Voice-bandwidth visual communication through logmaps: The Telecortex*

Richard S. Wallace
Courant Institute of Mathematical Sciences
New York University
New York, NY

Benjamin B. Bederson
Bell Communications Research
Morristown, NJ

Eric L. Schwartz
Cognitive and Neural Systems
Boston University
Boston, MA

Abstract

We present a robotic video telephone application of the Cortex-1 miniaturized space-variant active vision system. The embedded processor architecture of Cortex-1 enables it to implement a variety of functions not found in conventional video telephones, for example the camera tracks moving users with its pantilt mechanism. We also report an analog channel coding scheme to transmit logmap video images through band-limited analog channels such as the public switched telephone network (PSTN). The transmitter divides the voice frequency band into 768 channels, and modulates two values in quadrature on each channel. Some channels are reserved for special calibration signals enabling the receiver to recover both the phase and magnitude of the transmitted signal. The remaining channels carry pixel intensities. We synthesize the signal in the frequency domain and run the FFT algorithm to implement a fast conversion to a real, time-domain signal. A phase-lock loop keeps the receiver frame-synchronized with the transmitter. We constructed an experimental video telephone that sends 1376 pixel logmap images at 3.9 frames per second through the PSTN. Using the analog channel coding scheme, we achieve an effective data transfer rate in excess of 40000 bits per second.

*This work supported in part by DARPA Contract #N00014-90-C-0049, and AFOSR Contract #88-0275. An earlier version of this report was published in the *Proceedings of the IEEE Workshop on Applications of Computer Vision*, Palm Springs, CA, November, 1992. Please address correspondence to Richard S. Wallace, Courant Institute of Mathematical Sciences, New York University, 251 Mercer St., New York, NY 10003.

1 Introduction

We report the development of a video telephone application built on a miniaturized active vision system, Cortex-1 [3]. The vision system combines a logmap camera sensor, high-speed pantilt, microprocessors, a microcontroller, a telephone interface and a display. The logmap camera is space-variant, having only 1376 pixels organized in a logarithmic geometry, much like the human visual system. This radical reduction in pixel count makes possible the transmission of logmap images through band-limited analog channels, such as the Public Switched Telephone Network (PSTN). Our pantilt device, the Spherical Pointing Motor (SPM), nearly matches the performance of the human eye in speed and saccadic rate. The embedded processors in Cortex-1 faciltate the development of machine vision algorithms to control the SPM, resulting in a programmable robotic video telephone. We call the video phone application *Telecortex*.

Relying on its embedded processor power, the Telecortex overcomes one of the frequently cited obstancles to consumer acceptance of video telephones, namely, that the user remain fixed in one place [8] [9] . We have implemented a simple real-time motion-tracking program so that the camera sensor can follow moving objects, such as a video phone user walking around. We have also experimented with more sophisticated tracking algorithms, and expect the system to run pattern recognition algorithms such as face and gesture recognition, scene identification, and attentional algorithms.

1.1 Background

The history of video telephone technology and the development of the video teleconferencing industry is an important and interesting topic, but outside the scope of this paper. The International Teleconferencing Association [1] publishes a list of video teleconferencing vendors.

At least one telephone company, NTT, has embarked on a project to develop an *intelligent video codec*, a system that combines a computer vision system at the transmitter with a computer graphics system at the receiver [1] [11] . Their idea is to reduce the channel bit rate by sending high-level symbolic scene descriptions rather that images per se. Our goal was much less ambitious: to demonstrate that a low-cost machine vision system can function as a video telephone, compatible with the existing analog voice network, by relying on image processing and the logmap geometry to benefit the human interface.

2 A miniaturized active vision system

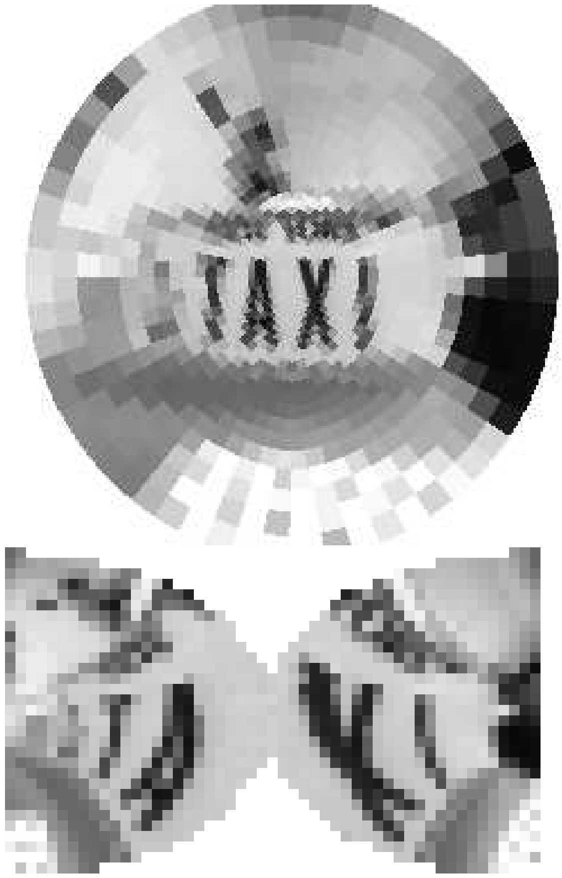
The video telephone is one application of our miniaturized active vision system, Cortex-1. This system consists of the emulated logmap sensor (which will be replaced by a custom VLSI sensor in the near future), a miniature pan-tilt mechanism, controller, general purpose processors, and display. The controller consists of a camera driver, a 2 MIPS programmable microcontroller (Motorola MC68332), a video display driver and up to 3 12MIPS digital signal processors (Analog Devices AD2101). The actuator and camera are mounted to the chassis ($14 \times 22 \times 22$ cm) and connected by twisted-pair cables. The prototype is powered from a standard 110 Volt AC line, but uses less than 25 watts and could be battery powered.

2.1 Camera sensor

The camera sensor consists of a miniature commercially available CCD and a custom lens assembly mounted in the SPM. The CCD image is converted to a logmap image containing 1376 pixels. Its maximum central resolution is 0.175 degrees per pixel and its horizontal field of view measures 33^o. The sensor has a fixed focus 4mm lens with a manually changeable aperture (3 sizes). Imaged objects more than 40mm away from the lens are in focus. The system outputs up to 32 frames per second and measures 256 gray levels per pixel. The camera head (CCD and lens assembly) measures only $8 \times 8 \times 10$mm.

One DSP runs the sensor emulation and forms the logmap image by averaging sets of CCD pixels. The

[1] ITA, Washington, D.C., 202-833-2549



a

b

Figure 1: (a) The inverse logmap image and (b) The forward logmap image. We enlarged the printed pixels in the logmap (b).

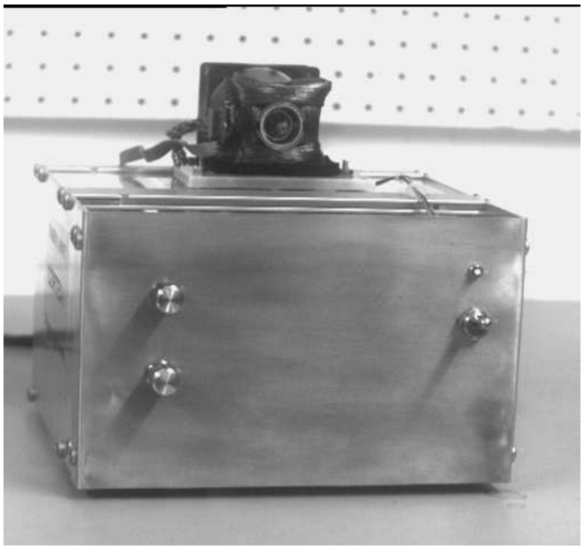


Figure 2: The video telephone is one application of our miniaturized active vision system, Cortex-1, shown here.

Figure 3: The Spherical Pointing Motor and camera sensor. The camera head (CCD and lens assembly) measures only $8 \times 8 \times 10$mm. The SPM takes up a volume $4 \times 5 \times 6$cm.

interface between the CCD and the DSP is a camera driver board which provides timing signals to the CCD and converts analog brightness levels to 8-bit digital data for the processors.

Another DSP board creates the video display signal, a standard RS-170 monochrome signal compatible with consumer TV sets in the U.S. and Japan. The third DSP implements the telephone interface.

2.2 Pan-tilt actuator

The Spherical Pointing Motor (SPM) is a novel pan-tilt actuator using three orthogonal motor windings to achieve open-loop pan-tilt actuation of the camera sensor in a small, low-power package. The SPM can orient the sensor through approximately 60^o pan and tilt, at speeds up to 1000 degrees per second. It measures $4 \times 5 \times 6$cm and weighs 170 grams. The control currents in the three motor windings vary under pulse width modulation control. We reported details of the SPM in two other papers [2,4].

3 The Telecortex

The Telecortex implements an analog channel coding technique to transmit the logmap images at about 4 frames per second. The radical reduction in pixel count achieved by the logmap sensor enables us to transmit through phone lines without any digital compression, save the compression implemenetd by the logmap itself.

The channel coding scheme divides the frequency band into 768 channels, and modulates two values in quadrature on each channel. Some channels are reserved for special calibration signals enabling the receiver to recover both the phase and magnitude of the transmitted signal. Although the technique bears some resemblence to frequency-division multiplexing (FDM) of quadrature amplitude modulation (QAM) digital channel coding [7] [10] [5] , our scheme modulates pixel values as the magnitudes of analog waves, relying the human user to implement error correction in his or her own visual system. Modem designers have demonstrated that FDM-QAM achieves the highest channel rates on the PSTN, but we have tried to push beyond this rate by allowing image pixel noise to vary with channel noise.

A PSTN line has a bandwidth of around $W =$ 3000Hz, a signal power level $P_{av} = 400$ times the noise level WN_0, (i.e. 26 dB). According to Shannon's expression, the channel capacity is

$$C = W \log(1 + \frac{P_{av}}{W N_0}) \quad (1)$$

[10], or 25942 bits per second. Other sources claim bit rates up to 49000, but even the best adaptive FDM-QAM modems in existence today achieve a bit rate of around 15000 per second, and only under optimum noise conditions. Modems need much of the excess capacity for error-detection and correction processes. FDM-QAM modems use adaptive algorithms that adjust C according to the signal-to-noise ratio on the line. Thus, under a digital channel code the frame rate (pixels per second) varies according to the SNR. With our analog transmission scheme the frame rate is always constant. What improves with SNR is the *picture quality.*

The Telcortex consists of a transmitter and receiver pair. The transmitter constructs a quadrature signal in the frequency domain, transforms the sigal to a real, time-domain signal, then modulates the signal onto the phone line through a codec. The receiver digitizes a distorted version of the time-domain signal, transforms it back to the frequency domain, and recovers the pixel brightnesses. The details of these functions are the topic of the next two subsections.

3.1 Transmitter functions

The transmitter constructs an analog signal in the frequency domain, then applies a fast Fourier transform (FFT) to convert the image signal to create a time domain signal. The transmitter divides the frequency band into 768 channels, and modulates two values in quadrature on each channel. Every 16th channel carries a special calibration signal enabling the receiver to recover both the phase and magnitude of the transmitted signal (see figure 5). The remaining channels carry the 8-bit pixel intensities. But before mapping pixel values to channels, the transmitter removes the image DC component and, becuase we chose calibration values that sum to zero, the resultant signal maximizes the signal-to-noise of the time domain waveform. One extra channel carries the image DC component. With the signal synthesized in the frequency domain, the FFT algorithm implements a fast conversion to a real, time-domain signal.

Then, it converts the time signal to voltage through an analog-to-digital converter (ADC). The ADC is coupled to the phone line through a Cermetek CH1817 telephone interface module.

The transmitter implements a sequence of these steps every 0.26 seconds:

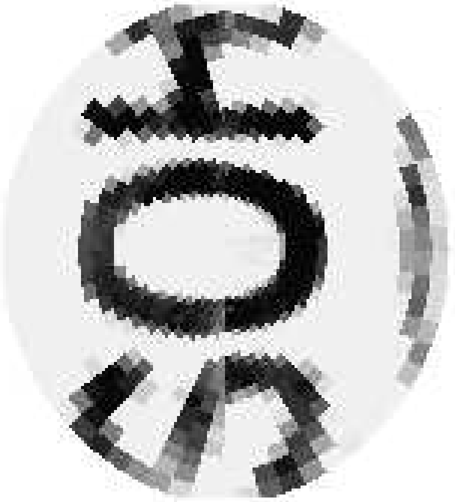


Figure 4: We use this logmap image to explain our analog channel coding technique.

1. Map pixels to frequencies and set calibration channels.
2. Make the frequency domain signal Hermetian.
3. Execute the fast Fourier transform.
4. Amplify the (real) signal in software.
5. Convert to mu-law and play.

The play operation actually overlaps with the other steps. Because the sampling rate of our ADC is only 8KHz, compared with a processor speed of 12MHz, the transmitter implements the play operation with an interrupt routine activated by a timer every 1.25×10^{-4} second.

3.2 Receiver functions

The receiver regenerates the picture signal by detecting the phase an amplitude distortion of the calibration channels, and calculating a set of complex constants to multiply with the received calibration values to obtain their original values. We assume that these constants change little between calibration channels,

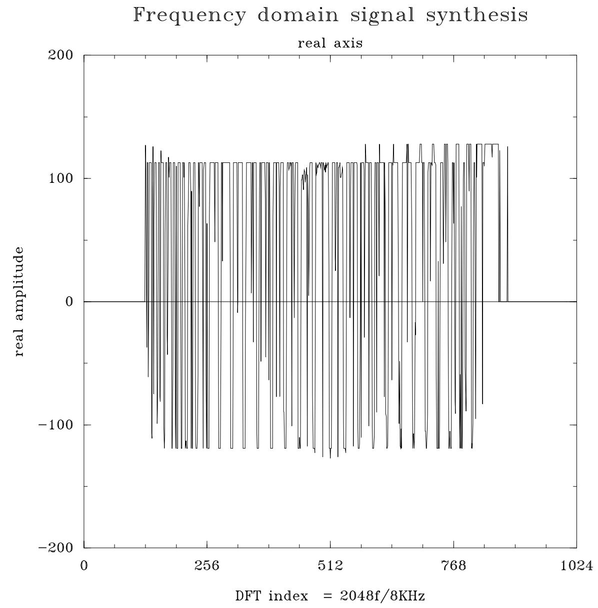


Figure 5: Frequency domain signal synthesis: real axis. The set of signals forming a sinusoidal pattern are the calibration channels. The nearly constant high and low values are the pure black and white pixels in the test image (previous figure). The sinusoidal pattern is formed by the calibration channels.

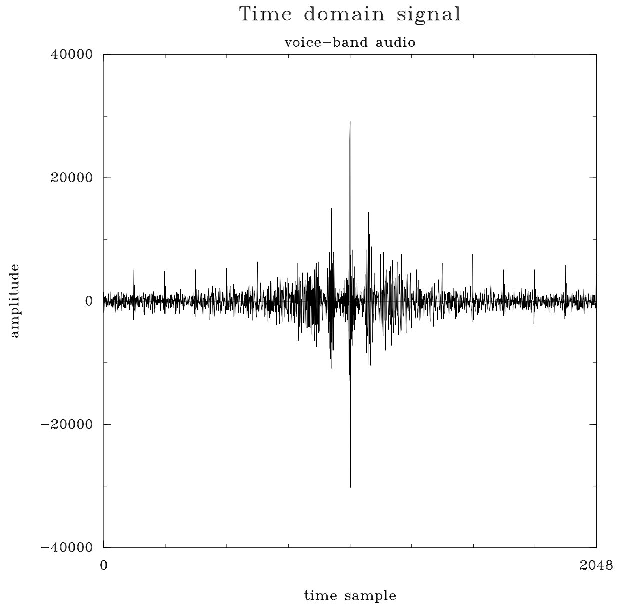


Figure 6: (Real) time domain signal transmitted

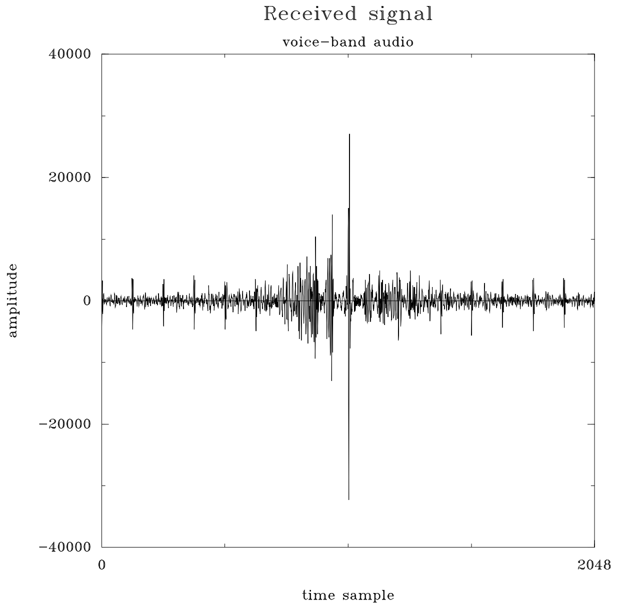


Figure 7: (Real) time domain signal received

and so we find their approximate value by linear interpolation. Figure 9 illustrates the interpolated calibration correction factors.

The receiver implements these functions every 0.26 seconds:

1. Frame-synchronize to transmitter.
2. Record the time-domain signal.
3. Execute the fast Fourier transform.
4. Solve for calibration constants and interpolate.
5. Recover the frequency domain signal.
6. Map frequencies to pixels and copy to display.

Frame synchronization is implemented by detecting the large spike in the image signal that arises from the DC component of the image itself (see figures 6 and 7). In pracitice we remove the DC component from the image before the frequency mapping, but we transmit a few seconds of high DC images in order for the receiver to frame-synchronize to the transmitter. One frame synchrony is achieved, the receiver follows the transmitter by phase-locking to one calibration channel.

Although we have developed a unique analog coding scheme well-suited to the noisy voice telephone channel, this scheme is by no means the only possible

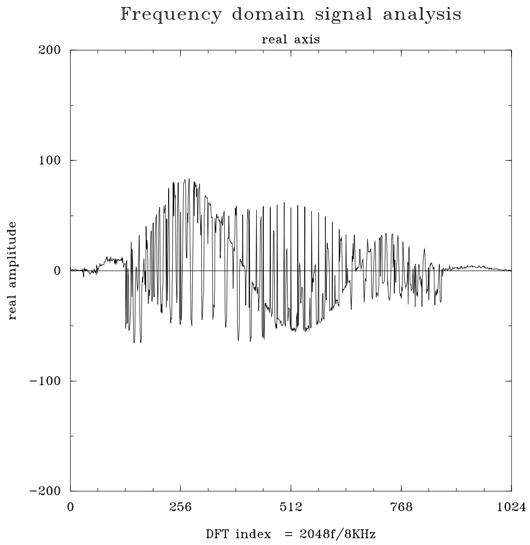


Figure 8: Frequency domain signal analysis: real axis. The received signal is distorted by time delay, voiceband attenuation, and other random noise.

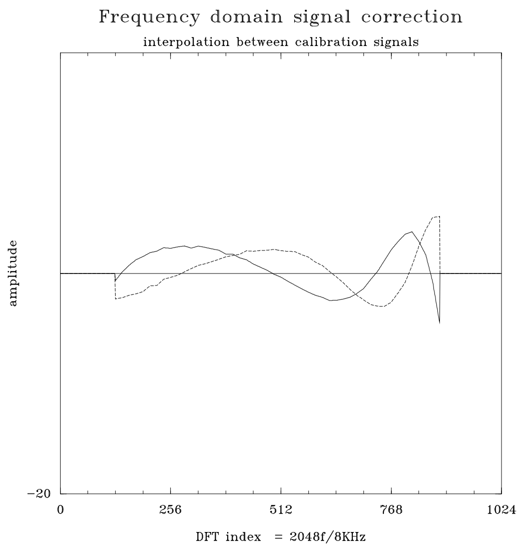


Figure 9: Frequency domain signal analysis: interpolated correction factors, real and imaginary axes overlaid.

one. But our telephone interface board contains all the electronics necessary to implement a digital modem. The DSP can run software to emulate a variety of modems.

4 Experiments

We built several versions of the Telecortex transmitters and receivers. The first version used a pair of Sun Sparcstation 1's to test our analog channel coding scheme. We used the Sun's audio port to play and record the analog image signal. At first we simply wired the audio out port of one Sun to the audio in port of the other. The transmitting Sun derived the logmap image from a digitized video image by executing the logmap transform in software.

After demonstrating the feasibility of the logmap video telephone, we proceeded to build a transportable system. We developed the telephone transmitter as an optional peripheral to the Cortex-1 system. The transmitter board consists of an AD2101 DSP chip, RAM and ROM, telephone codec, telephone line interface, digital inputs and outputs and analog buffering circuitry. The transmitter received logmap images from the Cortex-1 sensor through a 2MHz serial connection.

The Telecortex receiver used only two boards from the Cortex-1: the phone interface board and video display board. We developed a second, more compact telephone-in video-out receiver combining the audio and video functions on a single dual-processor board. This prototype, measuring $9'' \times 6'' \times 2''$, is small enough to fit in a briefcase The receiver is compatible with all analog telephones (at least in the U.S.) and all consumer NTSC television monitors.

We tested the Telecortex over long distance telephone connections, between Chicago and New York, Los Angeles and New York, New York and central Florida, and between western Massachusettes and New York. We also executed numerous tests through the local New York Telephone network. Each test was a success in the sense that the receiver displayed recognizeable images of people's faces and simple objects like pens, nametags, license plates and photographs.

5 Conclusion and discussion

We have constructed a prototype video telephone based on our miniaturized active vision system. This prototype has at least three advantages over conventional digital video teleconferencing systems: (1) the

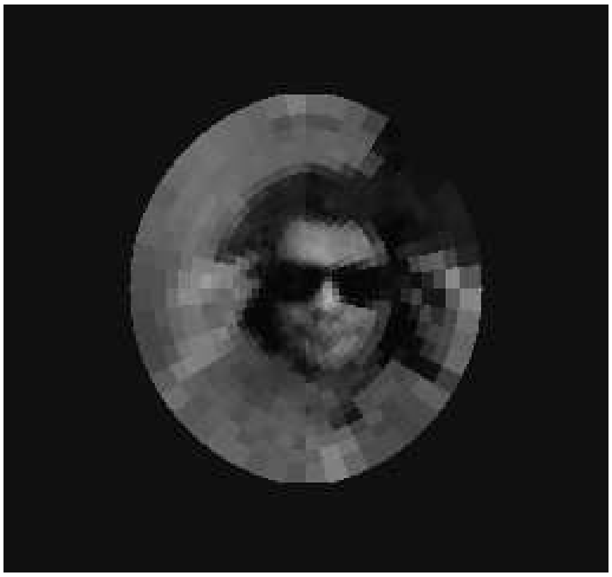


Figure 10: A logmap image transmitted in 0.26 seconds via a long-distance voice-grade telephone connection from Chicago to New York on August 5, 1991

Telecortex overcomes at least one major problem of consumer acceptance: that the user must remain in one place, (2) our analog channel coding technique maintains constant frame rate and works well over a variety of noisy analog voice channels, and (3) except for the SPM, the prototype is constructed of only low-cost commodity parts, resulting in a very inexpensive video phone system.

Our prototype contains no digital compression functions other than the compression afforded by the logmap itself. We hypothesize that standard digital image compression does not commute with the logmap transform, in the sense that the logmap represents a reduction by a factor of 30 compared with the TV image having the same field of view, but another factor of, say, 10 is not achievable using standard compression techniques. In some sense, solving the attention problem for the active vision transmitter means that each image should contain a high information content, or in other words, little intra-frame redundacny.

Motion compression is also not likely to reduce the bit rate of an active logmap video communication system. Unlike fixed video sensors, the active vision system's sensor constantly pans and tilts, so there is little pixel coherence from one frame to the next.

As the technologies of eye tracking [14] [12] , space-variant displays [13] [6] , and digital communications advance, there will be little point to transmitting uniform resolution images, because of the resolution wasted in peripheral pixels. Future video communications systems will use logmap format to conserve bandwidth by taking advantage of the space-variant character of the human visual system [2]

References

[1] Takaaki Akiomoto, Richard S. Wallace, and Yasuhito Suenaga.
Automatic creation of face model for human image from front and side views.
In *Proceedings of Imagecon 90*, Bordeaux, France, November 1990.

[2] Benjamin B. Bederson, Richard S. Wallace, and Eric L. Schwartz.
A miniature pan-tilt actuator: The spherical pointing motor.
Technical Report 264, Courant Institute, NYU, April 1992.

[3] Benjamin B. Bederson, Richard S. Wallace, and Eric L. Schwartz.
A miniaturized active vision system.
In *11th International Conference on Pattern Recognition*, August 1992.

[4] Benjamin B. Bederson, Richard S. Wallace, and Eric L. Schwartz.
Two miniature pantilt devices.
In *IEEE International Conference on Robotics and Automation*, May 1992.

[5] John A. C. Bingham.
The Theory and Practics of Modem Design.
Wiley-Interscience, 1988.

[6] Mark S. Franzblau.
Log mapped focal driven procedural rendering: an application of the complex log map to ray tracing.
Master's thesis, Courant Institute, New York University, 1991.

[7] Batoro Hirosaki, Satoshi Hasegawa, and Akio Sabato.
Advanced groupband data modem using orthoganally multiplexed qam technique.
IEEE Transactions on Communications, COM-34(6), June 1986.

[8] A. Michael Noll.
Teleconferencing target market.
IMR, 2(2), 1986.

[2] The authors are grateful for the assistance of Max Chien and Gil Engel in building and debugging the Telecortex system.

[9] A. Michael Noll.
The broadbandwagon.
Telecommunications Policy, September 1989.

[10] John G. Proakis.
Digital Communications.
McGraw-Hill, second edition, 1989.

[11] Tsutomu Sasaki, Richard S. Wallace, and Yasuhito Suenaga.
Discussion on a face recognition system based on image segmentation and labeling algorithm and pattern classification method.
In *Proceedings of 1990 Electronics, Information and Communications Society Fall National Conference*, 1990.
in Japanese.

[12] Akira Tonomo, Muneo Iida, and Yukio Kobayashi.
A tv camera system which extracts feature points for non-contact eye movement detection.
SPIE Optics, Illumination and Image Sensing for machine vision, 1194, 1989.

[13] Hiroyuki Yamaguchi, Muneo Iida, Akira Tonomo, and Fumio Kishino.
Picture quality of a large field visual field display with selective high resolution in foveal vision region.
ITEJ Technical Report, 14(12), 1990.
in Japanese.

[14] Hiroyuki Yamaguchi, Akira Tonomo, and Yukio Kobayashi.
Proposal for a large field visual display employing eye movement tracking.
SPIE Optics, Illumination and Image Sensing for machine vision, 1194, 1989.