Lecture 27:
3-D Reconstruction II:
Stereo

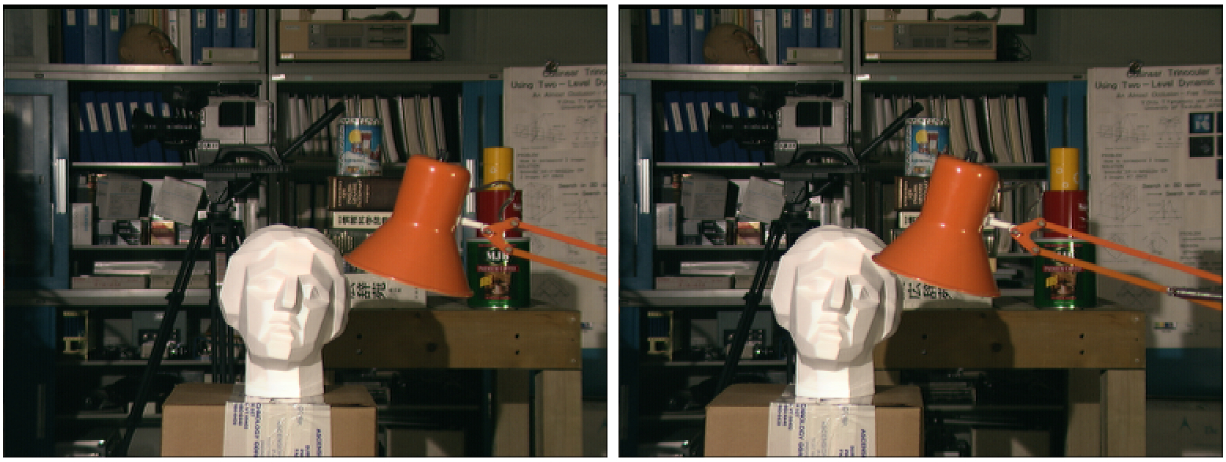
Contents

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29

Introduction

- So far, we have only investigated the projective geometry in the monocular case with a single pinhole camera.
- Considerering two cameras allows us to reconstruct the depth of a scene from the displacements between the two stereo images.
- This problem resembles the correspondence problem from optic flow estimation.
- However, we will see that the stereo geometry (also called *epipolar geometry*) creates an additional constraint.

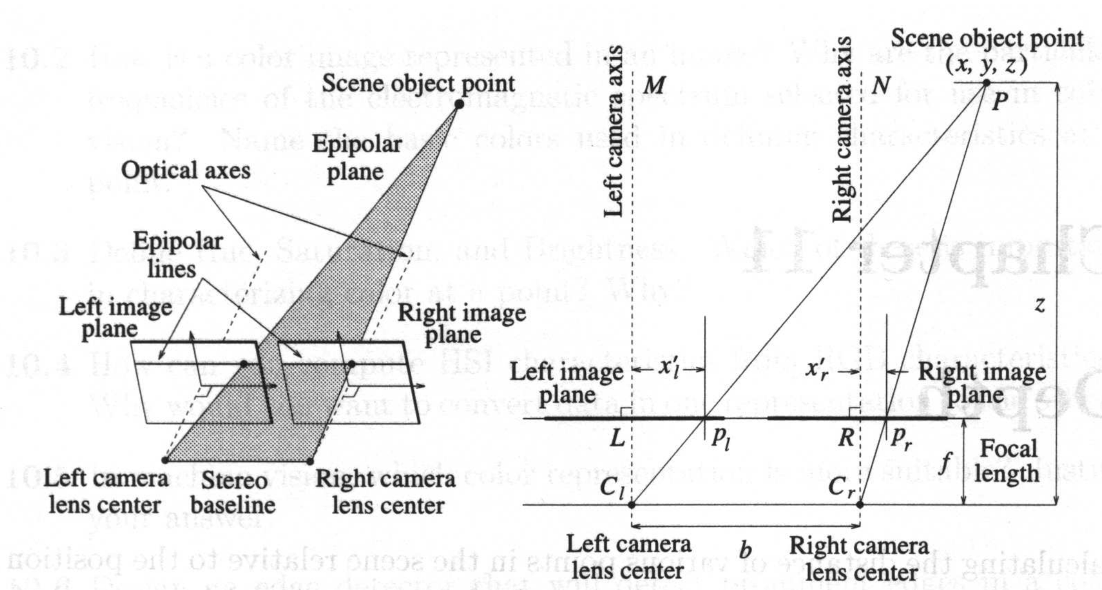
M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29



A stereo image pair from the Middlebury web page. The goal is to reconstruct the 3-D scene. Source: `http://cat.middlebury.edu/stereo/data.html`.

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29

Stereo Geometry for Orthoparallel Cameras



Stereo geometry. Authors: R. Jain, R. Kasturi, B. G. Schunck (1995).

A Simplified Model: Orthoparallel Cameras

- *orthoparallel cameras*: two identical cameras on the same height (y coordinate), with parallel optical axes
- *base line (Basislinie:)* connecting line between both optical centres (focal points)
- *base line distance (Basisliniendistanz) b:* distance between both optical centres
- *conjugated points (konjugierte Punkte):* two points in different images that result from the same 3-D scene point
- *disparity (Disparität):* distance between two conjugated points, if both images are superposed
- *epipolar plane (Epipolarebene):* plane through the scene point and both optical centres
- *epipolar lines (Epipolarlinien):* intersecting lines of the epipolar plane with both image planes; contain corresponding points

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29

Depth Computation

- Figure on Page 4, right:
  Put the origin of the coordinate system in the left camera lens centre C_l.
- From the similarity of the triangles PMC_l and $p_l LC_l$, it follows that

$$\frac{x}{z} = \frac{x'_l}{f}$$

- From the similarity of PNC_r and $p_r RC_r$ one obtains:

$$\frac{x-b}{z} = \frac{x'_r}{f}$$

- Elimination of x in both equations gives

$$z = \frac{bf}{x'_l - x'_r}.$$

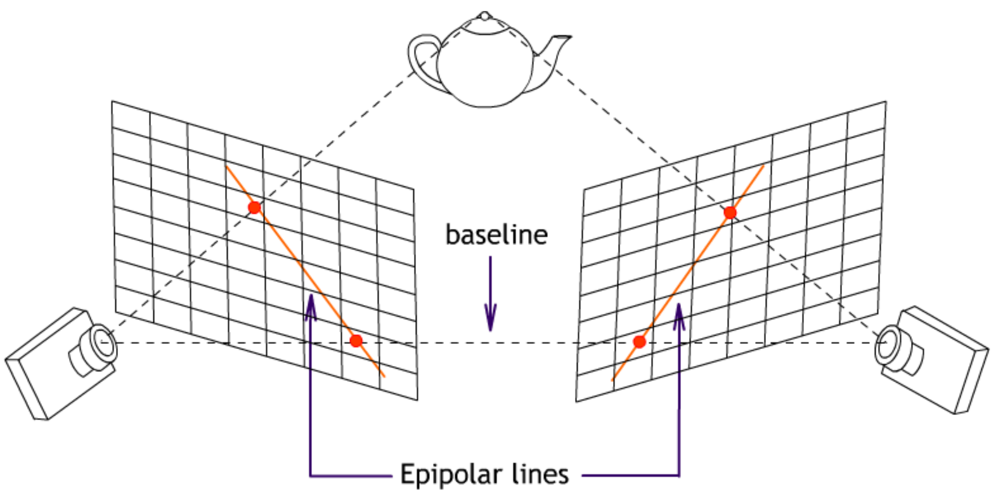
Stereo Geometry for Orthoparallel Cameras (4)

- This shows: If the baseline distance b and the focal length f is known in the orthoparallel case, the disparity $x_l' - x_r'$ allows to compute the depth z.
- The main problem is the reliable estimation of the disparity:
  - Often disparities can only be measured with pixel precision. This suggests to choose a large baseline distance.
  - On the other hand, this may lead to more occlusions and makes it more difficult to find correspondences between both images.

Stereo Geometry for Converging Cameras (1)

Stereo Geometry for Converging Cameras

- Also in the general case with so-called *converging cameras*, conjugated points can be found along the epipolar lines.
- In general, the two epipolar lines are no longer parallel.



Two converging cameras in arbitrary position and orientation. Author: N. Slesareva (2005).

Stereo Geometry for Converging Cameras (2)

Epipolar Constraint and Fundamental Matrix

- The pixel $\mathbf{m}_2$ in the second image that corresponds to the pixel $\mathbf{m}_1$ in the first image cannot lie everywhere.
- One can show that conjugated points $\mathbf{m}_1$ and $\mathbf{m}_2$ have to satisfy the *epipolar constraint*

  $$\tilde{\mathbf{m}}_2^\top \, F \, \tilde{\mathbf{m}}_1 = 0$$

  with a suitable 3×3 matrix F *(fundamental matrix)*.
- F has rank 2.
  It is thus not invertible, and offers 7 degrees of freedom (-scale,-rank).
  It is not possible to extract intrinsic and extrinsic parameters of the two cameras from F.

Stereo Geometry for Converging Cameras (3)

- If the fundamental matrix is known (*weakly calibrated system*) one knows for each pixel $\mathbf{m}_1$ in the first frame the corresponding epipolar line $\mathbf{l}_2$ in the second frame and vice versa:

  $$\tilde{\mathbf{m}}_2^\top \mathbf{l}_1 = 0 \qquad \text{with} \qquad \mathbf{l}_1 = F \, \tilde{\mathbf{m}}_1 \, ,$$
  $$\tilde{\mathbf{m}}_1^\top \mathbf{l}_2 = 0 \qquad \text{with} \qquad \mathbf{l}_2 = F^\top \tilde{\mathbf{m}}_2 \, .$$

  $\rightarrow$ *Reduced search space (1-D)* for stereo matching algorithms (*epipolar line*)
  Orthoparallel cameras even yield horizontal epipolar lines (search in x-direction)
- If the fundamental matrix is not known (*uncalibrated system*), one can estimate it from at least 7 point correspondences.

Stereo Geometry for Converging Cameras (4)

Estimation of the Fundamental Matrix

- Let us consider the epipolar constraint given by the equation

$$\tilde{\mathbf{m}}_2^\top \, F \, \tilde{\mathbf{m}}_1 = \begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix}^\top \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} = 0 \, .$$

Defining the following two vectors for correspondences and matrix entries

$$\begin{aligned} \mathbf{s} &= (x_1 \, x_2, \, y_1 \, x_2, \, x_2, \, x_1 \, y_2, \, y_1 \, y_2, \, y_2, \, x_1, \, y_1, \, 1)^\top \\ \mathbf{f} &= (f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33})^\top \end{aligned}$$

we can write the epipolar constraint in terms of the inner product

$$\tilde{\mathbf{m}}_\mathrm{r}^\top \, F \, \tilde{\mathbf{m}}_\mathrm{l} = \mathbf{s}^\top \mathbf{f} = 0 \, .$$

Stereo Geometry for Converging Cameras (5)

Two Types of Constraints

- *Correspondence Constraints:* For each known correspondence $(x_\mathrm{l}^i, y_\mathrm{l}^i) \leftrightarrow (x_\mathrm{r}^i, y_\mathrm{r}^i)$ we obtain one linear constraint

$$\mathbf{s}_i^\top \mathbf{f} = 0.$$

- *Rank-2 Constraint:* Additionally, we know that F has rank 2. Thus we obtain the following nonlinear (cubic) constraint

$$\begin{aligned} \det(F) \;=\;& f_{11}f_{22}f_{33} + f_{12}f_{23}f_{31} + f_{13}f_{21}f_{32} \\ & - f_{31}f_{22}f_{13} - f_{32}f_{23}f_{11} - f_{33}f_{21}f_{12} \;=\; 0 \, . \end{aligned}$$

- *Consequence:* We have two possibilities:
  - Use 8 correspondence constraints and enforce rank-2 constraint afterwards (linear problem).
  - Use 7 correspondence constraints and the rank-2 constraint (nonlinear problem, not discussed in this lecture).

Linear Approach

- *Idea:* Use $N \geq 8$ correspondence constraints and solve the resulting linear systems of equations given by

$$S\,\mathbf{f} = 0\,.$$

The lines of the $N \times 9$ matrix S are formed by the different constraints

$$S = \begin{pmatrix} \mathbf{s}_1^\top \\ \vdots \\ \mathbf{s}_N^\top \end{pmatrix}\,.$$

- *Problems:*
  - trivial solution $\mathbf{f} = 0$ always works (homogeneous system of equations)
  - no solution possible for $N > 8$ constraints (system is over-determined)
  - infinitely many solutions for less than 8 linearly independent constraints (can be avoided in advance by picking a suitable set of correspondences)

Total Least Squares Fit

- *Idea:* Sum up the squared deviations from the N constraints and minimise the resulting quadratic form
  (Longuet-Higgins 1981)

$$E(\mathbf{f}) = \sum_{i=1}^{N} (\mathbf{s}_i^\top \mathbf{f})^2 = \mathbf{f}^\top \left(\sum_{i=1}^{N} \mathbf{s}_i \mathbf{s}_i^\top \right) \mathbf{f} = \mathbf{f}^\top S^\top S\, \mathbf{f}$$

with explicit constraint $|\mathbf{f}| = 1$ to avoid the trivial solution $\mathbf{f} = \mathbf{0}$.

- *Minimisation:*. The solution that minimises $E(\mathbf{f})$ is given by the eigenvector $\mathbf{e}$ to the smallest eigenvalue λ of the symmetric 9×9 matrix

$$S^\top S = \sum_{i=1}^{N} \mathbf{s}_i \mathbf{s}_i^\top\,.$$

Enforcing the Rank-2 Constraint

- Enforce rank-2 constraint by setting the smallest singular value of F zero
- *Example:* Let us consider the singular value decomposition of F, where the singular values are assumed to be ordered, i.e. $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq 0$

$$F = UDV^T = U \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} V^T .$$

  Here, U and V are orthogonal matrices containing left and right eigenvectors.

- *Projection:* Setting the smallest singular value of F zero, i.e. $\sigma_3 := 0$, we obtain

$$F' = U \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & 0 \end{pmatrix} V^T$$

  which is the closest rank-2 matrix to original matrix F in the Frobenius norm.

Correlation-Based Methods

Basic Problem

- estimation of the disparity is a correspondence problem:
  find conjugated points
- similar to the optic flow problem
- main difference:
  search space is smaller, since conjugated points are located on the epipolar line

Correlation-Based Methods (2)

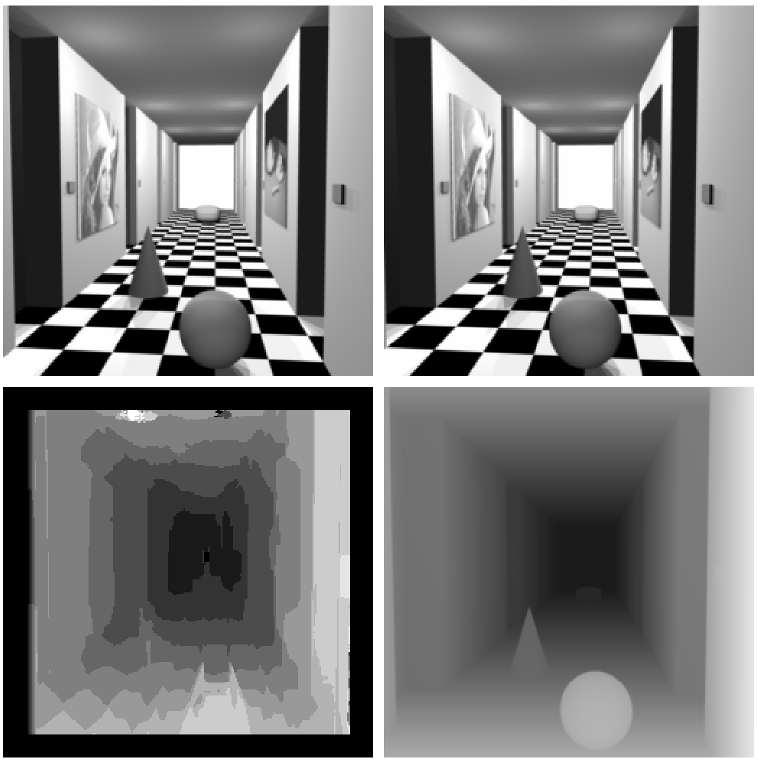
Correlation Method

- searches for corresponding local structures in both images
- The search can be restricted along epipolar lines.
  Often one also limits the largest displacement.
- The correlation coefficient may serve as a quality measure of the match.
  It compares the neighbourhood B_ρ around two points (x_0, y_0) and $(x_0 + u, y_0 + v)$.
- Let f_l, f_r denote the two images and $B_\rho(x_0, y_0)$ a neighbourhood around (x_0, y_0) with radius ρ. Then the *correlation coefficient (Korrelationskoeffizient)* is given by

$$\frac{\int_{B_\rho(x_0,y_0)} \left(f_l(x,y) - \bar{f}_l\right)\, \left(f_r(x{+}u, y{+}v) - \bar{f}_r\right)\, dx\, dy}{\sqrt{\int_{B_\rho(x_0,y_0)} \left(f_l(x,y) - \bar{f}_l\right)^2\, dx\, dy}\; \sqrt{\int_{B_\rho(x_0,y_0)} \left(f_r(x{+}u, y{+}v) - \bar{f}_r\right)^2\, dx\, dy}}$$

  where $\bar{f}_l$, $\bar{f}_r$ denote the mean values in both neighbourhoods.
- Similar to a normalised inner product $\frac{a^\top b}{|a|\,|b|}$ with values in $[-1, 1]$, the correlation coefficient attains values in $[-1, 1]$.

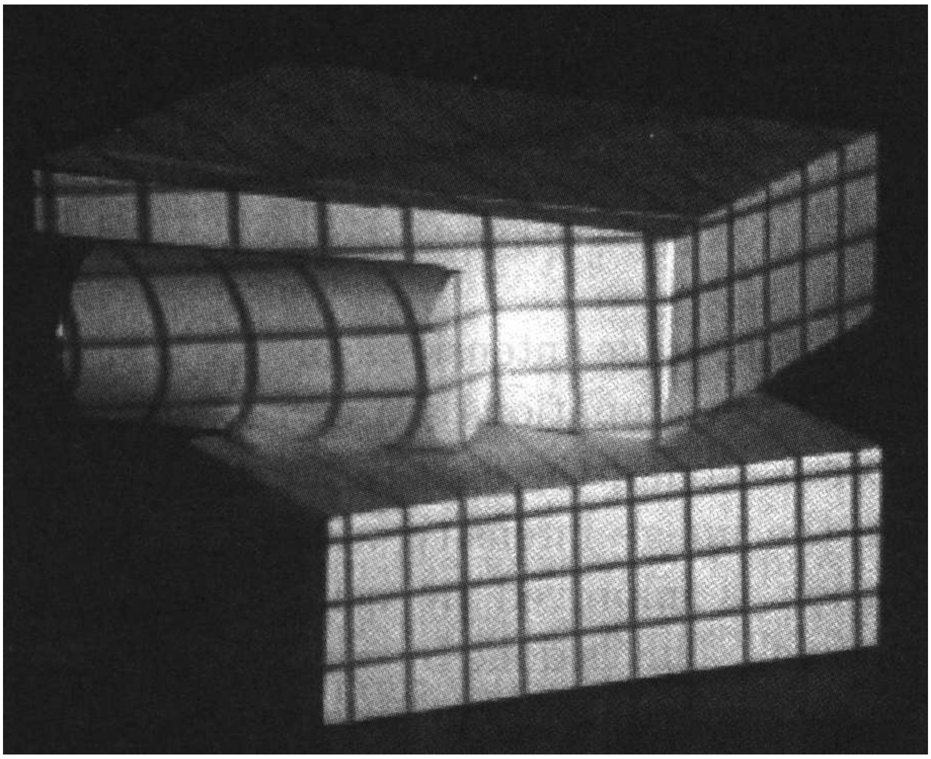
Correlation-Based Methods (3)



(a) Top left: Left image. (from `http://www-dbv.cs.uni-bonn.de/stereo_data/`). **(b) Top right:** Right image. **(c) Bottom left:** Disparity computed with a correlation window of size 17×17. **(d) Bottom right:** Ground truth. Authors: A. Bruhn, J. Weickert (2006).

Advantages and Shortcomings of Correlation-Based Methods

- simple and fast
- unreliable results in flat regions
- possibility of improving the quality:
  structured illumination



Structured illumination: Light patterns are projected onto the object in order to create additional features for matching both images. Authors: R. Jain, R. Kasturi, B. G. Schunck (1995).

Variational Methods

Assumptions:

- For simplicity of notation, consider the orthoparallel case:
  The epipolar lines are parallel to the x-axis, i.e. no disparities occur in y-direction.
- Grey value constancy assumption:
  Conjugated points have the same grey values.
- Smoothness assumption:
  The disparity u varies smoothly in space.

Energy Functional:

$$E(u) \;=\; \int_\Omega \left(\Big(f_r(x+u,y) - f_l(x,y) \Big)^2 + \alpha \, |\nabla u|^2 \right) dx\, dy.$$

Problem:

- The data term $(f_r(x+u,y) - f_l(x,y))^2$ cannot guarantee convexity.
  Since the disparities can be large, a Taylor linearisation is inappropriate.
- Thus, the energy functional may have many local minima.
- No global convergence of numerical methods:
  Different initialisations may lead to different local minima.

Basic Idea Behind a Possible Numerical Continuation Strategy

- Smooth the problem in order to reduce the number of local minima:
  Replace the image data f_r, f_l in the Euler–Lagrange equation

$$(f_r(x+u, y) - f_l(x, y))\, \partial_x f_r(x+u, y) - \alpha\, \Delta u \;=\; 0$$

  by Gaussian-smoothed versions $K_{\sigma_0} * f_r$ and $K_{\sigma_0} * f_l$.

- Solve this problem with an iterative numerical method (next page).
  If σ_0 is sufficiently large, the convergence is nonlocal.
  A good value for σ_0 is the expected order of magnitude for the disparity.

- Reduce the width of the Gaussian (e.g. $\sigma_1 := 0.95\,\sigma_0$) and solve the problem with the solution for σ_0 as initialisation.

- Continue this strategy until σ_k reaches the pixel scale.

Numerical Solution of the Euler–Lagrange Equation

- Discretise the derivatives with simple finite differences (cf. Lecture 18) and evaluate $\partial_x f_r(x+u, y)$ by means of linear interpolation.

- Solve the resulting nonlinear system of equations with a suitable iterative method (e.g. nonlinear Gauß–Seidel or linear Jacobi method after using a Taylor linearisation)

Advantages and Shortcomings of Variational Methods

- filling-in effect creates dense disparity maps of acceptable quality

- relatively high computational efforts

Possible Model Refinements

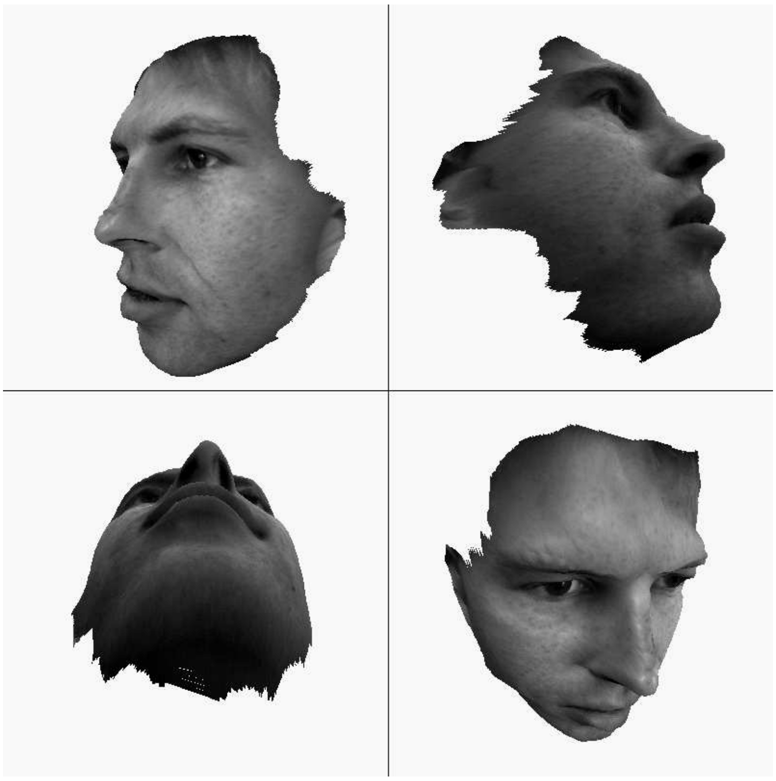
- nonquadratic smoothness terms that reduce smoothing across discontinuities of the image or its disparities.

Variational Methods (5)

A stereo image pair from INRIA Sophia-Antipolis. The epipolar lines are not horizontal in this case.

M I A
1 2
3 4
5 6
7 8
9 10
11 12
13 14
15 16
17 18
19 20
21 22
23 24
25 26
27 28
29

Variational Methods (6)



Four views of a stereo reconstruction with a variational method that reduces smoothing across image edges. Authors: L. Alvarez, R. Deriche, J. Sánchez, J. Weickert (2002).

M I A
1 2
3 4
5 6
7 8
9 10
11 12
13 14
15 16
17 18
19 20
21 22
23 24
25 26
27 28
29

Summary

- Conjugated points in two stereo images are located on the epipolar lines (epipolar constraint).
- The disparity allows to compute the depth information if the focal length and the baseline distance is known.
- Estimating the disparity is a correspondence problem similar to optic flow estimation with large displacements.
  However, the search space is restricted due to the epipolar constraint.
- Correlation methods are simple and fast, but may give poor results in flat regions.
  Structured illuminatation can increase the quality.
- Global methods based on variational approaches create dense disparity maps, but are computationally more demanding.

M I A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29

Literature

- R. Jain, R. Kasturi, B. G. Schunck: *Machine Vision*. McGraw-Hill, New York, 1995.
  (see Sections 11.1 and 11.2)
- L. Alvarez, R. Deriche, J. Sánchez, J. Weickert: Dense disparity map estimation respecing image discontinuities: a PDE and scale-space based approach. *Journal of Visual Communication and Image Representation*, Vol. 13, No. 1/2, pp. 3–21, 2002.
  `www-sop.inria.fr/rapports/sophia/RR-3874`
  (example of a variational approach)
- G. Xu, Z. Zhang: *Epipolar Geometry in Stereo, Motion and Object Recognition: A Unified Approach*. Kluwer, Dordrecht, 1996.
  (excellent book on stereo)
- R. Klette, K. Schlüns, A. Koschan: *Computer Vision: Three-Dimensional Data from Images*. Springer, Singapore, 1998.
  (Chapter 4 describes geometrical foundations and stereo reconstruction methods.)
- O. Faugeras, Q.-T. Luong, T. Papadopoulo: *The Geometry of Multiple Images*. MIT Press, Cambridge, MA, 2001.
  (Chapters 5-7 deal with stereo geometry)
- R. Hartley, A. Zisserman: *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, 2000.
  (Part II on stereo geometry)

Summary (3)

- Y. Ma, S. Soatto, J. Kosecká, S. S. Sastry: *An Invitation to 3-D Vision*. Springer, New York, 2002. *(Chapters 4–6 on stereo geometry)*
- Middlebury Stereo Vision Page: `http://vision.middlebury.edu/stereo/`. *(compares state-of-the-art algorithms)*