BUILDING THE THIRD DIMENSION USING CONSUMER-MARKET DIGITAL CAMERAS, CAMCORDERS, AND PHONES

Yuan-Fang Wang, Ph.D.*
Professor, Department of Computer Science
University of California
Santa Barbara, CA,USA

Founder
VisualSize Inc.
Goleta, CA

* Research supported in part by contracts and grants from US Army and US Navy

VisualSize

Introduction

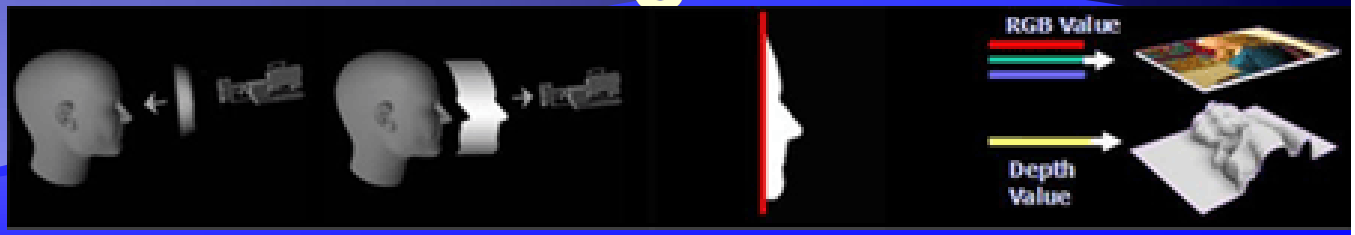
- Convergence of
  - Consumer-market cameras, camcorders, cell phones
  - Computational power
  - Storage capability
  - Communication bandwidth (Web)
- Rubber-meets-the-road validation and commercialization of Computer Vision algorithms
  - Over one billion cell phones and another billion digital cameras sold world wide each year, what cool CV algorithms can be used on them?

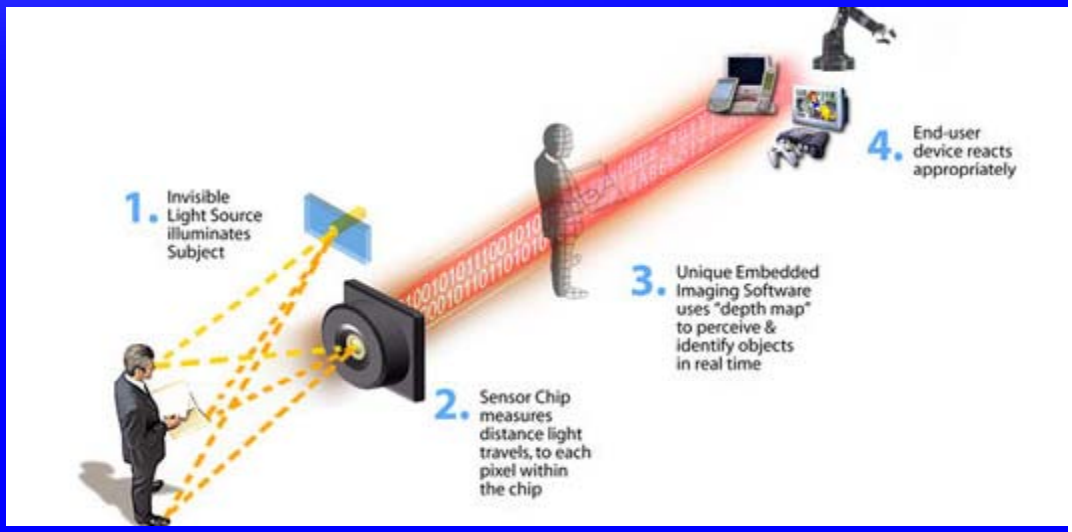
Third Dimension

- Digital photographs record appearance explicitly
- 3D objects usually have distinct
  - Appearance
  - Structure, and
  - Behavior traits
- Recovering the structure and behavior traits from mass-market camera pictures
  - Mostly about structure traits in this talk
  - Behavior (deformation) is much harder (e.g., behavior modeling of tumors in computer-assisted colonoscopy)

Alternative: Active Range Cameras

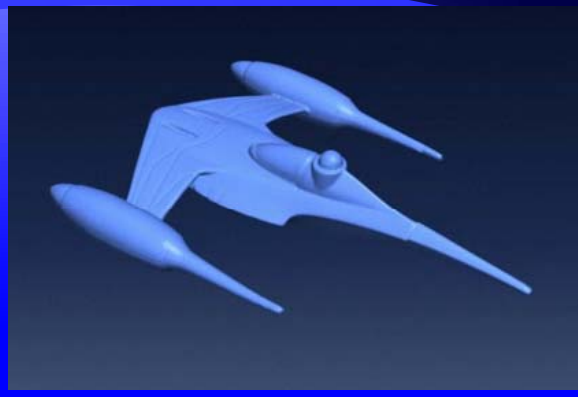
- Principles
  - Time of flight
  - Structured light
  - Phase shift detection
  - Laser, LCD
  - ASC, 3DV (Microsoft), PrimeSense (Microsoft Xbox), Canesta





Alternative: Active Range Cameras

- NextEngine ($2,995)
- escan3D ($7,995)
  - Sweeping laser line with triangulation



Active Range Cameras

- Niche markets in the foreseeable future
  - Cost
  - Size
  - Selection
  - Spatial resolution
  - Power consumption
  - Scanning speed
  - Availability of public-domain data

Alternative: Specialized Stereo Video Camera

- Stereo sensors (tyzx)
- Pros: Real-time
- Cons:
  - Narrow-base-line stereo
  - Poor depth resolution
  - No cross validation of 3D depth
  - Bulky, expensive, one-of-a-kind
  - Not consumer-market

VisualSize

Alternative: A Single Photograph

- Photowoosh
- Make3D (http://make3d.stanford.edu/)
  - Expensive, time-consuming off-line learning
  - Manual image marking on-line
  - Qualitative, coarse depth profile with significant error




Our Philosophy

- 3D Inference is an inherently "ill-posed", inverse problem
- Many unknowns, not enough constraints
- Solution 1:
  - Clever algorithms
  - Past experience (learning and inference)
- Solution 2:
  - Hard data (more images)
- Our claim: hard data trounce clever algorithms
  - Minuscule effort in data collection
  - Readily available computational power and storage space
  - Specificity of information enables practical, robust and efficient CV algorithms

Our Goals

- Consumer-market
- Multiple capabilities (one stop shopping)
- Hardware:
  - No calibration, specialized equipment used
- User:
  - point-shoot-upload, no training or expertise, arbitrary sensing configurations
- Complete systems, fully automated, end-to-end
- Avoid third-party licensing requirements

Visualsize's Many Solutions

Image stitching
Panorama building

Panorama Engine

Spatially aware
Image browsing

PhotoNav3D

Metrology

Metrology Engine

Discrete 3D structures

Dense 3D structures

PhotoModel3D

Structure Motion

Camera motion only

Camera motion + User-specified structure

Camera motion + Sparse structure

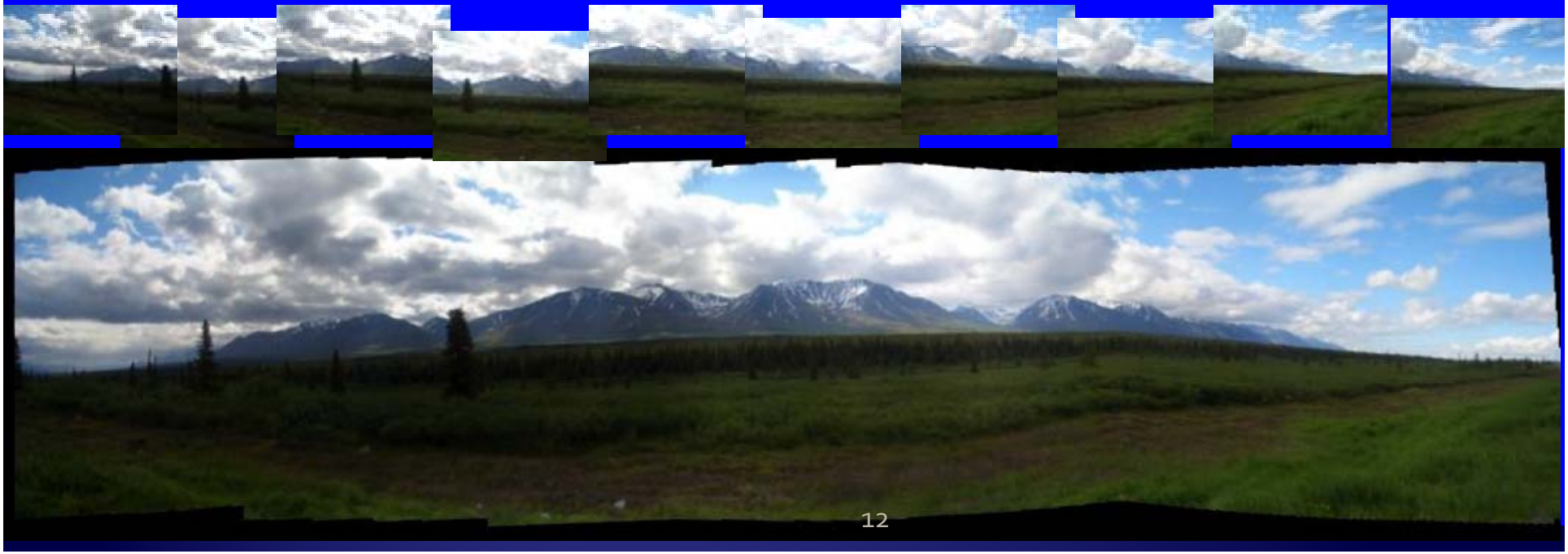
Camera motion + Dense structure

VisualSize

Panorama Building

http:/localhost/mosaic3d/index2.php

- A simple, global image registration method
- Pixel movements are explained by a single model (homography transformation)
  - Rotational only camera motion
  - Far-field images



Panorama Building (cont.)

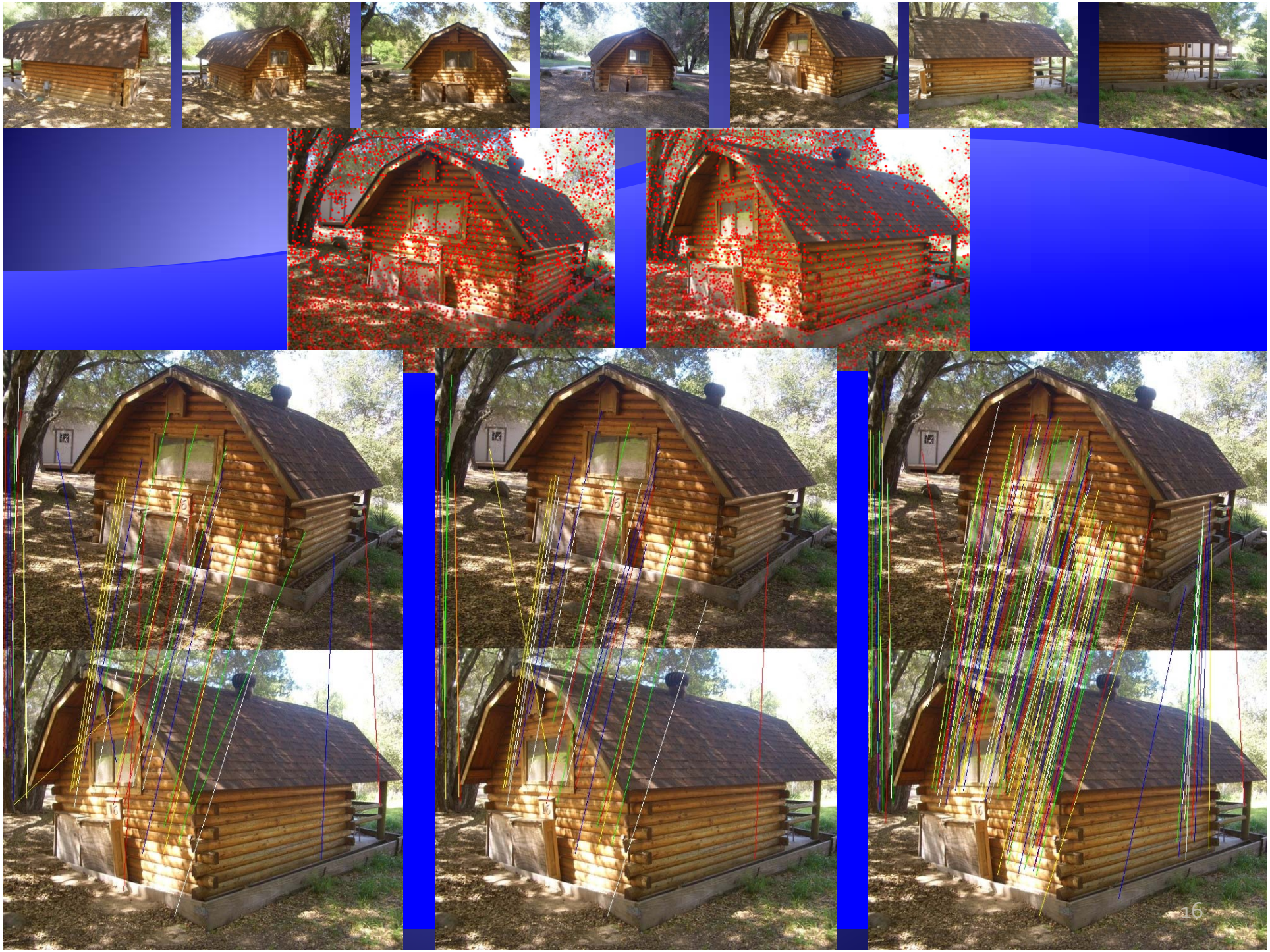
- How to address accumulation of registration error?
- How to estimate intrinsic camera parameters (cameras are not explicitly calibrated)?

Comparison (competitor) Analysis

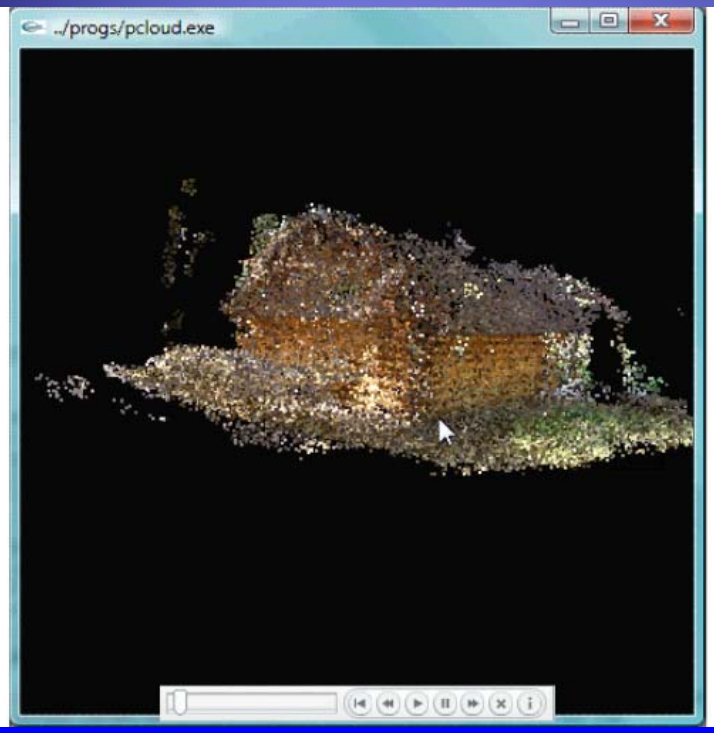
- Too many (PanoramaFactory, EasyPano, Autopano, Microsoft)
- “Me-too” technology
- Distinction
  - Web-based (Face book application)
  - Part of a complete 3D suite of algorithms, one-stop shopping

When Stitching is not Enough

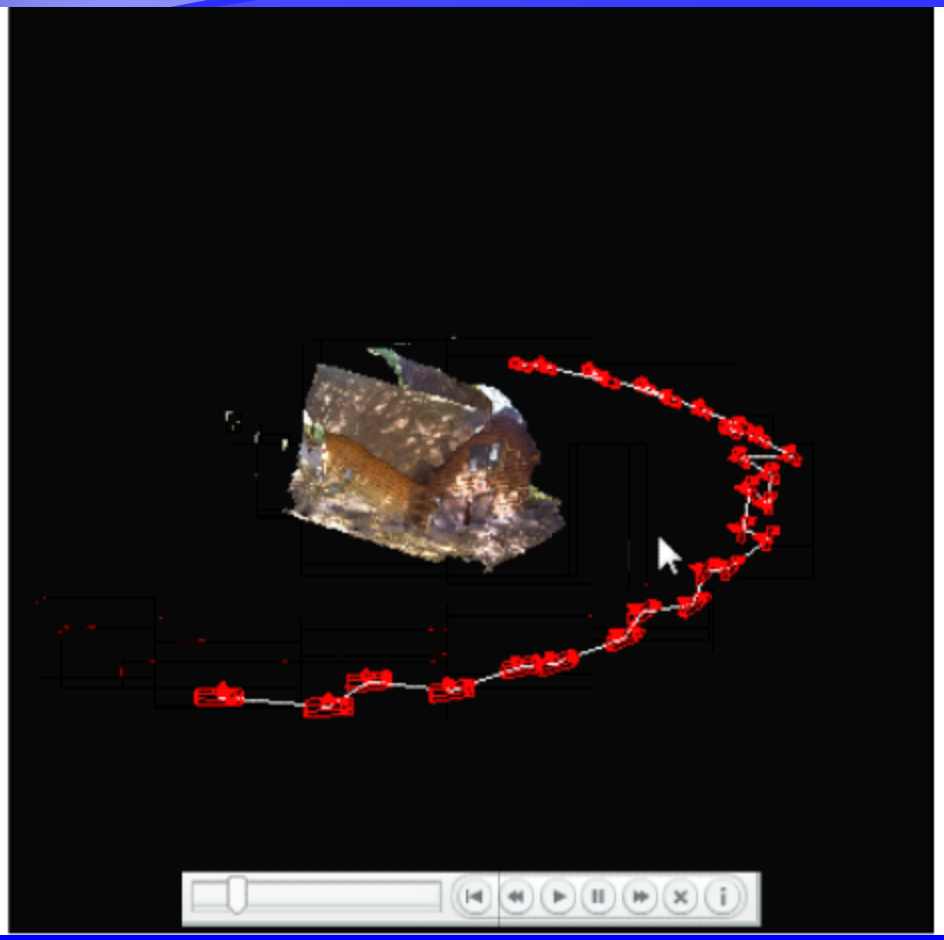
- CV + CG
  - Structure from motion
  - Discrete snapshots
  - Matching
  - SBA
  - Off–line
  - Batch
  - Dense maps
- Robotics
  - SLAM – simultaneous localization and mapping
  - Continuous video
  - Tracking
  - Extended Kalman Filter
  - On-line
  - Incremental
  - Sparse maps



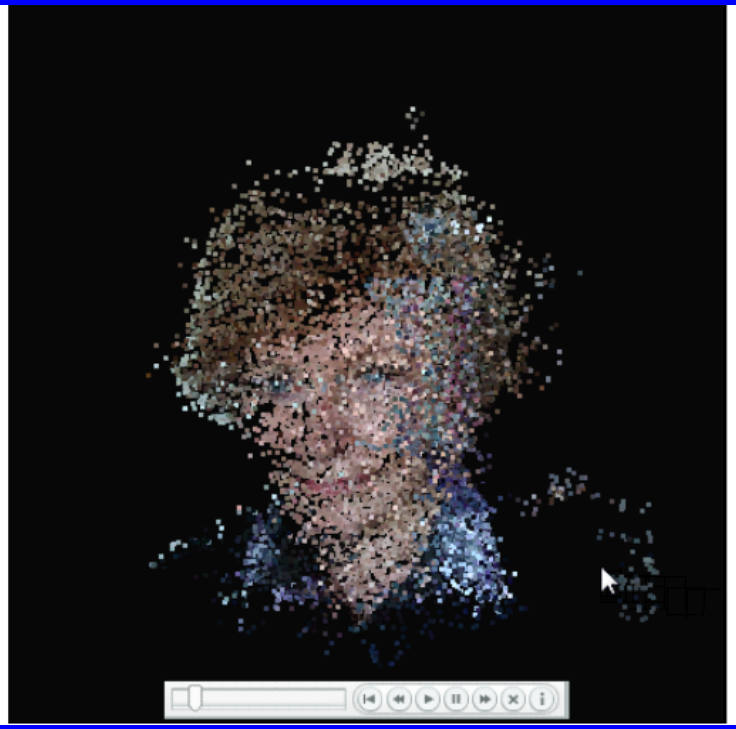
Object Structure



Camera Motion



More Examples



Difficulty - Robustness

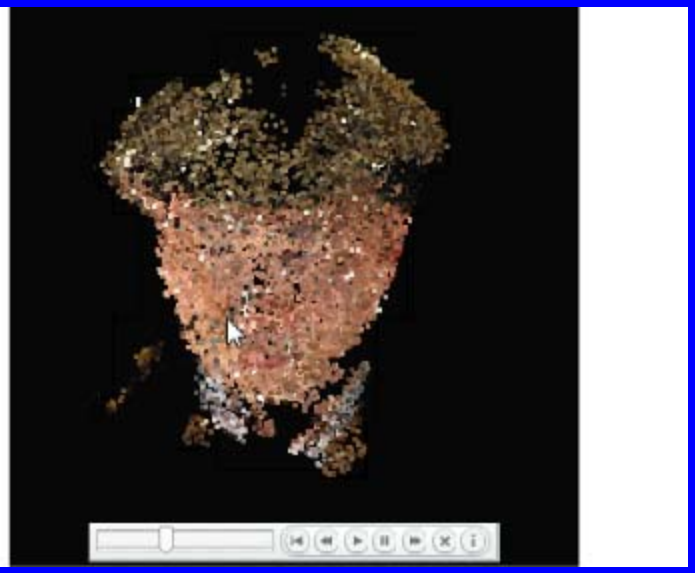
- Ill-posed, inverse problem
- No explicit camera calibration for consumer markets
- Unknown (partially-known) camera intrinsic parameters (whatever in JPEG header)
- Noise in feature locations
- Outliers
  - "Obvious" – those violate epipolar constraints
  - "Subtle" – those satisfy epipolar constraints (stereo cannot handle this!)
- Numerical stability

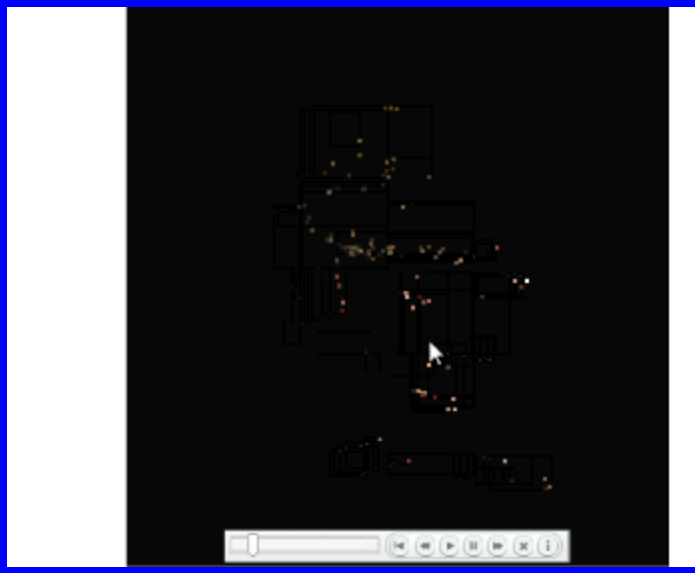
Difficulty - Efficiency

- Many variables
  - 6 (extrinsic) + >4 (intrinsic) for each camera shot
  - 3 (x,y,z) for each feature point

Trade-off

- Dense 3D point clouds
  - Slow, less robust
  - High recall, low precision
  - High false-positive, low false-negative
- Sparse 3D point clouds
  - Fast, more robust
  - Low recall, high precision
  - Low false-positive, high false-negative



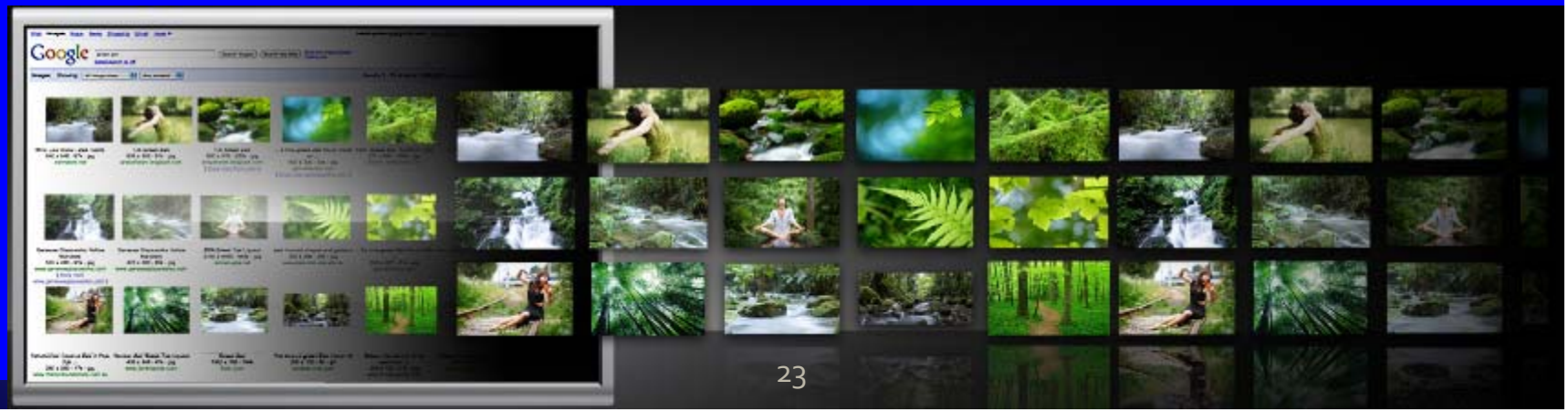


- Key contribution: **Robustness + Density + Efficiency**

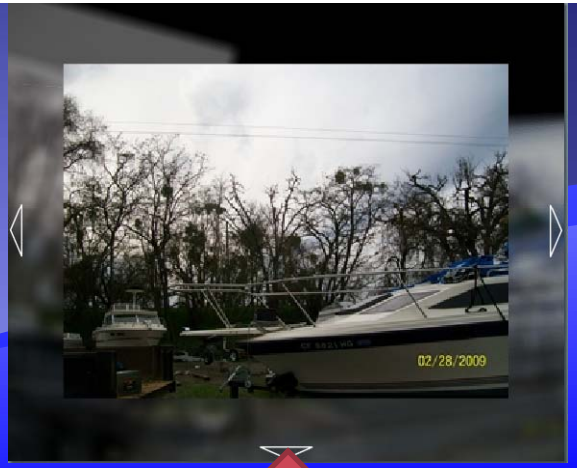
VisualSize

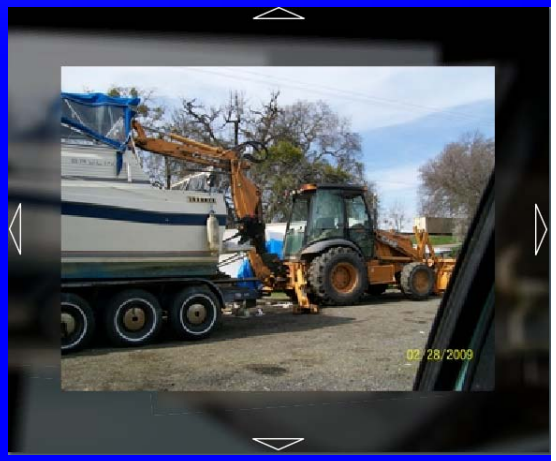
Spatially-Aware Image Browsing

- Recovered camera motion parameters give camera trajectory and view similarity
- Browse an image collection based on spatial adjacency & view similarity of the camera
- More flexible than panorama
- More robust than 3D models
- Cf cooliris.com (pretty graphics, no CV)



PhotoNav3D





VisualSize

http://localhost/photonav3d/viewer/tool.1.html

















http://localhost/photonav3d/viewer/test3.1.html

VisualSize

Comparison (Competitor) Study

http:/localhost/photonav3d/summary.html

- Against Microsoft Photosynth (photosynth.net)
- 27 data sets and over 800 images
  - Indoor and outdoor
  - Near-field, median-field, and far-field
  - Inside-out and outside-in
- How many images are reached (navigable)
- Beat Photosynth significantly (>40%) in 15
- Beat Photosynth slightly in 2
- Tie Photosynth in 9
- Slightly worse than Photosynth in 1 (22 vs. 23 photos)

3D Models

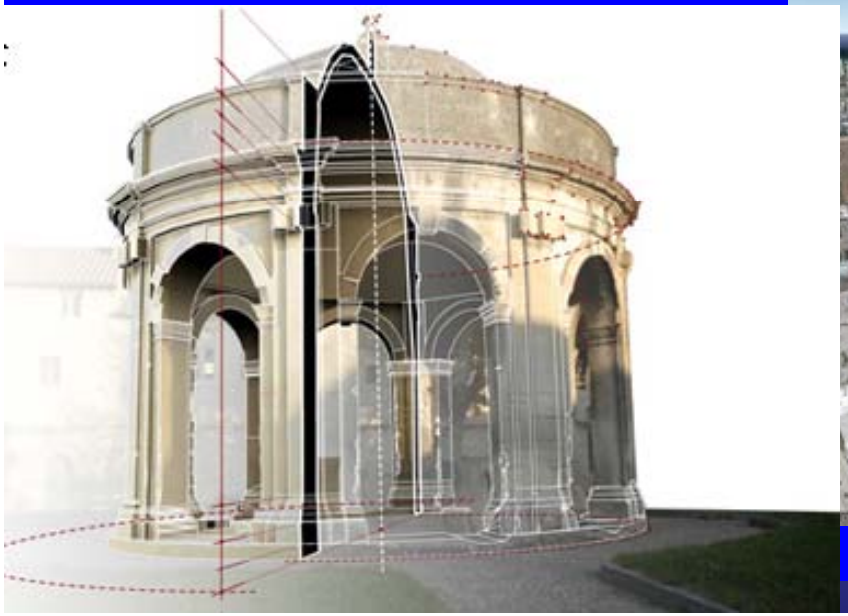
http://localhost/3ddemo/index.php

- Sparse (point cloud)
  - 3D positions of tracked/matched features
- Dense (textured surface)
  - Depth per pixel

Comparison (Competitor) Study

- Google Sketchup, AutoDesk Image Modeler
- Mostly for architectural design
- Interactive, extensive human interaction
- Steep learning curve

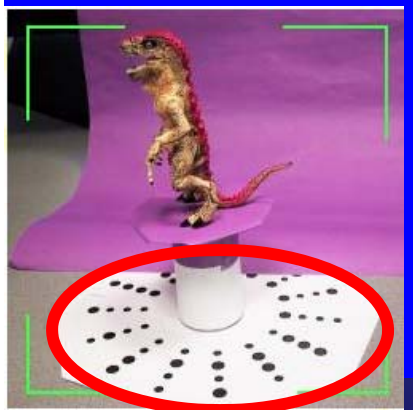
- PhotoModel3D
- Any 3D objects
- Fully automated
- Point, shoot, upload



Comparison (Competitor) Study

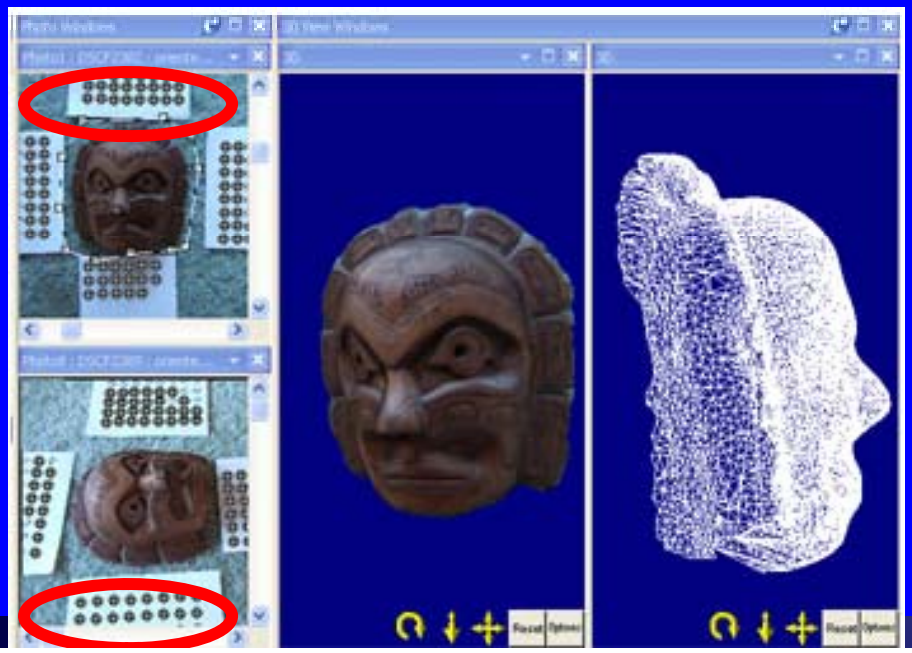
- 3dsom.com, strata.com
- Special registration markers
- Blue screen segmentation
- Silhouette-based volume intersection
- Interactive, extensive human interaction
- Small, complete objects

- PhotoModel3D
- Any 3D objects
- Fully automated
- Point, shoot, upload



Comparison (Competitor) Study

- PhotoModler
- Special registration markers
- Manual feature selection and registration
- Dated two-view stereo analysis (with manual interaction)



- PhotoModel3D
- Any 3D objects
- Fully automated
- Point, shoot, upload

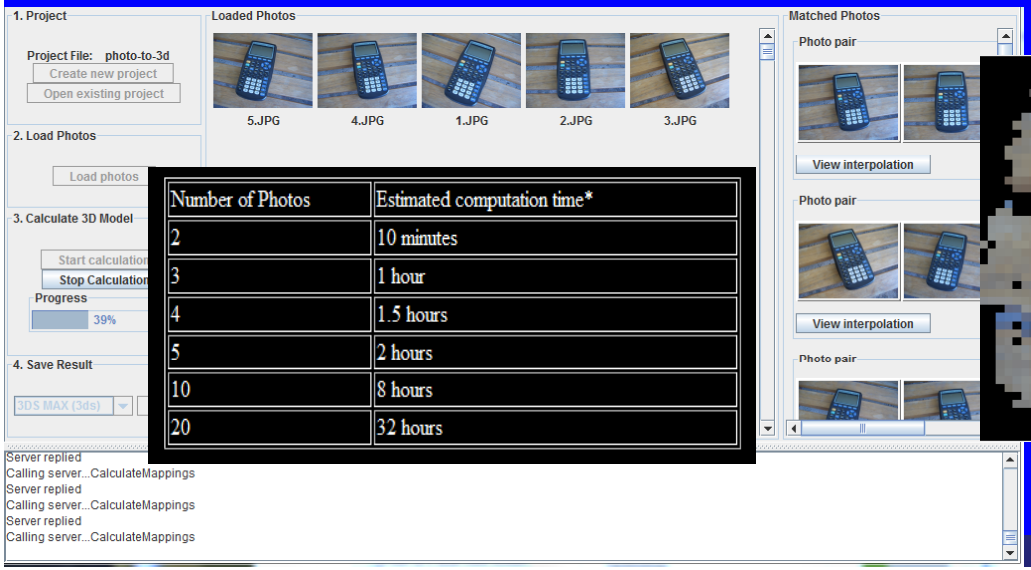
Comparison (Competitor) Study

Photo-to-3D.com

- Commercial services ended (licensing issues?)
- Slow, a simple 5-image VGA data (calc) took more than 1 hour (30 sec for PhotoModel3D)
- Dated stereo pair-wise analysis

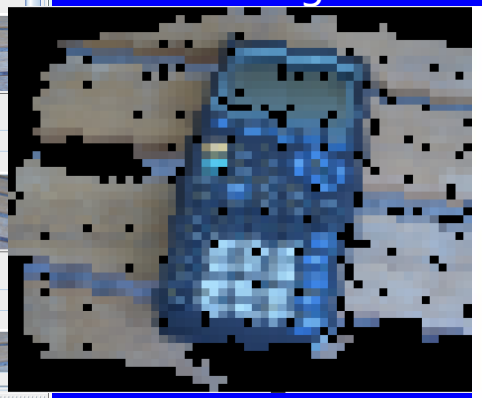
PhotoModel3D

- Any 3D objects
- Fully automated
- Point, shoot, upload
- Doesn't take forever ☺



Number of Photos	Estimated computation time*
2	10 minutes
3	1 hour
4	1.5 hours
5	2 hours
10	8 hours
20	32 hours

Photo-to-3d



PhotoModel3D

VisualSize

Comparison (Competitor) Study

- ARC 3D webservice
  - http://homes.esat.kuleuven.be/~visit3d/webservice/v2/download.php
  - Maarten Vergauwen and Luc Van Gool, "Web-Based 3D Reconstruction Service", Machine Vision Applications, 17, pp. 411-426, 2006
- Outdoor architecture scenes
  - Mostly planar surfaces
  - Feature-rich façade
  - Partial construction
  - Programs not working
  - Not responding to email

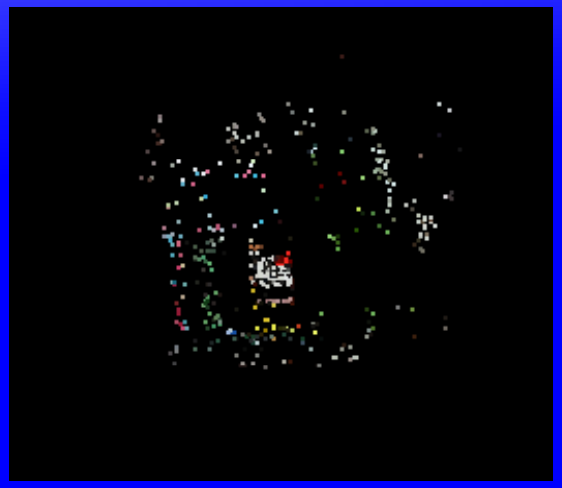
Comparison (competitor) Study

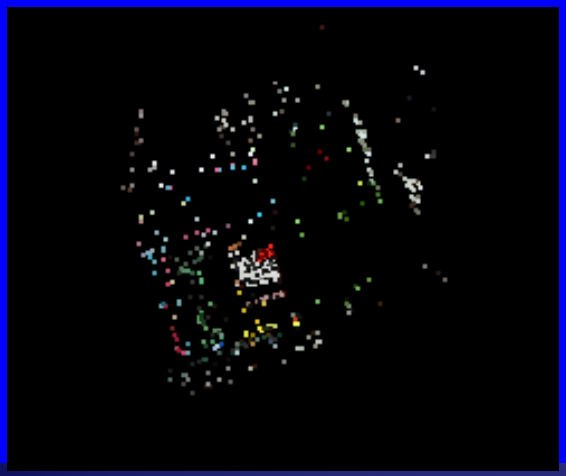
- Bundler (core of UWash/Microsoft Phototourism)
  - http://phototour.cs.washington.edu/bundler/
  - Noah Snavely, Steven M. Seitz, Richard Szeliski. **Photo Tourism: Exploring image collections in 3D**. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2006)*, 2006.
    Noah Snavely, Steven M. Seitz, Richard Szeliski. **Modeling the World from Internet Photo Collections**. *International Journal of Computer Vision* 2007.
  - Two standard Bundler data sets: Kermit and ET
- On an unloaded PC (Intel Core 2 Duo 2.66GHz, 2G) – only one core is used, no GPU acceleration

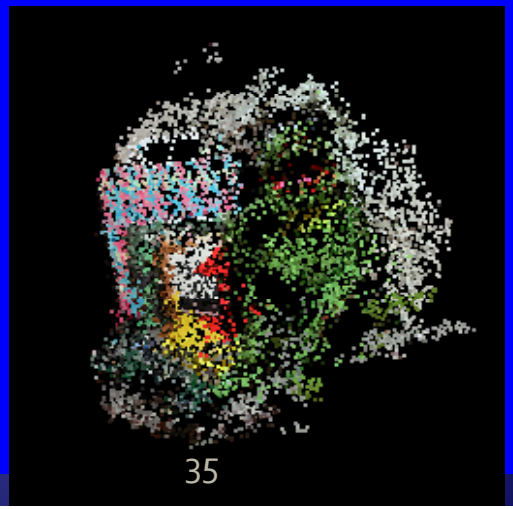
Runtime	# of images	Ours	Bundler
Kermit	11	1:09	0:44
ET	9	1:23	0:47
Knight	16	1:38	1:02

Density (# of 3D points)	# of images	Ours	Bundler
Kermit	11	8649	623
ET	9	8699	514
Knight	16	7381	412

Bundler

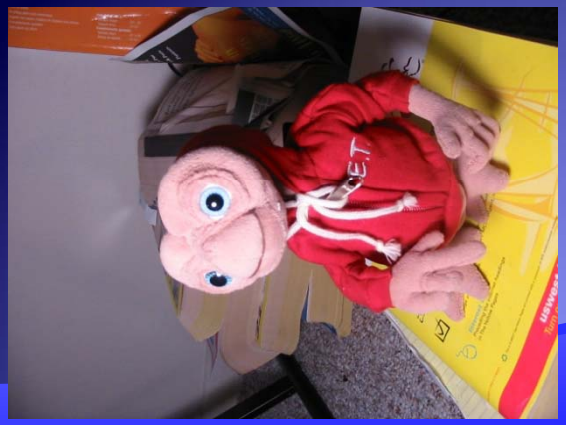


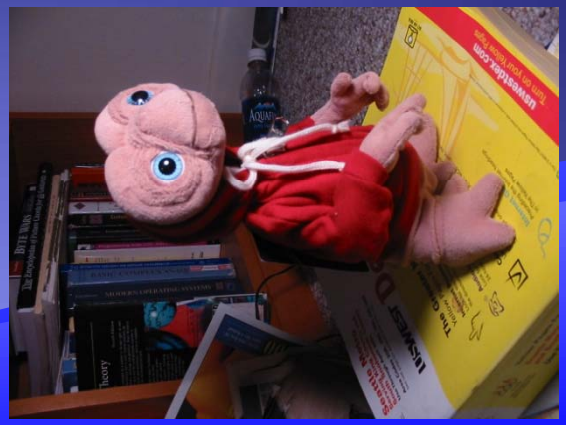




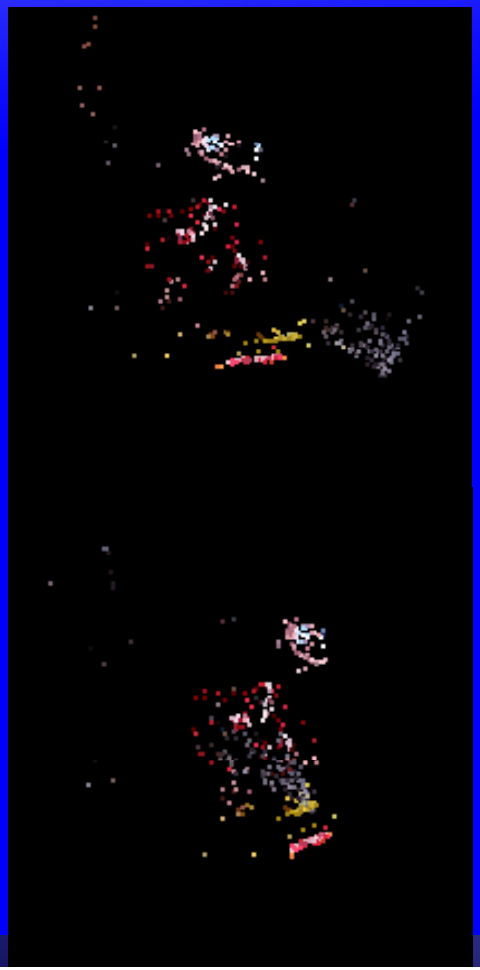
VisualSize

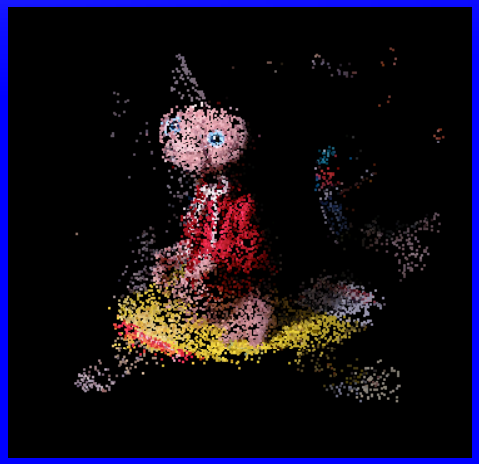
VisualSize

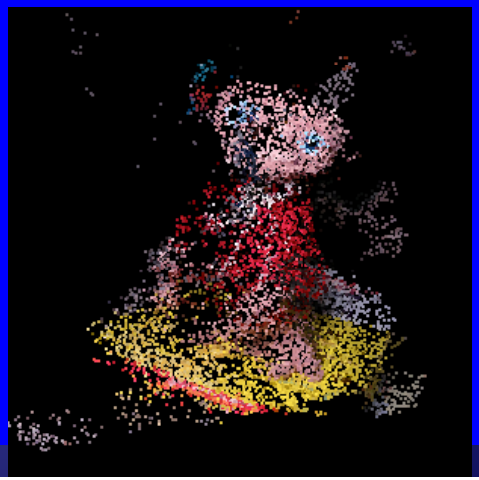




Bundler



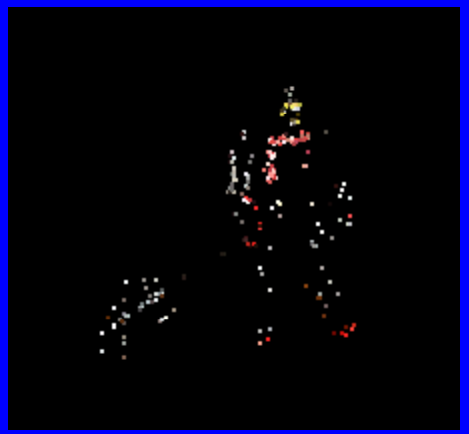




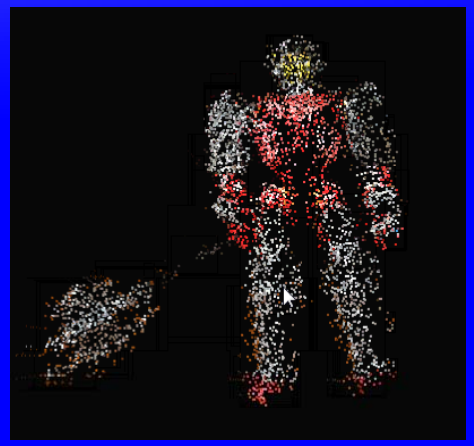
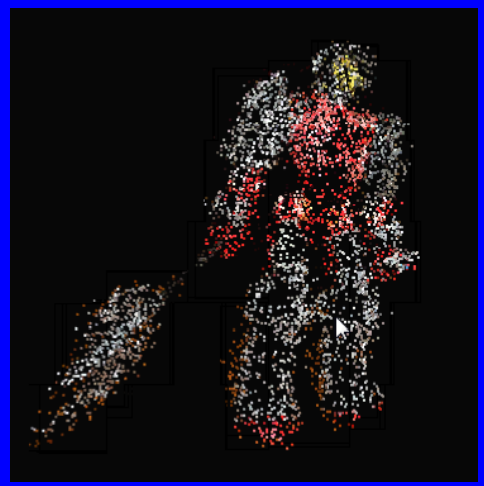
VisualSize

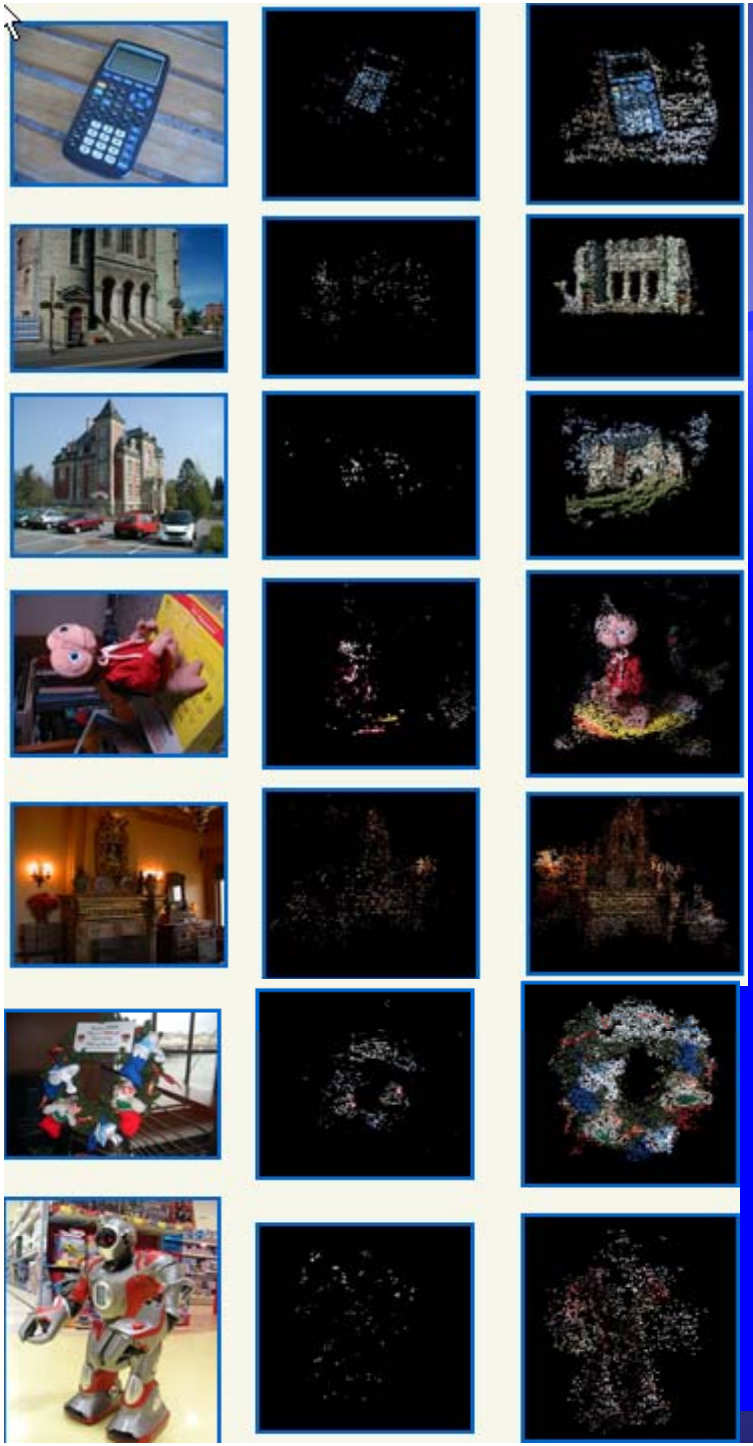


Bundler

VisualSize



Comparison

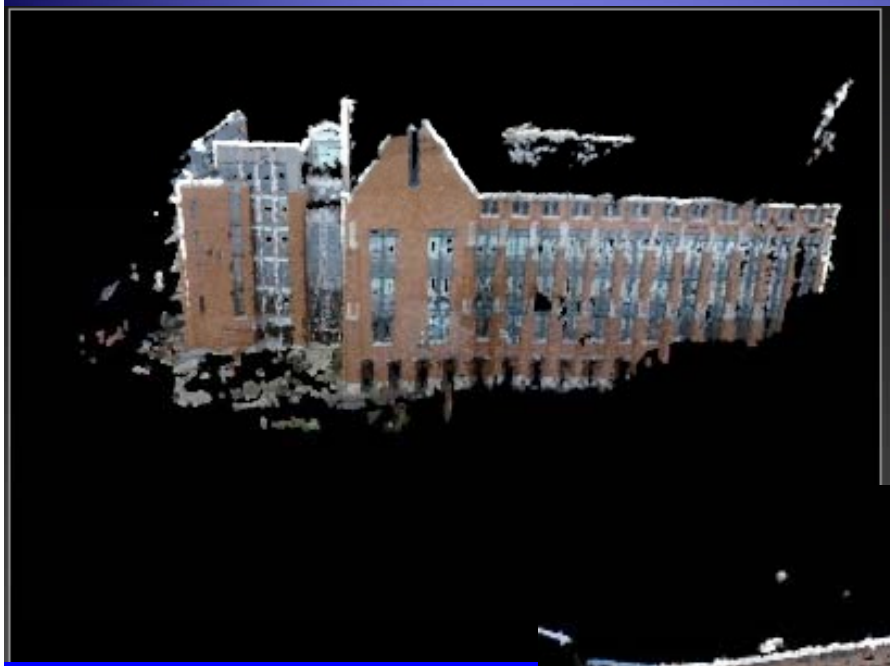
http://http://localhost/3ddemo/true3d/comparison/index.html/

- 30 data sets
- Indoor, outdoor, partial, 360^o
- As few as 5, 6, 7 images, as many as 64, 72, 88
- Uniform trend
  - Bundler is faster (~2x)
  - Ours is denser (~10x)
  - Ours requires orders-of-magnitude less # of photos

Left: sample images
Middle: Bundler's models
Right: Visualsize's models

VisualSize

Comparison with PhotoCity

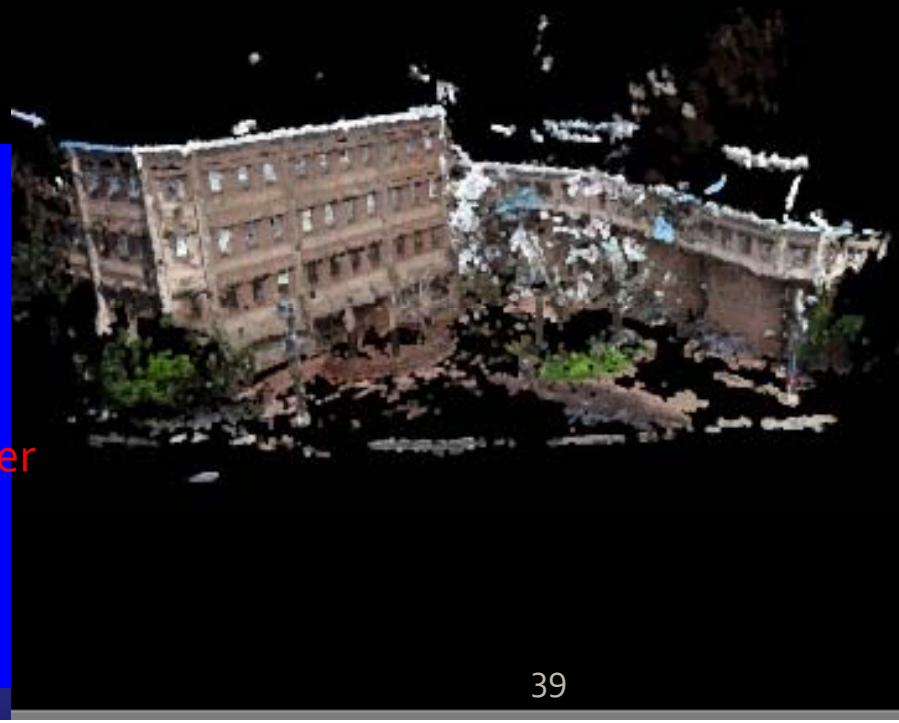


CSE Front Entrance

552 photos
151366 points

View in 3D

Bundler



Engineering Library

806 photos
218724 points

View in 3D

Bundler

Comparison with PhotoCity

Visualsize:
42 images, 99,937points

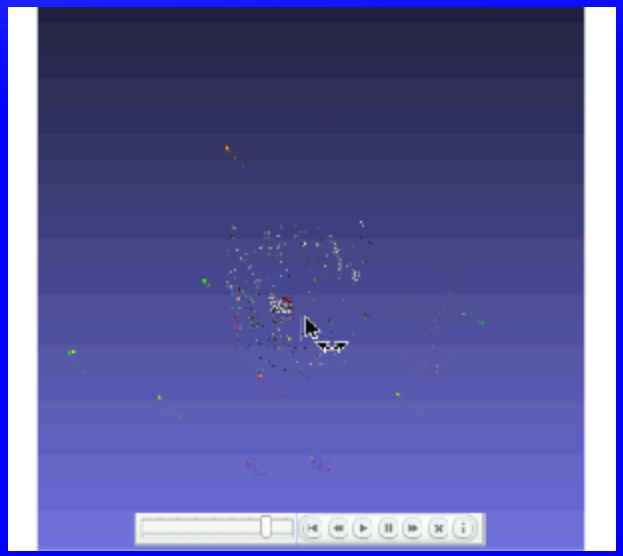
Visualsize:
34 images, 71,072 points

VisualSize

Comparison with Bundler + PMVS2

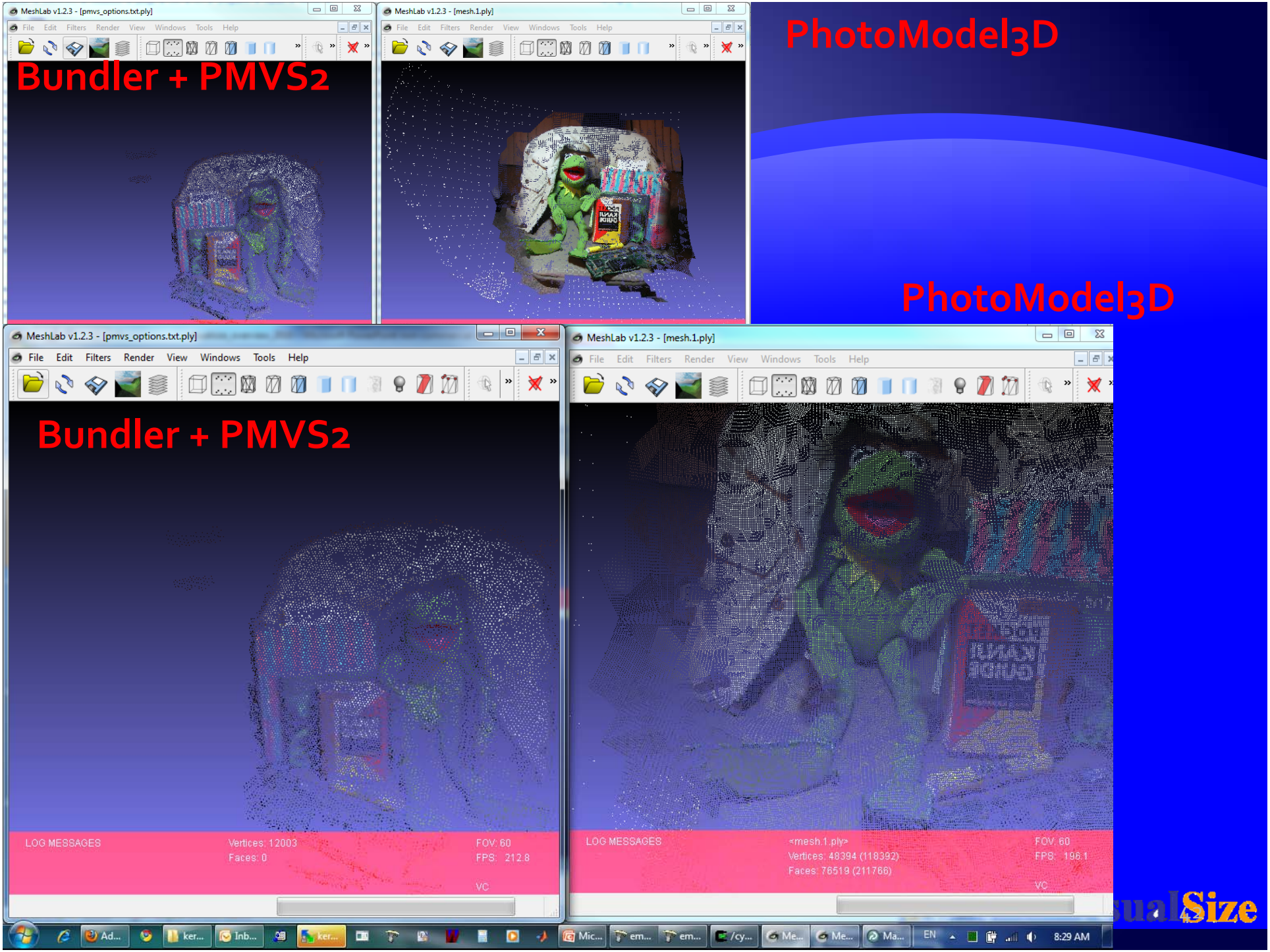
http://localhost/3ddemo/true3d/comparison/index.html#_PMVS2

- Bundler (Version 0.4, April 10, 2010): 0:46
- PMVS2 (July 13, 2010): 0:53
- Total (Bunlder+PMVS2): 1:39
- PhotoModel3D: 2:03



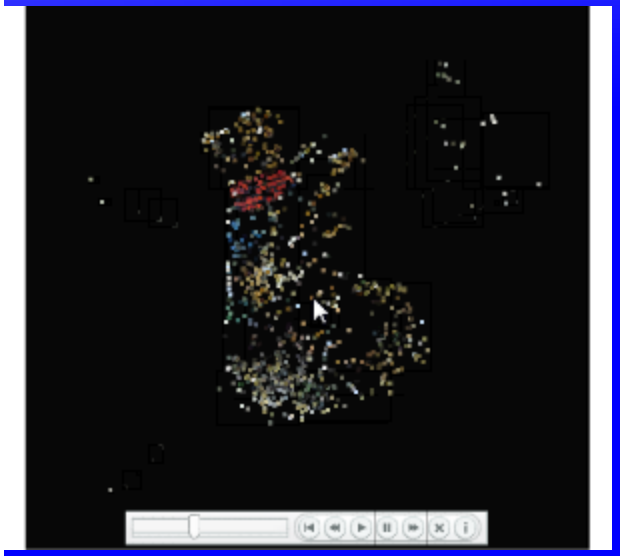
11 images, On an unloaded notebook (Intel Core 2 Duo 2.8GHz, 4G RAM) – only one core is used, no GPU acceleration

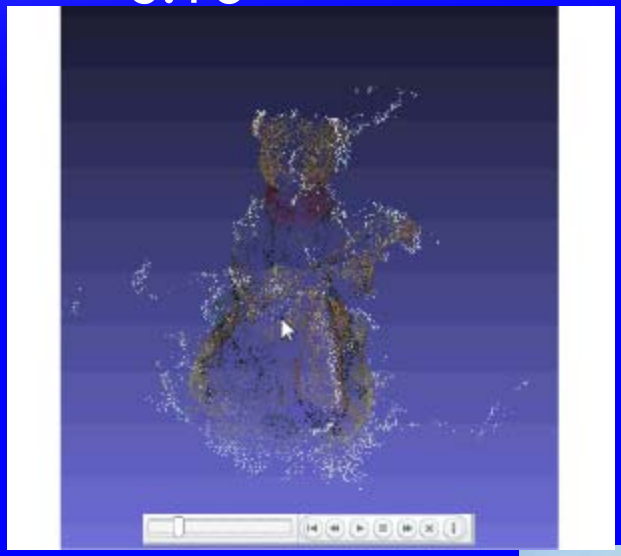
VisualSize


MeshLab v1.2.3 - [pmvs_options.txt.ply]
MeshLab v1.2.3 - [mesh.1.ply]
Bundler + PMVS2
PhotoModel3D
PhotoModel3D
Bundler + PMVS2
File Edit Filters Render View Windows Tools Help
LOG MESSAGES
Vertices: 12003
Faces: 0
FOV: 60
FPS: 212.8
VC
LOG MESSAGES
<mesh.1.ply>
Vertices: 48394 (118392)
Faces: 76519 (211766)
FOV: 60
FPS: 196.1
VC
8:29 AM

Comparison with Bundler + PMVS2

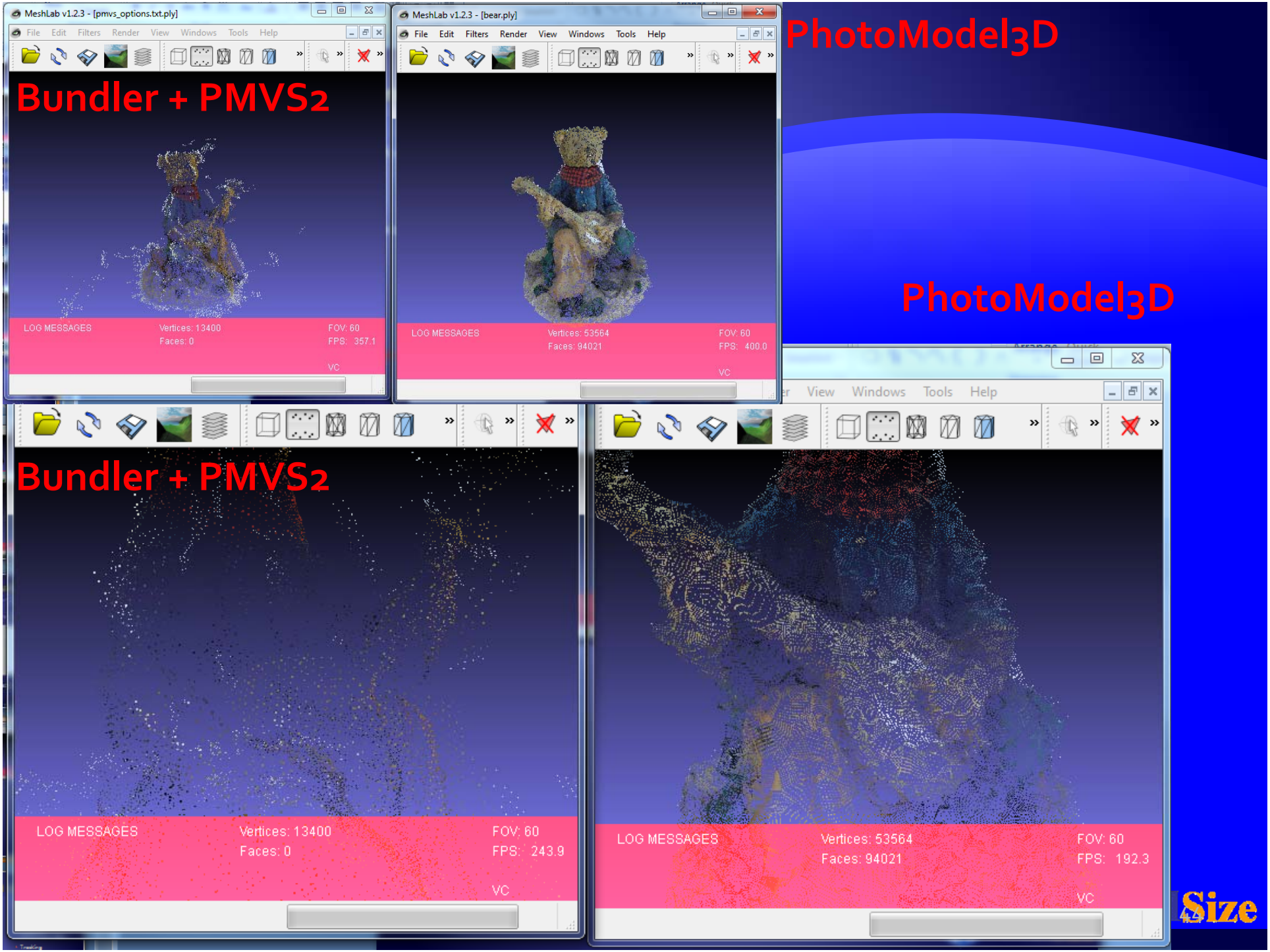
- Bundler (Version 0.4, April 10, 2010): 1:25
- PMVS2 (July 13, 2010): 1:53
- Total (Bunlder+PMVS2): 3:18
- PhotoModel3D: 2:46





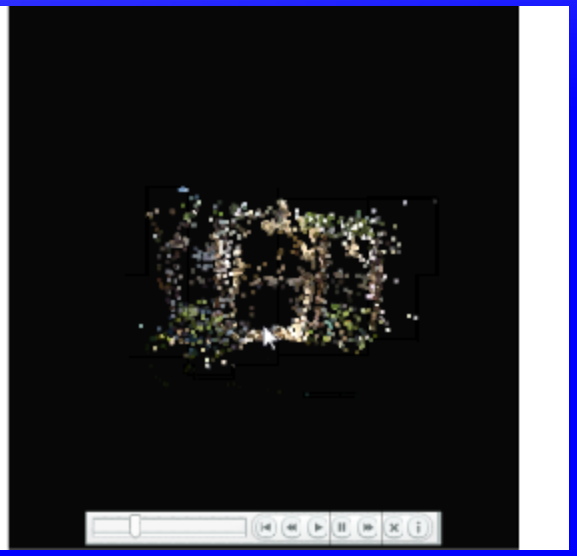
19 images: On an unloaded notebook (Intel Core 2 Duo 2.8GHz, 4G RAM) – only one core is used, no GPU acceleration

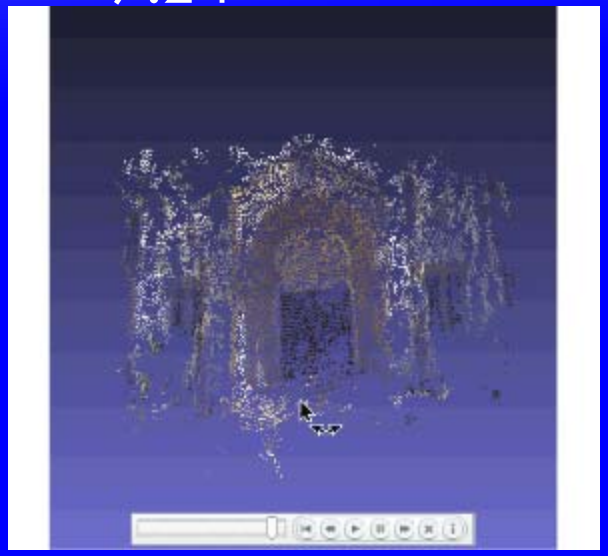
VisualSize

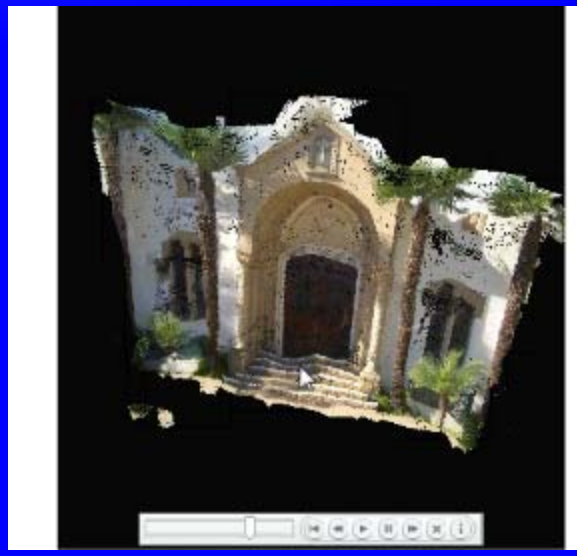

MeshLab v1.2.3 - [pmvs_options.txt.ply]
File Edit Filters Render View Windows Tools Help
Bundler + PMVS2
LOG MESSAGES
Vertices: 13400
Faces: 0
FOV: 60
FPS: 357.1
VC
MeshLab v1.2.3 - [bear.ply]
File Edit Filters Render View Windows Tools Help
LOG MESSAGES
Vertices: 53564
Faces: 94021
FOV: 60
FPS: 400.0
VC
PhotoModel3D
PhotoModel3D
Bundler + PMVS2
LOG MESSAGES
Vertices: 13400
Faces: 0
FOV: 60
FPS: 243.9
VC
View Windows Tools Help
LOG MESSAGES
Vertices: 53564
Faces: 94021
FOV: 60
FPS: 192.3
VC
Size

Comparison with Bundler + PMVS2

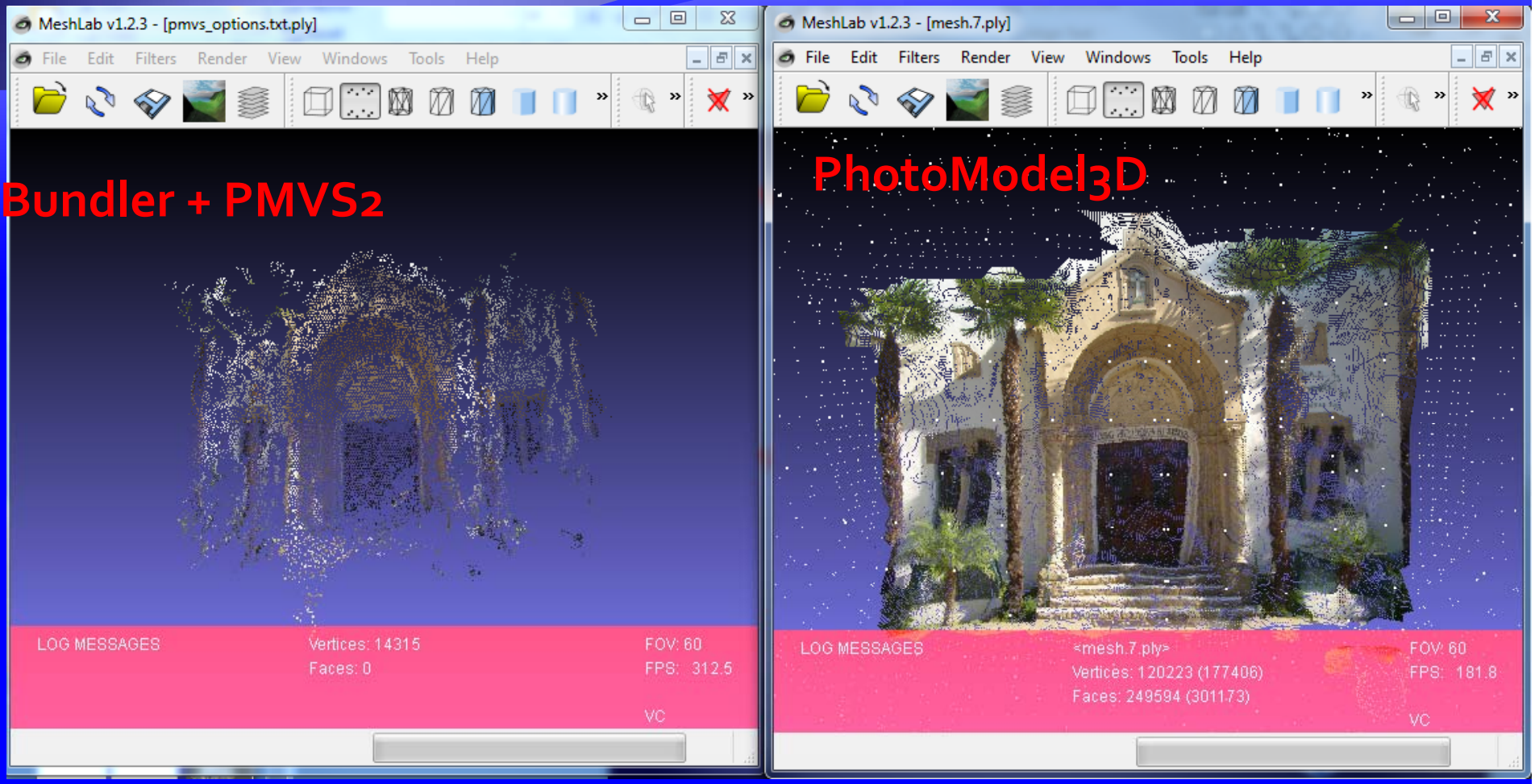
- Bundler (Version 0.4, April 10, 2010): 3:00
- PMVS2 (July 13, 2010): 4:24
- Total (Bunlder+PMVS2): 7:24
- PhotoModel3D: 10:03







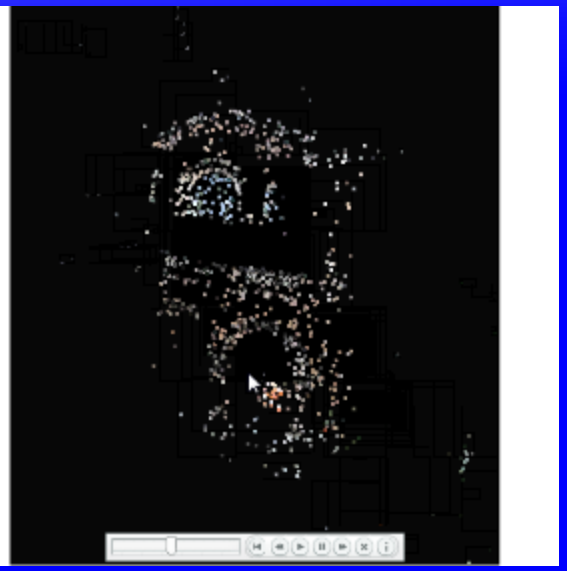
24 images: On an unloaded notebook (Intel Core 2 Duo 2.8GHz, 4G RAM) – only one core is used, no GPU acceleration

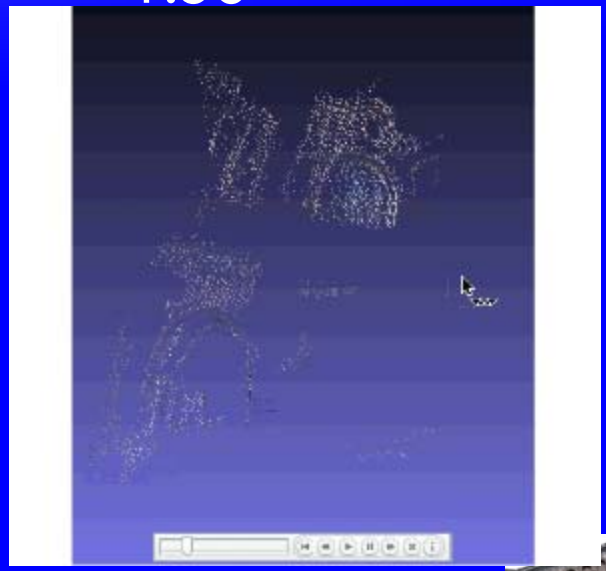

MeshLab v1.2.3 - [pmvs_options.txt.ply]
File Edit Filters Render View Windows Tools Help
Bundler + PMVS2
LOG MESSAGES
Vertices: 14315
Faces: 0
FOV: 60
FPS: 312.5
VC
MeshLab v1.2.3 - [mesh.7.ply]
File Edit Filters Render View Windows Tools Help
PhotoModel3D
LOG MESSAGES
<mesh.7.ply>
Vertices: 120223 (177406)
Faces: 249594 (301173)
FOV: 60
FPS: 181.8
VC

VisualSize

Comparison with Bundler + PMVS2

- Bundler (Version 0.4, April 10, 2010): 1:32
- PMVS2 (July 13, 2010): 0:24
- Total (Bunlder+PMVS2): 1:56
- PhotoModel3D: 4:39



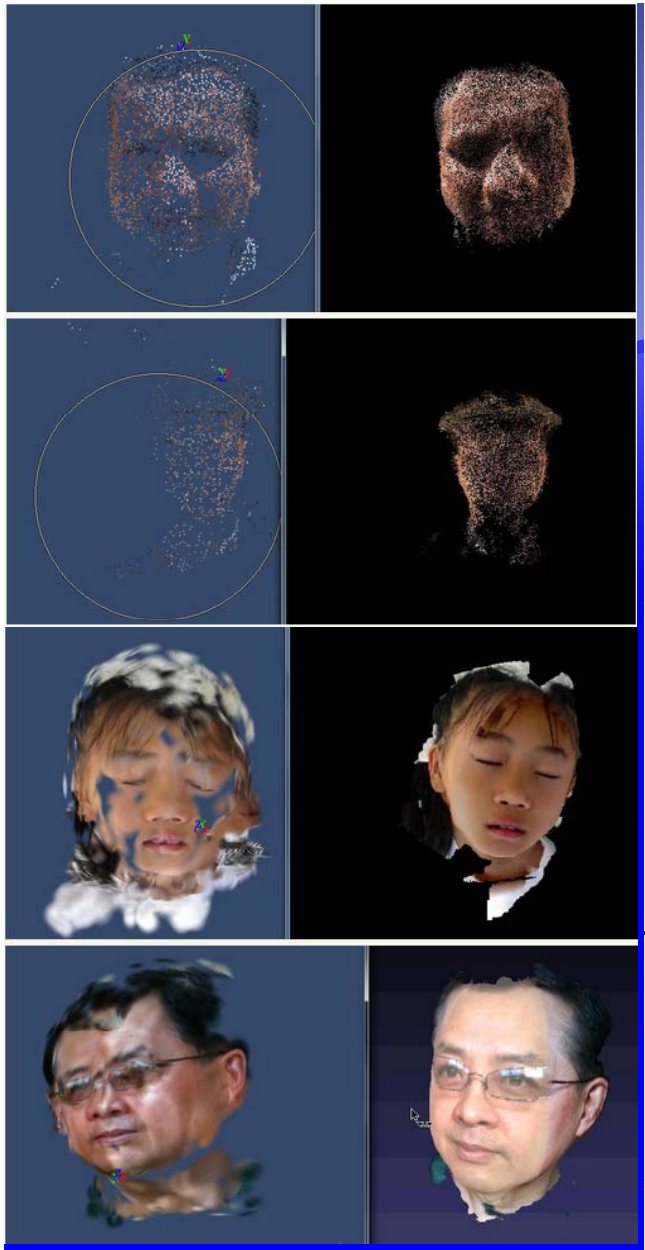


16 images: On an unloaded notebook (Intel Core 2 Duo 2.8GHz, 4G RAM) – only one core is used, no GPU acceleration

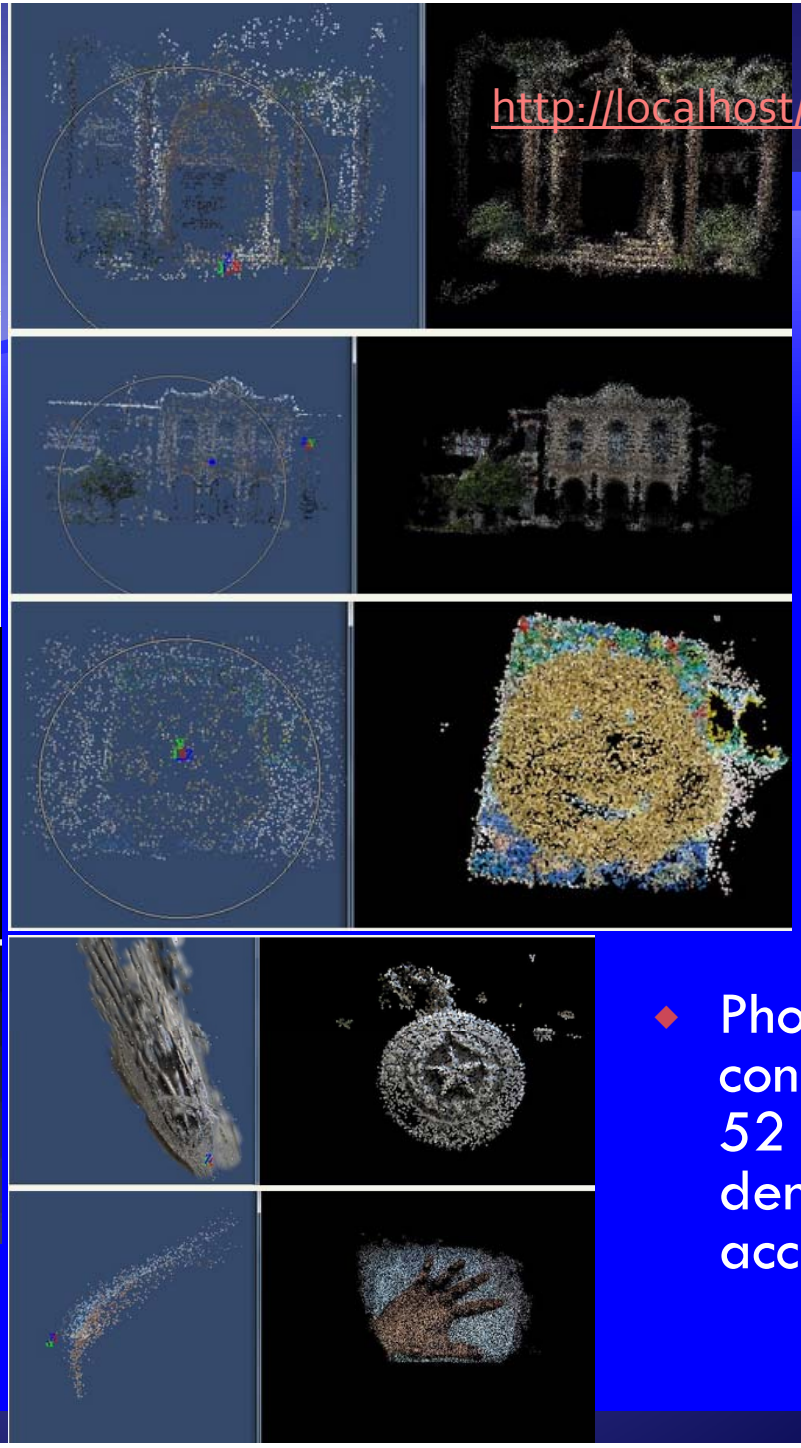
Comparison with Autodesk Photofly

http://localhost/3Ddemo/comparison

- "Photo Scenes" – Automated 3D models from digital photos (sans Visualsize, the only such 3D product in the world)
- Technologies acquired from Realviz (on May, 2008)
- RealViz (founded in 1998) technology transfer from INRIA (the ROBOTVIS research group head by Dr. Olivier Faugeras)
- Public release 7/22/2010 (after 12 years of R&D)
- 52 data sets
  - Faces/non-faces
  - Soft/hard objects
  - Shining/dull appearances
  - Fuzzy/smooth surfaces
  - Etc.



Left (blue): photofly
Right (black): PhotoModel3D



http://localhost/3ddemo/comparison/

- PhotoModel3D consistently (52 out of 52 sets) produces denser, visually accurate results

VisualSize

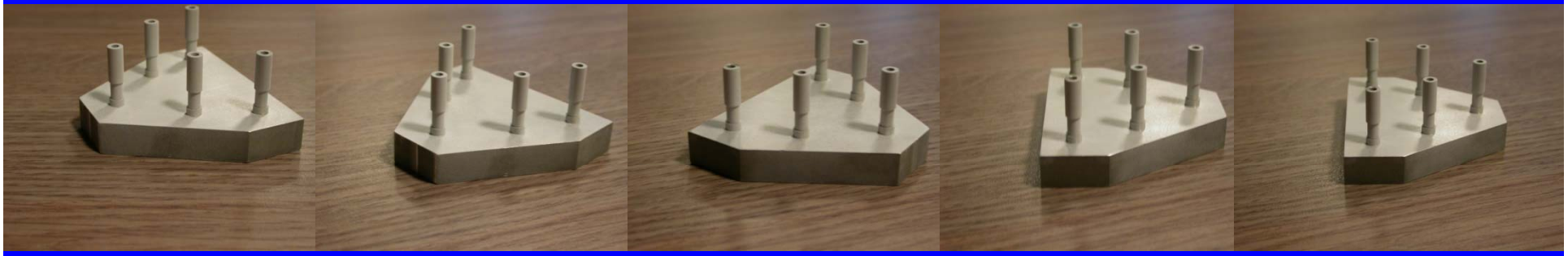
3D inference (SLAM vs. Modeling)

- CV + CG
  - Structure from motion
  - Discrete snapshots
  - Matching
  - SBA
  - Off–line
  - Batch
  - Dense maps

- Robotics
  - SLAM – simultaneous localization and mapping
  - Continuous video
  - Tracking
  - Extended Kalman Filter
  - On-line
  - Incremental
  - Sparse maps

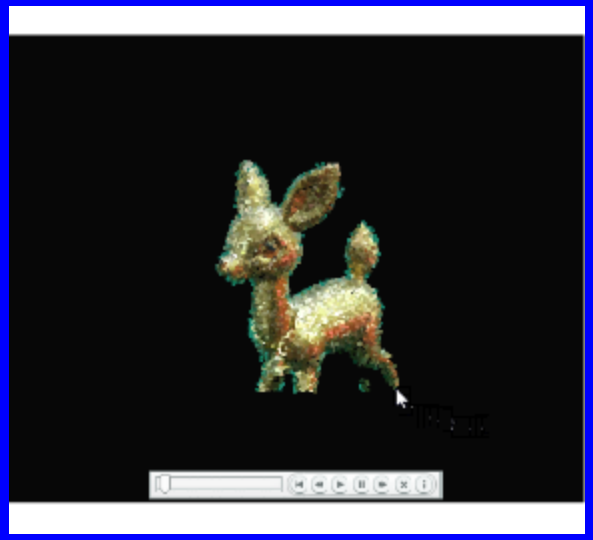
Failure Cases

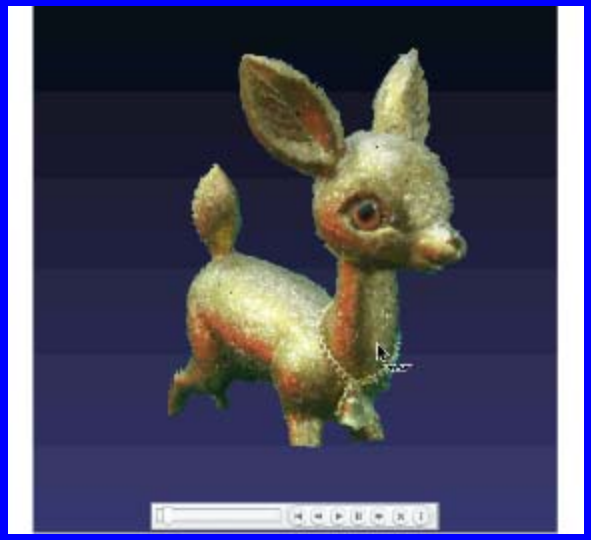
- Plain, texture-less surfaces
- Shining, transparent, translucent surfaces
- Deformable and moving objects



VisualSize

Failure Cases?



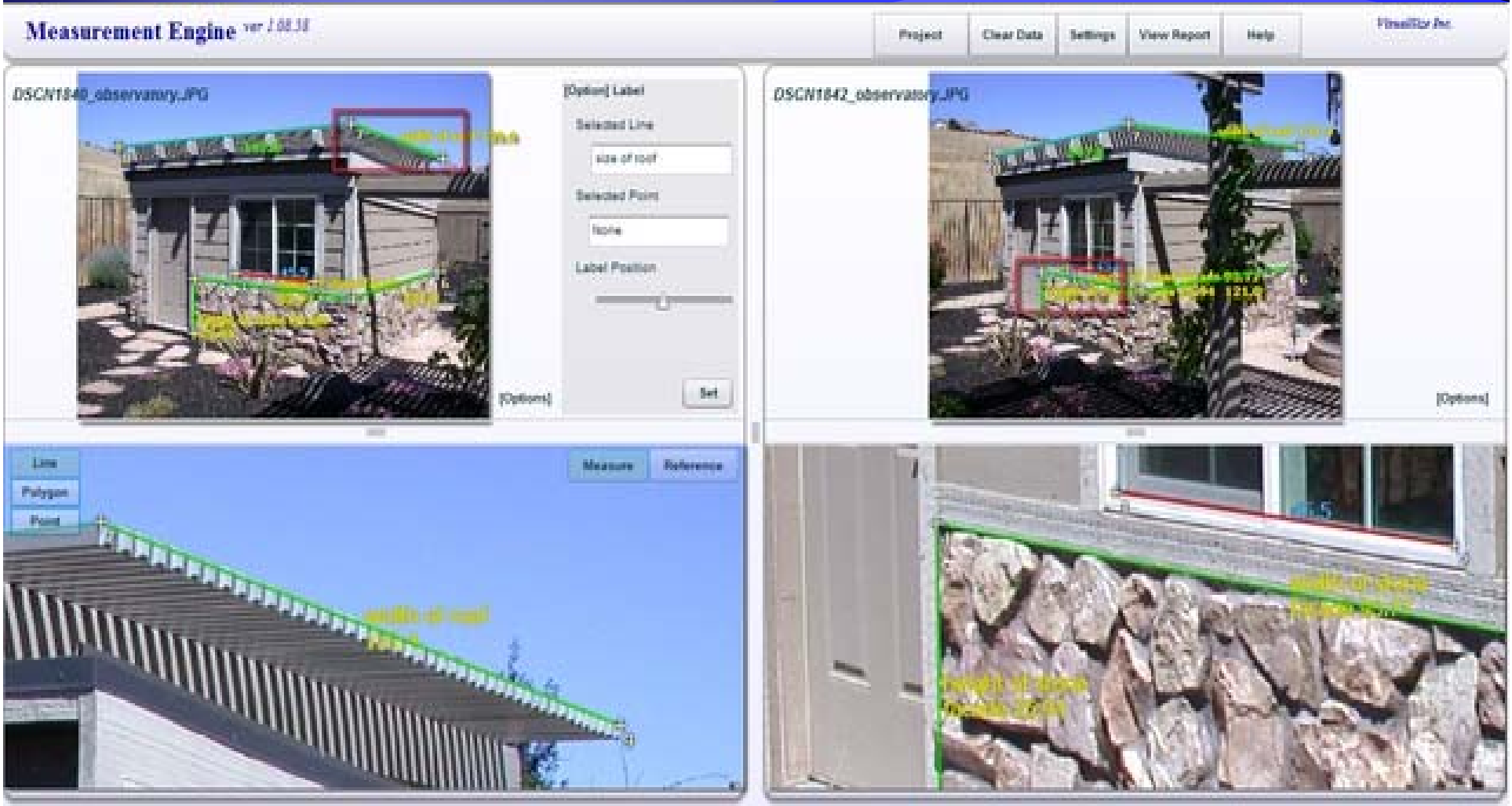


VisualSize

Metrology

- Automated camera motion analysis
- User-specified structure analysis
  - What do you want to measure?
- A single “reference” dimension must be known
- Useful for
  - Home improvement
  - Contracting
  - Cost estimation
  - Insurance damage claim

Metrology Engine



VisualSize

Comparison (competitor) Study

- Caveat:
  - Old programs (two years back)
  - Use only 2 images (without global optimization)
- Against
  - iWitness (http://www.iwitnessphoto.com/)
  - Pixdim (http://www.pixdim.com/)
  - Both competitors use “marker-based” registration system

iWitness

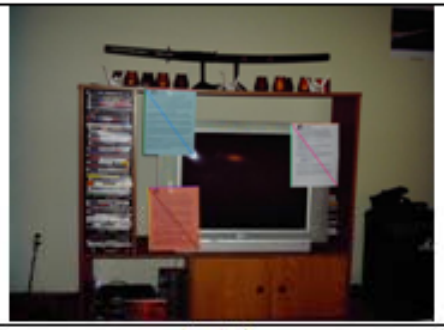
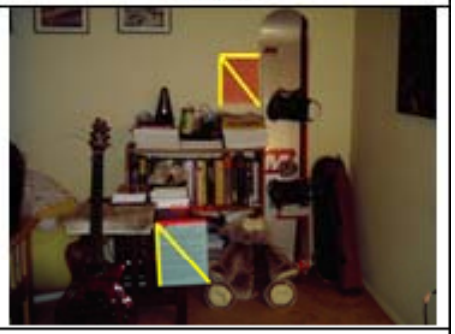
Pixdim

Comparison (competitor) Study

- Six *un*-calibrated consumer-market digital cameras
- 42 image pairs
- One reference of a known dimension per pair
- Over 200 line segments of varying lengths, positions, and orientations
  - Ground truth measured manually
  - Image locations measured using GUI of these packages
  - Average metrology error
    - 2.27% Visualsize
    - 20.62% iWitness
    - 33.03% Pixdim

VisualSize

Data sets for Metrology Accuracy Comparison

		
Data1	Data2	Data3-5
		
Data6-11	Data12	Data13-15
		
Data16	Data17	Data18, 27, and 28
		
Data19	Data20	Data21

Data sets for Metrology Accuracy Comparison

Data22

Data23

Data24

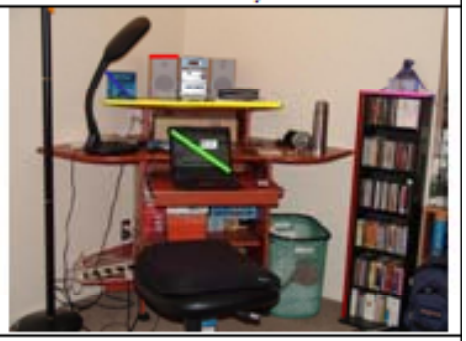
Data25

Data26

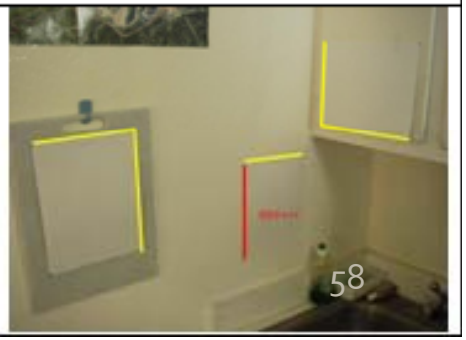
Data29, 30

Data31

Data32



Data33



Recap

- A suite of 3D algorithms for
  - Navigation and browsing of photos
  - 3D Metrology
  - Panorama
  - 3D Models
- Developed in-house and hold IPs
  - Bundler uses SIFT, LM, SBA, ANN with GNU GPL – not for commercial use
  - Furukawa's PMVS – again GNU GPL
- One stop shopping for 3D technologies
- Complete systems, end-to-end and fully automated
- World-class
  - Compared favorably with Microsoft/U Wash and Autodesk
  - Exhaustive Web search ("3D models from photos, 3D faces from photos") unearthed no other worthy competitors

Potential Applications

- Internet showroom and web sales
- Social networks
- Entertainment (movie and game) environment map, FX
- Virtual tourism and museums
- Insurance claim processing
- Crime scene analysis
- Realistic event simulation, surveillance
- Situation study, threat assessment, campaign planning
- Construction (roofing, floor, etc.) and home improvement (remodeling)
- Urban development, city planning
- 3D digital cameras, other hardware solutions

Specific Example: 3D face models

- Human nature: fascination with faces
- 10 to 20 images, point-shoot-upload
- <5 minutes from start to finish (1CPU core, no hardware acceleration)
- No pre-existing "fake" 3D face model to introduce artificial bias in 3D structure
- No active mechanism used
- Internet games
- Plastic surgery
- Telephony
- Social networks
- Security surveillance

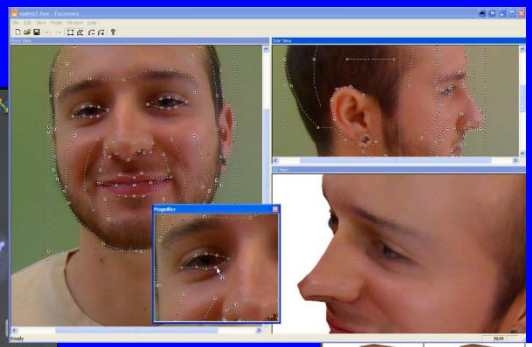
Comparison (Competitor) Study

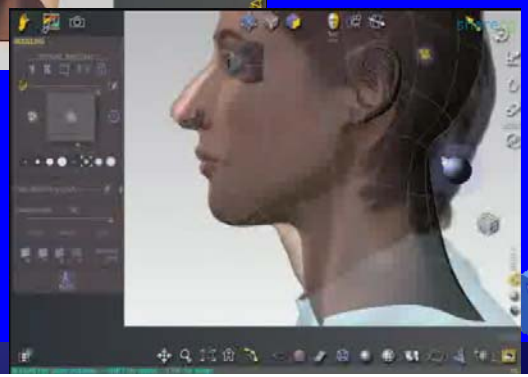
Facegen.com, FaceShop, quidam, thatsmyface.com , looxis.com

- Manual face editing and animation programs
- "Faking" 3D structure
  - texture mapping on existing 3D models, using manually entered fiducial points
- Bias

PhotoModel3D

- Real texture and structure
- No underlying model is used
- No bias
- Fast (<5 minutes)
- Not $299 ☺





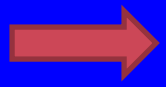
VisualSize

Comparison (Competitor) Study

- Inspeck.com, looxis.com
- Active projection systems
- Multiple projectors for full head capture
- Expensive, time consuming registration & calibration





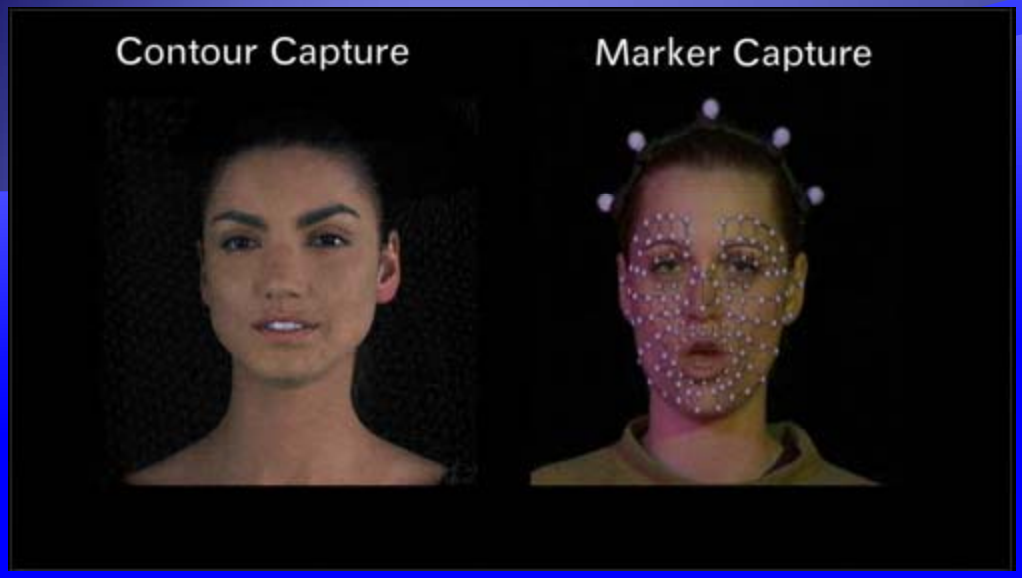



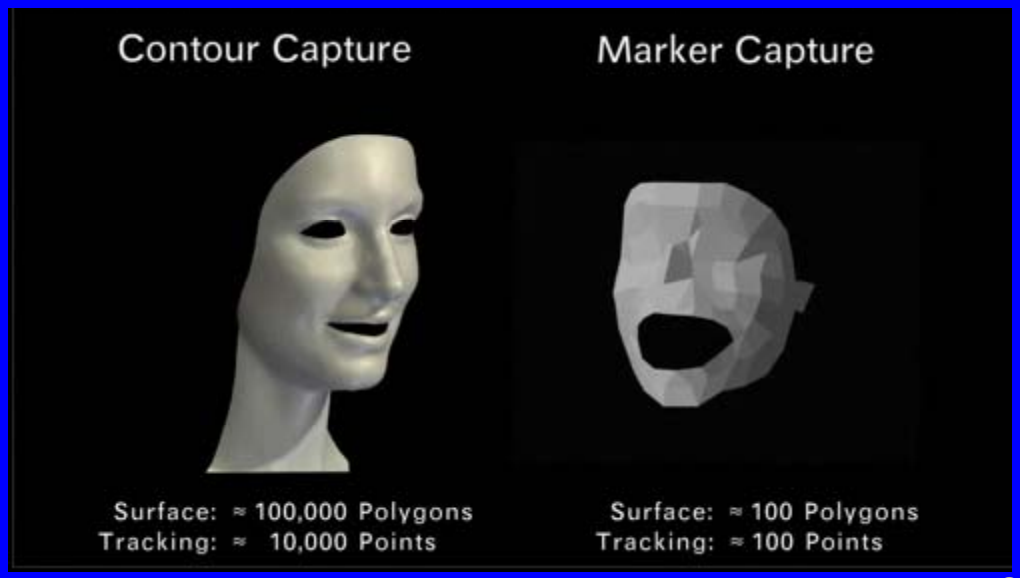
Comparison (Competitor) Study

- Mova.com (contour capture)
  - Phosphorescent makeup and dye (90-120fps flash)
  - Capture both bright and dark frames
  - Random phosphorescent patterns from multiple cameras for triangulation

VisualSize

Comparison (competitor) Study

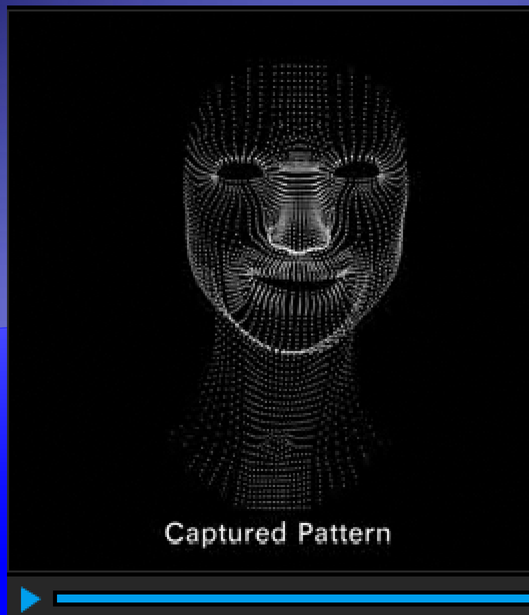


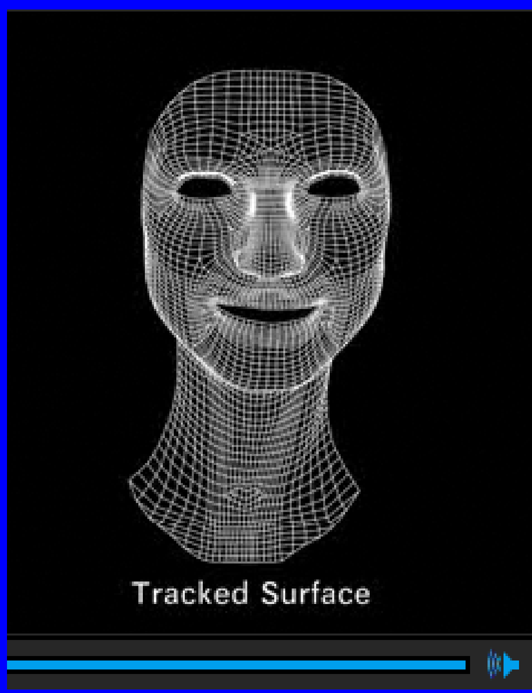


- Photomodel3D
  - Any single consumer-market digital camera
  - No markers
  - 18,761 points (dad)
  - 23,854 points (mom)
  - 30,404 points (daughter)

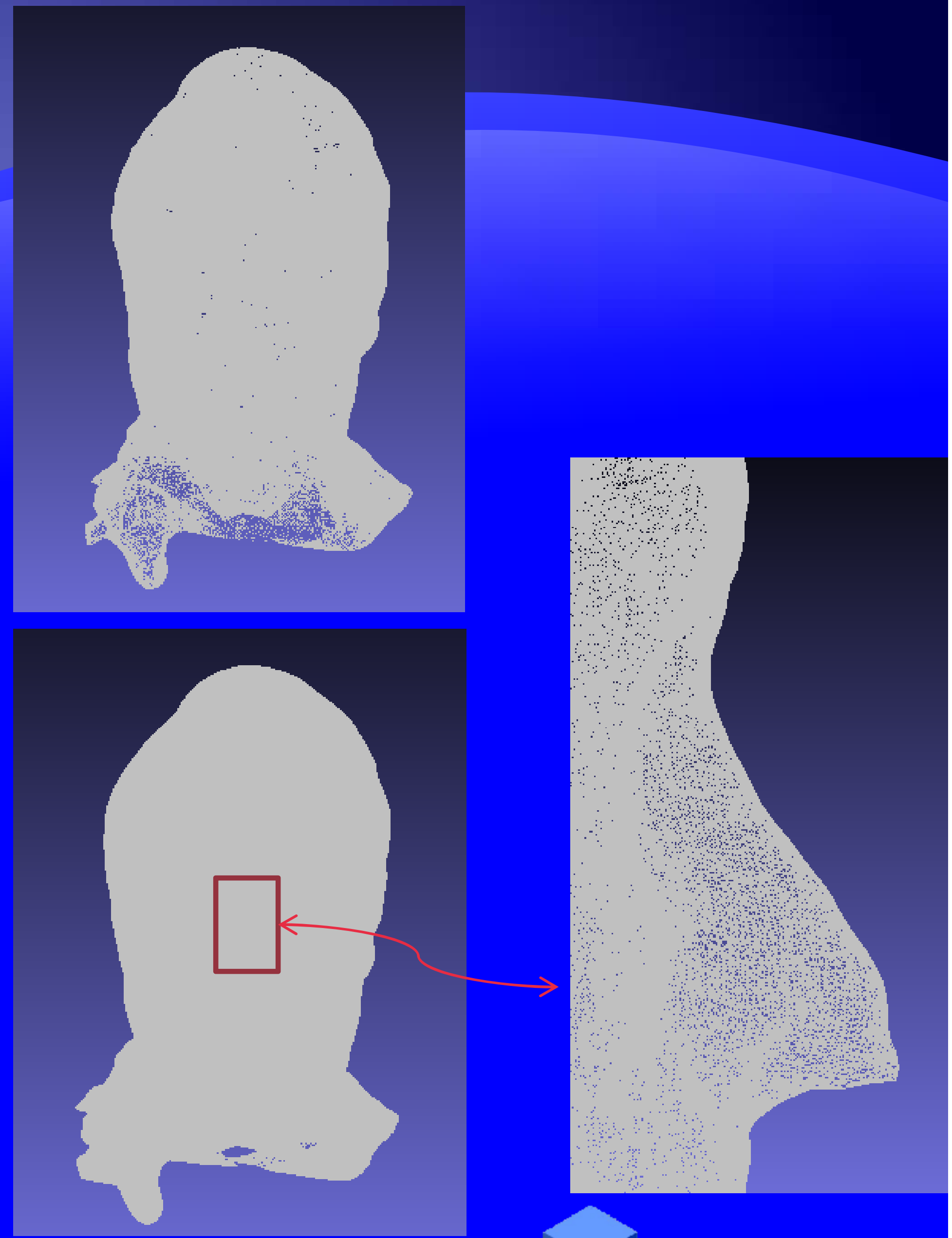
VisualSize

Mova.com

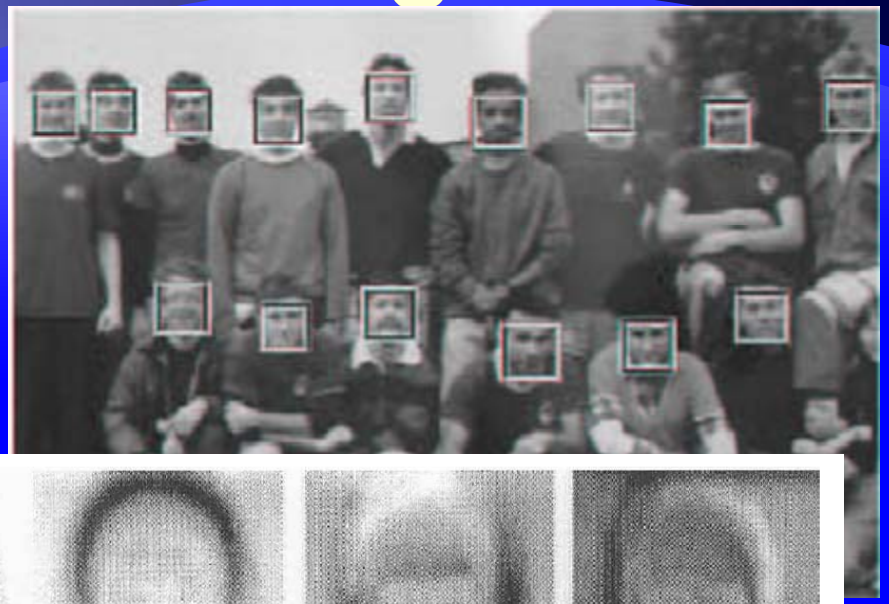


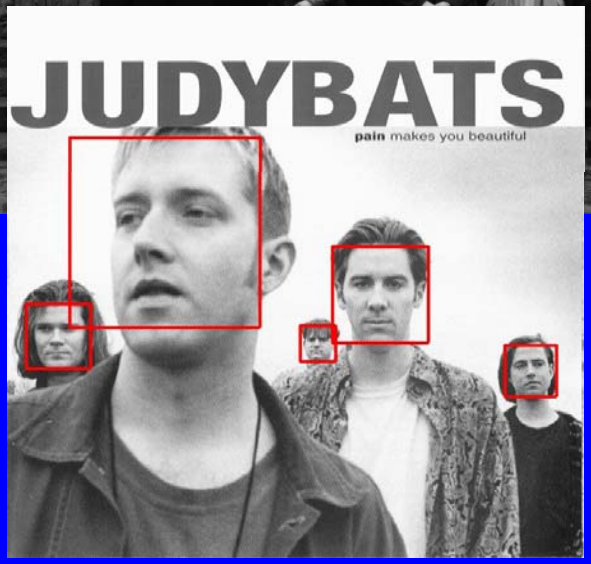


Visualsize



Face Detection and Recognition





- 2D, Frontal views
- Learning and training often necessary

VisualSize

Embedded Applications

State of the art

- Face detection
- Hit-and-miss (mostly misses ☺)
- Another auto-focusing solution
- Not recorded in images/headers
- Not used for recognition, search, categorization later
- Little improved experience
- No tangible byproducts, very limited enhanced experience (do you know/care your camera's autofocus mechanism?)

New possibility

- Face model
- Tangible byproducts, 3D face models for
  - Social networking
  - Internet games
  - Baby pictures in 3D
  - Fancy screen saver

Animation

- Our 3D face model is a snapshot, but it is ready to be animated (aka talking head)
- Concrete applications:
  - Voicemail: has messages read to you by caller's avatar (transcription, text-to-speech, face animation)
  - Teleconference over cell phones: pre-stored 3D face model of the caller
  - IM: has typed text messages read to you by caller's avatar

PERSONALIZATION IS THE KEY

- A talking head you don't even know is "gee-whiz"
- Personalized avatar provides the needed emotional connection to make technologies desirable
  - A child will get tired talking to an unknown avatar, but not to her parents
  - Teleconference with an unknown talking head provides little enhancement in user experience
- Inexpensive, consumer-market enabling 3D modeling technology
  - Consumer-market camera, efficient and robust solution
  - Only company with such face-modeling ability

VisualSize

Security Surveillance

State of the art

- Well-controlled environment:
- 2D
- Frontal
- Learning and training
- Preprocessing, cropping, normalization, etc. may be necessary
- Logic extensions: moderately controlled and un-controlled (Gang Hua, *et. al*, IEEE PAMI special issue on Real-World Face Recognition)

New possibility

- 3D
- Not necessarily frontal
- No offline learning
- No manual interaction
- A stable, forward projection process instead of unstable, inverse prediction process

Significantly More Examples at www.visualsize.com