DOE/ER/25277-1

Visual Communications for Heterogeneous Networks/Visually Optimized Scalable Image Compression

Final Report
September 1, 1995 — February 28, 2002, no-cost extension to February 28, 2003

Principal Investigator: Sheila S. Hemami

School of Electrical Engineering
Cornell University
224 Phillips Hall
Ithaca, NY 14853

hemami@ece.cornell.edu
Tel: (607) 254-5128
Fax: (607) 255-9072

DOE Grant Number DE-FG02-95ER25277

DISCLAIMER

1. Overview & Summary

The goal of this research is to enable reliable "universal access" to visual communications over the National Information Infrastructure (NII). All users, regardless of their individual network connection bandwidths, qualities-of-service, or terminal capabilities, should have the ability to access still images, video clips, and multimedia information services, and to use interactive visual communications services. To do so requires special capabilities for image and video compression algorithms: *scalability, reconstructibility,* and *network adaptivity.* Scalability allows an information service to provide visual information at many rates, without requiring additional compression or storage after the stream has been compressed the first time. Reconstructibility allows reliable visual communications over an imperfect network. Network adaptivity permits real-time modification of compression parameters to adjust to changing network conditions. Furthermore, to optimize the efficiency of the compression algorithms, they should be *visually optimal,* where each bit expended reduces the visual distortion. Visual optimality is achieved through first extensive experimentation to quantify human sensitivity to supra-threshold compression artifacts and then incorporation of these experimental results into quantization strategies and compression algorithms.

1.1. Research Goals

We have developed image and video compression algorithms that provide scalability, reconstructability, and network adaptivity, and we have developed compression and quantization strategies that are visually optimal at all bit rates.

A *scalable* compression algorithm allows an information service to provide visual information at many rates, without requiring additional compression or storage after the stream has been compressed the first time. Such an algorithm generates a single stream, from which many streams at different rates may be extracted. Extraction requires no additional processing. In the case of images, these extracted streams may vary in image size, providing *spatial scalability.* In the case of video, these extracted streams can vary both in frame size and in frame rate (*temporal scalability*). Instead of requiring multiple coded streams, the information provider can produce a stream for any requested rate by simply extracting it from the single stream. Information services therefore become "universal" — their user base is expanded, because they can service users of any bandwidth capability. Conversely, users gain affordable access to all information services, regardless of the users' bandwidth capabilities.

As network usage increases, the probability of packet delay and loss increases as well. Reconstruction provisions in a compression algorithm allow reliable visual communications over an imperfect network by allowing for effective "repair" of damaged visual information at the receiver.

Interactive services requires real-time encoding and transmission. Therefore, if the algorithm is appropriately designed, the video compression parameters can be modified to adapt to the collective bandwidths and packet handling characteristics of the network segments. Adaptivity is performed to optimize scalability and reconstruction performance for the current network conditions, providing the best visual quality communications.

Previous experiments to quantify HVS sensitivity to artifacts in wavelet-compressed images have been in the *sub-threshold* regime, in which distortion is non-perceivable and hence below

Contents

visibility thresholds. A limitation of these experiments is that they omit a phenomenon called *spatial masking*, in which the presence of one image component can affect the visibility threshold of another component. A second limitation is that these results typically provide only a single compression ratio, and the flexibility to provide a range of compression ratios is absent. The resulting ratio may or may not be appropriate for a specific application.

When compression artifacts become visible, the compression is no longer in the sub-threshold regime and is now *supra-threshold.* In this regime, we can only provide visually optimal compression, in which distortion visibility is minimized for a given bit-rate. Our goal has been to provide an understanding of supra-threshold distortion visibilities which can be used to provide a range of visually-optimized compression ratios.

1.2. Major Accomplishments

Three students (two of them women) have completed Ph.D. degrees with support from this grant, and a fourth will finish in May 2004; one student completed an M.S. degree (students are listed in Section 5).

Forty publications (11 journal papers and 29 conference papers) have resulted from this grant (publications are listed in Section 6). One of the conference papers was a prize paper at the *IEEE International Conference on Image Processing* in 2002.

The major technical results are summarized below.

Scalability & Adaptivity:

- The perceptually-scalable image compression (PSIC) algorithm provides robust transmission over packet-based networks, spatial scalability, and uses activity selection to maximize visual quality at all bit rates and resolutions.
- Smoothness-constrained wavelet-based image compression techniques have been extended to be adaptive to local regularity of images, providing both improved rate-distortion characteristics over and better visual performance than standard scalar quantizers.
- A study of temporal scalable techniques provided both theoretical and empirical proof that motion-compensated coding followed by frame discarding is the most efficient technique for providing temporal scalability.
- Based on the temporal scalability results, improved rate-control strategies for motion-compensated video compression algorithms have been developed; these strategies optimize parameter selection for variable-rate networks and hence permit adaptivity.

Reconstructability:

- Reconstruction-optimized source coding algorithms have been developed, which provide graceful degradation in the event of severe packet loss.
- Resynchronizing variable-length codes have demonstrated that robust lossless coding is possible at near the entropy rate.

Visual optimality:

- The human visual system responses to both simple and compound stimuli resulting from quantization of wavelet coefficients has been quantified, both without and with a natural image masker.
- A contrast-based quantization strategy has been proposed which produces visually superior images at low rate and visually equivalent images at higher rates (in comparison with current state-of-the-art compression techniques such as JPEG-2000). This strategy can be used in custom or standards-based compression algorithms.

2. Scalable and Adaptive Image and Video Compression

2.1. Perceptually-Scalable Image Compression

We have developed a multiresolution-based image coding technique that achieves high visual quality through perceptual-based scalability and robustness to transmission errors. The perceptually-based scalability is achieved by segmenting the image into perceptually significant activity regions (smooth, edge, and highly detailed regions) that are coded and packetized to facilitate the extraction of high quality images at different spatial resolutions and bit rates. The coding scheme takes advantage of the scalability provided by subband decompositions by coding the perceptually significant regions in each band independently. An extensive psychovisual study was performed to analyze the human visual system (HVS) sensitivity to the activity regions and to high frequency subbands at different levels of decomposition. The results of the study were used to design a perceptually-based bit allocation scheme. The coding scheme proposed is robust for transmission over packet-based networks as demonstrated by an analysis of packet losses over the Internet.

2.2. Smoothness-constrained Wavelet-based Image Compression

A prominent theorem in wavelet analysis relates rate of decay of the wavelet transform coefficients across scales to the mathematical Holder regularity of signals, which correlates well with visual smoothness. The Holder regularity essentially indicates the number of continuous derivatives of a function. If the energies of the wavelet coefficients in a given region decrease rapidly as the resolution varies from coarse to fine, the region is smooth. Slow decay indicates edge or texture activity.

We have demonstrated that the uniform scalar quantization used in many image compression algorithms can affect this regularity, creating spurious artifacts and high-frequency detail not present in the original image. When the magnitude of wavelet coefficients is rounded up during quantization, the smoothness of the signal is decreased. We have proposed a modified quantization scheme which guarantees that the reconstructed image is at least as smooth as the original image, thus eliminating false detail. This quantizer works by performing a soft-thresholding operation on the wavelet coefficients prior to quantization. In other words, the magnitudes of the coefficients are decreased uniformly by a certain threshold. Those wavelet coefficients with small magnitude are quantized to zero. Since the high-frequency wavelet coefficients tend to be quite small, the soft thresholding essentially removes high-frequency noise without significantly affecting the important high-frequency details such as strong edges.

2.3. A Comparison of Temporal Scalability Techniques

A temporally scalable video coding algorithm allows extraction of video of multiple frame rates from a single coded stream. In recent years, several video coding techniques have been proposed that provide temporal scalability using subband coding, both without and with motion compensation. With a two-band subband decomposition applied hierarchically, frame rates halve after each filtering operation. Alternatively, motion-compensated prediction (as used in MPEG) can provide temporal scalability and the same frame rates as temporal subband coding through strategic placement of reference frames and selective decoding of frames. We have compared three temporal coding techniques with respect to providing temporal scalability: temporal subband coding (TSB), motion-compensated temporal subband coding (MC-TSB), and motion-compensated prediction (MCP). Predicted rate-distortion performances at full- and lower frame rates and experimental quantitative and visual performances from coding several video sequences have been compared. The comparison is explicitly for temporal coding when the dimensionality of the subsequent source coding is held constant; any spatial or higher dimensional source coding can follow. In theory and in practice, MCP and MC-TSB always outperform TSB. For high-bit-rate full-frame-rate video, the performances of MCP and MC-TSB are approximately equivalent. However, to provide temporal scalability, MCP clearly provides the best performance in terms of visual quality, quantitative quality, and bit rate of the lower frame-rate video.

2.4. Rate-control for Motion-compensated Video Coders

Given that MCP maximizes the performance of temporal scalability, we have developed both within-frame and between-frame rate control strategies for MCP video coding. To provide within-frame rate-control, we have addressed jointly rate-distortion optimal selection of coding parameters in a general motion-compensated video coder. The general coder uses variable-block-size motion estimation and multimode residual coding. This is essentially the optimal bit-allocation problem for an individual frame at a given rate constraint. We have derived the general formulation and solution using the Lagrange multiplier method and dynamic programming, and have also demonstrated how the general theory can he adapted and applied to both an MPEG-like coder and a motion-compensated wavelet coder. Simulations demonstrate that both proposed coders outperform MPEG (TM5) by 0.7-1.3 dB at a variety of bit rates, with the gain provided by both better motion estimation and the joint-parameter optimization. The technique is applicable to MPEG-compliant coders with fixed block-size motion estimation and provides a gain of 0.5-0.7 dB over TM5. The optimization approach can also be applied to distortion-constrained coding, and therefore allows a fine tuning of either the rate or distortion to follow any desired profile.

Between-frame rate control is essential to ensure high-quality video transmission over asynchronous transfer mode (ATM) networks; both source and channel rates must not violate buffer or network traffic constraints in order to guarantee decoded video quality. To date, source and channel rate selection has been performed jointly because the rates are related through buffer and network constraints and therefore appear to be interdependent. We have shown that selection of these rates can be separated and the rate control problem is therefore simplified, allowing sequential source and channel rate selection without the need for iteration. A rate control algorithm has been proposed which implements noniterative, separate sequential selection of source and channel rates in order to minimize the distortion variation between frames, subject to all buffer and network constraints. In particular, the leaky bucket is used as the traffic policing mechanism. Achieving a target rate for individual frames is facilitated by using a rate-distortion optimized video coder which allows coding to an exactly specified rate or distortion. Channel rate selection is performed following source rate selection with the goal of providing smooth traffic to the network, thereby allowing a high statistical multiplexing gain. Experiments demonstrate that the proposed noniterative rate control algorithm, when implemented using the optimized coder, achieves consistent video quality while providing smooth traffic to the network.

3. Reconstructability

3.1. Reconstruction-optimized Lapped Orthogonal Transforms for Robust Image Transmission

Many reconstruction algorithms have been developed for fixed coding techniques. This work formulates and solves the dual problem-a block-based coding technique, namely a family of lapped orthogonal transforms (LOTs) is designed to maximize the reconstruction performance of a specified reconstruction algorithm. Mean-reconstruction, in which a missing coefficient block is replaced with the average of its available neighbors, is selected for its simplicity and ease of implementation. A reconstruction criterion is defined as the equal distribution of reconstruction errors across all transform coefficients, and a family of LOTs is then designed to meet the reconstruction criterion as well as consider the transform coding gain. Reconstruction capability and coding gain are traded off, and the LOT family consists of transforms that provide increasing reconstruction capability with lower coding gain. The reconstruction-optimized LOT family provides excellent reconstruction capability, and a transform can be selected based on the loss characteristics of the channel, the desired reconstruction performance, and the desired compression.

3.2. Resynchronizing Variable-length Codes and Incorporation into Compression Algorithms with Error Concealment

Resynchronizing variable-length codes (RVLCs) for large alphabets are designed by first creating resynchronizing Huffman codes and then adding an extended synchronizing codeword, and the RVLCs are applied to both JPEG and wavelet-based image compression. The RVLCs demonstrate the desired resynchronization properties, both at a symbol level and structurally so that decoded data can be correctly placed within an image following errors. The encoded images, when subject to both structural and statistical error detection and concealment, can tolerate BERs of up to 10^{-4} and are very tolerant of burst errors. The RVLC-JPEG images have negligible overhead at visually lossless bit rates, while the RVLC-wavelet overhead can be adjusted based on the desired tolerance to burst errors and typically ranges from 7 to 18%. The tolerance to both bit and burst errors demonstrates that images coded with such RVLCs can be transmitted over imperfect channels suffering bit errors or packet losses without channel coding for the image data, or with less channel coding than would be required if the encoded image data could tolerate no bit errors. While the overhead is nontrivial for the RVLC-wavelet images and the lower-rate RVLC-JPEG images, the encoded bitstreams do not have the firm restrictions on numbers or spacings of bit errors that some error correcting codes have, and hence provide more graceful degradation.

4. Achieving Visual Optimality

4.1. Characterization of Human Visual System Responses to Simple & Compound Suprathreshold Distortions

A psychophysical experiment to quantify human sensitivities to suprathreshold distortions caused by wavelet coefficient quantization in natural images was designed and performed. Quantization of the coefficients within a discrete wavelet transform subband gives rise to distortions in the reconstructed image which are localized in spatial frequency and orientation, and which are spatially correlated with the image. We have investigated the detectability of these distortions: contrast thresholds were measured for both simple and compound distortions presented in the unmasked paradigm and against two natural-image maskers. Simple and compound distortions were generated via uniform scalar quantization of one or two subband(s). Unmasked detection thresholds for simple distortions yielded contrast sensitivity functions similar to those reported for 1-octave Gabor patches. Detection thresholds for simple distortions presented against two natural-image backgrounds revealed that thresholds were elevated across the frequency range of 1.15-18.4 cycles/deg, with the greatest elevation for low-frequency distortions. Unmasked thresholds for compound distortions revealed relative sensitivities of 1.1-1.2, suggesting that summation of responses to wavelet distortions is similar to summation to gratings. Masked thresholds for compound distortions revealed relative sensitivities of 1.5-1.7, suggesting greater summation when distortions are masked by natural images.

4.2. Contrast-based Quantization Producing Visually Optimal Images

Modern lossy image compression algorithms exploit characteristics of the human visual system; however, whereas psychophysical results are commonly reported in terms of contrast, the majority of compression strategies incorporate contrast sensitivity only in an implicit fashion (e.g., by weighting quantizer step sizes). We developed a contrast-based quantization strategy for use in wavelet-based lossy image compression. Based on the results of our psychophysical experiments using wavelet subband quantization distortions and natural-image backgrounds, subbands are quantized such that the distortions in the reconstructed image exhibit specific root-mean squared (RMS) contrasts, and such that edge-structure is preserved across scale-space. Within a single, unified framework, the proposed contrast-based strategy yields images which are competitive in visual quality with results from current visually lossless approaches at high bit-rates, and which demonstrate improved visual quality over current visually lossy approaches at low bit-rates. This strategy operates in the context of non-embedded and embedded quantization, the latter of which yields a highly scalable codestream which provides the best-possible visual quality at all bit-rates; a specific application of the proposed algorithm to JPEG-2000 has been developed.

5. Graduates Resulting from this Grant

Greg Conklin, M.S., December 1998, "Theory and practice of scalable digital video coding"

Dr. W. Knox Carey, August 1999, "Applications of Wavelet Coefficient Decay"

Dr. (Ms.) Yan Yang, January 2000, "Rate Control for Video Coding and Transmission"

Dr. (Ms.) Marcia G. Ramos, May 2000, "Perceptually-Based Image Coding"

Damon M. Chandler, Ph.D. expected May 2004, "Contrast-based image compression and processing"

6. Publications Resulting from this Grant

Journal papers

1. S. S. Hemami, "Reconstruction-Optimized Lapped Orthogonal Transforms for Robust Image Transmission," *IEEE Trans. Circuits & Systems for Video Technology*, April 1996.
2. M. G. Ramos, S. S. Hemami, "Perceptually-Based Scalable Image Compression for Packet Networks," *Journal of Electronic Imaging, special issue on Image/Video Compression and Processing for Visual Communications,* Vol. 7, No. 3, pp. 453-64, July 1998.
3. G. C. Conklin, S. S. Hemami, "A Comparison of Temporal Scalability Techniques," *IEEE Transactions on Circuits and Systems for Video Technology,* Vol. 9, No. 6, pp. 909-19, September 1999.
4. W. K. Carey, D. B. Chuang, S. S. Hemami, "Regularity-Preserving Image Interpolation," *IEEE Transactions on Image Processing,* Vol. 8, No. 9, pp. 1293-7, September 1999.
5. W. K. Carey, S. S. Hemami, P. N. Heller, "Smoothness-Constrained Wavelet Image Compression," *IEEE Transactions on Image Processing*, Vol. 8, No. 12, pp. 1807-1811, December 1999.
6. S. S. Hemami, "Robust image transmission using resynchronizing variable-length codes," *IEEE Journal on Selected Areas in Communications, Special Issue on Error Resilient Image and Video Transmission,* Vol. 18, No. 6, pp. 927-39, August 2000.
7. Y. Yang, S. S. Hemami, "Generalized rate-distortion optimization for motion-compensated video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 10, No. 6, pp. 942-55, September 2000.
8. Y. Yang, S. S. Hemami, "Rate control for VBR video over ATM networks: simplification and implementation," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 11, No. 9, pp. 1045-58, September 2001.
9. M. G. Ramos, S. S. Hemami, "Supra-threshold wavelet coefficient quantization in complex stimuli: psychophysical evaluation and analysis," *Journal of the Optical Society of America:A*, October 2001.
10. D. M. Chandler, S. S. Hemami, "Effects of natural images on the detectability of simple and compound wavelet subband quantization distortions," to appear in *Journal of the Optical Society of America:A.*, July 2003.

11. D. M. Chandler, S. S. Hemami, "Dynamic contrast-based quantization for lossy wavelet image compression," submitted to *IEEE Transactions on Image Processing.*

Conference papers

12. M. Ramos, S. S. Hemami, "Edge-Adaptive JPEG Image Compression," *Proceedings of SPIE Conference on Visual Communications and Image Processing*, Orlando, Florida, March 1996.
13. S. S. Hemami, "Reconstruction-Optimized Lapped Orthogonal Transforms," *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Atlanta, Georgia, May 1996.
14. W. K. Carey, S. S. Hemami, P. N. Heller, "Smoothness-Constrained Wavelet Image Compression," *IEEE International Conference on Image Processing*, Lausanne, Switzerland, September 1996.
15. M. G. Ramos, S. S. Hemami, "Robust Image Coding with Perceptual-Based Scalability," *Proceedings Data Compression Industry Workshop '97*, Snowbird, Utah, March 1997.
16. G. J. Conklin, S. S. Hemami, "Multiresolution Motion Estimation," *ICASSP '97*, Munich, Germany, April 1997.
17. Y. Yang, S. S. Hemami, "Rate-Constrained Motion Estimation and Perceptual Coding," *Proceedings of IEEE Int. Conference on Image Processing*, Santa Barbara, CA, Oct. 1997, Vol. 1, pp. 81-4.
18. S. S. Hemami, "Visual Sensitivity Considerations for Subband Coding," *Proceedings of Thirty-first Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1997, Vol. 1, pp. 652-6.
19. W. K. Carey, L. A. Von Pischke, S. S. Hemami, "Rate-distortion Based Scalable Progressive Image Coding," *SPIE Conf. on Mathematics of Data and Image Coding, Compression, and Encryption*, San Diego, CA, July 1998, Vol. 3456, pp. 197-208.
20. Y. Yang, S. S. Hemami, "Rate-distortion-based Combined Motion Estimation and Segmentation," *IEEE Int. Conference on Image Processing*, Chicago, IL, Oct. 1998, Vol. 3, pp. 920-4.
21. M. G. Ramos, S. S. Hemami, "Activity-Selective SPIHT Coding," *Proceedings SPIE Visual Communications & Image Processing*, Vol. 3653, Pt. 1, San Jose, CA, Jan. 1999, pp. 315-26.
22. S. S. Hemami, T. Chang, R. Lau, "Resynchronizing Variable-Length Codes for Robust Image Transmission," *Data Compression Conference*, Snowbird Utah, April 1999, pp. 529.
23. Y. Yang, S. S. Hemami, "A rate-distortion optimized wavelet video coder," *Proceedings of Picture Coding Symposium*, Portland, OR, April 1999, pp. 61-4.
24. S. S. Hemami, "Distortion Analyses for Temporal Scalability Coding Techniques," *Proceedings of IEEE Int. Conf. on Image Processing,* Kobe, Japan, Oct. 1999.
25. Y. Yang, S. S. Hemami, "VBR Rate-distortion optimization for a motion compensated video coder," *Proceedings of IEEE Int. Conf. on Image Processing,* Kobe, Japan, Oct. 1999.
26. Y. Yang, S. S. Hemami, "Minmax Frame Rate Control using a Rate-Distortion-Optimized Wavelet Coder," *Proceedings of IEEE Int. Conf. on Image Processing,* Kobe, Japan, Oct. 1999.

27. S. S. Hemami, M. G. Ramos, "Wavelet coefficient quantization to produce equivalent visual distortion in complex stimuli," *Proceedings of SPIE Human Vision & Electronic Imaging*, San Jose, CA, Jan. 2000.
28. Y. Yang, S. S. Hemami, "Separate Source and Channel Rate Selection for Video over ATM," *Proceedings Data Compression Conference, Snowbird, UT,* March 2000.
29. M. Masry, S. S. Hemami, "Data Hiding in Images with Psychovisual Thresholding," *Proceedings IEEE Intl. Conf. Image Processing,* Vancouver, BC, September 2000.
30. M. G. Ramos, S. S. Hemami, "Perceptual Quantization for Wavelet-Based Image Coding," *Proceedings IEEE Intl. Conf. Image Processing,* Vancouver, BC, September 2000.
31. M. Eoin Buckley, M. G. Ramos, S. S. Hemami, S. B. Wicker, "Perceptually-based Robust Image Transmission over Wireless Channels," *Proceedings IEEE Intl. Conf. Image Processing,* Vancouver, BC, September 2000
32. M. Masry, S. S. Hemami, "Data Hiding in Images with Psychovisual Thresholding," *Proc. IEEE Intl. Conf. Image Processing 2000*, Vancouver, B.C., Canada, September 2000.
33. M. G. Ramos, S. S. Hemami, "Perceptual Quantization for Wavelet-Based Image Coding," *Proc. IEEE Intl. Conf. Image Processing 2000*, Vancouver, B.C., Canada, September 2000.
34. Y. Yang, S. S. Hemami, "Rate-distortion optimization for region- and object-based wavelet video coding," *Proceedings Thirty-fourth Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, October 2000.
35. S. S. Hemami, "Supra-threshold wavelet coefficient quantization in natural images: analysis and application to compression," *Int. Conf. on Image and Signal Processing*, Agadir, Morocco, May 2001.
36. S. S. Hemami, M. G. Ramos, "Quantifying Visual Distortion in Low-Rate Wavelet-Coded Images," *Proceedings IEEE Intl. Conf. Image Processing,* Thessaloniki, Greece, October 2001.
37. D. M. Chandler, S. S. Hemami, "Additivity models for suprathreshold distortion in quantized, wavelet-coded images," *Proceedings SPIE Human Vision and Electronic Imaging*, San Jose, CA, January 2002.
38. D. M. Chandler, S. S. Hemami, "Contrast-based quantization and rate control for wavelet-coded images," *Proceedings IEEE Intl. Conf. Image Processing,* Rochester, NY, September 2002. Winner (2nd prize), Best Student Paper Award.
39. D. M. Chandler, S. S. Hemami, "Suprathreshold image compression based on contrast allocation and global precedence," *SPIE Human Vision and Electronic Imaging 2003,* San Jose, CA, January 2003.
40. D. M. Chandler, M. A. Masry, S. S. Hemami, "Quantifying the visual quality of wavelet-compressed images based on local contrast, visual masking, and global precedence," to be presented at *Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 2003.