Many-Camera Systems:
How They Started at CMU
up to EyeVision at 2001 Superbowl

CVPR 2017 in Honolulu
Takeo Kanade
Robotics Institute
Carnegie Mellon University

Okutomi-Kanade Multi-baseline stereo theory 1990

Multi-camera Stereo

Outputs: Color Image 256 x 256 x 24 bits
Deph Image 200 x 200 x 8 bits at 30 fps
Soft Camera (aka Virtual Camera): Occlusion Problem

View Synthesis

Range
Intensity/Color

Soft Camera Output
Whole-Scene Modeling

We use many, many cameras!!
3D Dome  Analog system
with 10 BW Cameras - circa 1994
with 51 Color Cameras - circa 1995
Old and fun stories

Grad student vs. Robot

51 Analog Video tape recorders
Old and fun stories

Grad student vs. Robot

Synchronization – nontrivial problem
Old and fun stories

Grad student vs. Robot
Synchronization – nontrivial problem
Our own digitizers

Figure 1: Vertical blanking portion of a frame containing VITC data
Old and fun stories

Grad student vs. Robot
Synchronization – nontrivial problem
Our own digitizers
Calibration pattern on special non-stretchable paper
4D Full Body Surface Modeling

Stereo

Merging

Texture Map

...
Voxel Merging

Implicit function:

\[ F(x,y,z) \begin{cases} > 0 & \text{outside} \\ = 0 & \text{on the surface} \\ < 0 & \text{inside} \end{cases} \]

Depth Map 1

5m x 5m x 3m
1 cm³ voxel
500 x 500 x 300

Depth Map 2

[Levoy, etal]
[Ikeuchi, etal]
[Kanade, etal]
4D Modeling
One on One – circa 1995
4D Digitization (1995)

Example:
3-Man Basketball

Synthetic court

Input sequence

4D Model
Fully Digital “3D Room” ~2000

New 3D Room ~ 2003
4D Digitization (~2000):
Man-Sofa-Ball

*Digital 3D Room: 39 High Quality Cameras*
Real Time 4D Digitization

- 64 x 64 x 64
- 10 frames/sec (with 5 PCs)
- Avatar creation
Example 5:
Dance
(Aug 2000)

4D Digitization

Click to play
Markerless Motion Transfer

Input video

Processing

Output video

- martial arts master
  - kung-fu fighting
- ballet expert
  - dancing

- ballet expert
  - dancing
- martial arts master
  - kung-fu fighting
SubjectE performs SubjectS's THROW motion

Motion Transfer from SubjectS to SubjectE
EyeVision

at

Super Bowl

Movie "Matrix"-like replay anywhere in the field
EyeVision at Super Bowl XXXV 2001

33 robot cameras
Trailer and wires
Bigger ideas that didn’t happen?
Interview Room of the Future

Multi-modal
Interactive
Real-time feedback
MEMS-based Steerable Balloon

Cameras, GPS, Radio links
O-ita Stadium
“Free View Point” Project

with Yuichi Ohta (Tsukuba)
Hideo Saito (Keio)
Akimichi (Takenaka)
Seiki Inoue (NHK)
Computation vs. Bandwidth

- How much computation at the center or at the TV computer?
- Where/how the storage at the user end?
- What the most efficient intermediate representation?
- What interactions by users?
Computation?

Estimate:

500 x 500 x 500, 50 NTSC Cam, 30 f/s $\Rightarrow$ 100 Gflop

Progress:

Jan 1998: 10,000 x Realtime with a few SGIs
(14 days elapse time)

Feb 1999: 1,000 x Realtime with 20 PCs
(4 days elapse time)

Jan 2002: Modeling: $< 50$ x Realtime
Display: Realtime
Estimate:

500 x 500 x 500, 50 NTSC Cam, 30 f/s \rightarrow 100 \text{ Gflop}

0 0 0 1000 \text{ HDTV} 60 20 \text{ Pflop}

Progress:

Jan 1998: 10,000 x Realtime with a few SGI\text{s} (14 days elapse time)

Feb 1999: 1,000 x Realtime with 20 PCs (4 days elapse time)

Jan 2002: Modeling: \textless{} 50 x Realtime
Display: Realtime
Many-Camera Systems

“Numerosity is power.”

Early efforts:
- Primitives and Pretty good

Fun and useful

Challenges in devices, algorithms, computation, communication
AND application scenarios