

REMOTE DRIVING WITH REDUCED BANDWIDTH COMMUNICATION*

Frederick W. DePiero, Timothy E. Noell, and Timothy F. Gee
Oak Ridge National Laboratory
Robotics & Process Systems Division
P. O. Box 2008
Oak Ridge, Tennessee 37831

"The submitted manuscript has been authored by contractor of the U.S. Government under contract No. De-AC05-84OR21400. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes."

Presented at the

Sixth Annual Space Operation, Application, and
Research Symposium (SOAR '92)
Sponsored by the NASA Johnson Space Center
Houston, Texas

August 6, 1992

*Research sponsored by the U.S. Army Human Engineering Laboratory and managed by Martin Marietta Energy Systems, Inc., for the U.S. Department of Energy under contract DE-AC05-84OR21400.

REMOTE DRIVING WITH REDUCED BANDWIDTH COMMUNICATION*

Frederick W. DePiero, Timothy E. Noell, and Timothy F. Gee
Oak Ridge National Laboratory
Robotics & Process Systems Division
P. O. Box 2008
Oak Ridge, Tennessee 37831

ABSTRACT

Oak Ridge National Laboratory has developed a real-time video transmission system for low-bandwidth remote operations. The system supports both continuous transmission of video for remote driving and progressive transmission of still images. Inherent in the system design is a spatiotemporal limitation to the effects of channel errors. The average data rate of the system is 64,000 bits/s, a compression of approximately 1000:1 for the black and white National Television Standard Code video. The image quality of the transmissions is maintained at a level that supports teleoperation of a high-mobility multipurpose wheeled vehicle at speeds up to 15 mph on a moguled dirt track. Video compression is achieved by using Laplacian image pyramids and a combination of classical techniques. Certain subbands of the image pyramid are transmitted by using interframe differencing with a periodic refresh to aid in bandwidth reduction. Images are also foveated to concentrate image detail in a steerable region. The system supports dynamic video quality adjustments between frame rate, image detail, and foveation rate. A typical configuration for the system used during driving has a frame rate of ~ 4 Hz, a compression per frame of ~ 125:1, and a resulting latency of < 1s.

INTRODUCTION

The use of untethered teleoperated vehicles for many remote operations is greatly limited because of a need to use low-bandwidth communication links. Vehicle control over a low-bandwidth channel is a necessity for tactical operations which, for example, require a low signature. Low-bandwidth channels are also encountered in underwater operations and in space applications. The most notable difficulty in using low-bandwidth channels is the problem of video transmissions from the teleoperated vehicle back to the driver's station. Given the availability of a low-bandwidth video transmission system, tactical remote operations such as reconnaissance, surveillance, target acquisition, and convoys, for example, would all become much more feasible. Oak Ridge National Laboratory (ORNL) has developed a real-time video transmission system for these types of low-bandwidth remote operations. The system supports both continuous transmission of video for remote driving and progressive transmission of still images.

The difficulty arising in the transmission of video is its extremely high data rate. Standard black and white video requires 60 M bits per second (bps). State-of-the-art tactical communication links, for example, support data rates in the range of 16 to 64 Kbps. Hydrophone data rates are lower still. A minimum factor of 1000 in video data rate reduction is required for remote driving via these types of

*Research sponsored by the U.S. Army Human Engineering Laboratory and managed by Martin Marietta Energy Systems, Inc., for the U.S. Department of Energy under contract DE-AC05-84Or21400.

low-bandwidth channels. Additional challenges exist to the problem of low-bandwidth remote driving beyond the high compression requirement. Driving experience using the ORNL system has demonstrated the significance of latency (image age) on driver performance. Both a low and a constant value of latency is very important. Image latency is affected by the duration of the encoding and decoding processes and by the data rate of the communication channel. Hence, compression techniques that possess deterministic processing times are more applicable for remote driving. The latency requirement together with the high compression rates makes it extremely unlikely that lossless compression techniques will ever be able to provide the required system performance [1]. In light of the fact that low-bandwidth remote driving will suffer from degraded imagery, several human factors studies have examined driver performance under these types of conditions [2][3][4].

Even lossy schemes with a high compression per video frame cannot meet the 1000:1 requirement alone [5]. Some reduction in frame rate is also necessary. Another facet of ORNL's development focused on an image-simulation technique that smoothed the interframe discontinuities associated with a reduced frame rate. These discontinuities are, of course, even more pronounced when the vehicle is driven on rough terrain.

APPROACH

The ORNL system is a hybrid version of the Laplacian pyramid approach to image compression [6]. This method decomposes an image into a set of subimages, each containing a separate spatial frequency band. By stacking the subimages vertically, the shape of a pyramid can be formed and, hence, the term 'image pyramid.' The motivation for this type of approach has its foundation in studies of the human visual system [7][8]. These studies demonstrated the significance of edge information to visual sensing. The studies also revealed a reduced sensitivity to gray-level errors that are present at edges. These results imply that edges should remain present in an image but permit them to possibly contain errors in their intensity values. Pyramidal methods of compression provide this sort of highly selective image degradation.

The block Discrete Cosine Transform (DCT) is another approach that has gained popularity in applications requiring high compression rates [9][10][11]. However, when operated at high compression rates, it suffers from the problem of producing noticeable artifacts at the DCT block boundaries. Block artifacts are not produced by pyramidal methods.

Image pyramids have also been used for scene analysis [12][13][14]. These types of analyses were not part of the ORNL remote-driving project. However, by adopting a compression scheme based on a similar type of image decomposition, results of this project provide an opportunity for a synergistic combination of an analysis plus compression system. An analysis system that could detect nearby obstacles, for example, would be of great benefit to a remote driving system that suffers from degraded imagery.

The hybrid aspect of the ORNL video compression system stems from several extensions to the Laplacian method that have been employed. The overall system uses a combination of several classical compression techniques and image foveation. A foveated image has reduced detail in the peripheral areas. This process mimics the structure of the human eye by placing a region of highest image quality at the center of the operator's field of view. This technique reduces bandwidth while still providing the driver with a feel for the terrain that is passing. The foveal center can be moved by the operator in realtime to adjust to the changing requirements of driving or for some other dynamic aspect of the remote operation such as when a vehicle enters a surveillance mode.

In addition to steering the foveal center, the ORNL system addresses the problem of dynamic adjustments in a more general sense. Five different preprogrammed video-quality settings are provided. These allow the operator to use the available bandwidth in a manner as effective as possible, given changing needs of the remote operation. Trade-offs can be made between image detail and frame rate or between the size of the foveal area and its rate of peripheral degradation, for example.

COMPRESSION ALGORITHM

The first step in producing a Laplacian image pyramid is to create another pyramid known as the Gaussian pyramid. It is generated by recursively applying a low-pass kernel to an image. The low-pass kernel used approximates a two-dimensional Gaussian function [6]. It has a normalized cutoff frequency of $\pi/2.0$ and produces a result that is subsampled by a factor of two in each direction. In this manner, each subsequent subimage, or layer, of the Gaussian pyramid is reduced in area by a factor of four from the previous layer. The ORNL system uses four pyramid layers.

The Laplacian pyramid is formed from the Gaussian pyramid by subtracting adjacent layers. To subtract two layers, the smaller one is expanded in area by a factor of four and then subtracted from its adjacent higher frequency layer. Each layer of the Laplacian pyramid contains a separate spatial frequency band of the original image. Layers of the Laplacian pyramid are referred to as 'subbands' because of this frequency decomposition. Expansion of the Gaussian layers was achieved via pixel replication followed by the application of a 2×2 averaging kernel. This method differs from Burt's [6], which used the Gaussian kernel both for expansion and for the recursive low-pass filtering. Slightly higher compression ratios were achieved in the ORNL system by switching to the 2×2 averaging kernel for expansion.

Once the Laplacian pyramid has been constructed, a uniform quantizer is applied to each subband. Each subband received a different degree of quantization to take advantage of the varying degree of visual sensitivity to errors in different spatial frequency bands [7][8]. The quantized subbands are then foveated by simply clipping the contents of each band that resides outside a rectangular region (see Figure 1). The centers of each foveal rectangle is collocated in the final image, and the size of each rectangle is determined by the desired rate of foveal degradation. The foveated bands are positioned to produce a gradual shift in image quality from the foveal center out towards the periphery.

To take advantage of the temporal correlation of images, experiments were made in differencing peer subbands in subsequent images. These experiments were designed to determine which bands should be processed in this manner and then to examine the effect of the duration of the differencing process. The process incorporated a periodic refresh to ensure a temporal limitation to channel errors. Taped driving imagery was used for input. The original vehicle transporting the camera had a speed of 10 mph. During the experiments, the compression system ran at a frame rate ranging from 3 to 5 Hz. Under these conditions, the lowest frequency band yielded a 15 to 20% improvement, and the second lowest showed a 5 to 10% improvement. Each varied with the scene content encountered on the tape. The two higher frequency bands did not yield any improvement in the cases examined. Most likely because these bands experienced substantial interframe differences with the vehicle speed and frame rate examined so that a net gain was not realized. The refresh period for both of the lower frequency bands was selected to be four frames. The realized compression did not substantially increase with longer refresh periods. The four-frame interval was chosen as a compromise so that the temporal duration of a channel error was limited to 1s.

Classical lossless techniques were employed in two stages to compress the quantized and foveated subbands into a one-dimensional bit stream. First, a zero-run-length coding technique [15] was used to process each row of a band. This process replaces a sequence of zero values with a single symbol indicating the length of the zero run. Nonzero pixels remain unchanged in this operation. Some runs of nonzero pixels did occur in the subbands, but the vast majority of runs consisted of zero pixels. To simplify the coder, it was decided to restrict the formation of runs to be those of zero values only.

The second stage of lossless processing used a Huffman coder [15] to assign a variable-length code word to each zero run and to each nonzero pixel. The code book for the Huffman coder was generated by using statistics gathered from driving imagery. The ORNL system provided the capability of on-line code book generation. An operator could specify starting and stopping times for the accumulation of image statistics while driving. In this way, code books could be tuned in the field for a given terrain. Once generated on board the vehicle, the code book was transmitted (in a lossless mode) to the decoder so that operation could begin on each side of the system using the new code books. Each subband's image statistics varied because of the different quantizers used. Different code books had to be used in each of these cases and when subbands were refreshed rather than differenced. The performance of the Huffman code books dropped by as much as 25% because of variations in scene content.

Using the above two classical techniques yielded a compression slightly higher than the zeroth-order entropy [15] in the higher subbands. An entropy measure is commonly used to determine the theoretically highest compression possible. Note that the calculation assumes a completely random arrangement of symbols in the data set. The quantized bands are far from random. The dominant (nonzero) components in each layer are typically associated with edges in the input image, so layers tend to contain large expanses of zero pixels with small clusters of nonzero values. Hence, the zeroth-order entropy is not a completely accurate metric for the Laplacian subbands.

Each subband was transmitted in a separate packet. Upon reception at the decompression unit, the bit stream was decoded and the image data was painted into frame buffers. The process of collapsing a Laplacian pyramid to recreate an image requires recursive steps of expansion and addition [6]. Several options for scheduling the time to collapse pyramids were explored. Nominally, a pyramid collapse could occur after all the subbands have been accumulated on the decompression side. Another possibility is to collapse the pyramid after each band is received. In the latter approach, final images contain a temporally skewed set of subbands.

A simulation scheme was investigated that employed temporal skewing. Bands were transmitted in order from highest frequency to lowest. A pyramid collapse following the arrival of each new band produced a gradual transition in the image from old to new. The edges lead the change in the image to continually provide the operator with a sense of the changing scene conditions. This technique yielded a pseudoincrease in the frame rate seen by the vehicle operator. Similar concepts have been applied to the area of improved-definition television by using temporal interpolation between subbands of subsequent images. Interpolative techniques are well suited for open-loop systems. Teleoperated systems could benefit from the improved image quality of interpolative approaches but cannot tolerate the accompanying increased latency. The logic behind the simulation approach studied here was to take advantage of subbands immediately upon reception. The use of temporally skewed subbands was considered to be a good starting point for addressing the needs for image simulation in a closed-loop system. Unfortunately, the technique had to be disengaged in the field. The large changes in imagery produced when the vehicle traversed moguls often resulted in noticeable residual artifacts. It became apparent that a more extrapolative technique is required. Future work will address the smooth extrapolation problem.

Another use for temporally skewed bands arises in an opportunity to robustly handle channel dropouts. The packet transmissions of each subband tended to perform in an "all or nothing" fashion. Packets were identified by frequency band and by frame number. If any subbands were dropped, the collapse heuristic was capable of substituting the old version of the same band and collapsing to form a new image. Hence, to be useful, an entire set of bands was not required.

Several close relatives to the Laplacian pyramid have been studied for image compression [16][17][18]. The motivation for choosing the Laplacian over these other methods is due to the property of temporal skewing discussed above. The most notable contender to the Laplacian approach is the Quadrature Mirror Filter (QMF). The QMF kernel produces a somewhat more compact decomposition of images than does the Laplacian and, consequently, has been more closely examined in recent years. Some debate has arisen over the merits of the Laplacian versus the QMF kernel. Vetterli contends, for example, that the Laplacian is a superior choice overall because of improved results in the area of motion-adaptive compression [19]. The Laplacian approach was chosen for use with the ORNL system for closely related reasons. The Laplacian has a tolerance to misregistration of subbands during the collapse process. The high-pass version of the QMF kernel produces bands that must be precisely aligned prior to collapsing to prevent noticeable artifacts. Given that driving imagery is typically always varying and because of an interest in temporally skewing subbands, the Laplacian approach to pyramid generation was chosen for the ORNL system.

SYSTEM ARCHITECTURE

The final configuration of the ORNL system consists of a two Versa Module European (VME) racks, one for compression and one for decompression. The compression rack is mounted on board a high-mobility multipurpose wheeled vehicle (HMMWV) outfitted for teleoperation. The decompression rack is mounted in an environmental enclosure adjacent to a VME-based Sparc station. An operator interface runs on the Sparc, providing vehicle control functions. The driving station and teleoperated HMMWV were developed by Harry Diamond Laboratories (HDL). The ORNL compression system provides video support for an Automatic Target Acquisition (ATA) system on board the HMMWV. The compression system transmits seven high-resolution still images at the start of targeting operations. As the ATA system acquires targets, the compression system transmits small rectangular portions of images containing the tracked targets. These were typically ~40 x 60 pixel in size. At the operator station, the decompression system pastes targets at appropriate locations within the frame buffers. The contents of the tracking buffers are displayed on the operator control station. A socket-based custom protocol is used to communicate between the ORNL compression system, the operator control station, and the ATA system.

Each rack contains three single-board computers, Datacube image processing hardware, and memory cards. The first processor in the compression rack is responsible for controlling the Datacube equipment and for determining the image capture rate. The second processor performs the zero-run-length and Huffman coding. The third interfaces to the radio. The processors in the decompression rack also form a pipeline for images and perform symmetrical functions.

The packet radios use a spread-spectrum type of modulation and operate in the 902 to 928-MHz band. The units provide a wireless Ethernet bridge between the VME systems. The low-bandwidth communication channel is emulated by using a real-time clock to maintain a data rate at 64 Kbps. The average rate was also monitored at the receiving unit for verification. Fluxuations down to ~ 55 Kbps commonly seen were due to competing (RF) traffic. It was necessary to adjust the physical packet size of the transmitted data to improve the system's ability to coexist with other nearby transmitters in the same RF band. The physical packet size was adjusted by modifying the operating systems' Ethernet driver.

The ORNL system is capable of using either protocol with the Ethernet bridge. TCP sockets guarantee faithful delivery of all packets, in order, and will retry indefinitely to provide such. UDP sockets do not provide a similar guarantee. This flexibility provides the option to avoid exhaustively attempting to retransmit old subbands. In the event of a dropout, the system simply begins the transmission of the next new band. Given the tolerances of the Laplacian pyramid described above, bands can be dropped without severe consequences. At the demonstration of the system, it was operated in a TCP mode. Hence, the effect of channel errors on video quality was not a visual blemish. Rather, it was an increased delay in transmission for that band. It is believed that the design aspects affecting channel noise tolerance that were made part of the system were a worthwhile development effort, although they were not fully exercised at the first demonstration. It is anticipated that these aspects of the system will come to fruition with future versions of the system.

In addition to the two-rack compression system, a one-rack simulation system has also been developed. The simulator has been delivered to the U.S. Army's Human Engineering Laboratory (HEL) for human factors studies on remote driving. The simulator is capable of producing the same degraded imagery and of emulating the latency present in the two-rack version of the system. Future work at ORNL will incorporate the recommendations indicated by HEL's studies.

CONCLUSIONS

The ORNL video compression system was demonstrated in April 1992. The system supports teleoperation of an HMMWV at speeds up to 15 mph on a moguled dirt track. During driving tests, the compression per frame ranged from ~ 105:1 to 145:1, depending on scene content. The frame rate varied as a function of the realizable compression, ranging from 3 to 6 Hz. Latency of the system was determined to be ~ 1s. Future work in this area will address improvements to the compression algorithm, the problem of temporal extrapolation, and the transmission of color images.

ACKNOWLEDGMENTS

The authors thank Mr. Charles M. Shoemaker and Mr. Thomas W. Haduch for programmatic guidance throughout this project. The authors acknowledge the cooperative attitude and excellent development efforts by the members of the Advanced Sensor Systems Division of HDL.

REFERENCES

- [1] Hong, T. H., Nashman, M., and Chaconas K. (1988). "Image Compression Technology and Techniques," unpublished manuscript, National Bureau of Standards, Gaithersburg, Md.
- [2] McGovern, D. E. (1987). "Experiences in teleoperation of land vehicles" (SAND87-1980C), Albuquerque, N.M.: Sandia National Laboratories.
- [3] Shoemaker, C. M. (1990) "Low data rate control of unmanned ground vehicles," *proceedings: Association for Unmanned Vehicle Systems AUVS-90*, Dayton, Ohio.
- [4] Miller, D. P. (1987). "Evaluation of vision systems for teleoperated land vehicles," *proceedings IEEE Systems, Man and Cybernetics Conference*.

- [5] Herman, M., Chaconas, K., Nashman, M., and Hong, T. H. (1988). "Video compression for remote vehicle driving." *Proceedings SPIE Advances in Intelligent Robotics Systems: Mobile Robots III*, Cambridge, Mass.
- [6] Burt, P. J., and Adelson, E. H., "The Laplacian Pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, pp. 532-540, April 1983.
- [7] Carlson, C. R., and Cohen, R. W., *Visibility of displayed information*, Office of Naval Research, Technical Report, Contract N00014-74-C-0184, 1978.
- [8] Carlson, C. R., and Cohen, R. W., "A simple psychophysical model for predicting the visibility of displayed information," *Proc. Soc. Inform. Display*, pp. 229-246, 1980.
- [9] Ahmed, N., Natarajan, T., and Rao, K. R., "Discrete cosine transform," *IEEE Trans. Comput.*, vol. C-23, pp. 90-93, January 1974.
- [10] Polysongsang, A., and Rao, K. R., "DCT/DCPM Processing of NTSC composite video signal," *Proc. Int. Symp. Circuits Syst.*, Tokyo, Japan, July 17-19, 1979, pp. 981-982.
- [11] Ligtenberg, A., and Wallace, G. K., "The emerging JPEG compression standard for Continuous Tone Images: An Overview," *Proc. Picture Coding Symposium, Cambridge, Mass., March 1990*, pp. 6.1-1.—6.1-4.
- [12] Anderson, C. H., Burt, P. J., and van der Wal, G. S., "Change detection and tracking using pyramid transform techniques," *Proc. SPIE Conf. on Intell. Robots and Computer Vision*, Boston, pp. 72-78, 1985.
- [13] Bergen, J. R., and Adelson, E. H., "Hierarchical, computationally efficient motion estimation algorithm," *J. Opt. Soc. Am. A*, 4, p. 35, 1987.
- [14] Burt, P. J., "Multiresolution techniques for image representation, analysis, and 'smart' transmission," *Proc. SPIE Conf. on Visual Communications and Image Processing*, pp. 2-15, Philadelphia, Pa., November 1989.
- [15] Jain, A. K., "Image data compression: A review," *Proceeding of the IEEE*, vol. 69, pp. 349-384, March 1981.
- [16] Vetterli, Martin, "Multidimensional sub-band coding: some theory and algorithms," *Signal Processing*, 6(2):97-112, February 1984.
- [17] Adelson, E. H., Simoncelli, E., and Hingorani, R., "Orthogonal pyramid transforms for image coding," *Proceedings of SPIE*, October 1987.
- [18] Pentland, A., and Horowitz, B., "A practical approach to fractal-based image compression," Technical Note 152, M.I.T. Media Lab Vision and Modeling Group, Boston, Mass., November 1990.
- [19] Uz, K. M., Vetterli, M., and LeGall, D. J., "Interpolative multiresolution coding of advanced television with compatible subchannels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 1, no. 1, March 1991.

Figure 1 (Facing). An original image (upper left), a compressed image (upper right), and a Laplacian pyramid which has been quantized and foveated.