Design and Implementation of the University of Maryland Keck Laboratory for the Analysis of Visual Movement

Ross G. Cutler,* Ramani Duraiswami, J. Hench Qian and Larry S. Davis
University of Maryland, College Park, Maryland
rcutler@microsoft.com, {ramani,qian,lsd}@umiacs.umd.edu

Abstract

The Keck Laboratory for the Analysis of Vision Motion is a state-of-the art multi-perspective imaging laboratory recently established at the University of Maryland. In this paper, we describe the design and architecture of the lab, that is currently being used to support many computer vision studies. In particular, we discuss: camera synchronization, image resolution analysis, image noise analysis, stereo error analysis, video capture, lighting, calibration hardware.

1 Introduction

A goal of many vision researchers has been to establish laboratories that enable multi-perspective viewing of a reasonably large area. Viewing a 3D scene from multiple viewpoints provides a strong set of constraining relations that enable one to achieve accurate reconstructions of structure and motion. Such facilities can support research in both the development of new algorithms for 3D structure/motion, but can also support application oriented research such as that in human motion capture, gait analysis, model development, and smart rooms.

An early laboratory of this kind was established at Carnegie Mellon University by Prof. T. Kanade. While this laboratory allowed for many interesting proof-of-concept type demonstrations, its use for computer vision was hindered by some of the choices made in its design (image capture to videotape, and subsequent analysis by the digitizing the recordings, approximate synchronization of the cameras, interlaced image capture, etc.). We set out to build a 2nd generation laboratory that would take advantage of advances in hardware, and provide a source of high-quality multiperspective data for continuing research. With support from the Keck foundation, we converted this vision to reality in 1998-1999.

The resulting Keck Laboratory for the Analysis of Visual Movement ("Keck lab") is a multi-perspective imaging laboratory, containing 64 digital, progressive-scan cameras organized as sixteen relatively short baseline stereo rigs (see Figure 1). In each quadranocular rig, there are three monochromatic cameras and one color camera. Each group of 4 cameras is connected to a PC running Windows NT [1] that can collect imagery from the cameras at speeds of up to 85 frames per second (FPS). The PCs are all networked. The dimensions of the Keck laboratory are 7m$\times$7m$\times$3m. A panoramic view is shown in Figure 2.

This paper provides a description of the development of the laboratory, and provides an idea of its capabilities. Section 2 describes the hardware choices made in the development of the laboratory. Section 3 describes the software architecture of the Keck laboratory. Section 4 describes the performance bounds of the currently configured laboratory, while Section 5 compares our laboratory with some others. Section 6 concludes the paper.

2 Hardware

The hardware used in the Keck lab includes the following:

– Sixty-four digital, 85 FPS, progressive-scan cameras

* Present address: Microsoft Research, Redmond, Washington.

[1] In the text we refer to a number of commercial products. These are the trademarks or copyrights of their respective companies. Mention of a particular product should not be construed as an endorsement of the product.

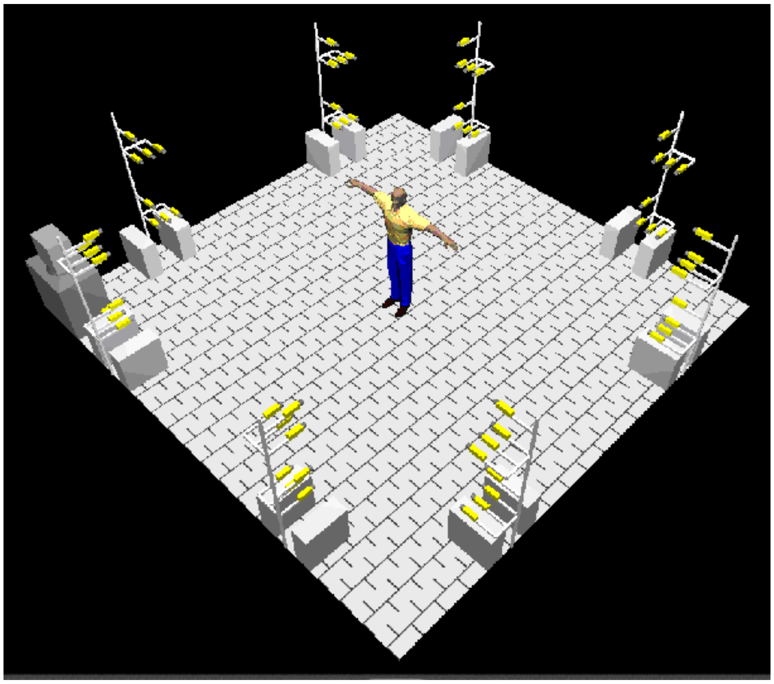

Fig. 1. Architecture of the Keck laboratory.

  - 48 grayscale, Kodak ES-310 and 16 color, Kodak ES-310C (Bayer color filter version of ES-310).
- Cameras mounted on custom, adjustable baseline quadranocular rigs, that ride on ITEM carriages fixed to the walls.
- 64 Schneider 8 mm C-mount lenses.
- 64 Matrox Meteor II Digital frame grabbers.
- 17 Dell 610 Precision Workstations (dual Pentium II Xeon 450 MHz with 1 GB SDRAM, expandable to 2 GB, a 9 GB SCSI Ultra II Wide hard drive, and an integrated 100 Mbps Ethernet interface).
- Data Translation DT340 Digital I/O board.
- Calibration objects (including a Peak Performance calibration frame, a timing calibration object, and a lighted stick).
- 3 Apex Outlook switches that share the monitor, keyboard and mouse among all the PCs.
- 3COM 100 Mbps 24-port network switch
- Blackbox RS-485 interface adapter
- Quantum 35 GB Digital Linear Tape drive (for data storage)
- Treadmill (for human gait analysis)
- 50" Pioneer PDP-V502MX plasma display (for demonstrating results).

2.1 Video data capture subsystem

A schematic of the Keck lab is shown in Figure 4.

Cameras A primary goal in the design of the Keck lab was to maximize captured video quality, while using commonly available hardware for economy. To meet this goal, uncompressed video is captured using digital, progressive scan cameras directly to PCs. Since the laboratory was to be used primarily for computer vision studies, it was important that the quality of the captured data be good. Consequently we chose a digital solution. The Kodak ES-310 cameras have a resolution of $648 \times 484 \times 8$ and can operate at up to 85 FPS in full frame progressive scan mode (speeds up to 140 FPS can be achieved using a smaller region of interest window). The ES-310 has a 10-bit digitizer for each pixel, in which the user can select which 8 bits are used for digital output, allowing for an effective gain in low-light conditions. The camera has a software controllable iris and shutter speed, is capable of being externally triggered, and can be controlled via an external serial interface (RS 232 or RS 485). Each CCD has a two-tap digitizer to achieve fast A/D output.

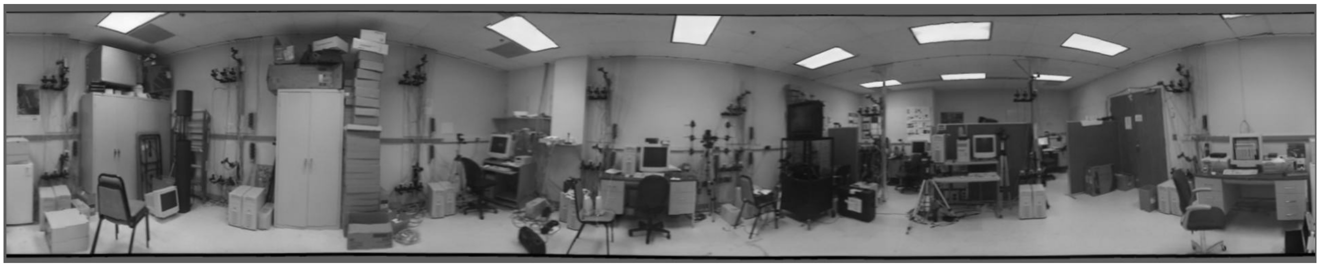

Fig. 2. A panoramic view of the Keck lab.

Forty-eight of the of cameras are grayscale (ES-310), while the remaining sixteen are color (ES-310C). The color cameras are identical to the grayscale ones, except for the addition of a Bayer color filter mask on the CCD, and an IR filter. To process the images from the color cameras, color pixel (RGB) values need to be spatially interpolated (e.g., see [5]).

A real-world problem with CCDs is that they often have defects. Our cameras were chosen so that each CCD can have at most 1 major pixel defect (a pixel that deviates from its surrounding pixels by more than 20% when imaging a uniformly illuminated object). The defect list for the cameras is stored as calibration data.

Fig. 3. Example images from four viewpoints.

Acquisition PCs Each acquisition PC runs Windows NT 4.0 SP4, and is equipped with 4 Matrox Meteor II digital frame grabbers. Each PC has 5 PCI slots. One slot was left free for any future cards (such as for gigabit networking, that is currently being implemented). The video display, keyboard and mouse outputs for each of the PCs are aggregated using Apex monitor switches.

The Keck lab was designed to capture uncompressed video sequences to both memory and disk. The data throughput requirements for various number of cameras and frame rates are shown in Table 1. The design of the Keck lab allows capturing uncompressed video to memory at up to 100 MB/s, and capturing to disk at up to 50 MB/s. In order to achieve the

# cameras	FPS	Throughput (MB/s)
1	30	8.9
4	30	35.9

# cameras	FPS	Throughput (MB/s)
4	60	71.8
4	85	101.7

Table 1. Data throughput requirements

required 50 MB/s disk throughput, 3 SCSI Ultra 2 Wide disks (Seagate Cheetah) are used in a RAID configuration. Double the disk throughput could be achieved by writing a custom frame grabber device driver, which would write the images directly to the SCSI controller, instead of buffering the images to memory (which requires transmitting them over the PCI bus twice) [4]. Newer dual 64-bit PCI bus systems (such as the Dell Precision Workstation 620) could write over 102MB/s to disk while capturing 4 cameras at full speed, without the need of a custom driver.

2.2 Controller

Trigger Generation All 64 cameras are frame synchronized using a TTL-level signal generated by a Data Translation DT340 card, and transmitted via coaxial cables to all the cameras. For video acquisition, the synchronization signal is used to simultaneously start all cameras. No timecode per frame is required.

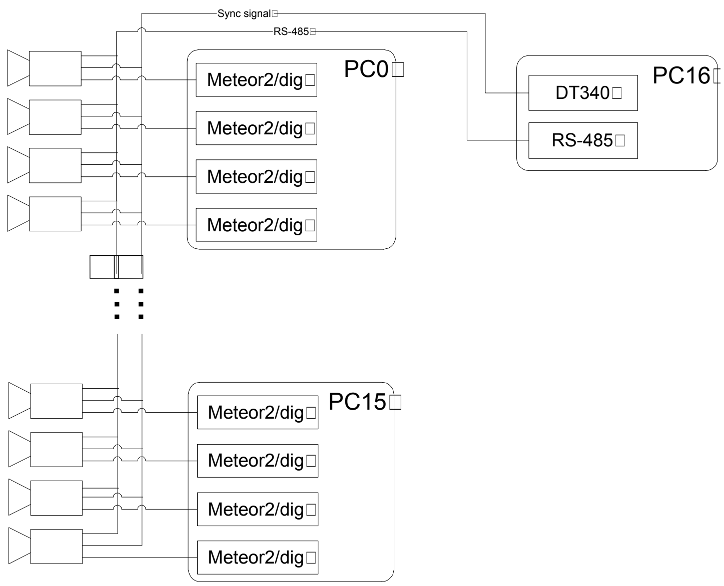

Fig. 4. Keck Lab Schematic

Camera Parameter control The configuration settings on each camera are controlled using a single RS-485 [1] network. The controller PC communicates to all cameras using the Blackbox RS-485 interface adapter. The RS-485 bus is a ribbon cable, which each camera is connected to. There are over 40 camera commands that can be given to the cameras for setting parameters such as the trigger mode, shutter speed, frame rate, f-stop, exposure time, and gain.

Controller PC The 17th PC is designated the controller PC. The DT-310 data acquisition card and the Black-box RS-485 cards are installed in this PC. The associated trigger software and camera control software also reside on it. This PC also has designated disk space that is used for the cloning software (described below). The monitor switch, the network switch, mouse, keyboard and a 21” display monitor, are all aggregated in one corner of the room around this PC, and are shielded from view of the cameras by using standard cubicle separators.

Light	Framerate (FPS)	Percent variation
Fluorescent	60	0.34
LightTech	60	0.083

Light	Framerate (FPS)	Percent variation
Fluorescent	85	3.6
LightTech	85	0.88

Table 2. Camera pixel variations (flicker) under different light sources and frame rates

2.3 Network

Each PC has an integrated 100 MBit network interface. The PCs are all connected to a 3COM 100 Mbit/s network switch. Tests indicate that sustained data throughputs of about 50 Mbit/s are possible. For purposes of data acquisition, this bandwidth is sufficient, as the network only carries simple messages generated by the controller software.

However, research is underway that employs the PCs as a computational cluster for running vision algorithms. To support this research, the PCs are being upgraded to employ a gigabit Alcatel switch and network interface cards. Preliminary testing with this setup indicates data throughputs of 350 Mbit/s.

2.4 Distributed Computing

The 450 MHz Pentium II is capable of 1800 MIPS, using MMX operations [4]. With dual CPUs per PC, the 34 available processors provides significant computational power for real-time computer vision applications. We use WinMPI for distributed computing. Further details may be found in [3]. Once the network upgrade is completed, the computers will also form a useful compute cluster for data analysis. Additionally, the Dell 610 is capable of being upgraded to faster Pentium III processors, which would further increase the computational capabilities. Additionally, the memory on each PC can be upgraded to 2GB.

2.5 Lighting

Our current setup uses standard fluorescent lights, augmented with high-frequency (25KHz) fluorescent (LightTech CFL 255) lamps and halogen lights. Because standard fluorescent lighting have a small component that oscillates at 60Hz, images captured with the Kodak ES310 flicker. To measure the magnitude of the flicker, a white sheet of paper was placed 9' away from a the light source, with the camera 5' from the paper. The ES310 exposure was set to 2ms, and the iris was adjusted to give images with 60% saturation. A pixel intensity $I(x, y, t)$ in the captured image sequence is a sinusoidal signal (plus noise) as a function of time. The percent of maximum deviation from the mean pixel intensity is given in Table 2. Under standard fluorescent lighting with images captured at 85 FPS, there is visibily noticeable flicker in an image sequence, which could affect the processing of the images; the same lighting at 60 FPS gives less than one tenth the flicker. Halogen light gives similar flicker magnitudes to standard fluorescent lighting.

2.6 Treadmill

While the laboratory provides a relatively large area (2.2 m square) over which a subject can be viewed by all cameras, such a distance would be covered quickly by a walking person. To facilitate human motion studies, a portable treadmill is used, with the treadmill's controller panel removed to eliminate occlusions.

3 Software

The Keck laboratory uses multiple Windows NT 4.0 SP4 workstations networked as a workgroup and communicating through a switch. One PC is designated the controller and runs control software, while the other 16 PCs are acquisition PCs that run the acquisition software.

3.1 Acquisition software

The acquisition software (see Figure 5) uses a custom DCOM server, `KeckServer`, which runs on each of the 16 PCs. The controller PC makes connections with each of the camera PCs, and sends and retrieves messages and images. The ICamera interface used for the `KeckServer` is:

– `HRESULT ICamera::SetDcfFile([in, string] char *dcf)` Sets the Matrox DCF file.
– `HRESULT ICamera::AllocBuffer([in] BOOL cameras[4], [in] long numFrames)` Allocates the number of image buffers for the specified cameras.
– `HRESULT ICamera::ReleaseBuffer()` Releases all previously allocated image buffers.

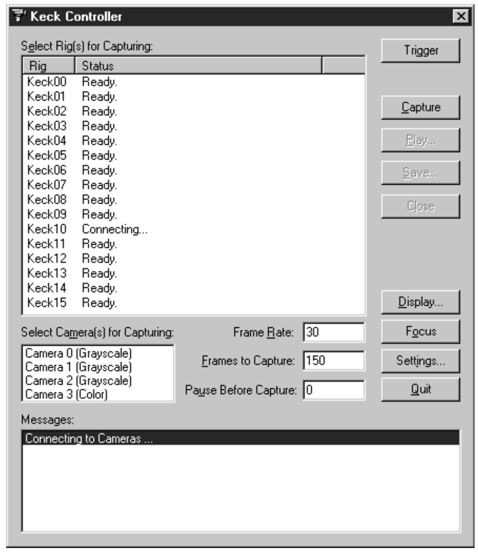

Fig. 5. Keck controller software

# cameras	FPS	Max duration (sec)
1	30	99.8
4	30	24.9
4	60	12.5
4	85	8.8

Table 3. Maximum capture durations

- `HRESULT ICamera::StartCapture()` Starts capturing; returns when all the image allocated buffers are filled.
- `HRESULT ICamera::SaveCapturedSeq([in] long format, [in, string] char *fileName)` Saves the captured sequence in the specified format.
- `HRESULT ICamera::GetCapturedImage([in] long cameraNumber, [in] long imageNumber, [out] char image[644*484])` Returns the captured image specified by `imageNumber`.
- `HRESULT ICamera::GetLiveImage([in] long cameraNumber, [out] char image[644*484])` Captures and returns a live image.

3.2 Cloning

Upgrading the OS/drivers can be a tedious task. We have developed a solution to clone the machines based around Disk Image Pro. One of the acquisition workstations is upgraded, and its disk cloned. Each workstation is booted to a DOS state (with networking) via a CD, and the cloning file copied from a file on the controller. A SID changer software is used then to replace the clone SID with the actual machine SID.

3.3 RS485 camera control software

The ES-310 can be configured using either a RS-232 or RS-485 interface. In the Keck Lab, we have designed a RS-485 network that is connected to all the 64 ES-310 cameras, and is controlled from the controller PC, which is equipped with the Blackbox RS-485 interface adapter.

4 Capabilities

The Keck lab is currently configured to capture up to 896 MB of video into upper memory (above the 128 MB allocated for Windows NT). This corresponds to 2995 frames of size $648 \times 484 \times 8$. The maximum capture durations are given in Table 3. The capture durations can be increased by a factor of 2.14 by expanding the PCs from 1 GB to 2 GB. As discussed earlier the PC cluster also forms a significant computational engine.

4.1 Stereo error analysis

The quadranocular camera nodes of the Keck lab are designed to facilitate stereo depth computations. The trinocular baseline is adjustable from 150 to 300 mm, using a specially designed rig for camera mounting. Additionally the rig itself can slide vertically along a rail on the wall on which it is mounted. This adjustability allows easy modification of the camera configuration for different applications, e.g., the cameras can be placed on the surface of a virtual sphere, as in Figure 6.

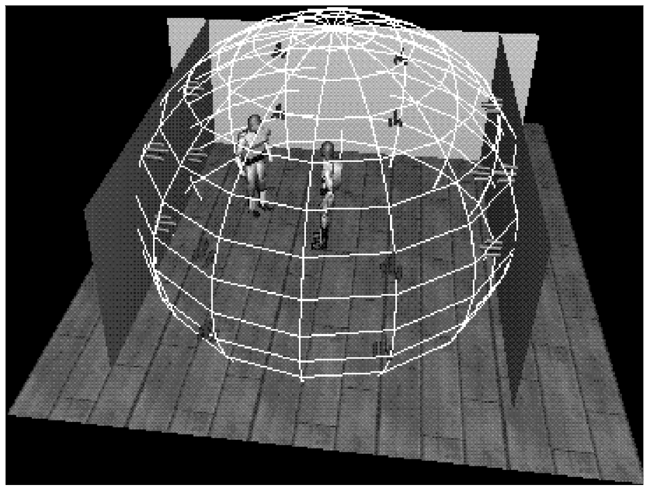

Fig. 6. The adjustable mountings allow the cameras to be positioned on a virtual sphere, if so desired.

With a 300 mm baseline, a distance of 2 m between the object and camera, and assuming single pixel correlation accuracy, the depth precision is 26 mm. The depth precision for a range of baselines is given in Figure 7.

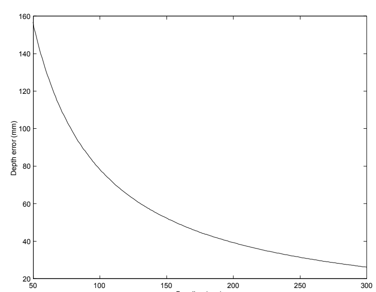

Fig. 7. Stereo error analysis

4.2 Lens distortion

In selecting the lenses, we considered both the field of view (FOV) and lens distortion. In general, as FOV increases, so does lens distortion. The image distortion for 3 commonly available C-mount lenses was determined using a grid pattern. From the images shown in Figure 8, the Schneider lens clearly had the least distortion. Moreover, the Schneider lens was the only lens tested that did not have significant defocus near the image perimeter. Note that while certain types of lens distortion (e.g., radial) can be corrected in software, image defocus cannot be, particularly within the constraints of a real-time system.

4.3 Calibration hardware

Large Frame To facilitate strong calibration of the camera system, a Peak Performance calibration frame is utilized (see Figure 9). The calibration frame contains 25 white balls (1" in diameter), each of which has a known location accuracy of 1 mm. Additional hardware, such as a 1 m length wand with LED's at known locations, are also used for projective calibration. Further research on fast techniques for calibrating the Keck laboratory is underway [2]

To check the accuracy of the camera synchronization, a special box with multiple time-synchronized lights was developed. This box has multiple rows of lights that blink in a specified pattern, with the blinking rate that can be set by the user. This device is also a useful tool to check the acquisition for frame drops.

Camera Internal Calibration Based on these specifications, the camera internal calibration include the regular calibration parameters (focal length, skew, offset), the pixel defect locations (if any) and their corrective gain ratio. In addition since the two taps can have different gains, we must also determine an inter-tap gain ratio that will equalize the rows. In addition, the color cameras must be color-calibrated, so that corresponding colors between cameras can be correlated. This is important for applications requiring the use of color (e.g., space-carving). This is done by capturing an image of a Gretar Macbeth color chart using each color camera, extracting the color values, and computing a color transformation for each camera from the camera colors to some base color values.

5 Related work

The most related multi-perspective camera laboratory is the CMU "3D Room" [6]. The CMU lab uses 51 analog single CCD cameras, which are now captured to PCs (they were previously captured to 51 VHS recorders). As a result of single CCD color interpolation, the effective luminance resolution is approximately 320x240 @ 30 FPS (for one field only; if both fields are analyzed, then 60 FPS is achieved, but the odd and even fields cannot be readily analyzed together. In comparison, the Keck laboratory achieves 648x484 luminance resolution at 85 FPS (up to 140 FPS if a smaller region-of-interest is used). Since the Kodak ES-310 cameras have a digital interface with a 10-bit digitizer (8-bits selectable), the S/N is significantly better than the analog CMU cameras (analog cameras typically get 6 good bits (36dB), while the ES-310 gets 8 good bits (48dB)).

After our work commenced we have heard of a number of other projects for multiple camera imaging. Notable among these are the work of Prof. Matsuyama at Kyoto University, where fewer pan tilt zoom cameras are used. However, that facility still uses analog cameras running at a lower frame rate, and the choice of pan-tilt and zoom cameras would place a significant calibration and registration burden on many computer vision applications.

6 Conclusions

We have described the design and implementation of the Keck lab, a state-of-the art multi-perspective imaging laboratory. The lab is a unique facility that is being used to support many studies, including multi-perspective 3D object reconstruction (e.g., as in Figure 10), motion analysis, human gait analysis, and real-time parallel processing for computer vision. The laboratory should be a rich source of high quality synchronized video data for the vision community for years to come.

7 Acknowledgments

The generous support of the W.M. Keck foundation is gratefully acknowledged. We would also like to thank Ms. Cecilia Kullman for managing the complex job of ordering all components for the lab with good humor and grace.

References

1. Http://www.rs-485.com/specs.htm.
2. P. Baker and Y. Aloimonos. Complete calibration of a multi-camera network. In *Proc. IEEE Workshop on Omnidirectional Stereo, Hilton Head, SC*, pages 134–141, 2000.
3. Eugene Borovikov and Larry Davis. A distributed system for real-time volume reconstruction. In *Proceedings of International Workshop on Computer Architectures for Machine Perception, Padova, Italy*, 2000.
4. Ross Cutler and Larry Davis. Developing real-time computer vision applications for intel pentium iii based windows nt workstations. In *FRAME-RATE*, 1999.
5. M. Jackowski, A. Goshtasby, S. Bines, D. Roesman, and C. Yu. Correcting the geometry and color of digital images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(10):1152–1158, 1997.
6. Takeo Kanade, Hideo Saito, and Sundar Vedula. The 3D room: Digitizing time-varying 3D events by synchronized multiple video streams. Technical Report CMU-RI-TR-98-34, CMU, 1998.
7. Yonatan Wexler. *Tensor Methods for Vision and Graphics*. PhD thesis, Department of Computer Science, University of Maryland, College Park, 2001.

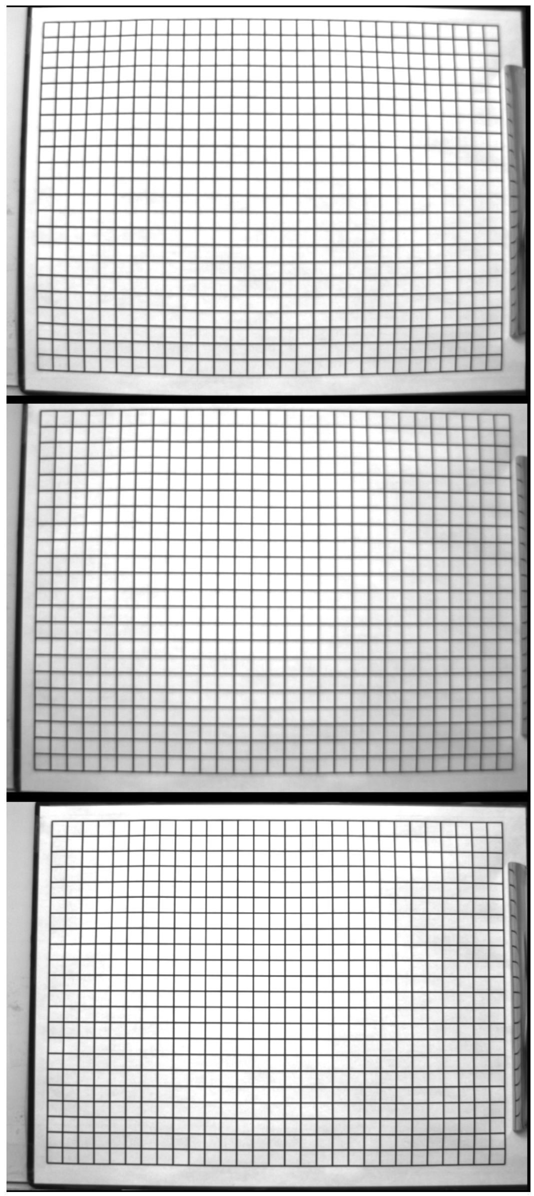

Fig. 8. Lens distortion analysis. Top: Cosmicar 6 mm, Middle: Canon 7.5 mm, Bottom: Schneider 8 mm.

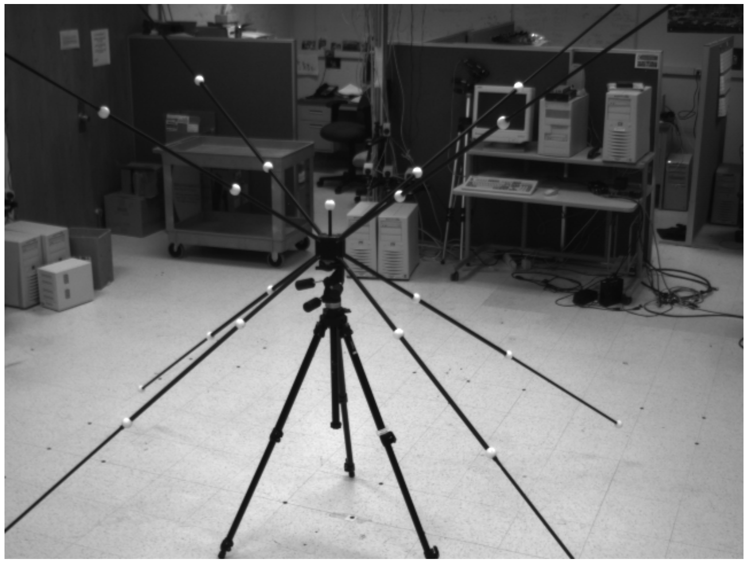

Fig. 9. Peak Performance calibration frame

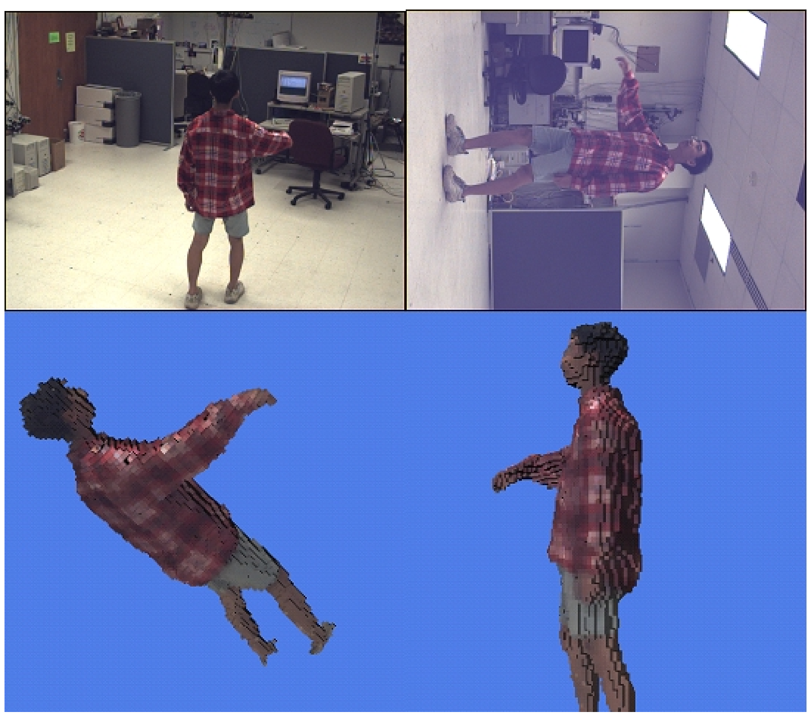

Fig. 10. Preliminary reconstructions using a fast algorithm developed by Y. Wexler [7].