Series/1-based videoconferencing system

by D. Anastassiou
M. K. Brown
H. C. Jones
J. L. Mitchell
W. B. Pennebaker
K. S. Pennington

Discussed is a new videoconferencing system that has been developed and deployed at several IBM locations. This system transmits high-quality monochrome, freeze-frame images over dial-up telephone lines between two (or three) dedicated videoconferencing rooms. There are two main system components. An IBM Series/1 provides control, communication, data compression, and storage, and a Grinnell GMR-270 image processing display system implements image acquisition, processing, and video buffering functions. Conference participants may choose either a basically black and white rendering of an image for fast transmission or a continuous-tone rendering with a longer transmission time. Details are given regarding the system configuration, function, and operation.

The increasing cost of business travel makes the ability to conduct business meetings via teleconferencing an attractive alternative. Reduction in cost of business meetings provides the current major impetus for the development of suitable teleconferencing and videoconferencing systems. Also, the development of these systems may substantially influence future office communications and result in a significant improvement in the ability of geographically remote business units to communicate, coordinate, and conduct joint programs. At the present time, our experience suggests that executives and professionals are willing to use some form of teleconferencing. Technological trends indicate that the functional capabilities of teleconferencing and videoconferencing systems will continue to improve and communication costs will decrease. Thus, teleconferencing functions of various types seem likely to emerge as a major component in offices of the future.

When it is viewed as an area of technical and business activity, teleconferencing may be divided into the following four major categories:

- Audio teleconferencing.
- Audio teleconferencing with graphics assist.
- Freeze-frame videoconferencing.
- Full-motion videoconferencing.

Recently, a fifth form of conferencing, namely, shared-screen computer conferencing, has also gained recognition as a useful form of business interaction.[1,2]

Audio teleconferencing via the telephone network originated more than forty years ago. It can be used for group conferences and has the definite advantage that individuals at many separate locations can be linked into such conferences with access to a telephone only. Audio teleconferencing is a useful tool, although it lacks the ability to communicate visual information.

Videoconferencing, on the other hand, attempts to replicate closely the normal environment of a business meeting, and thus minimizes the need for participants to develop a new set of communication skills. Full-motion color videoconferencing with large-screen projection monitors provides a meeting environment that most closely resembles the face-to-face environment present in most business meetings. However, the large transmission bandwidth required to provide an adequate full-motion rendition of the proceedings results in relatively high transmission costs. These costs, which presently are of the order of \$2400 per hour,[3] are appropriate for meetings with large numbers of participants—stockholder meetings, company-wide branch office meetings, tele-education, etc. The benefits of full-motion videoconferencing are far more difficult to justify on a basis of cost for the majority of small business meetings.

The key to this faster response is data compression.

Freeze-frame videoconferencing, as the name implies, does not give the conferees the benefits of a full-motion rendition of the events taking place at the remote meeting site. It does, however, provide a meeting environment in which the participants can exchange pictorial and graphic information, in addition to the normal voice communications. Usually, the freeze-frame images are transmitted over low-bandwidth communication channels. Thus, the cost per meeting hour is significantly lower than the cost of full-motion videoconferencing. A number of freeze-frame videoconferencing systems are capable of transmitting images in one to two minutes over ordinary telephone lines, the cost of which is in the range of \$40–\$90 per hour. This makes practical the application of videoconferencing to smaller business meetings. The one-to-two-minute transmission times, however, are considered by many meeting participants to inhibit the free exchange of information. Higher-bandwidth leased lines can be used to achieve higher-speed transmission, but they are significantly more expensive.

The Series/1-based videoconferencing system described in this paper transmits images by digital techniques faster than typical for systems utilizing dial-up telephone lines (4.8 kbits per second). The key to this faster response is data compression.

For exact reproduction of continuous-tone monochrome images, compression can reduce the time by a factor of only about three. If a high-contrast approximation of the original image is sufficient, the data can be reduced to a much greater extent. A good example of this is the current generation of digital facsimile machines, which first reduce the image to a black-and-white form and then further compress it by factors of ten to thirty to achieve subminute transmission times for a full page. For gray-scale images, tradeoffs are also made between faithfulness of the decompressed image to the original image and the amount of compression obtained. The process of maximizing the compression, while maintaining excellent image quality, is still an art. (Figures 2B, 4, 5, 6, and 7 are images produced by the videoconferencing system discussed in this paper.)

A key factor in the final image quality is the number of samples taken of the image. The Series/1 videoconferencing system uses an image format of 512 pixels per line (picture elements) and 480 lines per image. This is compatible with a National Television Standards Committee (NTSC) format and is well-matched to higher-performance cameras and displays. Image processing techniques are used to maximize image quality within the constraints of this format and the encoding techniques used.

Also, several aspects of the system discussed in this paper help provide good human factors. The system is quite easy to operate because most of the basic operations can be performed by pressing one or two buttons on a control console. Fast response is obtained by providing two modes of operation: a high-speed graphics mode (much like facsimile), which is suitable for line copy or text; and a slower continuous-tone mode for images that are not intrinsically black and white. Typical images of graphics materials (i.e., most of the visuals presented in business meetings) are transmitted in nine to twelve seconds. Typical gray-scale images (e.g., the persons in the videoconferencing rooms) are

Figure 1 A typical conference room

transmitted in 30–40 seconds, at 4.8 kbit-per-second rates with image quality sufficient for almost all needs. The image processing and compression techniques are sufficiently robust that most images can be sent by either technique with acceptable results.

The IBM videoconferencing system has been designed for use with dial-up lines. This makes possible immediate access (given that the dialed room is free) to any part of the system, without prior commitment to dedicated leased lines. Cost-attractiveness improves considerably when the telephone cost is incurred only with actual usage.

System functions

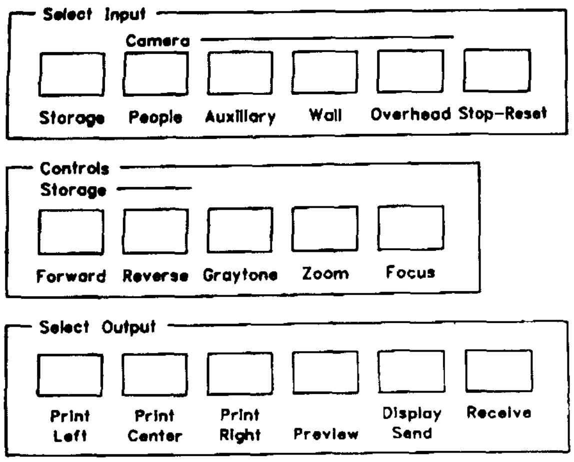
The IBM videoconferencing system is structured for communication between two (or three) dedicated rooms. The facilities in each room are designed for groups of up to seven people. Both audio and freeze-frame video information are exchanged over normal dial-up telephone links.

Figure 1 is a photograph of a typical videoconference room. The conferees position themselves around a tapered table facing the three 23-inch television monitors recessed into the wall at the front of the room. In two-way conferencing, one of the monitors displays an image of the people in the other room, another monitor displays an image of the most recently transmitted foil or similar graphics material, and the third monitor is used to display images recalled from storage. In three-way conferencing, this third monitor might be used for a "people picture" of the second remote room.

In each room, one of the participants operates a control console that is inset in the table, as shown in Figure 2A. A block diagram of the control console buttons is shown in Figure 2B. This console is used for camera selection, camera control, and functional

Figure 2 (A) Control console; (B) Diagram of the control console buttons

(A)



(B)

control of the system. The small (8-inch) screen associated with the console is used for monitoring live video from the selected camera and for review of images recalled from local storage.

The camera at the front of the room has a fixed-focus, wide-angle lens for viewing the people seated around the table. Images acquired from this camera are always sent as continuous-tone (gray-scale) images and are not retained on disk in either system.

There are two other cameras. One is mounted in the ceiling and pointed down at an area on the table adjacent to the control console as input for transparencies and other graphics on the table. The other camera is directed toward a chalkless board or an easel to pick up information written and displayed there. Both of these cameras have motorized zoom lenses that the operator adjusts from the control console using the ZOOM and FOCUS controls. For these latter two cameras, the image may be sent in either a gray-scale mode (8 bits per pixel) or as a primarily black-and-white image (1 to 2 bits per pixel). These modes are selectable, using the GRAYTONE button shown in the middle of Figure 2B. Once the transmission mode has been selected, pressing the SEND button starts the encoding and transmission of the image to the other site(s). As the image is transmitted, it is painted on the graphics monitor of the sending room. It is also stored on disk in both the sending and receiving systems. If STOP-RESET is pressed before transmission is completed, the transmission is terminated and the stored compressed data for that image are discarded by both sender and receiver. The RECEIVE light indicates that the system is receiving image data from another room.

The console operator in either room can review the images previously stored on disk by selecting input from STORAGE and sequentially pressing FORWARD or REVERSE. These images are viewed only on the console monitor, leaving the images on the large monitors at the front of the room undisturbed. If, however, SEND is depressed while input from STORAGE is selected, the image currently viewed on the small monitor in that room is decoded simultaneously in both rooms on the large monitors used for image recall.

The PREVIEW function has not yet been implemented. When available, it will allow the operator to view a processed image on the console monitor.

That image is to be transmitted and saved without disturbing the images on the room monitors until SEND is pressed.

A Tektronix 4634 Image Forming Module supplies hard copy from any of the monitors at the front of the room. The PRINT LEFT, PRINT CENTER, and PRINT RIGHT buttons control this function.

On another small control panel (not shown) is an END CONFERENCE button. Pressing this button once puts the system into local mode, allowing review of the conference images without the need for a communication link. Pressing the button a second time completely terminates the session. Images are saved on diskette at this time, if that option is requested. The image files on the hard disk are erased, and the system is recycled to start the next conference.

System configuration

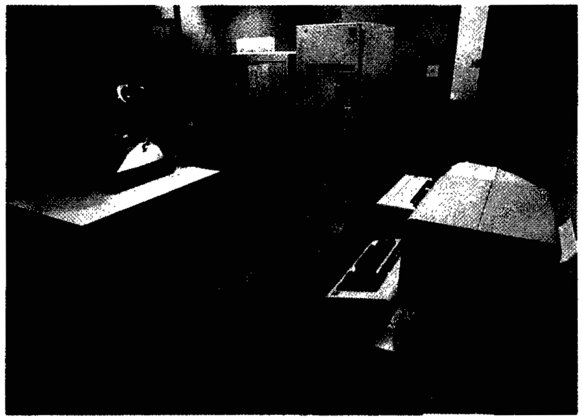
The hardware configuration of the system is shown pictorially in Figure 3. The heart of the videoconferencing system is an IBM Series/1, which is located

Communication tasks are handled by two IBM Bisync Single Line attachments.

in an adjacent equipment room, along with the image processing and display subsystems and other necessary hardware. In addition to carrying out the normal system functions, the Series/1 senses operator requests, controls the communication links, manages image storage functions, generates program instructions for the image processing and display subsystem, and does the actual compression and decompression of the image data.

The entire software system runs under CM1—the Communications Monitor software package—which, in turn, runs under the Realtime Programming System (RPS) in the Series/1. The compression and decompression and the image processing code comprise a separate software package that is called as a subroutine by control code. This package runs as an applications program under CM1.

Figure 3 Configuration of the videoconferencing system hardware



Communication tasks are handled by two IBM Bisync Single Line attachments. Each attachment communicates with an IBM 3845-12 encryption box and then through IBM 3864-2 modems to data phones that are connected to switched-network-voice-grade lines. The modems typically transmit at 4.8 kbits per second.

The data link between the Series/1 and Grinnell GMR270 is through the IBM Programmable Cycle-Stealing Digital Input/Digital Output (DI/DO) attachment. Since each image contains roughly one-quarter million bytes of data, the high-speed data link is essential. The processing of a graphics image requires more than 20 000 bytes of instructions to the image processor, which is another reason for the high-speed link.

The control console is monitored by the IBM Integrated Digital Input/Digital Output (DI/DO) attachment. The same attachment also controls the video switching as well as the lights that inform the operator about current system activities and status.

The system uses both an IBM 4963 disk drive and a IBM 4964 diskette drive for nonvolatile storage. The disk has a 29-megabyte capacity. The diskette can read/write single- or double-sided 8-inch diskettes. Typical graphics images compress to approximately 3000 bytes, while gray-scale images compress to about 15 000 bytes. A conference rarely requires more than twenty to thirty images, most of which are compressed in the graphics mode. Consequently, the images from a single conference can easily be stored during the conference, and a single 8-inch diskette is usually sufficient for long-term storage of conference images. As noted earlier, if the images are not saved on diskette they will be lost. For security reasons, images are erased when a conference is terminated. Images saved on diskette can be preloaded for a later conference if desired.

The IBM 4978 display and keyboard and the IBM 4974 printer are used primarily during system initialization and for diagnostic reporting. The audio link is through a Conference 2000 half-duplex system. This unit, consisting of an omnidirectional microphone and speaker system, is anchored at the center of the table.

The graphics image type is selected when the desired image is basically black and white in nature.

Videoconferencing has been in use in IBM for many years. An earlier system used a NEC DFP-754B-TR Telephone Video System, coupled with a small IBM computer that is not available commercially.[4] Images captured with this equipment have a format of 384 pixels by 254 lines. A requirement for the system that is being discussed in this paper is the ability to communicate with rooms having the lower-resolution format. The older rooms were retrofitted with Series/1s so that communication compatibility would be straightforward. In the older rooms, the Series/1 performs all the image processing and format translation, compression and decompression, system control, and image storage functions.

Image processing and compression

As noted earlier, two distinct types of images are transmitted: gray-scale images and graphics images. The graphics image type is selected when the desired image is basically black and white in nature. However, sections of the document containing material that does not fit this assumption (e.g., photographs) are usually rendered recognizably. The gray-scale algorithm is generally used when the image is basically continuous in tone. At the same time, any text is sent with virtually no loss of quality, compared to the original source image. Only an outline of the image processing and compression algorithms used in the system is presented here. (The algorithms will be described in greater detail in forthcoming publications.)

Gray-scale algorithm. When an image is compressed using the gray-scale algorithm, the full-frame video image is separated into two fields for the encoding process. The first field contains the even-numbered lines, and the second field contains the odd-numbered lines.

Encoding the first field. The first field is encoded by a Differential Pulse Code Modulation (DPCM)-related technique,[5] in which many modes of operation produce a variable-bit-rate code. To encode a current pixel, the pixel values to the left (A), above and to the left (B), and above (C) the current pixel are used to estimate the value of the current pixel (X). Each estimate is made according to the following prediction formula, which is widely used in DPCM schemes: $PRED = A + 1/2\,(C - B)$.

The prediction error, $X\text{-}PRED$, is then quantized according to a set of fixed quantization levels. The values A, B, and C just defined are quantized values rather than original values, because both the encoder and decoder must make the same prediction for the current pixel. For the prediction error, information from D (to the right of C on the line above) is also used. The prediction errors of previously encoded pixels are used to determine a quantity, $DIFF$, that measures the degree of confidence in this estimate. This quantity also measures the degree of smoothness in the local image area. $DIFF$ is expressed as follows: $DIFF = max\,(|E_a|, |E_b|, |E_c|, |E_d|)$. The Es are integer quantities that give the prediction error measured in allowed quantization steps, at the four previously coded pixel locations closest to the current pixel. The quantity $DIFF$ has been used to determine the quantization levels adaptively.[6] The same indicator is used here for a different purpose, namely to assign the number of bits N that are to be used to code the current pixel, while keeping the quantization levels constant.

For each pixel, an additional error bit is used to indicate whether accurate coding can occur with N

bits. If N bits are not sufficient, further correction bits are sent. In most cases, N is 0. Thus, the predicted value can be used as is, and only the error bit is sent. In this manner, the quality of a 5-bit (32 quantization levels) DPCM scheme is achieved at a much lower bit coding rate.

For additional compression, the states resulting from this multimode DPCM scheme are coded using a form of entropy coding called arithmetic coding,[7] in which the statistical redundancy of the error bits is exploited. This entropy coding technique allows any data compression problem to be separated into two independent parts: (1) finding a good statistical model for the entity to be coded, and (2) optimally coding the states resulting from that model. In the case of the error-bit coding, the probability of a bit being a 0 or a 1 is estimated from the error bits previously encoded. Coding occurs pixel by pixel, and the problem of interleaving arithmetically coded bits among the regularly coded bits is solved by arithmetically coding all bits and assuming uniform probability distributions for all except the error bits. The significant skew in the statistics of the error bits allows the overall bit rate to be reduced to a small number. The limit for the coding of a completely flat image is close to zero bits per pixel.

The gray-scale algorithm is generally used when the image is basically continuous in tone.

Encoding the second field. A full frame of video consists of two interlaced fields captured one-sixtieth of a second apart. While the algorithm just described is used for the first field, another form of DPCM algorithm is used for the second field. Normally, the first field is a good approximation to the complete image. Some videoconferencing systems transmit only a single field. The penalty in compression for transmitting the video image as two independent fields is typically less than twenty percent for the first-field algorithm. Consequently, a good approximation to the final image can be transmitted in about one-half the time the full frame requires.

There is a practical reason for not transmitting the second field with the same fidelity as the first field. If the subject moves during the video frame capture, the resulting two fields can be displaced from one another on the screen, giving rise to an annoying motion-induced flicker. That is, when the two fields are displaced, the interlace no longer provides an effective 60-hertz refresh rate. The flicker in the resulting image is distracting, unpleasant, and sometimes produces unintended comic effects. The algorithm used for the second field suppresses this flicker while retaining almost all of the quality of the full-frame video image.

The second-field algorithm is conceptually similar to the first field in that both use a predicted value *PRED* and an activity indicator *DIFF* for each pixel. For the second field, however, the current pixel is estimated from the quantized values of the pixels that are known to the decoder above (A) and below (B) from the first field,[8] as the average of these two values: $PRED = 1/2\ (A + B)$. (Note that the definitions of A and B differ for the first- and second-field algorithms.) The activity measure is obtained from the first field by a simple relation that is a measure of the vertical gradient magnitude in the image, as follows: $DIFF = max\ (|A - B|_{-1}, |A - B|_0, |A - B|_{+1})$ where the -1 and $+1$ refer to the nearest horizontal neighbors. All pixels for which the activity indicator *DIFF* has a value lower than some limit are forced to the value of *PRED*. This is the step that eliminates motion-induced flicker. Pixels with non-zero values of *DIFF* are adaptively DPCM-coded and compressed using Huffmann coding techniques. Typically, the second field can be compressed to about twenty percent of the data required for the first field. Although this approach incorporates, in theory, the risk of erroneously coding some pixels, the probability of such occurrence—after having accurately encoded the first field—is extremely small. At the same time, the savings in compression are very high.

Image preprocessing and postprocessing are essential components of the gray-scale image compression that greatly improve compression-versus-distortion performance. Image smoothing (noise reduction) filtering techniques increase the compressibility of an image by making the image more predictable. Such filtering is especially beneficial in conjunction with the gray-scale compression algo-

Figure 4 (A) Gray-scale image A; (B) Gray-scale image B; (C) Enlargement of gray-scale image B after the first-field decoding; (D) Enlargement of gray-scale image B after the second-field decoding

rithm previously given because the probability of the zero-bit-per-pixel mode increases and the statistics of the error bits are more skewed. Linear, low-pass filters, however, have the disadvantage of blurring the edges while reducing the noise. Therefore, a nonlinear filter is used to retain edge sharpness and still strongly smooth the image. This concept is similar to that described by Graham,[9] the main difference being the way in which nonlinearity is incorporated. In its simplest form, the value of a current pixel is modified by adding to it the properly weighted differences between the value of that pixel and the values of its two horizontal nearest neighbors, except that these differences are clipped to some maximum allowed limit. Filters with contributions from a larger neighborhood are structured similarly. The application of these filters with properly selected coefficients yields smoothed images that are nearly indistinguishable from the original. Compression is improved from thirty percent to more than two hundred percent, depending on the image. Finally, application of the same filtering scheme in the vertical direction is used as a postprocessing stage, in order to eliminate slight horizontal streaks that can occur in smooth areas of the reconstructed image.

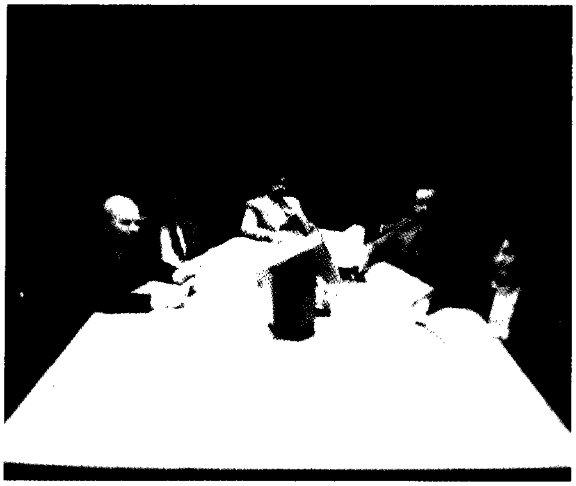
An original gray-scale image is shown in Figure 4A. The decoded image is shown in Figure 4B. This image has been compressed to 74 records of data (0.6 bits per pixel). Figures 4C and 4D show a section of the same image after first-field decoding and interpolation and after second-field decoding. The addition of the second-field information causes the "staircase" slanted edge to become sharp and continuous. Figure 5 is a reconstructed "people picture" from a videoconferencing room.

Graphics algorithm. When an image is compressed using the graphics algorithm, the original 8-bit image is transformed to a basically black and white image. It is then encoded exactly, using the Group III CCITT standard (Modified READ) coding algorithm.[10] (As will be explained shortly, more than two gray levels are actually used.) This transformation is traditionally done using a constant threshold selected from the histogram of the image, so that the black portions (usually graphics information or printed matter) are separated from the white portion (usually background). One problem with this approach is that the edges do not always appear crisp, smooth, and continuous. With the Cohu 4400 video camera used in our systems, some edge enhancement prior to thresholding is desirable. This is done using a Laplacian operator, which is defined as the difference between the current pixel and an average value of a small neighborhood around it. If the value of the Laplacian, properly scaled, is added to the actual value, the resulting image has enhanced high-frequency components, resulting in crisper edges and better definition of the high-frequency patterns typical of text. Noise is also enhanced by this operation, but it does not become deleterious because of filtering both before and after the edge enhancement.

The filtering operation used before the high-frequency enhancement is identical to that used to smooth streaks in the reconstructed gray-scale image. Such filtering reduces the occurrence of saw-toothed edges caused by camera vibration.

The constant-threshold approach is also affected by shading due to uneven lighting, lens vignetting, shadows, etc. Shading can result in a thresholded image in which printed characters have nonuniform stroke width. Some darker white segments of an image may even be thresholded to the black state. A shading correction has therefore been devised to transform a picture to one in which the background values have been flattened, thereby permitting proper thresholding to be performed easily.

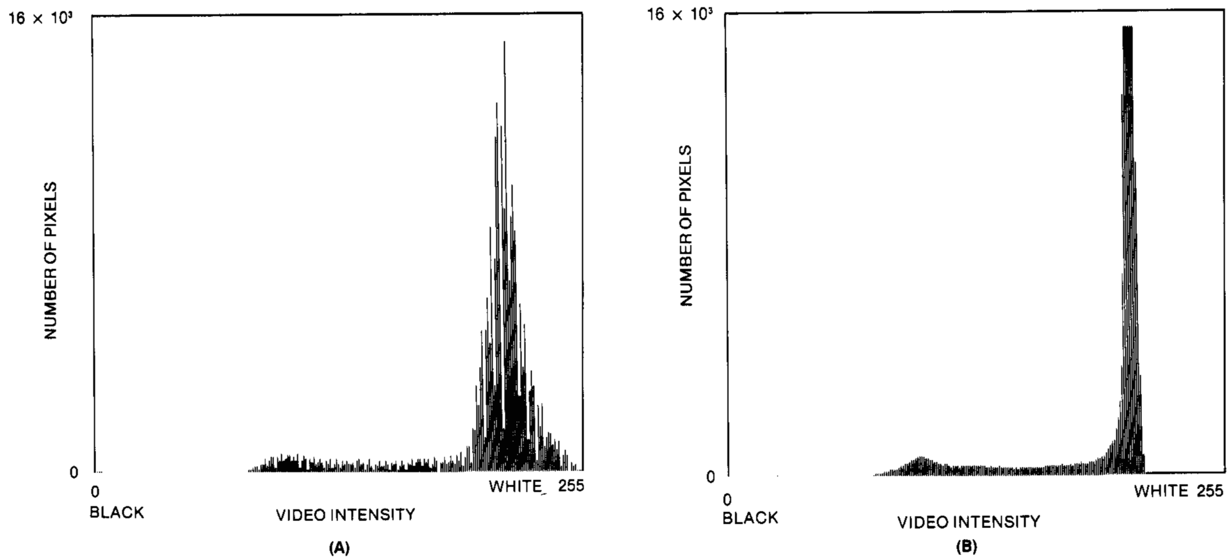
Figure 5 A typical "people picture"



The shading correction is basically a nonlinear adaptive high-pass filter in which a maximum white (or maximum black) value is found by sampling the neighborhood around each pixel. This sample, which is a good estimate of the background intensity value in the vicinity of the current pixel, is used to correct each pixel and force background areas to uniform luminance. The histograms of gray-level distribution for a typical image before and after this transformation are shown in Figures 6A and 6B. The histogram of the transformed image is now clearly separated into a narrow high peak, representing the background pixels, and a smaller clustering of darker values, representing the black pixels. After this transformation, thresholding to a bilevel image is straightforward.

At pixel locations at transitions between black and white in the bilevel image, more than two intensity levels are allowed. Elsewhere, however, the image is allowed to be bilevel only. The effect of increasing the number of allowed gray levels around the edges of characters is visually equivalent, when observed from some distance, to having increased the image resolution. Staircase effects are reduced, and details regarding the curvature of the character edges become more apparent. The information of the gray levels around the edges is small and is coded without large sacrifice in the total bit coding rate.

Figure 6 (A) Histogram of a typical graphics source image; (B) Histogram of shading-corrected graphics source image



Unlike the gray-scale algorithm, substantial changes occur during the transformation of the graphics images, as can be seen in Figures 7A (the original graphics image) and 7B (the reconstructed graphics form). This image has been compressed to five records, or about 0.04 bits per pixel. In black-and-white sections of the image, legibility is often enhanced during the transformation. Sections of an inherently gray-scale image can be rendered only approximately.

The operator chooses the mode of image transmission. However, there is considerable overlap between the ability of the gray-scale algorithm to accurately reproduce text and the ability of the graphics preprocessing to produce recognizable renderings of gray-scale images.

System performance

When we started, our goal was a nominal ten-second response for graphics images and a nominal forty-second response for gray-scale images. Transmission was to be on a dial-up telephone system (4.8 kbits per second). These goals were based in part on user surveys that indicated a need for a typical ten-second response and in part on estimates of what was possible for graphics and gray-scale images.

The actual system performance meets these goals. Times for various operational modes, shown in Figures 8A and 8B, are plotted as a function of the number of records (252 bytes of compressed image data per record). Record count is a good measure of image complexity. Times are for completed tasks. The time for sending an image (SEND), for example, is the interval from pressing the SEND button to completion of the image in the receiving room and acknowledgment of that fact to the sender. The time for recall from disk (RETRIEVE) is the interval from pressing SEND with STORAGE selected to complete decoding of the stored image in both rooms. In local mode (SEND(L) or RETRIEVE(L)), time is the interval required to accomplish the complete task without any communication over telephone lines.

The slope of the linear portion of the SEND data is that which would be expected from a 4.8 kbit-per-second line with normal bisync communication protocol. The deviation from linear behavior near the origin is due to two distinct effects. First, for simple images, the process becomes computationally bound, rather than being limited by transmission times. For gray-scale images compressed to fewer than forty records, the time required to process the 250 000 bytes of image data is longer than that required to send the compressed data. This is apparent from an examination of the

SEND(L) data. A similar effect occurs in the graphics mode for record counts of fewer than five. Second, an additional increase in times for the very simplest images occurs because data are generated too slowly to build up much overlap in the sending encoder and the receiving decoder. The increase does not occur in either the image recall or the local mode, because data do not have to be transmitted. The slope for both gray-scale and graphics encoders and decoders is about 0.12 seconds per record, if image data are not being transmitted. Faster transmission lines would, therefore, improve response somewhat, especially for very complex images.

If a third location is included in the conference, two monitors can be used for "people pictures" from the two remote sites and the remaining monitor can be used to display graphics material.

A significant amount of the time is required for fixed system overhead: disk file operations, image processing, reading and manipulating the 250 000 bytes of image data, initial messages required to start the image transmission process, and image postprocessing and disk-storage operations.

Typical graphics images, such as that shown in Figure 7B (five records), are usually completed in times of the order of ten seconds. Typical gray-scale images, such as pictures of persons in a videoconference room, are usually completed in less than forty seconds. To viewers in the rooms, the perceived time is actually slightly less than these plots indicate, especially for gray-scale images. The graphics images whose data points are greater than fifty records are from a half page of three-column magazine print that has more than one hundred twenty characters per line. These same images were used to generate gray-scale data points consisting of more than ninety records. Although the resulting graphics was legible, the text was too small to be read from the conference table. Source material with small-point-size text is beyond the capabilities of the system.

Communications

Most videoconferences take place between two locations, so that one of the three monitors displays the persons in the remote room and two monitors are

Figure 7 (A) An original graphics image; (B) A reconstructed graphics image

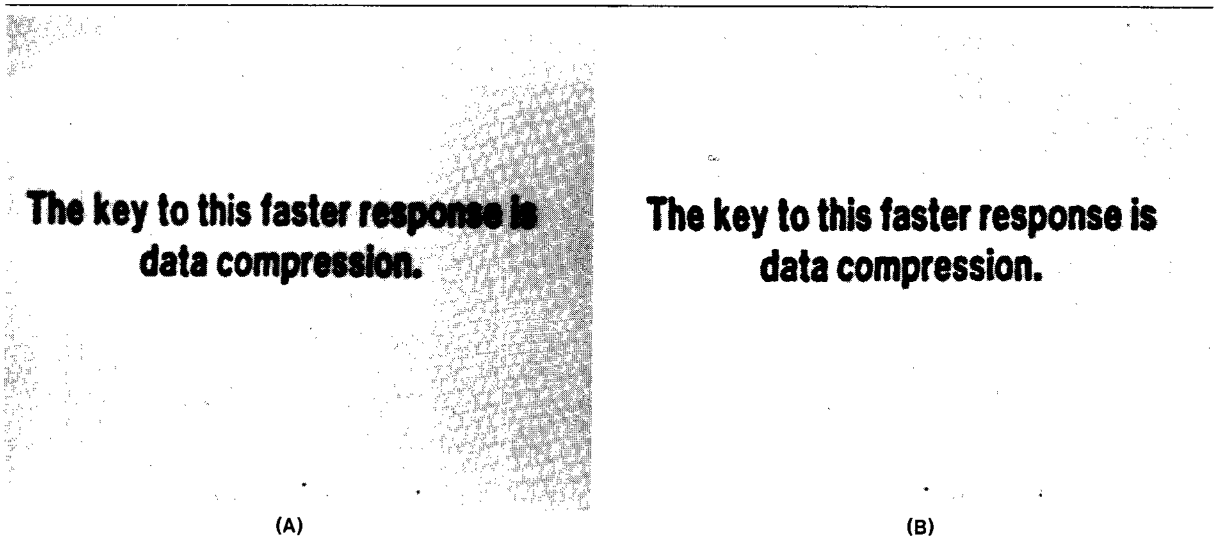
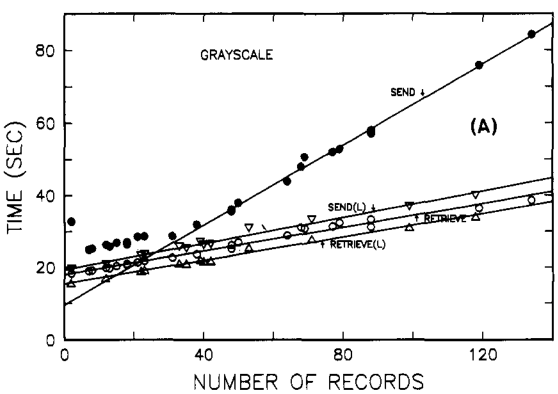
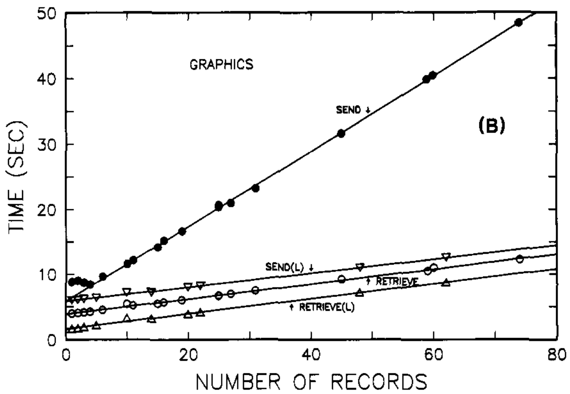


Figure 8 (A) System performance timing graph for gray scale; (B) System performance timing graph for graphics





available to display presentation material. If a third location is included in the conference, two monitors can be used for "people pictures" from the two remote sites and the remaining monitor can be used to display graphics material. The classification of monitors is a parameter that is determined by the room administrator when initializing the conference.

Most users find the ten-second response times associated with the transmission of business graphics to be very acceptable, and the use of low-bandwidth, dial-up lines results in a significant reduction in the transmission costs.

To set up a conference, a "host" location dials a remote location and passes conference parameters to it. Since the passing of parameters is handled by room administrators, the parameters are open to negotiation. If a third location is to be included in the conference, the second location administrator dials the third location, and repeats the parameter-passing and negotiation process.

Initially, the host location is in control of the conference. A picture can be sent by a location only when it is in control. If the SEND button is pressed at a location that is not in control, a request for control is sent to the controlling location. If that location is not sending a picture, it passes control to the requesting location.

By conference etiquette, the host sends a "people picture" to the other location(s), and the other location(s) send a "people picture" back. Additional "people pictures" may be sent at any time during a conference, but this is usually done only when persons enter or leave the room.

In a two-location conference, encoded data messages are sent directly from one location to the other. In a three-location conference, the middle location is the only location in direct contact with the other two locations. Therefore, when pictures are being sent by the middle location, the encoded data messages are sent directly to both of the other locations. When pictures are being sent from one of the end locations, the encoded data messages are sent to the middle location. The middle location then retransmits the message to the other end and retains a copy of the message for processing.

The *transmission process* is as follows. As soon as a segment of encoded data is generated, it is sent to the other location(s), if the communication line is available. If the line is not available (busy), the message is queued for transmission when the line becomes available. The encoding of the image continues, and segments of encoded data are queued until the line becomes available or until the painting of the image at the sending location becomes too far out of synchronization with the receiving location(s) to be useful. When the communications line becomes available, messages that have been queued are sent. Up to four messages are blocked and sent as one transmission. This blocking of messages increases the efficiency of the communication line.

Concluding remarks

In the development of the Series/1-based videoconferencing system presented in this paper, strong emphasis has been placed on three objectives: (1) producing image quality that is consistent with high-quality National Television Standards Committee (NTSC) format video equipment; (2) obtaining good human factors through a responsive yet flexible system; and (3) operating at low cost through the use of standard dial-up telephone line communication channels. Most users find the ten-second response times associated with the transmission of business graphics to be very acceptable, and the use of low-bandwidth, dial-up lines results in a significant reduction in the transmission costs.

This system is already operational in many IBM locations and continues to grow in both number and function. There are also technological trends that are expected to exert a major influence on videoconferencing. The video industry is evolving rapidly toward higher resolution and quality.[11] Satellites and fiber optics are expected to help reduce communication costs. The capabilities of solid state technologies continue to improve each year. At the same time, hardware costs are being reduced. These factors are inducing a trend toward higher levels of function in future videoconferencing systems.

Acknowledgments

We would like to acknowledge the help and influence that the following persons have had on the Series/1-based videoconferencing system. The work of G. Goertzel, R. Henshaw, and K. Anderson, of Research, Yorktown, on compilers and code development tools was essential to the implementation of the compression algorithms. W. Hampel, of CPD, Raleigh, was responsible for several aspects of the system control code, and E. Kubiliun, of SPD, Boca Raton, assisted in writing some of the early code. D. Olson, of NMD, Atlanta, oversaw this project, and R. Yellowlees, formerly of NMD, Atlanta, consistently encouraged and supported this effort. L. Leyden and H. Hagopian, of Corporate Communications, White Plains, pioneered teleconferencing in IBM and have continued with improvements in room design and site coordination. D. Lasher, W. Stelzenmuller, R. Greenhalgh, and C. Kunzinger, of CPD, Raleigh, developed and installed compression algorithms in an earlier IBM teleconferencing system. The authors also acknowledge the room installation and maintenance team, room administrators, site coordinators, and the many others who contributed to making this system a reality.

Cited references

1. L. S. Randall, "The shared graphic workspace: Interactive data sharing in a teleconference environment," *Proceedings, Twenty-fifth IEEE Computer Society International Conference*, September 20–23, 1982, Washington, DC, pp. 535–542.
2. C. W. Kelly, "An enhanced presence video teleconferencing system," *Proceedings, Twenty-fifth IEEE Computer Society International Conference*, September 20–23, 1982, Washington, DC, pp. 544–551.
3. T. B. Cross, "Computer teleconferencing—Being there without going there," *Telecommunications* **16,** No. 5, 73–78 (May 1982).
4. D. Lasher, W. Stelzenmuller, R. Greenhalgh, and C. Kunzinger, IBM Communication Products Division, Raleigh, NC, private communication.
5. A. N. Netravali and J. O. Limb, "Picture coding: A review," *Proceedings of the IEEE* **68,** No. 3, 366–406 (March 1980).
6. H. Mussmann and W. Erdmann, German Patent No. P2740945.6-1977.
7. J. Rissanen and G. G. Langdon, Jr., "Universal modeling and coding," *IEEE Transactions on Information Theory* **IT-27,** No. 1, 12–23 (January 1981).
8. B. G. Haskell, "Entropy measurements for nonadaptive and adaptive, frame-to-frame, linear-predictive coding of videotelephone signals," *The Bell System Technical Journal* **54,** No. 6, 1155–1174 (July–August 1975).
9. R. E. Graham, "Snow removal—A noise-stripping process for picture signals," *IRE Transactions on Information Theory* **IT-8,** 129–144 (1962).
10. R. Hunter and A. H. Robinson, "International digital facsimile coding standards," *Proceedings of the IEEE* **68,** No. 7, 854–867 (July 1980).
11. N. Mokhoff, "A step toward 'perfect' resolution," *IEEE Spectrum* **18,** No. 7, 56–58 (July 1981).

General references

D. Anastassiou, J. L. Mitchell, W. D. Pennebaker, and K. S. Pennington, "A new freeze-frame teleconferencing system," *Proceedings, Twenty-fifth IEEE Computer Society International Conference*, September 20–23, 1982, Washington, DC, p. 543.

"Special report—Teleconferencing today," *Communication News*, 43–68 (February 1982).

C. E. White, "Meeting business telecom needs," *Telecommunications* **16,** No. 10, 25–40 (September 1982).

T. B. Cross, "Teleconferencing can reduce need to travel," *The Office* **95,** No. 4, 100 and 115 (April 1982).

J. E. Thompson, "Visual services trial—The British telecom system for teleconferencing and new visual services," *British Telecommunications Engineering* **1,** 28–34 (April 1982).

H. Kaneko and T. Ishiguro, "Digital television transmission using bandwidth compression techniques," *IEEE Communications Magazine* **18,** 14–22 (July 1980).

M. Kikuchi and M. Yoshikawa, "New video conference system," *Japan Telecommunications Review* **22,** No. 2, 112–119 (April 1980).

T. Omachi, Y. Yakashima, T. Usubuchi, T. Sakabe, and H. Iijima, "High resolution image system for teleconferencing," *Proceedings, IEEE Computer Society International Conference,* June 1982, Philadelphia, PA, pp. 26.3.1–26.3.5.

M. Turoff and S. R. Hiltz, "Computer support for group versus individual decisions," *IEEE Transactions on Communications* **COM-30,** No. 1, 82–91 (1982).

S. R. Hiltz and M. Turoff, "The evolution of user behavior in a computerized conferencing system," *Communications of the ACM* **24,** No. 11, 737–751 (November 1981).

Reprint Order No. G321-5185.

Dimitris Anastassiou *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598.* Dr. Anastassiou received the Diploma in electrical engineering from the National Technical University of Athens, Greece, in 1974. He then received the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, in 1975 and 1979, respectively. Since 1978, he has been working at the Research Center as a research staff member, doing work in various aspects of image processing and coding. In 1982, he received an Outstanding Innovation Award for his contributions to the development of the IBM videoconferencing system.

Marvin K. Brown *IBM National Accounts Division, P.O. Box 1328, Boca Raton, Florida 33432.* Mr. Brown is a senior engineer in Industrial Automation Product Programs (IAPP). He joined IBM at Endicott in 1968 as part of the machine technology group engaged in the electromagnetic compatibility of the System/3. In 1969, he transferred to the machine technology group in Boca Raton to start and continue a similar group until joining the Sensor Based Installation Services (SBIS) in 1973. In the SBIS group, now called IAPP, he designed and coordinated the installation of customer application hardware for the System/7 and Series/1. Since 1980, he has had the hardware development responsibilities for the IBM internal videoconferencing system. Mr. Brown has a B.S. degree in Bible from Philadelphia College of Bible, a B.S. degree in electrical engineering from Purdue University, and an M.B.A. degree from Florida Atlantic University. He is registered as a professional engineer in the State of California.

Hugh C. Jones *IBM National Marketing Division, P.O. Box 2150, Atlanta, Georgia 30055.* Mr. Jones is currently a senior marketing support representative. Since joining IBM in 1963, he has held programming development positions in the Federal Systems Division, Field Engineering Division, General Systems Division, and National Marketing Division. Most of his work for the past five years has been with the Series/1 and the videoconferencing system. Mr. Jones has a B.A. in mathematics from Hastings College, Hastings, Nebraska.

Joan L. Mitchell *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598.* Dr. Mitchell graduated from Stanford University with a B.S. in physics in 1969. She received her M.S. and Ph.D. degrees in physics from the University of Illinois at Urbana-Champaign in 1971 and 1974, respectively. She joined the exploratory printing technologies group at the Thomas J. Watson Research Center immediately after completing her Ph.D. Dr. Mitchell has a patent in ultrasonic printing and a Research Division Award for her contributions to the origin and development of a new nonimpact printing technology. Since 1976, she has worked in the field of data compression. She received an Outstanding Innovation Award for two-dimensional data compression in 1978. The data compression scheme she developed with a colleague was proposed by IBM to the CCITT as a facsimile data compression standard. Dr. Mitchell also received an Outstanding Innovation Award for her contribution to the IBM videoconferencing system in 1982.

William B. Pennebaker *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598.* For the past few years Dr. Pennebaker has been managing a group doing research in areas related to image acquisition, processing, and compression. He joined the IBM Research Division in 1962, and has worked in the areas of low-temperature physics, thin films, display technology, printing technology, and image processing. He received an Outstanding Contribution Award for work on strontium titanate films, an Outstanding Invention Award for work on silicon nitride films, and an Outstanding Innovation Award for his contribution to the IBM videoconferencing system. Dr. Pennebaker obtained his B.S. in engineering physics from Lehigh University in 1957, and his Ph.D. in physics from Rutgers in 1962. He is a member of AIP, IEEE, SID, and AAAS.

Keith S. Pennington *IBM Research Division, Thomas J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598.* Dr. Pennington received a B.Sc. in physics from Birmingham University, England, in 1957, and a Ph.D. in physics from McMaster University, Canada, in 1961. He started his career in research at the Bell Telephone Laboratories, Murray Hill, NJ, in the field of holography. He developed multicolor holograms, and did early work in holographic interferometry and optical information processing. In 1967 Dr. Pennington joined the IBM Research Division, where he made several contributions to the development of improved holographic materials and techniques for three-dimensional screen analysis. He was appointed manager of exploratory terminal technologies in 1972, and, in this position, contributed to the development of new printing techniques. In 1979, Dr. Pennington was promoted to his present position in image technologies, where he has responsibility for several noncoded information projects, including projects related to teleconferencing, document scanning, and new printing processes. Dr. Pennington has written three book chapters related to holography and optical information processing. During 1971–1972, he served as participant and group leader for the National Academy of Sciences, Undersea Warfare Committee. He has received two Outstanding Contribution Awards and two Outstanding Innovation Awards. Dr. Pennington is the author of more than twenty-five patents.