3D photography

Digital Visual Effects, Spring 2008

Yung-Yu Chuang

2008/5/13

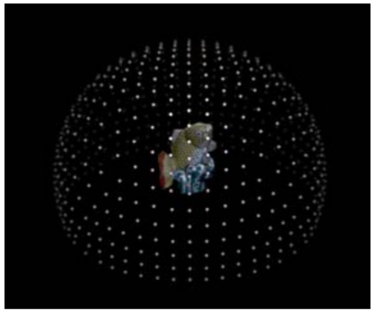
with slides by Szymon Rusinkiewicz, Richard Szeliski, Steve Seitz and Brian Curless

Announcements

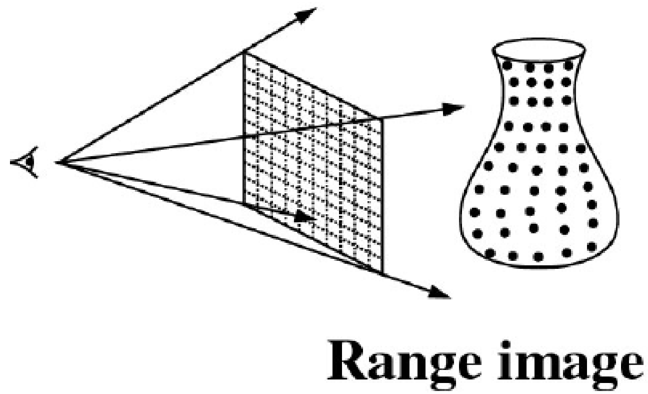
- Project #3 is due on 5/19. Two videos. Send us links only.
- We may have final project proposal presentation on 5/27. Please send me your team members and topic by 5/26.
- Final project demo day will be 1:30pm on 6/25 (Wed). Room to be announced.

3D photography

- Acquisition of geometry and material



Range acquisition



Range acquisition taxonomy

- range acquisition
  - contact
    - mechanical (CMM, jointed arm)
    - inertial (gyroscope, accelerometer)
    - ultrasonic trackers
    - magnetic trackers
  - transmissive
    - industrial CT
    - ultrasound
    - MRI
  - reflective
    - non-optical
      - radar
      - sonar
    - optical

Range acquisition taxonomy

- optical methods
  - passive
    - shape from X:
      - stereo
      - motion
      - shading
      - texture
      - focus
      - defocus
  - active
    - active variants of passive methods
      - Stereo w. projected texture
      - Active depth from defocus
      - Photometric stereo
    - time of flight
    - triangulation

Outline

- Passive approaches
  - Stereo
  - Multiview approach
- Active approaches
  - Triangulation
  - Shadow scanning
- Active variants of passive approaches
  - Photometric stereo
  - Example-based photometric stereo

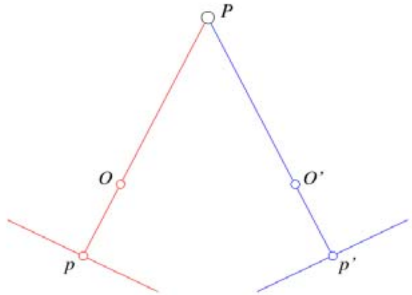
Passive approaches

Stereo

Public Library, Stereoscopic Looking Room, Chicago, by Phillips, 1923

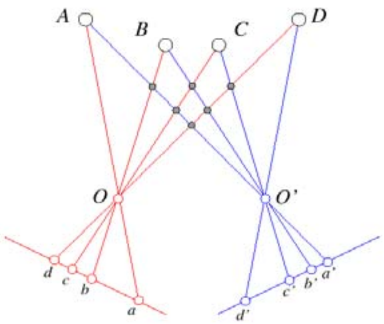
Stereo

- One distinguishable point being observed
  - The preimage can be found at the intersection of the rays from the focal points to the image points



Stereo

- Many points being observed
  - Need some method to establish correspondences

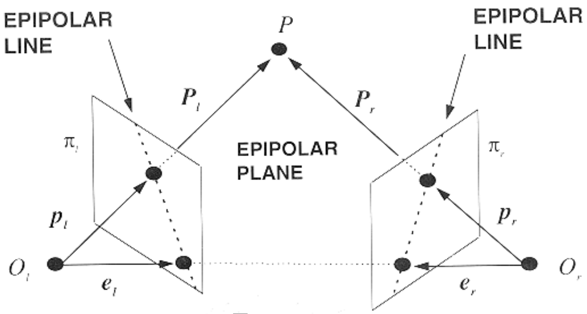


Components of stereo vision systems

- Camera calibration
- Image rectification: simplifies the search for correspondences
- Correspondence: which item in the left image corresponds to which item in the right image
- Reconstruction: recovers 3-D information from the 2-D correspondences

Epipolar geometry

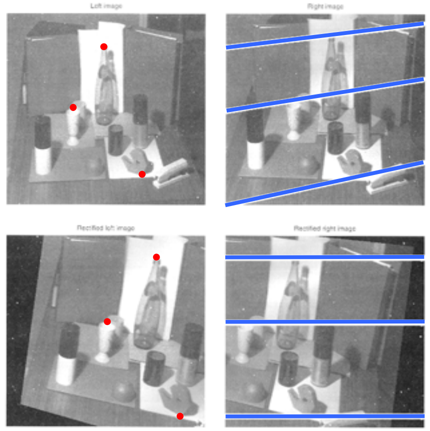
- Epipolar constraint: corresponding points must lie on conjugate epipolar lines
  - Search for correspondences becomes a 1-D problem



$$\mathbf{x}'^{\mathrm{T}} \mathbf{F} \mathbf{x} = 0$$

Image rectification

- Warp images such that conjugate epipolar lines become collinear and parallel to *u* axis

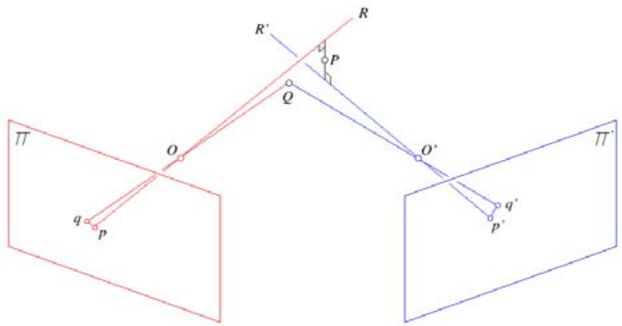


Disparity

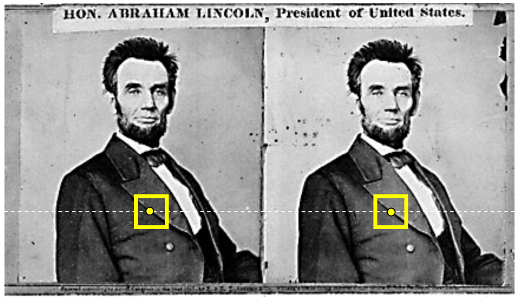
- With rectified images, disparity is just (horizontal) displacement of corresponding features in the two images
  - Disparity = 0 for distant points
  - Larger disparity for closer points
  - Depth of point proportional to 1/disparity

Reconstruction

- Geometric
  - Construct the line segment perpendicular to R and R' that intersects both rays and take its mid-point



Basic stereo algorithm



For each epipolar line

For each pixel in the left image

- compare with every pixel on same epipolar line in right image
- pick pixel with minimum match cost

Improvement: match ***windows***

Basic stereo algorithm

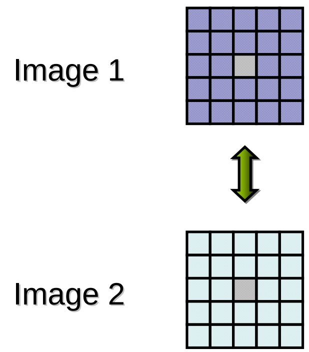
- For each pixel
  - For each disparity
    - For each pixel in window
      - Compute difference
  - Find disparity with minimum SSD

Reverse order of loops

- For each disparity
  - For each pixel
    - For each pixel in window
      - Compute difference
- Find disparity with minimum SSD at each pixel

Incremental computation

- Given SSD of a window, at some disparity

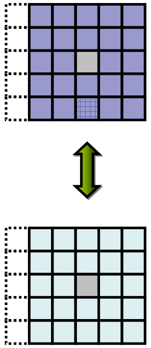


Incremental computation

- Want: SSD at next location

Image 1

Image 2

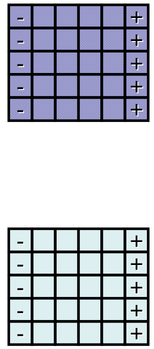


Incremental computation

- Subtract contributions from leftmost column, add contributions from rightmost column

Image 1

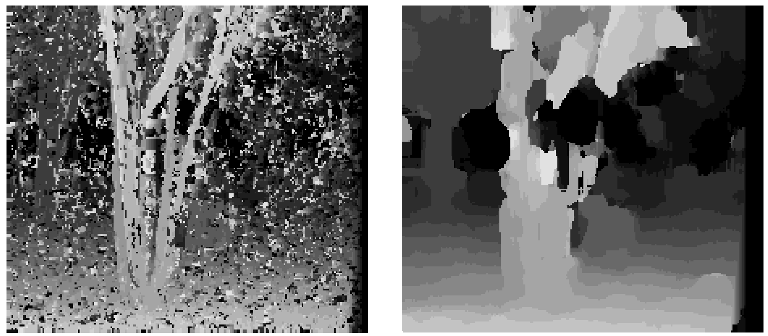
Image 2



Selecting window size

- Small window: more detail, but more noise
- Large window: more robustness, less detail
- Example:

Selecting window size



3 pixel window

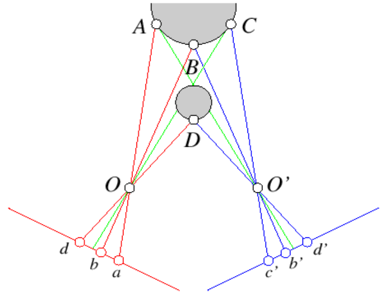
20 pixel window

Non-square windows

- Compromise: have a large window, but higher weight near the center
- Example: Gaussian
- Example: Shifted windows

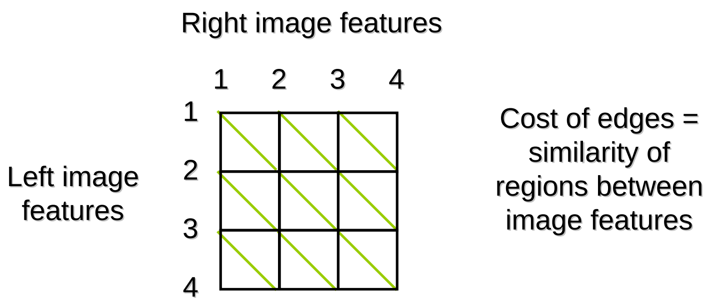
Ordering constraint

- Order of matching features usually the same in both images
- But not always: occlusion



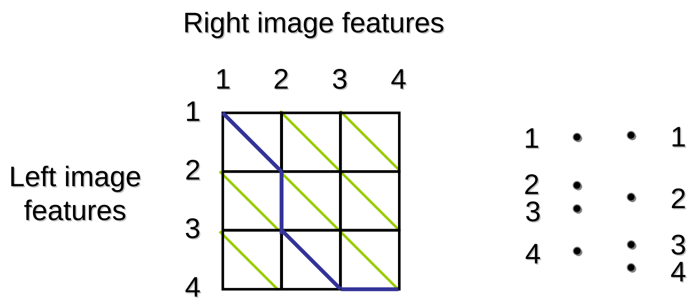
Dynamic programming

- Treat feature correspondence as graph problem



Dynamic programming

- Find min-cost path through graph



Energy minimization

- Another approach to improve quality of correspondences
- Assumption: disparities vary (mostly) smoothly
- Minimize energy function:

$$E_{data}+\lambda E_{smoothness}$$

- E_{data}: how well does disparity match data
- $E_{smoothness}$: how well does disparity match that of neighbors – regularization

Energy minimization

- If data and energy terms are nice (continuous, smooth, etc.) can try to minimize via gradient descent, etc.
- In practice, disparities only piecewise smooth
- Design smoothness function that doesn't penalize large jumps too much
  - Example: $V(\alpha,\beta)=\min(|\alpha-\beta|, K)$

Stereo as energy minimization

- Matching Cost Formulated as Energy
  - "data" term penalizing bad matches

$$D(x,y,d) = |\mathbf{I}(x,y) - \mathbf{J}(x+d,y)|$$

  - "neighborhood term" encouraging spatial smoothness

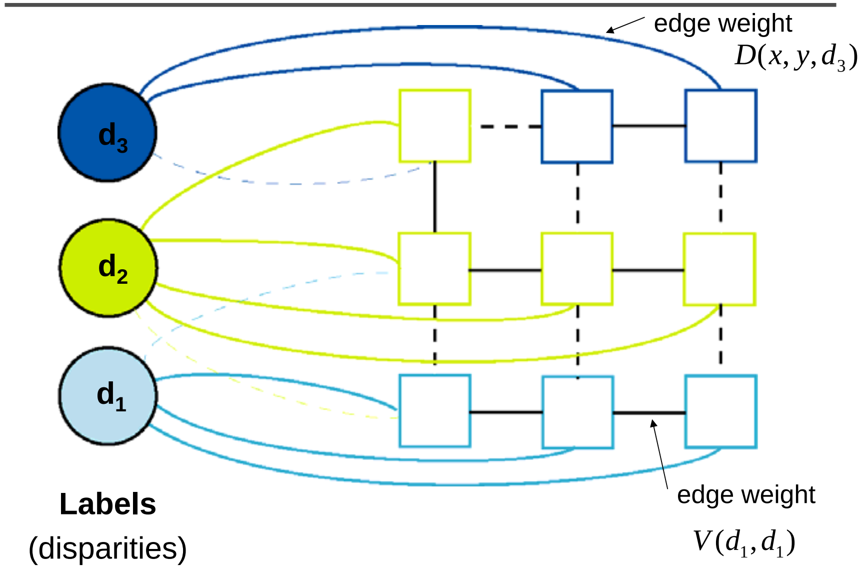
$$V(d_1,d_2) = \text{cost of adjacent pixels with labels d1 and d2}$$
$$= |d_1 - d_2| \quad \text{(or something similar)}$$

$$E = \sum_{(x,y)} D(x,y,d_{x,y}) + \sum_{neighbors\ (x1,y1),(x2,y2)} V(d_{x1,y1}, d_{x2,y2})$$

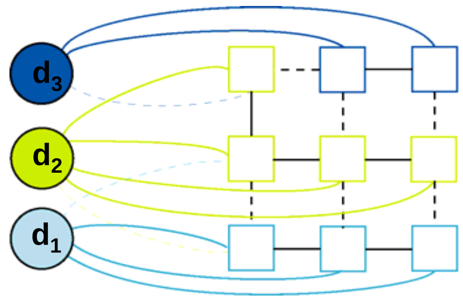
Energy minimization

- Hard to find global minima of non-smooth functions
  - Many local minima
  - Provably NP-hard
- Practical algorithms look for approximate minima (e.g., simulated annealing)

Energy minimization via graph cuts

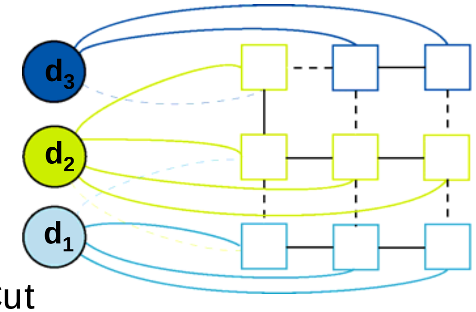


Energy minimization via graph cuts



- Graph Cost
  - Matching cost between images
  - Neighborhood matching term
  - Goal: figure out which labels are connected to which pixels

Energy minimization via graph cuts

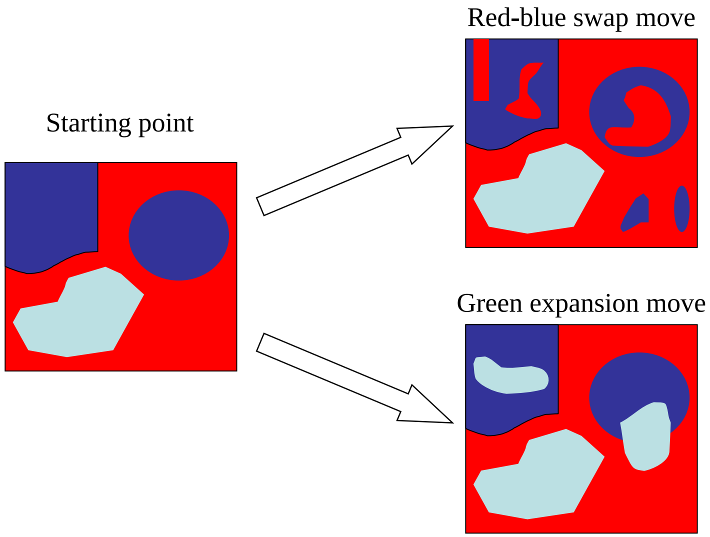


- Graph Cut
  - Delete enough edges so that
    - each pixel is (transitively) connected to exactly one label node
  - Cost of a cut: sum of deleted edge weights
  - Finding min cost cut equivalent to finding global minimum of energy function

Computing a multiway cut

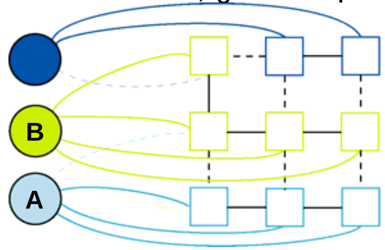
- With 2 labels: classical min-cut problem
  - Solvable by standard flow algorithms
    - polynomial time in theory, nearly linear in practice
  - More than 2 terminals: NP-hard
    [Dahlhaus *et al.*, STOC '92]
- Efficient approximation algorithms exist
  - Within a factor of 2 of optimal
  - Computes local minimum in a strong sense
    - even very large moves will not improve the energy
  - Yuri Boykov, Olga Veksler and Ramin Zabih, Fast Approximate Energy Minimization via Graph Cuts, International Conference on Computer Vision, September 1999.

Move examples

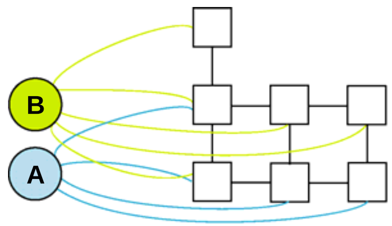


The swap move algorithm

1. Start with an arbitrary labeling
2. Cycle through every label pair *(A,B)* in some order
    - 2.1 Find the lowest *E* labeling within a single *AB*-swap
    - 2.2 Go there if *E* is lower than the current labeling
3. If *E* did not decrease in the cycle, we're done
   Otherwise, go to step 2



Original graph



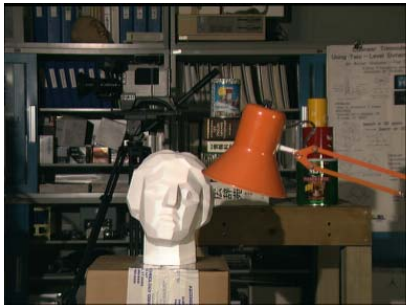
AB subgraph
(run min-cut on this graph)

The expansion move algorithm

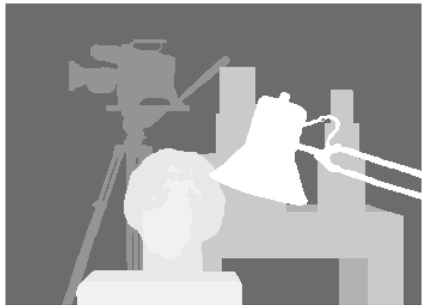
1. Start with an arbitrary labeling
2. Cycle through every label *A* in some order
    - 2.1 Find the lowest *E* labeling within a single *A*-expansion
    - 2.2 Go there if it *E* is lower than the current labeling
3. If *E* did not decrease in the cycle, we're done
   Otherwise, go to step 2

Stereo results

– Data from University of Tsukuba



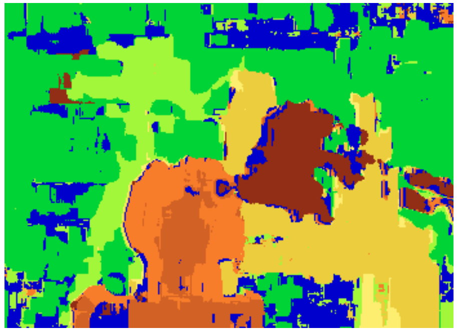
scene



ground truth

http://cat.middlebury.edu/stereo/

Results with window correlation



normalized correlation
(best window size)

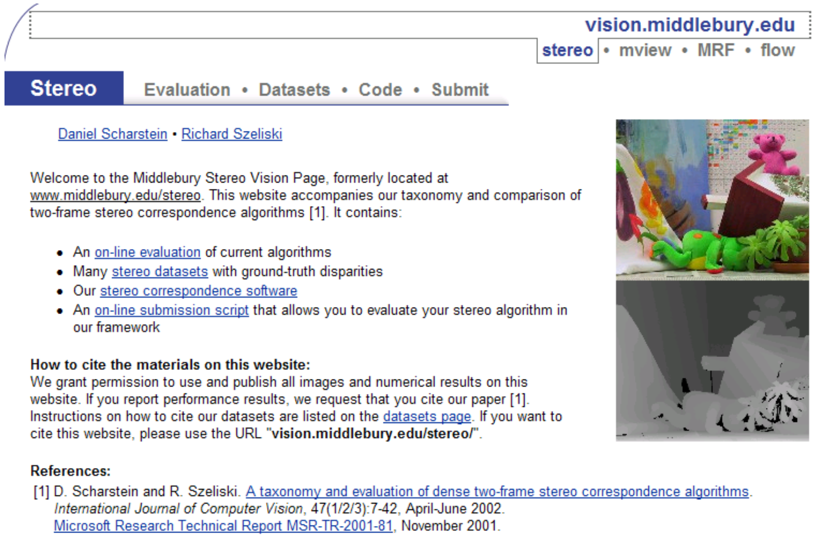
ground truth

Results with graph cuts

graph cuts
(Potts model *E*,
expansion move algorithm)

ground truth

Stereo evaluation



vision.middlebury.edu

stereo • mview • MRF • flow

Stereo   Evaluation • Datasets • Code • Submit

Daniel Scharstein • Richard Szeliski

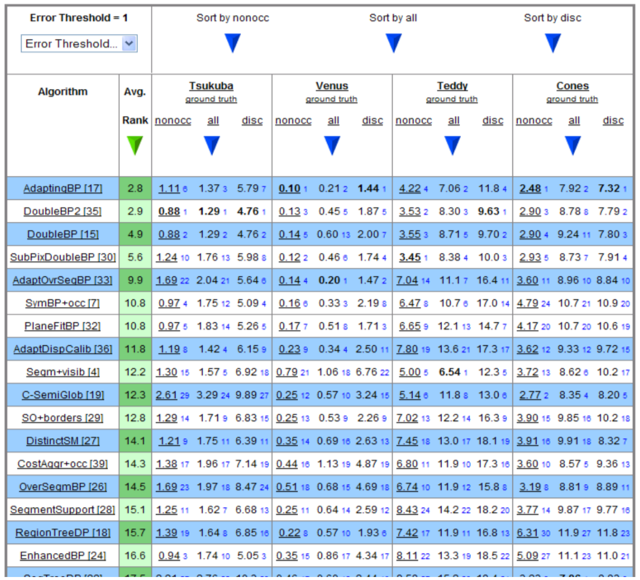
Welcome to the Middlebury Stereo Vision Page, formerly located at www.middlebury.edu/stereo. This website accompanies our taxonomy and comparison of two-frame stereo correspondence algorithms [1]. It contains:

- An on-line evaluation of current algorithms
- Many stereo datasets with ground-truth disparities
- Our stereo correspondence software
- An on-line submission script that allows you to evaluate your stereo algorithm in our framework

How to cite the materials on this website:
We grant permission to use and publish all images and numerical results on this website. If you report performance results, we request that you cite our paper [1]. Instructions on how to cite our datasets are listed on the datasets page. If you want to cite this website, please use the URL "**vision.middlebury.edu/stereo/**".

References:
[1] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1/2/3):7-42, April-June 2002. Microsoft Research Technical Report MSR-TR-2001-81, November 2001.

Stereo—best algorithms



Error Threshold = 1 — Error Threshold... — Sort by nonocc — Sort by all — Sort by disc

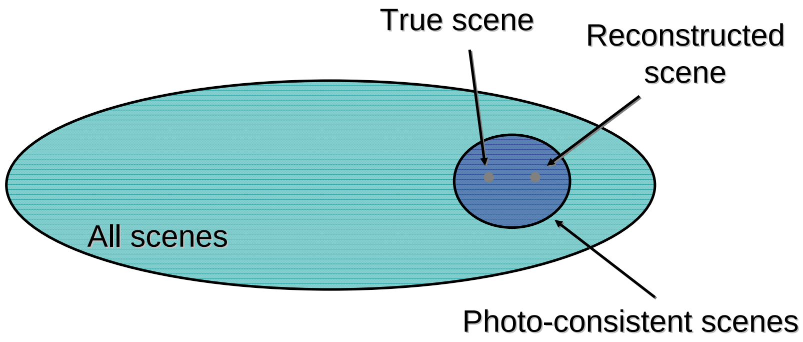
Algorithm	Avg. Rank	Tsukuba ground truth nonocc	Tsukuba all	Tsukuba disc	Venus ground truth nonocc	Venus all	Venus disc	Teddy ground truth nonocc	Teddy all	Teddy disc	Cones ground truth nonocc	Cones all	Cones disc
AdaptingBP [17]	2.8	1.11 6	1.37 3	5.79 7	**0.10** 1	0.21 2	**1.44** 1	4.22 4	7.06 2	11.8 4	**2.48** 1	7.92 2	**7.32** 1
DoubleBP2 [35]	2.9	**0.88** 1	**1.29** 1	**4.76** 1	0.13 3	0.45 5	1.87 5	3.53 2	8.30 3	**9.63** 1	2.90 3	8.78 8	7.79 2
DoubleBP [15]	4.9	0.88 2	1.29 2	4.76 2	0.14 5	0.60 13	2.00 7	3.55 3	8.71 5	9.70 2	2.90 4	9.24 11	7.80 3
SubPixDoubleBP [30]	5.6	1.24 10	1.76 13	5.98 8	0.12 2	0.46 6	1.74 4	**3.45** 1	8.38 4	10.0 3	2.93 5	8.73 7	7.91 4
AdaptOvrSegBP [33]	9.9	1.69 22	2.04 21	5.64 6	0.14 4	**0.20** 1	1.47 2	7.04 14	11.1 7	16.4 11	3.60 11	8.96 10	8.84 10
SymBP+occ [7]	10.8	0.97 4	1.75 12	5.09 4	0.16 6	0.33 3	2.19 8	6.47 8	10.7 6	17.0 14	4.79 24	10.7 21	10.9 20
PlaneFitBP [32]	10.8	0.97 5	1.83 14	5.26 5	0.17 7	0.51 8	1.71 3	6.65 9	12.1 13	14.7 7	4.17 20	10.7 20	10.6 19
AdaptDispCalib [36]	11.8	1.19 8	1.42 4	6.15 9	0.23 9	0.34 4	2.50 11	7.80 19	13.6 21	17.3 17	3.62 12	9.33 12	9.72 15
Segm+visib [4]	12.2	1.30 15	1.57 5	6.92 18	0.79 21	1.06 18	6.76 22	5.00 5	**6.54** 1	12.3 5	3.72 13	8.62 6	10.2 17
C-SemiGlob [19]	12.3	2.61 29	3.29 24	9.89 27	0.25 12	0.57 10	3.24 15	5.14 6	11.8 8	13.0 6	2.77 2	8.35 4	8.20 5
SO+borders [29]	12.8	1.29 14	1.71 9	6.83 15	0.25 13	0.53 9	2.26 9	7.02 13	12.2 14	16.3 9	3.90 15	9.85 16	10.2 18
DistinctSM [27]	14.1	1.21 9	1.75 11	6.39 11	0.35 14	0.69 16	2.63 13	7.45 18	13.0 17	18.1 19	3.91 16	9.91 18	8.32 7
CostAggr+occ [39]	14.3	1.38 17	1.96 17	7.14 19	0.44 16	1.13 19	4.87 19	6.80 11	11.9 10	17.3 16	3.60 10	8.57 5	9.36 13
OverSegmBP [26]	14.5	1.69 23	1.97 18	8.47 24	0.51 18	0.68 15	4.69 18	6.74 10	11.9 12	15.8 8	3.19 8	8.81 9	8.89 11
SegmentSupport [28]	15.1	1.25 11	1.62 7	6.68 13	0.25 11	0.64 14	2.59 12	8.43 24	14.2 22	18.2 20	3.77 14	9.87 17	9.77 16
RegionTreeDP [18]	15.7	1.39 19	1.64 8	6.85 16	0.22 8	0.57 10	1.93 6	7.42 17	11.9 11	16.8 13	6.31 30	11.9 27	11.8 23
EnhancedBP [24]	16.6	0.94 3	1.74 10	5.05 3	0.35 15	0.86 17	4.34 17	8.11 22	13.3 19	18.5 22	5.09 27	11.1 23	11.0 21

Volumetric multiview approaches

- Goal: find a model consistent with images
- "Model-centric" (vs. image-centric)
- Typically use discretized volume (voxel grid)
- For each voxel, compute occupied / free
  (for some algorithms, also color, etc.)

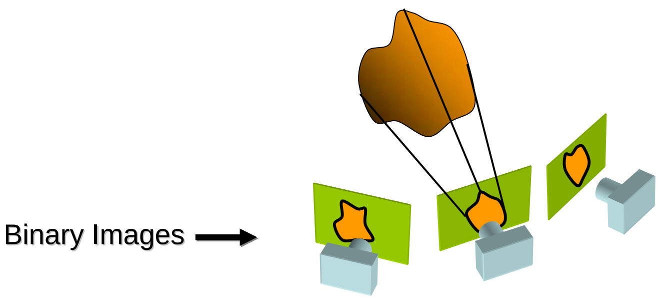
Photo consistency

- Result: not necessarily the correct scene
- Many scenes produce the same images



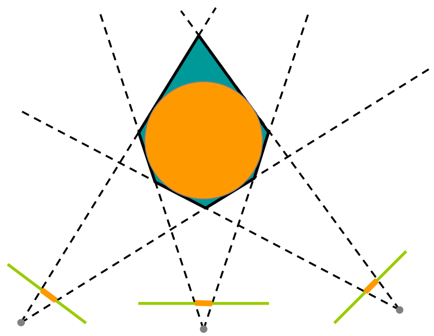
Silhouette carving

- Find silhouettes in all images
- Exact version:
  - Back-project all silhouettes, find intersection



Silhouette carving

- Find silhouettes in all images
- Exact version:
  - Back-project all silhouettes, find intersection



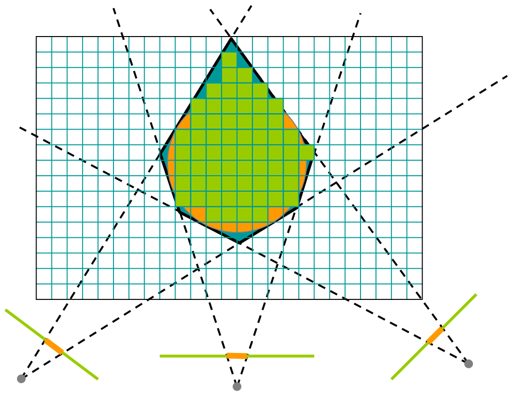
Silhouette carving

- Limit of silhouette carving is *visual hull* or *line hull*
- Complement of lines that don't intersect object
- In general not the same as object
  - Can't recover "pits" in object
- Not the same as convex hull

Silhouette carving

- Discrete version:
  - Loop over all voxels in some volume
  - If projection into images lies inside all silhouettes, mark as occupied
  - Else mark as free

Silhouette carving



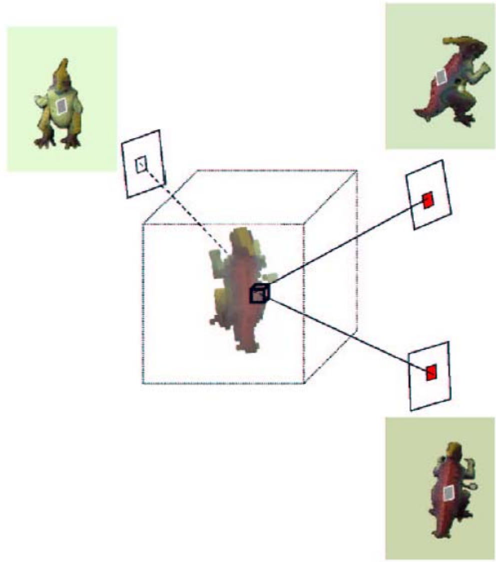
Voxel coloring

- Seitz and Dyer, 1997
- In addition to free / occupied, store color at each voxel
- Explicitly accounts for occlusion

Voxel coloring

- Basic idea: sweep through a voxel grid
  - Project each voxel into each image in which it is visible
  - If colors in images agree, mark voxel with color
  - Else, mark voxel as empty
- Agreement of colors based on comparing standard deviation of colors to threshold

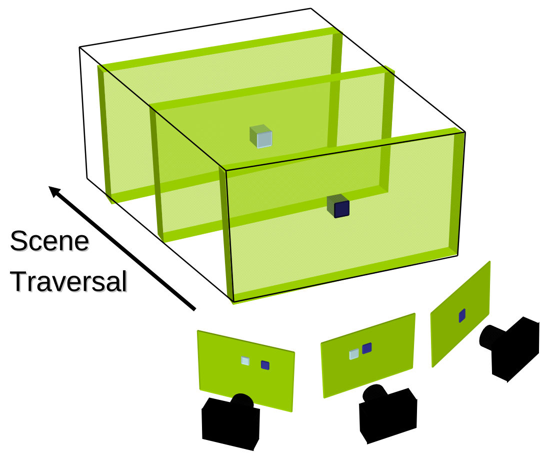
Voxel coloring and occlusion



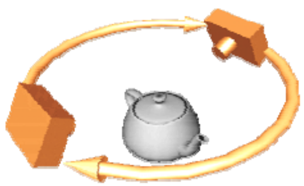
Voxel coloring and occlusion

- Problem: which voxels are visible?
- Solution: constrain camera views
  - When a voxel is considered, necessary occlusion information must be available
  - Sweep occluders before occludees
  - Constrain camera positions to allow this sweep

Voxel coloring sweep order



Voxel coloring camera positions



Inward-looking
Cameras above scene

Outward-looking
Cameras inside scene

Seitz

Image acquisition

Selected Dinosaur Images

Selected Flower Images

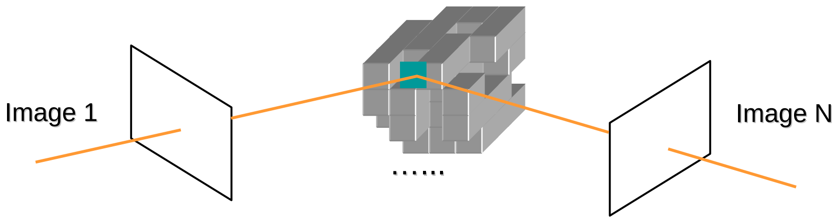
- Calibrated Turntable
- 360° rotation (21 images)

Voxel coloring results

Dinosaur Reconstruction
72 K voxels colored
7.6 M voxels tested
7 min. to compute
on a 250MHz SGI

Flower Reconstruction
70 K voxels colored
7.6 M voxels tested
7 min. to compute
on a 250MHz SGI

Space carving

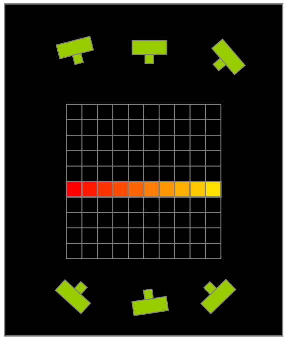


Initialize to a volume V containing the true scene
Choose a voxel on the current surface
Project to visible input images
Carve if not photo-consistent
Repeat until convergence

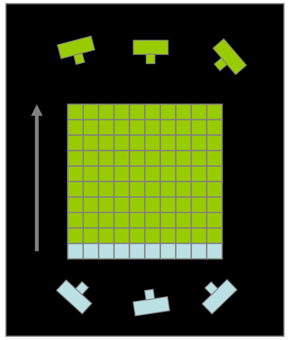
Multi-pass plane sweep

- Faster alternative:
  - Sweep plane in each of 6 principal directions
  - Consider cameras on only one side of plane
  - Repeat until convergence

Multi-pass plane sweep

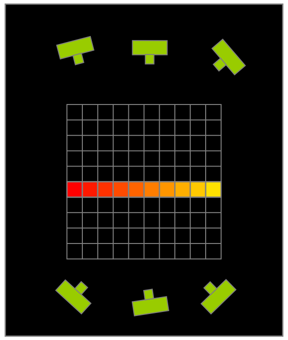


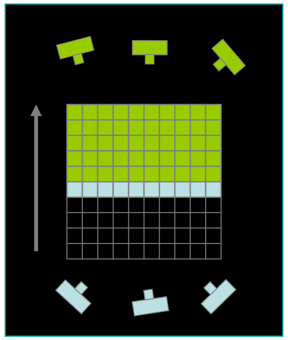
True Scene



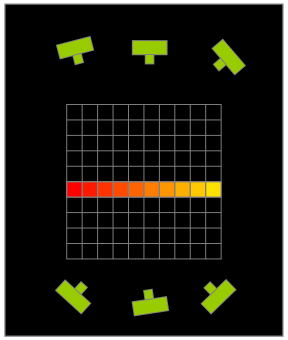
Reconstruction

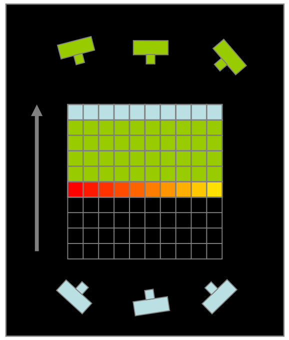
Multi-pass plane sweep



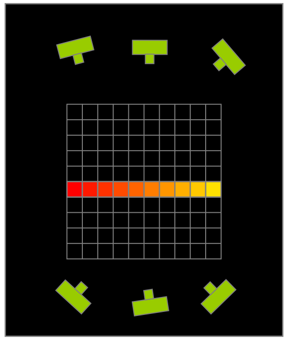


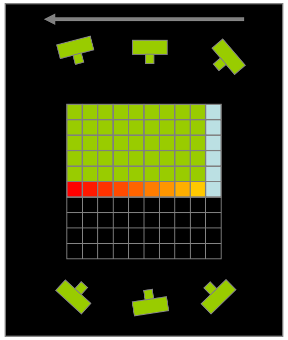
Multi-pass plane sweep



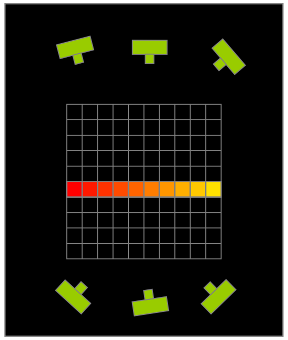


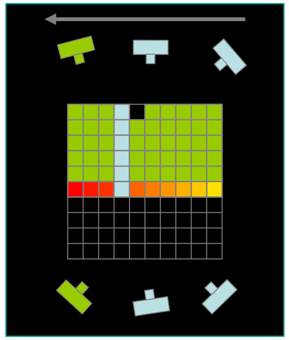
Multi-pass plane sweep





Multi-pass plane sweep

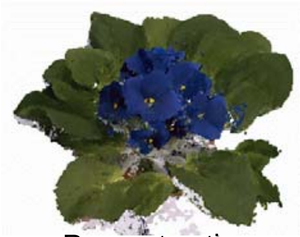




Space carving results: African violet

Input image (1 of 45)

Reconstruction



Reconstruction

Reconstruction

Space carving results: hand

Input image (1 of 100)

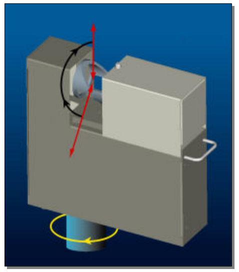
Reconstruction

Active approaches

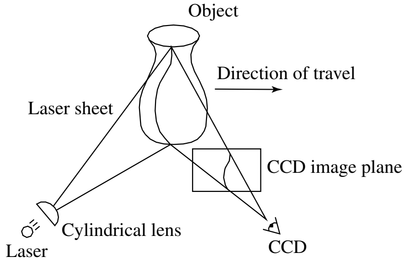
Time of flight

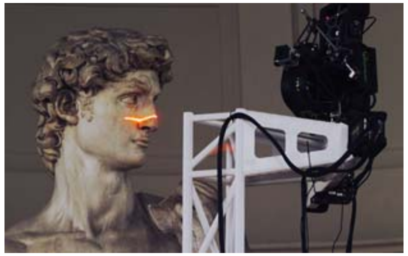
- Basic idea: send out pulse of light (usually laser), time how long it takes to return

$$r = \frac{1}{2} c \Delta t$$



Laser scanning (triangulation)





Digital Michelangelo Project
http://graphics.stanford.edu/projects/mich/

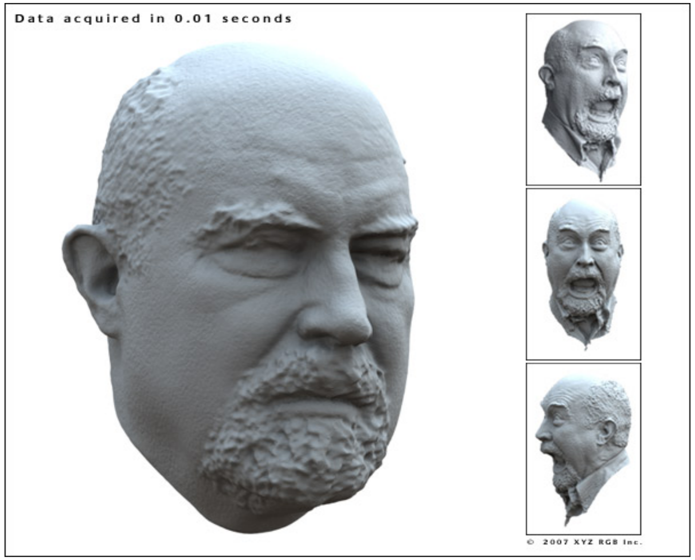
- Optical triangulation
  - Project a single stripe of laser light
  - Scan it across the surface of the object
  - This is a very precise version of structured light scanning
- Other patterns are possible

Cyberware

face and hand

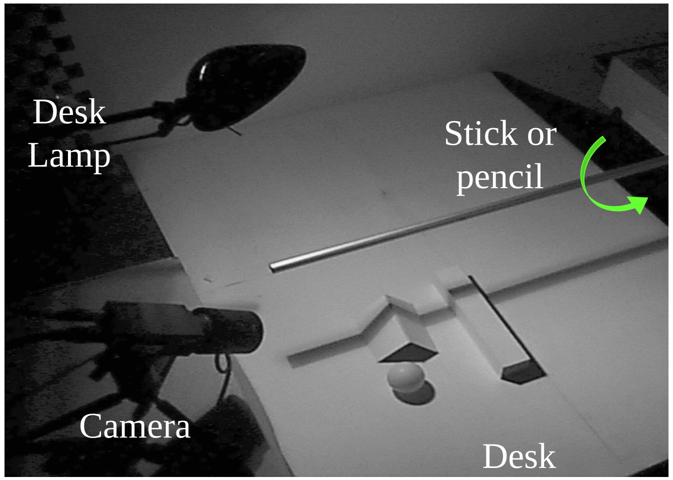
full body

XYZRGB



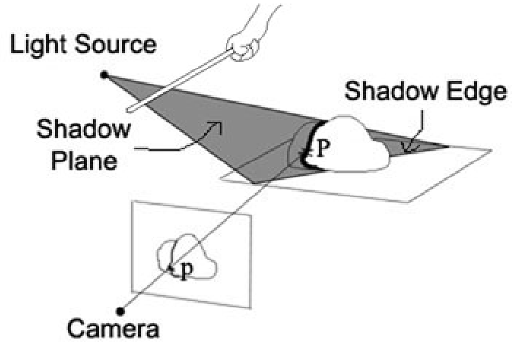
XYZRGB

Shadow scanning



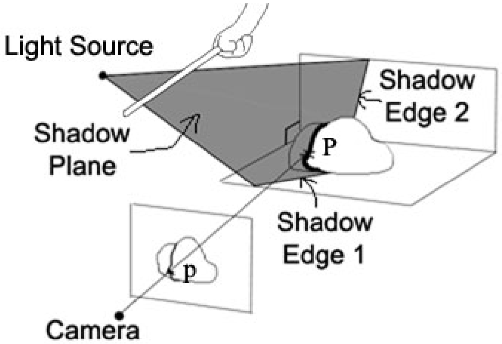
http://www.vision.caltech.edu/bouguetj/ICCV98/

Basic idea



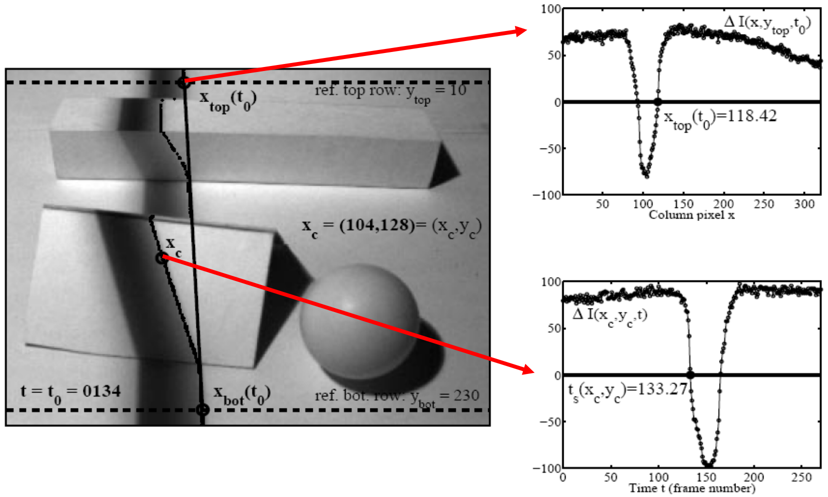
- Calibration issues:
  - where's the camera wrt. ground plane?
  - where's the shadow plane?
    - depends on light source position, shadow edge

Two Plane Version



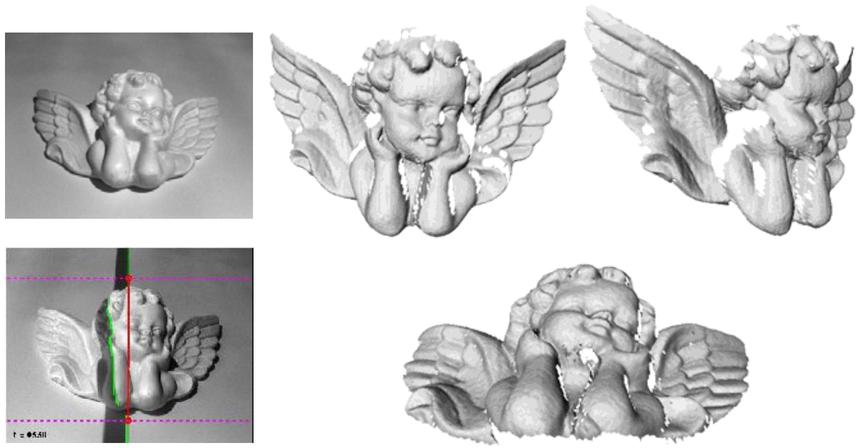
- Advantages
  - don't need to pre-calibrate the light source
  - shadow plane determined from two shadow edges

Estimating shadow lines



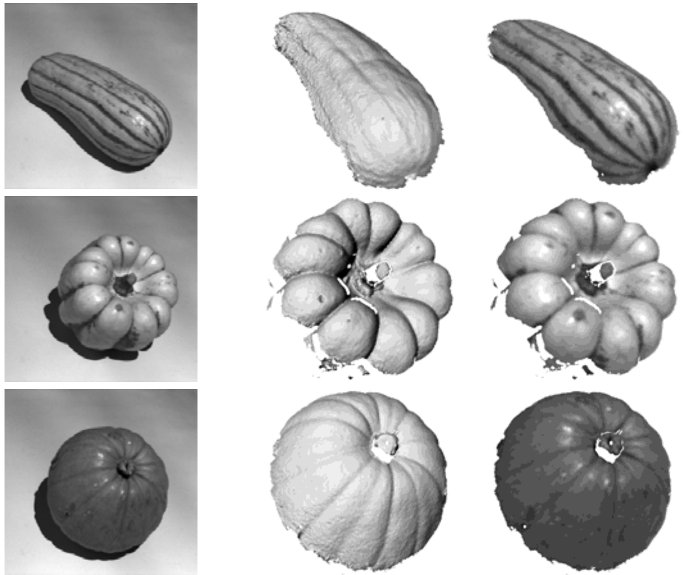
Shadow scanning in action

Results

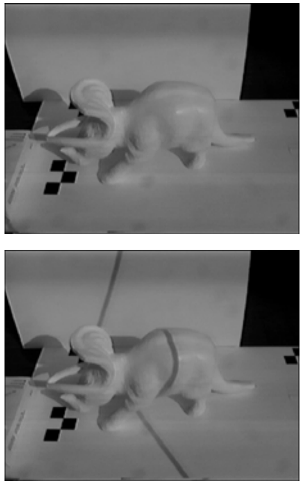


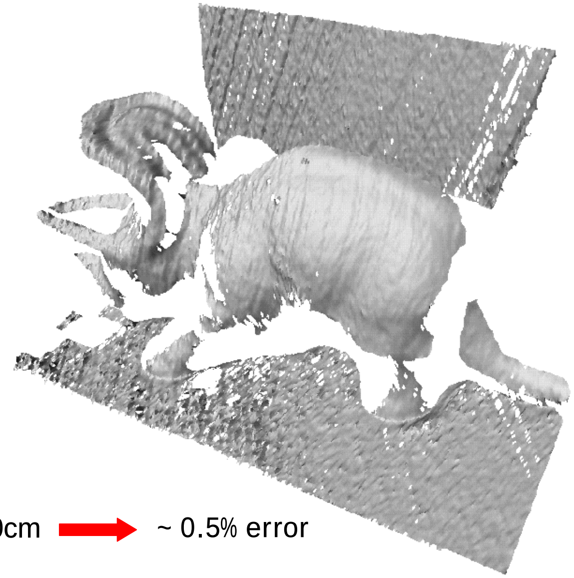
accuracy: 0.1mm over 10cm → ~ 0.1% error

Textured objects



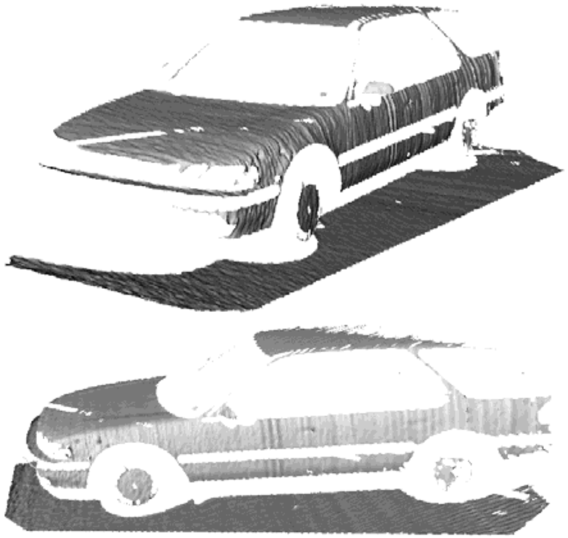
Scanning with the sun





accuracy: 1mm over 50cm → ~ 0.5% error

Scanning with the sun



accuracy: 1cm over 2m
→ ~ 0.5% error

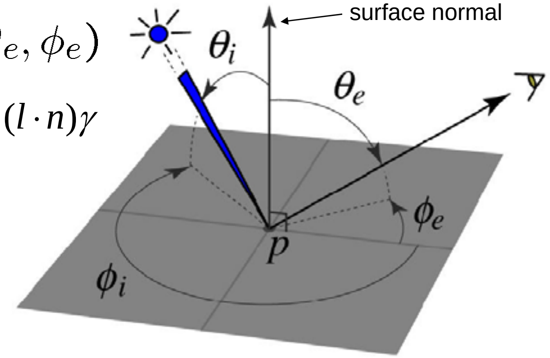
Active variants of passive approaches

The BRDF

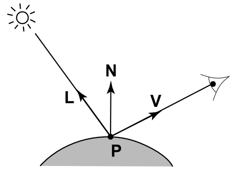
- The Bidirectional Reflection Distribution Function
  - Given an incoming ray (θ_i, ϕ_i) and outgoing ray (θ_e, ϕ_e) what proportion of the incoming light is reflected along outgoing ray?

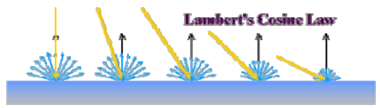
$$\rho(\theta_i, \phi_i, \theta_e, \phi_e)$$

$$I = \rho(l, v) \times (l \cdot n) \gamma$$



Diffuse reflection (Lambertian)



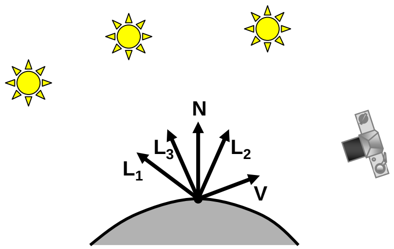


$$\rho(l, v) = k_d$$ ← **albedo**

$$I = k_d \mathbf{N} \cdot \mathbf{L}$$

Assuming that light strength is 1.

Photometric stereo



$$I_1 = k_d \mathbf{N} \cdot \mathbf{L_1}$$
$$I_2 = k_d \mathbf{N} \cdot \mathbf{L_2}$$
$$I_3 = k_d \mathbf{N} \cdot \mathbf{L_3}$$

- Can write this as a matrix equation:

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} = k_d \begin{bmatrix} \mathbf{L_1}^T \\ \mathbf{L_2}^T \\ \mathbf{L_3}^T \end{bmatrix} \mathbf{N}$$

Solving the equations

$$\underbrace{\begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix}}_{\substack{\mathbf{I} \\ 3 \times 1}} = \underbrace{\begin{bmatrix} \mathbf{L_1}^T \\ \mathbf{L_2}^T \\ \mathbf{L_3}^T \end{bmatrix}}_{\substack{\mathbf{L} \\ 3 \times 3}} \underbrace{k_d \mathbf{N}}_{\substack{\mathbf{G} \\ 3 \times 1}}$$

$$\mathbf{G} = \mathbf{L}^{-1} \mathbf{I}$$

$$k_d = \|\mathbf{G}\|$$

$$\mathbf{N} = \frac{1}{k_d} \mathbf{G}$$

More than three lights

- Get better results by using more lights

$$\begin{bmatrix} I_1 \\ \vdots \\ I_n \end{bmatrix} = \begin{bmatrix} \mathbf{L_1} \\ \vdots \\ \mathbf{L_n} \end{bmatrix} k_d \mathbf{N}$$

- Least squares solution:

$$\begin{aligned} \mathbf{I} &= \mathbf{LG} \\ \mathbf{L^T I} &= \mathbf{L^T L G} \\ \mathbf{G} &= (\mathbf{L^T L})^{-1} (\mathbf{L^T I}) \end{aligned}$$

- Solve for N, k_d as before

Trick for handling shadows

- Weight each equation by the pixel brightness:

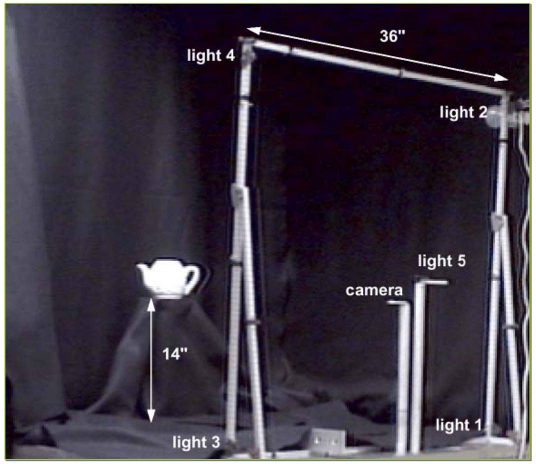
$$I_i(I_i) = I_i[k_d \mathbf{N} \cdot \mathbf{L_i}]$$

- Gives weighted least-squares matrix equation:

$$\begin{bmatrix} I_1^2 \\ \vdots \\ I_n^2 \end{bmatrix} = \begin{bmatrix} I_1 \mathbf{L_1}^T \\ \vdots \\ I_n \mathbf{L_n}^T \end{bmatrix} k_d \mathbf{N}$$

- Solve for N, k_d as before

Photometric Stereo Setup



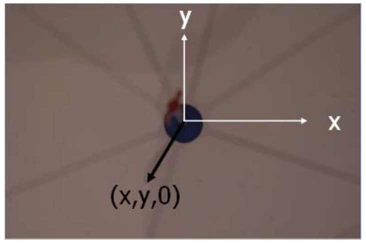
Procedure

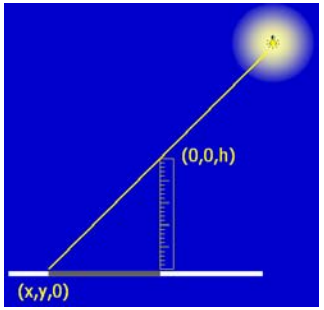
- Calibrate camera
- Calibrate light directions/intensities
- Photographing objects (HDR recommended)
- Estimate normals
- Estimate depth

Estimating light directions

- Trick: place a chrome sphere in the scene

  - the location of the highlight tells you where the light source is
- Use a ruler

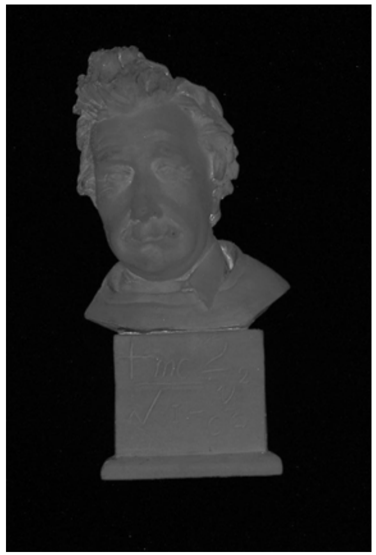




Photographing objects

Normalize light intensities

Estimate normals



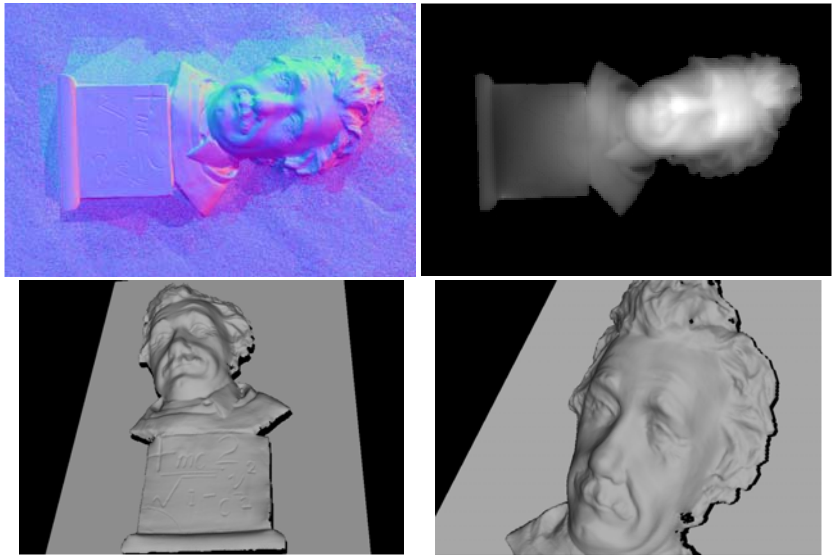
Depth from normals

$$(n_x, n_y, n_z) = \left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}, -1\right) = (p, q, -1)$$

$$\begin{aligned} E &= E_{data} + E_{smooth} + E_{cons} \\ &= \sum_{i,j} w_{data} * \left[\left(\frac{\partial z(i,j)}{\partial x} - p_{ij} \right)^2 + \left(\frac{\partial z(i,j)}{\partial y} - q_{ij} \right)^2 \right] \\ &+ \sum_{i,j} w_{smooth} * \left[\left(\frac{\partial^2 z(i,j)}{\partial x^2} \right)^2 + 2\left(\frac{\partial^2 z(i,j)}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 z(i,j)}{\partial y^2} \right)^2 \right] \\ &+ \sum_{(i,j)\in Cons} w_{cons} * \left(z(i,j) - c_{ij} \right)^2 \end{aligned}$$

$$E = \frac{1}{2} z^T A z - b^T z + c \quad \equiv \quad Az = b$$

Results

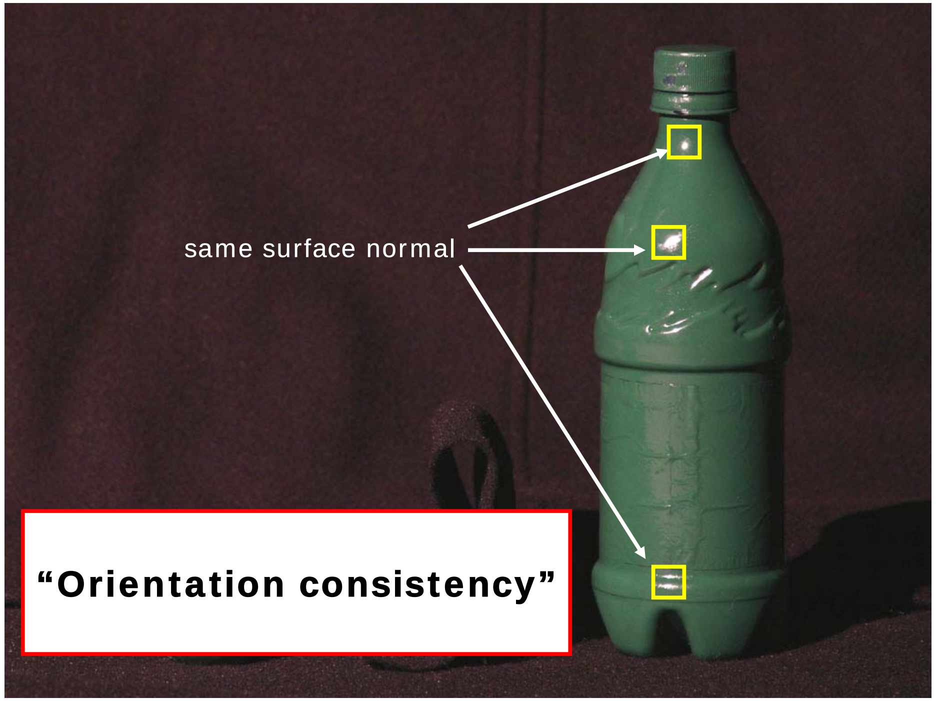


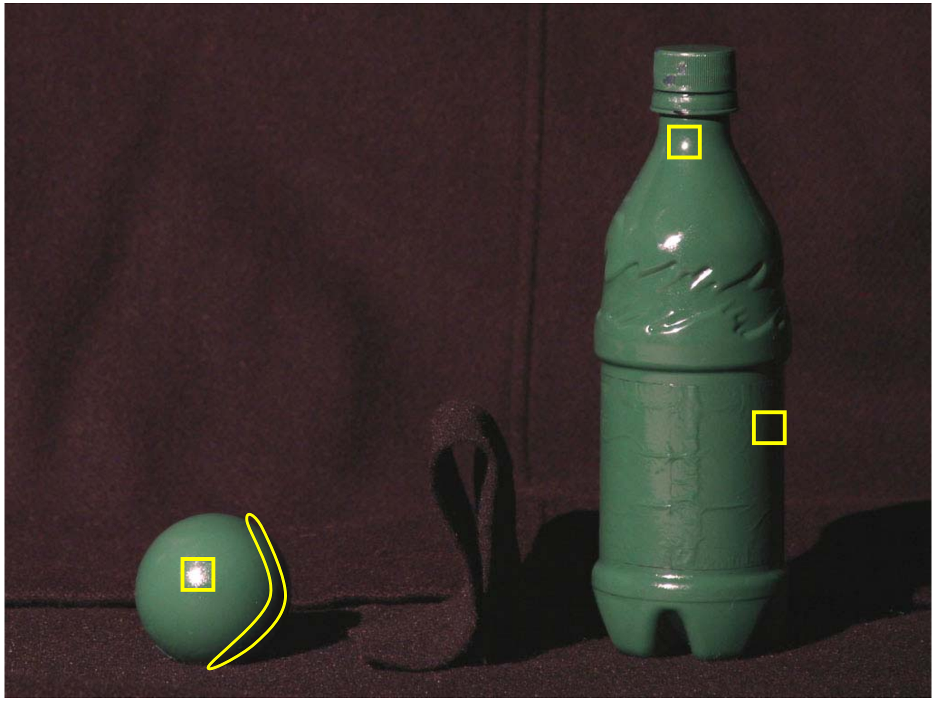
Limitations

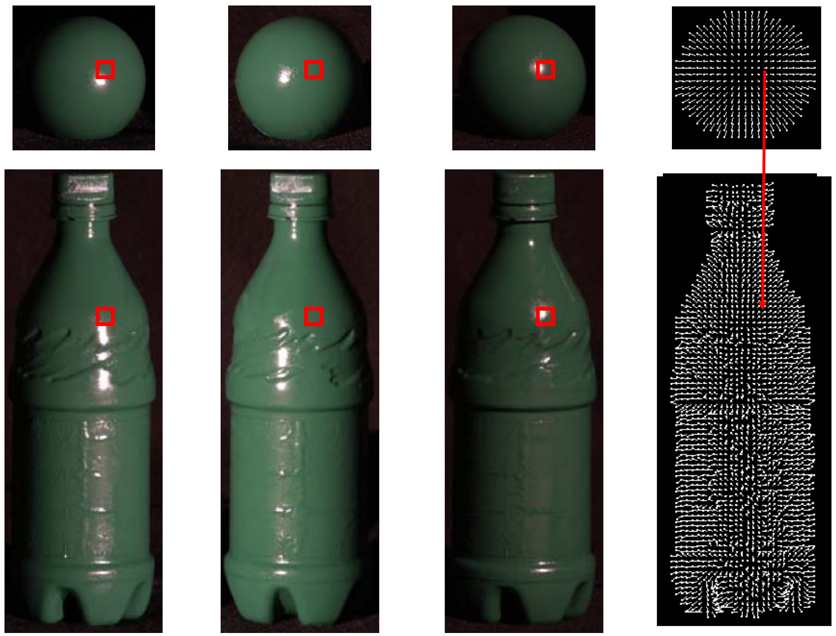
- Big problems
  - doesn't work for shiny things, semi-translucent things
  - shadows, inter-reflections
- Smaller problems
  - calibration requirements
    - measure light source directions, intensities
    - camera response function

Example-based photometric stereo

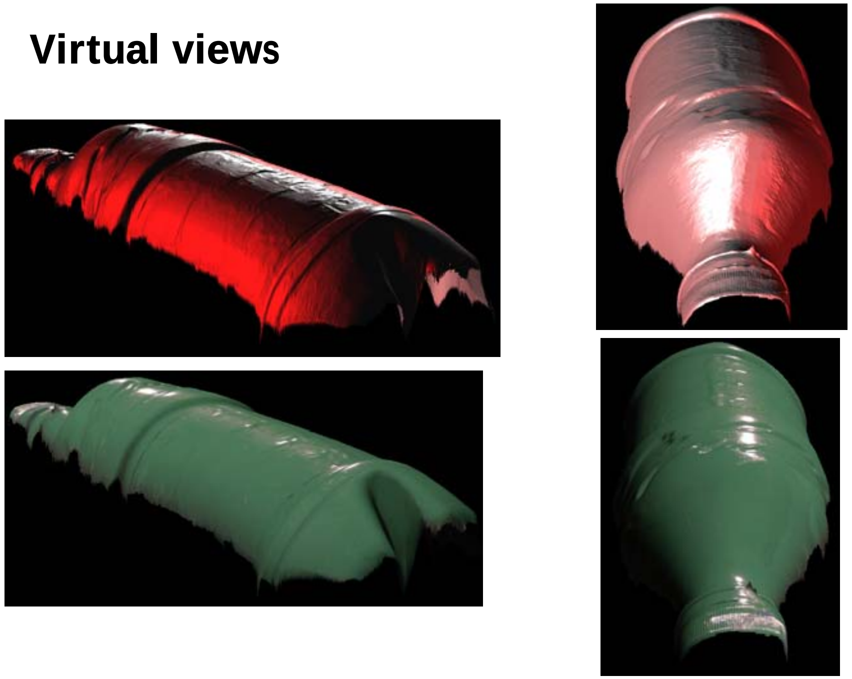
- Estimate 3D shape by varying illumination, fixed camera
- Operating conditions
  - any opaque material
  - distant camera, lighting
  - reference object available
  - no shadows, interreflections, transparency


same surface normal
“Orientation consistency”

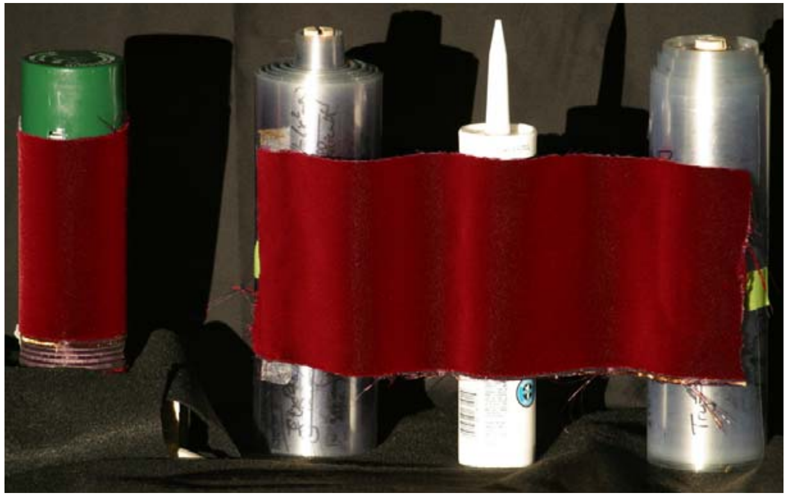




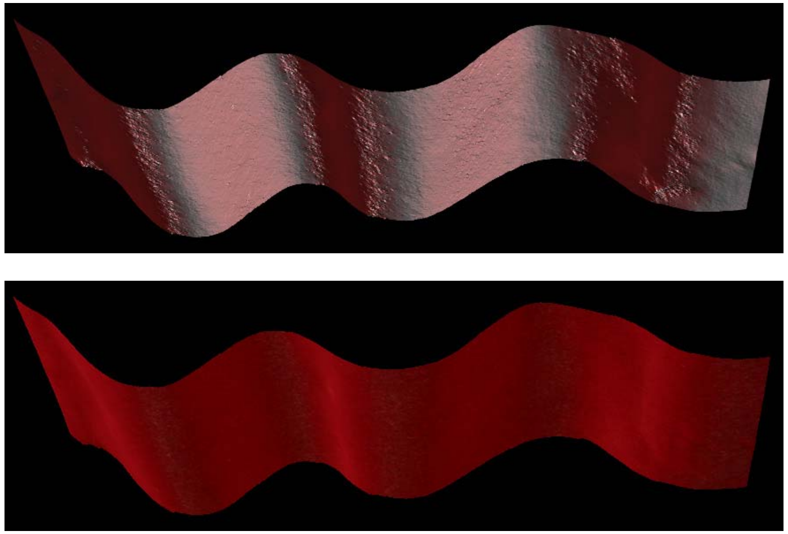
Virtual views



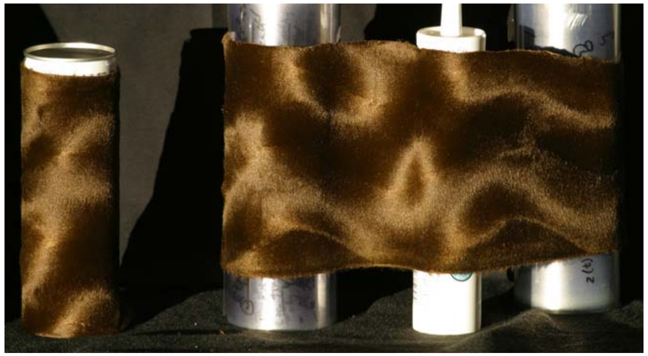
Velvet



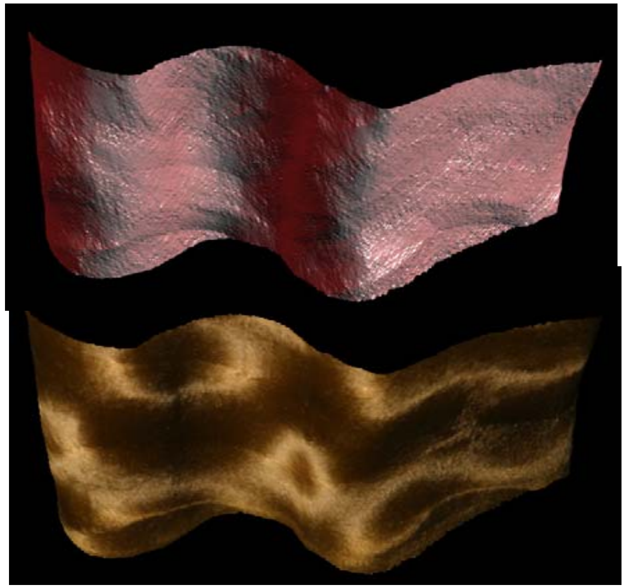
Virtual Views

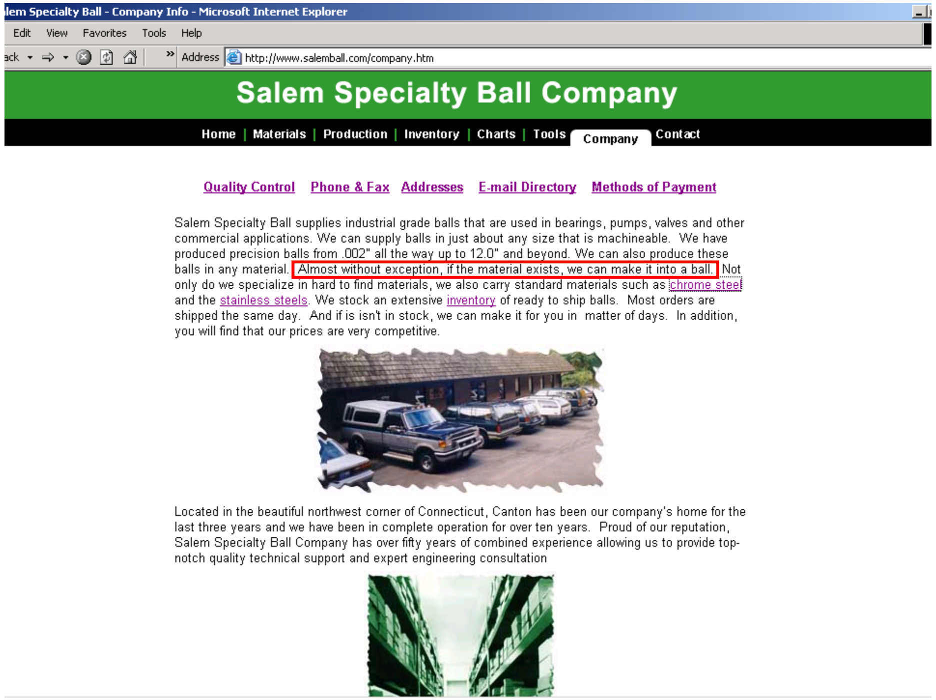


Brushed Fur



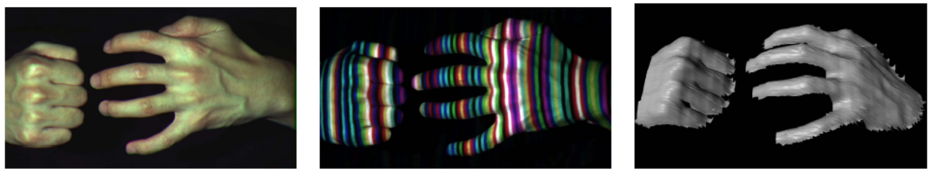
Virtual Views



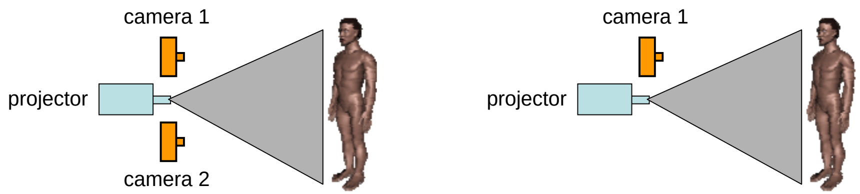


Active stereo with structured light

DigiVFX



Li Zhang's one-shot stereo



- Project "structured" light patterns onto the object
  - simplifies the correspondence problem

Spacetime Stereo

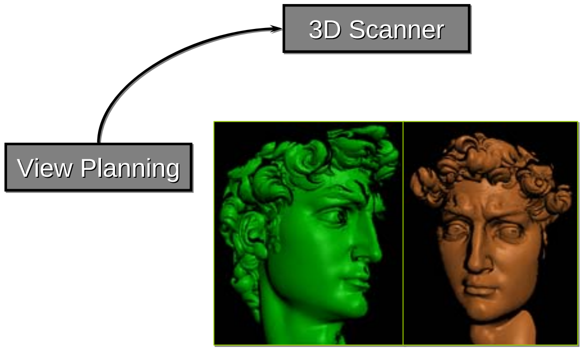
DigiVFX

http://grail.cs.washington.edu/projects/stfaces/

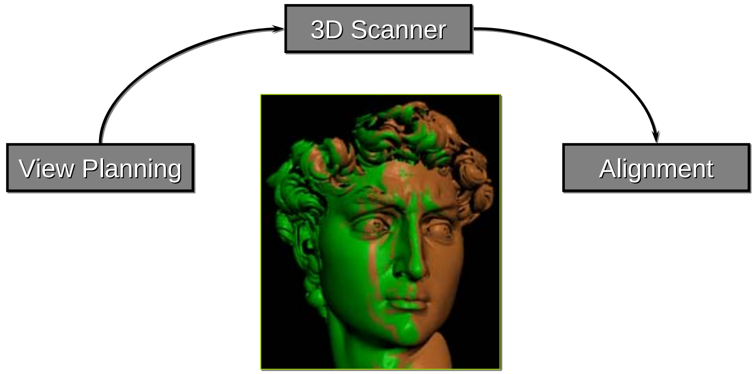
3D Model Acquisition Pipeline

DigiVFX

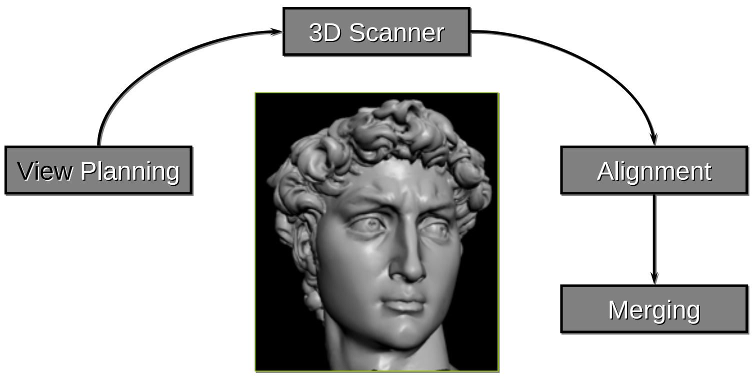
3D Model Acquisition Pipeline



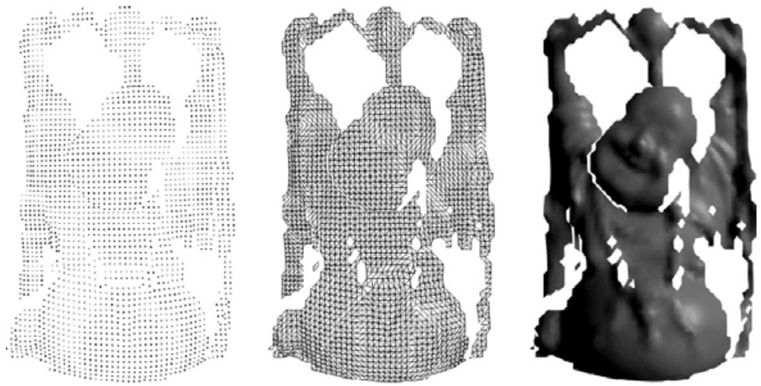
3D Model Acquisition Pipeline



3D Model Acquisition Pipeline

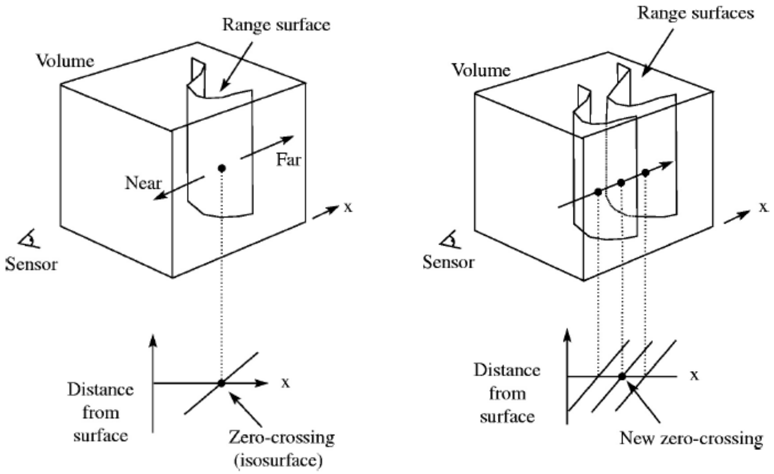


Volumetric reconstruction

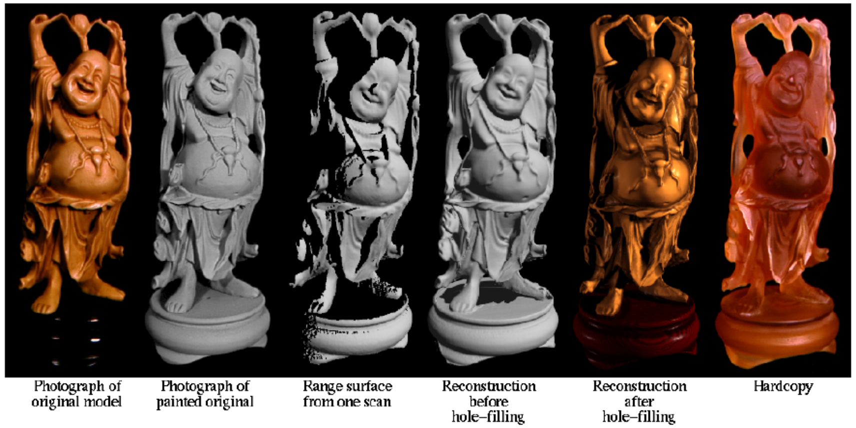


Range image | Tesellation | Range surface

Signed distance function



Results

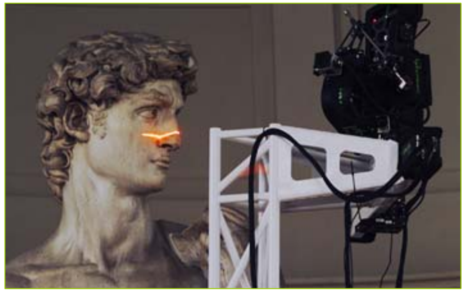


The Digital Michelangelo Project

- Goal: scan 10 sculptures by Michelangelo
- High-resolution (“quarter-millimeter”) geometry
- Stanford University, led by Marc Levoy

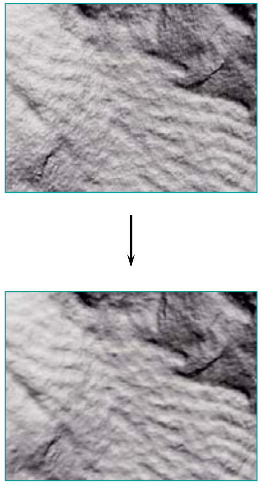
Systems, projects and applications

Scanning the David



height of gantry: 7.5 meters
weight of gantry: 800 kilograms

Range processing pipeline



- steps
  1. manual initial alignment
  2. ICP to one existing scan
  3. automatic ICP of all overlapping pairs
  4. global relaxation to spread out error
  5. merging using volumetric method

Statistics about the scan

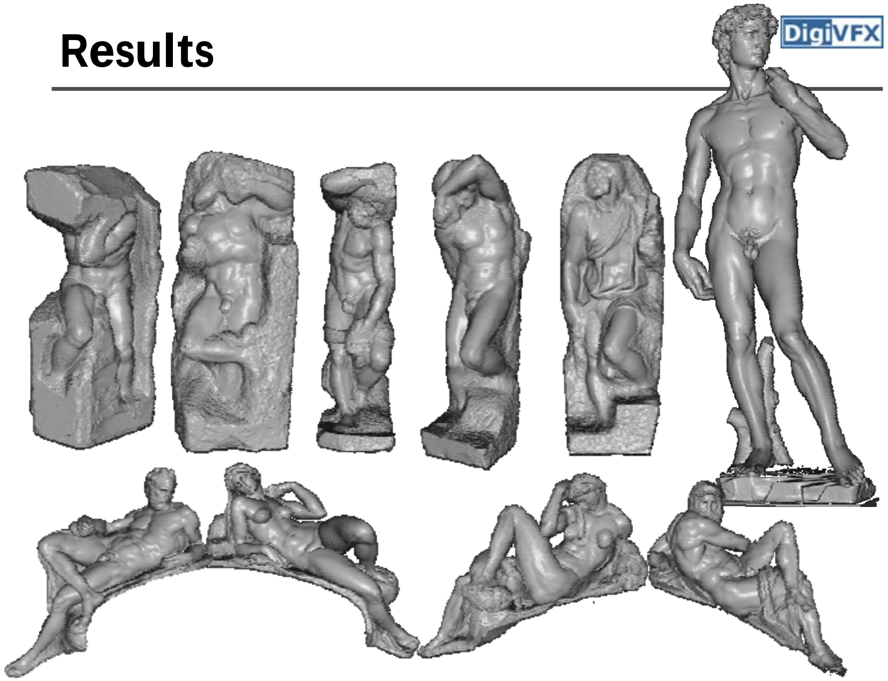
- 480 individually aimed scans
- 2 billion polygons
- 7,000 color images
- 32 gigabytes
- 30 nights of scanning
- 22 people

Comparison

photograph

1.0 mm computer model

Results



The Great Buddha Project

- Great Buddha of Kamakura
- Original made of wood, completed 1243
- Covered in bronze and gold leaf, 1267
- Approx. 15 m tall
- Goal: preservation of cultural heritage
- Institute of Industrial Science, University of Tokyo, led by Katsushi Ikeuchi

Scanner

- Cyrax range scanner by Cyra Technologies
- Laser pulse time-of-flight
- Accuracy: 4 mm
- Range: 100 m

Processing

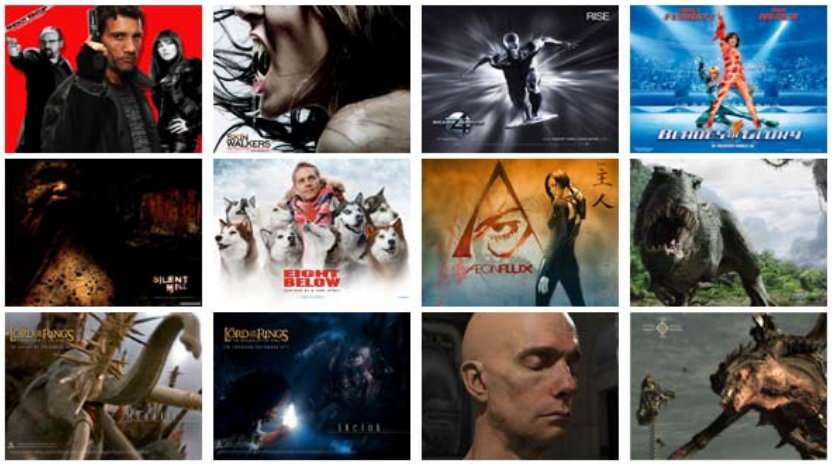
- 20 range images (a few million points)
- Simultaneous all-to-all ICP
- Variant of volumetric merging (parallelized)

Results

Applications in VFX

- 3D scanning
- Hybrid camera for IMAX
- View interpolation

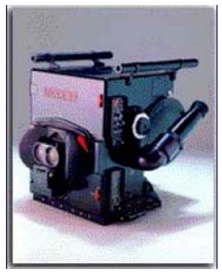
3D scanning



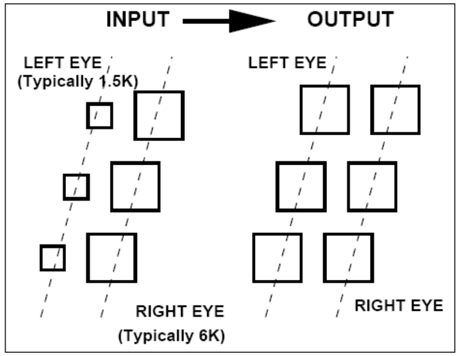
XYZRGB Inc.

IMAX 3D

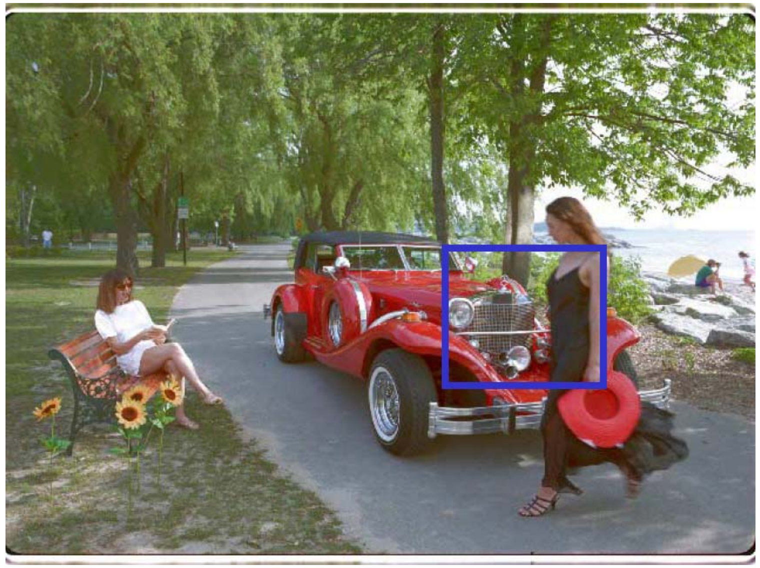
- 6K resolution, 42 linear bits per pixel
- For CG, it typically takes 6 hours for a frame
- 45-minute IMAX 3D CG film requires a 100-CPU rendering farm full-time for about a year just for rendering
- For live-action, camera is bulky (like a refrigerator)



Hybrid stereo camera



Live-action sequence



Hybrid input

left

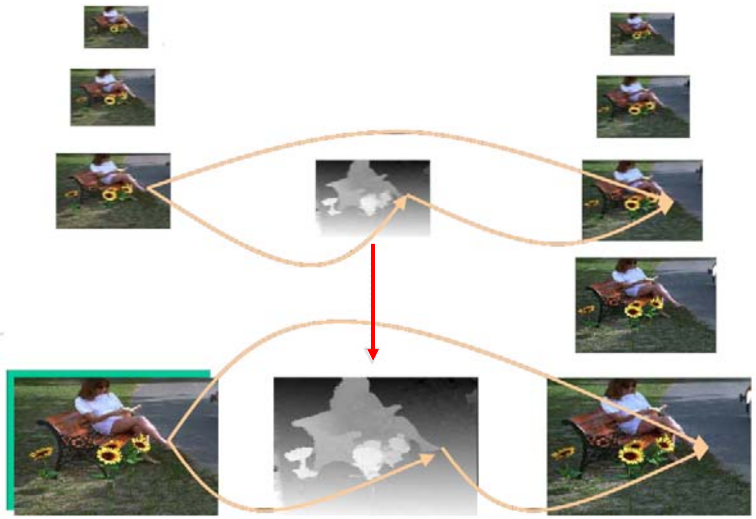
right

Hybrid input

left

right

Combine multiple hires to lores



Results

View interpolation

Bullet time video

View interpolation

High-quality video view interpolation

Final project

Final project

- Assigned: 5/14
- Due: 6/25 Wednesday
- Proposal and midterm report on 5/27

Final project

- Research (1-2 people)
- System (1-3 people)
- Film (3-4 people)

Research

- Define a problem and try to solve it
- You don't need to solve it all, but have to make a reasonable progress, for example, solve a simplified version.
- Find inspirations from SIGGRAPH/CVPR/ICCV papers

System

- Implement existing algorithm into a useful system such as implementing SIGGRAPH 2006/2007/2008 or CVPR/ICCV 2006/2007/2008 papers

Film

- It must be an "effect" film.
- You can use any tools as you want. But, I assume that you have to write some on your own.
- Find inspirations from

  Gatech's vfx course
  http://www.cc.gatech.edu/classes/AY2004/cs4480_spring/
  independent film makers
  http://www.peerlessproductions.com/
  ADs/films/YouTube
- Submit two videos, final and making-of.