W

Watermarking, Audio

Definition

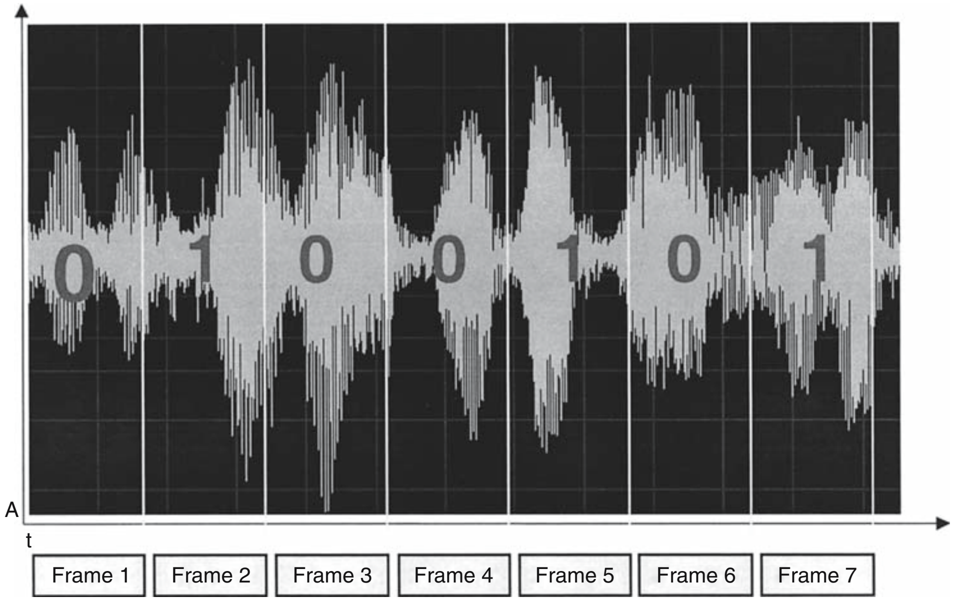
Digital audio watermarking is a technology to embed and retrieve information into and from digital audio data.

Audio watermarking techniques use common water methods explained in detail in the article on Digital Watermarking.

Raw audio data is commonly stored as PCM (Pulse code modulation) samples. Common sample sizes vary from 8 to 24 bits, sampling rates range from 8 kHz to 96 kHz. Most audio watermarking algorithms work on raw audio data. But there are others which aim at lossy compression audio formats which are commonly applied today for storage or transfer. Mp3 is the best know example for this. Some of these algorithms embed the information in some format specific information, like for example the scale factors of mp3 files. Other algorithms change the compressed files in such a way that the watermark can also be retrieved from the raw data. This is called bit stream embedding.

Common methods like LSB, pattern, statistical methods, patchwork methods, correlation of embedded noise [1] applied in image watermarking are also known for audio watermarking. In audio watermarking, individual bits are often embedded in time frames like shown in Fig. 1.

But there are also approaches using specific features of audio data, like echo hiding or phase coding. Echo hiding utilizes the fact that human perception cannot perceive sounds of small energy shortly after loud sounds. An echo hiding watermarking algorithm repeats small amounts of the cover when loud sounds occur with reduced energy. The delay between sound and copy can be used to transfer information. Phase coding is based on the fact that human perception for audio phases is weak, so changes in the phase can be used for transferring information.



Watermarking, Audio. Figure 1. Audio watermarking usually embeds single bits in discreet time frames.

Audio watermarking usually addresses watermarking of sound data. But there are also solutions for watermarking musical scores and MIDI data. Musical score watermarking is more similar to watermarking of simples images or drawings due to the low complexity of the cover data. MIDI watermarking also needs to deal with a small cover files with only few possibilities of data modification. LSB watermarking methods of attack velocity codes and small timing modifications have been introduced as suitable watermarking strategies for MIDI data.

Challenges, Attacks and Benchmarking

At audio watermarking, usually two types of attacks against the watermarking robustness are important: Lossy compression and analogue transfer. Most audio files are distributed in a lossy compression format as for example mp3, wma or ogg. The watermark embedded in the audio data should survive the different compression formats and bit rates as long as an acceptable audio quality is provided. Audio often leaves the digital domain when it is consumed or used, for example in radio networks, live recordings or tape copies. Some applications require the audio watermark to stay in the audio signal and be still retrieval when transferred back into the digital domain. Current state of the art watermarking algorithms provide a good robustness against both attacks.

The growing number of attacks against watermarking systems has shown the importance of efficient and reliable benchmarking to improve the quality of existing watermarking methods. General aspects on attacks and benchmarking can be found in the article on Digital Watermarking. A wide range of image watermarking evaluation approaches and benchmarking suites have been described in the literature by neglecting video watermarking techniques.

With StirMark Benchmark for audio [2], a well-defined benchmark for audio watermarking robustness and security has been introduced. The benchmark contains of a set of single geometric attacks in time and frequency domain. They simulate different signal processing effects by adding or removing signals or applying filtering, common in several watermarking applications. The set of attacks allows determining robustness or fragility of a watermark embedded into the audio signal to specific single manipulations or to its arbitrary combination.

Applications

While all watermarking applications known from image and video watermarking can also be applied to audio watermarking, the protection of copyright always had an important role in audio watermarking. The best-known example for this is the Secure digital music Initiative (SDMI)) which was a group of companies planning to develop a framework for secure digital music distribution. They tried to install a framework for protected playback, storing, and distribution of digital music. SDMI participants include music content, consumer electronics, information technology, and wireless telecommunication companies. Major music labels where as well included as for example the developers of the Napster software, but complains had been raised that only a small group of participants were involved in key-problem identification and management.

References

1. B.H. Tewfik, "Digital Watermarks for Audio Signals," Proceedings of the EUSIPCO-96, VIII European Signal Processing Conference, Trieste, Italy, pp. 473–480, September 1996.
2. M. Steinebach, F.A.P. Petitcolas, F. Raynal, J. Dittmann, C. Fontaine, C. Seibel, N. Fates, and C.L. Ferri, "StirMark Benchmark: Audio watermarking attacks," Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC 2001), Las Vegas, Nevada, pp. 49–54, April 2001.

Watermarking, Video

Definition

Digital video watermarking is a technology to embed and retrieve information into and from digital video data.

A variety of robust and fragile video watermarking methods have been proposed to solve the illegal copying and proof of ownership problems as well as to identify manipulations. Although a number of broad claims have been made in the field of robustness of various digital watermarking methods, it is still difficult to handle combined or non-linear geometric transformations. The methods can be divided into techniques working on compressed or uncompressed data. In particular video watermarking is based in general on the following concepts to hide a watermark by modifying some of its characteristics:

1. Spatial domain approach, also called native domain: embedding and detection are performed on spatial pixels values (luminance, chrominance, color space) or on the overall video frame characteristic,
2. Feature or salient point watermarking by modifying geometric properties of the video frames,
3. Frequency domain techniques where the spatial values are transformed, like DCT Discrete Cosine, FFT Fast Fourier Transform, Wavelets or fractals,
4. Quantization index modulation (QIM) watermarking,
5. Format-specific approaches like watermarking of structure elements like Facial Animation Parameter of MPEG-4 or motion vectors.

The robust watermarking approaches usually spread the watermarking information redundant over the overall signal representation in a non-invertible manner to enforce identification or verification of ownership or to annotate the video. For example the message is spread and encoded in the 2D FFT frequencies of each video frame. In most cases prior to transforming each frame to the frequency domain, the frame data is transformed from e.g., RGB space to Weber-Fechner YCbCr space. Generically, YCbCr space consists of a luminance component Y and two color difference components Cb and Cr. The Y component contains the luminance and black & white image information, while Cb represents the difference between R and Y and Cr represents the difference between Y and B. In YCbCr space most of the frame information is in the Y component. This representation is used also during MPEG compression. The MPEG algorithm grossly removes large portions of the Cb and Cr components without damaging the frame quality. The MPEG algorithm uses compression to reduce the Y component since it has more effect on the quality of the compressed frames, which is also used in most watermarking schemes. To avoid estimation attacks the watermark signal should be adoptive designed to the overall video frame sequence characteristics by facing the problem that equal or similar watermarking pattern for each frame could allow an estimation attack based on the slide visual differences between adjacent frames, different watermarking pattern could allow an estimation attack based on the similarities between similar visual frames, see for example in [1].

Data rates are measured in bits per frame. If the watermark is embedded into the compressed domain for example MPEG video, we count the embedded data rates in bits per I, B and P frame or per GOP.

For fragile video watermarking, relevant for authenticate the data in its authenticity and integrity, a fragile watermark signal can be spread over the overall video into manipulation sensitive video elements like LSB to detect changed and manipulated regions. The watermark for authentication purposes is often designed in an invertible (reversible) manner to allow reproduction of the original, see for example an analysis of the approaches in [2]. Today we find several fragile watermarking techniques to recognize video manipulations. In the moment most fragile watermarks are very sensitive to changes and can detect most possible changes in pixel values. Only few approaches address the so-called content-fragile watermarking relevant in applications with several allowed post production editing processes. For example [3] addresses the recognition of video frame sequence changes with the possibility of reproduction of the original frame sequence called self-watermarking of frame-pairs.

A further idea is for example from [4] suggesting to embed a visual content feature *M* into the video frame with a robust watermarking method. The content-fragile watermarking approach for video authentication tries to extract the frame characteristics of human perception, called content. The approach is for example to determine the edge characteristics of the single video frame. This characteristic is transformed into a feature code for the content-fragile digital watermark. The edge characteristics of a frame give a very good reflection of the frame content, because they allow the identification of object structures and homogeneity of the video. Dittmann [4] use the canny edge detector described as the most efficient edge separator. The author described several strategies for generating and verification of the edge based feature codes. From the general perspective the content feature *M* can be embedded directly or used as a seed to generate the watermarking pattern itself.

Benchmarking of Video Watermarks

The available benchmarking suites mainly cover image and audio watermarks and neglect video specific aspects. In [1], we find an evaluation of video watermarking sensitivity to collusion attacks and potential solutions. Video specific aspects of video watermarks for cinema applications are summarized in [5]. Various features were analyzed, including robustness to non intentional attacks such as MPEG compression, transcoding, analog to digital and digital-to-analog conversions, standard conversions (PAL – NTSC), and change of geometry.

References

1. K. Su, D. Kundur, and D. Hatzinakos, "A Novel Approach to Collusion-Resistant Video Watermarking," in E.J. Delp and P.W. Wong, (Eds.) Security and Watermarking of Multimedia Contents IV, Proceedings of SPIE, San Jose, California, Vol. 4675, January 2002.
2. S. Katzenbeisser and J. Dittmann, "Malicious Attacks on Media Authentication Schemes Based on Invertible Watermarks," Proceedings of SPIE Conference 5306: Security, Steganography, and Watermarking of Multimedia Contents VI, 2004.
3. B.G. Mobasseri and A. Evans, "Content-Dependent Video Authentication by Self-Watermarking in Color Space," Proceedings of Security and Watermarking of Multimedia Contents III, Electronic Imaging' 01, San Jose, pp. 35–44, January 2001.
4. J. Dittmann, "Content-fragile Watermarking for Image Authentication," in Ping Wah Wong, Edward J. Delp III, (Eds.) Security and Watermarking of Multimedia Contents III, Proceedings of SPIE, Vol. 4314, 2001, pp. 175–184.
5. B. Macq, J. Dittmann, and E.J. Delp, "Benchmarking of Image Watermarking Algorithms for Digital Rights Management," Proceedings of the IEEE, Special Issue on: Enabling Security Technology for Digital Rights Management, Vol. 92, No. 6, June 2004, pp. 971–984.

Wireless Video

Zhihai He[1], Chang Wen Chen[2]
[1]University of Missouri at Columbia, MO, USA
[2]Florida Institute of Technology, Melbourne, FL, USA

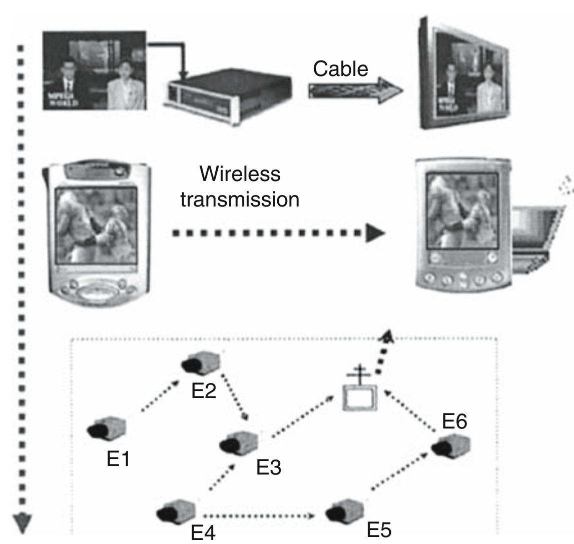
Definition

Wireless video refers to transporting video signals over mobile wireless links.

Introduction

The rapid growth of mobile wireless access devices, together with the success of wireless networking technologies, has brought a new era of video communications: transporting video signals over mobile wireless links. Transport of video content over mobile wireless channels is very challenging because the mobile wireless channels are usually severely impaired due to multipath fading, shadowing, inter-symbol interferences, and noise disturbances. Traditionally, vision has been the dominant medium through which people receive information. Visual information, coupled with intelligent vision processing, provides a rich set of important information for situational awareness and event understanding [1]. Incorporating the video capture, processing, and transmission capabilities into networked mobile devices will enable us to gather real-time visual information about the target events at large scales for situational awareness and decision making. This will create a potential impact through out the society via many important applications, including battle-space communication, video surveillance, security monitoring, environmental tracking, and smart spaces.

During the past decade, the video communication system has evolved from the conventional desktop computing and wired communication to mobile computing and wireless communication, as illustrated in Fig. 1. In this scenario, the live video is captured by a camera on a portable device. The video data is compressed on-board and transmitted to remote users through wireless channels. As the communication



Wireless Video. Figure 1. From desktop computing and wired communication to mobile computing and wireless communication, and furthermore to wireless sensor networks.

paradigm evolves from the conventional point-to-point, wired and centralized communication to the current wireless, distributed, ad hoc, and massive communication, the system becomes more and more complex. More specifically, such massive wireless communication networks often involve a large number of heterogeneous devices, each with different on-board computation speed, energy supply, and wireless communication capability, communicating over the dynamic and often error prune wireless networks. How to characterize and manage the communication behavior of each communication devices within the network, and how to coordinate their behaviors such that each operates in a contributive fashion to maximize the overall performance of the system as a whole remain a central challenging research problem.

Over the past few decades, extensive research has been conducted on various elements of the wireless video communication networks, such as video compression [2], mobile ad hoc protocol design [3], energy-aware routing [4], power management and topology control [5]. However, little research work has been done to bridge them into an integrated resource management and performance optimization framework. Developing efficient algorithms for real-time video compression and streaming over wireless networks to maximize the overall system performance under resource constraints has become one of the central research tasks in both signal processing and wireless communication research communities.

The ultimate goal in communication system design is to control and optimize the system performance under resource constraints. In mobile wireless video communication, video encoding and network communication operate under a set of resource constraints, including bandwidth, energy, and computational complexity constraints. To analyze the behavior of the mobile video communication system, manage its resources, and optimize the system performance, we need to study the intrinsic relationships between the resource constraints and the end-to-end video distortion. This study is called resource-distortion analysis. This resource-distortion analysis extends the traditional R-D analysis by considering new resource constraints. In this article, we shall analyze the major resource constraints in real-time video compression and streaming over wireless networks, and study the impact of these resource constraints on the overall system performance.

Rate Constraints

In traditional video communication applications, such as digital TV broadcast, and video-on-demand, video signals can be compressed offline, stored on a video server, and transmitted through the wired network to viewers upon request. In this case, the major constraint for video compression and communication is in the form of transmission bandwidth or storage space, which determines the output bit rate of video encoder [2, 6]. Therefore, the ultimate goal in this type of communication system design is to optimize the video quality under the rate constraint. To this end, rate-distortion (R-D) theories have been developed to model the relationship between the coding bit rate and signal distortion [7]. The R-D theory describes the performance limit of lossy data compression, and answers the following fundamental question: What is the minimum number of bits needed in compressing the source data at a given distortion level (or reconstruction quality).

During the last 50 years, R-D theories have been actively studied in the information theory literature, mainly focusing on performance bounds, including asymptotic analysis [7] and high rate approximation [8, 9]. It should be noted that theoretical analysis and analytical R-D performance bounds are likely to be found only for simple sources and simple encoding schemes. For complicated sources, such as 2-D images and 3-D videos, and sophisticated compression systems, such as JPEG and JPEG2000 image coding, MPEG-2, H.263, MPEG-4 [2], and H.264 [10] video encoding, this type of theoretical performance analysis is often inapplicable [11]. This is because: (1) Unlike 1-D text and acoustic data, whose compression characteristics can be easily captured by simple statistical models, such as Gaussian and Laplacian models, images and videos often exhibit very complicated source characteristics and correlation structure. The underlying scene structure of the 3-D environment, the time-varying motion patterns of scene objects, as well as the arbitrary camera movement, collectively define a very complicated source correlation structure in the video data. This type of correlation structure is often very difficult to be described by mathematical models. (2) Note that the major effort in image and video compression is to explore the spatiotemporal source correlation with various motion prediction and spatial transform techniques.

To explore the complicated source correlation structure of the video sequence, very sophisticated prediction and data representation techniques, such as multi-frame motion compensation, flexible macroblock (MB) size, intra prediction and mode decision [10], have been developed. These techniques, often seen to be ad hoc and difficult to be mathematically modeled, however have significant impact on the overall video compression performance. The difficulty in mathematical modeling of both the source characteristics and the compression system creates a significant gap between the information-theoretic R-D analysis and practices in rate control and quality optimization for video compression. To fill in this gap, over the past two decades, as more and more advanced image and video compression algorithms are being developed and finalized in international standards, a set of R-D analysis and modeling techniques algorithms for practical video compression have been developed [6, 11, 12].

The analysis and estimation of R-D functions have important applications in visual coding and communication. First, with the estimated R-D functions we can adjust the quantization setting of the encoder and control the output bit rate or picture quality according to channel conditions, storage capacity, or user's requirements [12]. Second, based on the estimated R-D functions, optimum bit allocation, as well as other R-D optimization procedures, can be performed to improve the efficiency of the coding algorithm and, consequently, to improve the image quality or video presentation quality [13].

Energy Constraints

In wireless video communication, video capture, compression and network streaming operate on the mobile devices with limited energy. A primary factor in determining the utility or operational lifetime of the mobile communication device is how efficiently it manages its energy consumption. The problem becomes even more critical with the power-demanding video encoding functionality integrated into the mobile computing platform [14]. Video encoding and data transmission are the two dominant power-consuming operations in wireless video communication, especially over wireless LAN, where the typical transmission distance ranges from 50 to 100m. Experimental studies show that for relative small picture sizes, such as QCIF (176×144) videos, video encoding consumes about 2/3 of the total power for video communication over Wireless LAN [14]. For pictures of higher resolutions, it is expected that the fraction of power consumption by video encoding will become even higher. From the power consumption perspective, the effect of video encoding is two-fold. First, efficient video compression significantly reduces the amount of the video data to be transmitted, which in turn saves a significant amount of energy in data transmission. Second, more efficient video compression often requires higher computational complexity and larger power consumption in computing. These two conflicting effects imply that in practical system design there is always a tradeoff among the bandwidth R, power consumption P, and video quality D. Here, the video quality is often measured by the mean square error (MSE) between the encoded picture and original one, also known as the source coding distortion. To find the best trade-off solution, we need to develop an analytic framework to model the power-rate-distortion (P-R-D) behavior of the video encoding system. To achieve flexible management of power consumption, we also need to develop a video encoding architecture which is fully scalable in power consumption.

Many algorithms have been reported in the literature to reduce the encoding computational complexity. Hardware implementation technologies have also been developed to improve the video coding speed. However, little research has been done to analyze the relationship between the encoder power consumption and its R-D performance [14].

Rate-distortion (R-D) analysis has been one of the major research focus in information theory and communication for the past few decades, from the early Shannon's source coding theorem for asymptotic R-D analysis of generic information data [7], to recent R-D modeling of modern video encoding systems [6, 11, 12]. For video encoding on the mobile devices and streaming over the wireless network, it is needed to consider another dimension, the power consumption, to establish a theoretical basis for R-D analysis under energy constraints. In energy-aware video encoding, the coding distortion is not only a function of the encoding bit rate as in the traditional R-D analysis, but also a function of the power consumption P. In other words:

$$D = D(R, P) \tag{1}$$

which describes the P-R-D behavior of the video encoding system. The P-R-D analysis answers the following question: for given bandwidth R and encoder power consumption level P, what is the minimum coding distortion one can achieve? Generally speaking, this is a theoretically difficult problem, because power consumption and R-D performance are concepts in two totally unrelated fields. However, for a specific video encoding system, for example MPEG-4 video coding, one can design an energy-scale video compression scheme, model its P-R-D behavior, and use this model to optimize the R-D performance under the energy constraint [15] (See the short article.).

Encoding Complexity and Encoder Power Consumption

In embedded video compression system design, the encoder power consumption is directly related to computational complexity of the encoder. In other words, the encoder power consumption Ps is a function of the encoder complexity $\boldsymbol{C_s}$, denoted by $\boldsymbol{P_s} = \boldsymbol{\Phi}(\boldsymbol{C_s})$, and this function is given by the power consumption model of the microprocessor [16]. To translate the complexity scalability into energy scalability, we need to consider the energy-scaling technologies in hardware design. To dynamically control the energy consumption of the microprocessor on the portable device, a CMOS circuits design technology, named dynamic voltage scaling (DVS), has been recently developed [17]. In CMOS circuits, the power consumption P is given by

$$P_s = V^2 \cdot f_{CLK} \cdot C_{EFF} \qquad (2)$$

where V, f_{CLK}, and C_{EFF} are the supply voltage, clock frequency, and effective switched capacitance of the circuits [17]. Since the energy is power times time, and the time to finish an operation is inversely proportional to the clock frequency. Therefore, the energy per operation E_{op} is proportional to V. This implies that lowering the supply voltage will reduce the energy consumption of the system in a quadratic fashion. However, lowering the supply voltage also decreases the maximum achievable clock speed. More specifically, it has been observed that fCLK is approximately linearly proportional to V [17]. Therefore, we have

$$P_s \propto f_{CLK}^3 \text{ and } E_{op} \propto f_{CLK}^2 \qquad (3)$$

It can be seen that the CPU can reduce its energy consumption substantially by running more slowly. For example, according to (3), it can run at half speed and thereby use only 1/4 of the energy for the same number of operations. This is the key idea behind the DVS technology. Variable chip makers, such as Intel [16], have recently announced and sold processors with this energy-scaling feature. In conventional system design with fixed supply voltage and clock frequency, clock cycles, and hence energy, are wasted when the CPU workload is light and the processor becomes idle. Reducing the supply voltage in conjunction with the clock frequency eliminates the idle cycles and saves the energy significantly. In this work, we just use this DVS technology and the related power consumption model to translate the computational complexity into the energy consumption of the hardware.

Video Compression and Transmission under Energy Constraints

As mentioned before, the energy supply of a mobile communication device is mainly used by video compression and wireless transmission. For the power consumption in video compression and streaming, we have the following two observations. Case A: If we decrease the encoder power consumption P_s, the coding distortion D_s increases, which is due to lack of enough video processing. That is, $P_s \downarrow$ ®$D_s \uparrow$. Case B: Since the total power consumption P_0 is fixed, and $P_0 = P_s + P_t$, where P_t is the transmission power. If we increase P_s, then P_t decreases. This implies that less bits can be transmitted because the transmission energy is proportional the number of bits to transmit. Therefore, $P_s \uparrow$ ® $p_t \downarrow$ ® $R_s \downarrow$ ® $D_s \uparrow$. It can be seen that when the encoding power P_s goes too low or too high, the encoding distortion D_s will be large. This implies that there exists an optimal power P_s that minimizes the video distortion D_s. In the following, based on a simplified power consumption model for wireless transmission, we study the performance of mobile video device. More specifically, we assume the transmission power is properly chosen such that the bit error rate at the receiver side is very low and the transmission errors can be neglected. In this case, the transmission power should be given by

$$P_t = \eta(d) \cdot R_s \text{ and } \eta(d) = w + \gamma d^{\pi} \qquad (4)$$

where R_s is the number of bits to be transmitted, d is the distance between the sensor node and the receiver (e.g., an AFN), and n is the path loss exponent [9, 44]. Therefore,

$$P_0 = P_s + P_t = P_s + \eta(d)R_s \tag{5}$$

and,

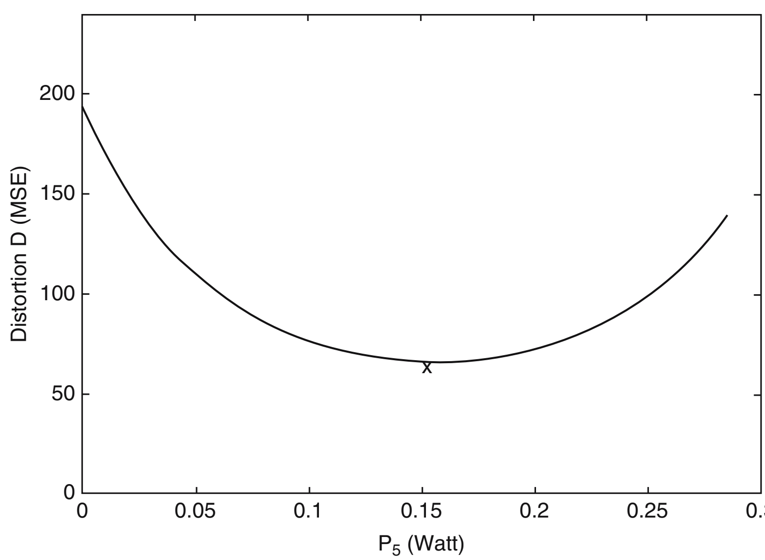
$$R_s = P_0 - P_s\eta(d) \tag{6}$$

Since the transmission errors are negligible, we have $D_t = 0$ and $D = D_s$. According to the P-R-D model,

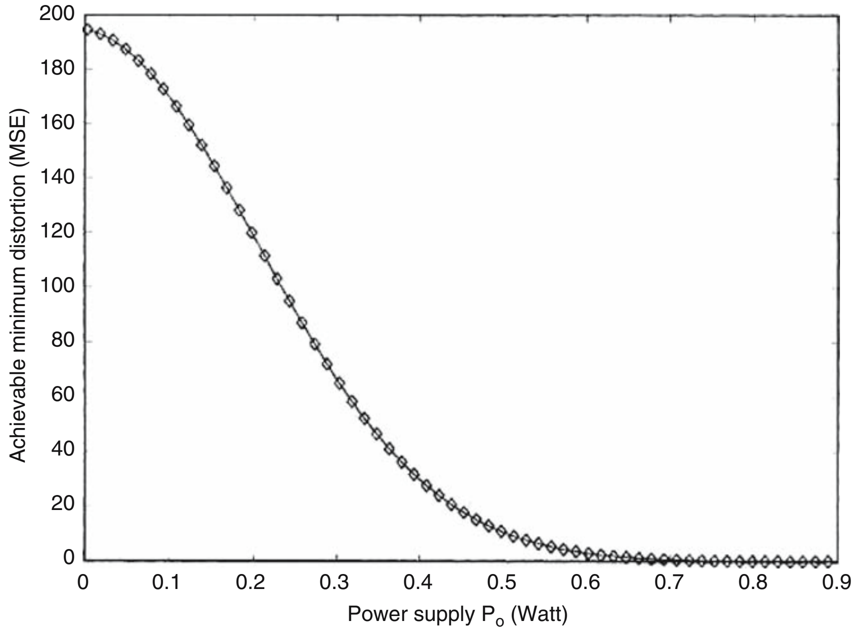
$$\begin{aligned} D &= D_s(R_s, P_s)|_{R_s = P_0 - P_s\eta(d)} \\ &= D_s(P_0 - P_s\eta(d), P_s)_{,0 \le P_s \le P_0} \end{aligned} \tag{7}$$

It can be seen that D is a function of P_s, denoted by $D(P_s)$. Using the P-R-D model, we compute the function $D(P_s)$in (7), and plot it in Fig. 2. Here, the power supply of the wireless video sensor is $P_0 = 0.3$ watts. This is a typical plot of $D(P_s)$. It can be seen that $D(P_s)$ has a minimum point, which is the minimum encoding distortion (or maximum video quality) that a mobile device can achieve, no matter how it allocates its power resource between video encoding and wireless transmission, given fixed total power supply. We call this minimum distortion as achievable minimum distortion (AMD). In Fig. 3, we plot the AMD as a function of the power supply P_0. For a given power supply P_0, the AMD indicates the lower bound on the video coding distortion, or the upper bound on the video quality of a mobile video device.

It should be noted that the AMD bound in (7) is derived based on a simplified model of the mobile video device, and this bound is not tight. More specifically, first, it has not considered the transmission errors [18]. The actual video distortion should consist of both the encoding distortion caused by Quantization loss and the transmission distortion caused by transmission errors [18]. Second, the analysis assumes that the bandwidth of wireless channel is always sufficiently large. Obviously, this is not true for video transmission over the wireless channel which has a limited and time-varying bandwidth. In actual performance analysis for video compression and streaming over mobile devices, it is needed to incorporate the bandwidth constraint and the transmission distortion model [18] into the AMD analysis, and study the achievable minimum distortion in video sensing over the error-prone wireless networks. Another important problem is the mobility of video communication device. The mobility of the device poses new issues in the AMD performance analysis, because it has to deal with various characteristics of the mobile wireless channel, including time-varying path loss, shadowing, and small-scale fading.



Wireless Video. Figure 2. Plot of $D(P_s)$ in (7) and illustration of the achievable minimum distortion, given fixed total power consumption P_0.



Wireless Video. Figure 3. Plot of AMD (P_0).

Wireless Multi-hop Delivery, Video Adaptation, and Security Issues

Wireless video applications often involve transport over a collection of multi-hop wireless nodes to reach the destination. A multi-hop network is dynamically self-organized and self-configured, with the nodes in the network automatically establishing and maintaining mesh connectivity among themselves. This feature brings many advantages to multi-hop networks such as low up-front cost, easy network maintenance, robustness, and reliable service coverage. However, limited network resource, severe interference/ contention among neighbor traffic, dynamic changing route, lack of QoS support, direct coupling between the physical layer and the upper layers, etc. pose many challenges for supporting video communication over wireless multi-hop networks.

Another significant challenge in wireless video is to effectively deal with the heterogeneity of the wireless links and mobile devices. The needs for video adaptation have becoming more important as the advance of the wireless video has become widespread. With regard to wireless video streaming, there are generally two issues in video adaptation: rate adaptation and robustness adaptation. The objective of rate adaptation is to intelligently remove some information from the video signal itself so that end-to-end resource requirement can be reduced. A popular approach to video rate adaptation is the design of the video transcoding algorithms to bridge between two different networks. The objective of robustness adaptation is to increase the capability of the compressed video for transmission over error prone wireless links. Both error resilient source coding and error control channel coding have been used to increase the robustness.

Because wireless video rely on the wireless networks in which users communicate with each other through the open air, unauthorized users may intercept content transmissions or attackers may inject malicious content or penetrate the network and impersonate legitimate users. This intrinsic nature of wireless networks has several specific security implications. Sensitive and valuable video content must be encrypted to safeguard confidentiality and integrity of the content and prevent unauthorized consumption. A wireless device usually has limited memory and computing power. Battery capacity is also at a premium. Growth in battery capacity has already lagged far behind the increase of energy requirement in a wireless device. The security concerns mentioned above make it even worse since security processing has to take away some of the premium computing resources and battery life. To address the peculiar security problems for wireless video with significant energy constraint, we will need to design lightweight cryptographic algorithms and encryption schemes, to include security processing instructions into the embedded processors, and to integrate scalable coding and encryption with error resilience.

Cross-References

- ► Power-Rate-Distortion Analysis for Wireless Video
- ► Security Issues in Wireless Video
- ► Wireless Multi-Hop Video Delivery
- ► Wireless Video Adaptation

References

1. R. Collins, A. Lipton, H. Fujiyoshi, and T. Kanade, "Algorithms for Cooperative Multi-Sensor Surveillance," Proceedings of the IEEE, Vol. 89, No. 10, October 2001, pp. 1456–1477.
2. T. Sikora, "The MPEG-4 Video Standard Verification Model," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 7, February 1997, pp. 19–31.
3. Z. He and S.K. Mitra, "A Unified Rate-Distortion Analysis Framework for Transform Coding," IEEE Transactions on Circuits and System for Video Technology, Vol. 11, December 2001, pp. 1221–1236.
4. C.E. Perkins, E.M. Belding-Royer, and S. R. Das, "Ad hoc on-Demand Distance Vector (AODV) Routing," IETF Internet Draft, Draft-Ietfmanet-Aodv-09.Txt, November 2001.
5. J. Chang and L. Tassiulas, "Energy Conserving Routing in Wireless Ad Hoc Networks," Proceedings INFOCOMM 2000, Tel Aviv, Israel, pp. 22–31, March 2000.
6. J. Pan, Y.T. Hou, L. Cai, Y. Shi, and S.X. Shen, "Topology Control for Wireless Video Surveillance Networks," Proceedings of ACM Mobicom 2003, San Diego, CA, September 2003.
7. W. Ding and B. Liu, "Rate Control of MPEG Video Coding and Recording by Rate-Quantization Modeling," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 6, February 1996, pp. 12–20.
8. T.M. Cover and J.A. Thomas, "Elements of Information Theory," Wiley, New York, 1991.
9. W. Bennett, "Spectra of Quantized Signals," Bell System Technical Journal, Vol. 27, July 1948, pp. 446–472.
10. A. Gersho, "Asymptotically Optimal Block Quantization," IEEE Transactions on Information Technology, Vol. IT-23, July 1979, pp. 373–380.
11. T. Wiegand, "Text of Committee Draft of Joint Video Specification (ITU-T Rec. H.264–ISO/IEC 14496-10 AVC)," Document JVTC167, Third JVT Meeting, Fairfax, Virginia, USA, May 2002.
12. Ortega and K. Ramchandran, "Rate-Distortion Methods for Image and Video Compression," IEEE Signal Processing Magazine, Vol. 15, No. 6, November 1998, pp. 23–50.
13. J. Ribas-Corbera and S. Lei, "Contribution to the Rate Control Q2 Experiment: A Quantizer Control Tool for Achieving Target Bit Rates Accurately," Coding of Moving Pictures and Associated Audio MPEG96/M1812ISO/IECJTC/SC29/WG11, Sevilla, Spain, February 1997.
14. K. Ramchandran and M. Vetterli, "Rate-Distortion Optimal Fast Thresholding with Complete JPEG/MPEG Decoder Compatibility," IEEE Transactions on Image Processing, Vol. 3, September 1994, pp. 700–704.
15. X. Lu, Y. Wang, and E. Erkip, "Power Efficient H.263 Video Transmission over Wireless Channels," Proceedings of the International Conference on Image Processing, Rochester, New York, September 2002.
16. Z. He, Y. Liang, L. Chen, I. Ahmad, and D. Wu, "Power-Rate-Distortion Analysis for Wireless Video Communication Under Energy Constraint," IEEE Transactions on Circuits and System for Video Technology, Vol. 15, No. 5, May 2005, pp.645–658.
17. Intel Inc, "IntelXScaleTechnology," Available at: http://www.intel.com/design/intelxscale.
18. T. Burd and R. Broderson, "Processor Design for Portable Systems," Journal of VLSI Signal Processing, Vol. 13, No. 2, August 1996, pp. 203–222.
19. Z. He, J. Cai, and C.W. Chen, Joint Source-Channel Rate-Distortion Analysis for Adaptive Mode Selection and Rate Control in Wireless Video Coding," IEEE Transactions on Circuits and System on Video Technology, Special Issue on Wireless Video, Vol. 12, June 2002, pp. 511–523.

Wireless Video Adaptation

Definition

Wireless streaming requires video coding to be robust to channel impairments and adaptable to the network and diverse scenarios; wireless video adaptation deals with rate adaptation and robustness adaptation.

With the rapid growth of wireless communications and the advance of video coding techniques, wireless video streaming is expected to be widely deployed in the near future. However, due to the characteristics of wireless networks such as high error rate, limited bandwidth, time-varying channel conditions, limited battery power of wireless devices and the diversity of wireless access networks and devices, wireless video streaming faces many challenges. In particular, from the coding point of view, wireless streaming requires video coding to be robust to channel impairments and adaptable to the network and diverse scenarios. Traditional media coding standards such as MPEG-2 and MPEG-4 are not suitable any more since they are targeted to a particular range of bit rates and a particular type of applications. Therefore, the needs for video adaptation have becoming more important as the advance of the wireless video has become widespread. With regard to wireless video streaming, there are generally two issues in video adaptation: rate adaptation and robustness adaptation. These issues are further elaborated in the following.

The objective of rate adaptation is to intelligently remove some information from the video signal itself so that end-to-end resource requirement can be reduced. Generally, video rate adaptation can be implemented in three ways. The first approach is to store many non-scalable bitstreams for each video sequence. Each bitstream is coded at different formats or different spatial/temporal/SNR (signal-to-noise ratio) resolutions. When a user requests to access the video sequence, the server can send the bitstream which is closet to the user's requirements. Although this method usually costs more storage spaces in the video server and the chosen bitstream may not satisfy the user's requirement exactly, it is widely used in practical systems due to its simplicity. Recent research works have focused on the how to efficiently switch among multiple non-scalable bitstreams in order to dynamically adapt to the time-varying network conditions

The second approach is a popular scheme based on video transcoding [1] including decreasing the spatial resolution, reducing the SNR, or down sampling the temporal frame rate through re-encoding, re-quantization, frame dropping and etc. Although transcoding is very flexible and does not require extra storage space, it needs complex extra processes and is not suitable for large-scale diverse users. The third approach is to use scalable video coding which has inherent ability to adapt video to different requirements. Scalable video coding schemes aim at encoding a video once and decode it at multiple reduced rates and resolutions to provide simple and flexible adaptability. However, since scalable video coding intends to cover a broad range of applications, the complexity of scalable video coding is typically very high, which may limit its usage on real-time wireless video applications.

The second important issue in wireless video adaptation is the robustness adaptation, for which we try to add controlled redundancy in video bitstreams for reliable transmission over wireless channels. Robust video adaptation can be generally implemented in two ways. The first approach is to build error resilience features into the coding scheme itself. Typical error resilience features for coding include the use of resynchronization markers to recover from decoding errors and intra-coded blocks to minimize error propagation. The second approach is to use the joint source-channel coding techniques which involve both source coding and channel coding to combat possible channel errors. We have seen extensive studies on FEC (forward error correction) based joint source-channel coding for robust video transmission. The common idea of most joint source-channel coding schemes is unequal error protection, i.e., the more important information is given more protection [2]. How to define and protect the important video information has been the main research focus in joint source-channel coding.

Besides the traditional low-level video adaptation, recent advance in video content analysis has introduced new space for wireless video adaptation such as semantic event based adaptation, structural-level adaptation, and video skimming [3]. For example, object-based video transcoding can be used to transmit a subset of objects for adaptive content delivery. On the other, recent advance in wireless network QoS supports has also brought in new challenges in wireless video adaptation such as cross-layer video adaptation.

Cross-References

► Wireless Video

References

1. A. Vetro, C. Christopoulos, and H. Sun, "Video Transcoding Architectures and Techniques: an Overview," IEEE Signal Processing Magazine, March 2003, pp. 18–29.
2. Z. He, J. Cai, and C.W. Chen, "Joint Source Channel Rate-Distortion Analysis for Adaptive Mode Selection and Rate Control in Wireless Video Coding," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 12, June 2002, pp. 511–523.
3. S.-F. Chang and A. Vetro, "Video Adaptation: Concepts, Technologies, and Open Issues," Proceedings of the IEEE, Vol. 93, January 2005, pp. 148–158.

WMV-9 Video Compression

Definition

Windows Media Video 9 (WMV-9) is a video codec developed by Microsoft, which is widely used for streaming media over Internet due to the popularity of MS Windows operating systems.

Since WMV-9 is a generic coder, many algorithms/tools of it can be used for a variety of applications under different operating conditions. Originally, three profiles were defined – Simple Profile, Main Profile, and Complex Profile. However, Complex Profile was dropped unofficially. Consequently, WMV-9 more focuses on compression technology for progressive video up to Main Profile, while VC-1, a derivative of WMV-9, is developed for broadcast interlaced video as well as progressive video. Those two technologies are almost identical in important tools except the interlace tool, and VC-1 is currently under standardization by the Society of Motion Picture and Television Engineer (SMPTE).

Key Compression Tools for WMV-9 Video

Like all other MPEG standards, WMV-9 is based on motion compensated transform coding. Originally YUV4:2:0 and YUV4:1:1 were defined as input formats for progressive and interlaced video, respectively. Since interlaced video is not considered with WMV-9 anymore, 8bit YUV4:2:0 is the only input format. There is no fixed GOP structure in WMV-9. And, I, P, B, BI and Skipped P are defined as pictures/frames. Unlike MPEG standards, I (Intra) frame doesn't happen periodically. Any reference can take on I or P (Predicted) frame (except the fist frame – I). Therefore, if there is no big scene change for a lengthy period of time, there could be only P frames as references. However, the number of B frames (Bi-directionally predicted frames) between two reference frames is fixed. Maximally, there could be seven B frames. BI frames are almost identical to I frames. If there is big scene change continuously, some B frames can not capture any similarity from two reference frames. In such cases, intra mode performance might be better than prediction mode performance – BI frame compression is a good choice in such cases. Since BI is not used as a reference, frame dropping based on the ASF file format is possible under certain conditions – any reasons like lack of computation or bandwidth. The last one is Skipped P frame – if the total length of the data comprising a compressed frame is 8 bits, then this signals that the frame was coded as a non-coded P frame in an encoder. A key compression tool in WMV-9 is adaptive block size transform. Transform block size can change adaptively, while size of motion compensation is either 16×16 or 8×8 in WMV-9. Note that this is quite the opposite to that of H.264. H.264 normally uses fixed size 4×4 or 8×8 transforms with various adaptive prediction size of motion compensation. There are four transform sizes – 8×8, 4×8, 8×4 and 4×4. The transforms are 16 bit transform where both sums and products of two 16 bit values produce results within 16 bits – the inverse transform can be implemented in 16 bit fixed point arithmetic. Note that the transform approximates a DCT, and norms of basis function between transforms are identical to enable the same quantization scheme through various transform types [1]. There are three options for motion compensation:

1. Half-pel or quarter-pel resolution motion compensation can be used,
2. Bi-cubic or bi-linear filter can be used for the interpolation, and
3. 16×16 or 8×8 block size can be used.

These are all combined into a single motion compensation mode to be represented in the Frame level. There is a Sequence layer mode FASTUVMC for motion vector computation of Chroma components. If this is on, computed Chroma MVs are all rounded to half-pel. Thus, interpolation for quarter points is not necessary for Chroma data. Quantization is defined with two parameters, generally specified in video standards – (Qp, Dead-zone). There are two choices about Dead-zone in WMV-9 – 3Qp and 5Qp.

There are two levels where this can be described:

1. Sequence header has QUANT field for this description – 3Qp or 5Qp for entire sequence.
2. Explicit option writes in each Picture header, or Implicit option is to describe it through PQindex.

In I frames, PQAUNT is applied to entire MBs. However, DQUANT is used to adaptively describe Qp in each MB in P/B frames. There are other options to use only two Qps for an entire frame depending on either boundary MB or non-boundary MB, too. There two techniques used in WMV-9 to reduce blocky effect around transform boundary – Overlapped Transform (OLT) smoothing and In Loop deblocking Filtering (ILF). OLT is a unique and interesting technique to reduce blocky effect based on a finely defined pre/post-processing pair. The idea is that two forward and inverse operations are defined in such a way that original data are recovered perfectly when operations are serially applied (forward and inverse).

The forward transform is to exchange information around boundary edges in adjacent blocks. The forward operation is performed before main coding stage. Let's say one block preserves relatively good edge data, while the other block looses details of edge data. In this case, the blocky effect is so visible. In decoder side, inverse operation is required to exchange the edge data back again to original data. By doing so, good quality and bad quality edges diffuse each other. Therefore, looking of blocky effect is much lessened. On the other hand, ILF is more or less heuristic way to reduce blocky effect. Blocky pattern is considered high frequency since abrupt value changes occur around block edges.

Considering original data quality, relatively simple low pass filtering is applied about block edges in ILF. ILF is performed on reference frames I and P. Thus, the result of filtering effects quality of the next pictures that use ILFed frames as references. Entropy coding used in WMV-9 is a similar to Context-Adaptive VLC. Based on Qp, from which the original quality can be guessed, a new set of VLC tables is introduced. Such examples include Mid-rate VLC tables and High-rate VLC tables. In addition, based on MVs, another set of VLC tables is introduced. Such examples include Low-motion DC differential tables and High-motion DC differential tables.

WMV-9 Video Specific Semantics and Syntax

There are five levels of headers inWMV-9 video bitstream syntax – Sequence, Picture, Slice (not clearly defined in WMV-9), MB and Block. Sequence header contains basic parameters such as profile, interlace, frame rate, bit rate, loop filter, overlap filter and some other global parameters. Picture header contains information about type of picture/BFraction/PQindex/LumScale /LumShift/DQUANT related/TTMBF/TTFRM/DCTACMBF/DCTACFRM, etc. BFraction data is relative temporal position of B that is factored into the computation of direct mode vectors. Note that this can be overridden with a value that has nothing to do with geometrical position. PQindex is interpreted for QS and quantizer types (3QP/5QP) in Implicit case, while quantizer types are explicitly defined in Sequence or Picture header in other cases. LumScale/LumShift are Intensity Compensation parameters. TTMBF is the flag that tells whether additional field for Transform Type is in MB level or Frame level. DCTACMBF is the flag that tells whether DCT AC Huffman table is defined in MB level or Frame level. Slice is not defined clearly in WMV-9. When STARTCODE is set in the Sequence header however, Slice header can be defined in long startcode to provide a mechanism for synchronization. MB header contains SkipMBbit/MVmodebit (1MV/4MV option)/MVDATA/TTMB, etc. MVDATA tells whether the blocks are coded as Intra or Inter type. If they are coded as Inter, then MVDATA indicates MV differentials. Block layer contains all transform coefficients related data. Sub-block pattern data is included to sub-divide the block.

Cross-References

► Video Coding Techniques and Standards

References

1. S. Srinivassan, P. Hsu, T. Holcomb, K. Mukerjee, S. Regunathan, B. Lin, J. Liang, M.-C. Lee, and J. Ribas-Corbera, "WMV-9: Overview and Applications," Signal Processing Image Communication, Vol. 19, No. 9, October 2004, pp. 851–875.

Workflow Computing

Definition

Workflow Management Systems (WfMS) have been defined as "technology based systems that define, manage, and execute workflow processes through the execution of software whose order of execution is driven by a computer representation of the workflow process logic" [1].

This limits the usability of WfMS in a world where constant adaptation to new situations is necessary and where teams are increasingly mobile and distributed. Workflow management systems are typically organizationally aware because they contain an explicit representation of organizational processes. In recent years there have been considerable attempts to merge workflow, groupware, and knowledge management technologies. Industrial research labs and product teams have made significant steps forward. A WfMS can impose a rigid work environment on users, which often has a consequence. One example is among users who perform time-consuming manual "work around" the consequence is lower efficiency and dissatisfaction with the system.

Workflow automation provides unique opportunities for enabling and tracking information flow as well as monitoring of work performance. As a consequence, WfMS enable continuous loops of sub processes such as goal setting, working, monitoring the work, measuring performance, recording and analyzing the outputs and evaluating the "productivity" of personnel. Users of WfMS often consider the controlling and monitoring possibilities as a "dark side" of these systems, which results in demotivating employees. A business process has well defined inputs and outputs and serves a meaningful purpose either inside or between organizations. Business processes and their corresponding workflows exist as logical models. When business process models are executed they have specific instances. When instantiated, the whole workflow is called a work case.

The WfMS enacts the real world business process for each process instance. A business process consists of a sequence of activities. An activity is a distinct process step and may be performed either by a human agent or by a machine. Any activity may consist of one or more tasks. A set of tasks to be worked on by a user (human agent or machine) is called work list. The work list itself is managed by the WfMS. The WfMC calls the individual task on the work list work item. To summarize, a workflow is the instantiated (enacted or executed) business process, either in whole or in parts. During enactment of a business process, documents, which are associated to tasks, are passed from one task participant to another. In most cases this passing of documents or executing applications is performed according to a set of rules. A WfMS is responsible for control and coordination such as instantiating the workflow, assigning human or non-human agents to perform activities, generating worklists for individuals, and routing tasks and their associated objects such as documents between the agents. For an in-depth analysis of Workflow computing we refer to [2] and for a discussion of the hybrid systems integrating Workflow and Groupware Computing we refer to [3].

Cross-References

► Collaborative Computing – Area Overview

References

1. Workflow Management Coalition (WfMC), Workflow Management Specification Glossary, Available at: http://www.wfmc.org.
2. D. Georgakopoulos, M. Hornick, and A. Sheth, "An Overview of Workflow Management: From Process Modeling to Workflow Automation Infrastructure," Distributed and Parallel Databases, Vol. 3, 1995, pp. 119–153.
3. S. Dustdar, "Caramba – A Process-Aware Collaboration System Supporting Ad Hoc and Collaborative Processes in Virtual Teams, Distributed and Parallel Databases," Special Issue on Teamware Technologies, Kluwer Academic, Hingham, MA, Vol. 15, No. 1, January 2004, pp. 45–66.