Video Communication – Advances and Applications

Bernd Girod
Telecommunications Laboratory
University of Erlangen-Nuremberg
Cauerstrasse 7, 91058 Erlangen, Germany
e-mail: girod@nt.e-technik.uni-erlangen.de

Invited Paper

ABSTRACT

This overview paper considers the applications driving today's progress in video communication and then reviews the principles behind advanced video compression technology, such as standardized in ITU-T Rec. H.263 or ISO MPEG. In addition, we look at the problem of robustness against packet loss, that has to be solved to transmit video over wireless channels or the Internet. Addressing promising directions for the future development of video compression, we discuss and illustrate the performance of multiframe prediction methods and advanced motion models.

1 INTRODUCTION

More than 100 years after 23-year-old engineering student Paul Nipkow's extraordinary vision of an "Electrical Telescope" [1], video communication from any point to any other at any time is finally in reach. Advanced video compression algorithms enable applications unthinkable only a few years ago, such as digital video broadcasting, video streaming over the Internet, or mobile videophones.

In recent years, several standards such as ITU-T H.261, H.263, and ISO MPEG-1, MPEG-2 have been introduced which address the compression of video for digital storage and communication services. They all build on the idea of a hybrid video coding scheme which comprises block-based motion-compensated prediction (MCP) and Discrete Cosine Transform (DCT) coding of the prediction error. Also the new MPEG-4 standard follows this approach.

In this paper, we review advances in video communication and compression in the last years. We start out by considering the applications that are driving today's progress in Section 2. Section 3 will then discuss the state-of-the-art in video compression technology as employed with today's standards. In particular, we emphasize rate-constrained coder control and recent error resiliency techniques that have been developed for video transmission over error-prone networks. Finally, in Section 4, we address promising directions of today's research.

2 EMERGING APPLICATIONS

Advanced digital video communication technology can be used as a superior substitute for traditional analog technology, as is the case for broadcasting, or it can enable new services that were not previously possible. We briefly discuss emerging applications in the following that motivate research and development of more advanced video communication technology today.

2.1 Digital Video Broadcasting

Spectrum for terrestial television broadcasting has always been a scarse resource. Even worse, television competes with other services for the allocation of frequency bands. Cable systems are used to provide additional bandwidth in most countries. In addition, direct satellite broadcasting of television programs has been gaining momentum. Transmission bandwidth will remain expensive in the future and should be used as efficiently as possible. With modern compression techniques, it is possible to increase the overall number of channels and transmit more programs at the same cost.

The picture quality provided by today's analog television systems is limited. Many of the impairments of the picture at the home receiver are due to noise interfering with the transmission and multipath propaga-

0-7803-5796-5/99/$10.00 © 1999 IEEE

tion. Digital video broadcasting, on the other hand, is robust against such effects over a wide range of conditions. To match the picture quality projected in movie theaters, high definition television (HDTV) systems with approximately twice the horizontal and vertical resolution have been developed, but it remains doubtful to what extent broadcasters will prefer more pixels over more channels.

Today, the ISO MPEG-2 standard is firmly established for Digital Video Broadcasting around the world [2] [3]. Support for line-interlaced scanning is still viewed as an important requirement by many, and, in that regard, MPEG-2 video is without competition. Typical bit-rates range from 2–8 Mbps for ITU-R 601 resolution material. Modern digital modulation schemes can squeeze up to about 20 Mbps through a terrestial TV broadcasting channel (6 MHz bandwidth), or up to about 40 Mbps through a cable or satellite TV channel [2] [3], hence the number of channels is readily multiplied by the use of digital technology.

2.2 Internet Video Streaming

A "video-on-demand" service operating over a future broadband interactive digital network could basically replace the local video rental store, delivering a compressed movie from a central data base to the home at any desired time with only a short latency. The bit-rates required would correspond to those for video broadcasting, with the important difference that the audience for each movie would be one or a few, rather than millions, and, therefore, the load on the network is increased by orders of magnitude.

In the mean-time, video-on-demand has taken off with much lower bit-rate (typically tens of kbps), and hence lower resolution and picture quality, in the form of Internet video streaming. First products appeared in 1995 and spread explosively. By the time this paper appears in print, the number of Internet video players installed on personal computers will be approximately 100 million. Live audiences for Internet video multicasting have been larger than one million viewers. The proprietary compression schemes employed resemble ITU-T Rec. H.263.

The Internet Protocol is evolving as the ubiquitous interface of open communication systems, thus enabling seamless communication across network boundaries and a broad range of new applications. Internet video streaming is well positioned to ultimately displace conventional "closed architecture" video broadcasting systems, once the bandwidth bottlenecks of today's Internet infrastructure have been removed.

2.3 Videoconferencing/Videotelephony

Videoconferencing systems operating over ISDN at bit-rates of px64 kbps (p=1,2) are in widespread use today in the business world. Most of them still follow the ITU-T Rec. H.261 for video compression, but the transition to the more advanced H.263 is proceeding quickly. As with Internet video streaming, the low bit-rates require very significant compression. Picture quality is typically acceptable at CIF resolution (Common Intermediate Format, 352×288 pixels for luminance) and frame rates up to 15 fps for limited motion videoconferences. The driving force for H.263 was the transmission of motion video at even lower bitrates over analog phone lines using, e.g., the V.34 standard at 28.8 kbps, to enable ubiquituous consumer videophones. Compared to H.261, H.263 requires about one half the bit-rate at the same picture quality [4]. Consumer videophones based on this standard are widely available today, typically operating at QCIF resolution (176×144 pixels for luminance) and frame rates up to 15 fps. Picture quality is acceptable for head-and-shoulder sequences with little motion.

There are in the order of 1 million operational videoconferencing and videotelephony systems in the world today. This number is small compared to the 100 million Internet video players or the more than one billion (analog) TV receivers. Nevertheless, videoconferencing and videotelephony has been a major technological driver for more advanced compression algorithms in the past, and will likely continue to play this role for some time in the future.

2.4 Wireless Video

Video communication is expected to be an integral part of future third-generation wireless communication services. The current vision includes a small handheld device that allows the user to communicate from anywhere in the world with anyone in a variety of formats (voice, data, image, and full-motion video) from virtually any geographic location. This next generation of wireless multimedia communicators is expected to be equipped with a camera, a microphone, and a liquid crystal color display, serving both as a videophone and computer screen. With progressing miniaturisation of components, wristwatch "Dick Tracy" communicators are expected to follow soon after.

Of all modalities desirable for future mobile multimedia systems, motion video is the most demanding in terms of bit-rate. Today's "second-generation" cellular telephony networks, such as Global System for Mobile Communications (GSM), typically provide 10 - 15 kbps, suitable for compressed speech, but too little for motion video. Fortunately, the standardisation of higher-bandwidth networks, such as Universal Mobile Telecommunications System (UMTS) [5] [6], is well underway, and, together with continued progress in video compression technology, wireless multimedia communicators with picturephone functionality and Internet videoserver access will soon be possible.

3 VIDEO COMPRESSION TECHNOLOGY

3.1 Motion-Compensated Hybrid Coding

All modern standardized video compression algorithms are motion-compensated hybrid coders. These coders estimate the displacement from frame to frame and transmit the motion vector field as side information in addition to the motion-compensated prediction error image. The prediction error image is compressed by an intraframe coder that exploits statistical dependencies between adjacent samples. The intraframe coder is an 8×8 DCT coder in all current video coding standards, but other schemes, such as subband coders or vector quantizers, can be used as well.

Motion-compensated hybrid coding can theoretically outperform an optimum intraframe coder by at most 0.8 bits/pixel in moving areas of an image, if motion compensation is performed with only integer-pixel accuracy [7] [8]. For half-pixel accuracy, as employed in most codecs today, this gain can be up to 1.8 bits/pixel. In addition, in nonmoving areas (or other parts of the image that can be predicted perfectly) no prediction error signal has to be transmitted and these areas can simply be repeated from a frame store, a technique often referred to as conditional replenishment.

Motion compensation works well for low spatial frequencies in the video signal; for high frequency components even a small inaccuracy of the motion compensation will render the prediction ineffective. Hence, it is important to spatially filter the prediction signal by a loop filter [7]. This loop filter is explicitly needed for integer-pixel accurate motion compensation. For subpixel accurate motion compensation, it can be incorporated into the interpolation kernel required to calculate signal samples between the original sampling positions [8]. The loop filter also improves prediction by acting as a noise reduction filter. Prediction can be further improved by combining multiple independently motion-compensated prediction signals. Examples are the bidirectionally predicted B-frames in MPEG and H.263 [9] or overlapped block motion compensation [10] that has also been incorporated in H.263.

3.2 Rate-Constrained Coder Control

Motion-compensated hybrid coders standardized today are forward-adaptive systems, i.e., the "intelligent" encoder decides about the motion vector and the coding mode, while the "dumb" decoder simply follows the encoder decisions. E.g., with H.263, the coder can not only adjust the quantizer stepsize, but choose between inter-prediction modes INTER, INTER-4V which consist of motion compensation and residual coding, as well as the UNCODED mode which indicates copying the macroblock from the previous frame without residual coding, and the INTRA mode where no reference to another frame is made [4]. As video compression standards only specify the bitstream syntax, a wide range of motion estimation, mode decision, and rate control algorithms can be employed. The impact of these, often proprietary, schemes on the resulting picture quality is substantial.

Especially at low bit-rates, motion compensation and mode decision are severely constrained by the limited bit-rate available to transmit the motion vector field as side information. The problem of the best rate-distortion trade-off for motion compensation can be elegantly solved by minimizing a Lagrangian cost function

$$J_{MOTION} = D_{DFD} + \lambda_{MOTION} R_{MOTION}, \qquad (1)$$

in which the distortion D_{DFD}, representing the mean squared prediction error, is weighted against the rate R_{MOTION} associated with the motion parameters [11]. Such a formulation is familiar, e.g., from entropy-constrained vector quantization (ECVQ) [12]. The Lagrange multiplier λ_{MOTION} incorporates the rate constraint and its value directly controls the rate-distortion trade-off.

A similar technique can be used for rate-constrained mode decision, by minimizing

$$J_{MODE} = D_{REC} + \lambda_{MODE} R_{TOTAL}, \qquad (2)$$

for each "macroblock" (16×16 luminance pixels and the corresponding chrominance pixels), see, e.g., [13].

Here, the distortion D_{REC} after encoding and reconstruction of the motion-compensated prediction error is weighted against bit-rate using another Lagrange multiplier λ_{MODE}. Different from rate-constrained motion estimation, the total bit-rate R_{TOTAL} for a particular macroblock mode is considered, including the rates for the macroblock header, motion vector and quantized DCT coefficients.

Simulation results illustrating the advantage of rate-constrained coder control in an H.263 codec are shown in Fig. 1. The baseline for our comparison is the TMN-2.0 codec, a publicly available software H.263 encoder [14]. The TMN-2.0 encoding scheme does not employ rate-constrained motion estimation and uses heuristic criteria for the mode decision. In contrast, our rate-distortion optimized H.263 coder uses rate-constrained motion estimation as well as rate-constrained mode decision. The rate-constrained motion vector and mode control technique outlined above has been adopted by the ITU-T as the latest H.263 reference model TMN-10 [15].

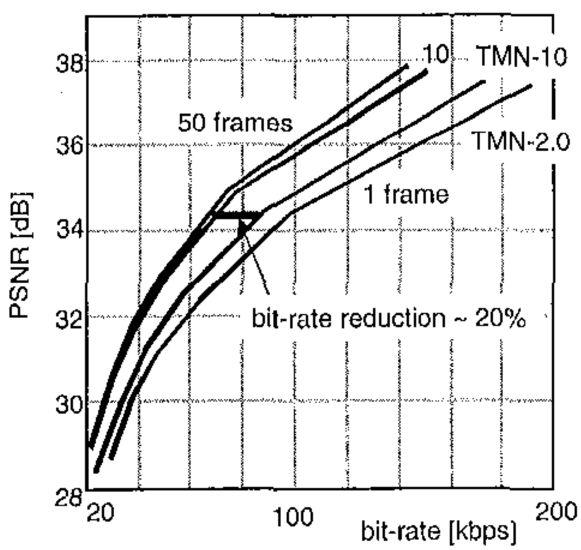

Figure 1: Rate-distortion curves for an H.263 coder with various enhancements. TMN-2.0: single-frame coder with heuristical coder control. TMN-10: single-frame coder with Lagrangian coder control. Other curves: multiframe prediction with 10 and 50 previous frames. Testsequence *Foreman*, QCIF, 10 fps.

3.3 Error-Resilient Video Coding

Over the last years, there has been an increased interest in techniques that allow to transmit video over error-prone channels. This is motivated mainly by video communication over congestion-sensitive packet networks, such as ATM or the Internet, or over mobile wireless radio links, where fading can lead to burst errors or even the intermittent loss of the connection. Without special measures, compressed video signals are extremely vulnerable against transmission errors. With interframe coding, the previous encoded and reconstructed video frame is used to predict the next frame. Therefore, the loss of information in one frame has considerable impact on the quality of the following frames. Since some residual transmission errors will inevitably corrupt the video bit-stream, this vulnerability precludes the use of interframe coding schemes designed for error-free channels without special measures. These measures have to be build into the video coding and decoding algorithms themselves and form the "last line of defense" if traditional error control techniques like FEC and ARQ fail.

A comprehensive review of the great variety of error control and concealment techniques that have been proposed during the last 10 - 15 years has been presented in an excellent paper by Wang and Zhu recently [16]. For example, one can partition the bit-stream into classes of different error sensitivity (often refered to as data partitioning) to enable the use of unequal error protection [17] [18] [19] [20]. Data partitioning has been included as an error resiliency tool in the MPEG-4 standard [21]. Unequal error protection can significantly increase the robustness of the transmission and provide graceful degradation of the picture quality in case of a deteriorating channel. Since unequal error protection does not incorporate information about the current state of the channel, the design of such a scheme is a compromise that accomodates a range of operating conditions. Feedback-based techniques, on the other hand, can adjust to the varying transmission conditions rapidly and make more effective use of the channel.

The ITU-T Study Group 16 has adopted feedback-based error control in their effort towards mobile extensions of H.263 [22]. The first version of H.263 already included *Error Tracking*, a technique that allows the encoder to accurately estimate interframe error propagation and adapt its encoding strategy to mitigate the effects of past transmission errors [23] [24].

With the second version, informally known as H.263+, an optional multiframe technique (Reference Picture Selection, Annex N) for feedback-based error control was adopted. Future enhancements, for example, data partitioning, unequal error protection, and reversible variable length coding, are under consideration for future versions of the standard.

4 PROMISING DIRECTIONS

4.1 Multiframe Prediction

In the early 1980s, video compression made the leap from intra-frame to inter-frame algorithms. Significantly lower bit-rates were achieved at the expense of memory and computational requirements that were two orders of magnitude larger. Today, with continuously dropping costs of semiconductors, we might soon be able to afford another leap by dramatically increasing the memory and computation power in video codecs. Algorithms to take advantage of increased memory and computation power are still in their infancy. This has been our motivation to initiate research into multiframe motion-compensated prediction.

Long-term memory MCP is a multiframe MCP method that extends the motion vector utilized in hybrid video coding by a variable time delay permitting the use of several decoded frames instead of only the previous one for block-based motion compensation. The frames inside the long-term memory which is built in synchrony at encoder and decoder are addressed by a combination of the codes for the spatial displacement vector and the variable time delay. As the additional transmission of the variable time delay potentially increases the bit-rate, we again need to employ rate-constrained parameter estimation minimizing a Lagrangian cost function (1).

Simulation results for integrating long-term memory prediction into an H.263 codec are also included in Fig. 1. The number of previous frames is increased to $M = 10$ and $M = 50$, respectively. Reconstruction PSNR improvements of about 1.5 dB are apparent in comparison to the single-frame TMN-10 H.263 coder. These PSNR improvements correspond to bit-rate savings of about 20 %. $M = 50$ corresponds to the previous 5 seconds at a frame rate of 10 fps. In the future, we might want to increase the number of previous frames to hundreds or thousands and also use a linear combination of previous frames, as suggested in [25]. For more details on long-term memory MCP, please

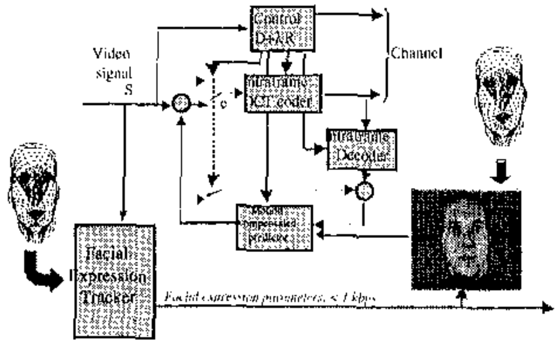

Figure 2: H.263 coder extended by multiframe prediction incorporating a 3-D head model.

refer to [26]. Interestingly, multiframe prediction can also improve the error resiliency of video coders [27]. ITU-T Study Group 16 has recently decided to incorporate long-term memory prediction as part of the upcoming revision of its low bit-rate video coding Recommendation, nick-named H.263++.

4.2 Advanced Motion Models

Video compression is a mature discipline today, resting solidly on the foundations of source coding theory. Often, practical schemes perform close to their information-theoretic bounds. Most of these bounds are, however, calculated on the basis of crude statistical models about the structure of images. As these models become more refined, compression ratios can improve further.

For simplicity, most practical video coding schemes today still employ block-wise constant motion compensation. More advanced schemes have been proposed that interpolate between motion vectors, employ arbitrarily shaped regions, or use triangular meshes to represent a smooth motion vector field. A particularly promising approach uses multiple affine motion models

$$d_x = ax + by + c \qquad (3)$$

$$d_y = dx + ey + f$$

where d_x and d_y are the horizontal and vertical components of the motion vector, x, y are the spatial coordinates, and a, b, c, d, e, f are six coefficients that have to be estimated and encoded to represent the motion vector field [28] [29]. Affine motion models can be elegantly incorporated into the multiframe prediction framework [30] and even combined with long-term memory prediction [31]. Bit-rate savings of up to 35% compared to H.263 TMN-10 have been reported in [31].

For videotelephony, we want to transmit the head-and-shoulder view of a talking person. More than 15 years ago, Forchheimer et al. have proposed a videotelephone system based on a computer-animated 3D head model [32] [33], and many groups have investigated such systems since [34]. Impressive progress has been made in the automatic tracking of facial expressions over the last few years, and a model-based videophone might become feasible soon [35]. A major drawback of such a system is still its limitation to a specific 3-D model and hence lack of generality. A solution to this problem has recently been reported by the author's group based on the multiframe prediction frame work discussed in the previous section. For the first time, we succeeded in incorporating a 3-D head model into an H.263 coder in such a way that the combination possesses both the efficiency of the model-based coder (if the model applies) and the generality of the H.263 waveform coder [36] [37] (Fig.2). The H.263 coder extended by the computer-animated 3-D head model achieves bit-rate savings compared to TMN-10 by around 35% at the same PSNR, while maintaining a noticeably higher subjective picture quality for head-and-shoulder sequences. Examples for the substantial picture quality improvement at the same bit-rate is shown in Fig.3.

5 CONCLUDING REMARKS

A presentation of the state-of-the-art cannot be complete in a short overview paper like this one, and the selection of a few promising research direction is obviously highly subjective. The fact that the selected directions coincide with the author's own research activities is, of course, not coincidental -- to be credible, they should! Beyond more efficient compression for motion video, there are many interesting open questions that arise when integrating video codecs into communication systems, where previously neglected requirements, such as robustness, delay, or random access, have to be taken into account. As in the past, we can expect new applications to present their specific set of new challenges that will continue to drive the progress in video communication technology.

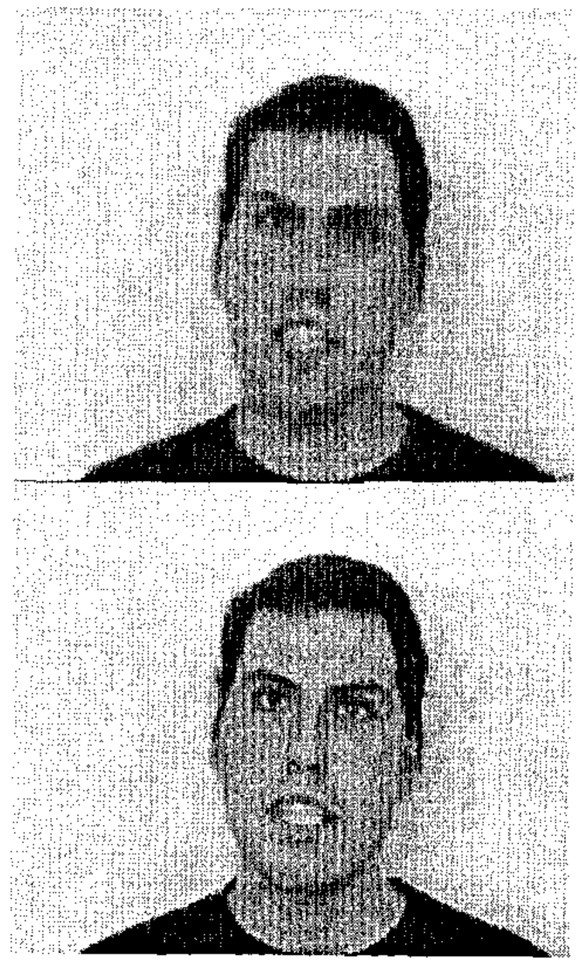

Figure 3: Picture quality improvement by extending H.263 by an animated 3-D head model. Top: H.263 TMN-10, CIF, 8.33 fps, PSNR 33.88 dB, 1680 bits for frame shown. Bottom: Multiframe prediction with 3-D head model PSNR 37.34 dB, 1682 bits for frame shown

References

[1] P. Nipkow, "Elektrisches Teleskop," German Reichspatent no. 30105, 1884.

[2] U. Reimers, "Digital Video Broadcasting," IEEE Communications Magazine, vol. 36, no. 6, pp. 104-110, June 1998.

[3] E. Petajan, "The HDTV Grand Alliance System," IEEE Communications Magazine, vol. 34, no. 6, pp. 126-132, June 1996.

[4] B. Girod, N. Färber, and E. Steinbach, "Performance of the H.263 Video Compression Standard," Journal of VLSI Signal Processing: Systems for Signal, Image, and Video Technology, vol. 17, pp. 101-111, Nov. 1997.

[5] E. Berruto, M. Gudmonson, R. Menolascino, W. Mohr, and M. Pizarroso, "Research Activities on UMTS Radio Interface, Network Architectures, and Planning," IEEE Communications Magazine, vol. 36, no. 2, pp. 82-95, Feb. 1998.

[6] D. Grillo (ed.), Special Section on Third-Generation Mobile Systems in Europe, IEEE Personal Communication Magazine, vol. 5, no. 2, pp. 5-38, April 1998.

[7] B. Girod, "The Efficiency of Motion-Compensating Prediction for Hybrid Coding of Video Sequences," IEEE Journal on Selected Areas in Communications, vol. SAC-5, no. 7, pp. 1140 - 1154, August 1987.

[8] B. Girod, "Motion-Compensating Prediction with Fractional-Pel Accuracy," IEEE Transactions on Communications, vol. 41, no. 4, pp. 604 - 612, April 1993.

[9] B. Girod, "Why B-Pictures Work: A Theory of Multihypothesis Motion-Compensated Prediction," Proc. IEEE International Conf. on Image Processing ICIP-98, Chicago, vol. II, pp. 213 - 217, Oct. 1998.

[10] M. T. Orchard and G. J. Sullivan, "Overlapped Block Motion Compensation: An Estimation-Theoretic Approach," IEEE Trans. Image Processing, vol. 3, no. 5, pp. 693 - 699, Sept. 1994.

[11] B. Girod, "Rate-constrained Motion Estimation", in Proc. Visual Communication and Image Processing VCIP '94, Chicago, IL, SPIE vol. 2308, pp. 1026-1034, Sept. 1994.

[12] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy-Constrained Vector Quantization," IEEE Trans. ASSP, vol. 37, no. 1, pp. 31-42, Jan. 1989.

[13] B. Girod, T. Wiegand, E. Steinbach , Markus Flierl, and Xiaozheng Zhang, "High-Order Motion Compensation for Low Bit-Rate Video", in Proc. EUSIPCO-98, Rhodes, Greece, Sept. 1998.

[14] Telenor Research, "TMN (H.263) Encoder/Decoder, Version 2.0," Download via anonymous ftp to `bonde.nta.no`, June 1997.

[15] G. J. Sullivan and T. Wiegand: "Rate-Distortion Optimization for Video Compression," IEEE Signal Processing Magazine, November 1998.

[16] Y. Wang and Q.-F. Zhu, "Error Control and Concealment for Video Communication: A Review," Proc. of the IEEE, vol. 86, no. 5, pp. 974 - 997, May 1998.

[17] R. Fischer, P. Mangold, R.M. Pelz, and G. Nitsche, "Combined Source and Channel Coding for Very Low Bitrate Mobile Visual Communication Systems," in Proc. Int. Picture Coding Symposium, pp. 231 - 236, 1996.

[18] K. Illgner and D. Lappe, "Mobile multimedia communications in a universal telecommunications network," in Proc. of SPIE Conf. on Visual Communications and Image Processing, pp. 1034-1043, 1995.

[19] B. Girod, U. Horn, and B. Belzer, "Scalable video coding with multiscale motion compensation and unequal error protection," in Multimedia Communications and Video Coding, Y. Wang, S. Panwar, S.P. Kim, and H.L. Bertoni (eds.), New York: Plenum Press, 1996, pp. 475-482.

[20] R. M. Pelz, "An unequal error protected px8 kbit/s video transmission for dect," in Vehicular Technology Conference, pp. 1020-1024, 1994.

[21] R. Talluri, "Error-Resilient Video Coding in the MPEG-4 Standard," IEEE Communications Magazine, vol. 36, no. 6, pp. 112 - 119, June 1998.

[22] N. Färber, B. Girod, and J. Villasenor, "Extensions of the ITU-T Recommendation H.324 for Error Resilient Video Transmission," IEEE Communications Magazine, vol. 36, no. 6, pp. 120 - 128, June 1998.

[23] E. Steinbach, N.Färber, and B. Girod, "Standard Compatible Extension of H.263 for Robust Video Transmission in Mobile Environments," IEEE Transactions on Circuits and Systems for Video Technology, vol. 7, no. 6, pp. 872-881, Dec. 1997.

[24] B. Girod, N. Färber, and E. Steinbach, "Error-Resilient Coding for H.263," in Insights into Mobile Multimedia Communication, D. Bull, N. Canagarajah, A. Nix (eds), Academic Press, pp. 445-459, 1999.

[25] M. Flierl, T. Wiegand, and B. Girod, "A Locally Optimal Design Algorithm for Block-Based Multi-Hypothesis Motion-Compensated Prediction," Proc. Data Compression Conference, Snowbird, UT, April 1998.

[26] T. Wiegand, X. Zhang, and B. Girod, "Long-Term Memory Motion-Compensated Prediction," IEEE Transactions on Circuits and Systems for Video Technology, vol. 9, no. 1, pp. 70-84, Feb. 1999.

[27] T. Wiegand, Niko Färber, and B. Girod, "Error-Resilient Video Transmission Using Long-Term Memory Motion-Compensated Prediction," IEEE Journal on Selected Areas in Communications, special issue on Error-Resilient Image and Video Coding, accepted.

[28] M. Karczewicz, J. Niewegkowski, and P. Haavisto, "Video Coding Using Motion Compensation with Polynomial Motion Vector Fields," Signal Processing: Image Communication, vol. 10, pp. 63-97, 1997.

[29] T. Wiegand, E. Steinbach, A. Stensrud, and B. Girod, "Multiple Reference Picture Video Coding Using Polynomial Motion Models," in Proc. SPIE Symposium on Visual Communications and Image Processing (VCIP'98), San Jose, CA, Jan. 1998.

[30] E. Steinbach, T. Wiegand, and B. Girod, "Using Multiple Global Motion Models for Improved Block-Based Video Coding," Proc. IEEE International Conference on Image Processing, ICIP-99, Kobe, Japan, October 1999.

[31] T. Wiegand, E. Steinbach, and B. Girod, "Long-Term Memory Prediction with Affine Motion Compensation," Proc. IEEE International Conference on Image Processing, ICIP-99, Kobe, Japan, October 1999.

[32] R. Forchheimer, O. Fahlander, and T. Kronander, "Low bit-rate coding through animation," Proc. PCS (Picture Coding Symposium) '83, pp. 113-114, Davis, California, March 1983.

[33] R. Forchheimer, O. Fahlander, and T. Kronander. "A semantic approach to the transmission of face images." Proc. Picture Coding Symposium (PCS '84), number 10.5, Cesson-Sevigne, France, July 1984.

[34] D. Pearson, "Developments in Model-Based Video Coding," Proc. IEEE, vol. 83, pp. 892-906, 1995.

[35] P. Eisert and B. Girod, "Analyzing Facial Expressions for Virtual Conferencing," IEEE Computer Graphics and Applications, Special issue: Computer Animation for Virtual Humans, pp. 70-78, September 1998.

[36] P. Eisert, T. Wiegand, and B. Girod, "Rate-Distortion Efficient Video Compression Using a 3-D Head Model," Proc. IEEE International Conference on Image Processing, ICIP-99, Kobe, Japan, Oct. 1999,

[37] P. Eisert, T. Wiegand, and B. Girod, "Rate-Distortion-Optimized Video Coding Using a 3-D Head Model," IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Three-Dimensional Video Technology, accepted.