Challenges of Wireless Digital Video
-
A Computer Vision Perspective

Reinhard Klette

The University of Auckland
New Zealand

with contributions by **Ji Sun** (*Wellington, New Zealand*),
Stefan Gehrig et al. (*Sindelfingen, Germany*),
Chris Croft and **Paul Dewar** (*Perth, Australia*), and
Felix Woelk et al. (*Kiel, Germany*)

Wireless Digital Video (starting about in 2000)

- concept not new to Computer Vision: e.g.
  digital photogrammetry (use of satellite data,…)
- there are already technologies for uncalibrated
  multi-view stereo analysis, for surveillance, …

New Challenges for Computer Vision

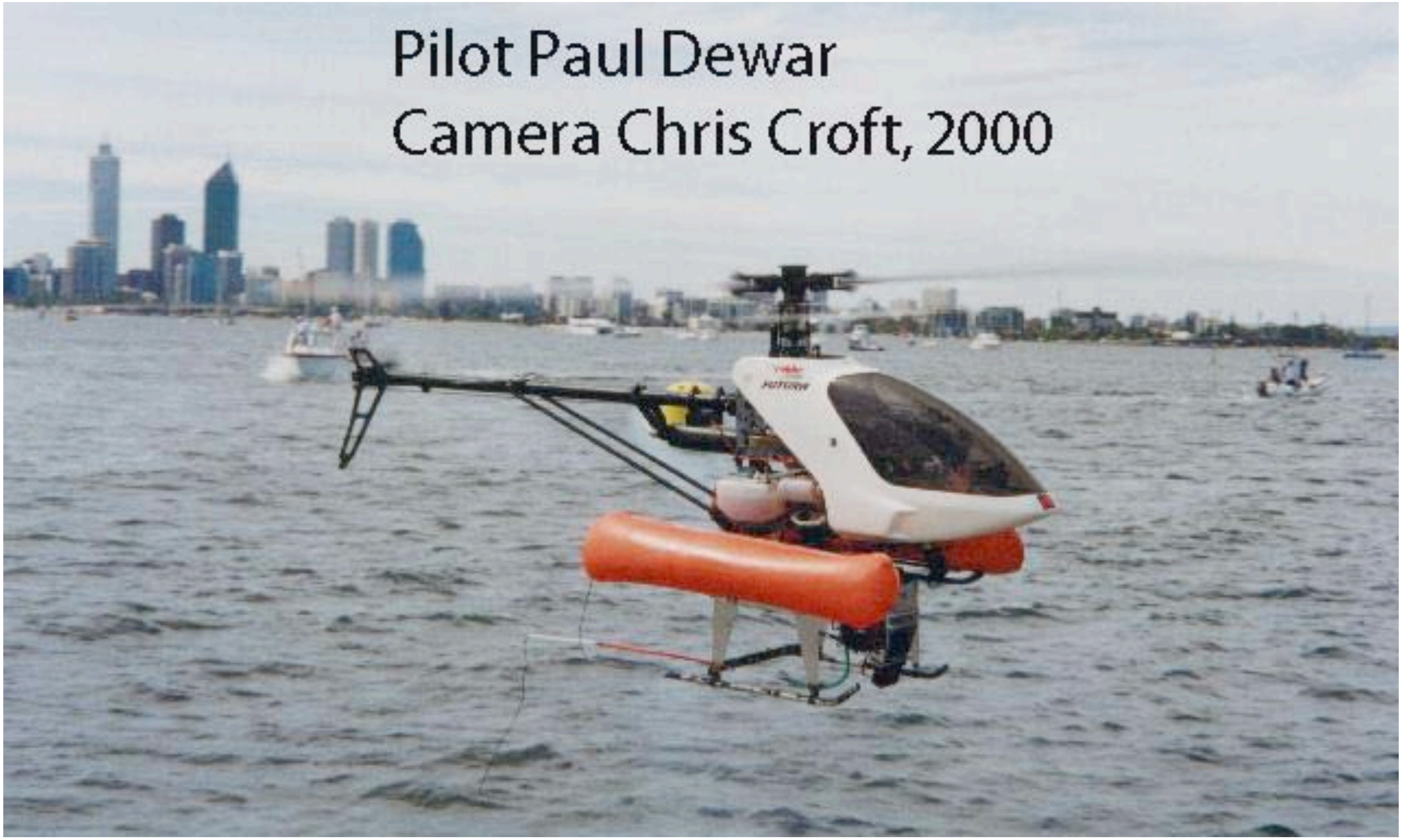
1. support application of wireless (WL) video
2. combine W+WL video into unified app's
3. apply WL video as a new technology for
   traditional areas (e.g., 3D scene documentation)

1. Support application of WL video

WL video for documentation, inspection, or visualization:

- civil engineering (inspection of elevated sites, inspection of demolition sites, etc.)
- reporting about aquatic events (white water canoeing, long-distance swimming, etc.)
- support movie production at difficult to reach locations
- and so forth

WL video transmitted from a helicopter



The University of Western Australia & Unmanned Vehicle Company
Chris Croft & Paul Dewar

Purpose-build **helicopter** (wing diameter: 1.0 m)

Standard wireless color **camera** (AUS$129.- in 2006)
with 3.6 mm board lens
in a purpose-build, remotely controlled PTU (pan-tilt unit)

Signal is fed to a **transmitter** (up to 1km) in the helicopter

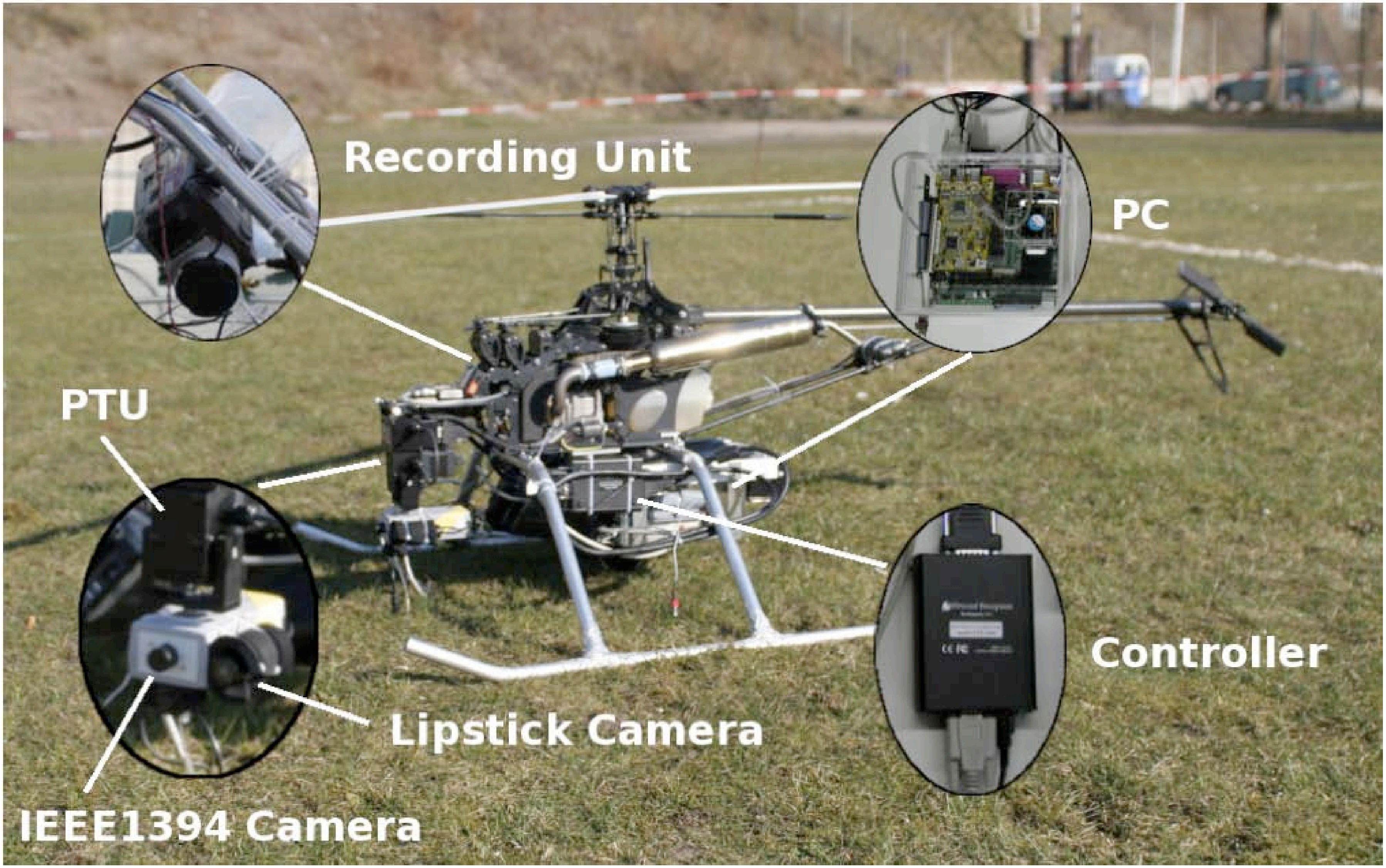
Video is **recorded** on the ground in standard PAL format
25 fps on a Sony Digital Handycam 8mm

Camera: **frame size** 440 pixel horizontal
Recording: 726 x 550, DV Standard

Civil engineering applications (pole inspection)

Reporting about an aquatic event (white water)

How to remove the need for a second operator?



Kiel University, Computer Science Department, 2003-2005
Ingo Schiller, Felix Woelk, and Reinhard Koch

Track object of interest automatically

i.e., no operator for camera PTU (= pan-tilt unit) needed

Approach:

Stabilize camera by an automatic tracking system (using the PTU for automatic camera attitude correction).

Method:

Use of color histogram similarity and of a Bayesian particle filter.

Important: accuracy of image data, processing frame by frame, bandwidth for transmission

An alternative approach **could (!)** be based on detecting the

dominant plane

in captured images

see CITR-TR-88, 2001 (Kawamoto and Klette)
and CITR-TR-111, 2002 (Kawamoto, Yamada, Imiya, and Klette)

"The dominant plane is a planar surface covering more than 50% of a frame, or being that planar surface which is represented in the image with the largest number of pixels."

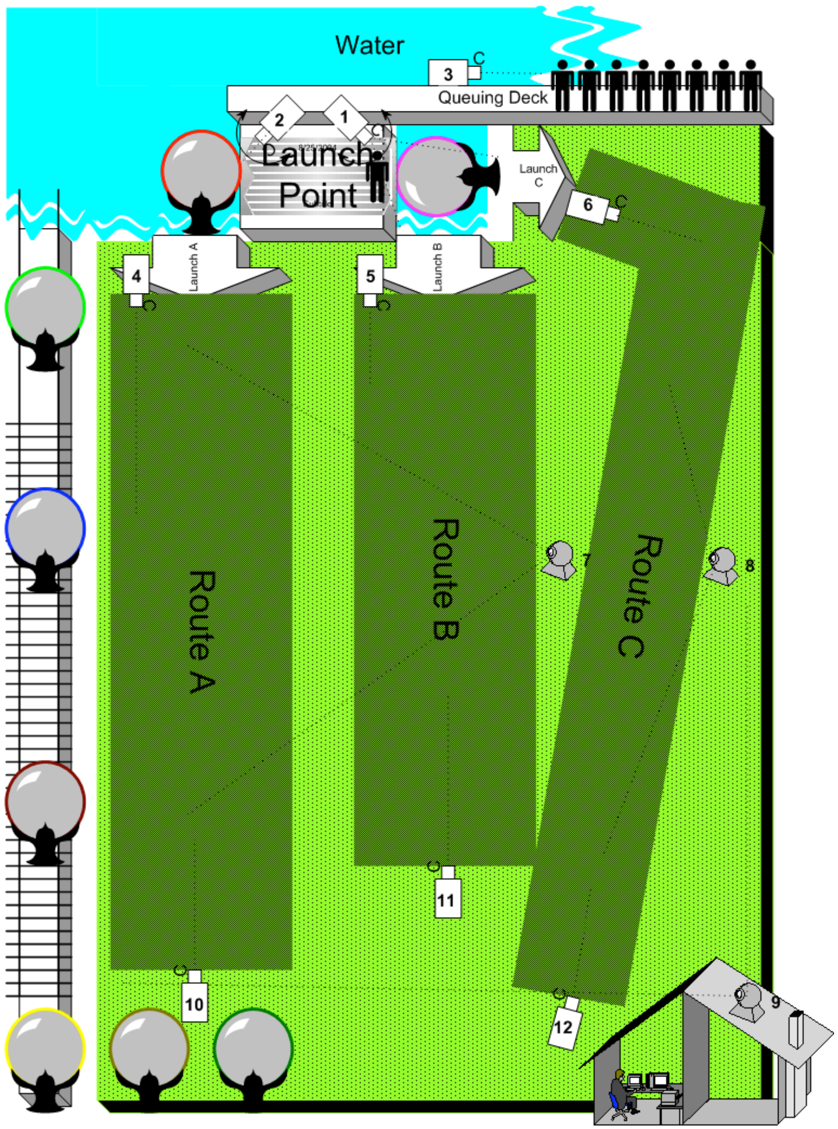
2. Combine (automatically) W+WL video

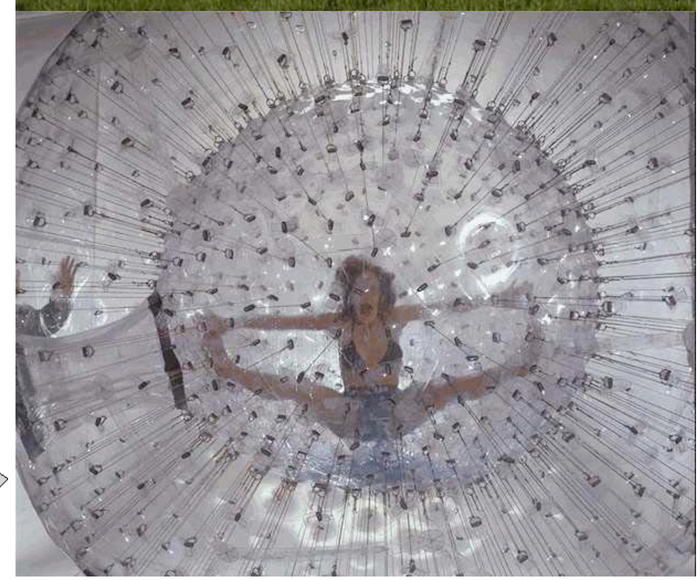
Multi-camera video systems for selected presentation of multi-view perspectives:

- of a sports or leisure activity taking place within a large area
- of a large-scale arts performance
- and so forth

Goal: combine footage from different cameras into one sequential final video, select based on captured activities (= automated video understanding)

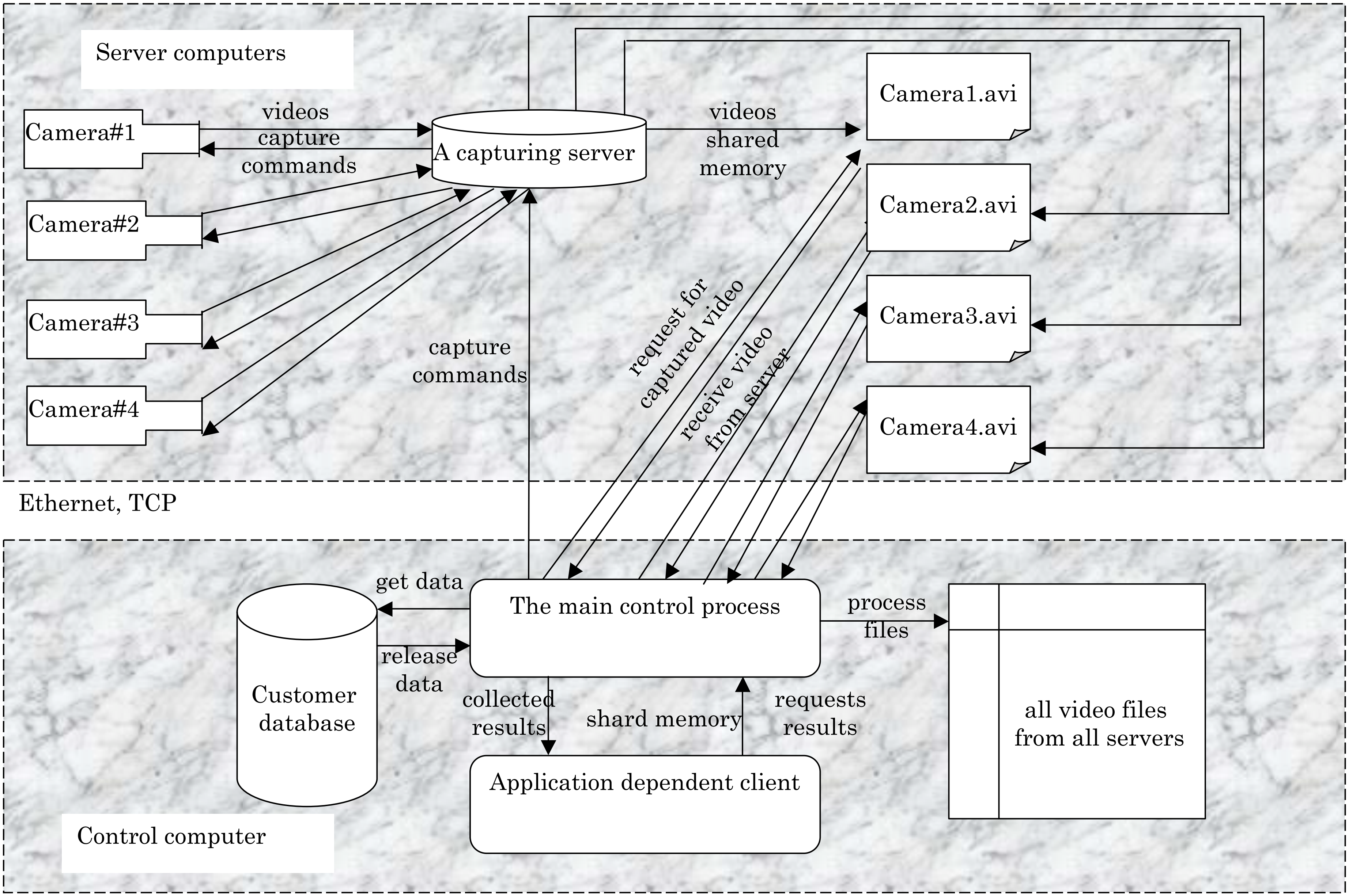
Example: video server for a Zorb site





The University of Auckland, Computer Science Department
Ji (Samuel) Sun, Lu (Lucy) Xia, Reinhard Klette

Sketch of video server system



Used WL video cameras

lightweight airborne video system LWV14 (or LWV14-T)

1 W amplifier, 1800mA NI-MH battery for 3-4 hours

10oz, transmitts 2,500 m (LWV14) or 4,500 m (LWV14-T)

450 line colour CCD, audio

Four ports video capture card

IVC-200G, an industrial four channels video capture card with GPIO module

Automated Video Generation

Multi-port software

e.g., SureLabs Stingray for a four ports surveillance system software

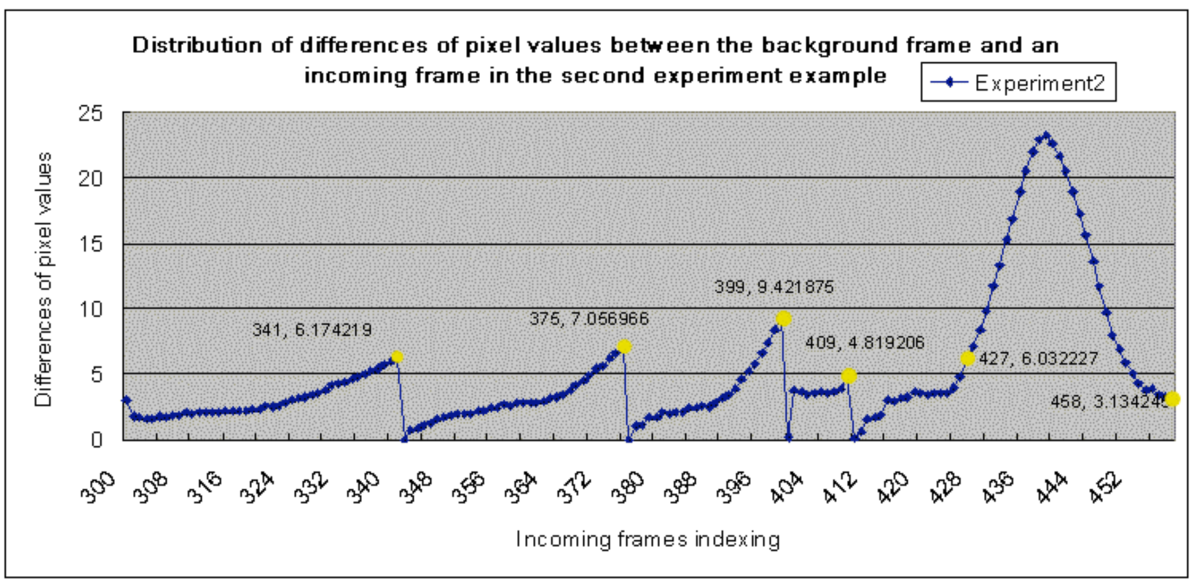
Two user controls at different locations, concurrent access to customers database

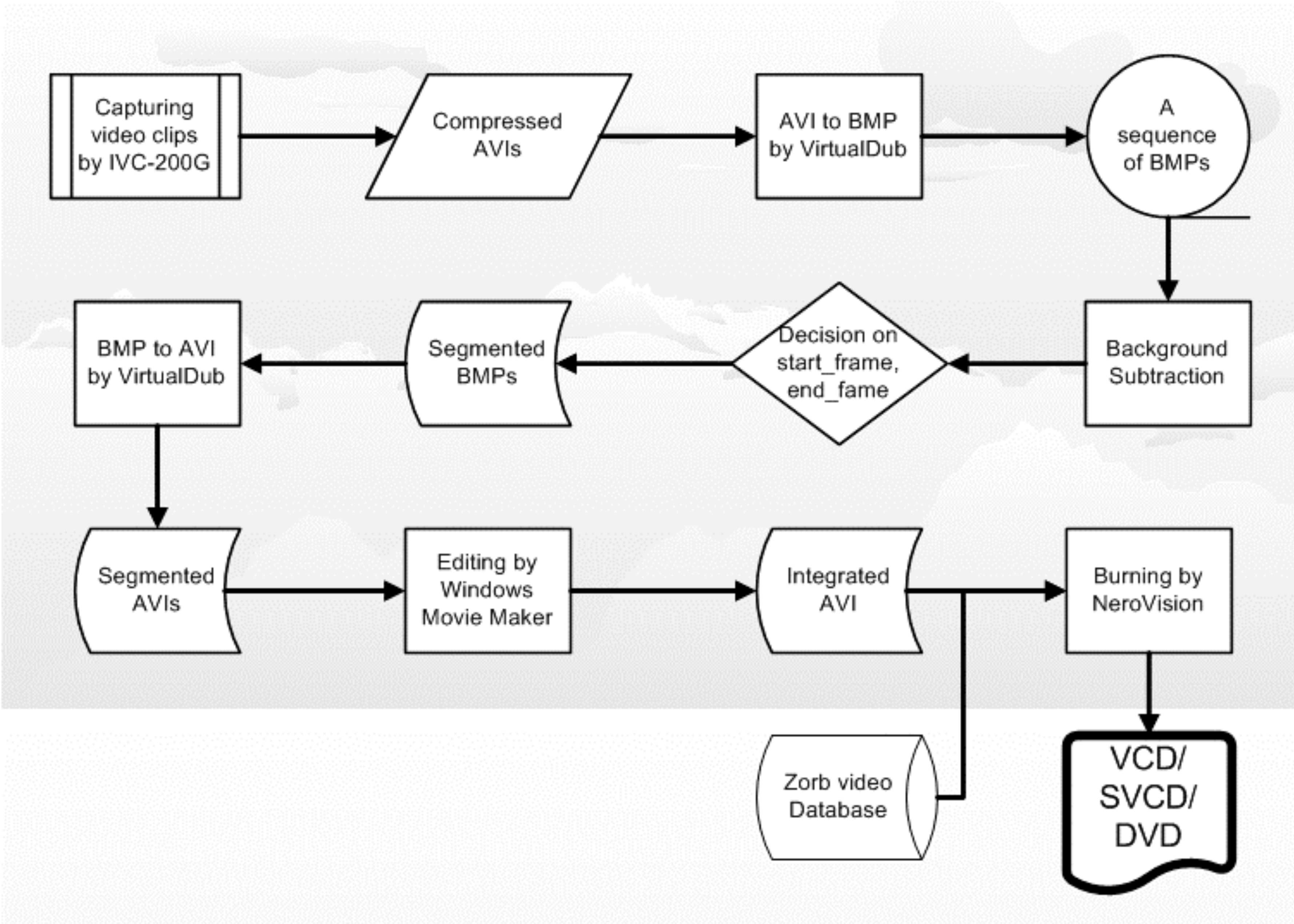
Video capture, monitor and record (file management system to transfer video files)

Interface to combine applications into integrated capturing and video generation

Time segmentation of video signals

Ensure that only relevant frame sequences are used for video clips





Capturing video clips by IVC-200G
Compressed AVIs
AVI to BMP by VirtualDub
A sequence of BMPs
Background Subtraction
Decision on start_frame, end_fame
Segmented BMPs
BMP to AVI by VirtualDub
Segmented AVIs
Editing by Windows Movie Maker
Integrated AVI
Burning by NeroVision
Zorb video Database
VCD/ SVCD/ DVD

Main problem: “quality” of WL image data
(2-layer plastic hull, object in irregular rotation etc.)
still insufficient for accurate video segmentation

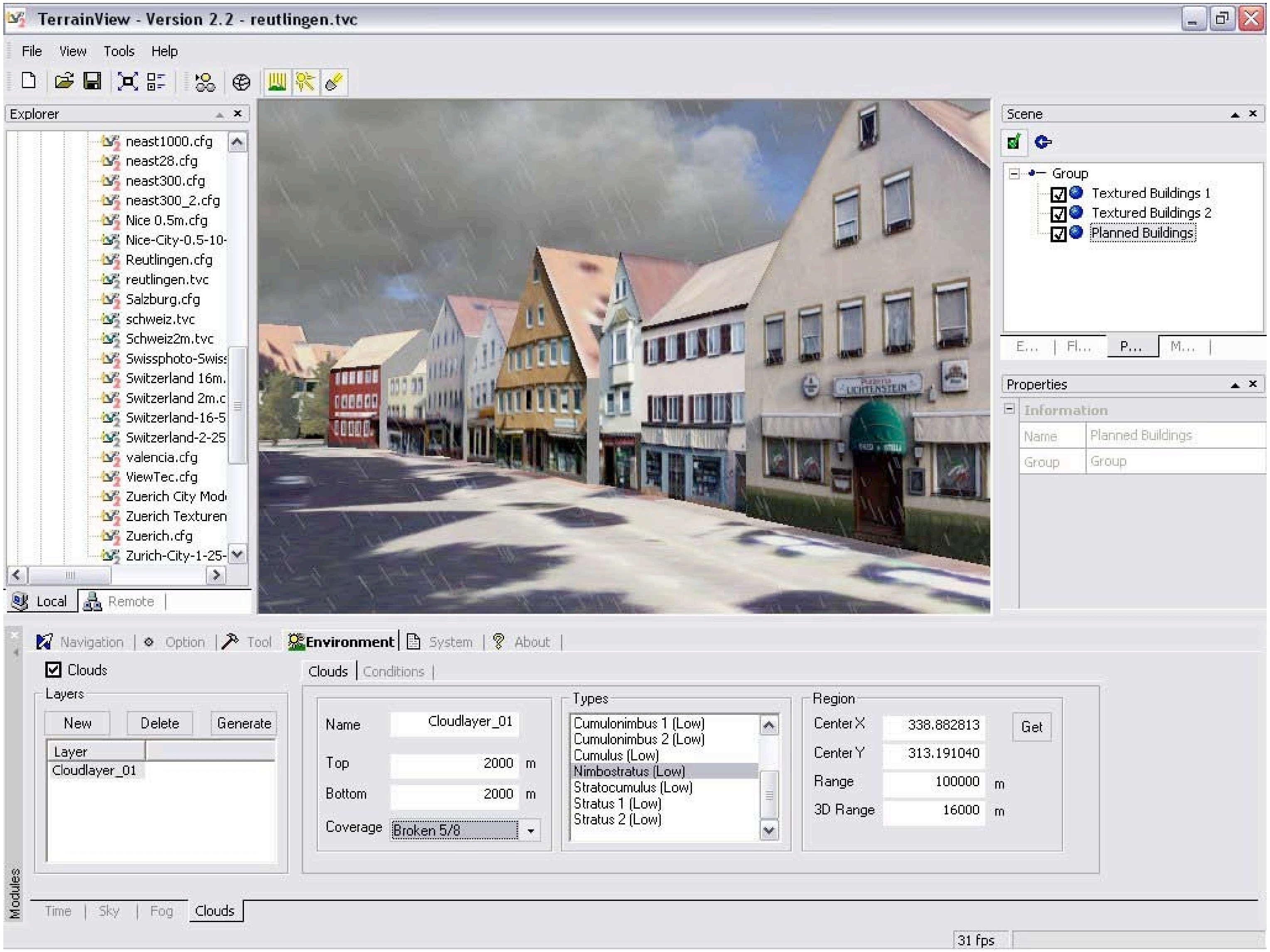
3. Use WL video for 3D Scene Recovery

Recently many projects (laser range finder, aerial imaging, etc.) already aim at:

- building large-scale 3D models of cities, suburbs,…
- 3D modeling of power lines, industrial sites, …
- 3D documentation of ("small") statues
- and so forth

Goal: recover the surface geometry of the given 3D scene, generate a surface model, use texture mapping and rendering for visualization

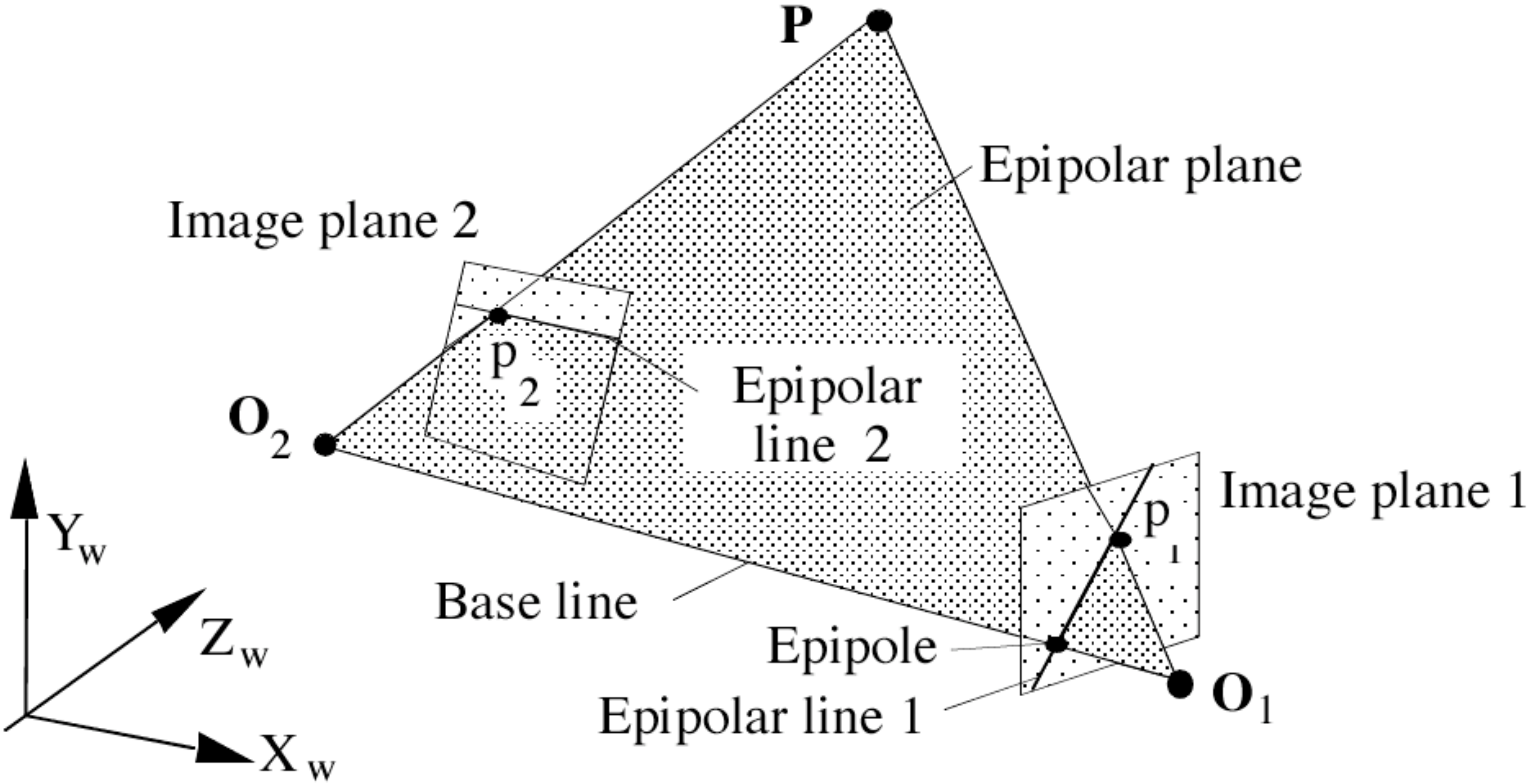
Use of computer graphics, orthophotos, or laser range finder data (see, e.g., CyberCity, Switzerland)



More details could be of interest
(CyberCity model of Los Angeles)

A “small” statue (5.5 m tall)
(Digital Micheangelo Project)

A WL camera flying around a "large" monument defines images in general positions:



Computer vision knows how to do

Structure from Motion - Multi-View Stereo

see, e.g., book by [Hartley and Zisserman, 2002]

1. Estimate essential matrix OR calibrate cameras
2. Geometric rectification of two cameras

   into standard epipolar geometry
3. Correspondence analysis for calculating disparities
4. Calculate depth or 3D surface points

   using the calculated disparities

Use of several calibrated wired cameras:

Daimler Chrysler Sindelfingen/Germany, Crash Test Analysis
Stefan Gehrig, Clemens Rabe, John Lin -- 300 ms at 1,000 pps

A brief intro into multi-view 3D scene recovery:

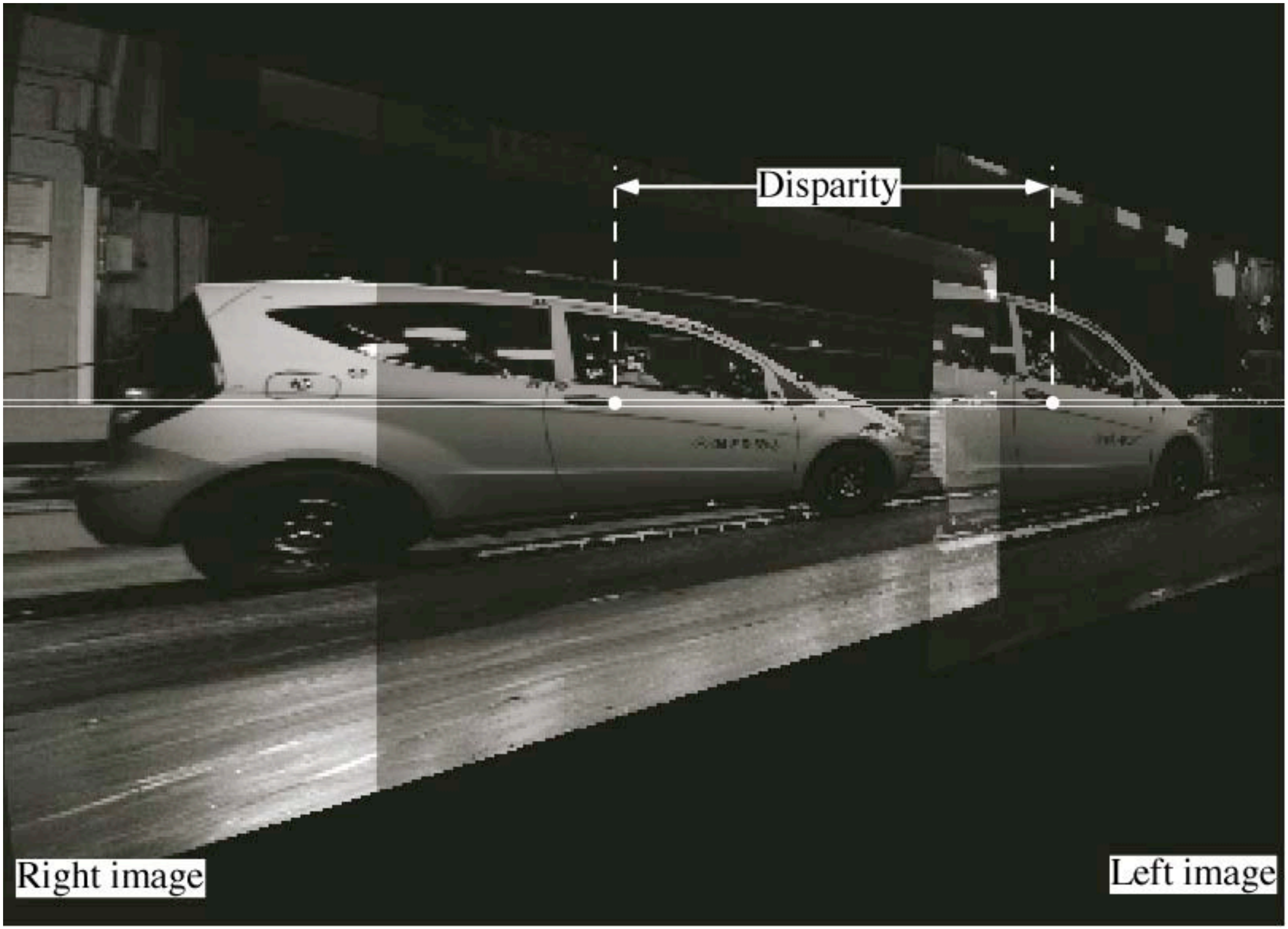
a. We start with two time-synchronised input images

b. Geometrically rectified images

(such as both cameras WOULD be ideally aligned)

c. Correspondence analysis along epipolar lines

d. From disparities to depth



e. Sparse 3D surface point reconstruction

Aim is:

dense stereo (dense matching points)

Requires/best if:

about Lambertian surfaces

search windows according to 3D geometry

accepting that there is
no ordering constraint etc.

BUT ALSO

Camera data have to be high-resolution, accurate, etc.

That’s what we want in computer vision:

High-Speed and High-Resolution Cameras

My subjective (2006) definition:

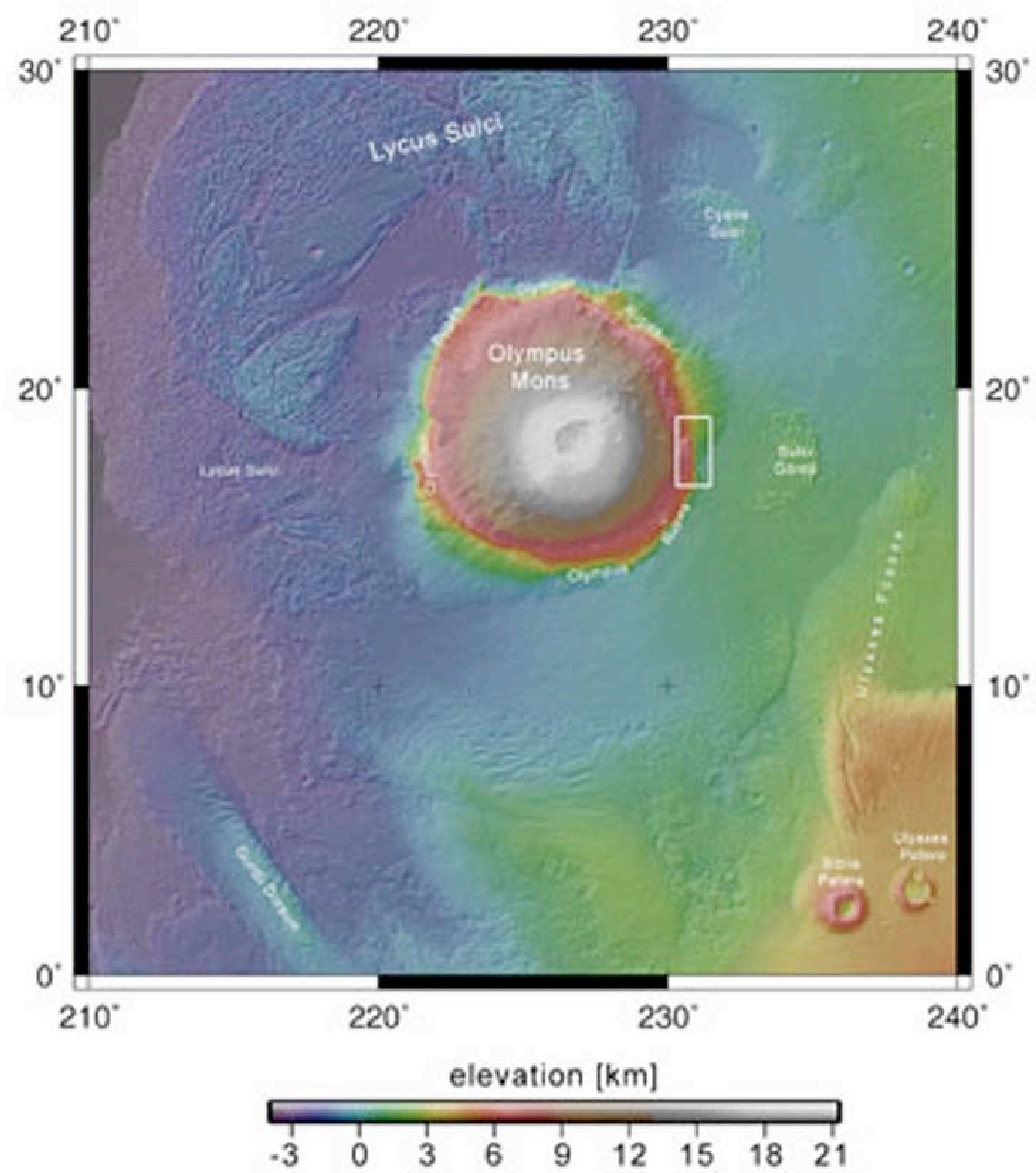
High-speed: 60 pps or more
(e.g., 300 pps at least for analyzing “slow” human motion)

High-resolution: at least 1Mpixel resolution (each picture)

The coming generation of (high-end) digital video cameras?

And: all that wireless, please

Here: the WL image sequences by NASA was "good"

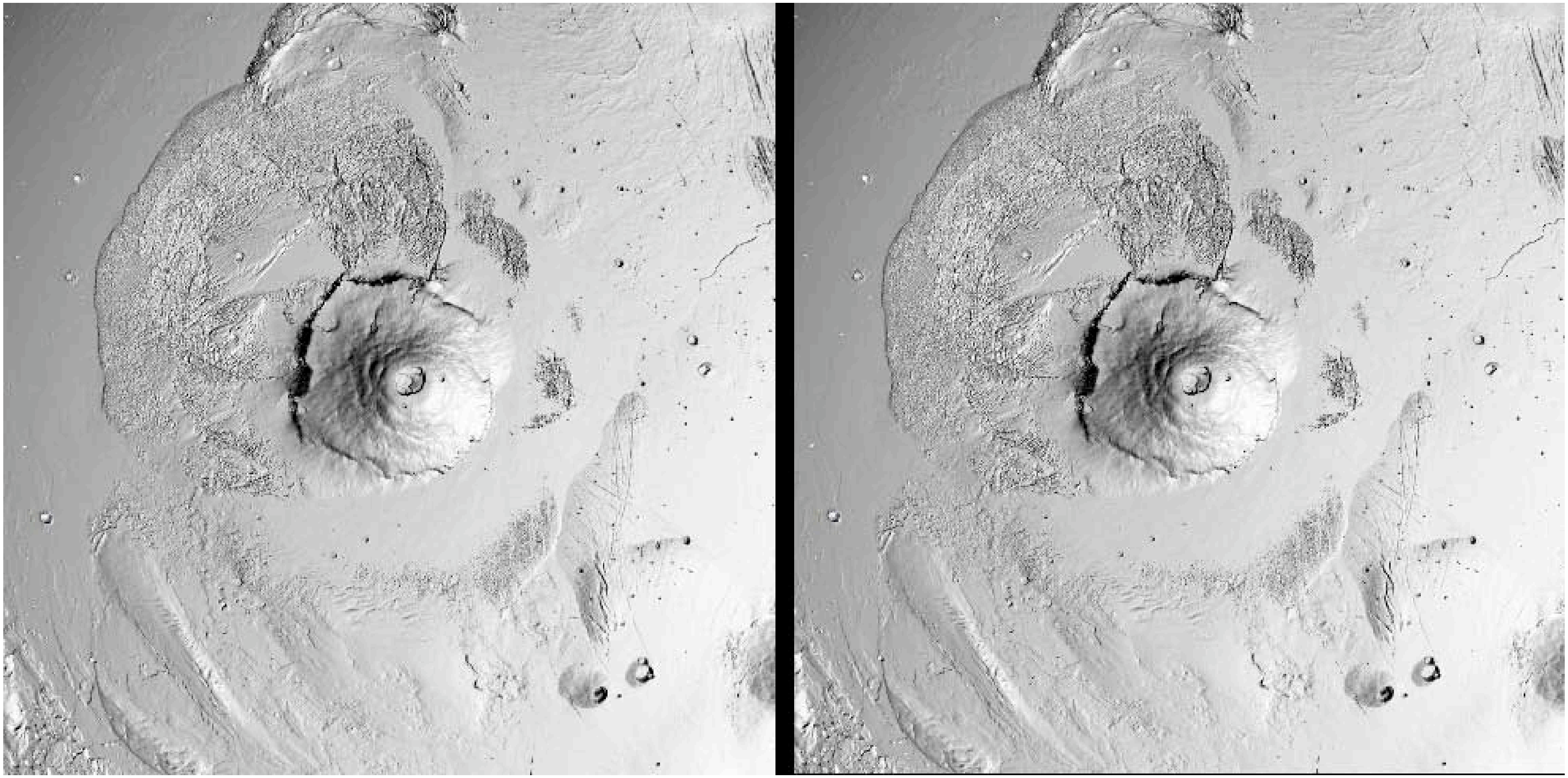


Olympus Mons (on Mars)

D = 500 km

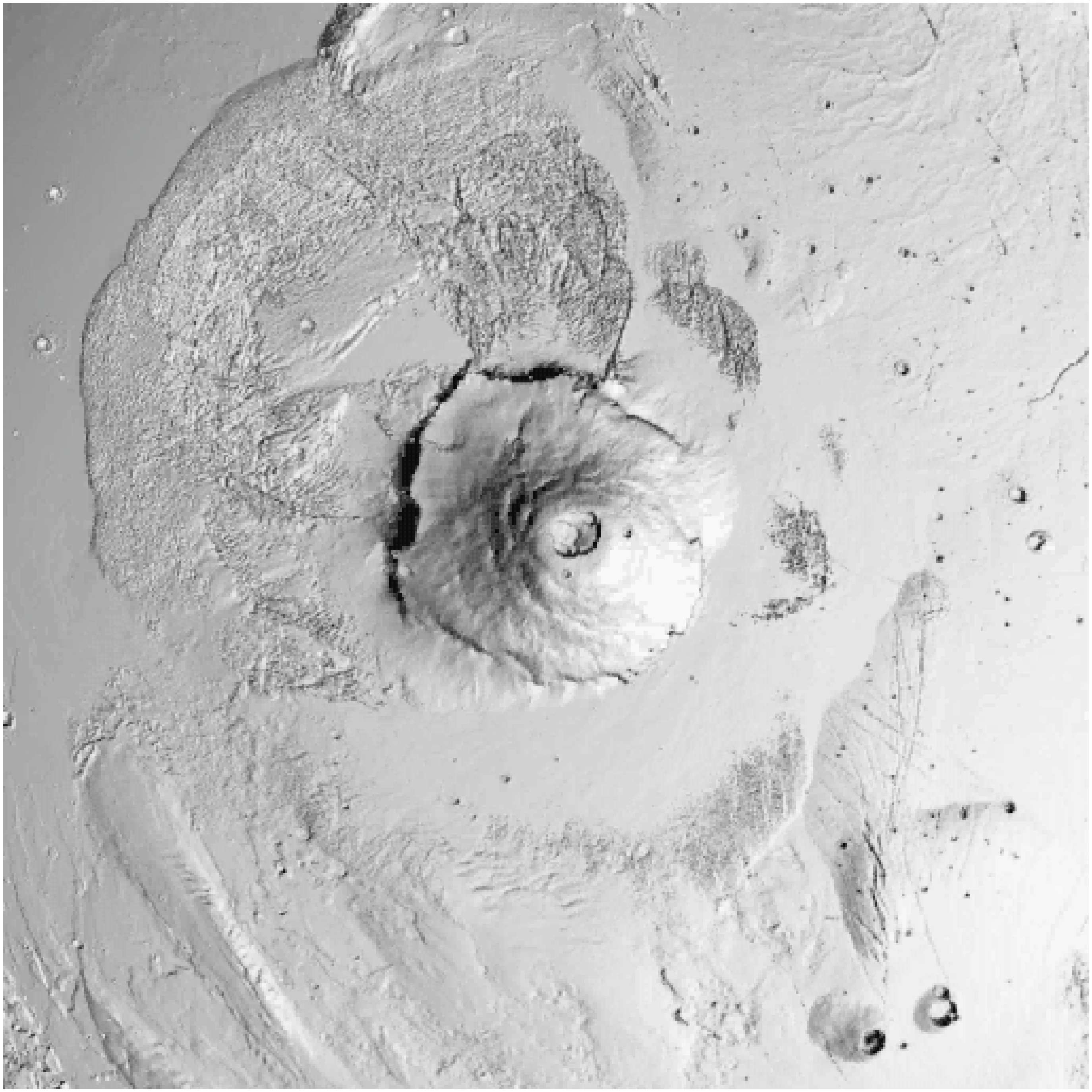
H = 25 km

Largest known volcano in our solar system.



Two input images downloaded in 2002 from a NASA website:

Now applying the computer vision approach as before, but using estimation of fundamental matrix instead of having extrinsic camera calibration available.



3D animation
CITR (2003)
based on
wireless "video"

Other 3D models
on the net:

- NASA (2004)
- DLR (2005)

4. Conclusions

WL video is a new challenge for computer vision.

Personally I am not dreaming about networks of WL cameras, but about applications such as:

- have WL video flying around large sculptures and generate an accurate 3D model of those
- have WL video "diving" into places where we do not have access otherwise
- use WL video for helping people, for example when searching for lost people on sea
- and so forth.

WL digital image sequence transmission has already a history in remote sensing, and this provides a valuable source.

Interesting new challenges are defined by

・ **extreme viewing conditions** (such as from a helicopter, or inside of a rotating object)
・ use of computer vision for **stabilizing or enhancing** transmitted video signals,
・ automated extractions of **relevant segments**, possibly combined into a purpose-designed movie, or
・ applications in computer vision using the new technology, for example for **3D shape recovery** using wireless video from a helicopter

Digital wireless video technology developed since 2000 for the general consumer market (e.g., surveillance), and this will be the real driving force.