Stereo Matching

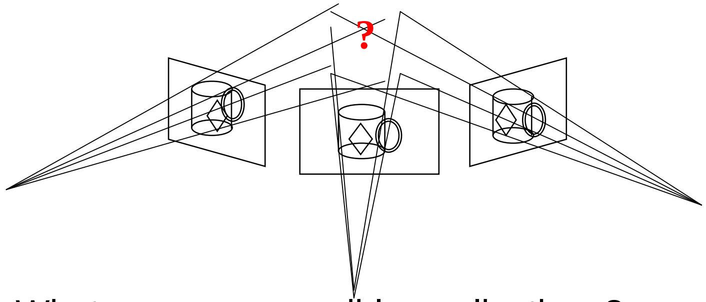
Computer Vision

CSE576, Spring 2005

Richard Szeliski

Stereo Matching

Given two or more images of the same scene or object, compute a representation of its shape



What are some possible applications?

Face modeling

From one stereo pair to a 3D head model

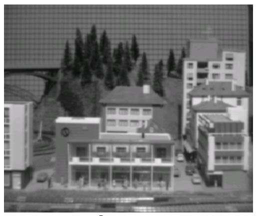
[Frederic Deverney, INRIA]

Z-keying: mix live and synthetic

Takeo Kanade, CMU (Stereo Machine)

View Interpolation

Given two images with correspondences, *morph* (warp and cross-dissolve) between them [Chen & Williams, SIGGRAPH'93]


input

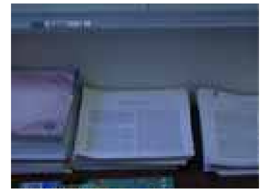

depth image


novel view

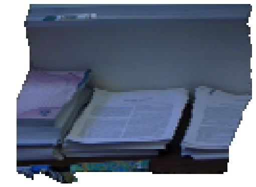
[Matthies,Szeliski,Kanade'88]

More view interpolation

Spline-based depth map


input

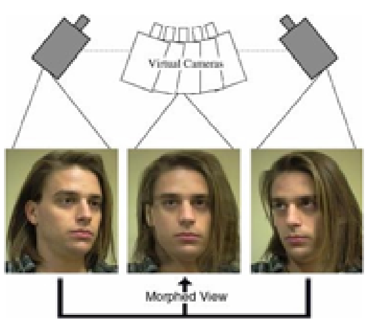

depth image


novel view

[Szeliski & Kang '95]

View Morphing

Morph between pair of images using epipolar geometry [Seitz & Dyer, SIGGRAPH'96]

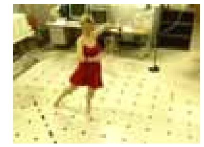
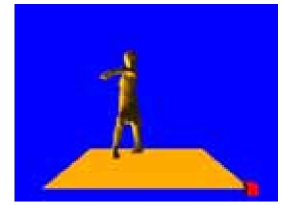


Video view interpolation

Virtualized Reality^{TM}

[Takeo Kanade *et al.*, CMU]

- collect video from 50+ stream
- reconstruct 3D model sequences

- steerable version used for SuperBowl XXV "eye vision"

http://www.cs.cmu.edu/afs/cs/project/VirtualizedR/www/VirtualizedR.html

Real-time stereo

Nomad robot searches for meteorites in Antartica
http://www.frc.ri.cmu.edu/projects/meteorobot/index.html

Used for robot navigation (and other tasks)

- Software-based real-time stereo techniques

Additional applications

- Real-time people tracking (systems from Pt. Gray Research and SRI)
- "Gaze" correction for video conferencing [Ott,Lewis,Cox InterChi'93]
- Other ideas?

Stereo Matching

Given two or more images of the same scene or object, compute a representation of its shape

What are some possible representations?

- depth maps
- volumetric models
- 3D surface models
- planar (or offset) layers

Stereo Matching

What are some possible algorithms?

- match “features” and interpolate
- match edges and interpolate
- match all pixels with windows (coarse-fine)
- use optimization:
  - iterative updating
  - dynamic programming
  - energy minimization (regularization, stochastic)
  - graph algorithms

Outline (remainder of lecture)

Image rectification

Matching criteria

Local algorithms (aggregation)

- iterative updating

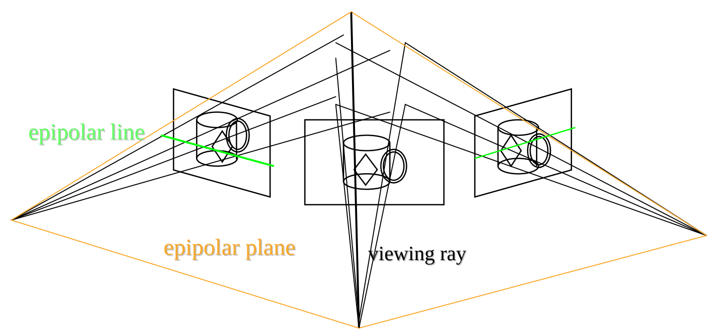
Optimization algorithms:

- energy (cost) formulation & Markov Random Fields
- mean-field, stochastic, and graph algorithms

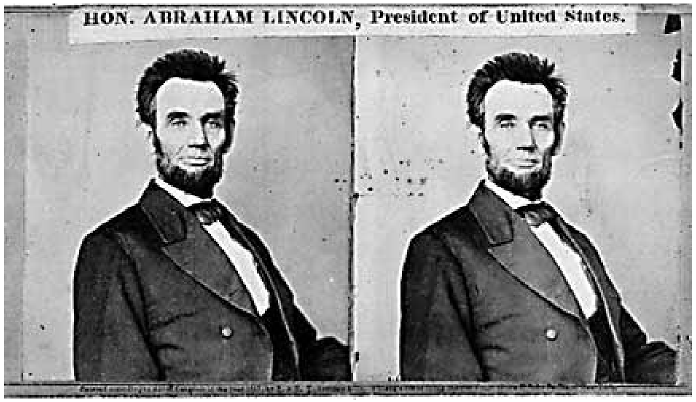
Multi-View stereo & occlusions

Stereo: epipolar geometry

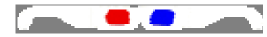
Match features along epipolar lines



Stereo image pair



Anaglyphs



http://www.rainbowsymphony.com/freestuff.html

(Wikipedia for images)

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

Stereo: epipolar geometry

for *two* images (or images with collinear camera centers), can find epipolar lines

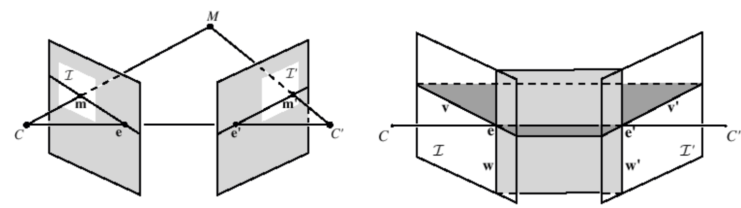
epipolar lines are the projection of the *pencil* of planes passing through the centers

Rectification: warping the input images (perspective transformation) so that epipolar lines are horizontal

Rectification

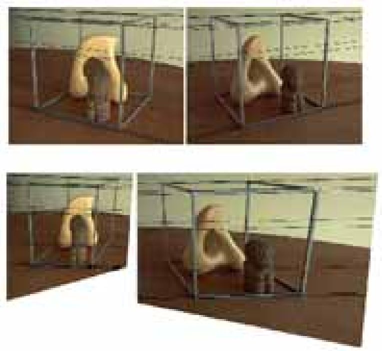
Project each image onto same plane, which is parallel to the epipole

Resample lines (and shear/stretch) to place lines in correspondence, and minimize distortion



[Loop and Zhang, CVPR'99]

Rectification

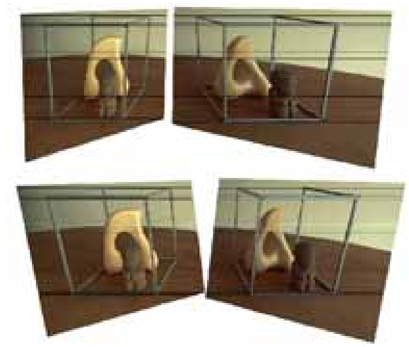


(a) Original image pair overlayed with several epipolar lines.

(b) Image pair transformed by the specialized projective mapping $\mathbf{H}_p$ and $\mathbf{H}'_p$. Note that the epipolar lines are now parallel to each other in each image.

BAD!

Rectification



(c) Image pair transformed by the similarity $\mathbf{H}_r$ and $\mathbf{H}'_r$. Note that the image pair is now rectified (the epipolar lines are horizontally aligned).

(d) Final image rectification after shearing transform $\mathbf{H}_s$ and $\mathbf{H}'_s$. Note that the image pair remains rectified, but the horizontal distortion is reduced.

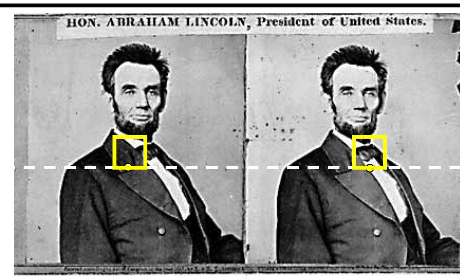
GOOD!

Finding correspondences

apply feature matching criterion (e.g., correlation or Lucas-Kanade) at *all* pixels simultaneously

search only over epipolar lines (many fewer candidate positions)

Your 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***

- This should look familar...

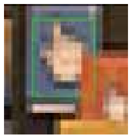
Image registration (revisited)

How do we determine correspondences?

- *block matching* or *SSD* (sum squared differences)

$$E(x,y;d) = \sum_{(x',y')\in N(x,y)} [I_L(x'+d,y') - I_R(x',y')]^2$$

d is the *disparity* (horizontal motion)



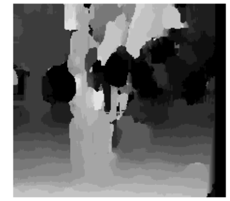

How big should the neighborhood be?

Neighborhood size

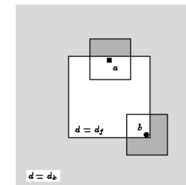
Smaller neighborhood: more details

Larger neighborhood: fewer isolated mistakes

w = 3



w = 20



Matching criteria

Raw pixel values (correlation)

Band-pass filtered images [Jones & Malik 92]

"Corner" like features [Zhang, …]

Edges [many people…]

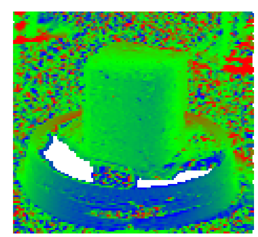
Gradients [Seitz 89; Scharstein 94]

Rank statistics [Zabih & Woodfill 94]

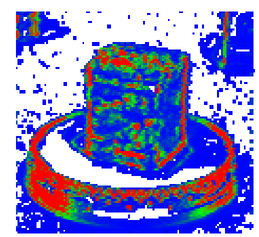
Stereo: certainty modeling

Compute certainty map from correlations

input



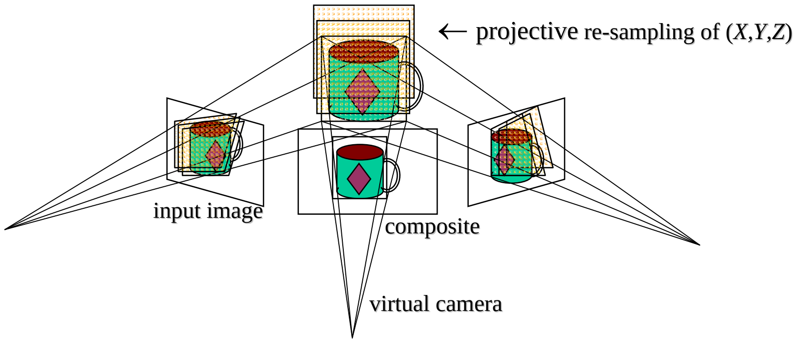
depth map



certainty map

Plane Sweep Stereo

Sweep family of planes through volume

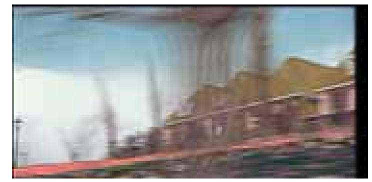


- each plane defines an image ⇒ composite homography

Plane Sweep Stereo

For each depth plane

- compute composite (mosaic) image — *mean*

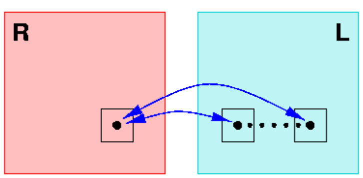


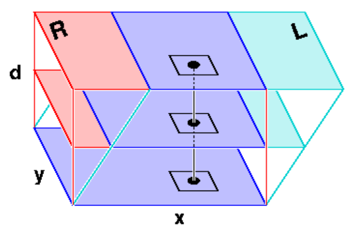
- compute error image — *variance*
- convert to confidence and aggregate spatially

Select winning depth at each pixel

Plane sweep stereo

Re-order (pixel / disparity) evaluation loops





for every pixel,
  for every disparity
    compute cost

for every disparity
  for every pixel
    compute cost

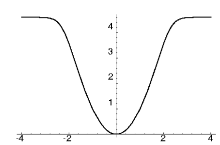
Stereo matching framework

1. For every disparity, compute *raw* matching costs

$$E_0(x, y; d) = \rho(I_L(x' + d, y') - I_R(x', y'))$$

Why use a robust function?

- occlusions, other outliers

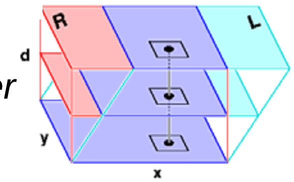


Can also use alternative match criteria

Stereo matching framework

2. Aggregate costs spatially

$$E(x,y;d) = \sum_{(x',y') \in N(x,y)} E_0(x',y',d)$$



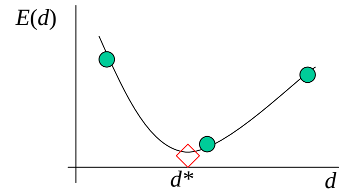
- Here, we are using a *box filter* (efficient moving average implementation)
- Can also use weighted average, [non-linear] diffusion…

Stereo matching framework

3. Choose winning disparity at each pixel

$$d(x,y) = \arg\min_d E(x,y;d)$$

4. Interpolate to *sub-pixel* accuracy



Traditional Stereo Matching

Advantages:

- gives detailed surface estimates
- fast algorithms based on moving averages
- sub-pixel disparity estimates and confidence

Limitations:

- narrow baseline $\Rightarrow$ noisy estimates
- fails in textureless areas
- gets confused near occlusion boundaries

Stereo with Non-Linear Diffusion

Problem with traditional approach:

- gets confused near discontinuities

New approach:

- use iterative (non-linear) aggregation to obtain better estimate
- provably equivalent to mean-field estimate of Markov Random Field

Linear diffusion

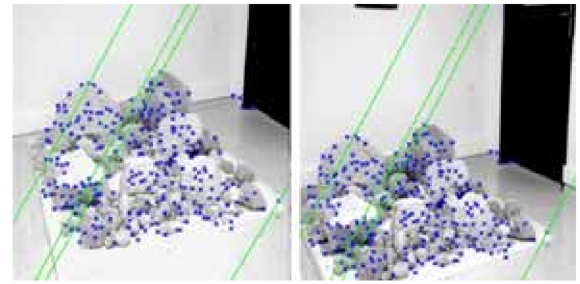
Average energy with neighbors + starting value

$$E(x,y,d) \leftarrow (1-4\lambda)E(x,y,d) + \lambda \sum_{(k,l)\in\mathcal{N}_4} E(x+k,y+l,d)$$
$$+\beta(E_0(x,y,d) - E(x,y,d))$$

window diffusion

Feature-based stereo

Match “corner” (interest) points



Interpolate complete solution

Data interpolation

Given a sparse set of 3D points, how do we *interpolate* to a full 3D surface?

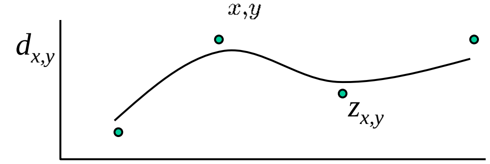
Scattered data interpolation [Nielson93]

- triangulate
- put onto a grid and fill (use pyramid?)
- place a *kernel function* over each data point
- minimize an energy function

Energy minimization

1-D example: approximating splines

$$\begin{aligned}
E_{\text{total}}(\mathbf{d}) &= E_{\text{data}}(\mathbf{d}) + \lambda E_{\text{smoothness}}(\mathbf{d}) \\
E_{\text{data}}(\mathbf{d}) &= \sum_{x,y}(d_{x,y} - z_{x,y})^2 \\
E_{\text{membrane}}(\mathbf{d}) &= \sum_{x,y}(d_{x,y} - d_{x-1,y})^2 \\
E_{\text{thin plate}}(\mathbf{d}) &= \sum_{x,y}(2d_{x,y} - d_{x-1,y} - d_{x+1,y})^2
\end{aligned}$$



Relaxation

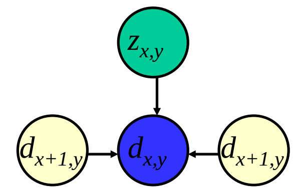
How can we get the best solution?

Differentiate energy function, set to 0

$$\begin{aligned}\frac{\partial E}{\partial d_{x,y}} &= 2(d_{x,y} - z_{x,y}) + \\ &\quad 2\lambda(2d_{x,y} - d_{x-1,y} - d_{x+1,y}) = 0 \\ d_{x,y} &\leftarrow \frac{1}{1+2\lambda}(z_{x,y} + d_{x-1,y} + d_{x+1,y})\end{aligned}$$

Relaxation

Iteratively improve a solution by locally minimizing the energy: *relax* to solution

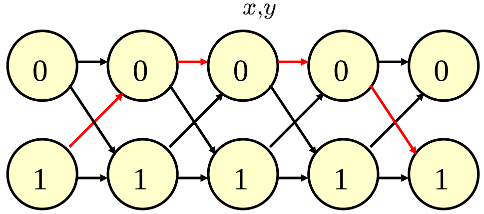


Earliest application: WWII numerical simulations

Dynamic programming

Evaluate best cumulative cost at each pixel

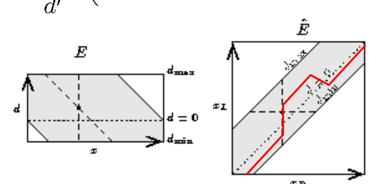
$$\begin{aligned}E_{\text{total}}(\mathbf{d}) &= E_{\text{data}}(\mathbf{d}) + \lambda E_{\text{smoothness}}(\mathbf{d}) \\ E_{\text{data}}(\mathbf{d}) &= \sum_{x,y}(d_{x,y} - z_{x,y})^2 \\ E_{\text{smoothness}}(\mathbf{d}) &= \sum_{x,y}|d_{x,y} - d_{x-1,y}|\end{aligned}$$



Dynamic programming

1-D cost function

$$\begin{aligned}E(\mathbf{d}) &= \sum_{x,y}\rho_P(d_{x+1,y} - d_{x,y}) + \sum_{x,y} E_0(x,y;d) \\ \tilde{E}(x,y,d) &= E_0(x,y;d) + \\ &\quad \min_{d'}\left(\tilde{E}(x-1,y,d') + \rho_P(d_{x,y} - d'_{x-1,y})\right)\end{aligned}$$



Dynamic programming

Disparity space image and min. cost path

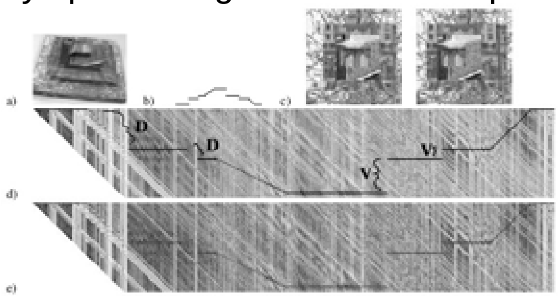


Fig. 4. This figure shows (a) a model of the stereo sloping wedding cake that we will use as a test example, (b) a depth profile through the center of the sloping wedding cake, (c) a simulated, noise-free image pair of the cake, (d) the enhanced, cropped, correlation DSI= representation for the image pair in (c), and (e) the enhanced, cropped, correlation DSI for a noisy sloping wedding cake (SNR = 18 dB). In (d), the regions labeled "D" mark diagonal gaps in the matching path caused by regions occluded in the left image. The regions labeled "V" mark vertical jumps in the path caused by regions occluded in the right image.

Dynamic programming

Sample result
(note horizontal streaks)

[Intille & Bobick]

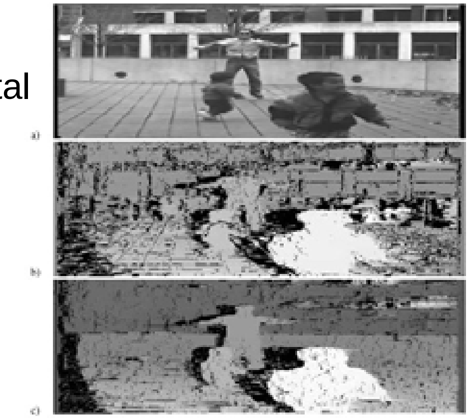
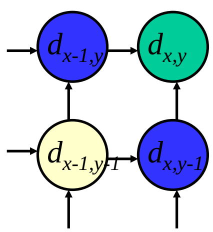


Fig. 12. Results of two stereo algorithms on Figure 1. (a) Original left image. (b) Cox *et al.* algorithm[14], and (c) the algorithm described in this paper.

Dynamic programming

Can we apply this trick in 2D as well?

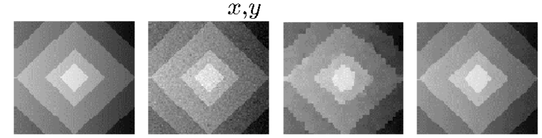


No: $d_{x,y-1}$ and $d_{x-1,y}$ may depend on different values of $d_{x-1,y-1}$

Graph cuts

Solution technique for general 2D problem

$$
\begin{aligned}
E_{\text{total}}(\mathbf{d}) &= E_{\text{data}}(\mathbf{d}) + \lambda E_{\text{smoothness}}(\mathbf{d}) \\
E_{\text{data}}(\mathbf{d}) &= \sum_{x,y} f_{x,y}(d_{x,y}) \\
E_{\text{smoothness}}(\mathbf{d}) &= \sum_{x,y} \rho(d_{x,y} - d_{x-1,y}) \\
&+ \sum_{x,y} \rho(d_{x,y} - d_{x,y-1})
\end{aligned}
$$



(a) original image (b) observed image (c) local min w.r.t. standard moves (d) local min w.r.t. α-expansion moves

Graph cuts

α-β swap

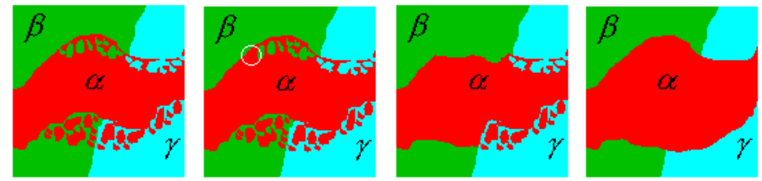
α expansion

modify smoothness penalty based on edges

compute best possible match within integer disparity

Graph cuts

Two different kinds of moves:



(a) initial labeling (b) standard move (c) α-β-swap (d) α-expansion

Bayesian inference

Formulate as statistical inference problem

Prior model $p_P(\boldsymbol{d})$

Measurement model $p_M(I_L, I_R | \boldsymbol{d})$

Posterior model

$$p_M(\boldsymbol{d} \mid I_L, I_R) \propto p_P(\boldsymbol{d})\ p_M(I_L, I_R | \boldsymbol{d})$$

Maximum a Posteriori (MAP estimate):

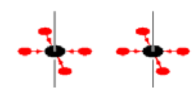
maximize $p_M(\boldsymbol{d} \mid I_L, I_R)$

Markov Random Field

Probability distribution on disparity field $d(x,y)$

$$p_P(d_{x,y}|\mathbf{d}) = p_P(d_{x,y}|\{d_{x',y'}, (x',y') \in \mathcal{N}(x,y)\})$$

$$p_P(\mathbf{d}) = \frac{1}{Z_P} e^{-E_P(\mathbf{d})}$$



$$E_P(\mathbf{d}) = \sum_{x,y} \rho_P(d_{x+1,y} - d_{x,y}) + \rho_P(d_{x,y+1} - d_{x,y})$$

Enforces *smoothness* or *coherence* on field

Measurement model

Likelihood of intensity correspondence

$$p_M(I_L, I_R|\mathbf{d}) = \frac{1}{Z_M} e^{-E_0(x,y;d)}$$

$$E_0(x,y;d) = \rho(I_L(x'+d, y') - I_R(x', y'))$$

Corresponds to Gaussian noise for quadratic ρ

MAP estimate

Maximize posterior likelihood

$$\begin{aligned} E(\mathbf{d}) &= -\log p(\mathbf{d}|I_L, I_R) \\ &= \sum_{x,y} \rho_P(d_{x+1,y} - d_{x,y}) + \rho_P(d_{x,y+1} - d_{x,y}) \\ &\quad + \sum_{x,y} \rho_M(I_L(x + d_{x,y}, y) - I_R(x,y)) \end{aligned}$$

Equivalent to *regularization* (energy minimization with smoothness constraints)

Why Bayesian estimation?

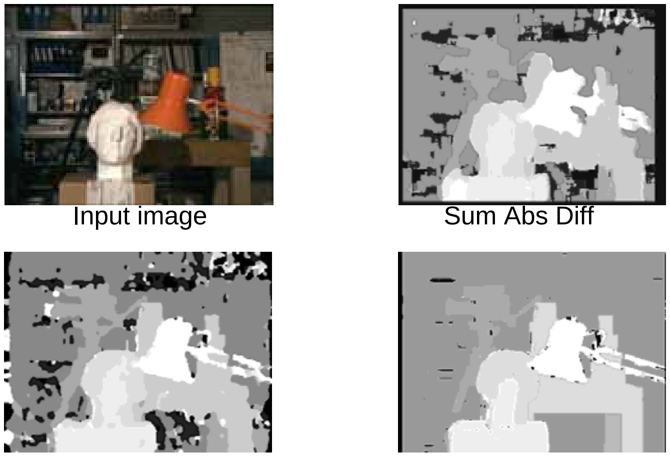
Principled way of determining cost function

Explicit model of noise and prior knowledge

Admits a wider variety of optimization algorithms:

- gradient descent (local minimization)
- stochastic optimization (Gibbs Sampler)
- mean-field optimization
- graph theoretic (actually deterministic) [Zabih]
- [loopy] belief propagation
- large stochastic flips [Swendsen-Wang]

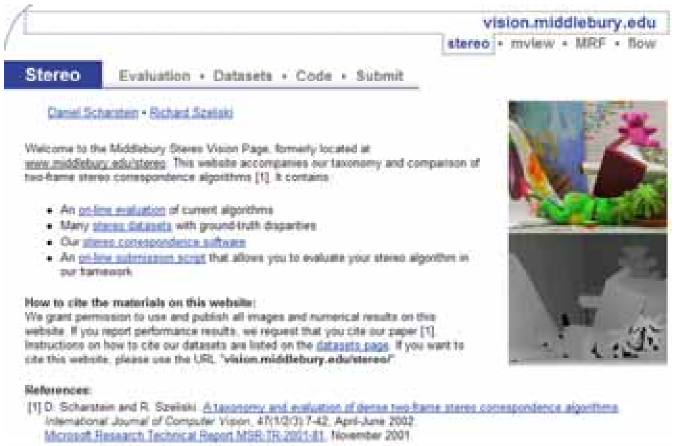
Depth Map Results



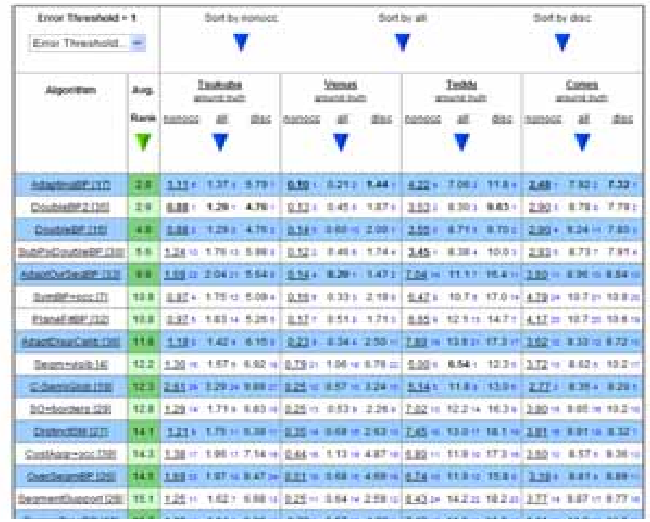
Input image Sum Abs Diff

Mean field Graph cuts

Stereo evaluation



Stereo—best algorithms



Error Threshold = 1 | Sort by nonocc | Sort by all | Sort by disc

Algorithm	Avg. Rank	Tsukuba nonocc	Tsukuba all	Tsukuba disc	Venus nonocc	Venus all	Venus disc	Teddy nonocc	Teddy all	Teddy disc	Cones nonocc	Cones all	Cones disc
AdaptingBP [17]	2.8	1.11	1.37	5.79	0.10	0.21	1.44	4.22	7.06	11.8	2.48	7.92	7.32
DoubleBP2 [35]	2.9	0.88	1.29	4.76	0.13	0.45	1.87	3.53	8.30	9.63	2.90	8.78	7.79
DoubleBP [15]	4.8	0.88	1.29	4.76	0.14	0.60	2.00	3.55	8.71	9.70	2.90	9.24	7.80
SubPixDoubleBP [30]	5.6	1.24	1.76	5.98	0.12	0.46	1.74	3.45	8.38	10.0	2.93	8.73	7.91
AdaptOvrSegBP [33]	9.8	1.69	2.04	5.64	0.14	0.20	1.47	7.04	11.1	16.4	3.60	8.96	8.84
SymBP+occ [7]	10.8	0.97	1.75	5.09	0.16	0.33	2.19	6.47	10.7	17.0	4.79	10.7	10.9
PlaneFitBP [32]	10.8	0.97	1.83	5.26	0.17	0.51	1.71	6.65	12.1	14.7	4.17	10.7	10.6
AdaptDispCalib [36]	11.6	1.19	1.42	6.15	0.23	0.34	2.50	7.80	13.6	17.3	3.62	9.33	9.72
Segm+visib [4]	12.2	1.30	1.57	6.92	0.79	1.06	6.76	5.00	6.54	12.3	3.72	8.62	10.2
C-SemiGlob [19]	12.3	2.61	3.29	9.89	0.25	0.57	3.24	5.14	11.8	13.0	2.77	8.35	8.20
SO+borders [29]	12.8	1.29	1.71	6.83	0.25	0.53	2.26	7.02	12.2	16.3	3.90	9.85	10.2
DistinctSM [27]	14.1	1.21	1.75	6.39	0.35	0.69	2.63	7.45	13.0	18.1	3.91	9.91	8.32
CostAggr+occ [39]	14.3	1.38	1.96	7.14	0.44	1.13	4.87	6.80	11.9	17.3	3.60	8.57	9.36
OverSegmBP [26]	14.9	1.69	1.97	8.47	0.51	0.68	4.69	6.74	11.9	15.8	3.19	8.81	8.89
SegmentSupport [28]	15.1	1.25	1.62	6.68	0.25	0.64	2.59	8.43	14.2	18.2	3.77	9.87	9.77

Traditional stereo

Advantages:

- works very well in non-occluded regions

Disadvantages:

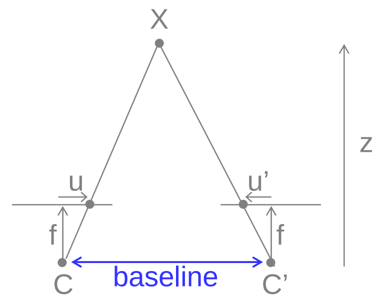
- restricted to two images (not)
- gets confused in occluded regions
- can't handle *mixed pixels*

Multi-View Stereo

Stereo Reconstruction

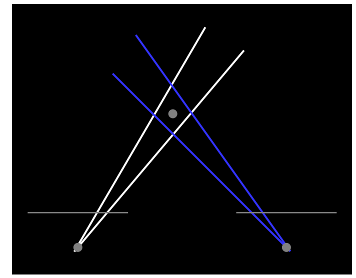
Steps

- Calibrate cameras
- Rectify images
- Compute disparity
- Estimate depth

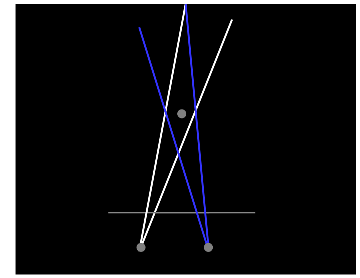


$$disparity = u - u' = \frac{baseline * f}{z}$$

Choosing the Baseline



Large Baseline



Small Baseline

What's the optimal baseline?

- Too small: large depth error
- Too large: difficult search problem

Effect of Baseline on Estimation

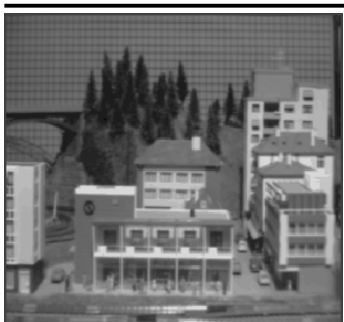
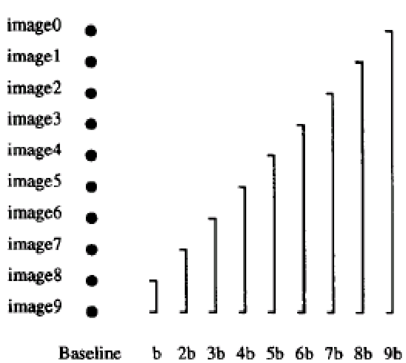


Figure 2: An example scene. The grid pattern in the background has ambiguity of matching.



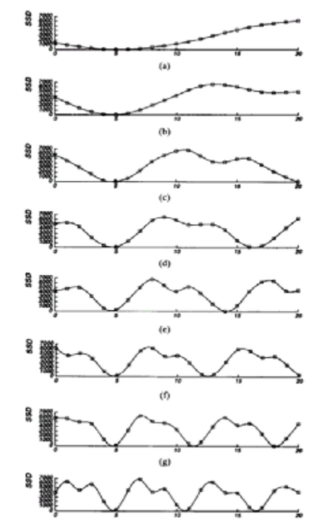


Fig. 5. SSD values versus inverse distance: (a) $B = b$; (b) $B = 2b$; (c) $B = 3b$; (d) $B = 4b$; (e) $B = 5b$; (f) $B = 6b$; (g) $B = 7b$; (h) $B = 8b$. The horizontal axis is normalized such that $8bF = 1$.

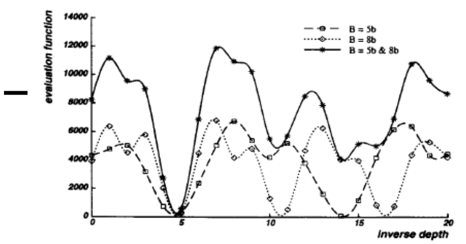


Fig. 6. Combining two stereo pairs with different baselines.

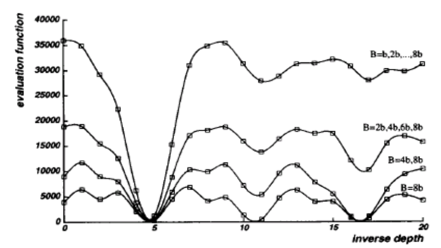


Fig. 7. Combining multiple baseline stereo pairs.

Multibaseline Stereo

Basic Approach

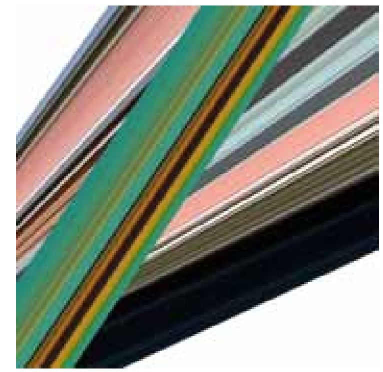
- Choose a reference view
- Use your favorite stereo algorithm BUT
  – replace two-view SSD with SSD over all baselines

Limitations

- Must choose a reference view
- Visibility: select which frames to match
  [Kang, Szeliski, Chai, CVPR'01]

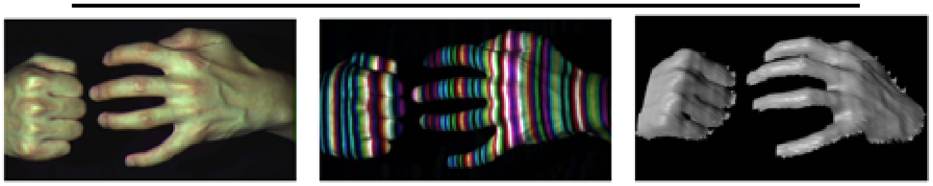
Epipolar-Plane Images [Bolles 87]

http://www.graphics.lcs.mit.edu/~aisaksen/projects/drlf/epi/

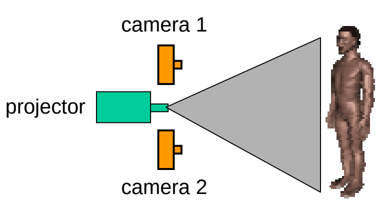


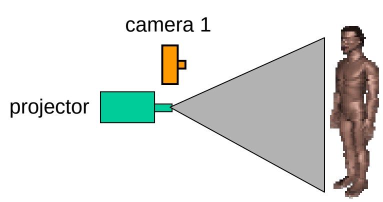
Lesson: Beware of ***occlusions***

Active stereo with structured light



Li Zhang's one-shot stereo





Project "structured" light patterns onto the object

- simplifies the correspondence problem

Spacetime Stereo

Li Zhang, Noah Snavely, Brian Curless, Steve Seitz

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

Summary

Applications

Image rectification

Matching criteria

Local algorithms (aggregation & diffusion)

Optimization algorithms

- energy (cost) formulation & Markov Random Fields
- mean-field; dynamic programming; stochastic; graph algorithms

Multi-View stereo

- visibility, occlusion-ordered sweeps

Bibliography

D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision, 47(1):7-42, May 2002.

R. Szeliski. Stereo algorithms and representations for image-based rendering. In British Machine Vision Conference (BMVC'99), volume 2, pages 314-328, Nottingham, England, September 1999.

G. M. Nielson, Scattered Data Modeling, IEEE Computer Graphics and Applications, 13(1), January 1993, pp. 60-70.

S. B. Kang, R. Szeliski, and J. Chai. Handling occlusions in dense multi-view stereo. In CVPR'2001, vol. I, pages 103-110, December 2001.

Y. Boykov, O. Veksler, and Ramin Zabih, Fast Approximate Energy Minimization via Graph Cuts, Unpublished manuscript, 2000.

A.F. Bobick and S.S. Intille. Large occlusion stereo. International Journal of Computer Vision, 33(3), September 1999. pp. 181-200

D. Scharstein and R. Szeliski. Stereo matching with nonlinear diffusion. International Journal of Computer Vision, 28(2):155-174, July 1998

Bibliography

Volume Intersection

- Martin & Aggarwal, “Volumetric description of objects from multiple views”, Trans. Pattern Analysis and Machine Intelligence, 5(2), 1991, pp. 150-158.
- Szeliski, “Rapid Octree Construction from Image Sequences”, Computer Vision, Graphics, and Image Processing: Image Understanding, 58(1), 1993, pp. 23-32.

Voxel Coloring and Space Carving

- Seitz & Dyer, “Photorealistic Scene Reconstruction by Voxel Coloring”, Proc. Computer Vision and Pattern Recognition (CVPR), 1997, pp. 1067-1073.
- Seitz & Kutulakos, “Plenoptic Image Editing”, Proc. Int. Conf. on Computer Vision (ICCV), 1998, pp. 17-24.
- Kutulakos & Seitz, “A Theory of Shape by Space Carving”, Proc. ICCV, 1998, pp. 307-314.

Bibliography

Related References

- Bolles, Baker, and Marimont, “Epipolar-Plane Image Analysis: An Approach to Determining Structure from Motion”, International Journal of Computer Vision, vol 1, no 1, 1987, pp. 7-55.
- Faugeras & Keriven, “Variational principles, surface evolution, PDE's, level set methods and the stereo problem", IEEE Trans. on Image Processing, 7(3), 1998, pp. 336-344.
- Szeliski & Golland, “Stereo Matching with Transparency and Matting”, Proc. Int. Conf. on Computer Vision (ICCV), 1998, 517-524.
- Roy & Cox, “A Maximum-Flow Formulation of the N-camera Stereo Correspondence Problem”, Proc. ICCV, 1998, pp. 492-499.
- Fua & Leclerc, “Object-centered surface reconstruction: Combining multi-image stereo and shading", International Journal of Computer Vision, 16, 1995, pp. 35-56.
- Narayanan, Rander, & Kanade, “Constructing Virtual Worlds Using Dense Stereo”, Proc. ICCV, 1998, pp. 3-10.