The 3D Room: Digitizing Time-Varying 3D Events by Synchronized Multiple Video Streams

Takeo Kanade, Hideo Saito, Sundar Vedula

CMU-RI-TR-98-34

December 28, 1998

The Robotics Institute

Carnegie Mellon University

Pittsburgh, Pennsylvania 15213-3890 USA

I. INTRODUCTION

The "3D room" is a facility for 4D digitization - capturing and modeling a real time-varying event into a computer as 3D representations which depend on time (1D). On the walls and ceiling of the room, a large number of cameras (currently 49) are mounted, all of which are synchronized with a common signal. A PC-cluster computer system (currently 17) can digitize all the video signals from the cameras simultaneously in real time as uncompressed and lossless color images of full frame rate (640x480x2x30 bytes per second). The images thus captured are used for research on Virtualized Reality™[1-6]. This digital 3D Room is a natural outgrowth of our previous 3D Dome [1,2], which was built in 1994 and has been used for a similar purpose, but was based on analog VCRs and thus limited to offline applications. This document describes the current 3D Room as of December 1998 – design, components and capabilities - built at Robotics Institute, Carnegie Mellon University (CMU).

II. THE CMU 3D ROOM

The CMU 3D Room is 20 feet (L) x 20 feet (W) x 9 feet (H). As shown in figure 1, 49 cameras are distributed inside the room: 10 cameras are mounted on each of the four walls, and 9 cameras on the ceiling. Figure 2 shows a panoramic view of the 3D Room. Figure 3 shows an overview of the digitizing system. All of the 49 cameras are synchronized by a single synchronizing signal. Since every image frame needs to be labeled by the time frame, the VITC (Vertical Interval Time Code) is embedded onto every image frame. The S-Video output from each camera, consisting of two separate signal lines of intensity (Y) and color difference (C), is sent to a time code translator, so that the time code can be embedded onto intensity signal of the S-Video for the time frame labeling. The computing system consists of one Control PC and a cluster of 17 Digitizing PCs. Each Digitizing PC contains 3 digitizer cards, and can simultaneously digitize up to 3 video inputs. The Control PC controls the 17 digitizing PCs for coordinating overall setup and timing for the whole digitization process.

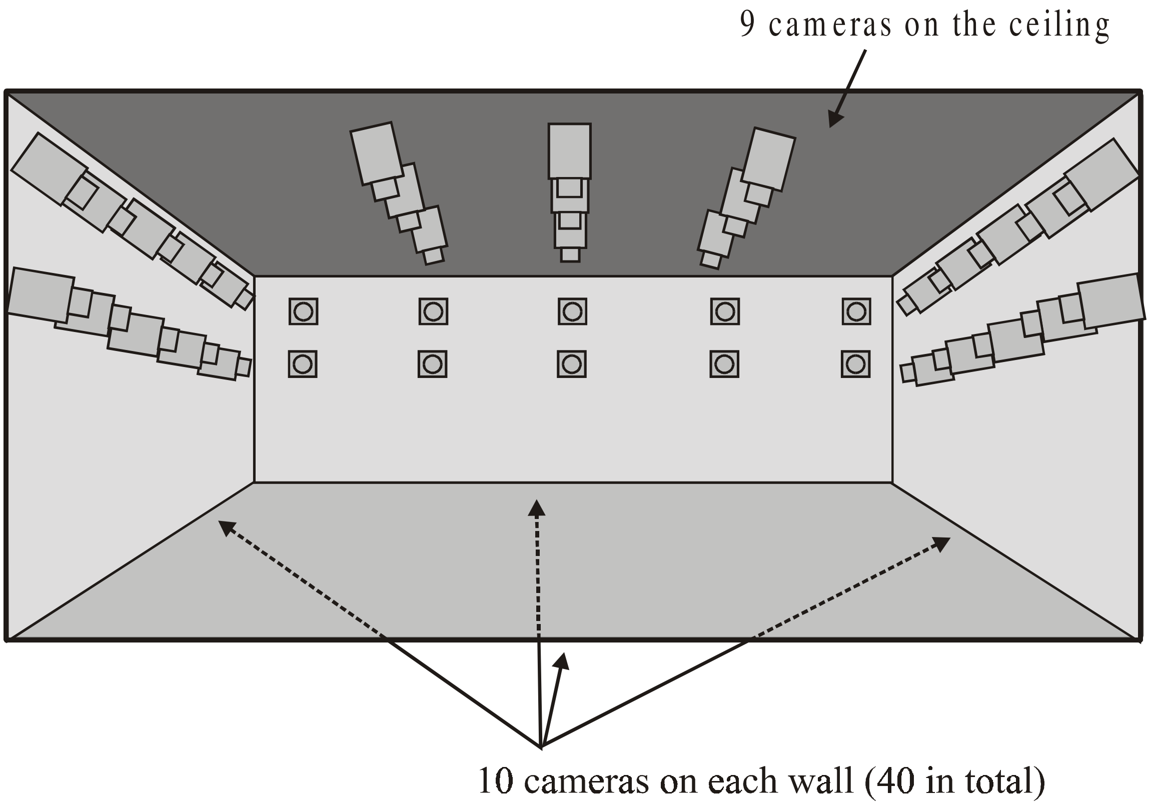

Figure 1 : Camera placement in the 3D Room.

Figure 2: Panoramic view of 3D Room in CMU.

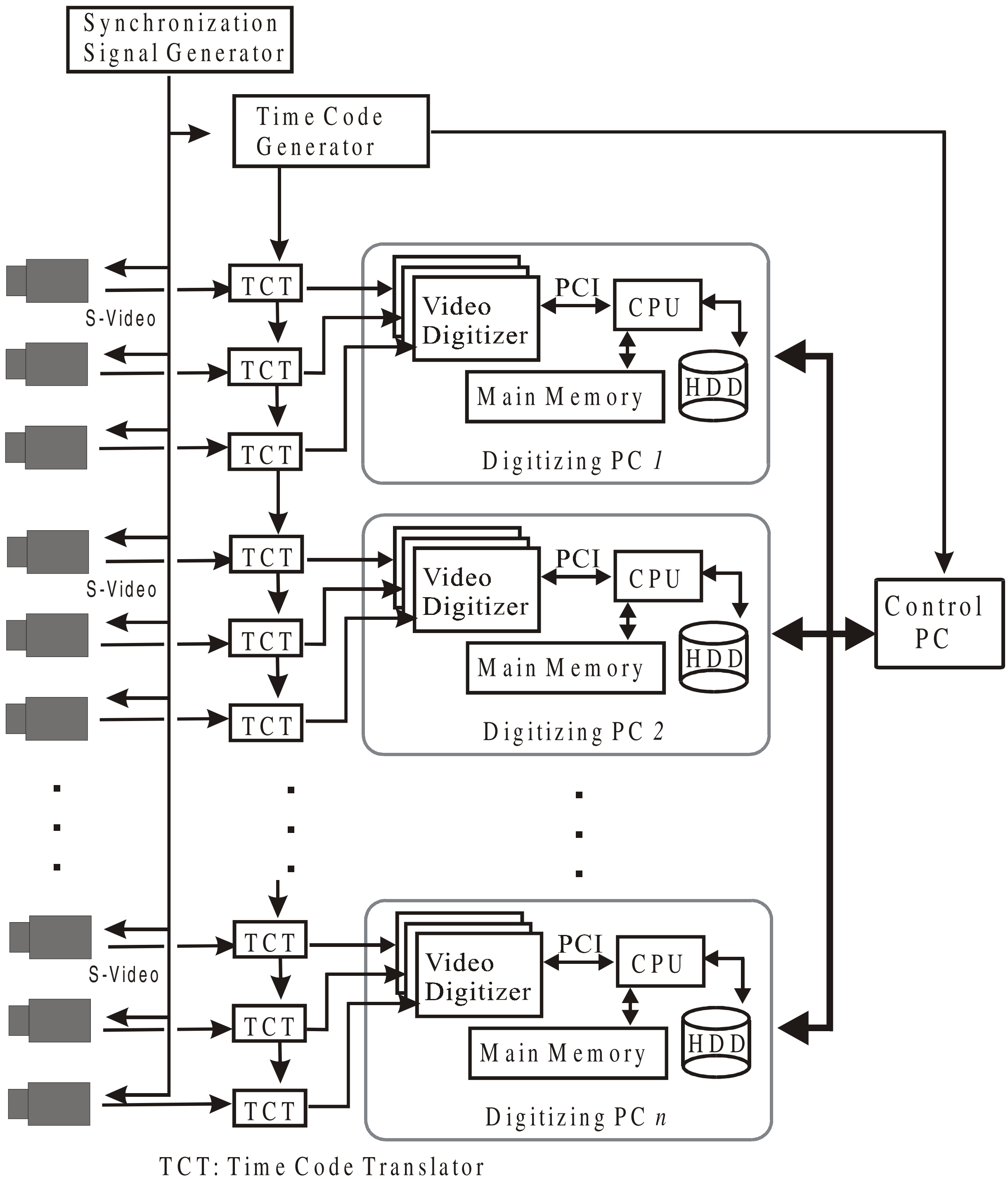

Figure 3: The digitization system of the 3D Room (ver.1, Dec. 98) consists of 49 synchronized cameras, one time code generator, 49 time code translators, 17 Digitizing PCs and one Control PC.

The detailed specifications of individual system components are as follows:

Camera

- Sony progressive scan 3CCD color camera DXC-9000 with 7.5mm~105mm zoom lens.
- JVC single CCD color camera TK-C1380U with lens of 6mm focal length.

VITC Unit Hardware

- Time code generator : HORITA LTC generator TG-50
- Time code translator : HORITA LTC-VITC Translator VG-50

LTC (Longitudinal Time Code)

SMPTE 80-bit longitudinal, recorded on an audio channel of video signal.

VITC (Vertical Interval Time Code)

SMPTE 90-bit, recorded onto two horizontal lines of the vertical interval of each video image field.

PC Hardware

- 266 MHz Intel Pentium II CPU
- Soyo-tek SY6BA 100 MHz 440BX Motherboard
- 512 Megabytes PC100 SDRAM
- 6.4 Gigabyte Western Digital Caviar Ultra IDE Hard Disk
- Imagenation PXC200 digitizer cards on PCI bus : Up to 3 cards on one PC

PC Software

- Microsoft Windows NT 4.0
- PXC200 Driver libraries from Imagenation Corporation
- Custom software for real-time digitization to memory and off-line writing to disk.

III. CAPABILITIES

A) Video Resolution

The analog video signal format is S-Video. The digital image format we use is YCrCb 4:2:2. The intensity image of Y is digitized at full resolution of 640x480 while color components, Cr and Cb, are digitized in half size of 320x480; thus on average, 2 bytes are used per pixel for representing the digitized image of YCrCb 4:2:2. Sampling color components at half resolution helps reduce the data rate of digitization, and is acceptable because human visual perception is not as sensitive to the spatial resolution of color components. Also, most vision algorithms use mainly intensity information for the purpose of matching and registration.

The data rate per video channel is

$$640 \times 480 \times 30 \text{ (fps)} \times 2 \text{ (bytes per pixel)} = 17.58 \text{ MBytes/sec.}$$

This is well under the PCI bus burst transfer rate of 132 MBytes/sec., and well over the 7 to 9 MBytes/sec transfer rate of a simple hard disk. Therefore as a simple and cost-effective solution we choose a method that uses real-time capturing into memory and off-line saving to the disks.

B) Capacity – Duration of Digitization

Each PC currently has 512 MBytes memory. The total memory area available for storing digitized images is about 480 MBytes after subtracting the system usage. The total number of images that can be stored in memory is therefore

$$480 \text{ MByte} / (640\text{x}480\text{x}2 \text{ Byte}) = 819 \text{ frames.}$$

If one PC handles digitization of three camera video outputs, up to 819/3 = 273 frames can be captured per camera at one time. This capacity corresponds to 9.1 seconds of duration at 30 full size frames per second. The digitization control program allows the user to choose smaller image formats or lower frame rates in order to extend the duration of digitization.

C) Interface

The control PC coordinates and monitors the whole digitization process, in addition to providing the user interface. Figure 4 shows the interface window, through which a user can specify the digitization parameters : frame rate, frame format, starting time, total duration, enabling/disabling of specific cameras, and so on. The current VITC time code is displayed on the window, so the user can easily specify the time code at which the digitization needs to begin. Clicking the “Grab” button sets the timer to start the digitization, and while waiting, the control PC beeps at an interval of 1 second, like an ordinary camera’s timer. Digitization starts at the specified time, and the beep turns to a continuous one till the end of digitization period. When the digitization is completed, each digitizing PC writes the image data out to disk, and the control PC collects information from the digitizing PCs to verify if the whole digitization was done as specified without any missing, or corrupted frames.

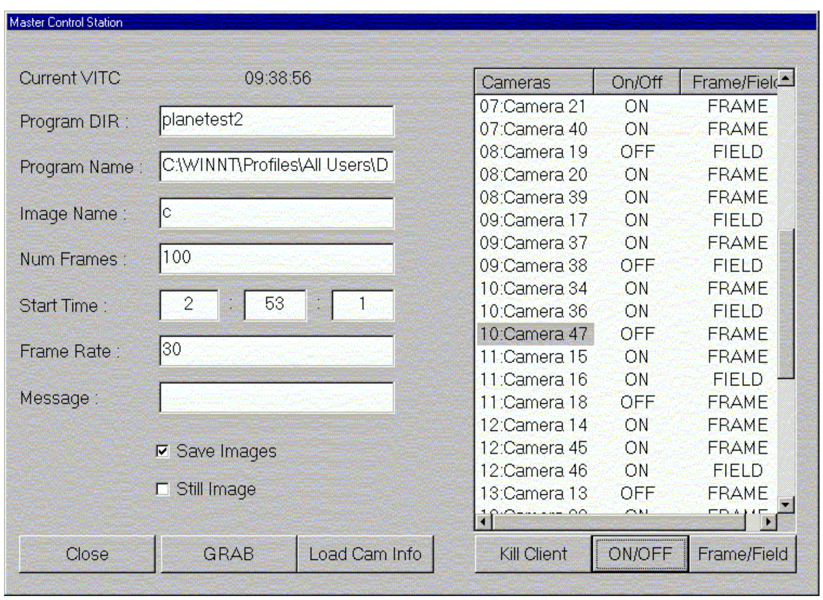

Figure 4 : Window of user interface for controlling digitization on all PCs. In this window, user can easily specify digitization parameters and different configuration for every camera.

D) Stability and Verification

The system performance has been verified by running 500 experiments, with each experiment involving a capture of 250 frames. On every experiment, the corrupt flag was checked, timecode was verified, both for consistency across all the cameras, and for regular increments between captured images. The stability of the system was thus satisfactorily verified.

IV. CONCLUSION

We have constructed the CMU 3D Room, a room with 49 cameras, whose output signals can be captured into a computer in real time as digital uncompressed, lossless, full frame images with color (640x480x2x30 bytes per second). The CMU 3D Room for Virtualized Reality [1-6] is a unique facility, and one of the first of its kind. This facility enables us to model a time-varying real 3D event into a computer, as is and in its entirely. The resulting models can be used for manipulating, altering and rendering the reality. We intend to make this facility and the data sets from it available for the vision research community in the near future.

ACKNOWLEDGEMENT

We thank Peter Rander, Makoto Kimura, Shigeyuki Baba, Ching-Kai Huang, and Peter Kioko for their help in the development of the CMU 3D Room. Intel Corporation, Sony Corporation and Matsushita Electric Industrial Company provided partial support for this project.

REFERENCES

[1] P.J. Narayanan, P. W. Rander, and T. Kanade, "Synchronous Capture of Image Sequences from Multiple Cameras", CMU-RI-TR-95-25, December 1995.

[2] T. Kanade, P. J. Narayanan, and P. W. Rander, "*Virtualized Reality*: Concepts and Early Results", *IEEE Workshop on the Representation of Virtual Scenes*, Boston, pp.69-76, June 1995.

[3] T. Kanade, P. W. Rander , and P. J. Narayanan, "*Virtualized Reality:* Constructing Virtual Worlds from Real Scenes", *IEEE Multimedia*, vol.4, no.1, pp.34-47, May 1997.

[4] P. W. Rander, P.J. Narayanan, and T. Kanade, "Recovery of Dynamic Scene Structure from Multiple Image Sequences", *Proc. IEEE Int'l. Conf. Multisensor Fusion and Integration for Intelligent Systems*, Washington D.C., pp.305-312, December 1996.

[5] P.J. Narayanan, P. W. Rander, and T. Kanade, "Constructing Virtual Worlds using Dense Stereo", *Proc.IEEE 6th Int'l. Conf. Computer Vision,* Bombay, pp.3-10, January 1998.

[6] S. Vedula, P.W. Rander, H. Saito, and T. Kanade, "Modeling, Combining, and Rendering Dynamic Real-World Events From Image Sequences", *Proc. 4th Conf. Virtual Systems and Multimedia*, Gifu Japan, vol.1, pp.326-332, November 1998.