Multiple Reference Motion Compensation: A Tutorial Introduction and Survey

Multiple Reference Motion Compensation: A Tutorial Introduction and Survey

Athanasios Leontaris

Dolby Laboratories, Inc.
Burbank, CA
USA
aleon@dolby.com

Pamela C. Cosman

University of California, San Diego
La Jolla, CA
USA
pcosman@ucsd.edu

Alexis M. Tourapis

Dolby Laboratories, Inc.
Burbank, CA
USA
atour@dolby.com

now
the essence of knowledge

Boston – Delft

Full text available at: http://dx.doi.org/10.1561/2000000019

Foundations and Trends® in Signal Processing

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is A. Leontaris, P. C. Cosman and A. M. Tourapis, Multiple Reference Motion Compensation: A Tutorial Introduction and Survey, Foundations and Trends® in Signal Processing, vol 2, no 4, pp 247–364, 2008

ISBN: 978-1-60198-252-0
© 2009 A. Leontaris, P. C. Cosman and A. M. Tourapis

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1-781-871-0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Foundations and Trends® in
Signal Processing
Volume 2 Issue 4, 2008

Editorial Board

Editorial Scope

Foundations and Trends® in Signal Processing will publish survey and tutorial articles on the foundations, algorithms, methods, and applications of signal processing including the following topics:

- Adaptive signal processing
- Audio signal processing
- Biological and biomedical signal processing
- Complexity in signal processing
- Digital and multirate signal processing
- Distributed and network signal processing
- Image and video processing
- Linear and nonlinear filtering
- Multidimensional signal processing
- Multimodal signal processing
- Multiresolution signal processing
- Nonlinear signal processing
- Randomized algorithms in signal processing
- Sensor and multiple source signal processing, source separation
- Signal decompositions, subband and transform methods, sparse representations
- Signal processing for communications
- Signal processing for security and forensic analysis, biometric signal processing
- Signal quantization, sampling, analog-to-digital conversion, coding and compression
- Signal reconstruction, digital-to-analog conversion, enhancement, decoding and inverse problems
- Speech/audio/image/video compression
- Speech and spoken language processing
- Statistical/machine learning
- Statistical signal processing
- classification and detection
- estimation and regression
- tree-structured methods

Information for Librarians

Foundations and Trends® in Signal Processing, 2008, Volume 2, 4 issues. ISSN paper version 1932-8346. ISSN online version 1932-8354. Also available as a combined paper and online subscription.

Foundations and Trends® in
Signal Processing
Vol. 2, No. 4 (2008) 247–364

DOI: 10.1561/2000000019

Multiple Reference Motion Compensation: A Tutorial Introduction and Survey

Athanasios Leontaris[1], **Pamela C. Cosman**[2]
and Alexis M. Tourapis[3]

[1] *Dolby Laboratories, Inc., Burbank, CA 91505-5300, USA, aleon@dolby.com*
[2] *Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093-0407, USA, pcosman@ucsd.edu*
[3] *Dolby Laboratories, Inc., Burbank, CA 91505-5300, USA, atour@dolby.com*

Abstract

Motion compensation exploits temporal correlation in a video sequence to yield high compression efficiency. Multiple reference frame motion compensation is an extension of motion compensation that exploits temporal correlation over a longer time scale. Devised mainly for increasing compression efficiency, it exhibits useful properties such as enhanced error resilience and error concealment. In this survey, we explore different aspects of multiple reference frame motion compensation, including multihypothesis prediction, global motion prediction, improved error resilience and concealment for multiple references, and algorithms for fast motion estimation in the context of multiple reference frame video encoders.

Contents

1

Introduction

Digital video compression has matured greatly over the past two decades. Initially reserved for niche applications such as videoconferencing, it began to spread into everyday life with the introduction of the Video CD and its accompanying Motion Pictures Experts Group MPEG-1 digital video compression standard in 1993 [51]. Home use became widespread in 1996, when the digital video/versatile disk (DVD) with MPEG-2 compression technology was introduced [48, 52]. Digital video compression also facilitated cable and IP-based digital TV broadcast. At the same time, the increase in Internet bandwidth fueled an unprecedented growth in Internet video streaming, while advances in wireless transmission made mobile video streaming possible.

An example video sequence consisting of two frames is shown in Figure 1.1. A frame contains an array of luma samples in monochrome format or an array of luma samples and two corresponding arrays of chroma samples in some pre-determined color sub-sampling format. These samples correspond to pixel locations in the frame. To compress these two frames, one can encode them independently using a still image coder such as the Joint Photographic Experts Group (JPEG) [50] standard. The two frames are similar (*temporally correlated*), hence more

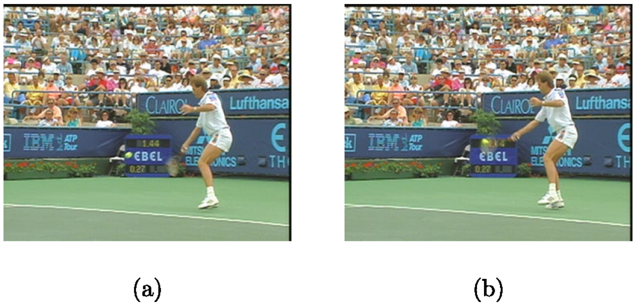
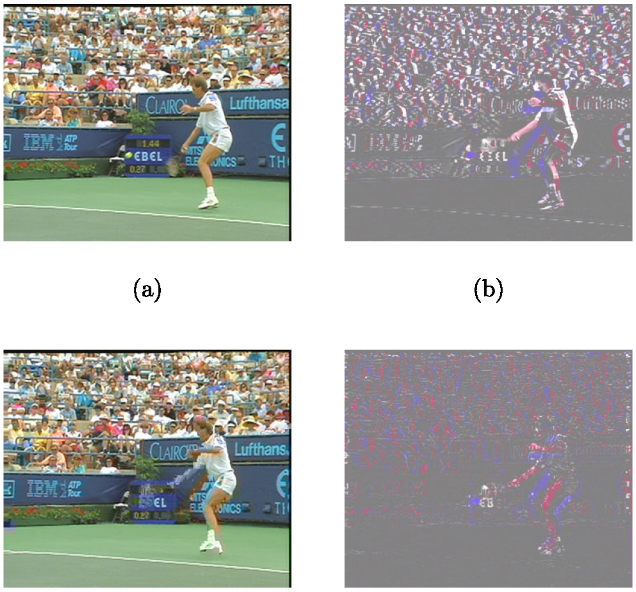


Fig. 1.1 The previous (a) and the current (b) frame of the video sequence.

compression can be obtained if we use the previous frame to help us compress the current frame. One way to do this is to use the previous frame to *predict* the current frame, and then to encode the difference between the actual current frame and its prediction. The simplest version of this process is to encode the difference between the two frames (i.e., subtract the previous frame from the current frame and encode that difference). In this case, the entire previous frame becomes the *prediction* of the current frame. Let i and j denote the spatial horizontal and vertical coordinates of a pixel in a rectangularly sampled grid in a raster-scan order. Let $f_n(i,j)$ denote the pixel with coordinates (i,j) in frame n. Let $\hat{f}_n(i,j)$ denote the predicted value of this pixel. The prediction value is mathematically expressed as $\hat{f}_n(i,j) = f_{n-1}(i,j)$. This technique is shown in the first row of Figure 1.2. For sequences with little motion such a technique ought to perform well; the difference between two similar frames is very small and is highly compressible. In Figure 1.1(b) for example, most of the bottom part of the tennis court will be highly compressed since the difference for these areas will be close to zero. However, there is considerable motion in terms of the player and camera pan from one frame to the next and the difference will be non-zero. This is likely to require many bits to represent.

1.1 Motion-Compensated Prediction

The key to achieving further compression is to *compensate* for this motion, by forming a better prediction of the current frame from some



(a) (b)

(c) (d)

Fig. 1.2 Motion compensated prediction. The top row shows the prediction which is the unaltered previous frame (a) and the resulting difference image (b) that has to be coded. The bottom row shows the equivalent prediction (c) and difference image (d) for motion compensated prediction. The reduction of the error is apparent.

reference frame. A frame is designated as a reference frame when it can be used for motion-compensated prediction. This prediction of the current frame, and subsequent compression of the difference between the actual and predicted frames, is often called *hybrid coding*. Hybrid coding forms the core of video coding schemes from the early compression standards such as ITU-T H.261 [103] and ISO MPEG-1 to the most recent ISO MPEG-4 Part 2 [53], SMPTE VC-1 [82], China's Audio Video Standard (AVS) [29], ITU-T H.263 [104], and ITU-T H.264/ISO MPEG-4 Part 10 AVC coding standards [1, 84].

When a camera pans or zooms, this causes *global motion*, meaning that all or most of the pixels in the frame are apparently in motion in some related way, differing from the values they had in the previous frame. When the camera is stationary but objects in the scene move, this is called *local motion*. To compensate for local motion, a frame is typically subdivided into smaller rectangular blocks of pixels, in which motion is assumed to consist of uniform translation. The translational motion model assumes that motion within some image region can be represented with a vector of horizontal and vertical spatial displacements. In *block-based* motion-compensated prediction (MCP), for each block b in the current frame, a *motion vector* (MV) can be transmitted to the decoder to indicate which block in a previously coded frame is the best match for the given block in the current frame, and therefore forms the prediction of block b. Let us assume a block size of 8×8 pixels. The MV points from the center of the current block to the center of its best match block in the previously coded frame. MVs are essentially addresses of the best match blocks in the reference frame, in this case the previous frame. Let $\mathbf{v} = (v_x, v_y)$ denote the MV for a block in frame n. For the pixels in that block, the motion-compensated prediction from frame $n - 1$ is written as $\hat{f}_n(i,j) = f_{n-1}(i + v_x, j + v_y)$. If the MV is $\mathbf{v} = (0,0)$, then the best match block is the co-located block in the reference frame. As Figure 1.1 shows, parts of the tennis court at the bottom part of the frame appear static, so the best match is found with the $(0,0)$ MV. However, there is substantial motion in the rest of the frame that can only be modeled with *non-zero* MVs.

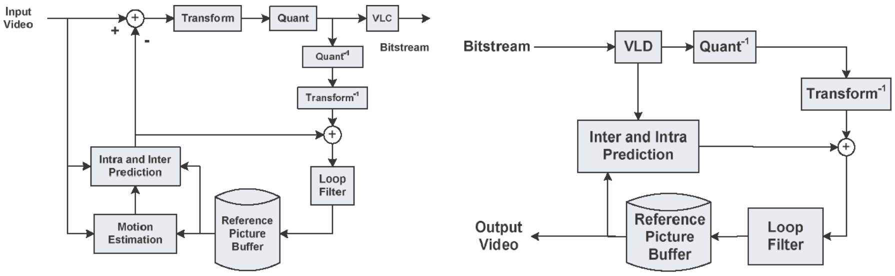
MVs or, in general, *motion parameters* are determined by doing a motion search, a process known as *motion estimation* (ME), in a reference frame. Assuming a search range of $[-16, +16]$ pixels for each spatial (horizontal and vertical) component, $33 \times 33 = 1089$ potential best match blocks can be referenced and have to be evaluated. The MV $\mathbf{v}$ that minimizes either the sum of absolute differences (SAD) or the sum of squared differences (SSD) between the block of pixels f in the current frame n and the block in the previous frame $n - l$ that is referenced by $\mathbf{v} = (v_x, v_y)$ may be selected and transmitted. Let b denote a set that contains the coordinates of all pixels in the block. The SAD and SSD are written as:

$$SAD = \sum_{(i,j) \in b} |f_n(i,j) - f_{n-l}(i + v_x, j + v_y)| \tag{1.1}$$

$$SSD = \sum_{(i,j) \in b} (f_n(i,j) - f_{n-l}(i + v_x, j + v_y))^2 \tag{1.2}$$

To form the MCP of the current frame, the blocks that are addressed through the MVs are copied from their original spatial location, possibly undergoing some type of *spatial filtering* (more on that in Section 3.5), to the location of the blocks in the current frame, as shown in Figure 1.2(c). This prediction frame is subsequently subtracted from the current frame to yield the *motion-compensated difference frame* or, in more general terms, the *prediction residual* in Figure 1.2(d). Obviously, if the MCP frame is very similar to the current frame, then the prediction residual will have most of its values close to zero, and hence require fewer bits to compress, compared to coding each frame with JPEG or subtracting the previous frame from the current one and coding the difference. One trade-off is an increase in complexity since ME is costly. The prediction residual is typically transmitted to the decoder by transforming it using a discrete cosine transform (DCT), rounding off the coefficients to some desired level of precision (a process called quantization) and sending unique variable-length codewords to represent these rounded-off coefficients. Along with this difference information, the MVs are transmitted to the decoder, requiring some additional bit rate of their own. For content with sufficient temporal correlation, the overall bit rate requirements are much less than without the use of MCP.

A diagram of a hybrid codec is illustrated in Figure 1.3. The decoder uses the MVs to obtain the motion compensated prediction blocks from some previously decoded reference frame. Then, the decoded prediction



(a) (b)

Fig. 1.3 Hybrid video (a) encoder and (b) decoder.

residual block is added to the MCP block to yield the current decoded block. This is repeated until the entire frame has been reconstructed. The reconstructed frame at the decoder may not be identical with the original one, because of the quantization used on the residual blocks.

Note that MCP for a block is also known as *inter prediction* since inter-frame redundancy is used to achieve compression. When combined with coding of the prediction residual it is called *inter-frame* coding. When a block is encoded independently of any other frame, this is known as *intra-frame* coding. Usually, intra-frame coding involves some kind of *intra-frame prediction* or *intra prediction*, which is predicting a block using spatial neighbors. This might involve using the DC coefficient of a transform block as a prediction of the DC coefficient of the next transform block in raster-scan order (as in JPEG). Or it might involve prediction of each pixel in a block from spatial neighbors using one of several possible directional extrapolations (as in H.264). In general, inter-frame coding enables higher compression ratios but is not as error resilient as intra-frame coding, since, for inter-frame coding, decoding the current frame depends on the availability of the reference frame. Video frames (or equivalently pictures) that use intra-frame coding exclusively to encode all blocks are called *intra-coded* or *I-coded frames*, while frames that allow the use of either intra-frame or inter-frame coding from some reference frame are known as *P-coded frames*. P-coded frames have been traditionally constrained to reference past frames in display order (as in the early standards H.261, H.263, MPEG-1, MPEG-2, and MPEG-4 Part 2). Finally, *B-coded frames* allow bi-directional prediction from one past and one future frame in display order in addition to intra-frame or inter-frame coding. Note that referencing future frames in display order generally involves transmitting frames out of order. For example, frame 3 can be encoded after frame 1 and then frame 2 can be encoded making reference to both frames 1 and 3. A simple illustration is shown in Figure 1.4. Note that B-coded frames were further extended in H.264/MPEG-4 AVC [1] to provide for a more generic form of bi-prediction without any restrictions in direction. Detailed information on bi-predictive coding is found in Section 4.1.

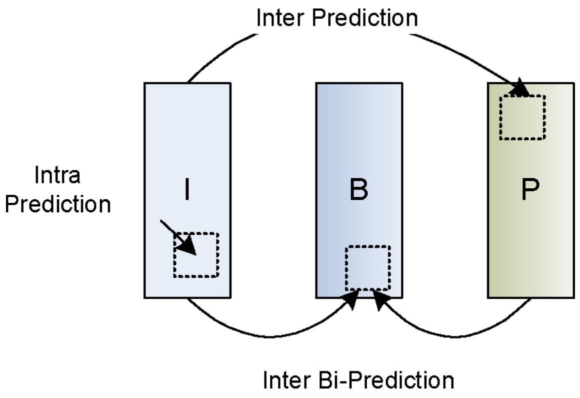


Fig. 1.4 An example of different prediction schemes.

1.2 Outline

Block-based MCP traditionally made use of a single previous frame as a reference frame for motion-compensated prediction, while for B-coded frames a single future frame was used jointly with the previous frame in display order to produce the best prediction for the current frame. However, motion search does not have to be limited to one frame from each prediction direction. Temporal correlation can be often nontrivial for temporally distant frames. In this article, the term multiple-reference frame motion compensation encompasses any method that uses combinations of more than one reference frame to predict the current frame. We also discuss cases where reference frames can be synthesized frames, such as panoramas and mosaics, or even composite frames that are assembled from parts of multiple previously coded frames. Finally, we note that we wish to decouple the term reference frame from that of a decoded frame. While a decoded frame can be a reference frame used for MCP of the current frame, a reference frame is not constrained to be identical to a decoded frame. The first treatise of the then state-of-the-art in multiple-reference frames for MCP is [109]. This work is intended to be somewhat broader and more tutorial. The article is organized as follows. Section 2 describes background, mosaic, and library coding, which preceded the development of modern multiple-reference techniques. Multiple-frame motion compensation is treated in

Section 3, while the almost concurrent development of multihypothesis prediction, often seen as a superset of multiple-reference prediction, is investigated in Section 4. The commercialization and rapid deployment of multiple-reference predictors has been hampered by the increased complexity requirements for motion estimation. Low complexity algorithms for multiple-frame motion search are covered in Section 5. The uses of multiple references for error resilience and error concealment are discussed in Sections 6 and 7, respectively. An experimental evaluation of some of the advances discussed in this work is presented in Section 8. This survey is concluded with Section 9. Appendix A provides the reader with additional information on rate-distortion optimization and Lagrangian minimization. Note that this work disregards the impact of each prediction scheme on decoder complexity.

References

[1] Advanced video coding for generic audiovisual services. ITU-T Recommendation H.264, 2005.

[2] M. E. Al-Mualla, C. N. Canagarajah, and D. R. Bull, "On the performance of temporal error concealment for long-term motion-compensated prediction," in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, pp. 376–379, September 2000.

[3] M. E. Al-Mualla, C. N. Canagarajah, and D. R. Bull, "Simplex minimization for single- and multiple-reference motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 12, pp. 1209–1220, December 2001.

[4] S. M. Amiri and I. V. Bajic, "A novel noncausal whole-frame concealment algorithm for video streaming," in *Proceedings of Tenth IEEE International Symposium on Multimedia*, pp. 154–159, February 2008.

[5] S. Belfiore, M. Grangetto, E. Magli, and G. Olmo, "Concealment of whole-frame losses for wireless low bit-rate video based on multiframe optical flow estimation," *IEEE Transactions on Multimedia*, vol. 7, no. 2, pp. 316–329, April 2005.

[6] G. Bjontegaard. Response to Call for Proposals for H.26L. ITU-T SG16/Q15, VCEG, Document Q15-F-11, November 1998.

[7] G. Bjontegaard. Enhancement of the Telenor proposal for H.26L. ITU-T SG16/Q15, VCEG, Document Q15-G-25, February 1999.

[8] G. Bjontegaard. Calculation of Average PSNR Differences between RD Curves. ITU-T SG16/Q6, VCEG-M33, April 2001.

[9] F. Bossen and A. Tourapis, “Method and apparatus for video encoding and decoding using adaptive interpolation,” U.S. Patent Application, US2006/0294171A1, December 2006.

[10] S. Brofferio and V. Corradi, “Videophone coding using background prediction,” in *Proceedings of European Signal Processing Conference*, vol. 2, pp. 813–816, 1986.

[11] M. Budagavi and J. D. Gibson, “Multiframe block motion compensated video coding for wireless channels,” in *Proceedings of Asilomar Conference on Signals, Systems and Computers*, pp. 953–957, 1996.

[12] M. Budagavi and J. D. Gibson, “Random lag selection in multiframe motion compensation,” in *Proceedings of IEEE International Symposium on Information Theory*, p. 410, August 1998.

[13] M. Budagavi and J. D. Gibson, “Multiframe video coding for improved performance over wireless channels,” *IEEE Transactions on Image Processing*, vol. 10, no. 2, pp. 252–265, February 2001.

[14] A. Chang, O. C. Au, and Y. M. Yeung, “A novel approach to fast multiframe selection for H.264 video coding,” in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, pp. 413–416, May 2003.

[15] V. Chellappa, P. C. Cosman, and G. M. Voelker, “Dual frame motion compensation with uneven quality assignment,” in *Proceedings of IEEE Data Compression Conference*, pp. 262–271, March 2004.

[16] V. Chellappa, P. C. Cosman, and G. M. Voelker, “Error concealment for dual frame video coding with uneven quality,” in *Proceedings of IEEE Data Compression Conference*, pp. 319–328, March 2005.

[17] M.-J. Chen, Y.-Y. Chiang, H.-J. Li, and M.-C. Chi, “Efficient multi-frame motion estimation algorithms for MPEG-4 AVC/JVT/H.264,” in *Proceedings of IEEE International Symposium on Circuits and Systems*, May 2004.

[18] Y. Chen, K. Yu, J. Li, and S. Li, “An error concealment algorithm for entire frame loss in video transmission,” in *Proceedings of Picture Coding Symposium*, December 2004.

[19] G. Cheung, “Near-optimal multipath streaming of H.264 using reference frame selection,” in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, pp. 653–656, September 2003.

[20] G. Cheung and W. Tan, “Loss-compensated reference frame optimization for multi-path video streaming,” in *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 844–847, July 2005.

[21] H. Chung and A. Ortega, “Low complexity motion estimation algorithm by multiresolution search for long-term memory motion compensation,” in *Proceedings of IEEE International Conference on Image Processing*, vol. 2, pp. 261–264, September 2002.

[22] H. Chung, A. Ortega, and A. Sawchuk, “Low complexity motion estimation for long term memory motion compensation,” in *Proceedings of SPIE Visual Communication and Image Processing*, January 2002.

[23] H. Chung, D. Romacho, and A. Ortega, “Fast long-term motion estimation for H.264 using multiresolution search,” in *Proceedings of IEEE International Conference on Image Processing*, vol. 1, pp. 905–908, September 2003.

[24] Codecs for videoconferencing using primary digital group transmission. ITU-T Recommendation H.120, 1993.

[25] S. Cui, Y. Wang, and J. E. Fowler, "Multihypothesis motion compensation in the redundant wavelet domain," in *Proceedings of IEEE International Conference on Image Processing*, vol. 2, pp. 53–56, September 2003.

[26] C. Duanmu, M. O. Ahmad, and M. N. S. Swamy, "A continuous tracking algorithm for long-term memory motion estimation," in *Proceedings of IEEE International Symposium on Circuits and Systems*, vol. 2, pp. 356–359, May 2003.

[27] F. Dufaux and F. Moscheni, "Background mosaicking for low bit rate video coding," in *Proceedings of IEEE International Conference on Image Processing*, vol. 1, pp. 673–676, September 1996.

[28] J. Fan, Z. Zhang, and Y. Chen, "A new error concealment scheme for whole frame loss in video transmission," in *Proceedings of Picture Coding Symposium*, 2007.

[29] L. Fan, S. Ma, and F. Wu, "Overview of AVS video standard," in *Proceedings of IEEE International Conference on Multimedia and Expo*, vol. 1, pp. 423–426, 2004.

[30] M. Flierl and B. Girod, "Multihypothesis motion-compensated prediction with forward-adaptive hypothesis switching," in *Proceedings of International Picture Coding Symposium PCS*, pp. 195–198, April 2001.

[31] M. Flierl and B. Girod, "Multihypothesis motion estimation for video coding," in *Proceedings of IEEE Data Compression Conference*, pp. 341–350, March 2001.

[32] M. Flierl and B. Girod, "Generalized B pictures and the draft H.264/AVC video-compression standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 587–597, July 2003.

[33] M. Flierl, T. Wiegand, and B. Girod, "A locally optimal design algorithm for block-based multi-hypothesis motion-compensated prediction," in *Proceedings of IEEE Data Compression Conference*, pp. 239–248, April 1998.

[34] M. Flierl, T. Wiegand, and B. Girod, "Rate-constrained multi-hypothesis motion-compensated prediction for video coding," in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, pp. 150–153, September 2000.

[35] T. Fukuhara, K. Asai, and T. Murakami, "Very low bit-rate video coding with block partitioning and adaptive selection of two time-differential frame memories," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 1, pp. 212–220, February 1997.

[36] S. Fukunaga, T. Nakai, and H. Inoue, "Error-resilient video coding by dynamic replacing of reference pictures," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 3, pp. 1503–1508, November 1996.

[37] B. Girod, "Motion-compensating prediction with fractional-pel accuracy," *IEEE Transactions on Communications*, vol. 41, no. 4, pp. 604–612, April 1993.

[38] B. Girod, "Rate-constrained motion estimation," in *Proceedings of SPIE Visual Communication and Image Processing*, vol. 2308, pp. 1026–1034, 1994.

[39] B. Girod, “Efficiency analysis of multi-hypothesis motion-compensated prediction for video coding,” *IEEE Transactions on Image Processing*, vol. 9, no. 2, pp. 173–183, February 2000.

[40] M. Gothe and J. Vaisey, “Improving motion compensation using multiple temporal frames,” in *Proceedings of IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, vol. 1, pp. 157–160, May 1993.

[41] D. Hepper, “Efficiency analysis and application of uncovered background prediction in a low bit rate image coder,” *IEEE Transactions on Communications*, vol. 38, no. 9, pp. 1578–1584, September 1990.

[42] D. Hepper and H. Li, “Analysis of uncovered background prediction for image sequence coding,” in *Proceedings of Picture Coding Symposium*, pp. 192–193, 1987.

[43] J. R. Hidalgo and P. Salembier, “Long term selection of reference frame sub-blocks using MPEG-7 indexing metadata,” in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, pp. 669–672, April 2007.

[44] M. Horowitz, Demonstration of H.263++ Annex U Performance. ITU-T SG16/Q15, VCEG, Document Q15-J-11, May 2000.

[45] S. Hsu and P. Anandan, “Hierarchical representations for mosaic-based video compression,” in *Proceedings of Picture Coding Symposium*, pp. 395–400, March 1996.

[46] Y. Huang, Z. Liu, S. Goto, and T. Ikenaga, “Adaptive edge detection pre-process multiple reference frames motion estimation in H.264/AVC,” in *Proceedings of International Conference on Communications, Circuits and Systems*, pp. 787–791, July 2007.

[47] Y.-W. Huang, B.-Y. Hsieh, T.-C. Wang, S.-Y. Chien, S.-Y. Ma, C.-F. Shen, and L.-G. Chen, “Analysis and reduction of reference frames for motion estimation in MPEG-4 AVC/JVT/H.264,” in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 3, pp. 145–148, April 2003.

[48] Information technology — Generic coding of moving pictures and associated audio information: Video. ITU-T Recommendation H.262, 2000.

[49] M. Irani, S. Hsu, and P. Anandan, “Mosaic-based video compression,” in *Proceedings of SPIE Digital Video Compression: Algorithms and Technologies*, vol. 2419, February 1995.

[50] ISO/IEC 10918-1:1992 Information technology — Digital Compression and Coding of Continuous-Tone Still Images — Requirements and Guidelines. ISO/IEC JTC1/SC29/WG10, 1992.

[51] ISO/IEC 11172-1:1993 Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbits/s. ISO/IEC JTC1/SC29/WG11, 1993.

[52] ISO/IEC 13818-2:2000 Information technology — Generic coding of moving pictures and associated audio information: Video. ISO/IEC JTC1/SC29/WG11, 1994.

[53] ISO/IEC 14496-2:1999 Information technology — Coding of audio-visual objects — Part 2: Visual. ISO/IEC JTC1/SC29/WG11, 1999.

[54] B. Jung, B. Jeon, M.-D. Kim, B. Suh, and S.-I. Choi, "Selective Temporal Error Concealment Algorithm for H.264/AVC," in *Proceedings of IEEE International Conference on Image Processing*, October 2004.

[55] C. Kim and C.-C. J. Kuo, "Efficient temporal search range prediction for motion estimation in H.264," in *Proceedings of IEEE 7th Workshop on Multimedia Signal Processing*, pp. 1–4, October 2005.

[56] C.-S. Kim, R.-C. Kim, and S.-U. Lee, "Robust transmission of video sequence using double-vector motion compensation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 9, pp. 1011–1021, September 2001.

[57] C.-S. Kim, S. Ma, and C.-C. J. Kuo, "Fast H.264 motion estimation with block-size adaptive referencing (BAR)," in *Proceedings of IEEE International Conference on Image Processing*, pp. 57–60, October 2006.

[58] M.-D. Kim, S.-I. Choi, and S.-W. Ra, "Hybrid error concealment method for H.264/AVC," in *Proceedings of the 7th International Conference on Advanced Communication Technology*, pp. 408–411, 2005.

[59] S. H. Kim, Y. K. Kim, K. H. Yang, and S. U. Lee, "Multiple reference frame based scalable video coding for low-delay Internet transmission," in *Proceedings of International Workshop on Very Low Bitrate Video Coding*, 2001.

[60] W.-Y. Kung, C.-S. Kim, and C.-C. J. Kuo, "Analysis of multi-hypothesis motion compensated prediction for robust video transmission," in *Proceedings of IEEE International Symposium on Circuits and Systems*, May 2004.

[61] R. Kutka, "Content-adaptive long-term prediction with reduced memory," in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, pp. 817–820, September 2003.

[62] W. M. Lam, A. R. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 417–420, 1993.

[63] P.-J. Lee and M.-L. Lin, "Fuzzy logic based temporal error concealment for H.264 video," *ETRI Journal*, vol. 28, no. 5, pp. 574–582, October 2006.

[64] Y.-C. Lee, Y. Altunbasak, and R. M. Mersereau, "A temporal error concealment method for MPEG coded video using a multi-frame boundary matching algorithm," in *Proceedings of IEEE International Conference on Image Processing*, vol. 1, pp. 990–993, October 2001.

[65] A. Leontaris and P. C. Cosman, "Video compression for lossy packet networks with mode switching and a dual-frame buffer," *IEEE Transactions on Image Processing*, vol. 13, no. 7, pp. 885–897, July 2004.

[66] A. Leontaris and A. M. Tourapis, "Weighted prediction alternative to adaptive interpolation mechanisms," Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, JVT-AB033, July 2008.

[67] X. Li, E. Q. Li, and Y.-K. Chen, "Fast multi-frame motion estimation algorithm with adaptive search strategies in H.264," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2004.

[68] Y. Liang, M. Flierl, and B. Girod, "Low-latency video transmission over lossy packet networks using rate-distortion optimized reference picture selection," in *Proceedings of IEEE International Conference on Image Processing*, vol. 2, pp. 181–184, September 2002.

[69] Y. J. Liang, E. Setton, and B. Girod, “Channel-adaptive video streaming using packet path diversity and rate-distortion optimized reference picture selection,” in *Proceedings of IEEE Workshop on Multimedia Signal Processing*, pp. 420–423, December 2002.

[70] Z. Liang, J. Zhou, O. C. Au, and L. Guo, “Content-adaptive temporal search range control based on frame buffer utilization,” in *Proceedings of IEEE 8th Workshop on Multimedia Signal Processing*, pp. 399–402, October 2006.

[71] S. Lin, S. Mao, Y. Wang, and S. Panwar, “A reference picture selection scheme for video transmission over ad-hoc networks using multiple paths,” in *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 96–99, August 2001.

[72] S. Lin and Y. Wang, “Error resilience property of multihypothesis motion-compensated prediction,” in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, pp. 545–548, September 2002.

[73] Y. Liu, J. Bu, C. Chen, L. Mo, and K. He, “Multiframe error concealment for whole-frame loss in H.264/AVC,” in *Proceedings of IEEE International Conference on Image Processing*, vol. 4, pp. 281–284, September 2007.

[74] N. Mukawa and H. Kuroda, “Uncovered background prediction in interframe coding,” *IEEE Transactions on Communications*, vol. 33, no. 11, pp. 1227–1231, November 1985.

[75] H. G. Musmann, P. Pirsch, and H.-J. Grallert, “Advances in picture coding,” *Proceedings of IEEE*, vol. 73, no. 4, pp. 523–548, April 1985.

[76] S. Nogaki and M. Ohta, “An overlapped block motion compensation for high quality motion picture coding,” in *Proceedings of IEEE International Symposium on Circuits and Systems*, pp. 184–187, May 1992.

[77] M. T. Orchard and G. J. Sullivan, “Overlapped block motion compensation: An estimation-theoretic approach,” *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 693–699, September 1994.

[78] I. Park and D. W. Capson, “Dynamic reference frame selection for improved motion estimation time in H.264/AVC,” in *Proceedings of IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 97–100, March 2008.

[79] Y. O. Park, C.-S. Kim, and S.-U. Lee, “Multi-hypothesis error concealment algorithm for H.26L video,” in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, pp. 465–468, September 2003.

[80] A. Puri, R. Aravind, B. G. Haskell, and R. Leonardi, “Video coding with motion-compensated interpolation for CD-ROM applications,” *Signal Processing: Image Communication*, vol. 2, no. 2, pp. 127–144, 1990.

[81] F. H. P. Spaan, R. L. Lagendijk, and J. Biemond, “Error robust video coding based on H.263,” in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, pp. 472–476, October 1998.

[82] Standard for Television: VC-1 Compressed Video Bitstream Format and Decoding Process. SMPTE 421M, 2006.

[83] Y. Su and M.-T. Sun, “Fast multiple reference frame motion estimation for H.264,” in *Proceedings of IEEE International Conference on Multimedia and Expo*, vol. 1, pp. 695–698, October 2004.

[84] G. Sullivan, T. Wiegand, D. Marpe, and A. Luthra, Text of ISO/IEC 14496-10 Advanced Video Coding 3rd Edition. ISO/IEC JTC 1/SC 29/WG11 N6540, July 2004.

[85] G. J. Sullivan, "Multi-hypothesis motion compensation for low bit-rate video coding," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 437–440, April 1993.

[86] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, vol. 15, pp. 74–90, November 1998.

[87] C. N. Taylor and S. Dey, "ORBit: An adaptive data shaping technique for robust wireless video clip communication," in *Proceedings of 37th Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 1567–1571, November 2003.

[88] C.-W. Ting, W.-H. Lam, and L.-M. Po, "Fast block-matching motion estimation by recent-biased search for multiple reference frames," in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, pp. 1445–1448, October 2004.

[89] C.-W. Ting, L.-M. Po, and C.-H. Cheung, "Center-biased frame selection algorithms for fast multi-frame motion estimation in H.264," in *Proceedings of IEEE International Conference on Neural Networks and Signal Processing*, pp. 1262–1265, December 2003.

[90] M. Tiwari and P. Cosman, "Dual frame video coding with pulsed quality and a lookahead window," in *Proceedings of IEEE Data Compression Conference*, pp. 372–381, March 2006.

[91] M. Tiwari and P. C. Cosman, "Selection of long-term reference frames in dual-frame video coding using simulated annealing," *IEEE Signal Processing Letters*, vol. 15, pp. 249–252, 2008.

[92] Y. Tomita, T. Kimura, and T. Ichikawa, "Error resilient modified inter-frame coding system for limited reference picture memories," in *Proceedings of Picture Coding Symposium*, pp. 743–748, September 1997.

[93] A. M. Tourapis, "Enhanced predictive zonal search for single and multiple frame motion estimation," in *Proceedings of SPIE Visual Communication and Image Processing*, January 2002.

[94] A. M. Tourapis, O. C. Au, and M. L. Liou, "Highly efficient predictive zonal algorithms for fast block-matching motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 10, pp. 934–947, October 2002.

[95] A. M. Tourapis, H. Y. Cheong, O. C. Au, and M. L. Liou, "N-Dimensional zonal algorithms. The future of block based motion estimation?," in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, pp. 506–509, October 2001.

[96] A. M. Tourapis, F. Wu, and S. Li, "Enabling VCR functionalities in streaming media (intelligent interactive streaming I2 stream)," in *Proceedings of IEEE International Conference on Information Technology: Research and Education*, pp. 1–5, August 2003.

[97] H.-Y. C. Tourapis and A. M. Tourapis, "Fast motion estimation within the H.264 codec," in *Proceedings of IEEE International Conference on Multimedia and Expo*, vol. 3, pp. 517–520, July 2003.

[98] Y.-C. Tsai and C.-W. Lin, "H.264 Error resilience coding based on multihypothesis motion compensated prediction," in *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 952–955, July 2005.

[99] W. Tu and E. Steinbach, "Proxy-based reference picture selection for real-time video transmission over mobile networks," in *Proceedings of IEEE International Conference on Multimedia and Expo*, July 2005.

[100] S. C. Vadapalli, H. Shetiya, and S. Sethuraman, "Efficient alternative to intra refresh using reliable reference frames," in *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 124–127, July 2007.

[101] N. Vasconcelos and A. Lippman, "Library-based image coding," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. v, pp. 489–492, April 1994.

[102] A. Vetro, P. Pandit, A. Smolic, and Y.-K. Wang, Joint draft 7.0 on multiview video coding. Joint Video Team of ISO/IEC MPEG and ITU-T VCEG, JVT-AA209, April 2008.

[103] Video codec for audiovisual services at p x 64 kbit/s. ITU-T Recommendation H.261, 1993.

[104] Video coding for low bit rate communication. ITU-T Recommendation H.263, 2005.

[105] Y.-K. Wang, M. M. Hannuksela, and M. Gabbouj, "Error resilient video coding using unequally protected key pictures," in *Proceedings of International Workshop on Very Low Bitrate Video Coding*, pp. 290–297, September 2003.

[106] H. Watanabe and S. Singhal, "Windowed motion compensation," in *Proceedings of SPIE Visual Communication and Image Processing VCIP*, vol. 1605, pp. 582–589, November 1991.

[107] T. Wiegand, Proposed draft on Annex U for enhanced reference picture selection. ITU-T SG16/Q15, VCEG, Document Q15-H-30, August 1999.

[108] T. Wiegand, N. Färber, K. Stuhlmüller, and B. Girod, "Error-resilient video transmission using long-term memory motion compensated prediction," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 1050–1062, June 2000.

[109] T. Wiegand and B. Girod, *Multi-frame Motion-compensated Prediction for Video Transmission.* Norwell, MA: Kluwer Academic Publishers, September 2001.

[110] T. Wiegand, B. Lincoln, and B. Girod, "Fast search for long-term memory motion-compensated prediction," in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, pp. 619–622, October 1998.

[111] T. Wiegand, E. Steinbach, and B. Girod, "Affine multipicture motion-compensated prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 2, pp. 197–209, February 2005.

[112] T. Wiegand, E. Steinbach, A. Stensrud, and B. Girod, "Multiple reference picture coding using polynomial motion models," in *Proceedings of SPIE Visual Communication and Image Processing VCIP*, vol. 3309, no. 2, pp. 134–145, January 1998.

[113] 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, February 1999.

[114] T. Wiegand, X. Zhang, B. Girod, and B. Andrews, Long-term memory motion-compensated prediction. ITU-T SG16/Q15, VCEG, Document Q15-C-11, December 1997.

[115] S. W. Wu and A. Gersho, "Joint estimation of forward and backward motion vectors for interpolative prediction of video," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 684–687, September 1994.

[116] X. Yuan, "Hierarchical uncovered background prediction in a low bit rate video coder," in *Proceedings of Picture Coding Symposium*, p. 12.1, 1993.

[117] K. Zhang and J. Kittler, "A background memory update scheme for H.263 video codec," in *Proceedings of European Signal Processing Conference*, vol. 4, pp. 2101–2104, September 1998.

[118] K. Zhang and J. Kittler, "Using background memory for efficient video coding," in *Proceedings of IEEE International Conference on Image Processing*, vol. 3, pp. 944–947, October 1998.

[119] K. Zhang and J. Kittler, Proposed amendments for Annex U on enhanced reference picture selection. ITU-T SG16/Q15, VCEG, Document Q15-H-09, August 1999.

[120] R. Zhang, S. L. Regunathan, and K. Rose, "Video coding with optimal inter/intra-mode switching for packet loss resilience," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 6, pp. 966–976, June 2000.

[121] J. Zheng and L.-P. Chau, "Error-resilient coding of H.264 based on periodic macroblock," *IEEE Transactions on Broadcasting*, vol. 52, no. 2, pp. 223–229, June 2006.

[122] X. Zhou and C.-C. J. Kuo, "Robust streaming of offline coded H.264/AVC video via alternative macroblock coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 4, pp. 425–438, April 2008.

[123] X.-J. Zhu and S.-A. Zhu, "Fast mode decision and reduction of reference frames for H.264 encoder," in *Proceedings of International Conference on Control and Automation*, vol. 2, pp. 1040–1043, June 2005.