

Stereo Vision III

Introduction to Computer Vision
CSE 152
Lecture 15

CSE152, Spring 2012

Intro Computer Vision

Stereo Vision Outline

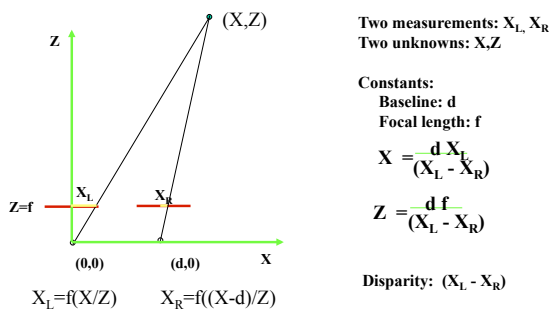
- Offline: Calibrate cameras & determine “epipolar geometry”
- Online
 1. Acquire stereo images
 2. Rectify images to convenient epipolar geometry
 3. Establish correspondence
 4. Estimate depth

CSE152, Spring 2012

Intro Computer Vision

BINOCULAR STEREO SYSTEM

Estimating Depth\
2D world with 1-D image plane



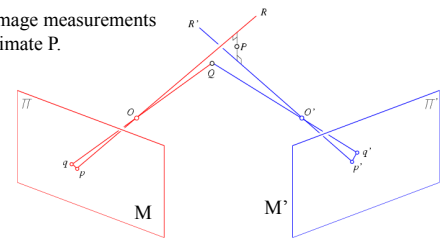
CSE152, Spring 2012

(Adapted from Hager)

Intro Computer Vision

Reconstruction: General 3-D case

Given two image measurements
 p and p' , estimate P .



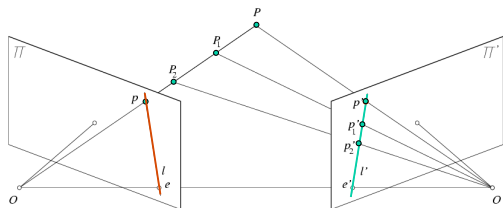
- Linear Method: find P such that

$$\begin{cases} p \times MP = 0 \\ p' \times M'P = 0 \end{cases} \iff \begin{pmatrix} [p]_{\times} M \\ [p']_{\times} M' \end{pmatrix} P = 0$$
 Where M is camera matrix

- Non-Linear Method: find Q minimizing $d^2(p, q) + d^2(p', q')$
where $q = MQ$ and $q' = M'Q$

CSE152, Spring 2012

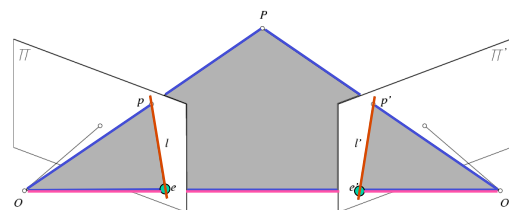
Intro Computer Vision



- Potential matches for p have to lie on the corresponding epipolar line l' .
- Potential matches for p' have to lie on the corresponding epipolar line l .

CSE152, Spring 2012

Intro Computer Vision



- Baseline
- Epipolar Plane
- Epipoles
- Epipolar Lines

CSE152, Spring 2012

Intro Computer Vision

Skew Symmetric Matrix & Cross Product

- The cross product $\mathbf{a} \times \mathbf{b}$ of two vectors $\mathbf{a}$ and $\mathbf{b}$ can be expressed as a matrix vector product $[\mathbf{a}_\times] \mathbf{b}$ where $[\mathbf{a}_\times]$ is the skew symmetric matrix:

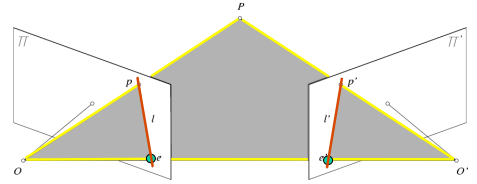
$$[\mathbf{a}_\times] = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}$$

- A matrix $\mathbf{S}$ is skew symmetric iff $\mathbf{S} = -\mathbf{S}^T$

CSE152, Spring 2012

Intro Computer Vision

Epipolar Constraint: Calibrated Case



$$\vec{Op} \cdot [\vec{OO'} \times \vec{Op}] = 0 \Rightarrow \mathbf{p} \cdot [\mathbf{t} \times (\mathcal{R}\mathbf{p}')] = 0 \quad \text{with} \quad \begin{cases} \mathbf{p} = (u, v, 1)^T \\ \mathbf{p}' = (u', v', 1)^T \\ \mathcal{M} = (\mathbf{Id} \quad \mathbf{0}) \\ \mathcal{M}' = (\mathcal{R}^T, -\mathcal{R}^T \mathbf{t}) \end{cases}$$

Note: Coordinates of $\mathbf{p}$ and $\mathbf{p}'$ are "calibrated" so that $(u, v, 1)$ is a 3-D vector pointing in the direction of $\mathbf{p}$.

Essential Matrix
(Longuet-Higgins, 1981)

$$\mathbf{p}^T \mathcal{E} \mathbf{p}' = 0 \quad \text{with} \quad \mathcal{E} = [\mathbf{t}_\times] \mathcal{R}$$

CSE152, Spring 2012

Intro Computer Vision

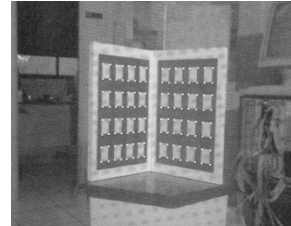
Two ways to estimate the Essential Matrix

1. Calibration-based
2. Eight-Point Algorithm

CSE152, Spring 2012

Intro Computer Vision

Calibration-Based Method



1. From image of known calibration fixture, determine intrinsic parameters and extrinsic relation of two cameras.
2. Compute the relative position and orientation $(\mathbf{R}, \mathbf{t})$ of the two cameras from $\mathbf{R}_1, \mathbf{R}_2, \mathbf{t}_1, \mathbf{t}_2$
3. Compute the Essential Matrix $\mathcal{E} = [\mathbf{t}_\times] \mathbf{R}$

CSE152, Spring 2012

Intro Computer Vision

The Eight-Point Algorithm (Longuet-Higgins, 1981)

$$[u, v, 1] \begin{bmatrix} E_{11} & E_{12} & E_{13} \\ E_{21} & E_{22} & E_{23} \\ E_{31} & E_{32} & E_{33} \end{bmatrix} \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = 0 \Rightarrow [uu' \quad uv' \quad u \quad uu' \quad vv' \quad v \quad u' \quad v' \quad 1] \begin{bmatrix} E_{11} \\ E_{12} \\ E_{13} \\ E_{21} \\ E_{22} \\ E_{23} \\ E_{31} \\ E_{32} \\ E_{33} \end{bmatrix} = 0$$

• Set E_{33} to 1
• Consider 8 points $(u_i, v_i), (u'_i, v'_i), i=1..8$

$$\begin{bmatrix} u_1 u'_1 & u_1 v'_1 & u_1 & u_1 u'_1 & v_1 v'_1 & v_1 & u'_1 & v'_1 \\ u_2 u'_2 & u_2 v'_2 & u_2 & u_2 u'_2 & v_2 v'_2 & v_2 & u'_2 & v'_2 \\ u_3 u'_3 & u_3 v'_3 & u_3 & u_3 u'_3 & v_3 v'_3 & v_3 & u'_3 & v'_3 \\ u_4 u'_4 & u_4 v'_4 & u_4 & u_4 u'_4 & v_4 v'_4 & v_4 & u'_4 & v'_4 \\ u_5 u'_5 & u_5 v'_5 & u_5 & u_5 u'_5 & v_5 v'_5 & v_5 & u'_5 & v'_5 \\ u_6 u'_6 & u_6 v'_6 & u_6 & u_6 u'_6 & v_6 v'_6 & v_6 & u'_6 & v'_6 \\ u_7 u'_7 & u_7 v'_7 & u_7 & u_7 u'_7 & v_7 v'_7 & v_7 & u'_7 & v'_7 \\ u_8 u'_8 & u_8 v'_8 & u_8 & u_8 u'_8 & v_8 v'_8 & v_8 & u'_8 & v'_8 \end{bmatrix} \begin{bmatrix} E_{11} \\ E_{12} \\ E_{13} \\ E_{21} \\ E_{22} \\ E_{23} \\ E_{31} \\ E_{32} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Solve E_{11} to E_{32} --
The Essential Matrix

CSE152, Spring 2012

Intro Computer Vision

Properties of the Essential Matrix

$$\mathbf{p}^T \mathcal{E} \mathbf{p}' = 0 \quad \text{with} \quad \mathcal{E} = [\mathbf{t}_\times] \mathcal{R}$$

- $\mathcal{E} \mathbf{p}'$ is the epipolar line associated with $\mathbf{p}'$.
- $\mathcal{E}^T \mathbf{p}$ is the epipolar line associated with $\mathbf{p}$.
- $\mathcal{E} \mathbf{e}' = 0$ and $\mathcal{E}^T \mathbf{e} = 0$.
- The epipoles $\mathbf{e}$ and $\mathbf{e}'$ are the Eigenvectors of the matrix $\mathbf{F}$ and $\mathbf{F}'$ corresponding to the smallest Eigenvalue
- $\mathcal{E}$ is singular.
- $\mathcal{E}$ has two equal non-zero singular values (Huang and Faugeras, 1989).

CSE152, Spring 2012

Intro Computer Vision

The Eight-Point Algorithm (Longuet-Higgins, 1981)

$$\begin{bmatrix} E_{11} & E_{12} & E_{13} \\ E_{21} & E_{22} & E_{23} \\ E_{31} & E_{32} & E_{33} \end{bmatrix} \begin{bmatrix} u' \\ v' \\ 1 \end{bmatrix} = 0 \quad \Rightarrow \quad \begin{bmatrix} uu' & uv' & u & uu' & vv' & v & u' & v' & 1 \end{bmatrix} \begin{bmatrix} E_{11} \\ E_{12} \\ E_{13} \\ E_{21} \\ E_{22} \\ E_{23} \\ E_{31} \\ E_{32} \\ E_{33} \end{bmatrix} = 0$$

- Following, the text, view this as system of homogenous equations in E_{11} to E_{33}
- Solve as Eigenvector corresponding to the smallest Eigenvalue of matrix created from the image data.

Equivalent to solving

Minimize:

$$\sum_{i=1}^n (\mathbf{p}_i^T \mathbf{E} \mathbf{p}_i)^2$$

under the constraint

$$|\mathbf{E}|^2 = 1.$$

CSE152, Spring 2012

Intro Computer Vision

The Fundamental Matrix

The epipolar constraint is given by: $\mathbf{p}^T \mathbf{E} \mathbf{p}' = 0$ with $\mathbf{E} = [\mathbf{t}_\times] \mathbf{R}$

where $\mathbf{p}$ and $\mathbf{p}'$ are 3-D coordinates of the image coordinates of points in the two images.

Without calibration, we can still identify corresponding points in two images, but we can't convert to 3-D coordinates. However, the relationship between the calibrated coordinates $(\mathbf{p}, \mathbf{p}')$ and uncalibrated image coordinates $(\mathbf{q}, \mathbf{q}')$ can be expressed as $\mathbf{p} = \mathbf{A}\mathbf{q}$, and $\mathbf{p}' = \mathbf{A}'\mathbf{q}'$

Therefore, we can express the epipolar constraint as:

$$(\mathbf{A}\mathbf{q})^T \mathbf{E} (\mathbf{A}'\mathbf{q}') = \mathbf{q}^T (\mathbf{A}^T \mathbf{E} \mathbf{A}') \mathbf{q}' = \mathbf{q}^T \mathbf{F} \mathbf{q}' = 0$$

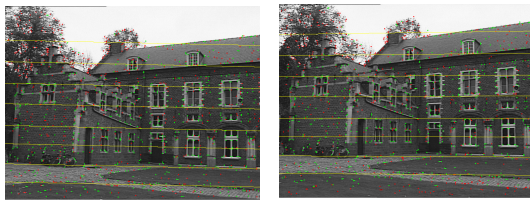
where $\mathbf{F}$ is called the Fundamental Matrix.

Can estimate $\mathbf{F}$ using 8 point algorithm WITHOUT CALIBRATION

CSE152, Spring 2012

Intro Computer Vision

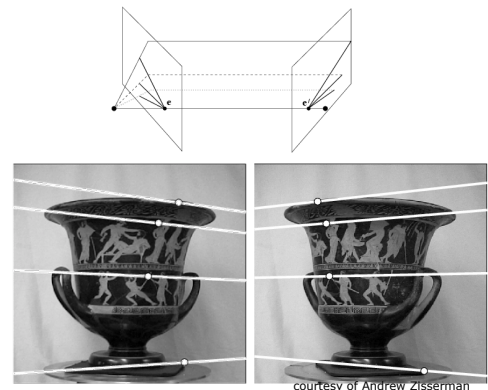
Epipolar geometry example



CSE152, Spring 2012

Intro Computer Vision

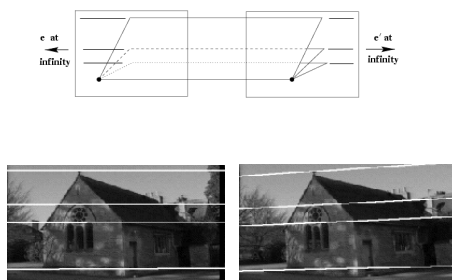
Example: converging cameras



CSE152, Spring 2012

courtesy of Andrew Zisserman

Example: motion parallel with image plane

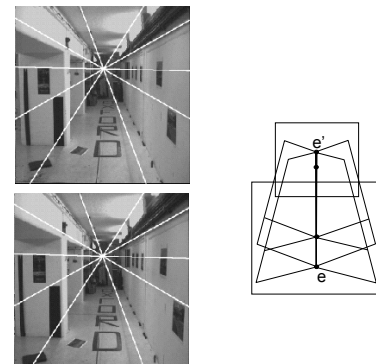


(simple for stereo → rectification)

courtesy of Andrew Zisserman

CSE152, Spring 2012

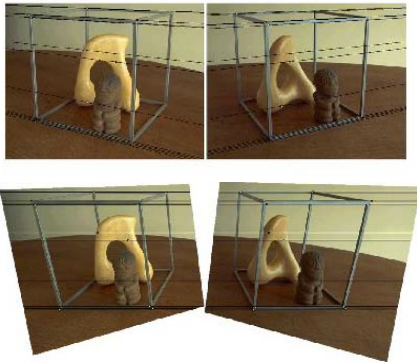
Example: forward motion



CSE152, Spring 2012

courtesy of Andrew Zisserman

Rectification

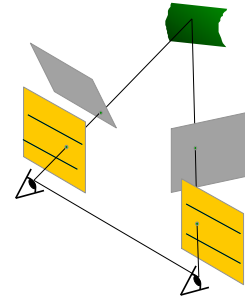


CSE152, Spring 2012

Intro Computer Vision

Rectification

Given a pair of images, transform both images so that epipolar lines are scan lines.

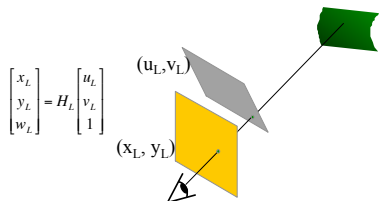


CSE152, Spring 2012

Intro Computer Vision

Rectification

Under perspective projection, the mapping from a plane to a plane is given by a projective transformation (aka homography).

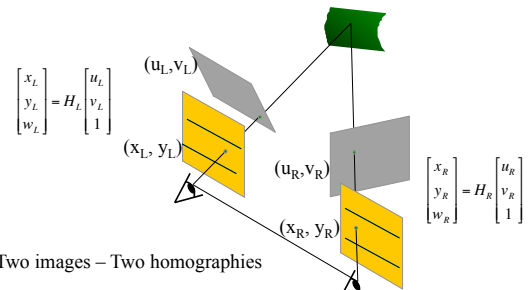


CSE152, Spring 2012

Intro Computer Vision

Rectification

Under perspective projection, the mapping from a plane to a plane is given by a projective transformation (aka homography).



Two images – Two homographies

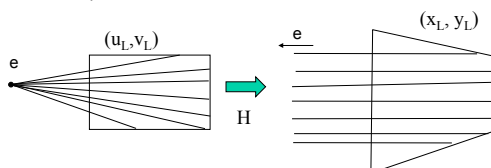
CSE152, Spring 2012

Intro Computer Vision

Image pair rectification

Simplify stereo matching by warping the images

Apply projective transformation so that epipolar lines correspond to horizontal scanlines



H should map epipole e to (1,0,0), a point at infinity

H should minimize image distortion

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = H e$$

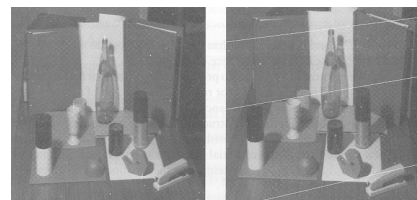
Note that rectified images usually not rectangular
See Section 7.3.7 for specific method

CSE152, Spring 2012

Intro Computer Vision

Rectification

Given a pair of images, transform both images so that epipolar lines are scan lines.



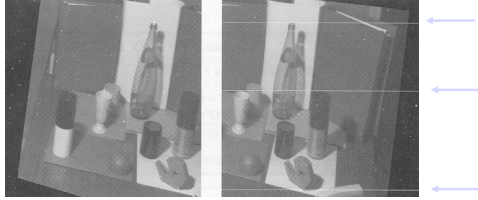
Input Images

CSE152, Spring 2012

Intro Computer Vision

Rectification

Given a pair of images, transform both images so that epipolar lines are scan lines.



Rectified Images

See Section 7.3.7 for specific method

CSE152, Spring 2012

Intro Computer Vision

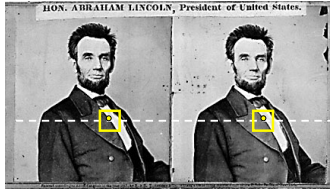
Features on same epipolar line



CSE152, Spring 2012

Intro Computer Vision

Using epipolar & constant Brightness constraints for stereo matching



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
- This will never work, so:

match **windows**

(Seitz)

CSE152, Spring 2012

Intro Computer Vision

Finding Correspondences



$W(p_l)$

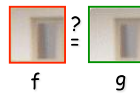


$W(p_r)$

CSE152, Spring 2012

Intro Computer Vision

Comparing Windows:



$$\left. \begin{aligned} SSD &= \sum_{[i,j] \in R} (f(i,j) - g(i,j))^2 \\ C_{fg} &= \sum_{[i,j] \in R} f(i,j)g(i,j) \end{aligned} \right\} \text{Most popular}$$

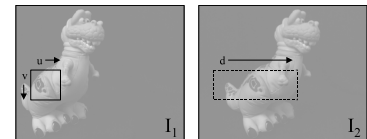
For each window, match to closest window on epipolar line in other image.

(Camps)

CSE152, Spring 2012

Intro Computer Vision

Correspondence Search Algorithm



```

For i = 1:rows
  for j = 1:ncols
    best(i,j) = -1
    for k = mindisparity:maxdisparity
      c = Match_Metric(I1(i,j), I2(i,j+k), winsize)
      if (c > best(i,j))
        best(i,j) = c
        disparities(i,j) = k
      end
    end
  end
end

```

$O(\text{rows} * \text{ncols} * \text{disparities} * \text{winx} * \text{winy})$

CSE152, Spring 2012

Intro Computer Vision

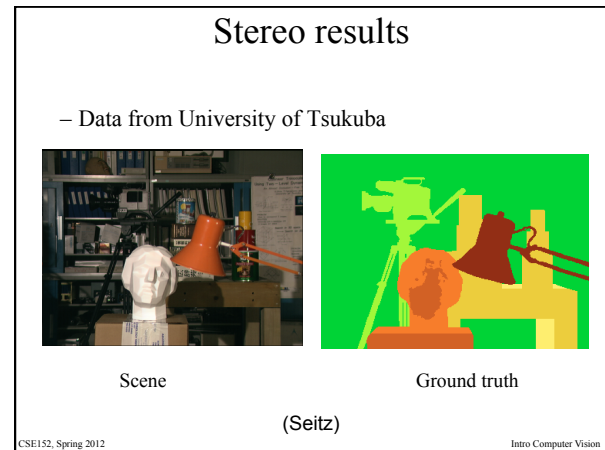
Match Metric Summary

MATCH METRIC	DEFINITION
Normalized Cross-Correlation (NCC)	$\frac{\sum_{u,v} (I_1(u,v) - \bar{I}_1)(I_2(u+d,v) - \bar{I}_2)}{\sqrt{\sum_{u,v} (I_1(u,v) - \bar{I}_1)^2} \sqrt{\sum_{u,v} (I_2(u+d,v) - \bar{I}_2)^2}}$
Sum of Squared Differences (SSD)	$\sum_{u,v} (I_1(u,v) - I_2(u+d,v))^2$
Normalized SSD	$\frac{\sum_{u,v} (I_1(u,v) - \bar{I}_1)(I_2(u+d,v) - \bar{I}_2)}{\sqrt{\sum_{u,v} (I_1(u,v) - \bar{I}_1)^2} \sqrt{\sum_{u,v} (I_2(u+d,v) - \bar{I}_2)^2}}$
Sum of Absolute Differences (SAD)	$\sum_{u,v} I_1(u,v) - I_2(u+d,v) $
Zero Mean SAD	$\sum_{u,v} I_1(u,v) - \bar{I}_1 - (I_2(u+d,v) - \bar{I}_2) $
Rank	$I_1(u,v) = \sum_n I_1(m,n) < I_1(u,v)$ $\sum_{u,v} (I_1(u,v) - I_2(u+d,v))$
Census	$I_1(u,v) = \text{BITSTRING}_i(I_1(u,v) < I_1(m,n))$ $\sum_{u,v} \text{HAMMING}(I_1(u,v), I_2(u+d,v))$

These two are actually the same

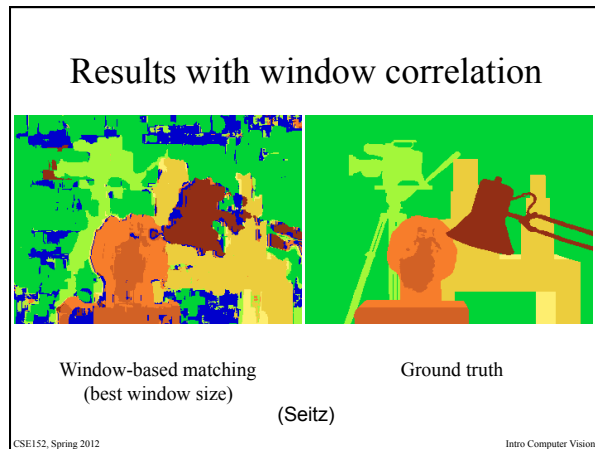
CSE152, Spring 2012

Intro Computer Vision



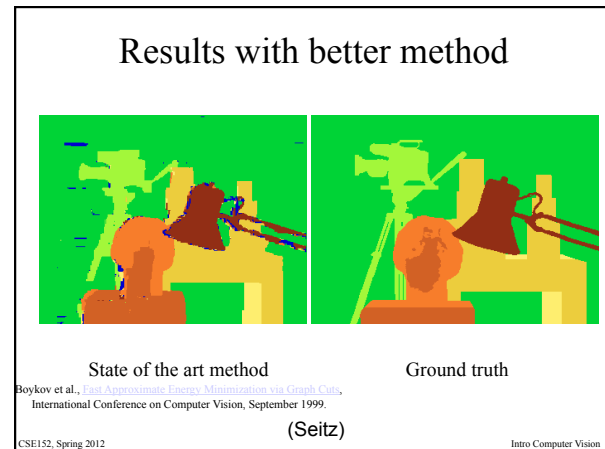
CSE152, Spring 2012

Intro Computer Vision



CSE152, Spring 2012

Intro Computer Vision



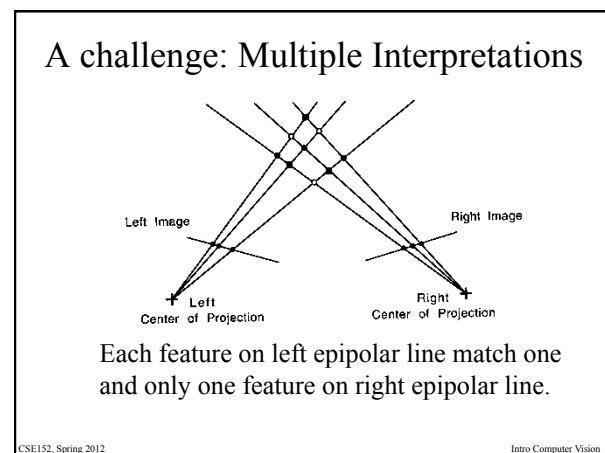
CSE152, Spring 2012

Intro Computer Vision

- ### Some Issues
- Epipolar ordering
 - Window size
 - Ambiguity
 - Window shape
 - Lighting
 - Half occluded regions

CSE152, Spring 2012

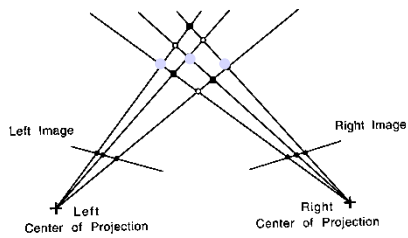
Intro Computer Vision



CSE152, Spring 2012

Intro Computer Vision

Multiple Interpretations

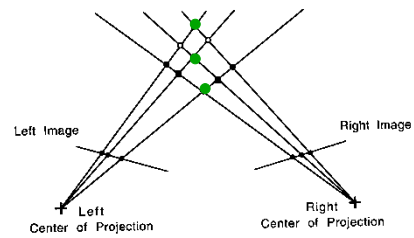


Each feature on left epipolar line match one and only one feature on right epipolar line.

CSE152, Spring 2012

Intro Computer Vision

Multiple Interpretations

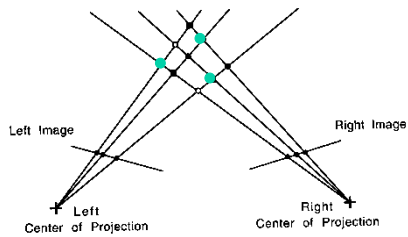


Each feature on left epipolar line match one and only one feature on right epipolar line.

CSE152, Spring 2012

Intro Computer Vision

Multiple Interpretations

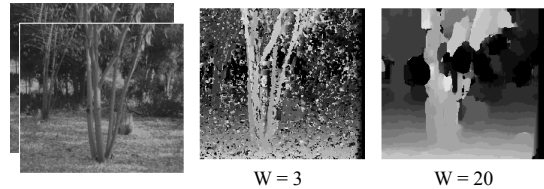


Each feature on left epipolar line match one and only one feature on right epipolar line.

CSE152, Spring 2012

Intro Computer Vision

Window size



$W = 3$

$W = 20$

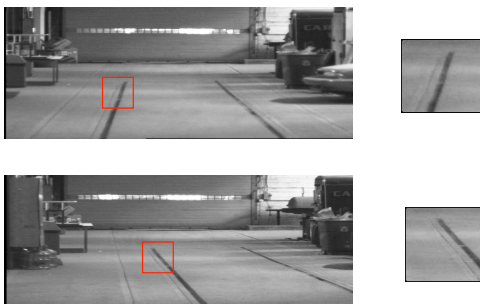
- Effect of window size Better results with *adaptive window*
 - T. Kanade and M. Okutomi, *A Stereo Matching Algorithm with an Adaptive Window, Theory and Experiment*, Proc. International Conference on Robotics and Automation, 1991.
 - D. Scharstein and R. Szeliski, *Stereo Matching using Dynamic Programming*, International Journal of Computer Vision, 28(2): 155-174, July 1998

(Seitz)

CSE152, Spring 2012

Intro Computer Vision

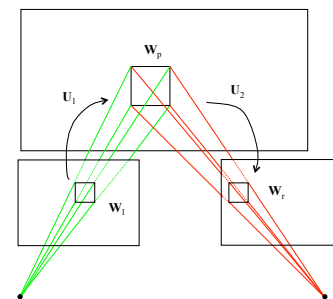
Window Shape and Forshortening



CSE152, Spring 2012

Intro Computer Vision

Window Shape: Fronto-parallel Configuration



CSE152, Spring 2012

Intro Computer Vision

Lighting Conditions (Photometric Variations)



$W(P_l)$



$W(P_r)$

CSE152, Spring 2012

Intro Computer Vision

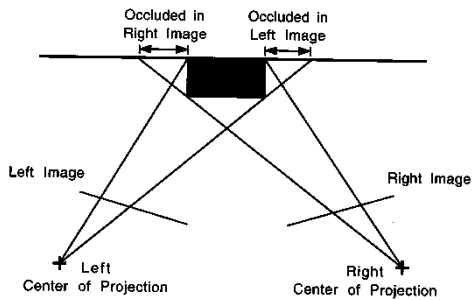
Ambiguity



CSE152, Spring 2012

Intro Computer Vision

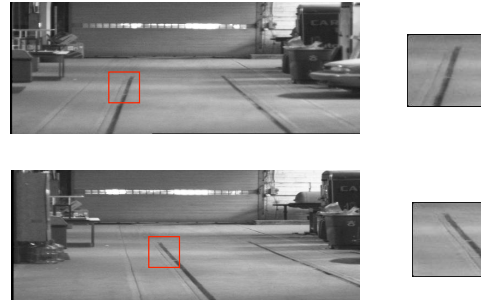
Problem of Occlusion



CSE152, Spring 2012

Intro Computer Vision

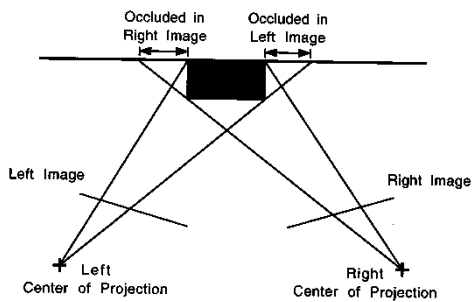
Window Shape and Foreshortening



CSE152, Spring 2012

Intro Computer Vision

Problem of Occlusion



CSE152, Spring 2012

Intro Computer Vision

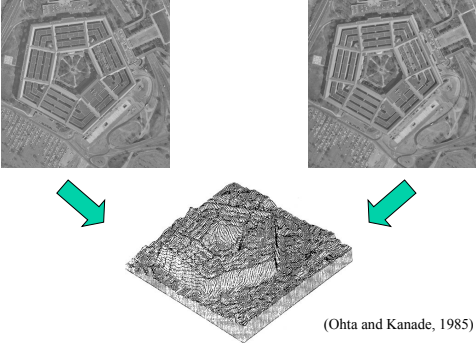
Stereo Constraints

CONSTRAINT	BRIEF DESCRIPTION
1-D Epipolar Search	Arbitrary images of the same scene may be rectified based on epipolar geometry such that stereo matches lie along one-dimensional scanlines. This reduces the computational complexity and also reduces the likelihood of false matches.
Monotonic Ordering	Points along an epipolar scanline appear in the same order in both stereo images, assuming that all objects in the scene are approximately the same distance from the camera.
Image Brightness Constancy	Assuming Lambertian surfaces, the brightness of corresponding points in stereo images are the same.
Match Uniqueness	For every point in one stereo image, there is at most one corresponding point in the other image.
Disparity Continuity	Disparities vary smoothly (i.e. disparity gradient is small) over most of the image. This assumption is violated at object boundaries.
Disparity Limit	The search space may be reduced significantly by limiting the disparity range, reducing both computational complexity and the likelihood of false matches.
Fronto-Parallel Surfaces	The implicit assumption made by area-based matching is that objects have fronto-parallel surfaces (i.e. depth is constant within the region of local support). This assumption is violated by sloping and creased surfaces.
Feature Similarity	Corresponding features must be similar (e.g. edges must have roughly the same length and orientation).
Structural Grouping	Corresponding feature groupings and their connectivity must be consistent.

CSE152, Spring 2012

(From G. Hager) Intro Computer Vision

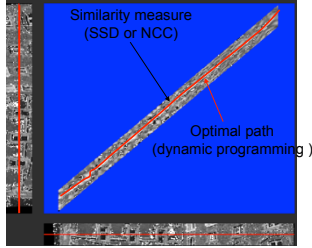
Stereo Matching using Dynamic Programming



(Ohta and Kanade, 1985)

Reprinted from "Stereo by Image and Intensity Search," by Y. Ohta and T. Kanade, IEEE Trans. on Pattern Analysis and Machine Intelligence, 9(2):139-154 (1985), © 1985 IEEE.
CSE152, Spring 2012 Intro Computer Vision

Stereo matching



Similarity measure (SSD or NCC)

Optimal path (dynamic programming)

Constraints

- epipolar
- ordering
- uniqueness
- disparity limit
- disparity gradient limit

Trade-off

- Matching cost (data)
- Discontinuities (prior)

(From Pollefeys)

(Cox et al. CVGIP'96; Koch'96; Falkenhagen '97; Van Meerbergen, Vergauwen, Pollefeys, VanGool IJCV'02)

CSE152, Spring 2012 Intro Computer Vision

Some Challenges & Problems

- Photometric issues:
 - specularities
 - strongly non-Lambertian BRDF's
- Surface structure
 - lack of texture
 - repeating texture within horopter bracket
- Geometric ambiguities
 - as surfaces turn away, difficult to get accurate reconstruction (affine approximate can help)
 - at the occluding contour, likelihood of good match but incorrect reconstruction

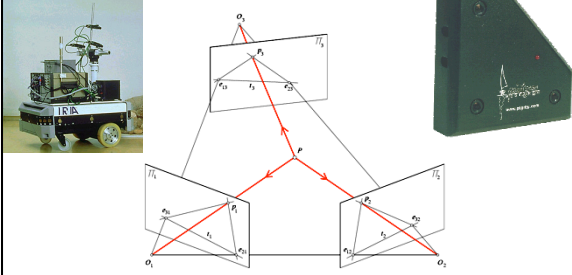
CSE152, Spring 2012 Intro Computer Vision

Variations on Binocular Stereo

1. Trinocular Stereopsis
2. Helmholtz Reciprocity Stereopsis

CSE152, Spring 2012 Intro Computer Vision

Trinocular Epipolar Constraints



These constraints are not independent!

$$\begin{cases} p_1^T \mathcal{E}_{12} p_2 = 0 \\ p_2^T \mathcal{E}_{23} p_3 = 0 \\ p_3^T \mathcal{E}_{31} p_1 = 0 \end{cases} \quad e_{31}^T \mathcal{E}_{12} e_{32} = e_{12}^T \mathcal{E}_{23} e_{13} = e_{23}^T \mathcal{E}_{31} e_{21} = 0$$

CSE152, Spring 2012 Intro Computer Vision

More on stereo ...

[The Middlebury Stereo Vision Research Page](http://cat.middlebury.edu/stereo/)
<http://cat.middlebury.edu/stereo/>

Recommended reading

D. Scharstein and R. Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. IJCV 47(1/2/3):7-42, April-June 2002. [PDF file](#) (1.15 MB) - includes current evaluation. Microsoft Research Technical Report MSR-TR-2001-81, November 2001.

Myron Z. Brown, Darius Burschka, and Gregory D. Hager. Advances in Computational Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8): 993-1008, 2003.

CSE152, Spring 2012 Intro Computer Vision