Landmarks in 3D Computer Vision

Steve Seitz
Google
University of Washington

NSF Workshop on Frontiers of Computer Vision
August 21, 2011

Sources

Contributors

- Wolfgang Förstner
- Berthold Horn
- Shree Nayar
- Bill Triggs
- Bob Woodham
- Richard Hartley
- Jitendra Malik
- Rick Szeliski
- George Vosselman
- Andrew Zisserman

References

- Robert Burtch's history of photogrammetry
- Thomas Buchanan, Photogrammetry and projective geometry-an historical survey
- Bill Triggs et al., Bundle Adjustment - A Modern Synthesis

Disclaimer

This list is biased and incomplete!

- it represents my perspective, informed by several experts
  - I selected for "highest impact" results
- Many important results omitted for brevity
- I will post other experts' lists online, along with these slides

A bit of pre-history

Leonardo da Vinci
(1452–1519)

"Perspective is nothing else than the seeing of an object behind a sheet of glass, smooth and quite transparent, on the surface of which all the things may be marked that are behind this glass. All things transmit their images to the eye by pyramidal lines, and these pyramids are cut by the said glass. The nearer to the eye these are intersected, the smaller the image of their cause will appear"

Johann Heinrich Lambert
(1728–1777)

Photometry
Camera resectioning

Carl Friedrich Gauss
(1777–1855)

Least squares
(and many other things!)

Charles Wheatstone
(1802–1875)

Stereopsis

Early photogrammetry

Mathematical foundations

- least squares [Gauss 1794]
- projective geometry [Poncelet 1822]
- horopter [Vieth 1818; Muller; Helmholtz; ...]
- multiview relations [Hauck, Hempe, 1850-60s; Kruppa 1913]

Practical use

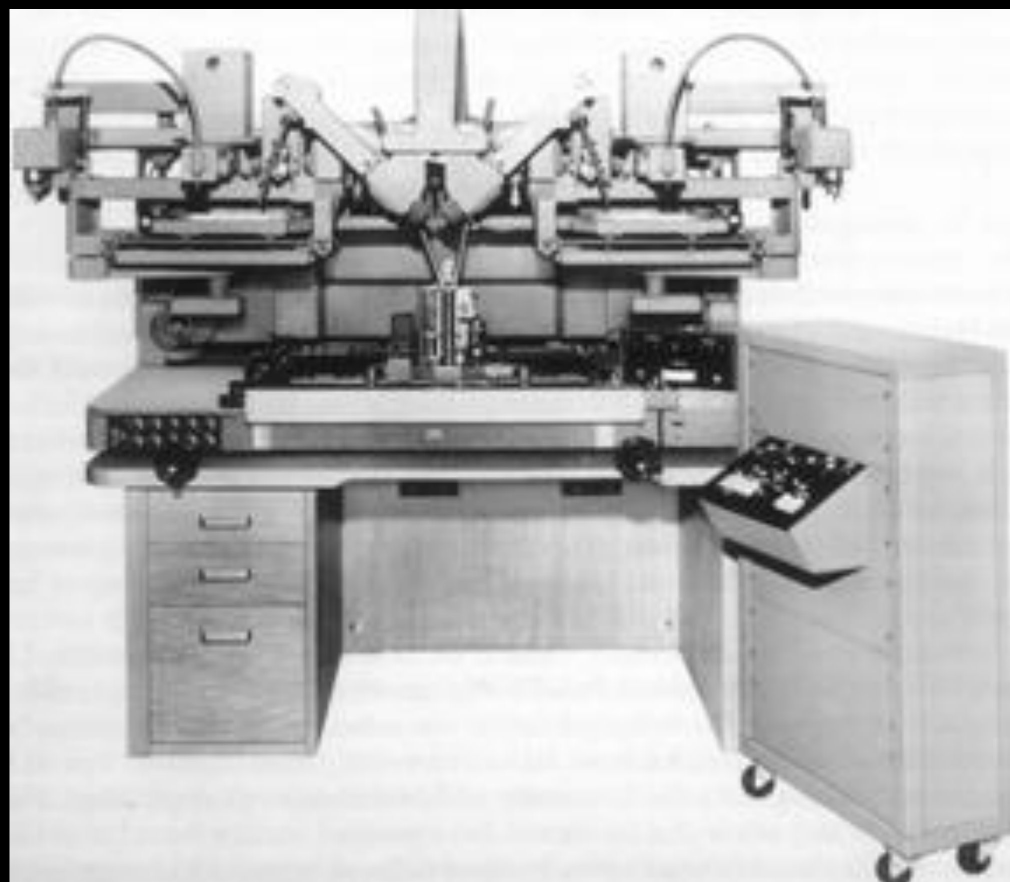
- geodesy, map-making date back hundreds of years
  - 1856: manual bundle adjustment proves Mt. Everest highest mountain on earth (Great Arc of India survey)
- Invention of photography, notably Daguerreotype [1837]

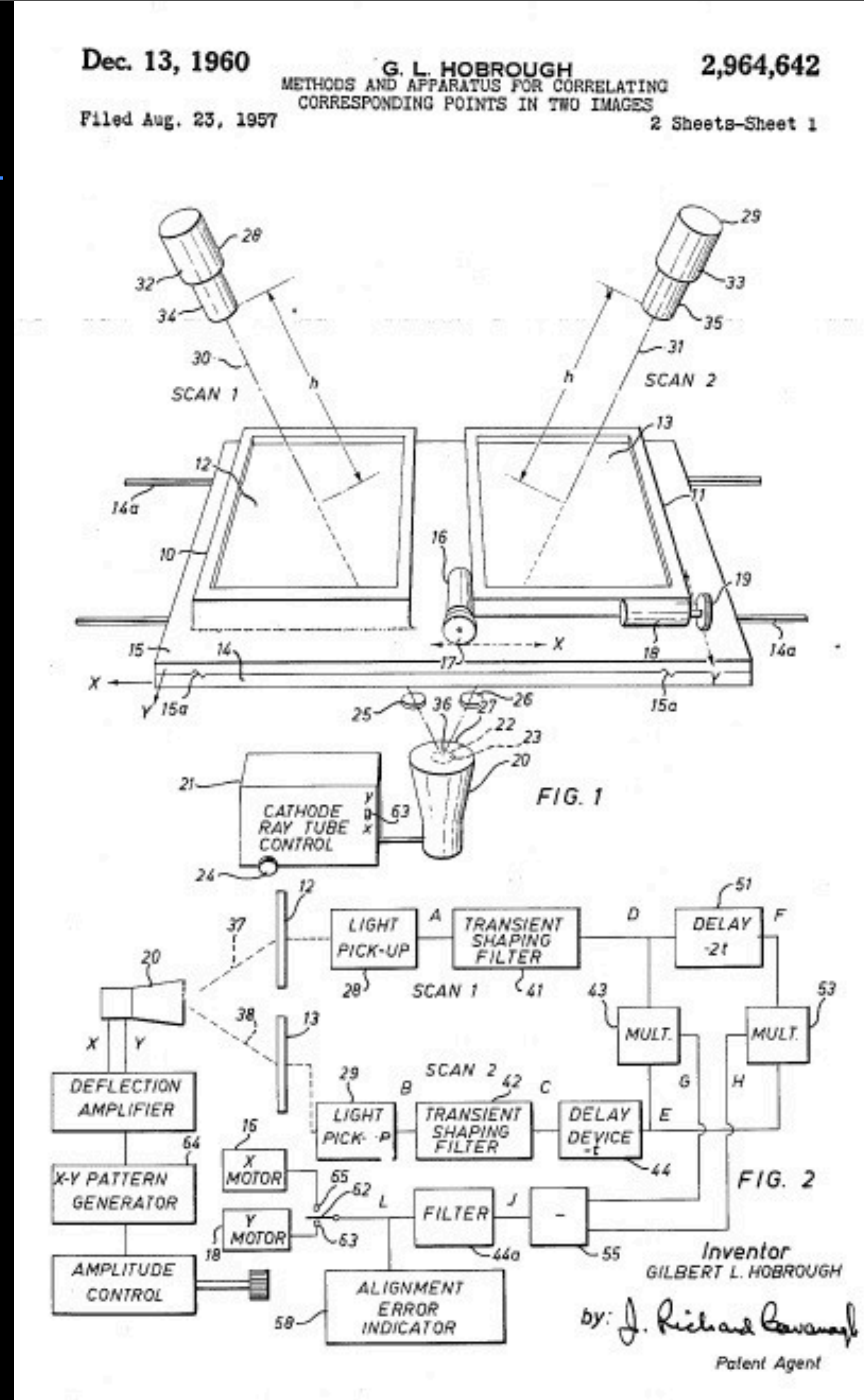
Digital bundle adjustment [late 1950’s]

- required manual specification of fiducial points
- Pioneered by U.S. Air Force [Duane C. Brown et al., 1957–9]

3D Computer Vision

1957 Stereo [Hobrough]



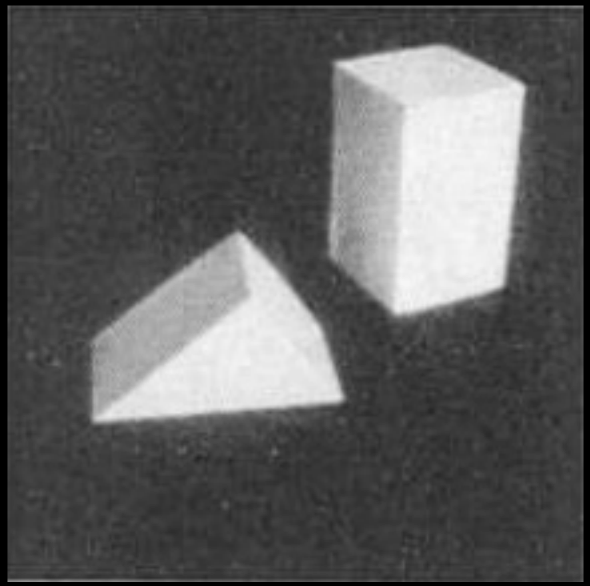


In 1957, Gilbert Hobrough demonstrated an analog implementation of stereo image correlation (patent shown right). This led to the creation of the Raytheon-Wild B8 Stereomat (pictured above).

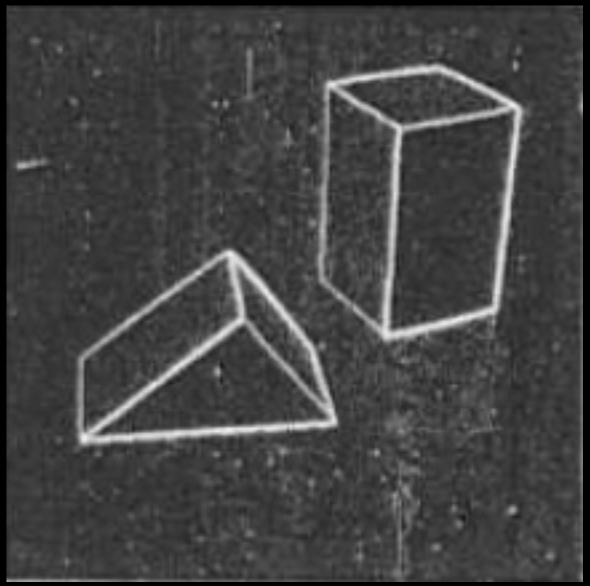
Forerunner of the *Gestalt Photo Mapper* [Hobrough 1967], used by the U.S. Geological survey to create Digital Elevation Maps and orthoimages. This used a 6-level pyramid with warp refinement at each level.

1963: Blocks World [Roberts, MIT Ph.D]

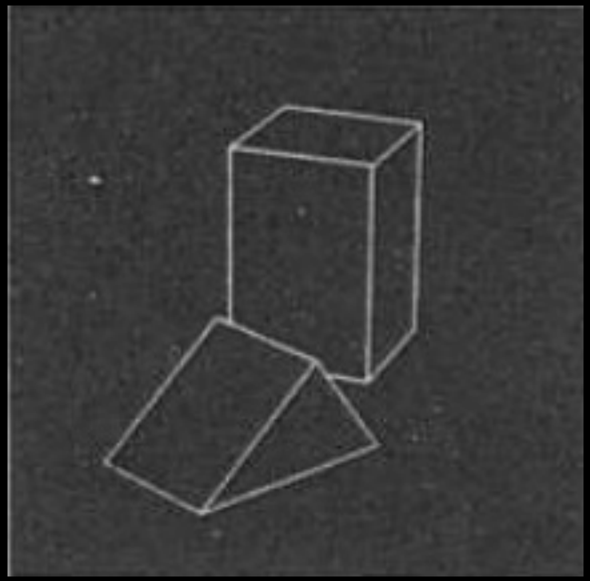
Larry Roberts
"Father of Computer Vision"
"Father of ARPANET"



input image



2x2 gradient operator



computed 3D model
rendered from
new viewpoint

Larry Roberts PhD Thesis, MIT, 1963, Machine Perception of Three-Dimensional Solids

Blocks world++

Culmination: 1970 Copy-demo at MIT

- vision system recovers the structure of a blocks scene, robot plans and builds an exact copy from another set of blocks

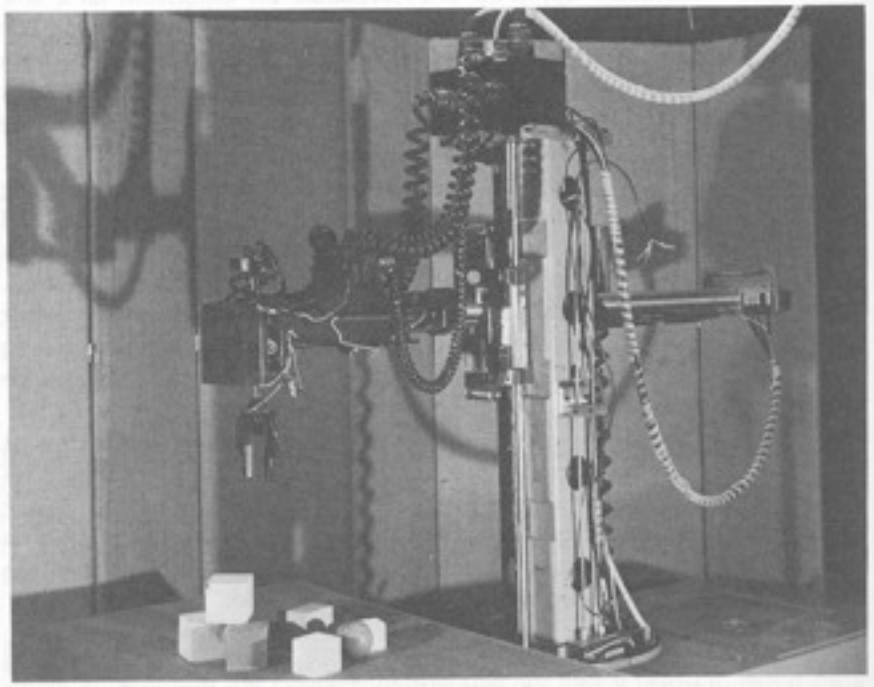


Figure 15-3. The mechanical manipulator used in the "copy-demo," one of the first projects in which visual information was used to plan the motions of an industrial robot. (Photo by Steve Slesinger.)

Winston, Horn, Freuder

Ambitious AI system

- vision + planning + manipulation
- weak link: low-level edge finding not good enough for task
  - led to more attention on low-level vision

1970: Shape-from-shading [Horn, MIT Ph.D., (AITR 232)]

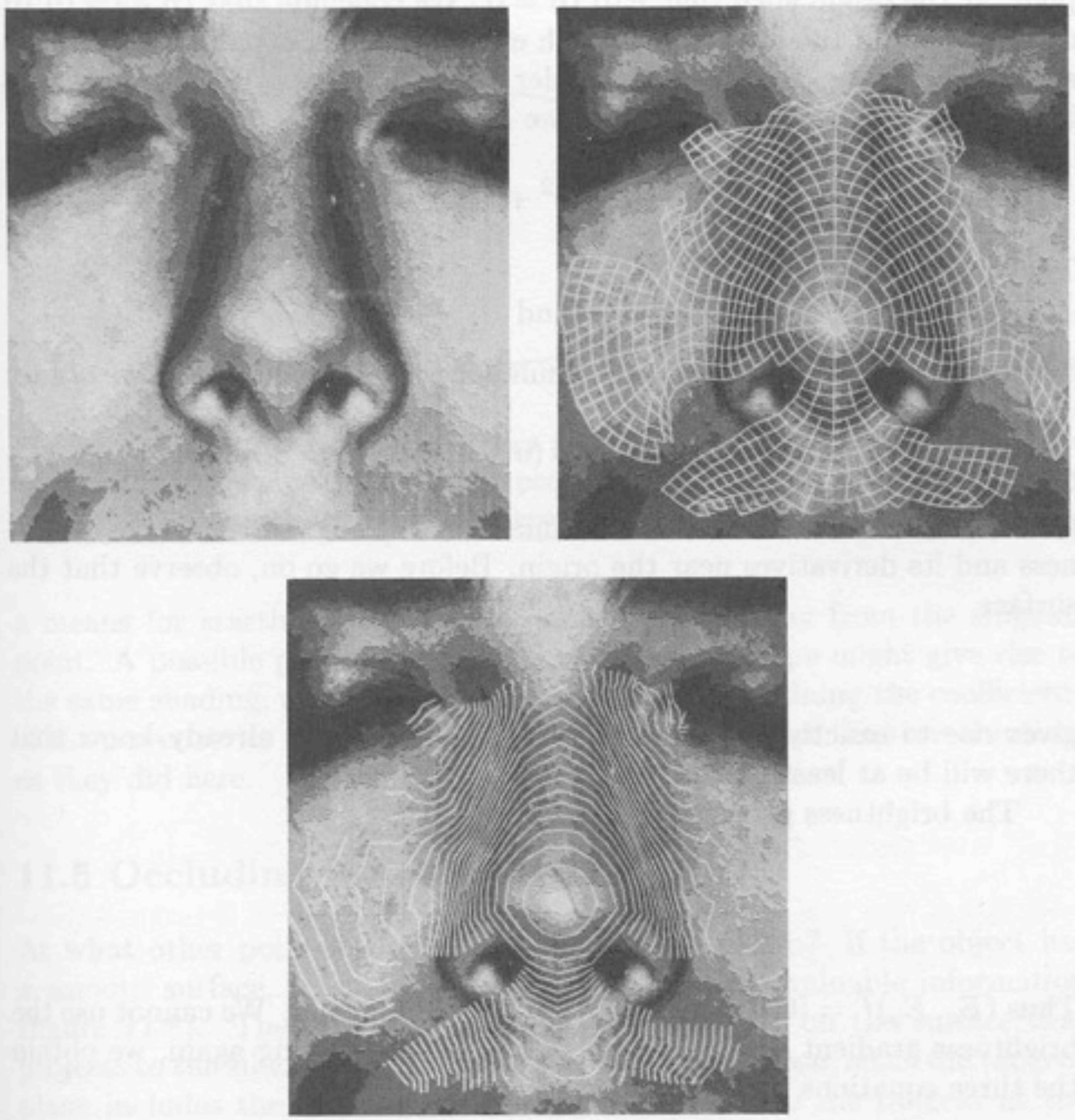


Figure 11-7. The shape-from-shading method is applied here to the recovery of the shape of a nose. The first picture shows the (crudely quantized) gray-level image available to the program. The second picture shows the base characteristics superimposed, while the third shows a contour map computed from the elevations found along the characteristic curves.

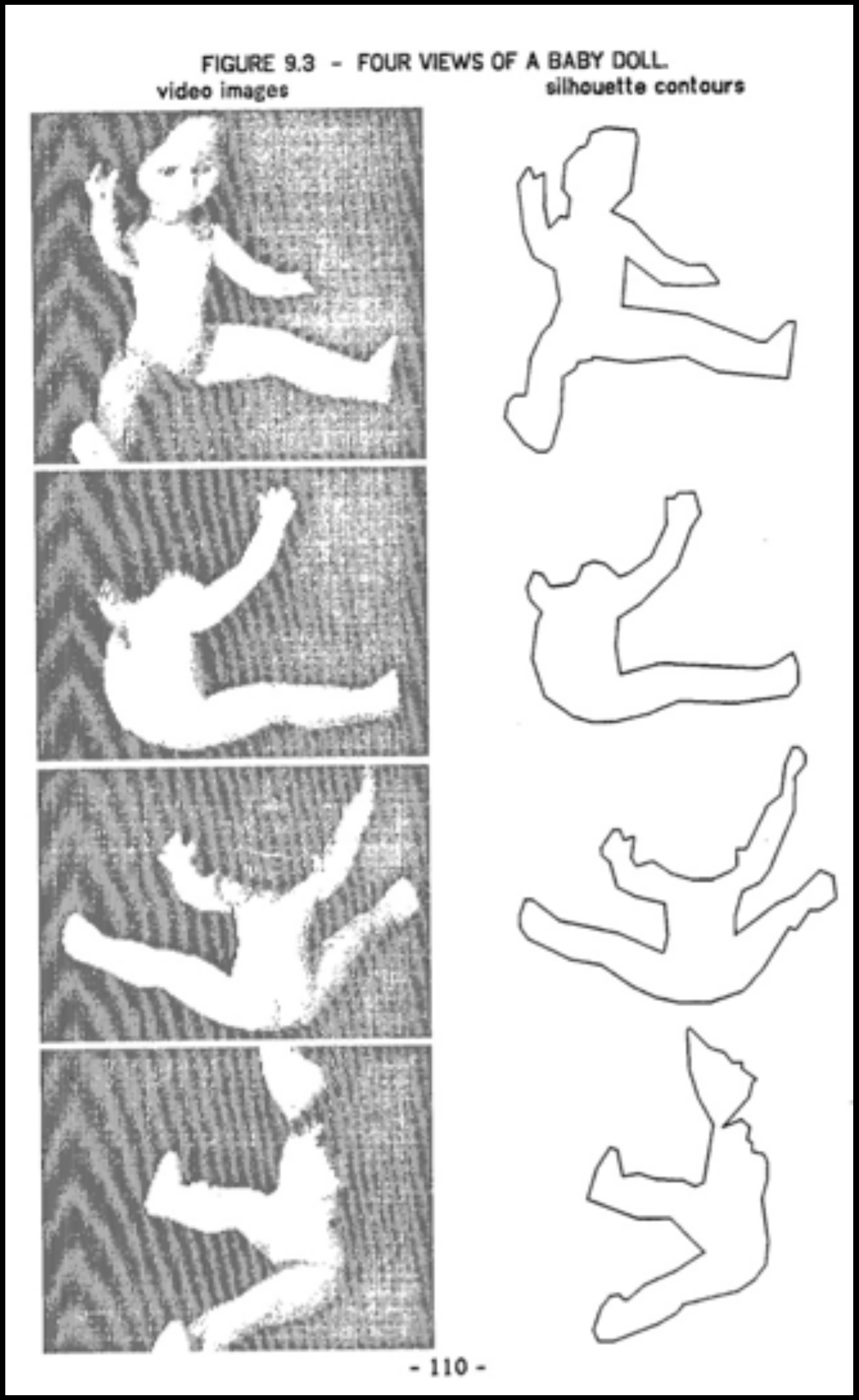
Also known as *photoclinometry*

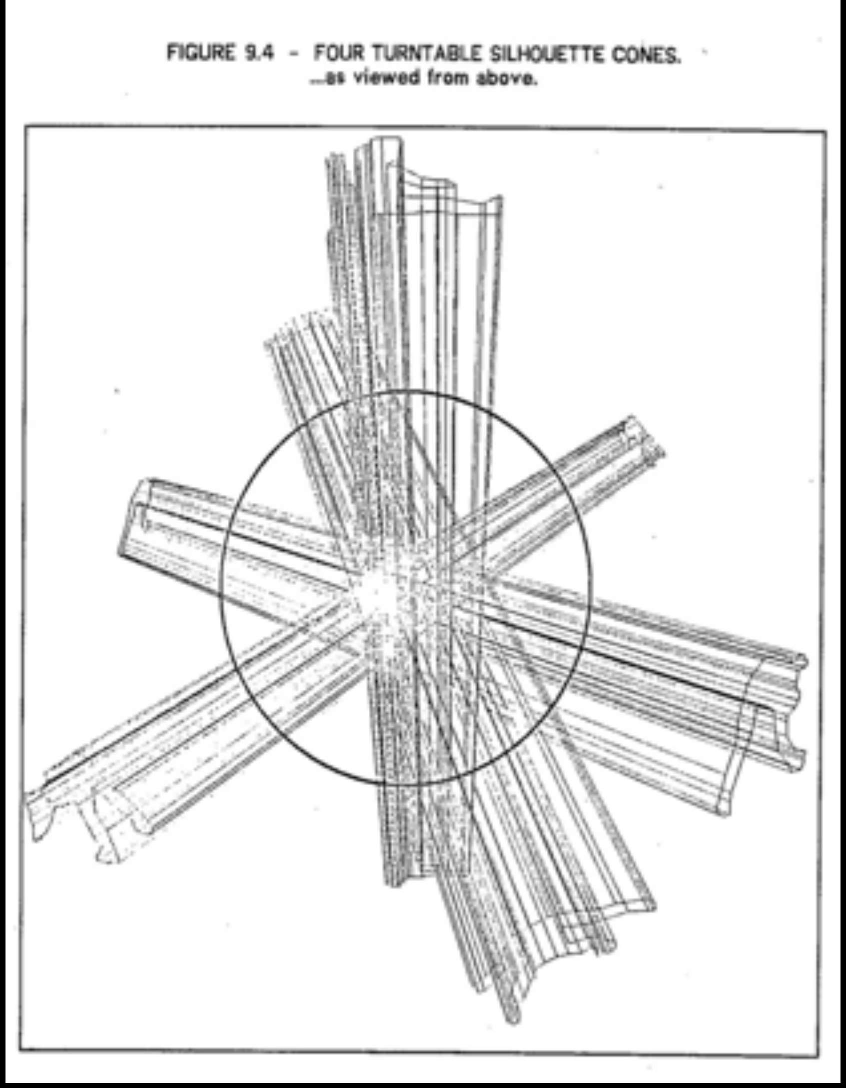
- measure planetary geometry
- Origins from mid 1960's
  - T. Rindfleisch, Photometric Method for Lunar Topography, Photometric Engineering, 1966.
- SfS vs. photoclinometry discussion in Horn 1989

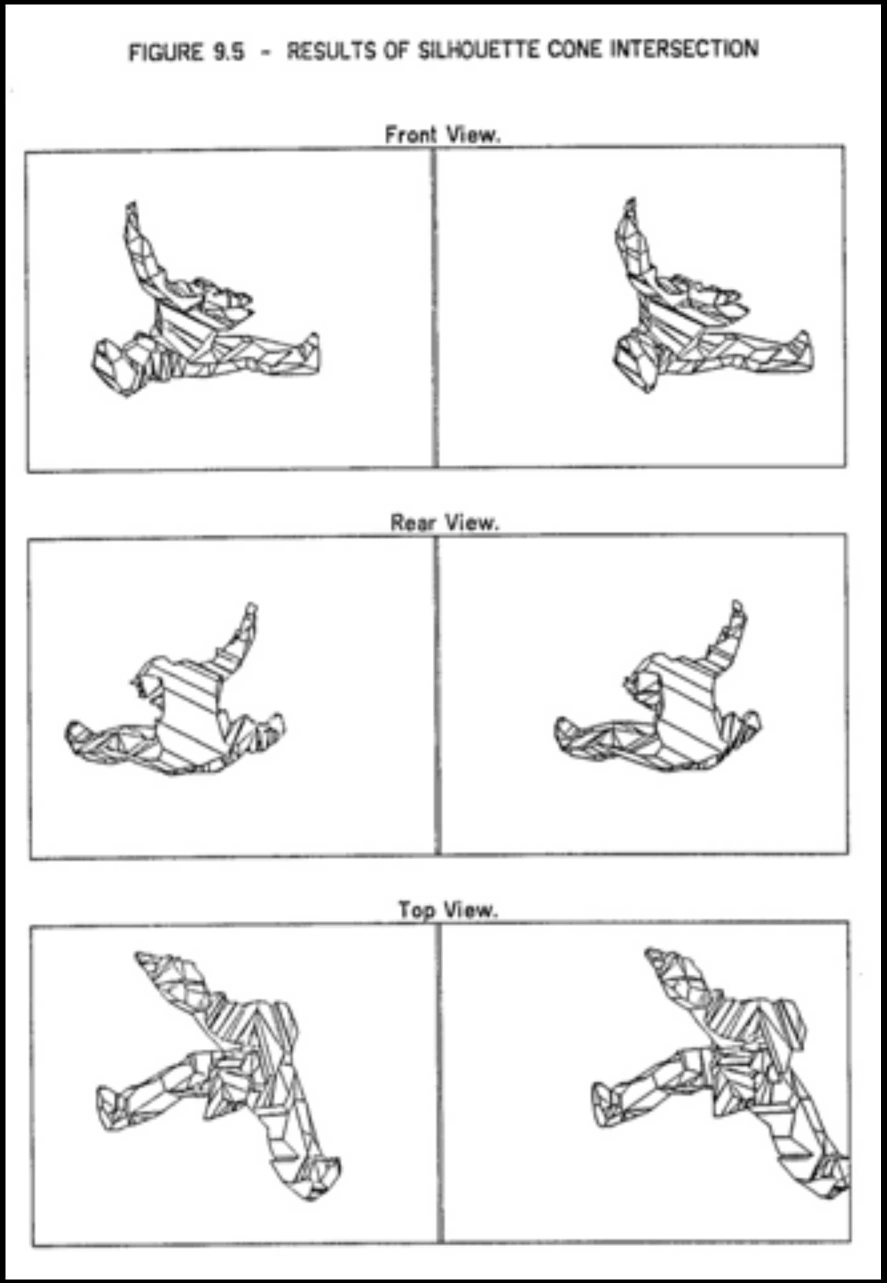
Ernst Mach [1865] formulated image irradiance equation

- concluded shape inference was impossible because of instabilities

1974: shape from silhouettes [Baumgart, Ph.D]



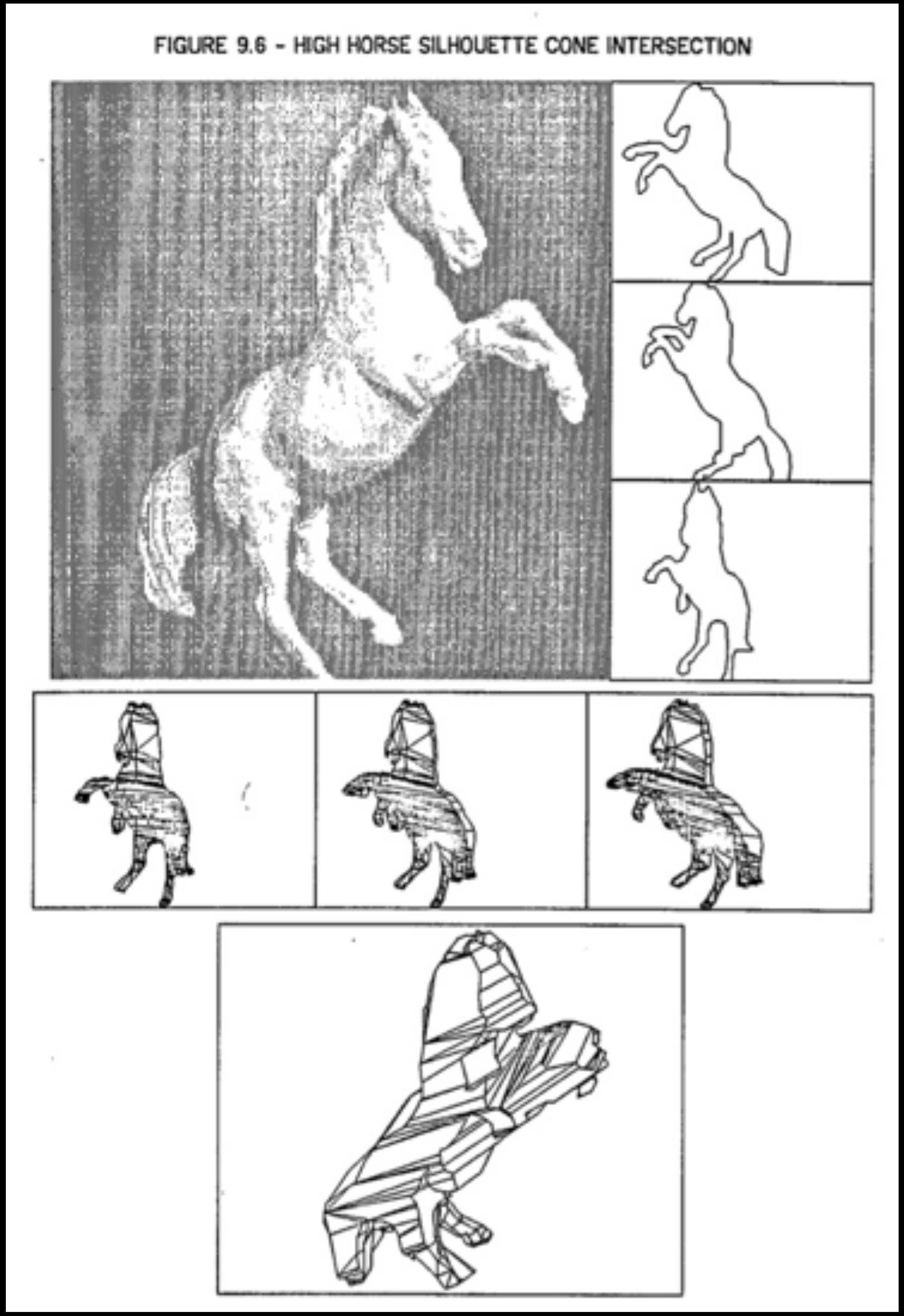




reconstruction
(fuse as stereo)

Bruce G. Baumgart, Geometric Modeling for Computer Vision. Stanford PhD Thesis (CS-TR-74-463)

1974: shape from silhouettes [Baumgart, Ph.D]



Bruce G. Baumgart, Geometric Modeling for Computer Vision. Stanford PhD Thesis (CS-TR-74-463)

Shape-from-silhouettes...

Other notable results include:

Theory: visual hull

- A. Laurentini, "The visual hull concept for silhouette-based image understanding". *IEEE Trans. Pattern Analysis and Machine Intelligence.* 1994, pp. 150–162.

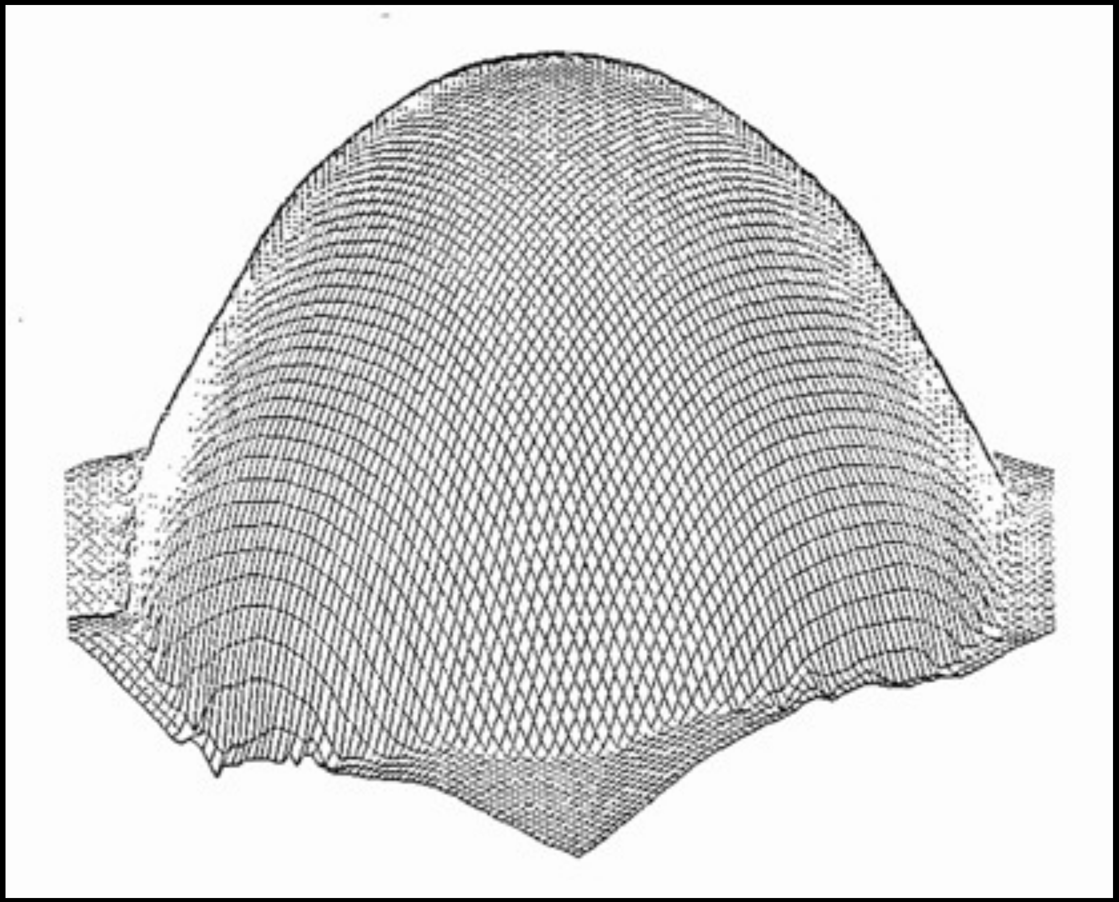
Efficient Algorithms

- Richard Szeliski. Rapid octree construction from image sequences *CVGIP: Image Understanding*, 58(1):23-32, July 1993.

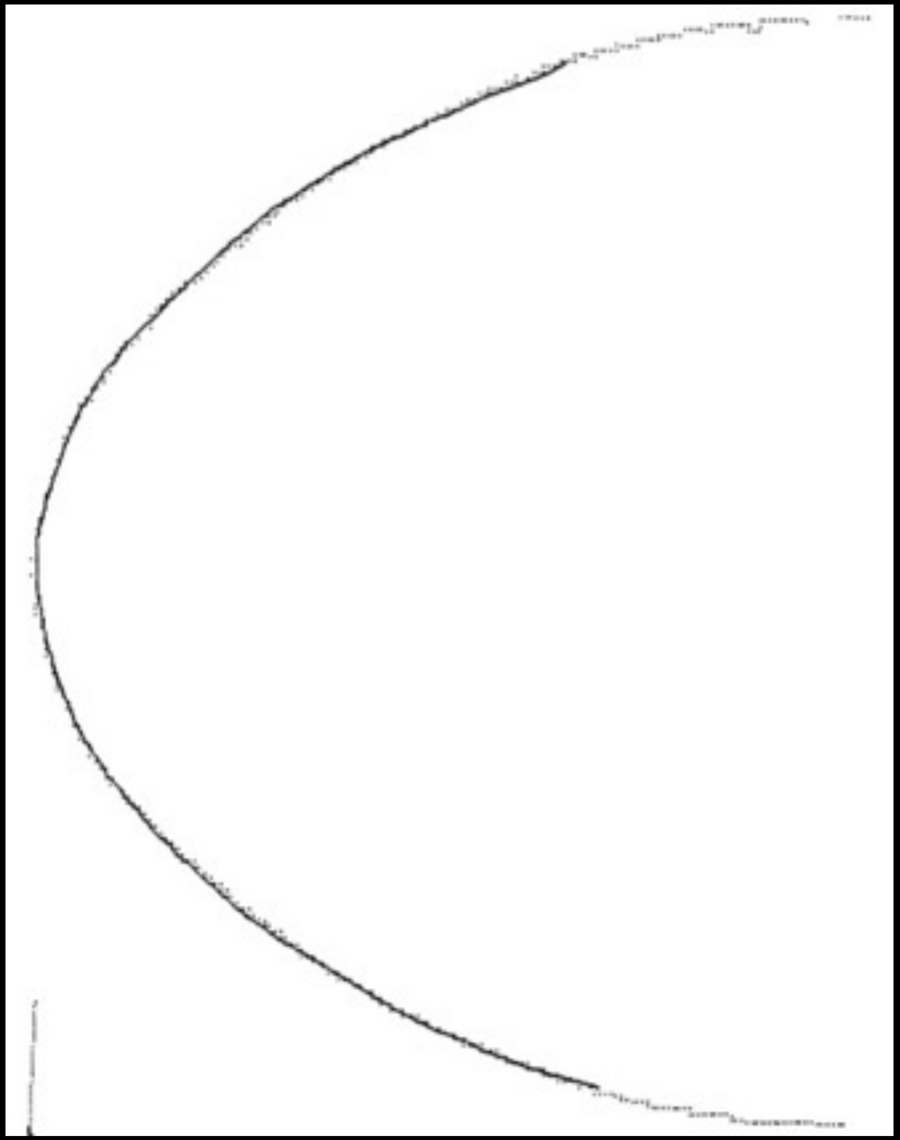
Usage in graphics

- W. Matusik, C. Buehler, R. Raskar, L. McMillan, and S. Gortler, Image-Based Visual Hulls. In Proc. SIGGRAPH 2000.

1977: Photometric Stereo [Woodham, MIT Ph.D.]



Reconstruction of a Wooden "Egg", William Silver, 1980



profile: reconstruction (solid), ground truth (dotted)

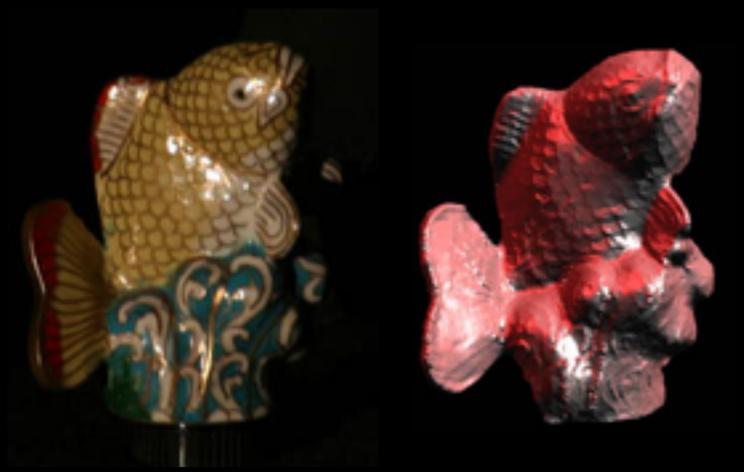
Multiview shape-from-shading

- Introduced by Bob Woodham: 1977 PhD, 1980 opt. Eng paper
  - first implementation by William Silver: 1980, MIT master's thesis
- Unprecedented detail, accuracy (requires 3+ views)
- Works for ANY isotropic BRDF (Lambertian not required)

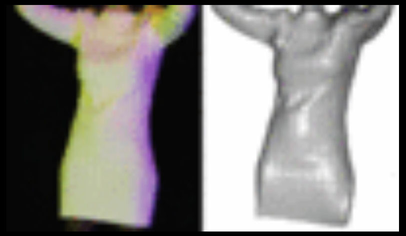
Photometric stereo…

Other notable results include:

- Shape from Interreflections,
  - Nayar et al., 1990, pdf
- Example-based Photometric Stereo
  - Hertzmann & Seitz, 2003, website



- Moving scenes
  - Hernandez et al., 2007, pdf video



- Retrographic sensing
  - Johnson & Adelson, 2009, website

1981: Essential Matrix [Longuet-Higgins, Nature]

H. Christopher Longuet-Higgins (September 1981). "A computer algorithm for reconstructing a scene from two projections". *Nature* **293** (5828): 133–135

3x3 Matrix mapping points to epipolar lines

- corresponding points p, p′ satisfy $p^T E p' = 0$
- camera matrices can be computed from E

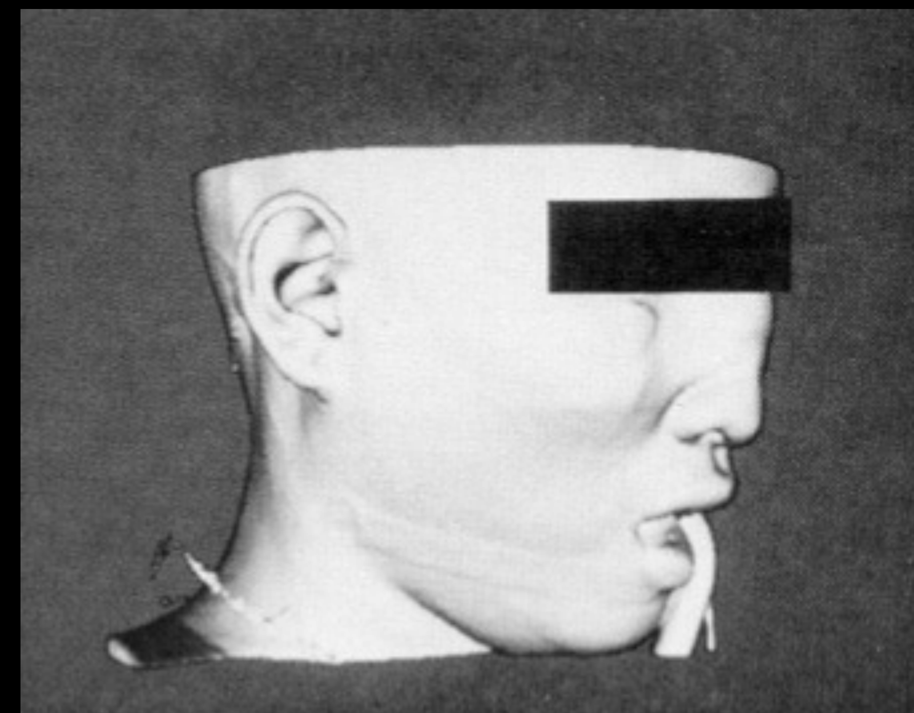
Historical precedents

- Chasles, Hesse, Sturm
  - introduced key ideas 100 years earlier [1863-9]
- Kruppa's "Structure-from-motion" theorem [1913]
  - rediscovered by Ullman [1977]

Sparked field of "multi-view geometry" in the 1990s

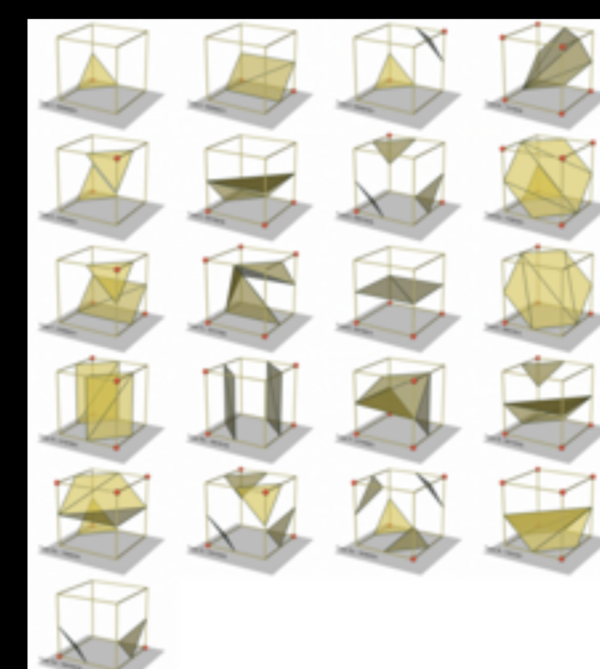
- Fundamental matrix [Faugeras; Hartley ECCV 1992], uncalibrated case, song!
- Trifocal tensor [Hartley; Shashua 1995], 3 view case
- Self-calibration, stratification, [Faugeras, ECCV 1992]

1987: Marching Cubes [Lorensen & Cline, SIGGRAPH]

From Volume to Surface mesh

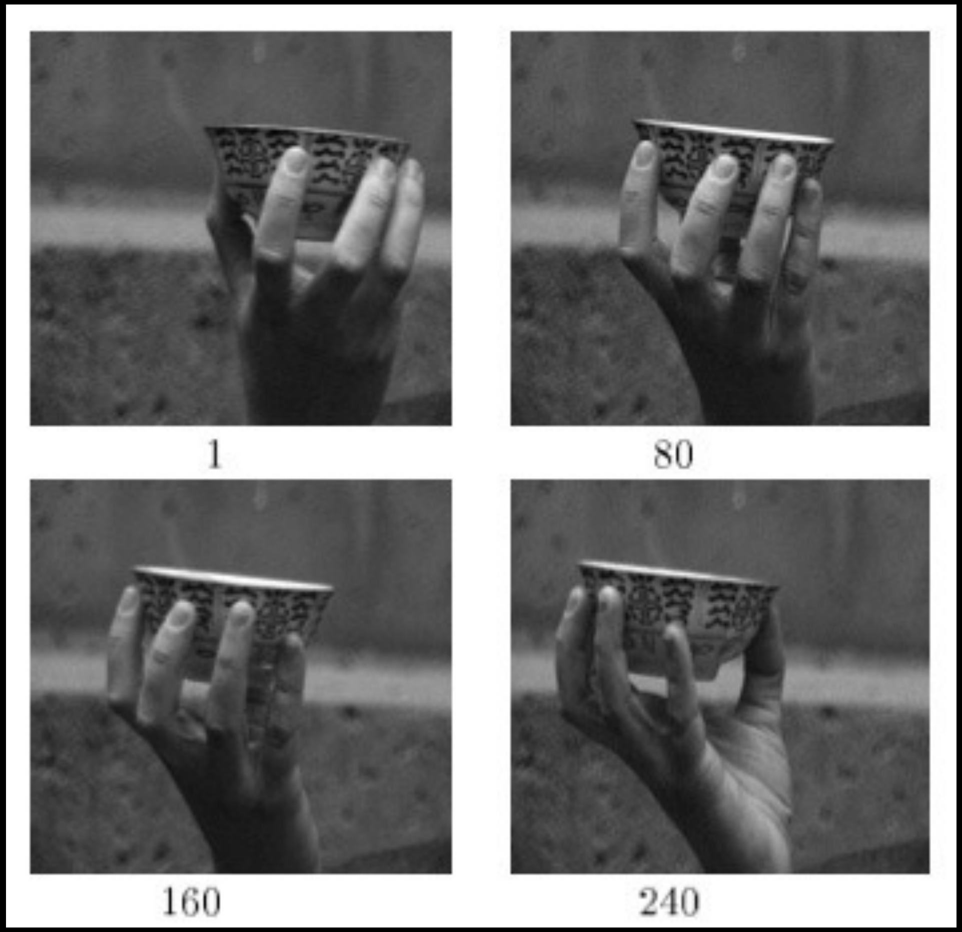
- Start at voxel containing surface
- Add polygon(s) based on configuration table
  - earlier: 1970's Hummel & Zucker, 3D edge finding
- March to next voxel

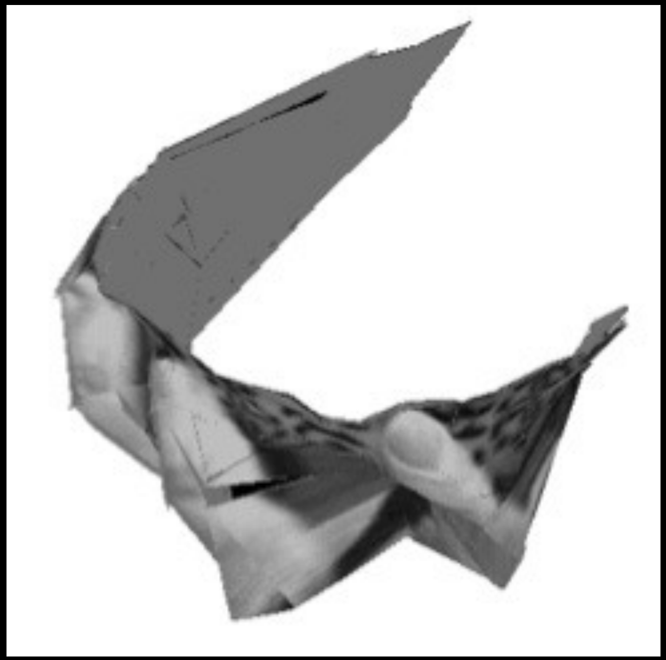


To this day, still dominant meshing alg!

[Tomasi & Kanade, ICCV]

1990: Structure-from-motion by factorization





Elegant "1 line" solution: W = MS

- optimal under affine (orthographic) model
- many extensions
  - multibody [Costeiri, ICCV 1995],
  - flow [Irani, ICCV 1999]
  - nonrigid [Bregler, CVPR 2000], [Brand, CVPR 2001]

[Besl & McKay; Chen & Medioni; Zhang]

1992: Iterative Closest Points (ICP)

Figure from Szymon Rusinkiewicz

Range Scan matching

- Invented simultaneously by multiple authors
  - Besl, McKay, "A Method for Registration of 3-D Shapes," PAMI 1992
  - Chen, Medioni, "Object Modelling by Registration of Multiple Range Images," International Journal of Image and Vision Computing, 1992.
  - Z. Zhang, Iterative point matching for registration of free-form curves, Research Report 1658, INRIA Sophia-Antipolis.
- Many variants (see Rusinkiewicz survey, 3DIM 01)
- Still in wide-spread use

1995: Depth from Defocus [Nayar et al., ICCV]

video

Real-time range sensor

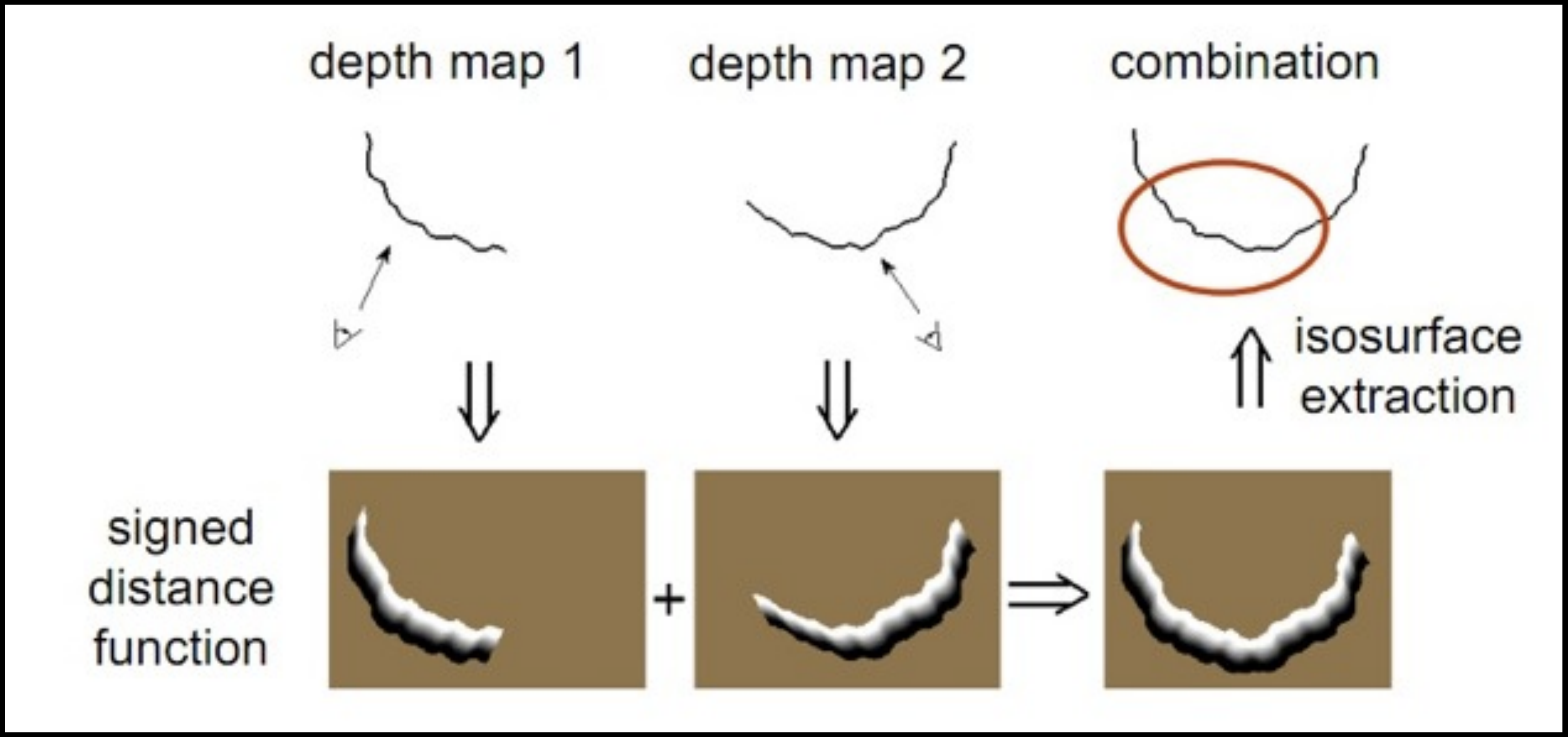
- Take two images with different focus blur
- Introduced "telecentric optics"
- 30Hz, accurate to 0.3%

1996: Façade [Debevec, Taylor, Malik, SIGGRAPH]

Introduced "Image-based Modeling"

- Key innovations include
  - View-dependent texture mapping
  - Model-based stereo
- Inspired Canoma, Google Sketchup, the Matrix effects

1996: Range scan merging [Curless; Hilton]

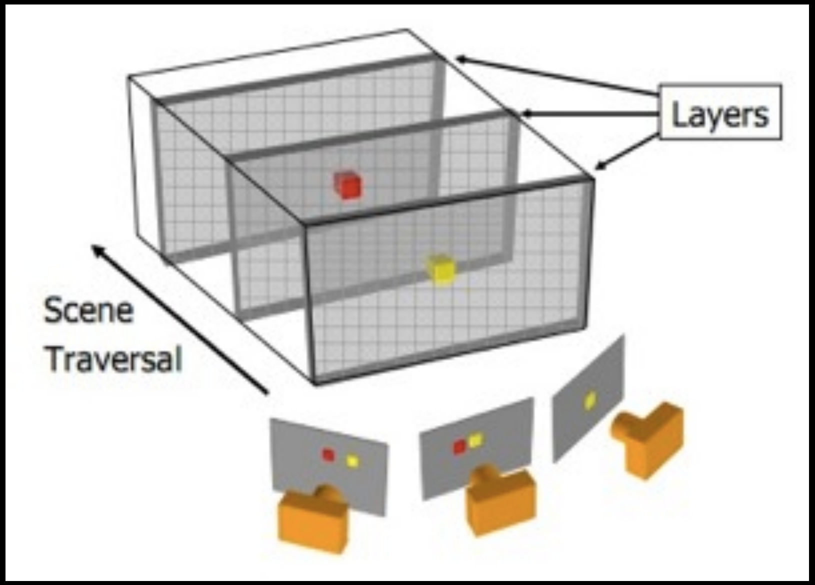


Curless & Levoy, SIGGRAPH 1996
Hilton et al., ECCV 1996

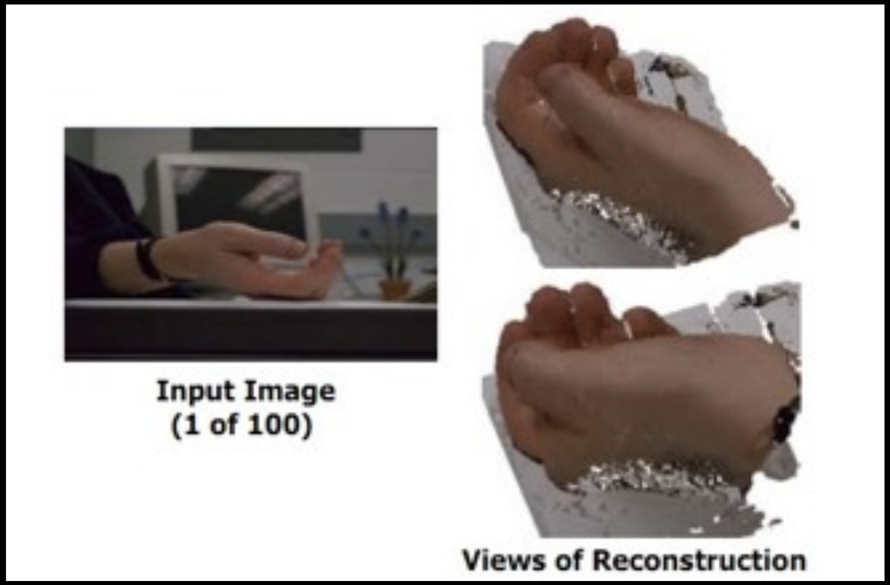
Optimal merging of range scans

- elegant and efficient: sum up signed distance xforms
- method of choice for a decade
  - until Poisson Surface Reconstruction [Kazhan et al, 2006]

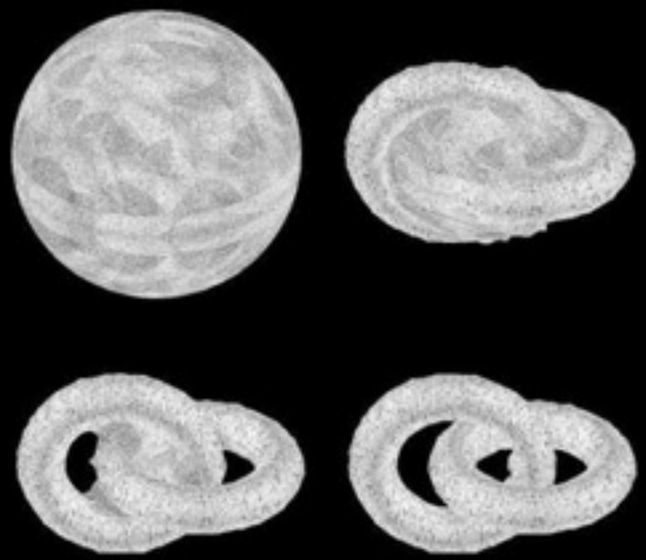
1997: Multi-view Stereo



Voxel Coloring
Seitz & Dyer, CVPR 1997



Space Carving
Kutulakos & Seitz, ICCV 1999
Fromherz & Bichsel, ISPRS 1995

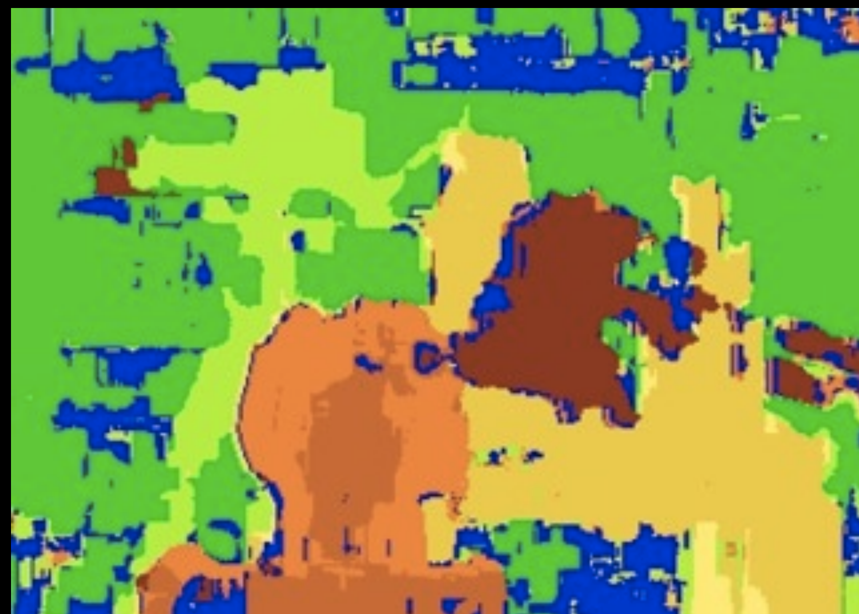


Level-set stereo
Faugeras & Keriven, ECCV 1998

Space Carving + Level Set Stereo

- reconstruct 3D directly rather than image matching
  - key work in photogrammetry: object-based least squares correlation [Helava; Ebner 1988], also Grün & Baltsavias: Geometrically constrained least squares matching PERS, 1988.
- proper modeling of visibility
- provable convergence properties

1998: Graph cut Stereo [Boykov, Veksler, Zabih, CVPR]



window matching

graph cuts

Global optimization for MRFs [Geman & Geman, 1984]

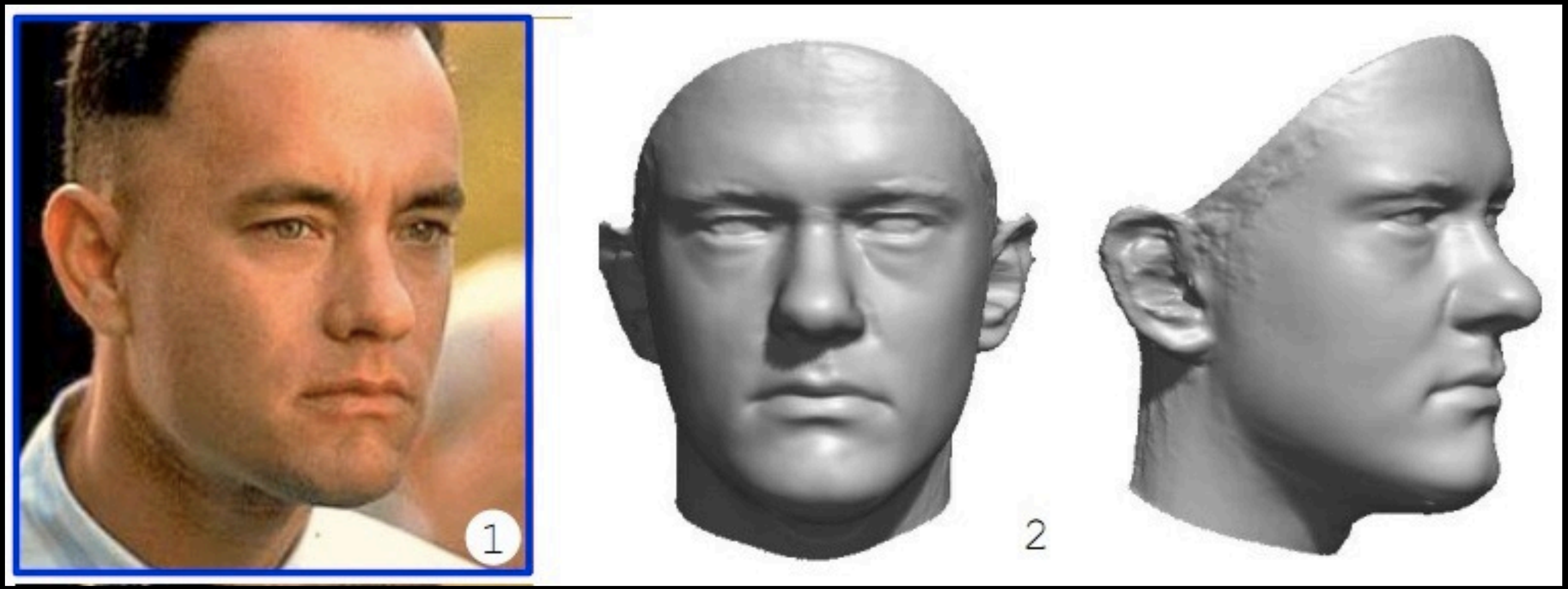
- ushered in the "graph cuts era" for computer vision
  - applications to many other problems
- extends dynamic programing stereo to 2D
  - [Baker & Binford, IJCAI, 1981]
- Also, [Roy & Cox, ICCV 1998]

1998: Uncalibrated 3D [Pollefeys, Koch, Van Gool, ICCV]

3D stereo models from uncalibrated video

- culmination of many research advances
  - tracking, SfM, self-calibration, stereo, texture-mapping
    - used SfM approach by [Beardsley, Torr, Zisserman., ECCV 1996]
- major milestone for projective reconstruction research

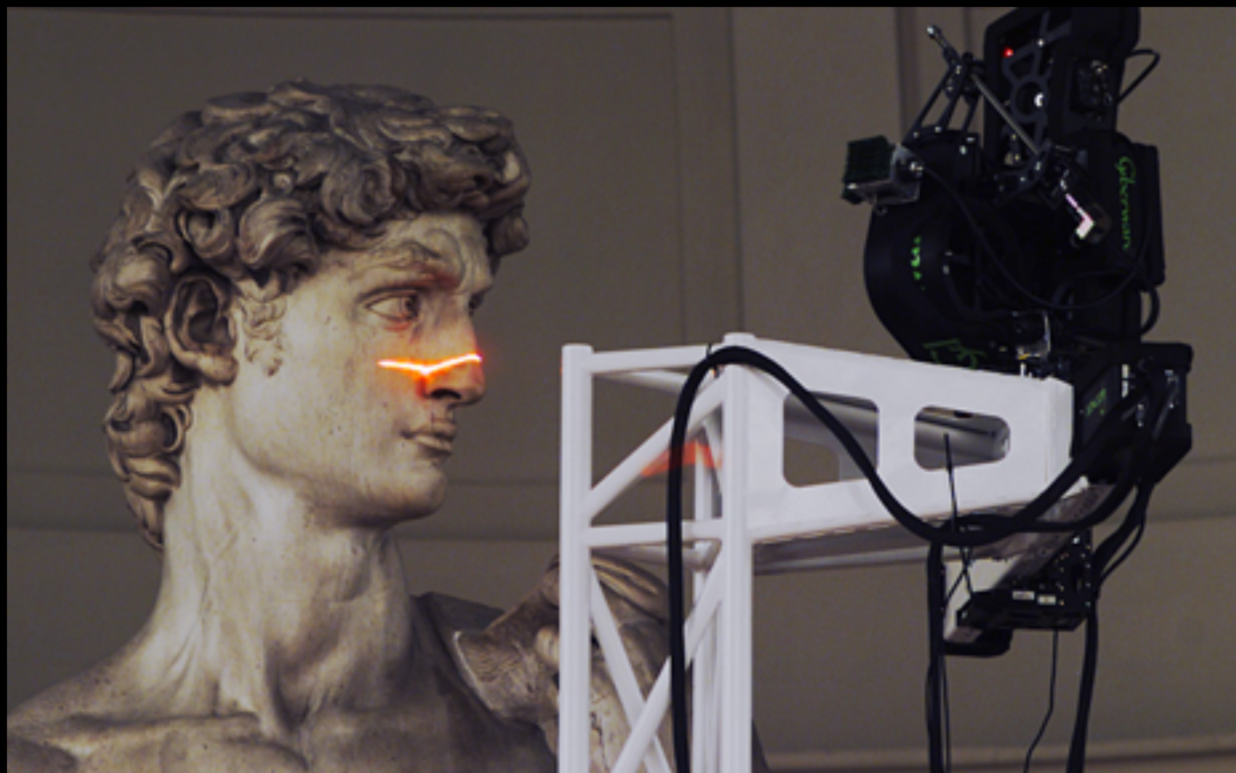
1999: Morphable Models [Blanz & Vetter, SIGGRAPH]

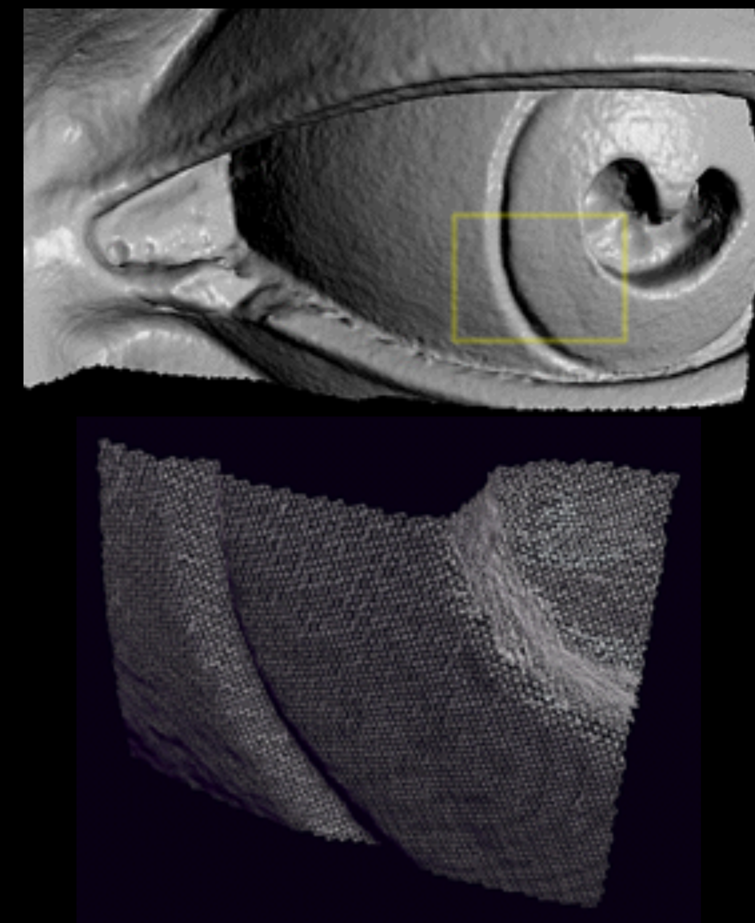


Single-view 3D face reconstruction

- linear combination of 200 laser scans of faces
- builds on Active Appearance Models [Cootes et al., 1998]
- stunningly good results (even 10+ years later)
  - convincing case for "model-based" approaches

1999: Digital Michelangelo Project [Levoy et al., SIGGRAPH]





Capture a full sculpture at 1/4mm precision

- 2 billion polygons, 250GB data
- took a month of laser scanning (at night) by a large team
- Also see the *Pieta project*
  - F. Bernardini, I. Martin, J. Mittleman, H. Rushmeier, G. Taubin. Building a Digital Model of Michelangelo's Florentine Pieta'. IEEE Computer Graphics & Applications, Jan/Feb. 2002, 22(1), pp. 59-67.

1999: Camera tracking ("Match Move")

Jurassic Park

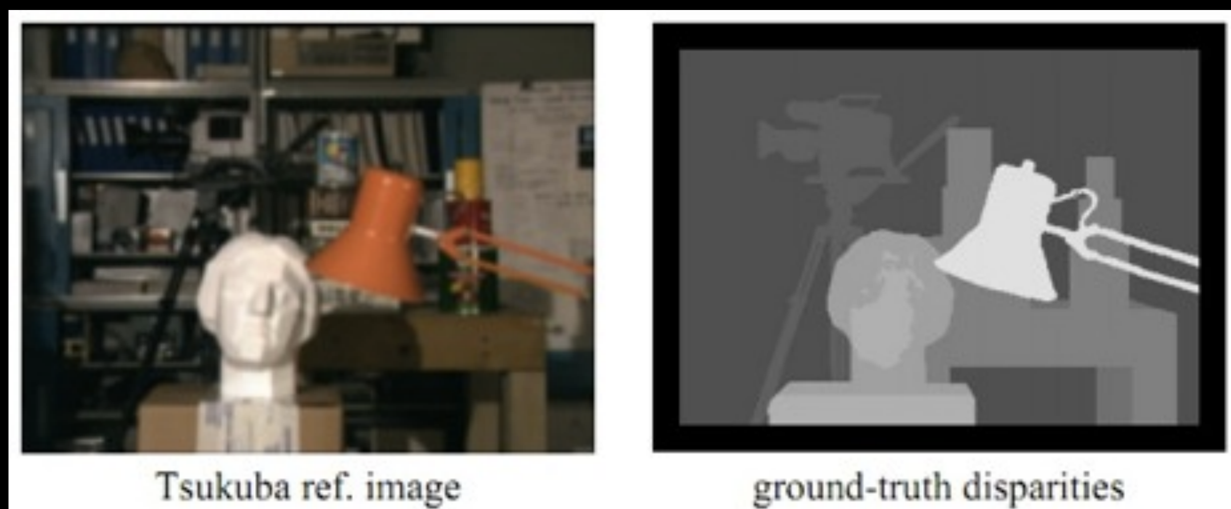
First automated 3D tracking systems

- (for GENERAL video)
  - feature tracking, bundle adjustment, CG overlays
  - Widespread use for movie special effects

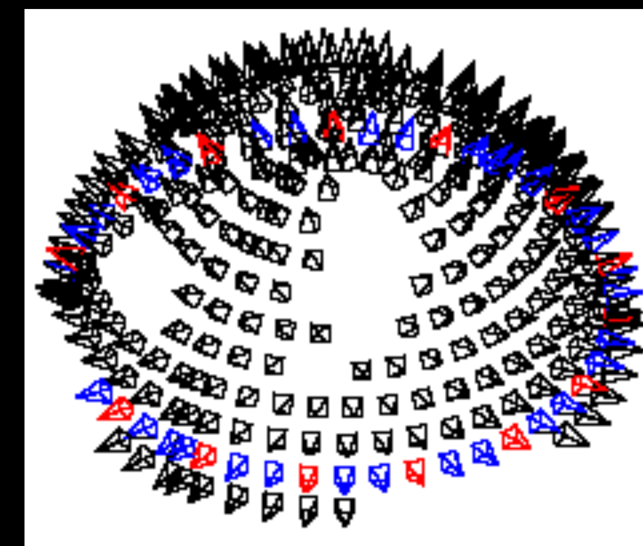
Pioneering Commercial Systems

- MatchMover [1999], grew out of Inria [L. Robert et al.]
- Boujou [2001], grew out of Oxford [Fitzgibbon et al.]
- Concurrent work in photogrammetry community

2001: Middlebury Benchmarks [Scharstein et al.]



Binocular stereo
Scharstein et al. 2001

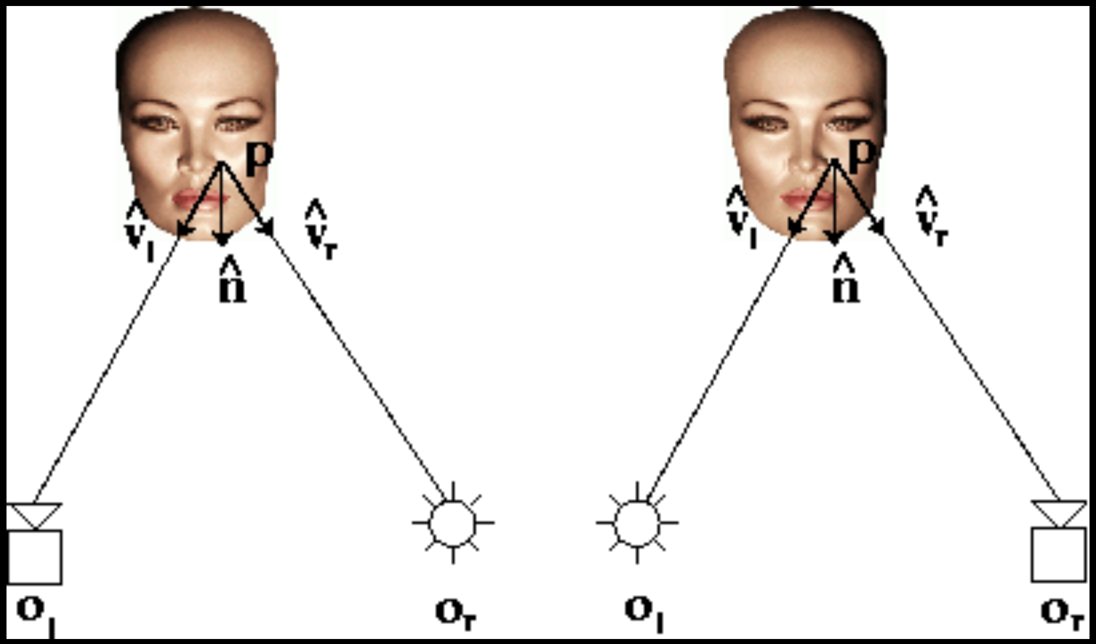


Multiview stereo
Seitz et al. 2006

Rigorous evaluation of stereo algorithms

- Major step forward for the field
- Ushered in an era of rapid progress

2002: Helmholtz Stereo [Zickler, Belhumeur, Kriegman, ECCV]



First BRDF-invariant stereo method

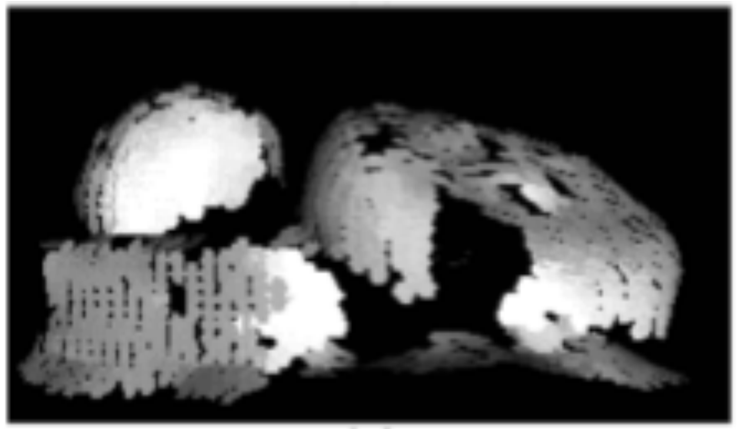
- exploits "Helmholtz" reciprocity
  - for each point, ratio of incident to reflected light is the same if you swap the camera and image

2003: Stereo = Laser Scan [Hernandez & Schmitt]

Silhouette and Stereo Fusion for 3D Object Modeling
Hernández, Schmitt, 3DIM 2003

Laser-scan quality 3D from stereo

- *robust* pair-wise sub-pixel image correlation + volumetric fusion
- first of new wave of extremely robust multi-view stereo methods: 0.25% error (.36mm)
  - see mview evaluation and online benchmarks

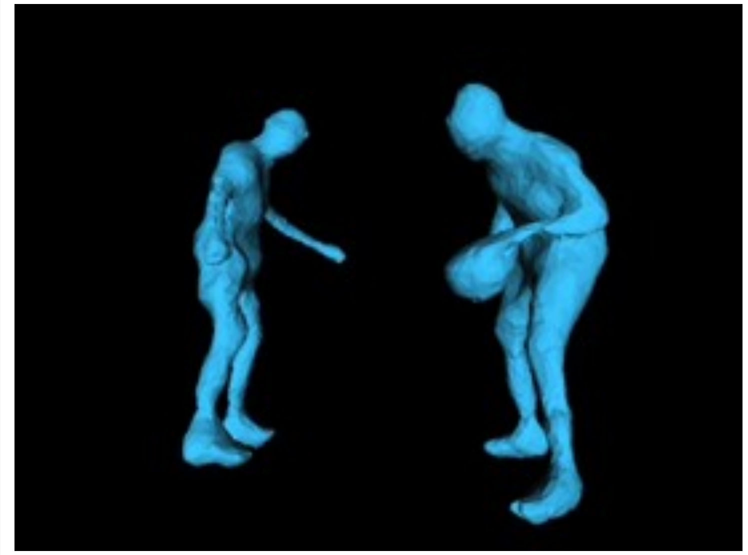


Fua

1995

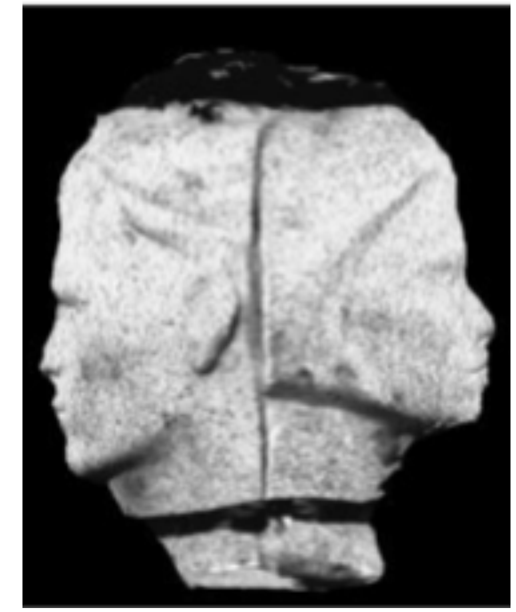
Seitz, Dyer

1997



Narayanan, Rander, Kanade

1998

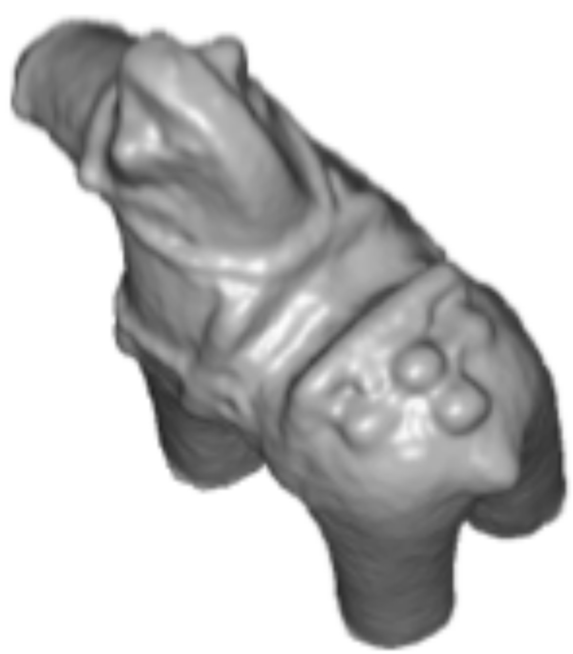


Faugeras, Keriven

1998

Hernandez, Schmitt

2004

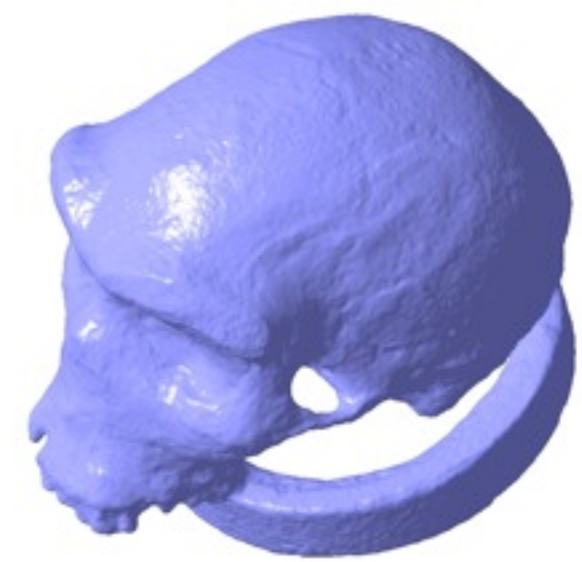


Vogiatzis, Torr, Cipolla

2005

Pons, Keriven, Faugeras

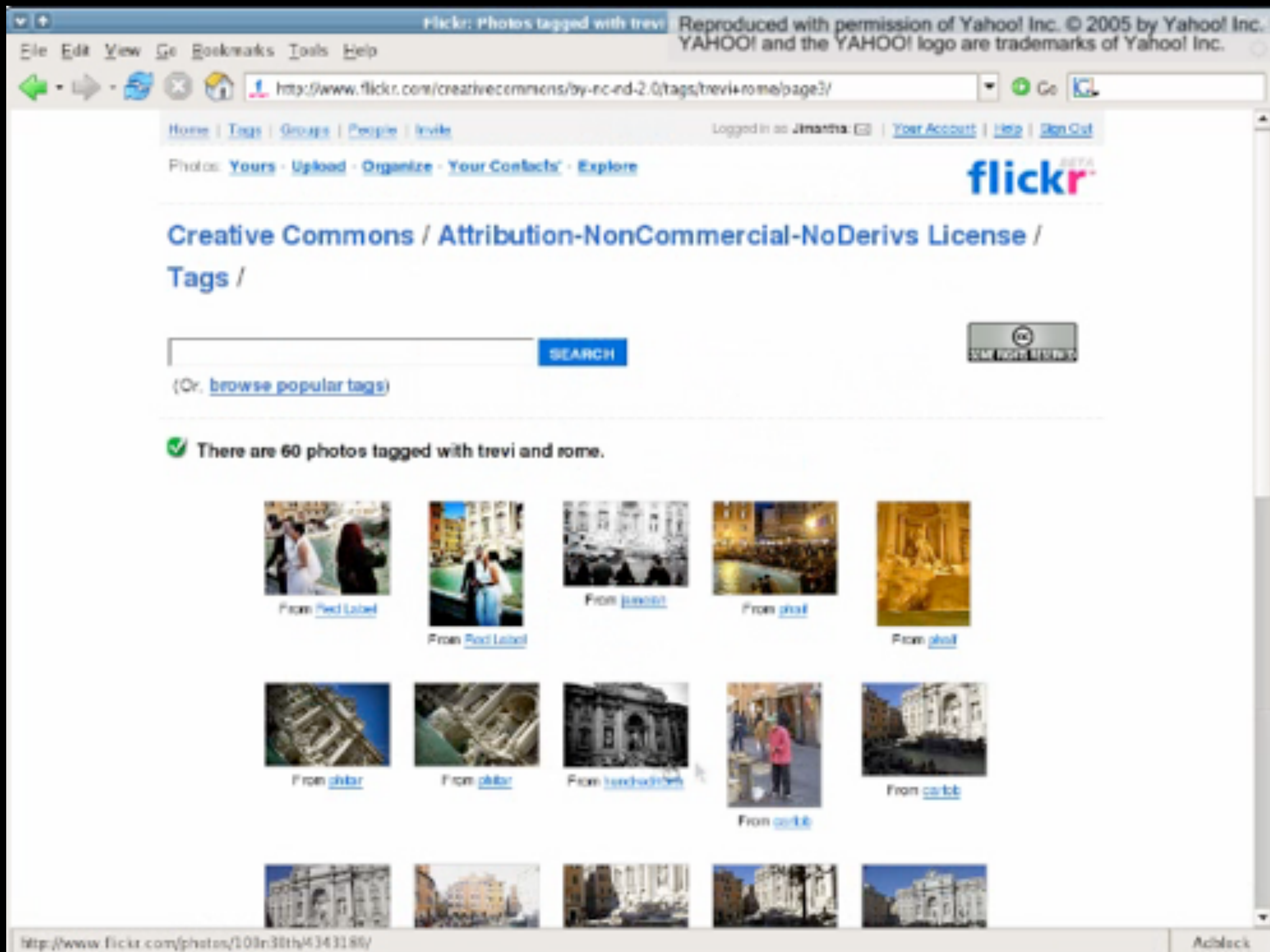
2005



Furukawa, Ponce

2006

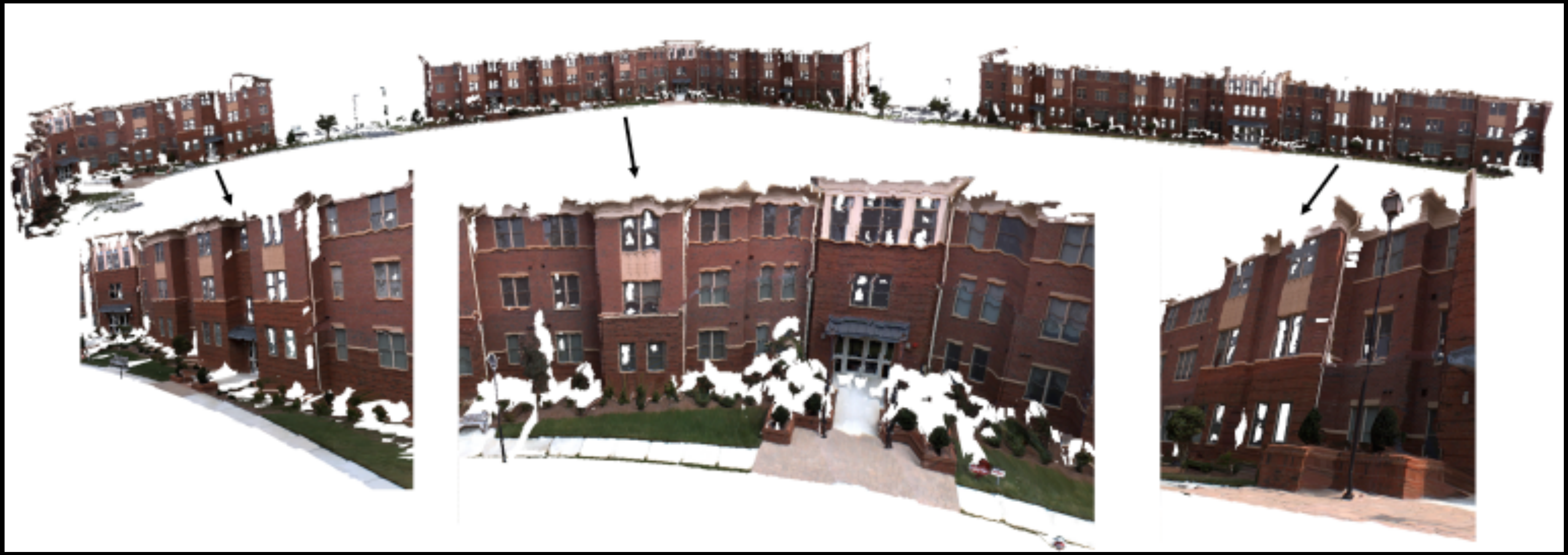
2006: Photo Tourism [Snavely, Seitz, Szeliski, SIGGRAPH]



3D reconstruction from Internet photos

- led to Microsoft's Photosynth
- key ingredients: SIFT [Lowe 1999], bundle adjustment
- earlier work: [Schaffalitzky & Zisserman 2002]; [Brown & Lowe 2005]
  - Schaffalitzky, Zisserman: Multi-view Matching for Unordered Image Sets, or "How Do I Organize My Holiday Snaps?". ECCV, 2002
  - Brown, Lowe: Unsupervised 3D Object Recognition and Reconstruction in Unordered Datasets. 3DIM 2005

2006: Urbanscape [Nister, Pollefeys, et al.]



video

Real-time 3D scene capture

- 4 cameras + GPS/IMU
  - feature tracking, triangulation, stereo, depth fusion, meshing

2008: Aerial 3D

Zebedin, Bauer, Karner, Bischof,
Fusion of Feature- and Area-Based Information for Urban Buildings Modeling from Aerial Imagery, ECCV

Fully automated pipeline ("Automated Facade")

- road/tree detection, stereo, fit building models

2009: Rome-in-a-day [Agarwal, Snavely, Simon, Seitz, Szeliski, ICCV]

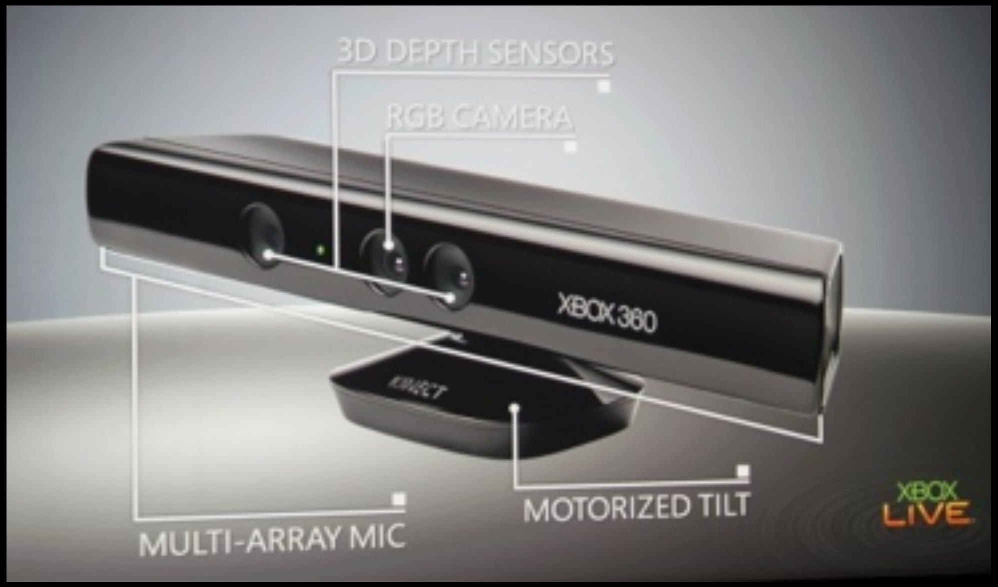
Dubrovnik

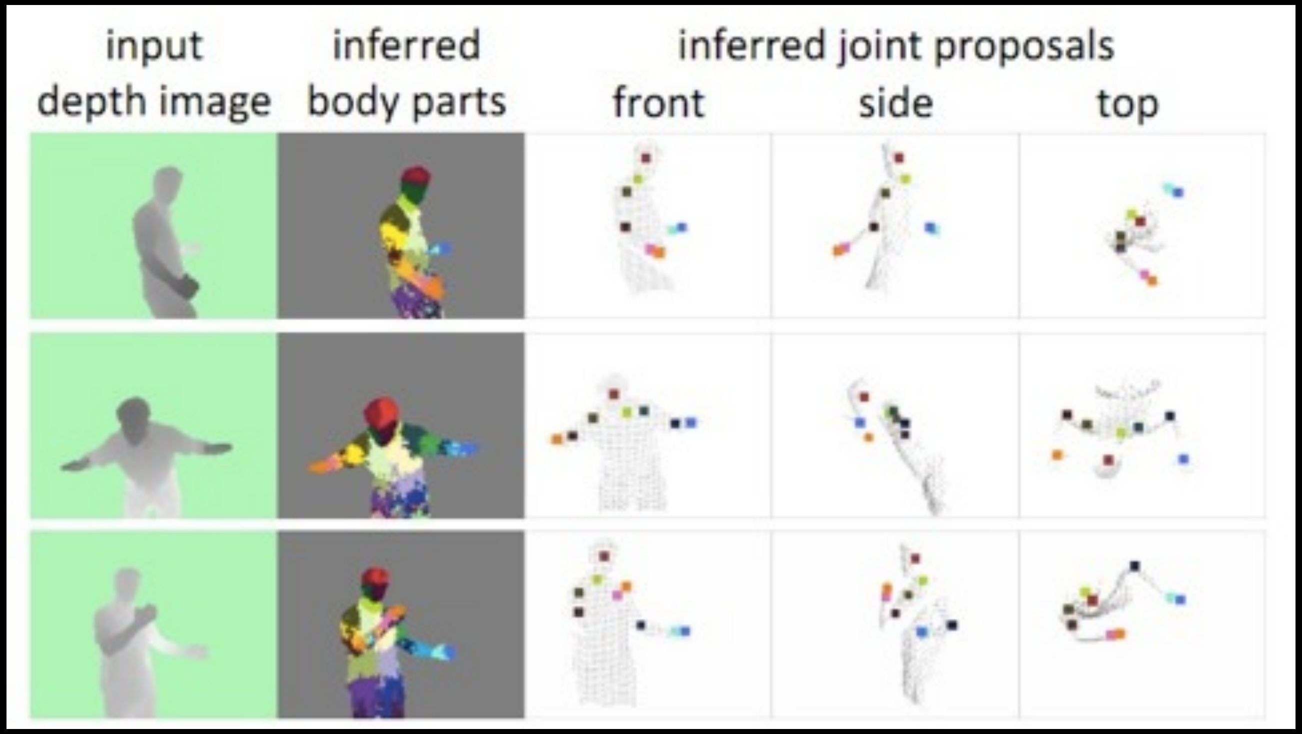
City-scale SfM

- ~500 cores ~200K images, one day of processing
- 3 cities: Rome, Venice, Dubrovnik
- also see “Rome in a Cloudless Day”, Frahm et al, ECCV 2010

2011: Kinect

Shotton, Fitzgibbon, Cook, Sharp, Finocchio, Moore, Kipman, Blake, Real-Time Human Pose Recognition in Parts from a Single Depth Image, *CVPR*





Body pose from single depth image

- 200 fps!
- Fastest selling electronic device in history (beat ipad)!

2013: Digital Michelangelo from a few photos

Key: knowledge-based hole filling

2015: Models of everything

Public repository of the world's geometry

- people, places, and things
- interiors and exteriors
- community contributed
- quality improves over time

2015: Reconstruct me

From my photo collection

- all my poses, expressions, body shapes over time

2020: Inverse CAD

CAD models from photographs

- what an architect would call a model
- capture the salient features, not the extraneous details
- recognize and label everything in the scene
- easy to edit, modify

2020: Visual Turing Test

Reconstructions indistinguishable from reality

- interactive, real-time
- at close range
- from sparse photos

2030: computer > human