Reference, Intra Refresh and Partial Encode architectures have been reduced by 70%, 66%, and 80%, respectively. We note that further reduction in complexity is still possible in some parts, such as the VLC/VLD operations, which are quite time-consuming and more challenging to optimize.

B. Quality Analysis

To compare the quality of the transcoding architectures, we consider two test sequences, *Highway* and *Road*. Both sequences have a resolution of 720x480 interlace and were encoded as MPEG-2 bitstreams at 6Mbps with N=15, M=3. A total of 900 frames are considered for both sequences.

Each bitstream is transcoded using each of the three architectures under consideration. Since B-frames are dropped in the transcoders, the total number of output frames is one-third the original number of frames. The target bit-rate for the transcoders is 384kbps.

Figs. 4 and 5 compare the PSNR of the transcoded outputs on a frame-basis for the *Highway* and *Road* sequences, respectively. It can be observed from these plots that the both the Intra Refresh and Partial Encode architectures provide a close match to the Reference architecture.

To further compare the quality of the transcoded outputs, a side-by-side comparison of frame 144 from the *Highway* sequence is shown in Fig. 6. From these figures, we can observe that the quality of the Intra Refresh and Partial Encode architectures are indeed close in quality to the Reference method.

V. Conclusion

This paper presents a complexity-quality analysis of various transcoding architectures for reduced spatial resolution conversion. We believe that this analysis provides useful information for others working in this area. The proposed architectures offer acceptable alternatives to the reference architecture. As an additional contribution we provide a simplified DCT-domain down-conversion method, which reduces the time for this process by more than 70%.

References

[1] P. Yin, A. Vetro, H. Sun and B. Liu, "Drift compensation architectures for reduced resolution transcoding," *Proc. SPIE Conf. on Visual Communications Image Processing*, San Jose, CA, Jan. 2001.

[2] K. Stuhlmuller, N. Farber, M. Link and B. Girod, "Analysis of Video Transmission over Lossy Channels," *J. Select Areas of Communications*, June 2000.

[3] A. Vetro, H. Sun, P. DaGraca, and T. Poon, "Minimum drift architectures for three-layer scalable DTV decoding," *IEEE Trans. Consumer Electronics*, Aug 1998.

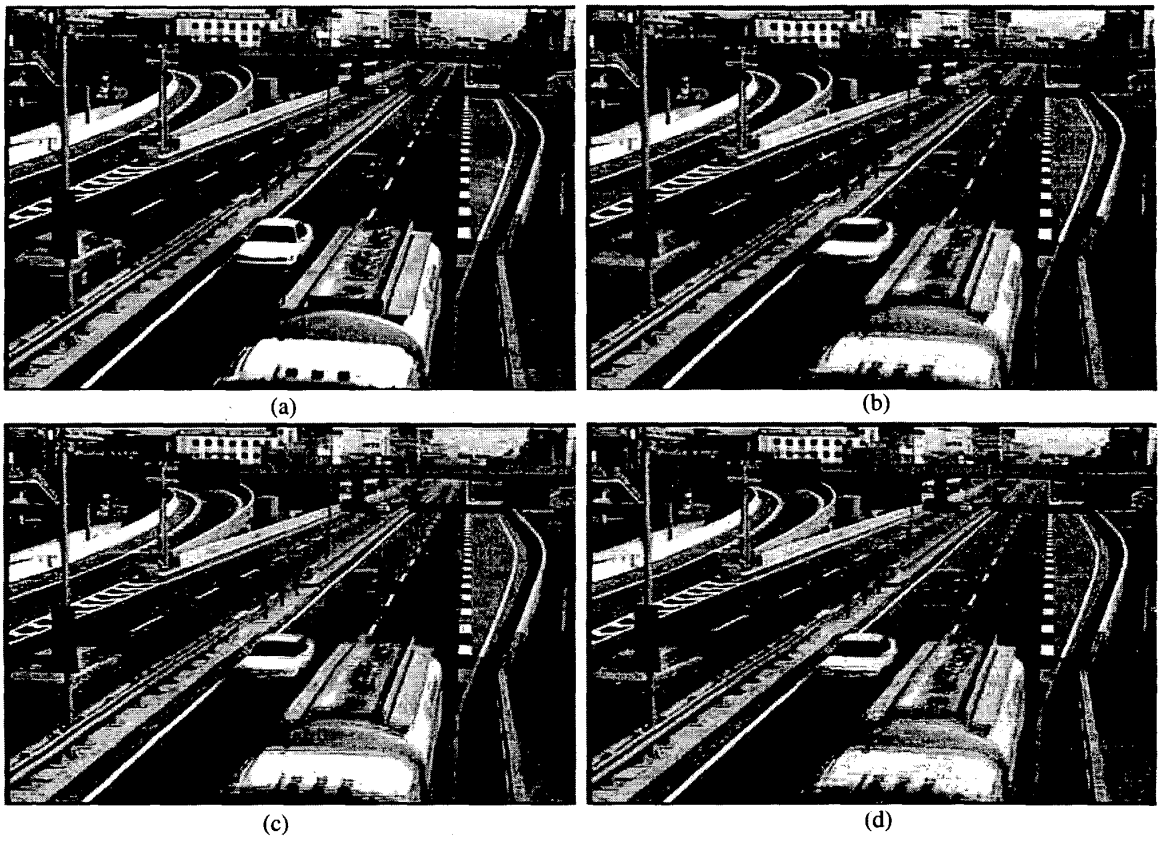
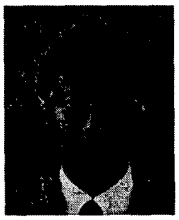


Fig. 6 Quality comparison for frame 144 of *Highway* sequence (a) Original, (b) Reference, (c) Intra Refresh, and (d) Partial Encode.

Anthony Vetro (S'91, M'96) received the B.S., M.S. and Ph.D. degrees in Electrical Engineering from Polytechnic University, Brooklyn, NY.

He joined Mitsubishi Electric Research Laboratories, Murray Hill, NJ, in 1996, and is currently a Senior Principal Member of the Technical Staff. Upon joining Mitsubishi, he worked on algorithms for down-conversion decoding and was later involved in the development of a single-chip HDTV receiver that employed these algorithms. More recently, his work has focused on the encoding and transport of multimedia content, with emphasis on video transcoding, rate-distortion modeling and optimal bit allocation. Since 1997, he has been an active participant in MPEG, contributing to the development of the MPEG-4 and MPEG-7 standards. He is now an editor for Part 7 of the MPEG-21 standard: Digital Item Adaptation.

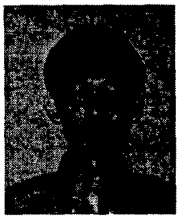
Dr. Vetro has been a member of the Technical Program Committee for the International Conference on Consumer Electronics since 1998, serving as Publicity Chair in 1999 and 2000, and Tutorials Chair in 2002. He serves on the AdCom of the IEEE Consumer Electronics Society and on the Publications Committee of the IEEE Transactions on Consumer Electronics. He is a member of the Technical Committee on Visual Signal Processing and Communications of the IEEE Circuits and Systems Society, and a member of the Editorial Board for the Journal of VLSI Signal Processing Systems, published by Kluwer.



Toshihiko Hata received his B.E. and M.E. degrees in Communication Engineering from Osaka University, and his Ph.D. degree in Graduate School of Science and Technology, Kobe University, in 1980, 1982, and 2001, respectively.

He is a senior researcher of Advanced Technology R&D Center, Mitsubishi Electric Corporation. He has been developing advanced surveillance and monitoring systems utilizing multimedia technologies. His research interests include multimedia systems, video database and multimedia networking.

Dr. Hata is a member of IEEE Computer Society, the Institute of Electronics, Information and Communication Engineers (IEICE), and the Information Processing Society of Japan (IPSJ).



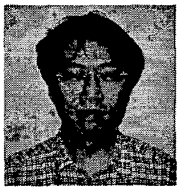
Naoki Kuwahara received his B.E. and M.E. degrees from Osaka City University in 1995 and 1997, respectively.

He is a research engineer of Advanced Technology R&D Center, Mitsubishi Electric Corporation. He has been developing advanced surveillance and monitoring systems utilizing multimedia technologies. His research interests include video processing, multimedia systems and multimedia networking.

Mr. Kuwahara is a member of Information and Communication Engineers (IEICE).

Hari Kalva received a Ph.D. from the Dept. of Electrical Engineering, Columbia University in 2000, M.S. in Computer Engineering from Florida Atlantic University in 1994, and B.Tech. in Electronics and Communications engineering from N.B.K.R. Institute of Science and Technology, S.V. University, India in 1991.

He is currently a consultant with Mitsubishi Electric Research Labs, Murray Hill, NJ, where he is working on different projects including MPEG-2 to MPEG-4 video transcoder optimization. He is a co-founder of Flavor Software Inc., developing MPEG-4 based solutions for content creation and distribution. From Jan 1995 to Aug 1996, he was a research staff member with the ADVENT project at Columbia University working on the design and development of the Columbia's VOD testbed. His standards development activities include participation in the DAVIC and the MPEG-4 Systems standardization. His research interests include multimedia communication and distribution and universal multimedia access.



Shun-ichi Sekiguchi received the M.S. degree in electrical engineering from Waseda University, Japan.

He joined Mitsubishi Electric Corporation in 1992, and has been working on R&D for video coding technologies and appliances. He has been participating in MPEG meetings and has made technical contributions actively in the area of MPEG-4 and MPEG-7. He was a loaned technical staff of Multimedia Laboratories of NTT DoCoMo during 1999-2001, where he contributed to R&D activities on mobile multimedia applications and related standardization activity. His current research interests include signal processing, advanced video coding, and content delivery architecture.

Mr. Sekiguchi is a member of IEICE and ITE.