Video Transcoding: An Overview of Various Techniques and Research Issues

Ishfaq Ahmad, *Senior Member, IEEE*, Xiaohui Wei, *Student Member, IEEE*, Yu Sun, *Student Member, IEEE*, and Ya-Qin Zhang, *Fellow, IEEE*

***Abstract*—One of the fundamental challenges in deploying multimedia systems, such as telemedicine, education, space endeavors, marketing, crisis management, transportation, and military, is to deliver smooth and uninterruptible flow of audio-visual information, anytime and anywhere. A multimedia system may consist of various devices (PCs, laptops, PDAs, smart phones, etc.) interconnected via heterogeneous wireline and wireless networks. In such systems, multimedia content originally authored and compressed with a certain format may need bit rate adjustment and format conversion in order to allow access by receiving devices with diverse capabilities (display, memory, processing, decoder). Thus, a transcoding mechanism is required to make the content adaptive to the capabilities of diverse networks and client devices. A video transcoder can perform several additional functions. For example, if the bandwidth required for a particular video is fluctuating due to congestion or other causes, a transcoder can provide fine and dynamic adjustments in the bit rate of the video bitstream in the compressed domain without imposing additional functional requirements in the decoder. In addition, a video transcoder can change the coding parameters of the compressed video, adjust spatial and temporal resolution, and modify the video content and/or the coding standard used. This paper provides an overview of several video transcoding techniques and some of the related research issues. We introduce some of the basic concepts of video transcoding, and then review and contrast various approaches while highlighting critical research issues. We propose solutions to some of these research issues, and identify possible research directions.**

***Index Terms*—Frequency domain, heterogeneous video systems, H.26X, MPEG-X, motion vector refinement, spatial domain, video transcoding.**

I. INTRODUCTION

VIDEO transcoding performs one or more operations, such as bit rate and format conversions, to transform one compressed video stream to another. Transcoding can enable multimedia devices of diverse capabilities and formats to exchange video content on heterogeneous network platforms such as the Internet. One scenario is delivering a high-quality multimedia source (such as a DVD or HDTV) to various receivers (such as PDAs, Pocket PCs, and fast desktop PCs) on wireless and wireline networks. Here, a transcoder (placed at the transmitter, receiver or somewhere in the network) can generate appropriate bitstream threads directly from the original bitstream without having to decode and re-encode. To suit available network bandwidth, a video transcoder can perform dynamic adjustments in the bit-rate of the video bitstream without additional functional requirements in the decoder. Another scenario is a video conferencing system on the Internet in which the participants may be using different terminals. Here, a video transcoder can offer dual functionality: provide video format conversion to enable content exchange, and perform dynamic bit rate adjustment to facilitate proper scheduling of network resources. Thus, video transcoding is one of the essential components for current and future multimedia systems that aim to provide *universal access*[13].

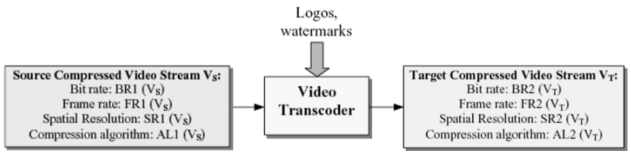
Logos,
watermarks
Source Compressed Video Stream V_S:
Bit rate: BR1 (V_S)
Frame rate: FR1 (V_S)
Spatial Resolution: SR1 (V_S)
Compression algorithm: AL1 (V_S)
Video Transcoder
Target Compressed Video Stream V_T:
Bit rate: BR2 (V_T)
Frame rate: FR2 (V_T)
Spatial Resolution: SR2 (V_T)
Compression algorithm: AL2 (V_T)

Fig. 1. Video transcoding operations.

Currently, several video compression standards exist for different multimedia applications. Each standard may be used in a range of applications but is optimized for a limited range. H.261, H.263, H.263+ designed by ITU (International Telecommunication Unit) are aimed for low-bit-rate video applications such as videophone and videoconferencing. MPEG standards are defined by ISO (International Organization for Standardization). MPEG-2 is aimed for high bit rate high quality applications such as digital TV broadcasting and DVD, and MPEG-4 is aimed at multimedia applications including streaming video applications on mobile devices. As the number of applications increases and various networks such as wireline and wireless integrate with each other, inter-compatibility between different systems and different platforms are becoming highly desirable. Transcoding is needed both within and across different standards to allow the interoperation of multimedia streams. As shown in Fig. 1, adjustment of coding parameters of compressed video, spatial and temporal resolution conversions, insertion of new information such as digital watermarks or company logos, and enhanced error resilience can also be done through transcoding.

Scalable coding is another approach to enable bit-rate adjustment. Traditional scalability in video compression can be of three types: SNR scalability, spatial scalability, and temporal scalability. To achieve different levels of video quality, the video

Manuscript received April 17, 2003; revised April 3, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Suh-Yin Lee.

I. Ahmad and X. Wei are with the Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX 76019 USA (e-mail: iahmad@cse.uta.edu; xhwei@cse.uta.edu).

Y. Sun was with the Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, TX 76019 USA. She is now with the Department of Computer Science, University of Central Arkansas, Conway, AR 72035 USA (e-mail: yusun@mail.uca.edu).

Y.-Q. Zhang is with the Mobile and Embedded Devices Division, Microsoft Corporation, Seattle, WA 98052 USA (e-mail: yzhang@microsoft.com).

Digital Object Identifier 10.1109/TMM.2005.854472

source is first encoded with low PSNR, low spatial resolution, or low frame-rate to form a base layer. The residual information between the base layer and the original input is then encoded to form one or more enhancement layers. Additional enhancement layers enhance the quality by adding the residual information. However, if pre-encoded video is used, scalable coding is inflexible since the number of different predefined layers is limited[1] and the bit-rate of the target video cannot be reduced lower than the bit-rate of the base layer. Thus, scalability alone does not solve the bit-rate adjustment problem.

This paper provides a comprehensive survey of video transcoding techniques. We discuss various research issues arising in transcoding and illustrate them using an *architectural approach*. An architecture, which can be implemented in hardware or software, shows various algorithmic modules, as well as their operations. We present several transcoding architectures with varying levels of efficiency and functional modules. We categorize these architectures and present various examples within a category. We discuss various outstanding issues and provide future directions. The organization of this paper is as follows. Section II provides the basic requirements and functionalities of transcoding. Section III classifies various transcoding architectures and discusses the basic problems. Sections IV and V describe techniques of homogeneous transcoding (with similar standard) and heterogeneous video transcoding (between different standards), respectively. Section VI reviews some research issues. Section VII concludes the paper with final remarks.

II. Requirements and Functionalities

The first and most important challenge in the context of a video conferencing is to provide transcoding on the fly with real-time speed and without any interruption of video flow [17], [49]. There are three basic requirements in transcoding [2], [42]: 1) the information in the original bitstream should be exploited as much as possible; 2) the resulting video quality of the new bitstream should be as high as possible, or as close as possible to the bitstream created by coding the original source video at the reduced rate; 3) in real-time applications, the transcoding delay and memory requirement should be minimized to meet real-time constraints.

A video transcoder can provide several functions, including adjustment of bit rate and format conversion. We illustrate these functionalities and their classification in Fig. 2.

Homogeneous transcoding performs conversion between video bitstreams of the same standard. A simple technique to transcode a video to lower bit rate is to increase the quantization step at the encoder part in the transcoder [35], [43]. Spatial resolution can be done in a number of ways (see Fig. 3) [24]. One possibility is to transcode from normal video to a video containing only the region of interest. Fig. 4 illustrates that a transcoder can down-sample a scene to the object of interest (determined through meta information). This may be done using some *meta* information. In subsampling, filtering and pixel averaging to reduce spatial resolution [24], [30] problems arise when passing motion vectors directly from the decoder to the encoder. Thus, motion vectors need to be refined [32], [37]. Frame-rate conversion is needed when the end-system supports only a lower frame-rate. With dropped frames, the incoming motion information is invalid because they point to the frames that do not exist in the transcoded bitstream.

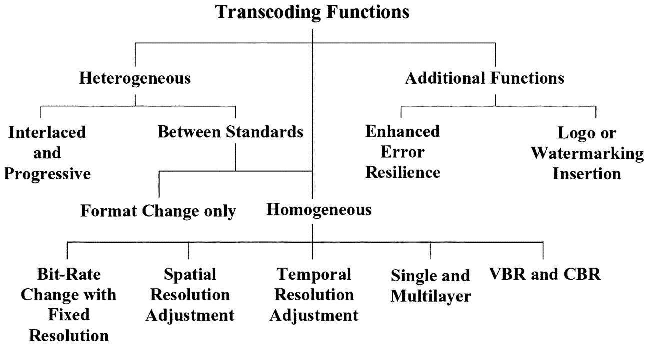
Transcoding Functions
Heterogeneous
Additional Functions
Interlaced and Progressive
Between Standards
Enhanced Error Resilience
Logo or Watermarking Insertion
Format Change only
Homogeneous
Bit-Rate Change with Fixed Resolution
Spatial Resolution Adjustment
Temporal Resolution Adjustment
Single and Multilayer
VBR and CBR

Fig. 2. Various transcoding operations and their classification.

	WorkStation/LAN	PC/Dialup	TV Browser	Gray PDA	B/W PDA
Video					
Color	24 bit color	24 bit color	256 colors	4 bit gray	B/W
Size	352×288	176×144	176×144	160×120	128×96
Bits	64KB	33KB	8KB	4KB	0.6KB

Fig. 3. Various ways of spatial transcoding.

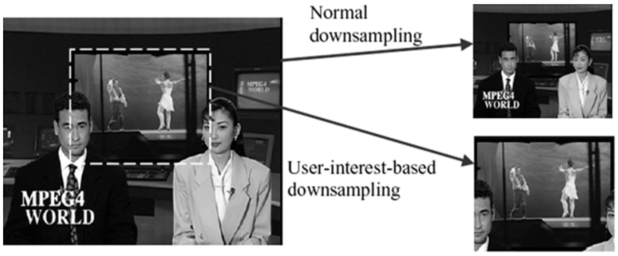

Fig. 4. Transcoding with normal down-sampling and with interest-based object.

A heterogeneous video transcoder provides conversions between existing and future video coding standards. It provides syntax conversion between these standards. Further, a heterogeneous video transcoder may also provide the functionalities of homogeneous transcoding. Transcoding may include additional functions such as error-resilience and logo or watermark insertion. These functions will be described in the paper subsequently.

III. Video Transcoding Architectures

A. Open-Loop Transcoder and Closed-Loop Transcoder

The most straightforward transcoding architecture is to cascade the decoder and encoder directly as shown in Fig. 5(a). In this architecture, the incoming source video stream (V_S) is fully decoded, and then re-encoded the decoded video into the target video stream (V_T) with desirable bit-rate or format, with no degradation in the visual quality due to transcoding. The more

[1] MPEG-4 FGS allows more flexible control.

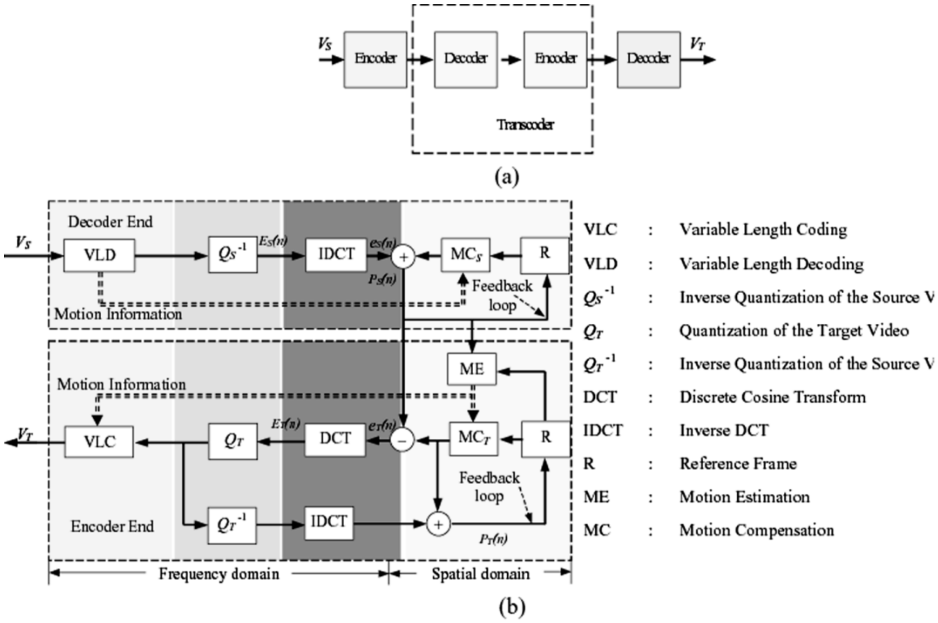

Fig. 5. Cascaded decoder and encoder transcoder: (a) function and (b) details.

detailed manifestation of the cascaded transcoder is shown in Fig. 5(b).

In predictive coding, a coded video frame is predicted from other frames and only the prediction error (residue error) is coded. For the decoder to operate properly, the video frames reconstructed and stored in decoder predictor must be exactly the same as those in the encoder predictor. Decoding of a transcoded video would result in errors if the predictors of the decoder are different from those of the original encoder; these errors would accumulate with time through the whole group of pictures (GOP). The error accumulation resulting from encoder/decoder predictor mismatch is called "drift" error [7].

In order to understand how the drift error comes, let us consider the architectures of the cascaded decoder and encoder transcoder in Fig. 5(b) and an open-loop transcoder with re-quantization scheme in Fig. 6.

From Fig. 5(b), we can get

$$P_S(n) = e_S(n) + \mathrm{MC}_S(P_S(n-1)) \quad (1)$$

$$e_T(n) = P_S(n) - \mathrm{MC}_T(P_T(n-1)) \quad (2)$$

$$\begin{aligned} E_T(n) &= \mathrm{DCT}(e_T(n)) \\ &= \mathrm{DCT}(P_S(n) - \mathrm{MC}_T(P_T(n-1))) \quad \text{from (2)} \\ &= \mathrm{DCT}(e_S(n) + \mathrm{MC}_S(P_S(n-1)) \\ &\quad - \mathrm{MC}_T(P_T(n-1))) \quad \text{From (1)} \\ &= \mathrm{DCT}(e_S(n)) + \mathrm{DCT}(\mathrm{MC}_S(P_S(n-1)) \\ &\quad - \mathrm{MC}_T(P_T(n-1))) \\ &= E_S(n) + \mathrm{DCT}(\Delta\mathrm{MC}) \quad (3) \end{aligned}$$

where

$$\Delta\mathrm{MC} = \mathrm{MC}_S(P_S(n-1)) - \mathrm{MC}_T(P_T(n-1)).$$

In Fig. 6, the open-loop transcoder starts with the de-quantization of the DCT coefficients using the original quantizer levels (Q_S). These coefficients are re-encoded with a different quantizer (Q_T) for output bit rate reduction. From Fig. 6, we get

$$E'_T(n) = E_S(n) \quad \text{or} \quad E_S(n) = E'_T(n). \quad (4)$$

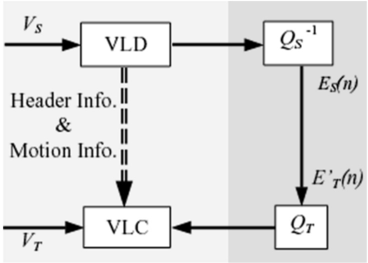

Fig. 6. Open-loop transcoder with re-quantization scheme.

Comparing (4) with (3), the drift error of frame $n, d_e(n)$ can be expressed as

$$E_T(n) = E'_T(n) + \mathrm{DCT}(\Delta\mathrm{MC})$$

and

$$d_e(n) = E_T(n) - E'_T(n) = \mathrm{DCT}(\Delta\mathrm{MC}).$$

We can see that $d_e(n)$ represents an error in the reference picture that is used for motion compensation (MC). This error may be caused by re-quantization, elimination of some nonzero DCT coefficients, or by integer truncation [47]. In video compression, Intra-coded frames (I frames) are encoded without reference frame, MC is not needed in encoding I frames, so the transcoding of I frames is not subject to the drift. Bi-directionally predictive coded frames (B frames) are not used for predicting future frames [7]. Therefore, the transcoding of B frames does not contribute to the propagation and accumulation of the drift. The drift error is only caused by the transcoding operation of INTER coded frames, and can accumulate through a GOP, the quality deterioration gradually increases until the next I-frame refreshes the video scene [1], [3], [47].

Open-loop transcoders contain no feedback loop in the transcoding architecture for compensating the drift error. They aim for minimum transcoding complexity, and thus only modify the encoded DCT coefficients to reduce the overall bit rate [1]. Open-loop transcoders include selective transmission

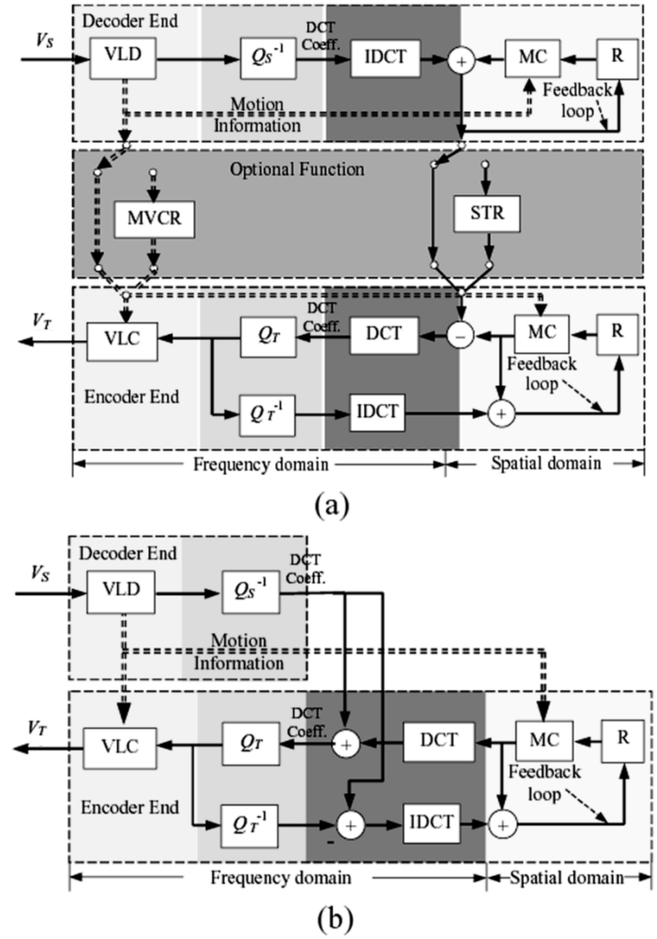

Fig. 7. SDTA with motion vector reused. (a) SDTA with STR. (b) Simplified SDTA without STR.

and re-quantization. Selective transmission [8], [34] discards high frequency DCT coefficients of a block. Re-quantization architectures re-quantize the motion compensated residue errors to adapt to the bit-rate requirement [34], as shown in Fig. 6. Both approaches operate in the frequency domain and are rather simple to implement. Both of them change the residue error and alter the content in the decoder predictor. Therefore, when the decoder decodes the video processed by an open-loop algorithm, the predictors would be different from those of the original encoder, leading to drift errors.

Closed-loop transcoders contain a feedback loop in the transcoding architecture in order to correct the transcoding distortion (see Figs. 7 and 8 as examples) by compensating the drift in the transcoder [2], [17], [34]. We will focus on the closed-loop architectures in the following subsections and classify them in various categories.

B. Spatial-Domain Video Transcoding

Fig. 5(b) shows a spatial-domain transcoding architecture (SDTA) that can perform dynamic bit-rate adaptation via the rate-control at the encoder side. This architecture is flexible since the decoder-loop and the encoder-loop can be totally independent of each other (e.g., they can operate at different bit-rates, frame-rates, picture resolutions, coding modes, and even different standards). This architecture is drift-free, but its computational complexity is high for real-time applications.

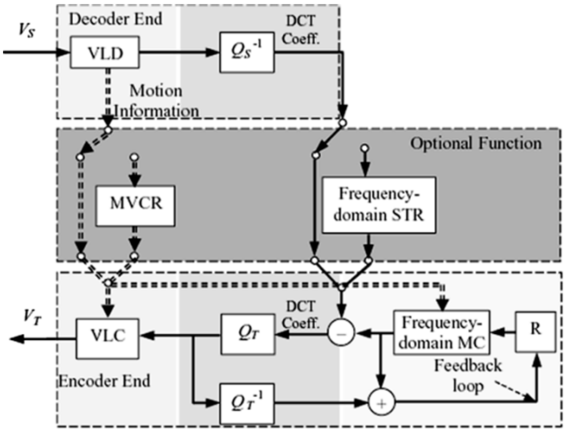

Fig. 8. Frequency domain transcoder architecture (FDTA).

Since a pre-encoded video stream arriving at a transcoder already carries useful information such as the picture type, motion vectors (MV), quantization step-size, bit-allocation statistics, etc., it is possible to construct transcoders with different complexity and performance in terms of coding efficiency and video quality. Intuitively, most of the motion information and the mode decision information received in the video decoder can be reused in the video encoder without introducing significant degradation on visual quality. Thus, motion estimation, the most time-consuming operation in video encoding which accounts for 60%–70% of the encoder computation [30], is avoided. This leads to an SDTA that can reuse MVs [shown in Fig. 7(a)]. This architecture saves the motion estimation operation, which is the most time-consuming module. The pre-encoded source video is decoded in the spatial-domain by performing variable-length decoding (VLD), inverse quantization (Q_S^{-1}), IDCT, and motion compensation. In the encoder, the motion compensated residue errors are encoded into frequency-domain through DCT, re-quantization (Q_T), and variable length coding (VLC). The motion compensation operation at the encoding end is also performed in the spatial domain for the prediction operation. The MV reuse approach is useful in complexity reduction for motion estimation in video transcoding [17].

The architectures in Fig. 7(a) and the Figs. 8 and 9 include two optional functional blocks placed between the decoder and encoder: spatial/temporal resolution reduction (STR) module and MV composition and refinement (MVCR) module. STR allows the source video to be transcoded to target video with different spatial/temporal resolution accordingly. MVRC is needed to adjust the MVs when STR is applied. When transcoding without spatial/temporal resolution reduction, the SDTA architecture can be further simplified into Fig. 7(b), [2], in which only one feedback loop is employed.

C. Frequency-Domain Transcoding

Exploiting the structural redundancy of the architecture in Fig. 7 and the linearity of the DCT/IDCT, a structurally simpler but functionally equivalent frequency-domain transcoding architecture is possible [8], which can be further simplified [2], [4], [23], as shown in Fig. 8. In this architecture, only VLD and inverse quantization are performed to get DCT value of each block in the decoder end. At the encoder end, the motion

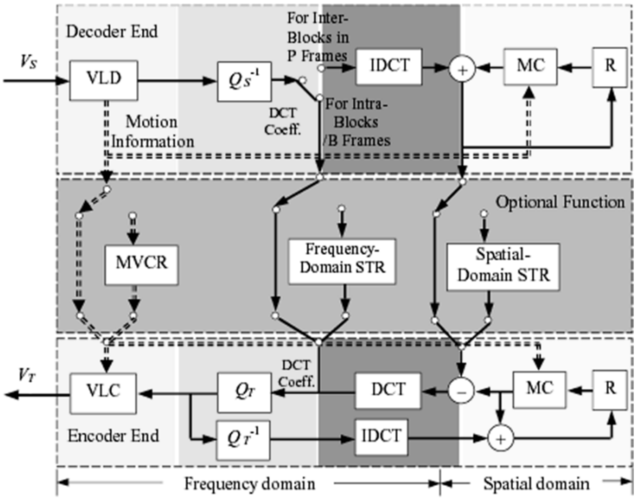

Fig. 9. Hybrid-domain transcoding architecture (HDTA).

compensated residue errors are encoded through re-quantization, and VLC. The reference frame memory in the encoder end stores the DCT values after inverse quantization, that are then fed to the frequency-domain MC module to reduce drift error. This is referred to as frequency-domain transcoding architecture (FDTA).

In this architecture, motion compensation is performed in the frequency domain using a MV reusing algorithm. Detail frequency domain MC algorithm can be found in [2] and [31]. An FDTA may need less computation but suffer from the drift problem due to nonlinearity operations, which includes subpixel motion compensation, and DCT coefficients clipping during MC. FDTAs also lack flexibility and are mostly fitted for bi-rate transcoding. Recently, researchers have studied frequency-domain motion estimation that may eliminate some of these constraints [18].

D. A Hybrid-Domain Transcoding Architecture

Various transcoding algorithms provide tradeoff between the computational complexity and reconstructed video quality. In order to reduce the computational complexity while maintain the reconstructed video quality, ME should be omitted and DCT/IDCT should be avoided if possible. For example, the architecture in [45] uses MC for P frames only. I frames are intra coded, which need no ME and MC, and thus, IDCT/DCT for I frames can be omitted in principle. But since I frames are the anchors for subsequent P and B frames, the IDCT at the decoder stage, inverse quantization and IDCT at the encoder stage for I frames are still needed to reconstruct the reference frames, while DCT at the encoder stage can be omitted. Since P frames are also the anchors for the following P and B frames, MC, DCT, and IDCT cannot be omitted. For B frames, which are not the reference frames for the subsequent frames, drift error generated in B frames would not propagated through the video sequence, so MC of B frames can be removed without introducing significant degradation on visual quality of reconstructed pictures. Thus, DCT/IDCT in all B frames can be omitted, and the transcoding of B frames can be directly done in the DCT domain.

We can further reduce the transcoding delay without degrading the video quality in this architecture. P frames with frequent scene changes and rapid motion may contain a large number of INTRA blocks. One can further omit the IDCT/DCT and MC operation of these INTRA blocks in P frames. In other words, blocks of I and B pictures and INTRA blocks of P pictures are transcoded in frequency-domain, the spatial-domain motion compensation is done only when the block is inter block in P frames. We call this transcoding architecture as hybrid domain transcoding architecture (HDTA), as shown in Fig. 9.

From the simulation results in [45], compared to SDTA with MV reused, the HDTA has less complexity, which speeds up the transcoding operation, but has the expense of some degradation in picture quality. Compared with frequency domain transcoder, this transcoder performs DCT/IDCT and MC when the block is INTER block in P frames, which may increase the transcoding delay but has better visual quality.

IV. Homogeneous Video Transcoding

Homogeneous transcoding performs conversion between video bitstreams of the same standard. A high quality source video may be transcoded to a target video bitstream of lower quality, with different spatial/temporal resolutions, and different bit rates. The following subsections describe some of the research issues in homogeneous transcoding.

A. Reducing Bits With Fixed Resolution

For fixed spatial and temporal resolution, we can reduce the bit rate using the following two techniques:

Re-Quantization: A simple technique to transcoding a video to lower bit rate is to increase the quantization step at the encoder part in the transcoder [26], [35], [43]. This decreases the number of nonzero quantized coefficients thus decreasing the amount of bits in the outgoing bitstream. Requantizing is a good compromise between the complexity and reconstructed image quality, and can control the bit-rate reduction.

Selective Transmission: Since most of the energy is concentrated at the lower frequency band of an image, discarding (truncating) some of the higher ac frequency coefficients [1], [30], [34] can preserve the picture quality, but may introduce a blocking effect in the reconstructed target video.

B. Spatial Resolution Reduction

Reduction in spatial resolution can obviously lower the bit rate. In this subsection, we describe some common video transcoding techniques.

Filtering and Subsampling: Filtering and subsampling are common techniques to reduce spatial resolution [24], [30], [48]. Shanableh [30] proposed a filter that can be used both horizontal and vertical directions for luminance and chrominance; the image is then down-sampled by dropping every alternate pixel in the both horizontal and vertical directions.

Pixel Averaging: Pixel averaging [30] is another common technique in which every m × m pixels are represented by a single pixel of their average value. Pixel averaging is the simplest method but the reconstructed pictures may become blurred.

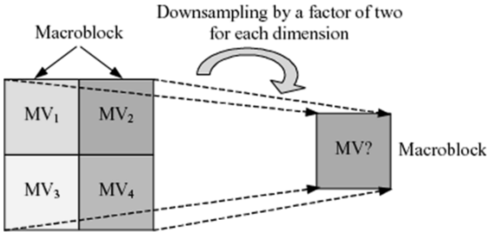

Fig. 10. Four motion vectors being down sampled to one.

Discarding High Order DCT Coefficients: To reduce the spatial resolution, some algorithms [30], [34], [38] remove the higher ac frequency coefficients. When transcoding four macroblocks into one macroblock, Tan [38] adopted a DCT decimation method in which every four input blocks of 8×8 pixels, corresponding to an area of 16×16 pixels, are first DCT transformed. DCT decimation delivers better quality for image down-sampling over filtering or pixel-averaging, but for large bit rate reduction greater than 25%, this method produces poor-quality blocky pictures [34].

Motion Vector Composition and Refinement: In subsampling, filtering and pixel averaging to reduce resolution, problems arise when passing MVs directly from the decoder to the encoder. Fig. 10 shows how multiple MVs are merged to a single MV when resolution is reduced by a factor of two in each dimension. The single MV, if composed properly, should reflect the precision of four MVs as much as possible.

Various methods are proposed to compose a single motion vector from multiple motion vectors:

1) **Random**: Selecting one of the incoming MVs in random [3]. The method is fast but inefficient.
2) **Mean**: Taking the average or mean of some MVs. These MVs can be in the corresponding area in the source video, or have the same direction [32], [48], or have some correlation between the neighboring macroblocks [30]. This technique may yield poor results if the magnitude of one of the input MVs is significantly larger than the rest.
3) **Weighted Average (WA)**: Taking the weighted average of the incoming MVs, where each MV is weighted by the spatial activity of the perspective prediction error [32], [33]. This method is prone to noise in candidate MVs and may bias the MV when original MVs are aimed in various directions.
4) **Weighted Median (WM)**: Extracting the motion vector situated in the middle of the rest of the MVs by computing the Euclidean distances between each MV [3], [30], [46]. This method yields good performance, but requires substantial computation in determining the median MV.
5) **DCmax**: Composing an MV by the corresponding MV with maximum dc coefficients of residual blocks in the source video [5]. This method takes a little more computation than the Mean, but yields better performance than the Mean and the WA.

In all of these methods, the magnitude of the new MV is scaled down by a factor (normally by 2) to reflect the spatial resolution transcoding. Takahashi [37] applied a MV composition technique for arbitrary ratio spatial resolution scaling by selecting a single MV from multiple MVs of MPEG-2 MBs. Since simple motion-vector reuse schemes may introduce considerable quality degradation in many applications, and the MV composition methods are suboptimal, MV refinements techniques at the encoder end are proposed [3], [30], [37]. Since the passed MVs will almost be the same as the recalculated ones, we can refine them to get more appropriate values. The refinement can be done in a small search window around the passed MV [49].

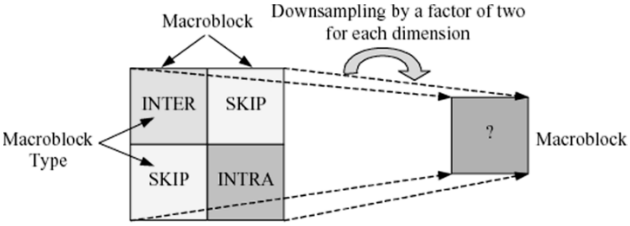

Fig. 11. Four macroblock types downsampled to one.

MB Coding Mode Decision: The decision modes obtained from the high-quality original bitstream are not optimum for re-encoding at the reduced rate in rate reduction by requantizing. In the worst case, macroblocks might be coded in the wrong mode. For instance, a macroblock that should be SKIPPED at the encoder of the transcoder, due to a larger quantization making all coefficients zero, could be coded as an INTER macroblock since it was coded as an INTER macroblock at the transmitter.

To solve this problem, Sun [34] proposed to always re-evaluate the macroblock type at the encoder of the transcoder. Bjork [3] adopted the following method for the macroblock type.

a) If it was coded as INTRA (at the transmitter) again code it in INTRA.
b) If it was coded as SKIPPED again code it as SKIPPED.
c) If it was coded in INTER, check to see if all coefficients are zero and if they are coded as SKIPPED, else check again whether the macroblock has to be coded in INTRA or INTER mode.

Problems also arise when passing the coding type of four macroblocks as one directly from the decoder to the encoder when doing a resolution reduction by a factor of two in each dimension; Fig. 11 illustrates this problem.

The procedure to handle such a situation [24] is as follows.

1) If there exists at least one INTRA type among the four MBs then pass it as INTRA; pass as INTER type if there is no INTRA MB and at least one INTER MB; pass as SKIP if all MBs are of the SKIP type.
2) Re-evaluate the MB types in the encoder.

If the four original macroblocks are all intra-coded, the new macroblock is also intra-coded. Otherwise, the new macroblock is inter-coded. Those intra-coded macroblocks, which do not provide any motion information, are viewed as inter-coded blocks with zero-valued MV [32].

C. Temporal Resolution Reduction

Reduction in frame rate may save bits that can be used in the remaining frames to maintain acceptable overall picture quality for each frame. In addition, frame-rate conversion is needed when the end-system supports only a lower frame-rate. With dropped frames, the incoming MVs are not valid because they

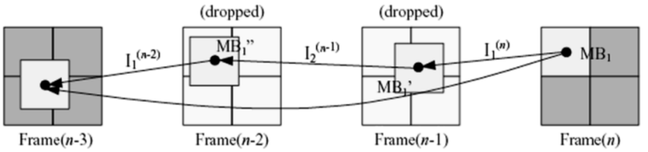

Fig. 12. FDVS composition scheme.

point to the frames that do not exist in the transcoded bitstream. One has to derive a new set of MVs taking into account the MVs of the dropped frames.

Bilinear Interpolation: Hwang [12] developed a bilinear interpolation method to estimate the MVs from the current frame to the previous nonskipped frame given the MVs between every adjacent frame are known. The newly located position based on this interpolated MV serves as the new search center, thus reducing the search range. The number of skipped frames and the accumulated magnitudes of their MVs decide the size of the search area.

Forward Dominant Vector Selection (FDVS): As shown in Fig. 12, this method proposed by Youn and Sun [49] selects the dominant MV (defined as the MV carried by a macroblock that has the largest overlapping segment with the block pointed by the incoming MV) from the four neighboring macroblocks. The best-matched area pointed by the MV of the current macroblock occurring after a dropped frame overlaps with at most four macroblocks in the previous dropped frame. The MV of the macroblock with the largest overlapping portion is selected and added to the current MV. This process is repeated each time a frame is dropped until a new set of MVs is composed for the first encoded frame after the frame dropping. Due to the dropped frames, the distance between the current frame and previous anchor frame becomes larger, and the extracted macroblock type may not be suitable.

Telescopic Vector Composition (TVC): This technique [30] accumulates all the MVs of the corresponding macroblocks of the dropped frames and add each resultant composed MV to its correspondence in the current frame. This technique also carries out new macroblock decision and MV refinement.

Activity-Dominant Vector Selection (ADVS): This algorithm [5] utilizes the activity of the macroblock to decide the choice of the MV. The activity information of a macroblock is represented by counting the number of nonzero quantized DCT coefficients of covered 8 × 8 residual blocks; other statistics, such as the sum of the absolute values of DCT coefficients, etc. These quantities are proportional to the spatial-activity measurement. The higher the activity of the macroblock, the more significant will be the motion of the macroblock. Since the quantized DCT coefficients of prediction errors are available in the incoming bitstream of transcoder, the computation for counting the nonzero coefficients is very little.

Among the above methods, the bilinear interpolation needs multiple memories to store the incoming MVs of all the dropped frames. The FDVS approach can achieve better performance with less computation than the bilinear interpolation. Another advantage of FDVS over the bilinear interpolation scheme is that when multiple frames are dropped, it can be processed in the forward order, eliminating the need for multiple memories needed to store the incoming MVs of all the dropped frames. TVC needs less computation than FDVS, while its picture quality is little lower or the same as that of FDVS. In most cases, the FDVS and ADVS would choose the same MV. ADVS is superior to the FDVS, especially for the high motion case but has a little more complexity than FDVS. Fung and Chan [10] proposed a frequency-domain frame rate reduction transcoding scheme with a direct addition (DA) of the DCT coefficients for MBs, which are coded without MC, to deactivate most of the complex modules of the transcoder.

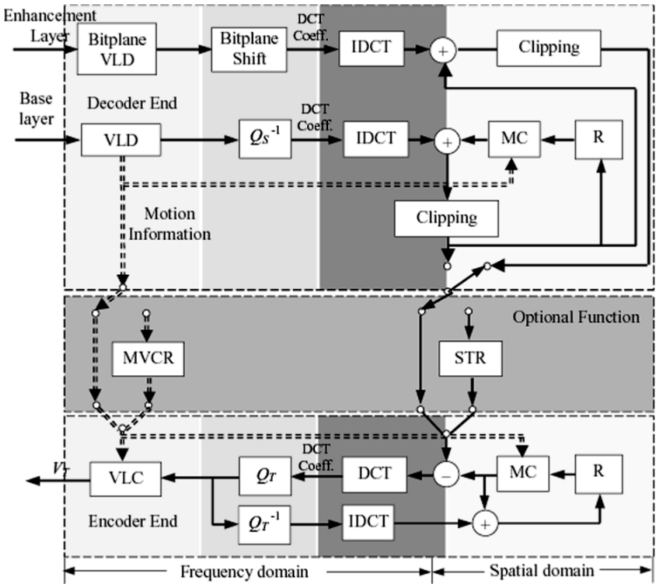

Fig. 13. FGS transcoder with MV reused.

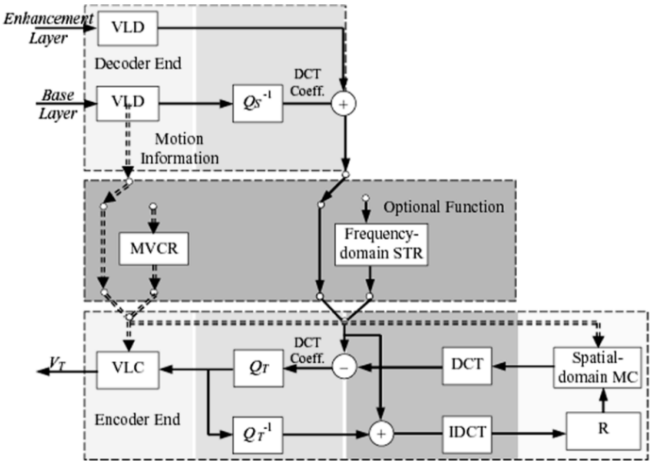

Fig. 14. Partial FDTA FGS transcoder.

D. Transcoding Between Multiple and Single Layers

Multi-layered coding provides scalable video quality [19]. For example, MPEG-4 has standardized a scalable coding scheme referred to as fine granularity scalability (FGS) [15]. With this scheme, a base layer and an enhancement layer bitstreams are generated. The base is coded with the usual motion-compensated DCT techniques and is compliant to the Advanced Streaming Profile, while the enhancement is coded with a bit-plane coding method that is supported in the FGS

TABLE I
KEY FEATURES OF VARIOUS VIDEO COMPRESSION STANDARDS

Features	H.263+, H.263++	MPEG-2	MPEG-4	H.264
Profiles	H.263 Annexes Support	Simple, Main, 4:2:2 SNR Scalable, Spatially Scalable, High	Simple, Simple Scalable, Core, Main, N-bit, Fine Granularity Scalable, etc.	Baseline, Main, Extended
Basic Approach	Block Based	Block Based	Object Based	Block Based
Processing Unit	Macroblock	Macroblock	Macroblock	Macroblock
Data Component	Luminance and Chrominance	Luminance and Chrominance	Luminance and Chrominance	Luminance and Chrominance
Picture Types	I, P, B, PB	I, P, B	I, P, B	I, P, B, SP, SI
I frames	Fewer (compression)	More (random access)	More (random access)	Fewer (compression)
Entropy Coding	VLC, SAC	VLC	VLC	UVLC or CABAC
ME Accuracy	½ pixel	½ pixel	Up to ¼ pixel	¼ pixel
Motion Vectors	Can point to outside picture	Inside reference picture only	Can point to outside picture	Can point to outside picture
Block Transform	8×8 DCT	8×8 DCT	8×8 DCT	4×4 Integer DCT
Multiple Reference Frames	Supported Annex U	Not supported	Not supported	Supported
Motion Estimation Block Size	8×8 or 16×16	8×8 or 16×16	8×8, 16×16, 16×8	(16×16, 16×8, 8×16, 8×8, 8×4, 4×8, 4×4)
Intra Block Prediction	NO	NO	NO	YES
Support formats	Progressive	Progressive & interlaced	Progressive & interlaced	Progressive & interlaced
Prediction Modes	Frame only	Field, frame	Field, frame	Field, frame
Support resolutions	SQCIF, QCIF, CIF, 4CIF, and 16CIF	From SQCIF to HDTV	Typically from SQCIF to 'Studio' resolutions (4k × 4k pixels)	From SQCIF to HDTV
De-block Filter	Annex J, 8×8 block boundary	NO	YES, 8×8 block boundary, post filter	YES, 4×4 block boundary, in loop

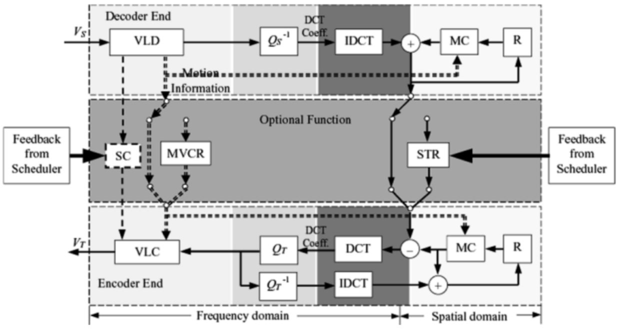

Fig. 15. Heterogeneous video transcoder.

Profile. The key advantage to this coding method is that it allows the enhancement layer bitstream to be truncated into any number of bits within each frame to provide partial enhancement proportional to the number of bits decoded for each frame [19]. The quality of the reconstructed frame is proportional to the number of enhancement bits received [22].

In [21], Liang and Tan proposed an SDTA to cascaded FGS decoder and single-layer encoder with MV reused as shown in Fig. 13. In [22], Lin proposed a simplified partial FDTA employing the information provided by the enhancement video to improve the quality of the target video, as shown in Fig. 14.

V. HETEROGENEOUS VIDEO TRANSCODING

A heterogeneous video transcoder provides conversions between various standards, for instance, MPEG-2 to H.263 transcoder, MPEG-2 to MPEG-4 transcoder, H.263 to MPEG-4 transcoder, etc., Further, a heterogeneous video transcoder may also provide the functionalities of homogeneous transcoding [9], [11], [30]. Several techniques aimed for homogeneous transcoding can also be exploited in heterogeneous transcoding.

A. Main Issues in Heterogeneous Transcoding

A heterogeneous transcoder needs a syntax conversion module, and may change the picture type, picture resolution, directionality of MVs, and picture rate. A heterogeneous transcoder must adjust the features of the incoming video to enable the features of the outgoing video [9]. Due to spatial-temporal subsampling, and different encoding format of the output sequence, the encoder and decoder motion compensation loops in a heterogeneous transcoder are more complex. A number of differences exist among various video coding standards (see Table I).

B. Generic Heterogeneous Transcoder

A generic heterogeneous transcoder is showed in Fig. 15. In this architecture, syntax conversion (SC) is needed to convert the syntax of source video to that of the target video. A higher resolution decoder decodes the incoming bitstream. The extracted MVs are then post-processed according to the desired output encoding structure, and if required, they are properly scaled down to suit the lower spatial-temporal resolution encoder. In

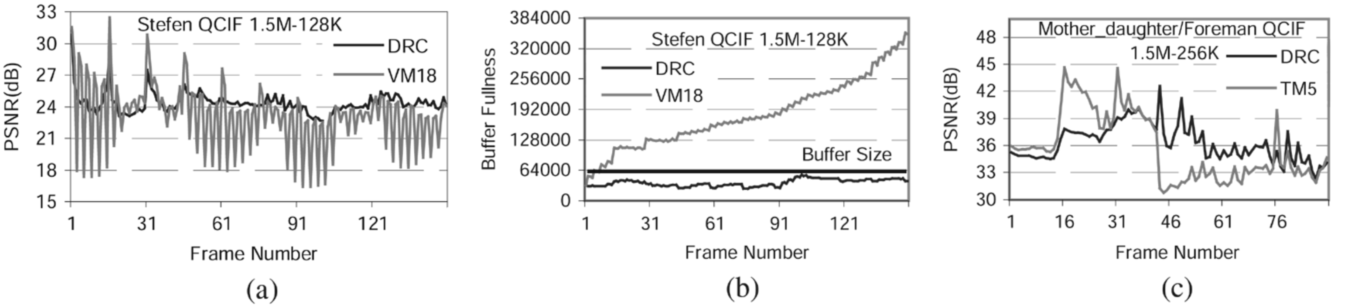

Fig. 16. Transcoding experimental results. (a) PSNR with frame skipping. (b) Buffer fullness with frame skipping. (c) Scene change at frame 42.

case post-processing is not sufficient, the extracted MVs are refined to improve the encoding efficiency. The decoded pictures are accordingly down-sampled spatially or temporally, and the down-sampled images are encoded with the new MVs. Since the incoming MVs are re-employed, and other encoding decisions, such as macroblock types can be extracted from the incoming bitstream, the architecture of this transcoder can be further simplified.

In this architecture, the MVs of the incoming bitstream are employed in the outgoing one, the extracted MVs have to be converted to be compatible with the encoding nature of the output bitstream. Note that the nature of extraction of the MVs and their usage depend on the picture type. The algorithm proposed in [30] assuming the motion between the pictures is uniform, such that the forward and the reverse MVs are images of each other, or an inter-frame MV is a scaled version of a larger picture distance and so on. In case no MV is found, one might either use a (0, 0) MV or in the worst-case intra-frame code the underlying macroblock. In [30], encoding format of MPEG-1, 2 is first transcoded into H.261/H.263, the algorithm adopted the incoming motion parameters of a sub GOP of up to three frames to produce several candidate MVs for the outgoing picture. Then all the estimated MVs are compared, and the one that gives the least coding error in terms of sum of absolute differences (SAD) is chosen. The best MV was then refined by half-pixel (or one pixel) motion estimation to produce near-optimum results. Second, transcoding from encoding format of H.261/H.263 into H.263 PB frames. The new MVs of both P and B frames were calculated in a similar manner to that used in the first case. After the new MVs are obtained, and spatial or temporal reduction is performed, the encoder in the heterogeneous transcoder can code every picture according to the picture type of the new format.

VI. Related Research Issues

This section discusses some related research issues that can benefit transcoding.

A. Rate Control in Transcoding

The goal of rate control in video coding and transcoding is to achieve a target bit-rate with good and consistent visual quality. Rate control for transcoding a pre-compressed bitstream may exploit certain information extracted from compressed bit streams to assist in bit-rate regulation. This information can be motion estimation, input bit rate/output bit rate, INTRA/INTER mode decision, and picture complexity. In the widely used MPEG-2 Test Model 5 (TM5) [14], a picture complexity measure characterizes the difficulty in coding a picture, such that the target number of bits for coding that picture is proportional to its complexity. One can compute the complexity measures of pictures from an input bitstream, and then use them for bit-allocation in recoding these pictures. The rate control algorithm used in [39] belongs to this category. However, the complexity measures defined in TM5 are dependent on the coding bit-rate. Therefore, the complexity measure calculated from the input video bitstream at the input bit-rate may not be suitable to directly serve as the complexities for coding the pictures at the output bit-rate. Xin [46] proposed a scheme to estimate the picture complexities of an output video using the coding statistics computed from the input video stream. A Lagrangia-based rate-distortion optimization technique has been exploited for bit allocation during transcoding [2], but it is suitable for nonreal-time application due to high computational complexity.

In MPEG-2 to MPEG-4 transcoding, the conventional TM5 and MPEG-4 VM18 [20] rate control strategy are not suitable for transcoding due to the following reasons: TM5 cannot resort to changing the temporal coding parameter to obtain the desired bit rate. VM18 adopts Mean Absolute Difference (MAD) as coding complexity to estimate the target bit allocation, but it is difficult to obtain MAD in video transcoding, since MC and ME are normally not performed (to reduce transcoding complexity) and most transcoding processing is performed in the frequency domain.

In [36], we presented an effective rate control scheme for MPEG-2 to MPEG-4 transcoding using a DCT-based encoding complexity. The proposed algorithm (named DRC) determines the spatial coding parameters to realize very low target birate transcoding. In addition, the algorithm effectively minimizes the buffer overflow or underflow. Experimental results presented in Fig. 16(a) and (b) shows the picture quality and buffer fullness, respectively, for Stefen video sequence (which contains rapid motion). Fig. 16(c) shows how the transcoder reacts when there is sudden scene change. Tables II and III show the effects of rate control on transcoding using TM5, VM18 and our algorithm, without and with frame skipping, respectively.

B. Error-Resilient Transcoding for Video Over Wireless Channel

Error resilience is the ability of a bitstream to accommodate to the channel conditions and yet produce acceptable quality. An error-resilient transcoder can improve video quality in the

TABLE II
TRANSCODING EXPERIMENTAL RESULTS WITHOUT FRAME SKIPPING

Sequence	Conversion (Bps)	Actual Bi-rate (Kbps)			PSNR (dB)		
		TM5	VM18	DRC	TM5	VM18	DRC
Container (QCIF)	1500K~1024K	1024.77	1020.15	1024.44	28.27	28.66	28.69
Mobile (QCIF)	1500K~1024K	1024.77	1020.15	1024.44	28.27	28.66	28.69
Stefan (CIF)	3000K~1024K	1026.29	1142.11	1023.44	29.39	28.48	29.80
Table (CIF)	3000K~384K	384.90	467.07	383.85	29.11	29.15	29.26
News (CIF)	3000K~512K	513.82	531.47	511.11	35.78	35.32	36.09
Flower (CCIR601)	6000K~3500K	3515.49	3945.09	3502.61	27.11	25.93	27.47

TABLE III
TRANSCODING EXPERIMENTAL RESULTS WITH FRAME SKIPPING

Sequence	Conversion (Bps)	Actual Bi-rate (Kbps)		PSNR (dB)		Skipped Frames	
		VM18	DRC	VM18	DRC	VM18	DRC
Foreman (QCIF)	1500K~84K	84.18	84.54	27.97	28.22	33	0
Stefan (QCIF)	1500K~128K	141.68	128.13	23.19	24.33	49	1
News (CIF)	3000K~256K	254.56	256.03	31.47	31.77	25	0
Mobile (CIF)	3000K~640K	641.92	638.64	22.81	23.54	34	20

presence of errors while maintaining the input bit rate over wireless channels. Reyes [28] describe a method to maintain quality for video transcoding for wireless channels. The method is based on analytical models that characterize how corruption propagates in a video that is compressed using motion-compensated encoding and subjected to bit errors. Dogan [6] used adaptive intra refresh (AIR) and feedback control signaling (FCS) methods to improve the error resilience of compressed video in the transcoding operation. The AIR method prevents the error propagation within a video stream by using a pre-determined number of intra MBs. The FCS method uses a feedback signal from receiver to adapt the encoding scheme.

C. Logo Insertion Scheme in Video Transcoding

A useful feature in video transcoding is to insert a logo in the bitstream. Logos can be transparent or nontransparent. Since a logo affects only a part of the video picture sequence and a small portion of the picture, one can reuse the incoming MVs during logo insertion for the part unaffected by logo. For MBs that are affected by the logo, the reference MBs may no longer be the best matching MB [29]. Various strategies can be developed for modifying the MVs and changing the macroblock prediction mode in the log areas and the logo-affected parts separately. Panusopone [27] proposed several schemes to adapt the MVs to make the effect of logo insertion smaller. To ensure good perceptual video quality, the quantization scales in the logo area can also be modified [44].

D. Object-Based Transcoding

A key advantage of object-based coding schemes is that the quality of each video object may be varied based on its complexity and available bandwidth [40], [41]. Various techniques can be employed to reduce the rate depending on the ratio of incoming to outgoing rate, and since the goal is to provide the best overall quality for objects of varying complexity, the degradation of each object need not be the same. Under severe network conditions, a transcoder may consider dropping less relevant objects from the scene. Vetro proposed a dynamic programming approach and a meta-data based approach in object-based transcoding [40], showing promising results.

E. Transcoding to H.264

The new H.264 is substantially different from previous MPEG and ITU standards (see Table I). It is a joint effort of MPEG and ITU with the first version finalized in May 2003 [16], and aims to deliver far more high-quality video at all bit rates, including low bit rates. It can also operate in a low-delay mode to adapt to real-time communications applications like video conferencing, while allowing higher processing delay where there are less time constraints, for video storage, server-based streaming etc. It also provides the tools needed to deal with packet loss in packet networks and bit errors in error-prone wireless networks. In H.264, video compression techniques are quite different from those in the previous video compression standards. The syntax and the algorithms used in H.264 are so different that transcoding a video compressed by traditional DCT-based standards to H.264 will face many difficulties, especially to perform transcoding in the frequency domain. The following issues are worthy to be studied in the H.264 related transcoding.

- As shown in Table I, intra block prediction is performed only in H.264. Thus, the DCT coefficients from a source video coded with other standards cannot be fed to H.264 encoder for re-quantization without any transformation. Transcoding of an I frame requires techniques for intra block prediction in the compressed domain.
- H.264 employs 4×4 Integer transformation, which is different from 8×8 DCT used in other standards.

Transcoding of 8 × 8 DCT coefficients to 4 × 4 Integer transformation would be an interesting research problem.

- The prediction block structures and MV prediction coding algorithms in H.264 are different from the previous standards. Thus, the MVs extracted from a source video coded with other standards are not appropriate for the H.264 target video. New algorithms are needed to acquire more adequate MVs of the target video.

VII. Conclusion

Video transcoding is a core technology for providing universal multimedia access by the Internet users with different access links and devices. This paper reviewed several existing video transcoding techniques. Various transcoding architectures provide tradeoff between the computational complexity and reconstructed video quality. SDTAs provide the best video quality but with more complexity, while FDTAs provide a bit lower quality but with lower complexity. HDTAs take advantages of both architectures to provide a tradeoff between complexity and video quality. In general, homogeneous transcoding techniques can also be used in heterogeneous cases. With the introduction of object-based coding concept in MPEG-4, object-based transcoding architectures and techniques offer important research directions. Video compression algorithms used in the standardizing H.264 are very different from that of in the previous traditional video compression standards. To obtain inter-compatibility between H.264 and other standards, H.264 related transcoding would become a more challenge issue in the future research of video transcoding.

References

[1] P. A. A. Assuncao and M. Ghanbari, "Transcoding of single-layer MPEG video into lower rates," *Proc. Inst. Elect. Eng.—Vis., Image Signal Process.*, vol. 144, no. 6, pp. 377–383, Dec. 1997.

[2] ——, "A frequency-domain video transcoder for dynamic bitrate reduction of MPEG-2 bit streams," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 8, pp. 953–967, Dec. 1998.

[3] N. Bjork and C. Christopoulos, "Transcoder architecture for video coding," *IEEE Trans. Consumer Electron.*, vol. 44, pp. 88–98, Feb. 1998.

[4] S.-F. Chang and D. G. Messerschmitt, "Manipulation and composition of MC-DCT compressed video," *IEEE J. Select. Areas Commun.*, vol. 13, no. 1, pp. 1–11, Jan. 1995.

[5] M.-J. Chen, M.-C. Chu, and C.-W. Pan, "Efficient motion-estimation algorithm for reduced frame-rate video transcoder," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 4, pp. 269–275, Apr. 2002.

[6] S. Dogan, A. Cellatoglu, M. Uyguroglu, A. H. Sadka, and A. M. Kondoz, "Error-resilient video transcoding for robust internetwork communications using GPRS," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 453–464, Jun. 2002.

[7] S. Dogan and A. H. Sadka, "Video transcoding for inter-network communications," in *Compressed Video Communications*. Chichester, U.K.: Wiley, 2002, 0-470-84 312-8, pp. 215–256.

[8] A. Eleftheriadis and D. Anastassiou, "Constrained and general dynamic rate shaping of compressed digital video," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, Washington, DC, Oct. 1995, pp. 396–399.

[9] N. Feamster and S. Wee, "An MPEG-2 to H.263 transcoder," in *SPIE Int. Symp. Voice, Video, and Data Communications*, Boston, MA, Sep. 1999.

[10] K.-T. Fung, Y.-L. Chan, and W.-C. Siu, "New architecture for dynamic frame-skipping transcoder," *IEEE Trans. Image Process.*, vol. 11, no. 8, pp. 886–900, Aug. 2002.

[11] K.-L. Huang, Y.-S. Tung, J.-L. Wu, P.-K. Hsiao, and H.-S. Chen, "A frame-based mpeg characteristics extraction tool and its application in video transcoding," *IEEE Trans. Consum. Electron.*, vol. 48, no. 3, Aug. 2002.

[12] J.-N. Hwang and T.-D. Wu, "Motion vector re-estimation and dynamic frame-skipping for video transcoding," in *Conf. Rec. 32nd Asilomar Conf. Signals, System & Computer*, vol. 2, 1998, pp. 1606–1610.

[13] Internet Transcoding for Universal Access [Online]. Available: http://www.research.ibm.com/networked_data_systems/transcoding/

[14] *ISO/IEC-JTC1/SC29/WG11*, Test model 5, MPEG93/457, Apr. 1993.

[15] *ISO/IEC 14 496-2:1999/FDAM4*, ISO/IEC JTC1/SC 29/WG11 N3904, Jan. 2001.

[16] *Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC)*, May 27, 2003. Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6).

[17] G. Keesman, R. Hellinghuizen, F. Hoeksema, and G. Heideman, "Transcoding of MPEG bitstreams," *Signal Processing: Image Commun.*, vol. 8, no. 6, pp. 481–500, 1996.

[18] U.-V. Koc and K. J. R. Liu, "Discrete-cosine/sine-transform based motion estimation," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, Austin, TX, Nov. 1994, pp. 771–775.

[19] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 301–317, Mar. 2001.

[20] W. Li, J.-R. Ohm, M. Schaar, H. Jiang, and S. Li, "MPEG-4 video verification model V18.0," in *ISO/IEC JTC1/SC29/WG11 N3908*, Pisa, Italy, Jan. 2001.

[21] Y.-Q. Liang and Y.-P. Tan, "Methods and needs for transcoding MPEG-4 fine granularity scalability video," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 4, ISCAS 2002, May 2002, pp. 719–722.

[22] Y.-C. Lin, C.-N. Wang, T.-H. Chiang, A. Vetro, and H. Sun, "Efficient FGS to single layer transcoding," in *Proc. IEEE Int. Conf. Consumer Electronics*, Los Angeles, CA, Jun. 2002, pp. 134–135.

[23] S.-Z. Liu and A. C. Bovik, "Local bandwidth constrained fast inverse motion compensation for DCT-domain video transcoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 5, pp. 309–319, May 2002.

[24] R. Mohan, J. R. Smith, and C. Li, "Adapting multimedia internet content for universal access," *IEEE Trans. Multimedia*, vol. 1, no. 1, pp. 104–114, Mar. 1999.

[25] A. Morel and A. Bourge, "Spatial and temporal filtering in a low-cost MPEG bit rate transcoder," in *Proc. IEEE Int. Conf. Acoustics 2001*, vol. 3, 2001, pp. 1885–1888.

[26] Y. Nakajima, H. Hori, and T. Kanoh, "Rate conversion of MPEG coded video by re-quantization process," in *Proc. IEEE Int. Conf. Image Processing (ICIP) 1995*, vol. 3, Washington, DC, Oct. 1995, pp. 408–411.

[27] K. Panusopone, X. Chen, and F. Ling, "Logo insertion in MPEG transcoder," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing 2001*, vol. 2, Salt Lake City, UT, May 2001, pp. 981–984.

[28] G. D. L. Reyes, A. R. Reibman, S.-F. Chang, and J. C.-I. Chuang, "Error-resilient transcoding for video over wireless channels," *IEEE J. Select. Areas Commun.*, vol. 18, no. 6, pp. 1063–1074, Jun. 2000.

[29] N. Roma and L. Sousa, "Insertion of irregular-shaped logos in the compressed DCT domain," in *Proc. 14th Int. Conf. Digital Signal Processing (DSP) 2002*, vol. 1, Jul. 2002, pp. 125–128.

[30] T. Shanableh and M. Ghanbari, "Heterogeneous video transcoding to lower spatial-temporal resolutions and different encoding formats," *IEEE Trans. Multimedia*, vol. 2, no. 2, pp. 101–110, Jun. 2000.

[31] ——, "Transcoding architectures for DCT-domain heterogeneous video transcoding," in *Proc. IEEE Conf. Image Processing 2001*, vol. 1, Oct. 2001, pp. 433–436.

[32] B. Shen, I. K. Sethi, and B. Vasudev, "Adaptive motion-vector resampling for compressed video down scaling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 6, pp. 929–936, Sep. 1999.

[33] G. Shen, B. Zeng, Y.-Q. Zhang, and M. L. Liou, "Transcoder with arbitrarily resizing capability," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS) 2001*, vol. 5, Sydney, NSW, Australia, May 2001, pp. 25–28.

[34] H. Sun, W. Kwok, and J. W. Zdepski, "Architectures for MPEG compressed bitstream scaling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 2, pp. 191–199, Apr. 1996.

[35] M.-T. Sun, T.-D. Wu, and J.-N. Hwang, "Dynamic bit allocation in video combining for multipoint conferencing," *IEEE Trans. Circuits Syst. II*, vol. 45, no. 5, pp. 644–648, May 1998.

[36] Y. Sun, X. Wei, and I. Ahmad, "Low delay rate-control in video transcoding," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS) 2003*, vol. 2, Bangkok, Thailand, May 2003, pp. II-660–II-663.

[37] K. Takahashi, K. Satoh, T. Suzuki, and Y. Yagasaki, "Motion vector synthesis algorithm for MPEG2-to-MPEG4 transcoder," in *, Proc. SPIE—Visual Communications and Image Processing*, vol. 4310, San Jose, CA, Jan. 2001, pp. 872–882.

[38] K.-H. Tan and M. Ganbari, "Layered image coding using the DCT pyramid," *IEEE Trans. Image Process.*, vol. 4, no. 4, pp. 512–516, Apr. 1995.

[39] P. N. Tudor and O. H. Werner, "Real-time transcoding of MPEG-2 video bit streams," in *IEE Int. Broadcasting Convention*, Amsterdam, The Netherlands, 1997, pp. 286–301.

[40] A. Vetro, H. F. Sun, and Y. Wang, "Object-based transcoding for adaptable video content delivery," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 3, pp. 387–401, Mar. 2001.

[41] A. Vetro, H. F. Sun, and Y. Wang, "MPEG-4 rate control for multiple video objects," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 1, pp. 186–199, Feb. 1999.

[42] X. Wang, W. Zheng, and I. Ahmad, "MPEG-2 to MPEG-4 transcoding," in *Proc. Workshop and Exhibition on MPEG-4*, San Jose, CA, Jun. 18–20, 2001, pp. 83–86.

[43] O. Werner, "Requantization for transcoding of MPEG-2 intraframes," *IEEE Trans. Image Process.*, vol. 8, no. 2, pp. 179–191, Feb. 1999.

[44] S. Xiao, L. Lu, J. L. Kouloheris, and C. A. Conzales, "A low cost and efficient logo insertion scheme in MPEG video transcoding," in *Proc. SPIE—Visual Communications and Image Processing*, vol. 4671, San Jose, CA, Jan. 2002, pp. 172–179.

[45] R. Xie, J. Liu, and X. Wang, "Efficient MPEG-2 to MPEG-4 compressed video transcoding," in *Proc. SPIE—Visual Communications and Image Processing*, vol. 4671, San Jose, CA, Jan. 2002, pp. 192–201.

[46] J. Xin, M.-T. Sun, K. Chun, and B. S. Choi, "Motion re-estimation for HDTV to SDTV transcoding," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS) 2002*, vol. 4, Geneva, Switzerland, May 2002, pp. 715–718.

[47] P. Yin, A. Vetro, H. Sun, and B. Liu, "Drift compensation architectures and techniques for reduced resolution transcoding," in *Proc. SPIE—Visual Communications and Image Processing*, vol. 4671, San Jose, CA, Jan. 2002.

[48] P. Yin and M. Wu, "Video transcoding by reducing spatial resolution," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, 2000, pp. 972–975.

[49] J. Youn, M.-T. Sun, and C.-W. Lin, "Motion vector refinement for high-performance transcoding," *IEEE Trans. Multimedia*, vol. 1, no. 1, pp. 30–40, Mar. 1999.

Ishfaq Ahmad (SM'03) received the B.Sc. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 1985, and the M.S. degree in computer engineering and Ph.D. degree in computer science from Syracuse University, Syracuse, NY, in 1987 and 1992, respectively.

He is currently a Full Professor of Computer Science and Engineering in the Department of Computer Science and Engineering, University of Texas (UT) at Arlington. Prior to joining UT Arlington, he was an Associate Professor in the Computer Science Department at Hong Kong University of Science and Technology (HKUST). At HKUST, he was also the Director of the Multimedia Technology Research Center, an officially recognized research center that he conceived and built from scratch. The center was funded by various agencies of the Government of the Hong Kong Special Administrative Region as well as local and international industries. With more than 40 personnel including faculty members, postdoctoral fellows, full-time staff, and graduate students, the center engaged in numerous R&D projects with academia and industry from Hong Kong, China, and the U.S. The particular areas of focus in the center are video (and related audio) compression technologies, video telephone and conferencing system. The center has commercialized several of its technologies to its industrial partners worldwide. His recent research focus has been on developing parallel programming tools, scheduling and mapping algorithms for scalable architectures, heterogeneous computing systems, distributed multimedia systems, video compression techniques, and web management. His research work in these areas is published in over 125 technical papers in refereed journals and conferences.

Dr. Ahmad has participated in the organization of several international conferences and is an Associate Editor of *Cluster Computing*, the *Journal of Parallel and Distributed Computing*, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, *IEEE Concurrency*, and *IEEE Distributed Systems Online*. He received Best Paper Awards at Supercomputing 90 (New York), Supercomputing'91 (Albuquerque), and the 2001 International Conference on Parallel Processing (Spain).

Xiaohui Wei (S'03) received the B.S. and M.S. degrees in electrical engineering from Zhongshan University (Sun Yat-Sen University), Guangzhou, China, in 1997 and 2000, respectively. She has been pursuing the Ph.D. degree in computer science and engineering at the University of Texas at Arlington since 2002.

From July 2000 to January 2002, she was an Engineer with China Telecom-Guangzhou R&D Center, Guangdong Telecommunications Academy of Science and Technology. Her main research interests include video compression, multimedia communication, video fusion and registration and image processing.

Yu Sun (S'03) received the B.S. and M.S. degrees in computer science from the University of Electronic Science and Technology of China, Chengdu, China, in 1996 and the Ph.D. degree in computer science and engineering from the University of Texas at Arlington in 2004.

From 1996 to 1998, she was a lecturer in the Department of Computer Science at Sichuan Normal University, China. Since August 2004, she has been an Assistant Professor in the Computer Science Department, University of Central Arkansas, Conway. Her main research interests include video compression, multimedia communication, and image processing.

Ya-Qin Zhang (S'87–M'90–SM'93–F'97) received the B.S. and M.S. degrees in electrical engineering from the University of Science and Technology of China (USTC) in 1983 and 1985, respectively, and the Ph.D. degree in electrical engineering from George Washington University, Washington, DC, in 1989. He also received executive business training from Harvard University, Cambridge, MA.

He is the Corporate Vice President of Microsoft Corporation's Mobile and Embedded Devices Division, Seattle, WA. He was previously the Managing Director of Microsoft Research Asia, Beijing, China, and the Director of the Multimedia Technology Laboratory, Sarnoff Corporation, Princeton, NJ (formerly David Sarnoff Research Center and RCA Laboratories). He was with GTE Laboratories, Inc., Waltham, MA, and Contel Technology Center, VA, from 1989 to 1994. He has been engaged in research and commercialization of MPEG2/DTV, MPEG4/VLBR, and multimedia information technologies. He has more than 70 U.S. patents—granted or pending—and has authored or contributed to more than a dozen books and 300 influential technical papers and journal articles. Many of the technologies he and his team developed have become the basis for start-up ventures, commercial products, and international standards. He has been a key contributor to the ISO/MPEG and ITU standardization efforts in digital video and multimedia.

Dr. Zhang served as the Editor-in-Chief for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY from July 1997 to July 1999. He was the Chairman of Visual Signal Processing and Communications Technical Committee of IEEE Circuits and Systems. He serves on the Editorial Boards of seven other professional journals and over a dozen conference committees. He has received numerous awards, including several industry technical achievement awards and IEEE awards such as the Jubilee Golden Medal. He was named the "Research Engineer of the Year" in 1998 by the New Jersey Engineering Council for his "leadership and invention in communications technology, which has enabled dramatic advances in digital video compression and manipulation for broadcast and interactive television and networking applications." He received the Prestigious National Award of "The Outstanding Young Electrical Engineering of 1998," given annually to one Electrical Engineer in the U.S.