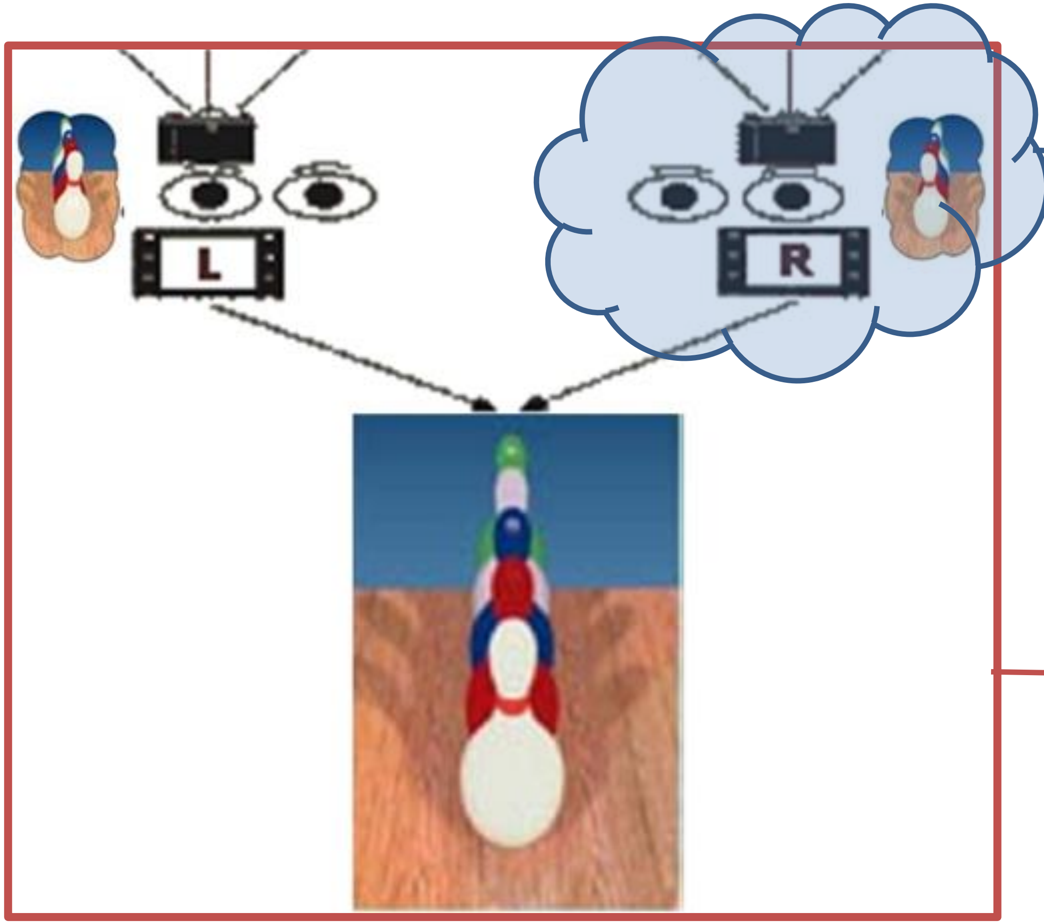
Stereo Vision
(Two-View Geometry)

Gang Pan
Zhejiang University
gpan@zju.edu.cn

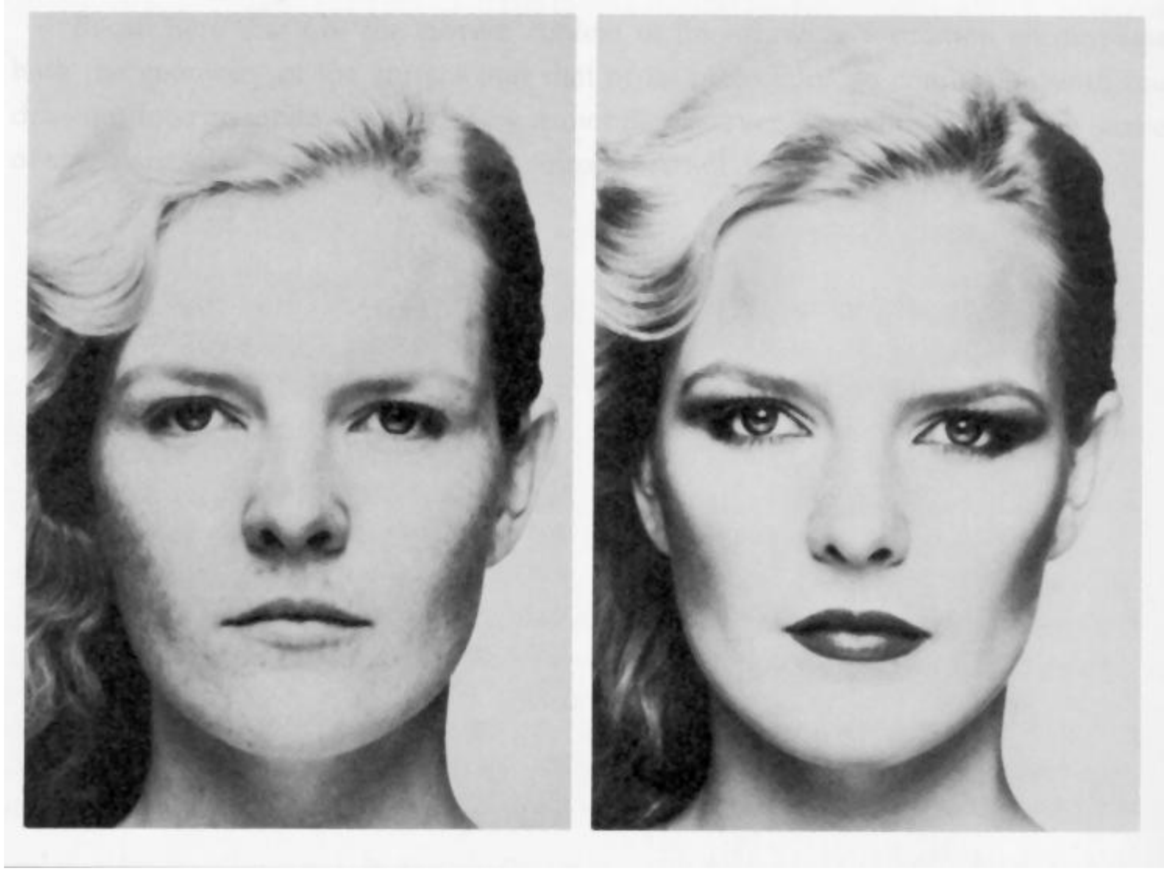


- **Single-view geometry**
- **Camera projection**

- **Two-view geometry**
- **Stereo**

Visual cues

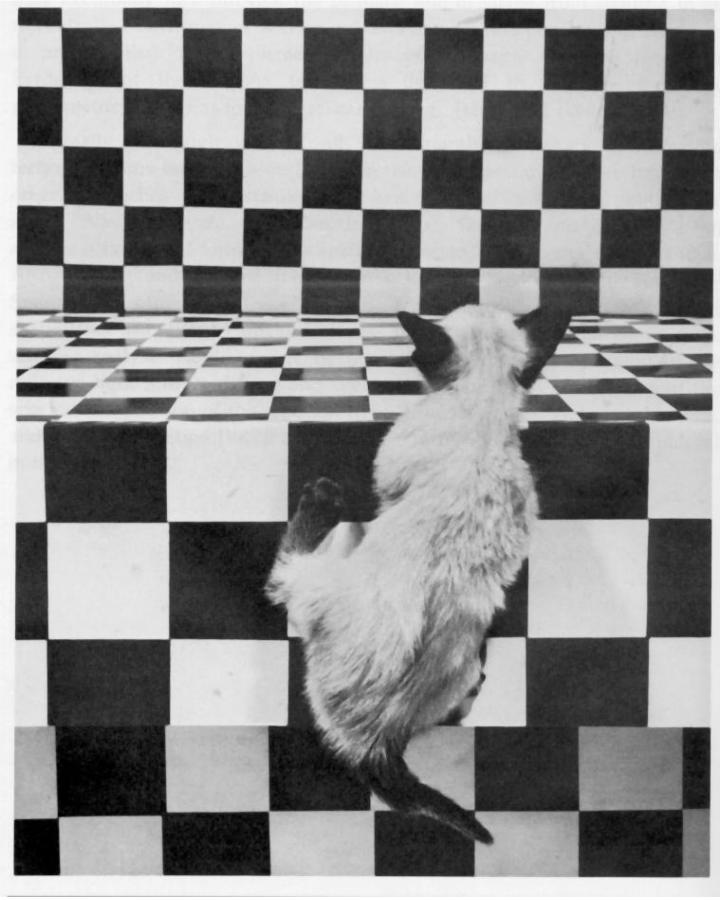
- Shading



Merle Norman Cosmetics, Los Angeles

Visual cues

- Shading
- Texture



The Visual Cliff, by William Vandivert, 1960

Visual cues

- Shading
- Texture
- Focus

From *The Art of Photography*, Canon

Visual cues

- Shading
- Texture
- Focus
- Motion

Visual cues

- Shading
- Texture
- Focus
- Motion

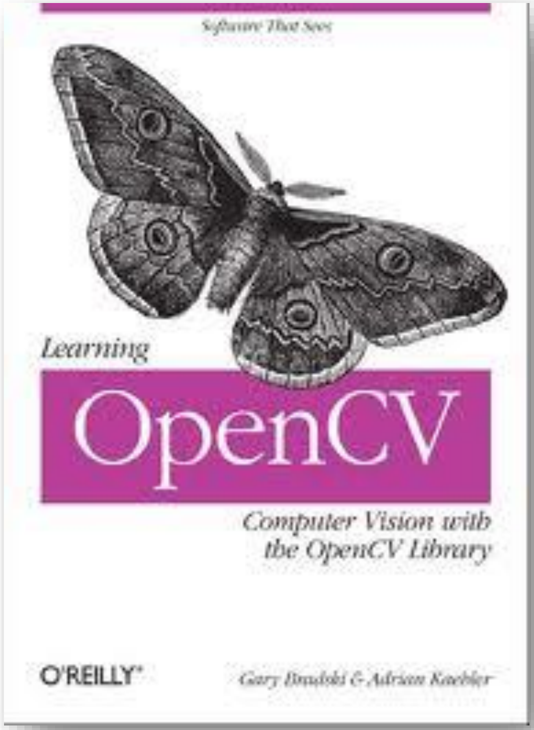
- Others:
  - Highlights
  - Shadows
  - Silhouettes
  - Inter-reflections
  - Symmetry
  - Light Polarization
  - ...

Shape From X

- X = shading, texture, focus, motion, ...
- In this class we'll focus on stereo: motion between two images

Outline

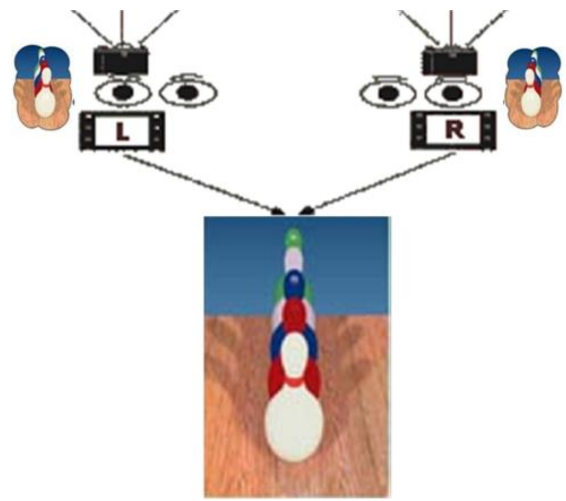
- Overview stereo imaging
- Techs in stereo imaging
- Applications



Chapter 12 Projection and 3D Vision

Outline

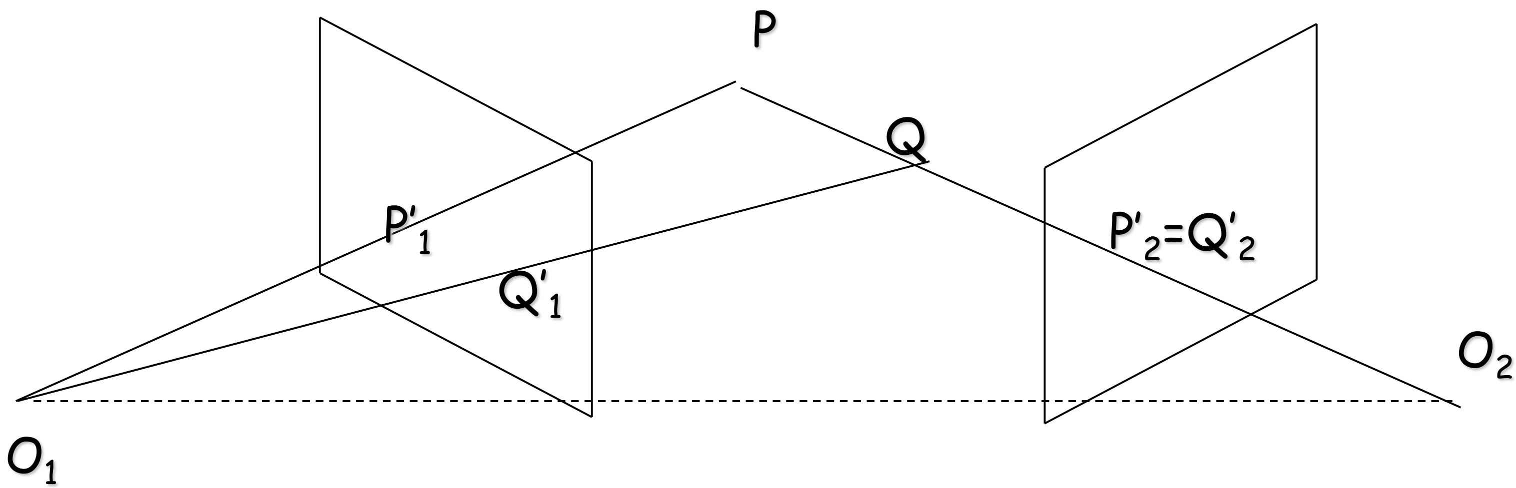
- Overview stereo imaging
  - What, why and how
  - Interesting stereo work
- Techs in stereo imaging
- Applications



What is Stereo?

- ‘Stereopsis’ – wiki
  - Stereo means solidity（实体感）, and opsis means vision or sight.
  - Is the process in visual perception leading to the sensation of **depth** from the **two slightly different projections** of the world onto the retinas of the two eyes.
  - For a computer, replace the eyes with two cameras.

Recovering Depth Information:

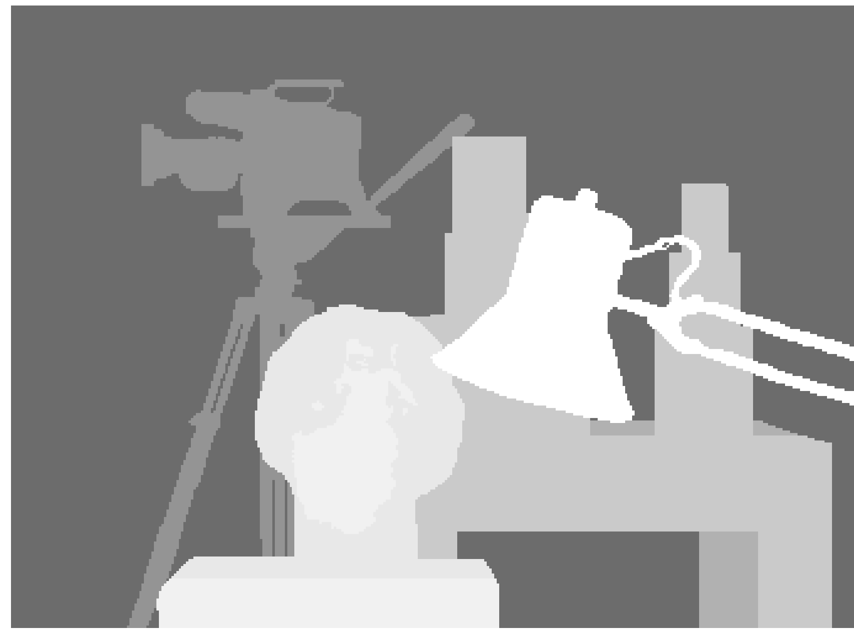


Depth can be recovered with two images and triangulation.

Stereo results

– Data from University of Tsukuba（筑波大学）

Scene



Ground truth

(Seitz)

http://www.well.com/~jimg/stereo/stereo_list.html

Grauman

Why?

- Important for applications in automation.
  - Mobile robot
  - Remote measuring
- Reconstructing sth is hot.
  - Augmented reality
- At least, we can have fun with it ☺

Real-time stereo

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

- Used for robot navigation (and other tasks)
  - Several software-based real-time stereo techniques have been developed (most based on simple discrete search)

Video view interpolation

Stereo photography and stereo viewers

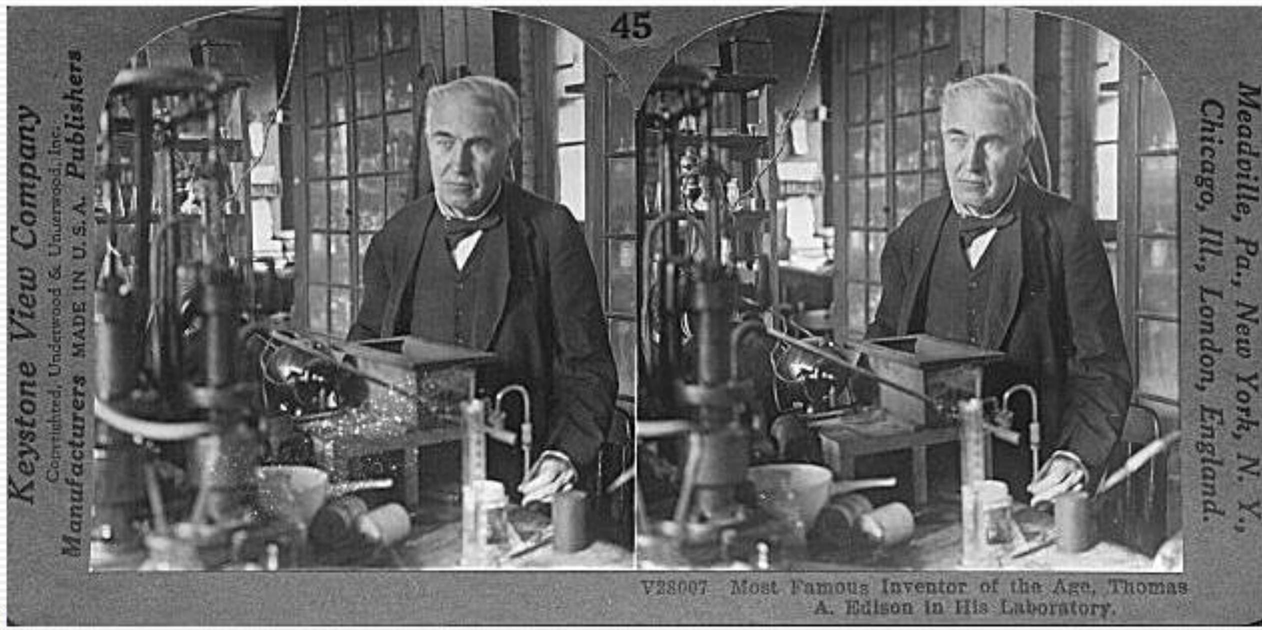
Take two pictures of the same subject from two slightly different viewpoints and display so that each eye sees only one of the images.

Invented by Sir Charles Wheatstone, 1838

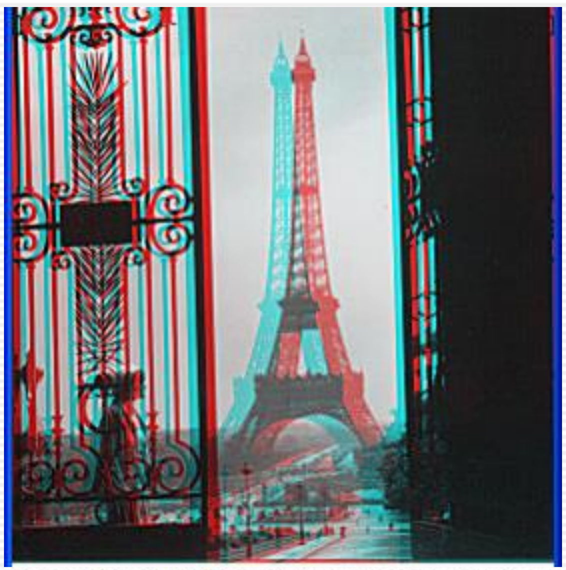
Image courtesy of fisher-price.com

Grauman

© Copyright 2001 Johnson-Shaw Stereoscopic Museum



http://www.johnsonshawmuseum.org



http://www.johnsonshawmuseum.org

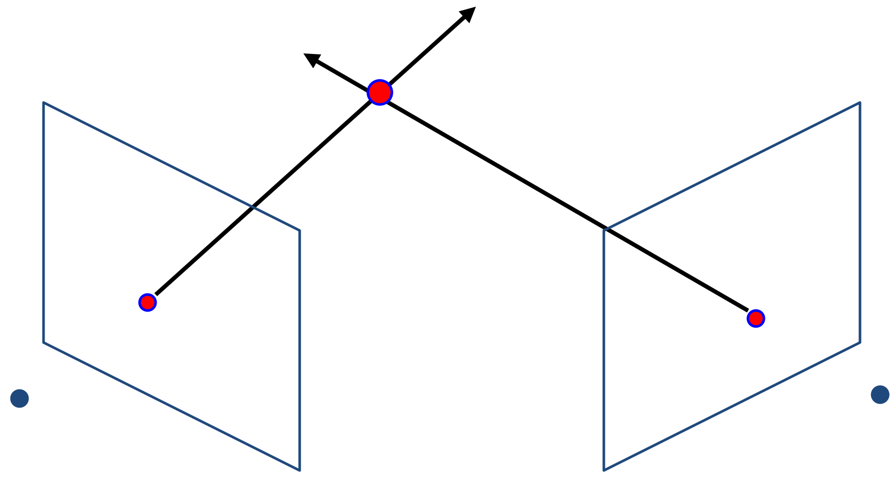
Public Library, Stereoscopic Looking Room, Chicago, by Phillips, 1923

Teesta suspension bridge-Darjeeling, India

Woman getting eye exam during immigration procedure at Ellis Island, c. 1905 - 1920 , UCR Museum of Phography

Mark Twain at Pool Table", no date, UCR Museum of Photography

Depth with stereo: basic idea



Basic Principle: Triangulation

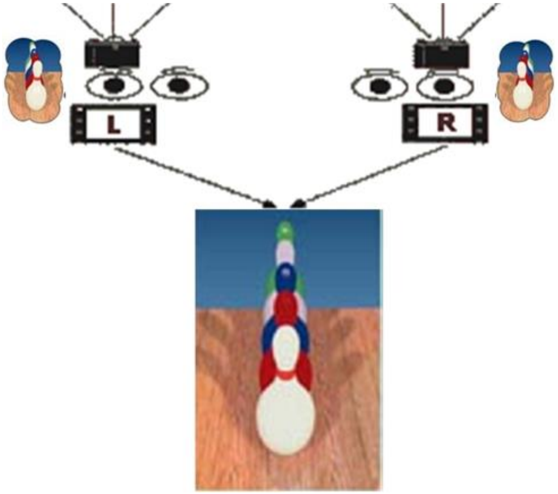
- Gives reconstruction as intersection of two rays
- Requires
  - camera pose (calibration)
  - ***point correspondence***

How to Do Stereo?

- Four major steps:
  - **Undistortion**
    - remove distortions -> undistorted images
  - **Rectification**
    - adjust cameras -> the two images row-aligned
  - **Correspondence**
    - find the same features in the two images -> disparity
  - **Reprojection**
    - triangulation -> a depth map

Outline

- Overview stereo imaging
- Techs in stereo imaging
  - Triangulation
  - Epipolar geometry
  - Stereo calibration
  - Stereo rectification
  - Matching correspondence
- Applications

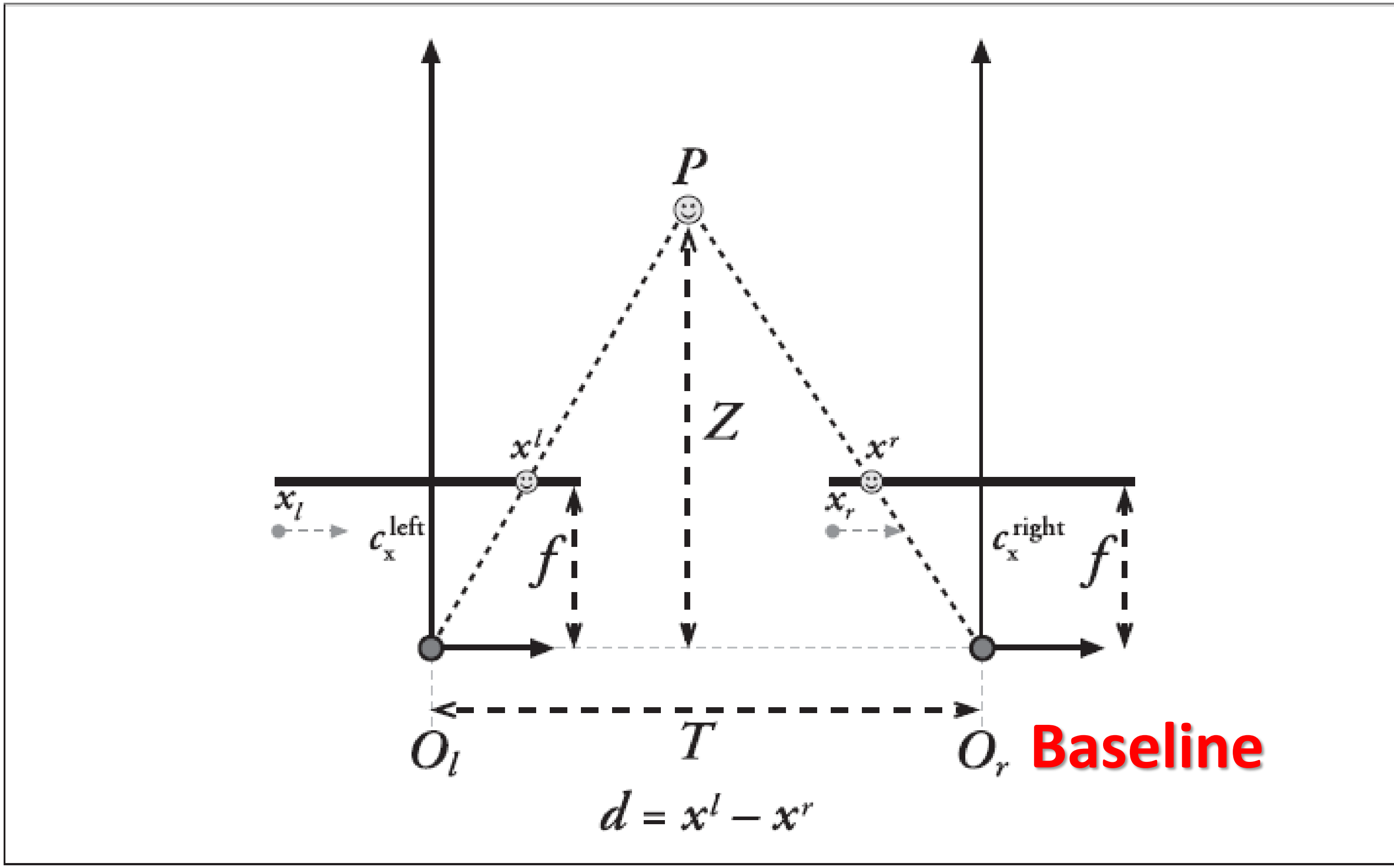


Triangulation

- **Assume** we have a stereo rig (two cameras with the same ***f*** in this case) that is perfectly
  - Undistorted
  - Aligned -- *frontal parallel*
  - Measured
- **Assume** we know the 2D-positions of a 3D-point in the two images.
- Similar triangles!

$$\frac{T-(x^l-x^r)}{Z-f}=\frac{T}{Z} \quad \Rightarrow \quad Z=\frac{fT}{x^l-x^r}$$

Triangulation



Disparity

Figure 12-4. With a perfectly undistorted, aligned stereo rig and known correspondence, the depth Z can be found by similar triangles; the principal rays of the imagers begin at the centers of projection O_l and O_r and extend through the principal points of the two image planes at c_l and c_r

Depth Resolution

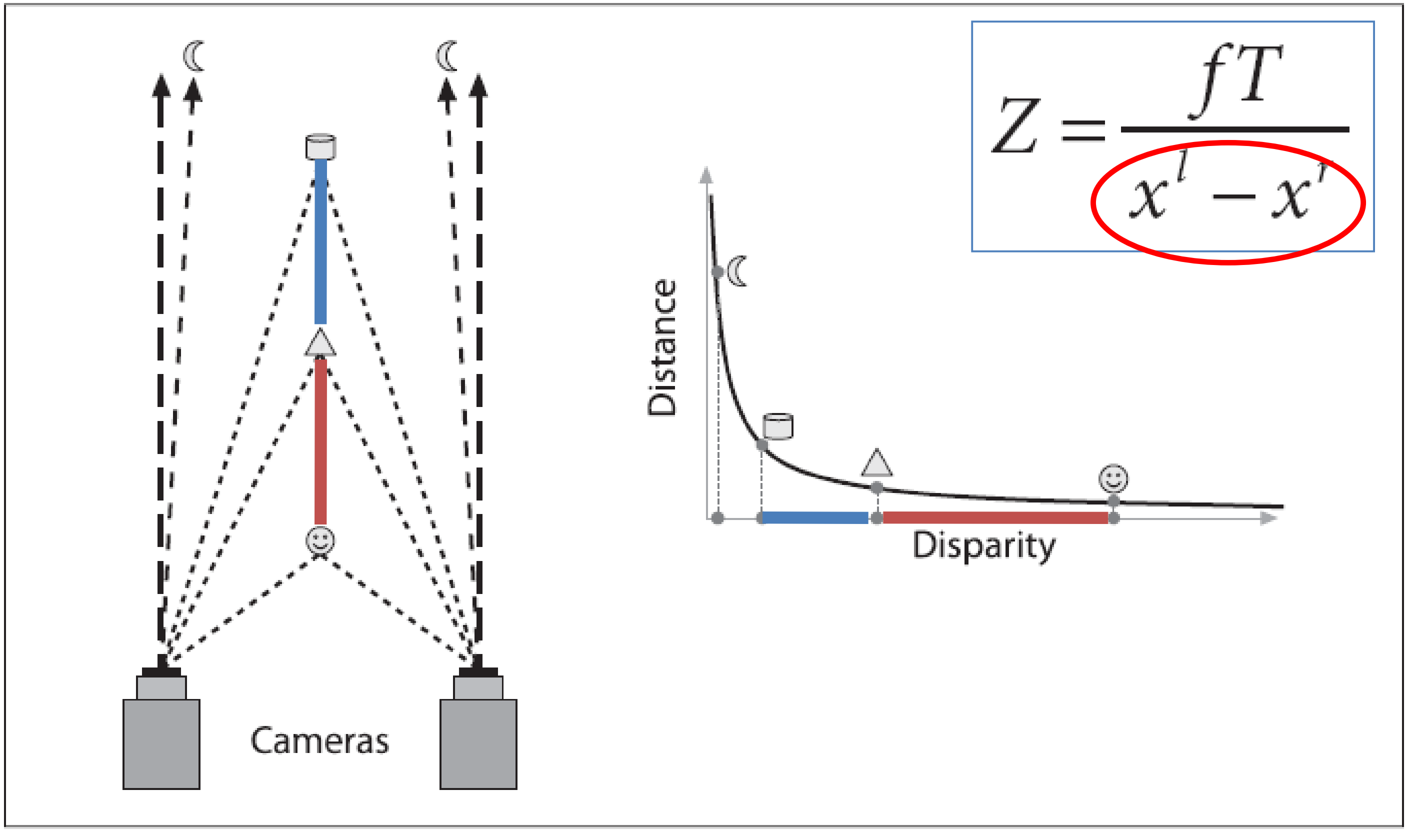
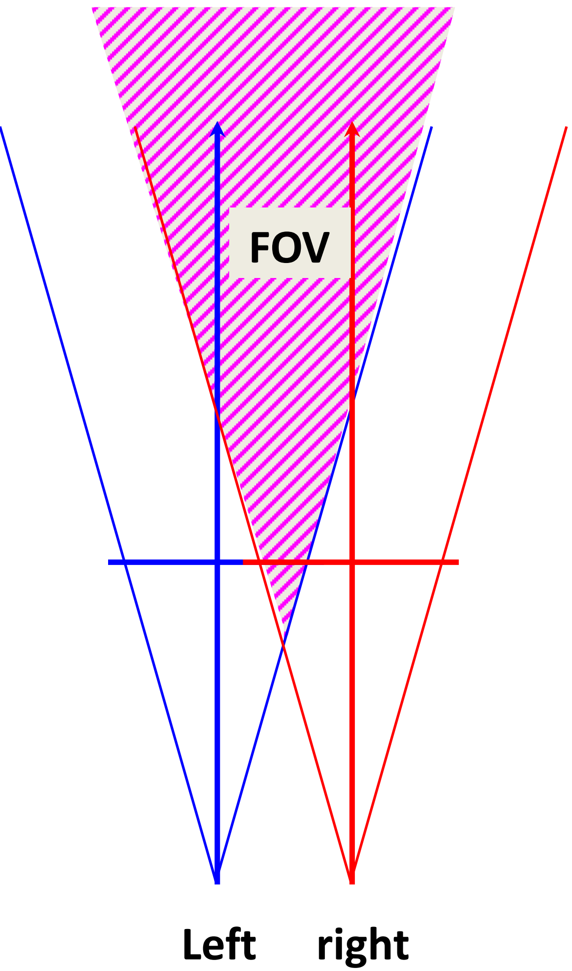


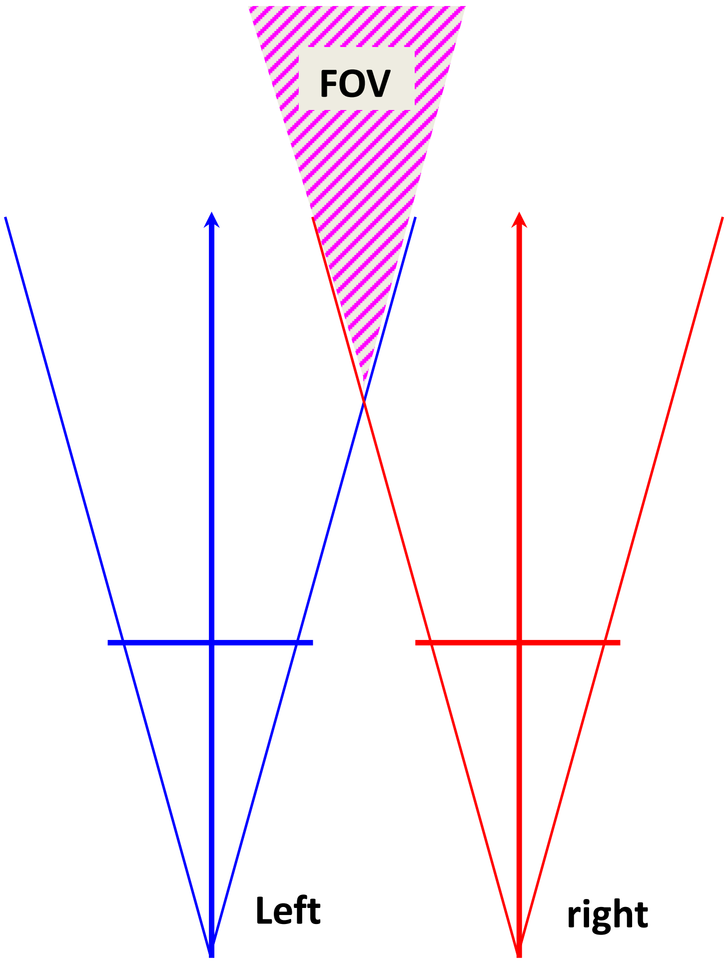
Figure 12-5. Depth and disparity are inversely related, so fine-grain depth measurement is restricted to nearby objects

Depth Resolution



$$Z = \frac{fT}{x^l - x^r}$$

- Short baseline
  - large common FOV
  - large depth error
- Long baseline
  - small depth error
  - small common FOV
  - More occlusion problems



Coordinates in OpenCV

Ideal

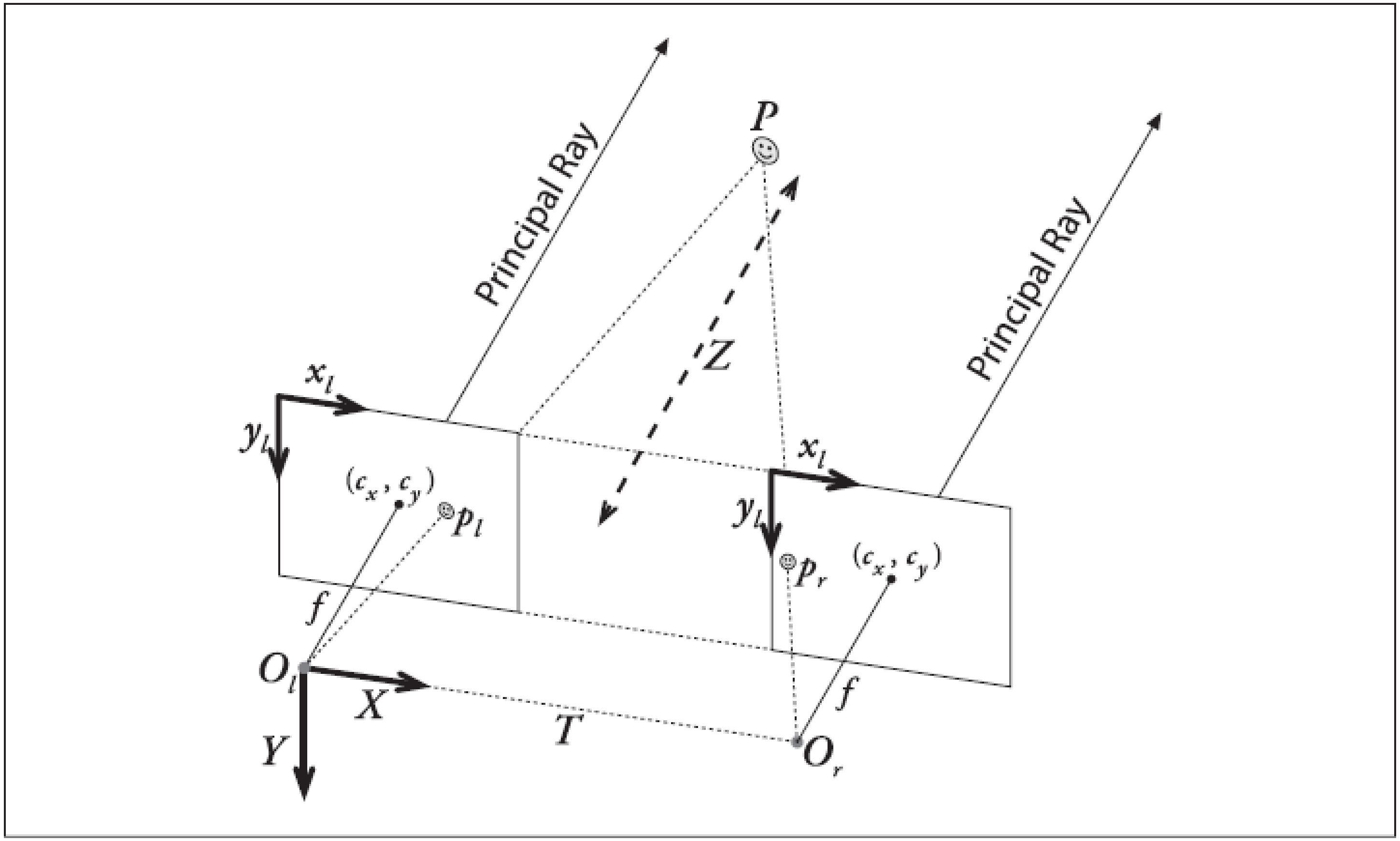


Figure 12-6. Stereo coordinate system used by OpenCV for undistorted rectified cameras: the pixel coordinates are relative to the upper left corner of the image, and the two planes are row-aligned; the camera coordinates are relative to the left camera's center of projection

Our Real World & **Rectification**

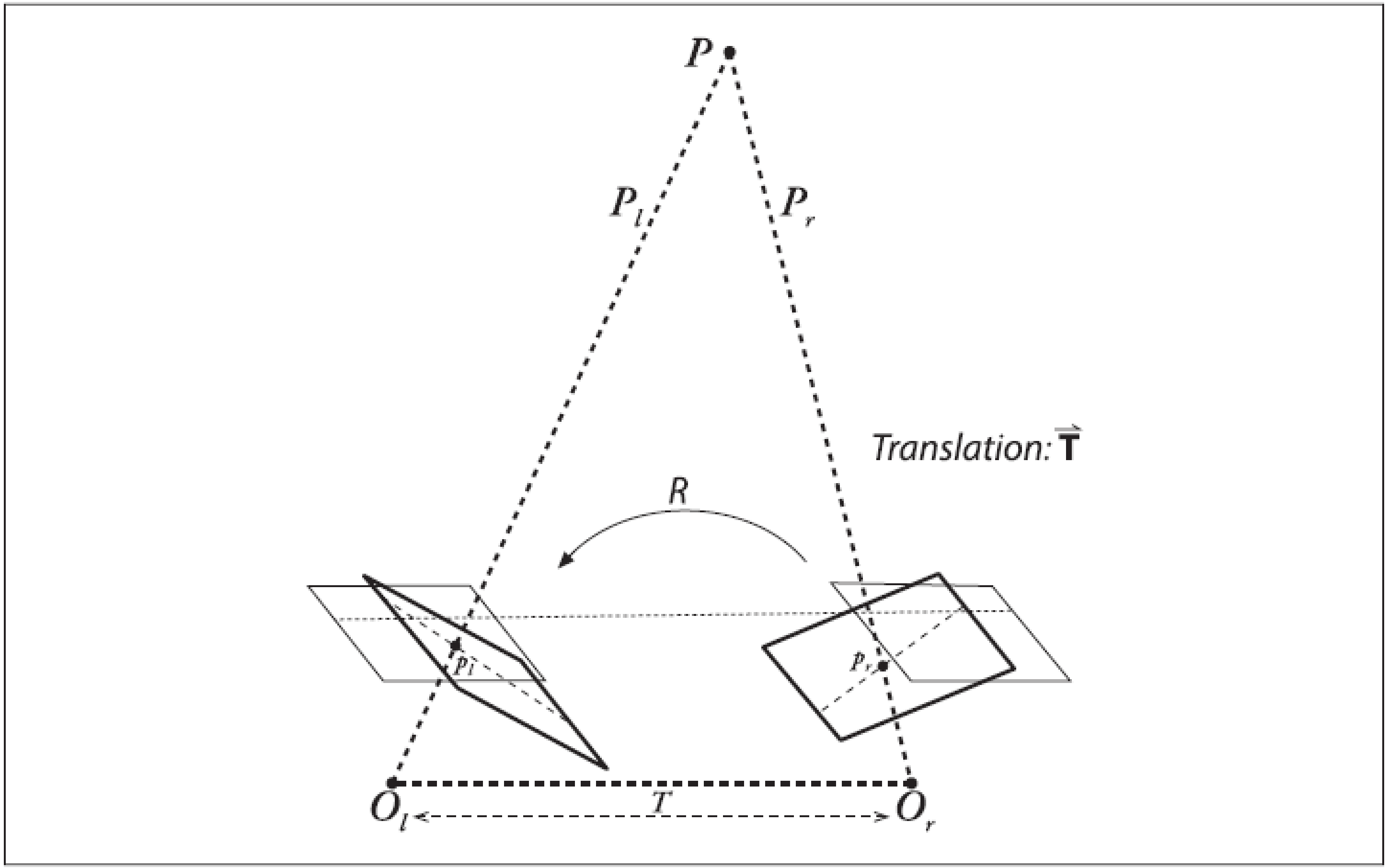


Figure 12-7. Our goal will be to mathematically (rather than physically) align the two cameras into one viewing plane so that pixel rows between the cameras are exactly aligned with each other

Rectification

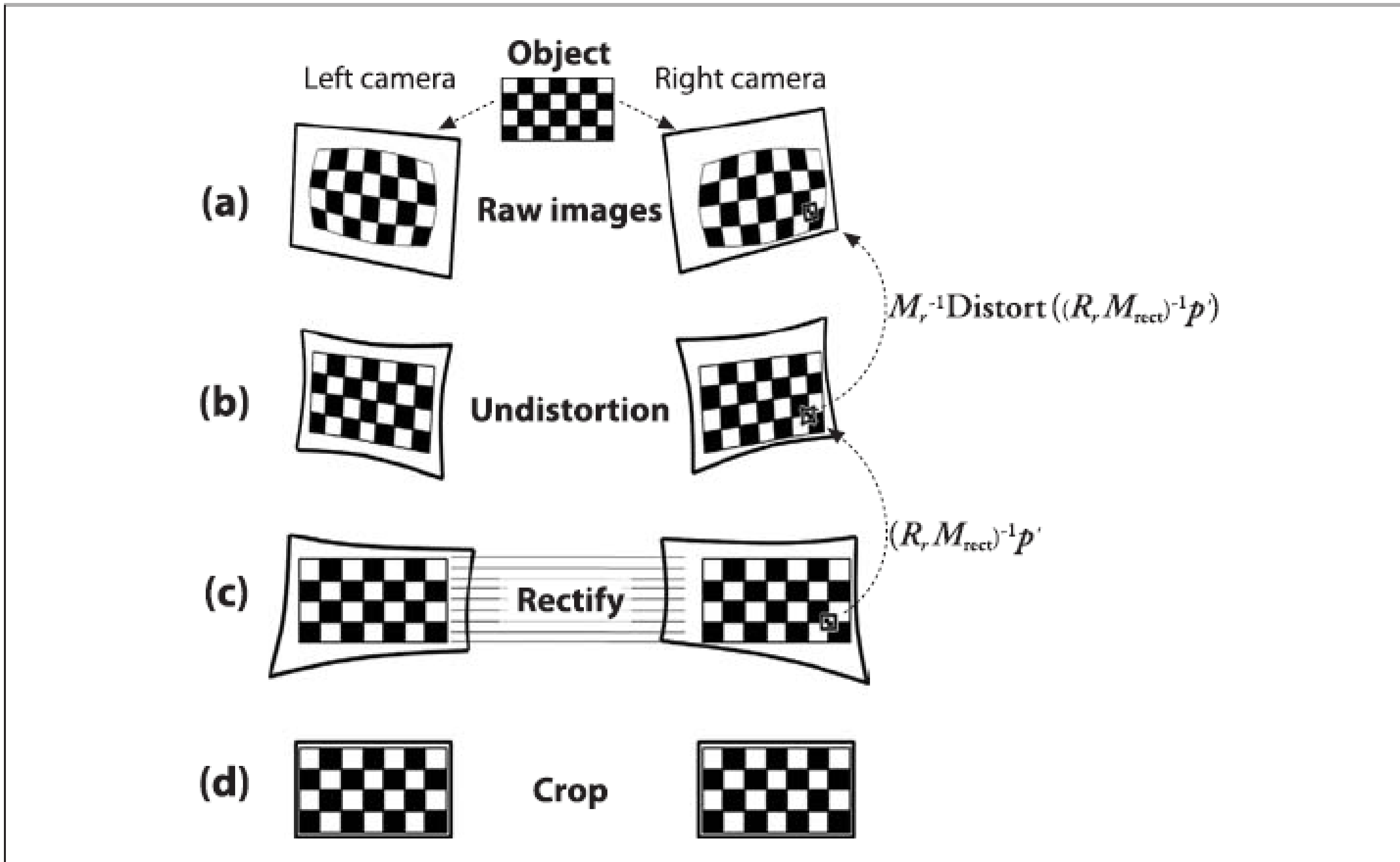
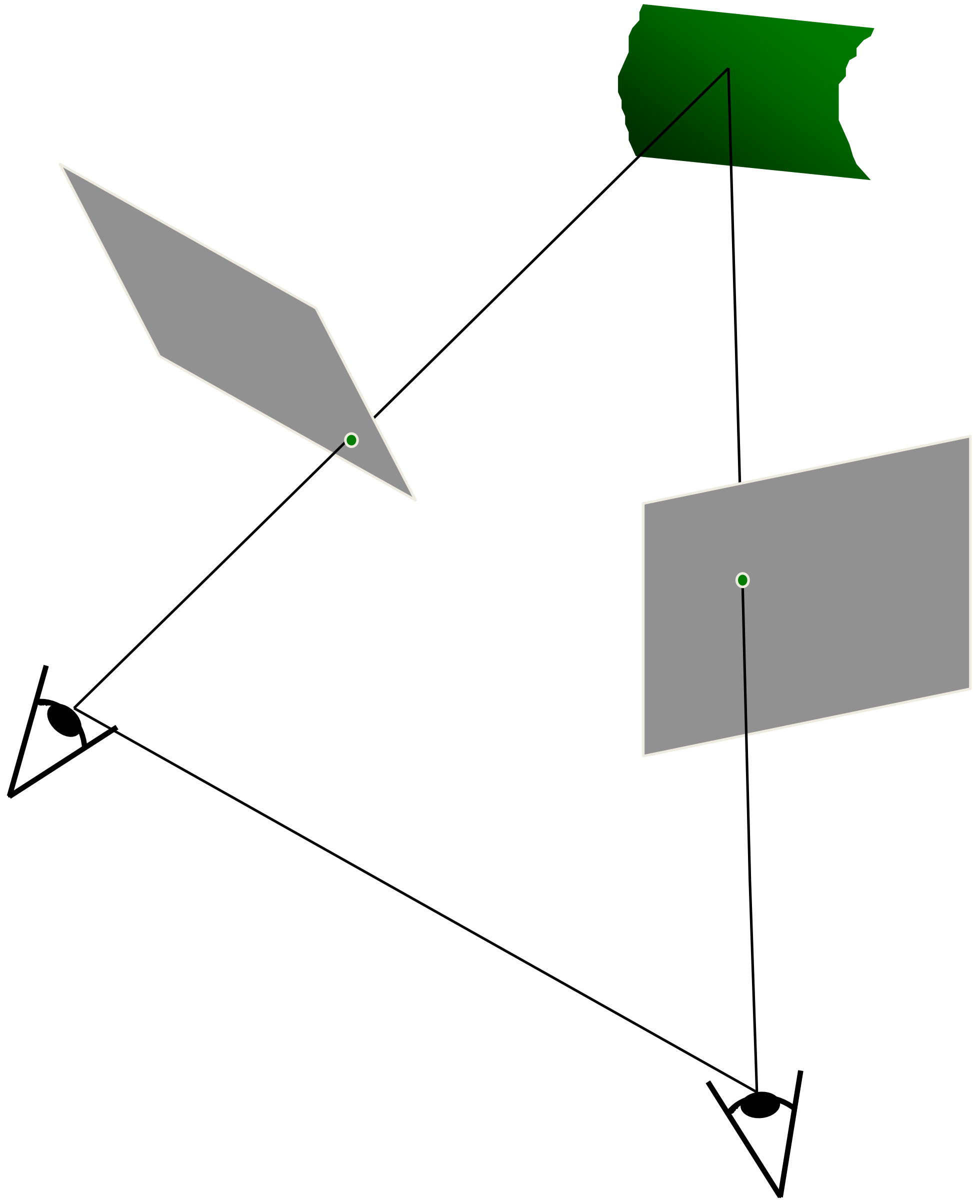
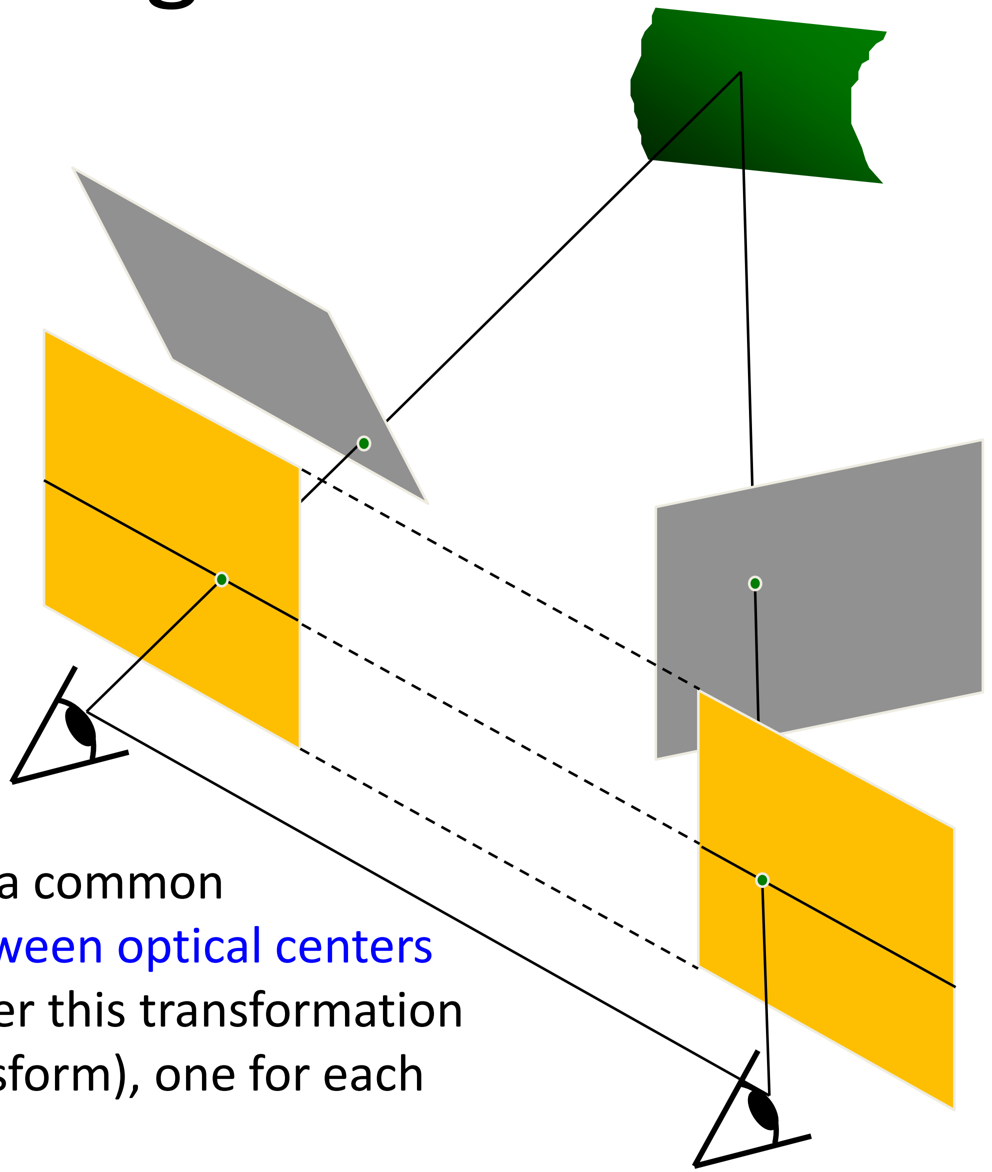


Figure 12-11. Stereo rectification: for the left and right camera, the raw image (a) is undistorted (b) and rectified (c) and finally cropped (d) to focus on overlapping areas between the two cameras; the rectification computation actually works backward from (c) to (a)

Stereo image rectification



Stereo image rectification



- reproject image planes onto a common plane parallel to the line between optical centers
- pixel motion is horizontal after this transformation
- two homographies (3x3 transform), one for each input image reprojection.

Epipolar Geometry

The basic geometry of a stereo imaging system

极几何(极线几何、核面几何、对极几何)

—— 表述两个相机成像关系的几何

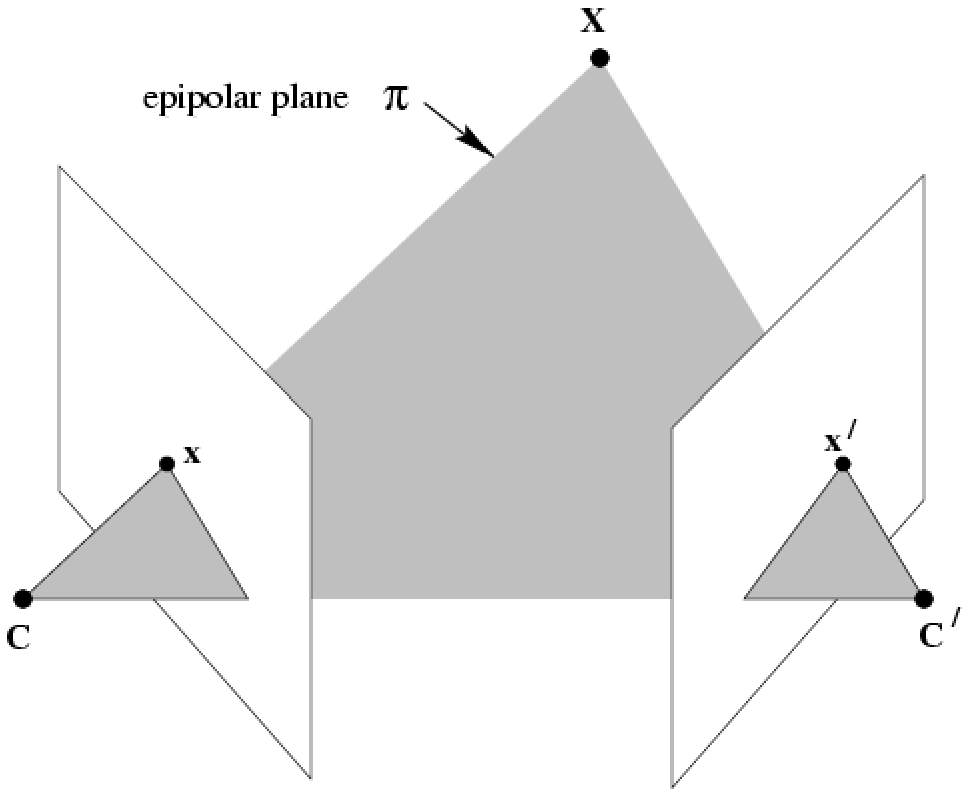
Three questions in advance:

(i) Correspondence geometry: Given an image point **x in the first view**, how does this constrain the position of the corresponding point **x' in the second image**?

(ii) Camera geometry (view): Given a set of corresponding image points $\{x_i \leftrightarrow x'_i\}$, i=1,…,n, what are the cameras P and P' for the two views? Or what is the geometric transformation between the views?

(iii) Scene geometry (structure): Given corresponding image points $x_i \leftrightarrow x'_i$ and cameras P, P', what is the position of the point X in space?

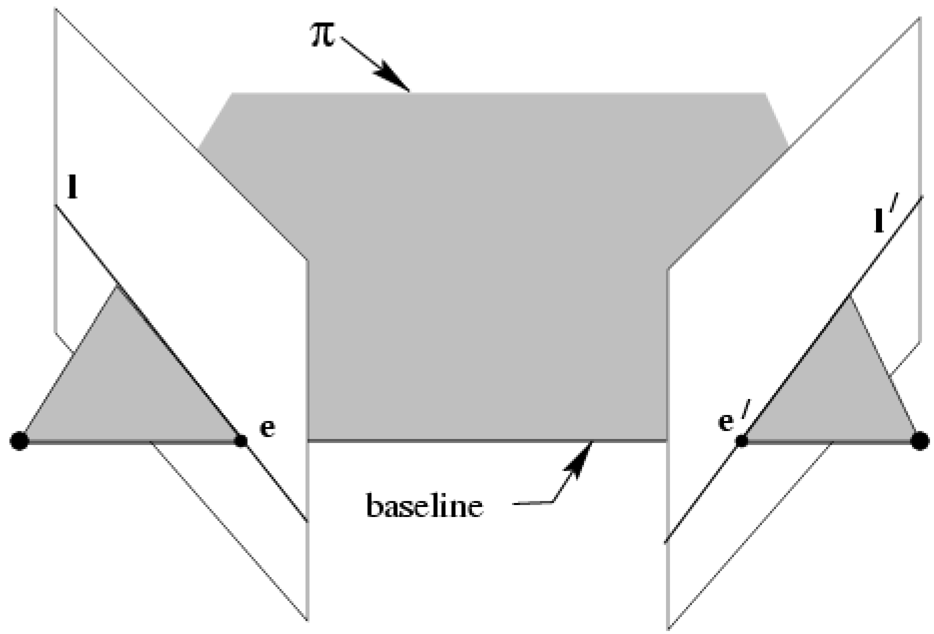
The epipolar geometry



C, C’, x, x’ , and X are coplanar

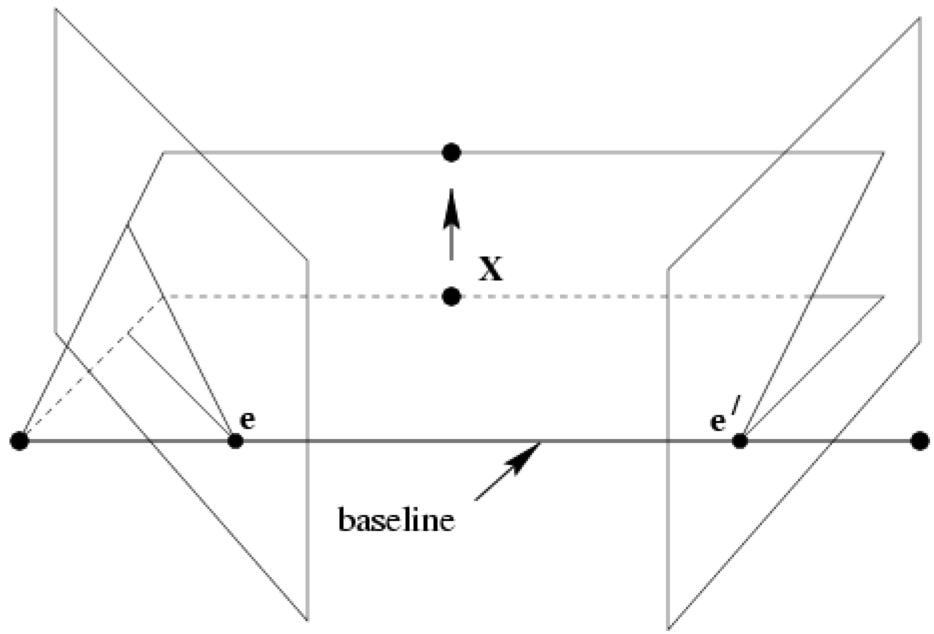
相机光心，物体像点和物体坐标点，五个点，三条线，共一面

The epipolar geometry



All points on π project on l and l'

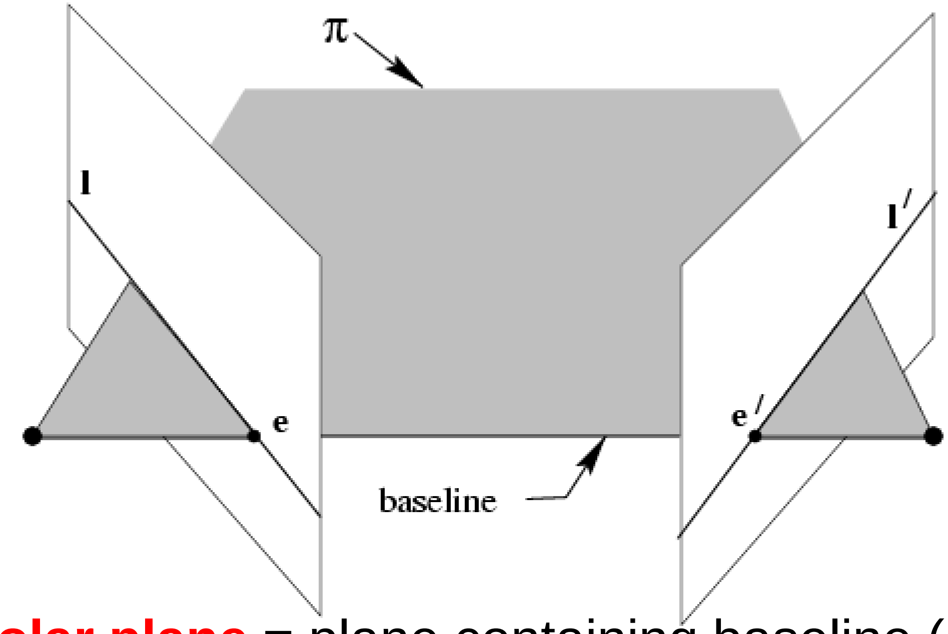
The epipolar geometry



Family of planes π and lines l and l' intersect in e and e'

The epipolar geometry: summay

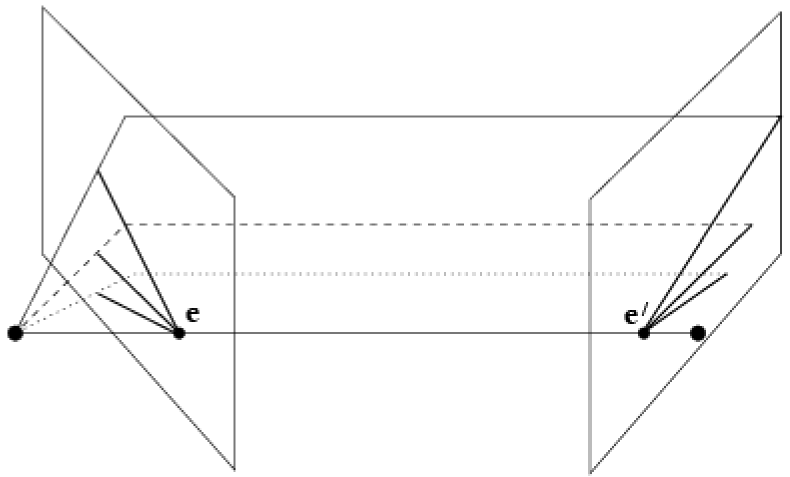
epipoles e,e'
= intersection of baseline with image plane
= projection of projection center in other image

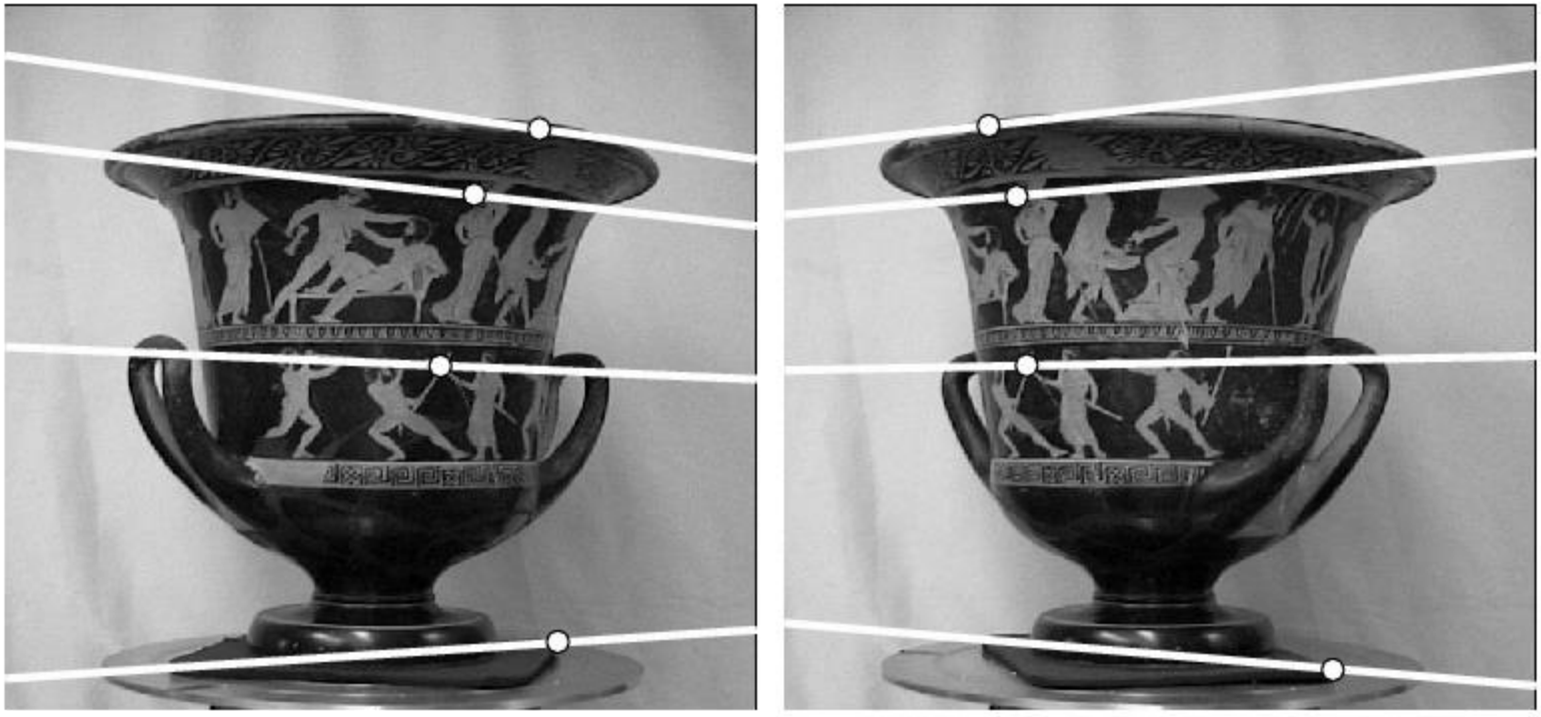


an **epipolar plane** = plane containing baseline (1-D family)

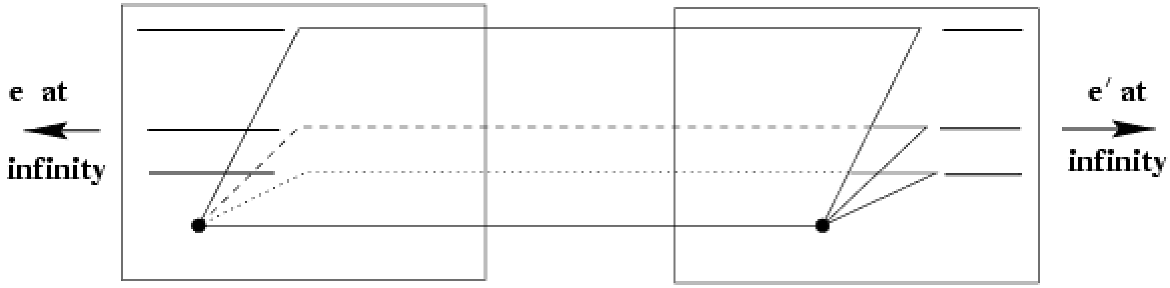
an **epipolar line** = intersection of epipolar plane with image
(always come in corresponding pairs)

Example: converging cameras





Example: motion parallel with image plane



Why Epipolar Geometry

- *Epipolar constraint*
  - Given a point in one image, its matching view in the other image ***must lie along*** the corresponding epipolar line

- Make corresponding easier
  - **2D search → 1D search**
  - Remove spurious correspondences, more robust

Transformation Between Cameras

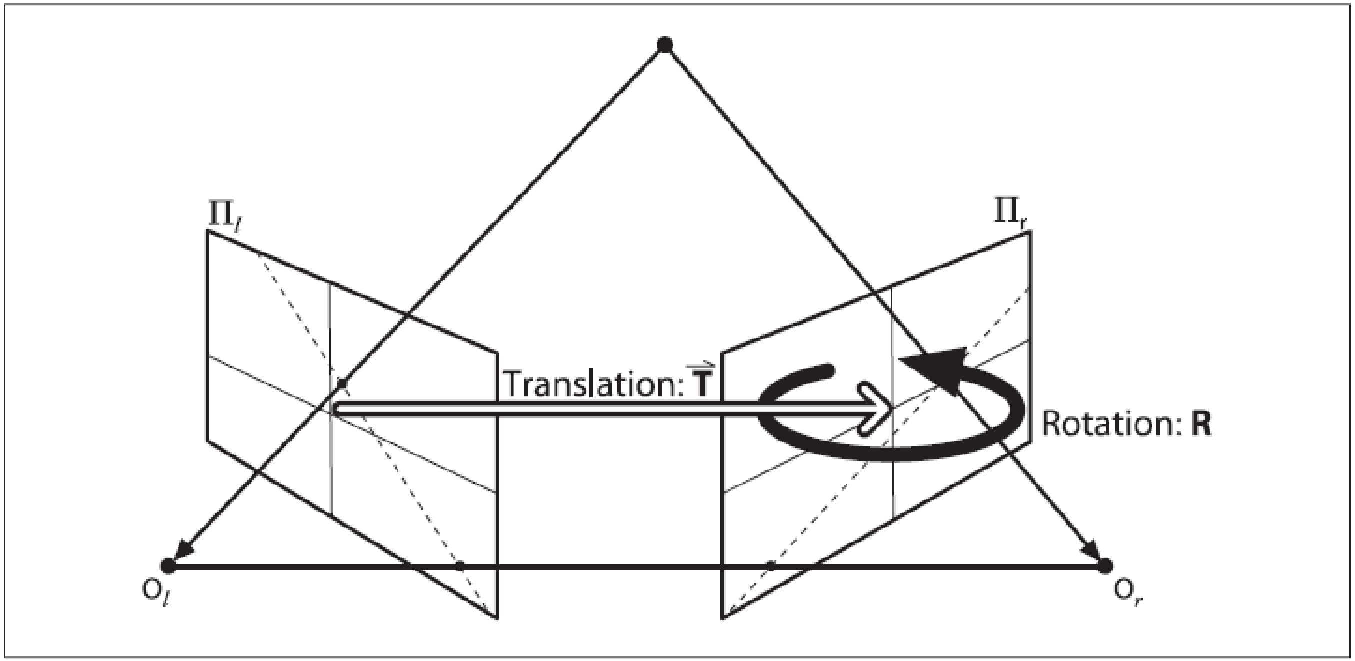


Figure 12-9. The essential geometry of stereo imaging is captured by the essential matrix E, which contains all of the information about the translation T and the rotation R, which describe the location of the second camera relative to the first in global coordinates

- **Essential matrix: *E*** $\quad (P_r)^{\mathrm{T}} E P_l = 0$
- **Fundamental matrix: *F*** $\quad q_r^{\mathrm{T}} F q_l = 0 \qquad F = (M_r^{-1})^{\mathrm{T}} E M_l^{-1}$

Find Fundamental Matrix with OpenCV

Value of method	Number of points	Algorithm
CV_FM_7POINT	$N = 7$	7-point algorithm
CV_FM_8POINT	$N \geq 8$	8-point algorithm
CV_FM_RANSAC	$N \geq 8$	RANSAC algorithm
CV_FM_LMEDS	$N \geq 8$	LMedS algorithm

```
int cvFindFundamentalMat(
  const CvMat* points1,
  const CvMat* points2,
  CvMat*       fundamental_matrix,
  int          method              = CV_FM_RANSAC,
  double       param1              = 1.0,
  double       param2              = 0.99,
  CvMat*       status              = NULL
);
```

Stereo Calibration

- The process of computing the geometrical relationship between the two cameras in space.
  - ***R*, *T***
- Make cameras *frontal parallel*
- Very similar to camera calibration
- Using epipolar constraints for verification

```
bool cvStereoCalibrate(
    const CvMat*    objectPoints,
    const CvMat*    imagePoints1,
    const CvMat*    imagePoints2,
    const CvMat*    npoints,
    CvMat*          cameraMatrix1,
    CvMat*          distCoeffs1,
    CvMat*          cameraMatrix2,
    CvMat*          distCoeffs2,
    CvSize          imageSize,
    CvMat*          R,
    CvMat*          T,
    CvMat*          E,
    CvMat*          F,
```

Stereo Rectification

- Make cameras *row-aligned*
- Many algorithms to use
  - **Hartley's algorithm**
    - Without stereo calibrations $q_r^{\mathrm{T}} F q_l = 0$
    - Good for online processing
  - **Bouguet's algorithm**
    - With calibration parameters: rotation and translation
    - More accurate

Stereo Rectification

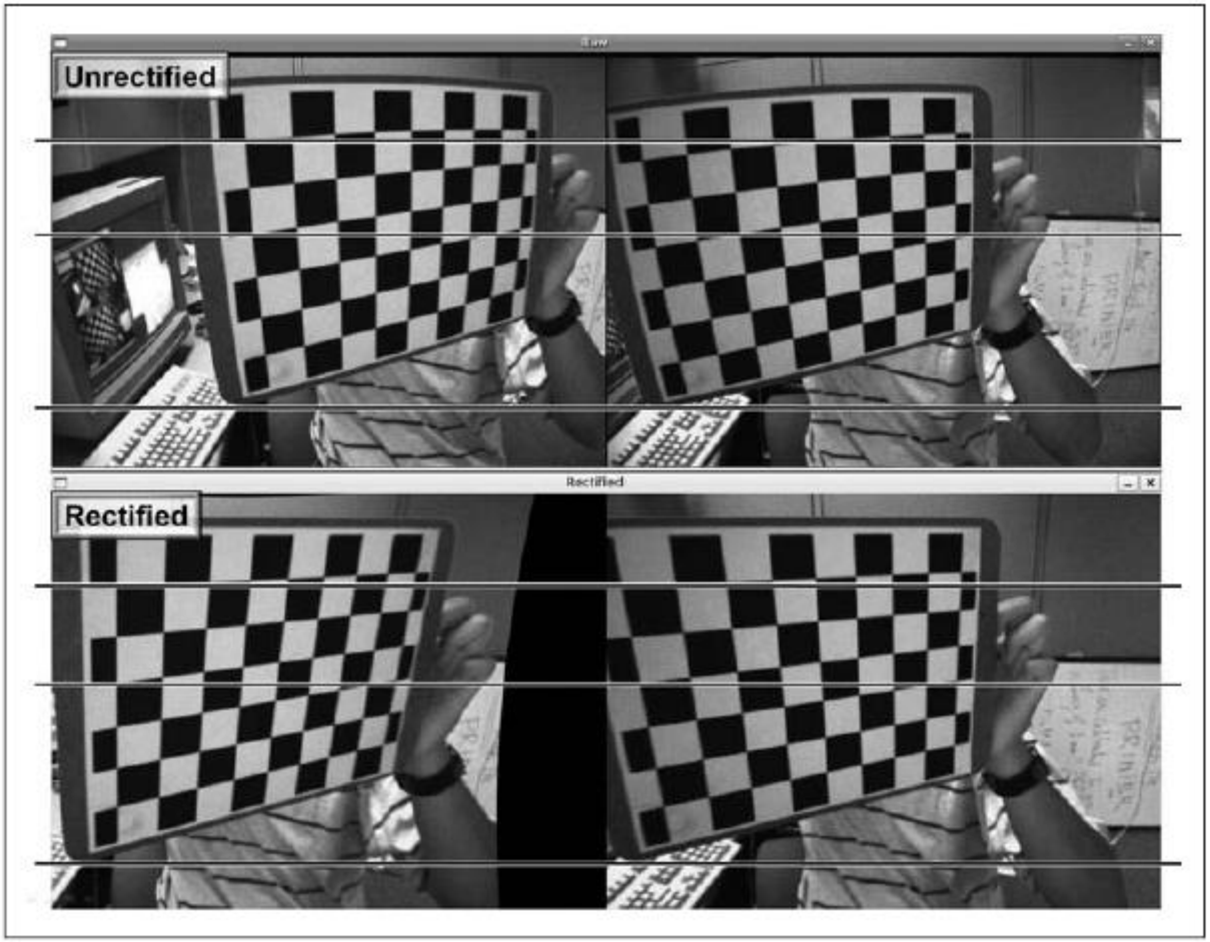
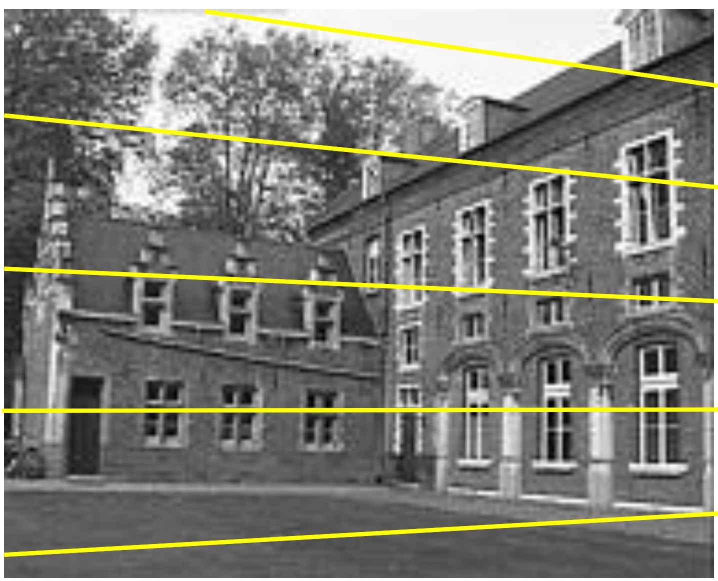
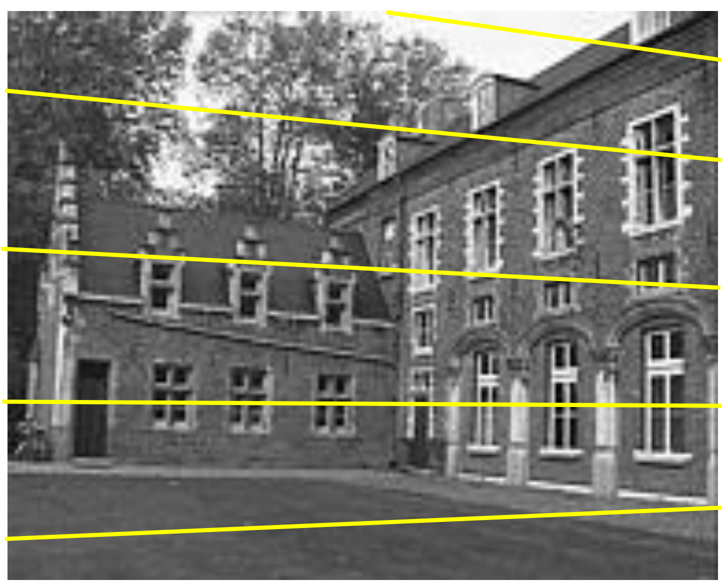


Figure 12-12. Stereo rectification: original left and right image pair (upper panels) and the stereo rectified left and right image pair (lower panels); note that the barrel distortion (in top of chessboard patterns) has been corrected and the scan lines are aligned in the rectified images

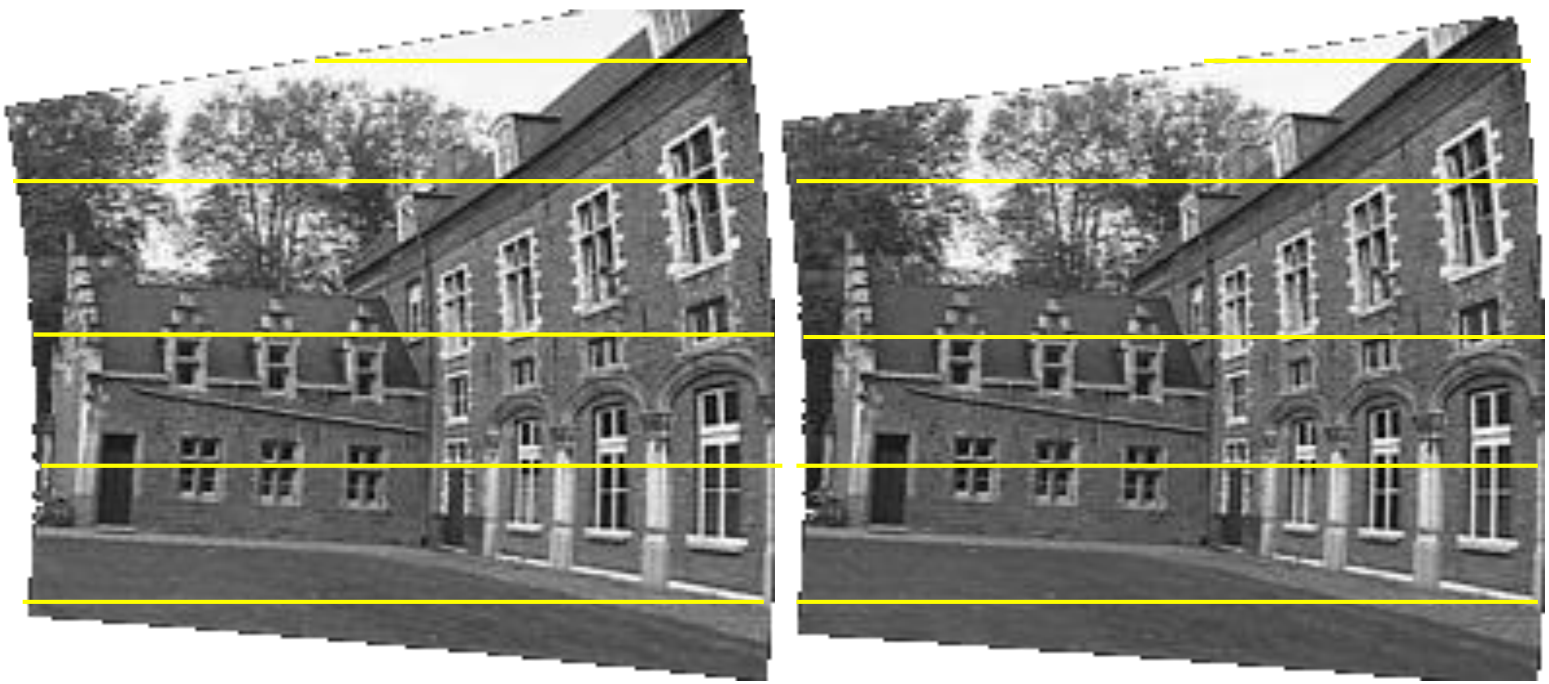
Stereo Rectification





courtesy of Marc Pollefeys

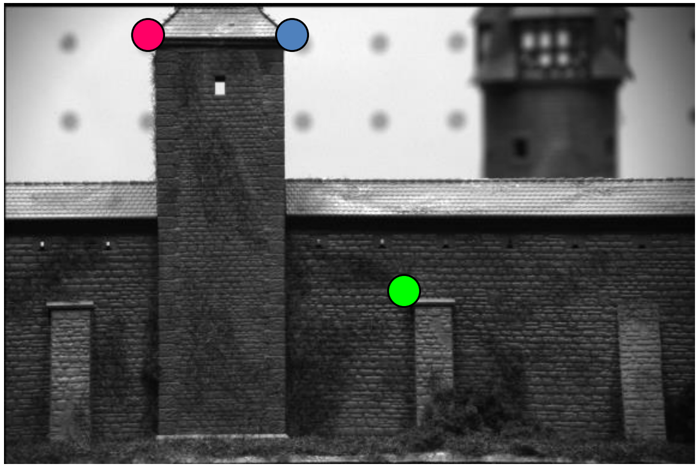
Stereo Rectification

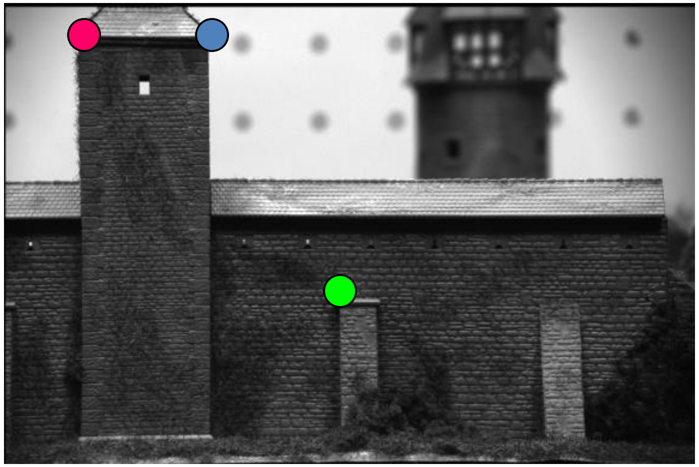


courtesy of Marc Pollefeys

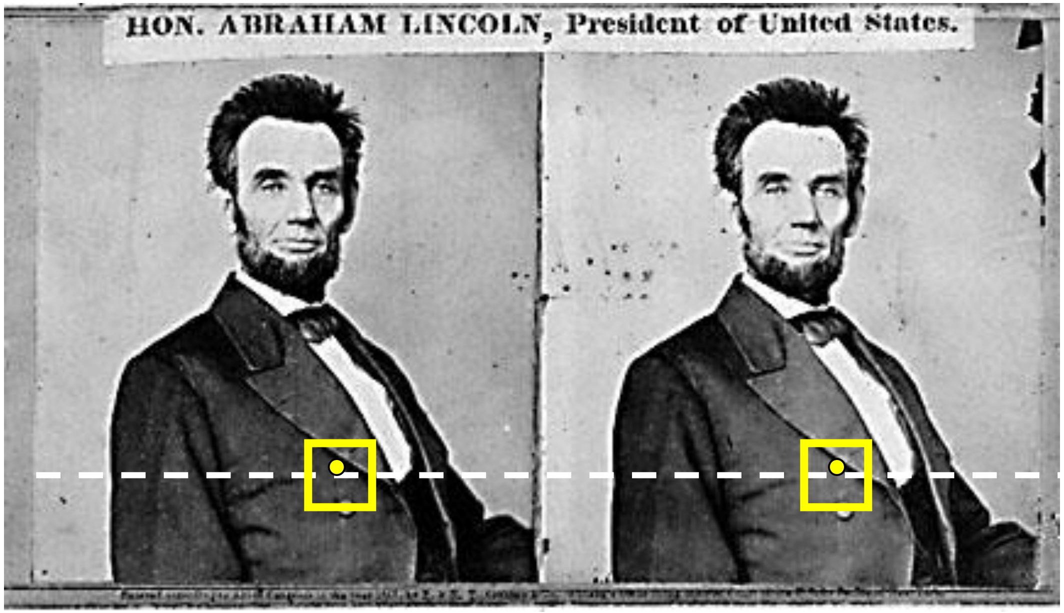
Correspondence: Stereo Matching

- The last step towards “**disparity**”
- Typical low-level vision problems
  - Finding features
  - Matching features





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:

Improvement: match ***windows***

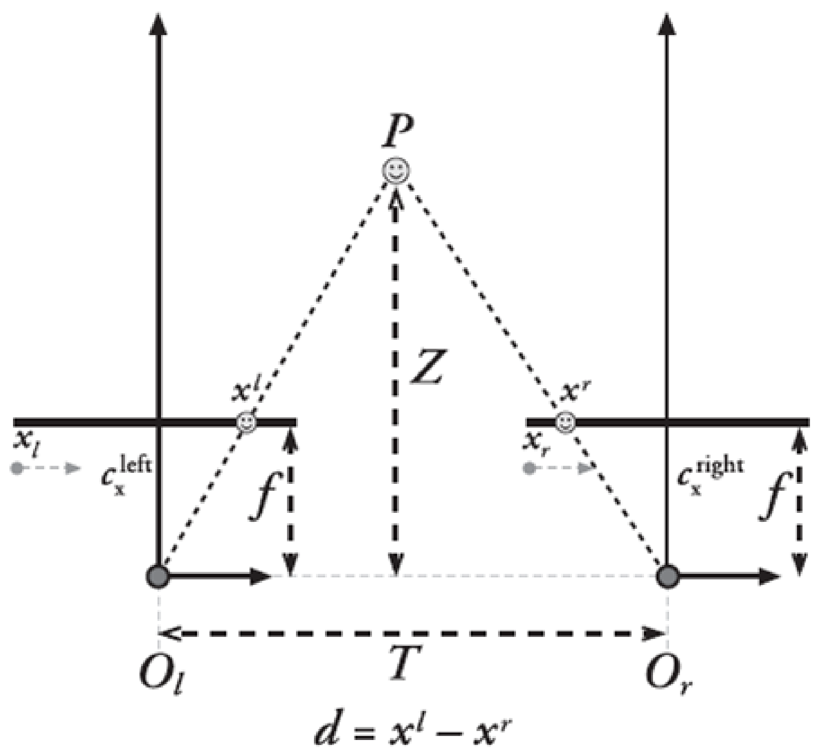
Correspondence Approach

- Local Matching Approach:
  - Block matching(块匹配，区域匹配)
  - Gradient-based optimization（基于剃度的优化）
  - Feature matching
- Global Matching Approach:
  - Dynamic programming（动态规划）
  - Intrinsic curves（本征曲线）
  - GC (Graph Cuts)（图分割）
  - Nonlinear diffusion（非线性扩散）
  - Belief propagation（置信度传播）
  - Correspondenceless methods（非对应性方法）

Back to Triangulation

- The image pair are already
  - Undistorted
  - Frontal parallel
  - Row-aligned
  - Corresponding

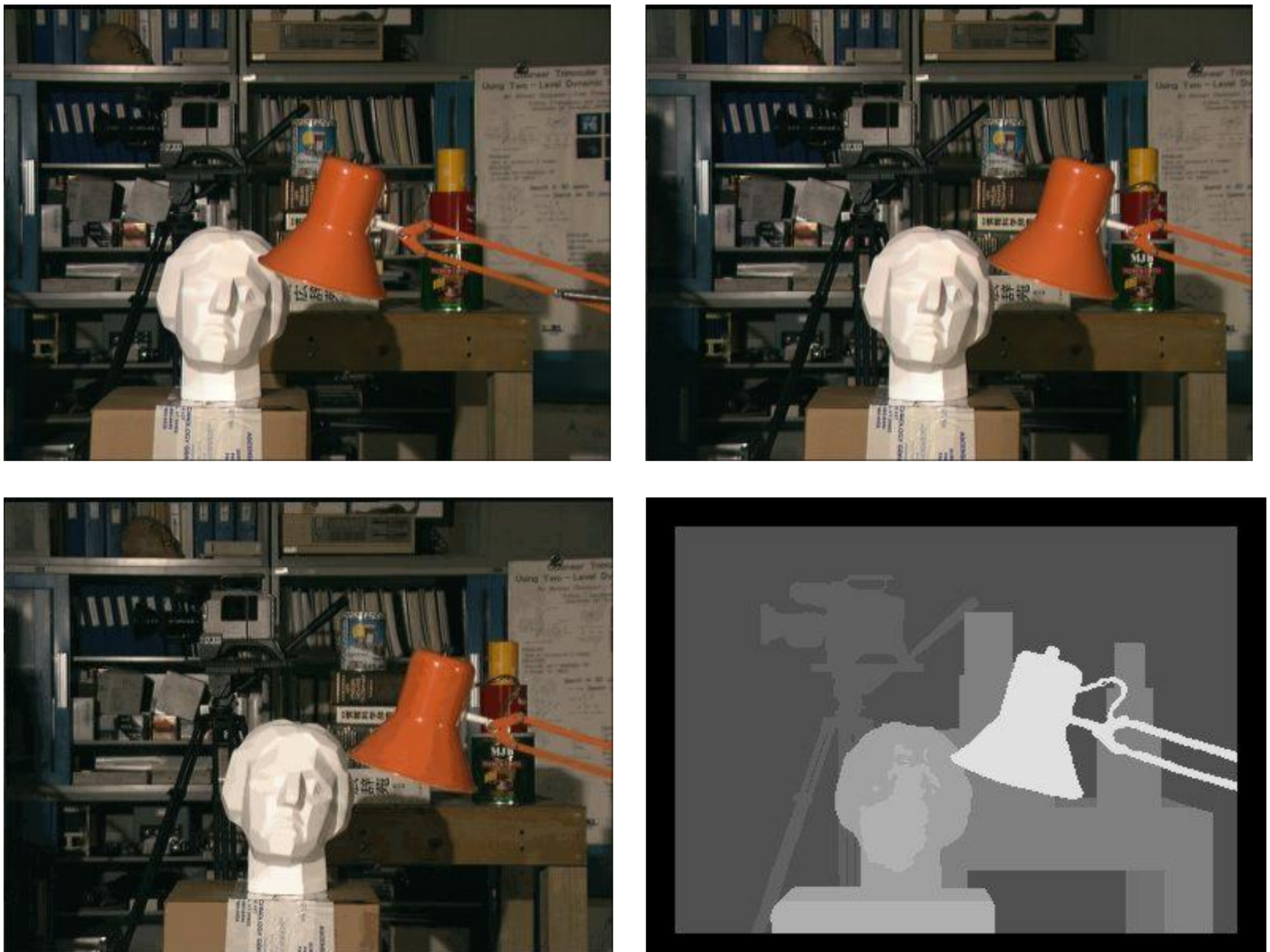
- **Disparity is there**



Evaluation

http://vision.middlebury.edu/stereo/

Original stereo pairs

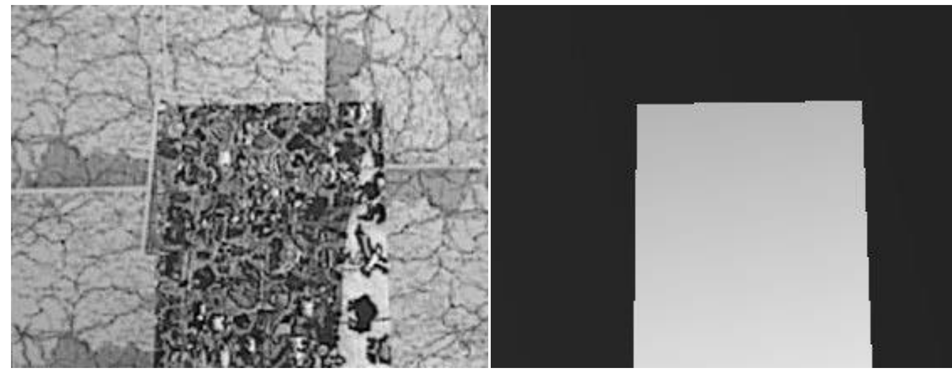


Evaluation

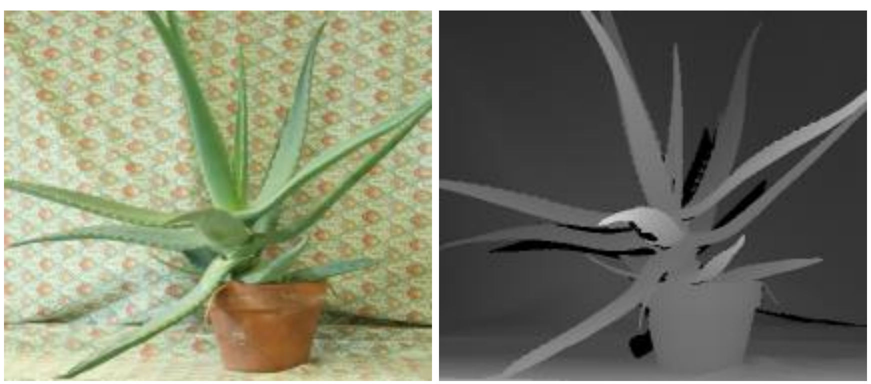
http://vision.middlebury.edu/stereo/

Venus

Map



Aloe



Art

Evaluation

http://vision.middlebury.edu/stereo/

vision.middlebury.edu/s

vision.middlebury.edu/stereo/eval3

Stereo | **Evaluation** • Datasets • Code • Submit

Middlebury Stereo Evaluation - Version 3

Mouseover the table cells to see the produced disparity map. Clicking a cell will blink the ground truth for comparison. To change the table type, click the links below. For more information, please see the description of new features.

Submit and evaluate your own results. See snapshots of previous results. See the evaluation v.2 (no longer active).

Set: test dense test sparse training dense training sparse
Metric: bad 0.5 bad 1.0 bad 2.0 bad 4.0 avgerr rms A50 A90 A95 A99 time time/MP time/GD
Mask: nonocc all

plot selected show invalid **Reset sort** **Reference list**

Date	bad 2.0 (%) Name	Res	Weight Avg	Austr MP: 5.6 nd: 290 im0 im1 GT nonocc	AustrP MP: 5.6 nd: 290 im0 im1 GT nonocc	Bicyc2 MP: 5.6 nd: 250 im0 im1 GT nonocc	Class MP: 5.7 nd: 610 im0 im1 GT nonocc	ClassE MP: 5.7 nd: 610 im0 im1 GT nonocc	Compu MP: 1.5 nd: 256 im0 im1 GT nonocc	Crusa MP: 5.5 nd: 800 im0 im1 GT nonocc	CrusaP MP: 5.5 nd: 800 im0 im1 GT nonocc	Djemb MP: 5.7 nd: 320 im0 im1 GT nonocc	DjembL MP: 5.7 nd: 320 im0 im1 GT nonocc	Hoops MP: 5.7 nd: 410 im0 im1 GT nonocc	Livgrm MP: 5.9 nd: 320 im0 im1 GT nonocc	Nkuba MP: 5.5 nd: 570 im0 im1 GT nonocc	Plants MP: 5.6 nd: 320 im0 im1 GT nonocc	Stairs MP: 5.2 nd: 450 im0 im1 GT nonocc
05/12/16	PMSC	H	**6.87** 1	**3.46** 1	**2.68** 1	6.19 7	**2.54** 1	**6.92** 1	**6.54** 1	**3.96** 1	4.04 2	**2.37** 1	13.1 4	12.3 2	12.2 3	16.2 10	5.88 3	10.8 5
10/19/16	LW-CNN	H	7.23 2	4.65 3	3.95 4	5.30 3	2.63 2	11.2 9	7.86 2	4.32 3	4.22 3	2.43 3	12.2 2	13.4 4	13.6 8	14.8 8	**4.72** 1	12.0 9
04/12/16	MeshStereoExt	H	7.29 3	4.41 2	3.98 5	5.40 4	3.17 5	10.0 4	8.89 6	4.62 4	4.77 5	3.49 11	12.7 3	12.4 3	10.4 2	14.5 6	7.80 8	8.85 2
05/28/16	APAP-Stereo	H	7.46 4	5.43 4	4.91 16	5.11 2	5.17 10	21.6 16	9.50 8	4.31 2	4.23 4	3.24 10	14.3 7	**9.78** 1	**7.32** 1	13.4 3	6.30 4	**8.46** 1
01/19/16	NTDE	H	7.62 5	5.72 8	4.36 8	5.92 5	2.83 3	10.4 5	8.02 3	5.30 5	5.54 6	2.40 2	13.5 5	14.1 6	12.6 5	13.9 5	6.39 5	12.2 10
08/28/15	MC-CNN-acrt	H	8.29 6	5.59 7	4.55 11	5.96 6	2.83 3	11.4 10	8.44 5	8.32 9	8.89 13	2.71 4	16.3 9	14.1 7	13.2 7	13.0 2	6.40 6	11.1 7
11/03/15	MC-CNN+RBS	H	8.62 7	6.05 10	5.16 20	6.24 8	3.27 6	11.1 8	8.91 7	8.87 11	9.83 18	3.21 9	15.1 8	15.9 9	12.8 6	13.5 4	7.04 7	9.99 3
09/13/16	SNP-RSM	H	8.98 8	5.46 5	4.85 14	6.50 10	3.37 7	10.4 6	10.1 11	8.73 10	9.37 16	3.58 12	14.3 6	14.7 8	14.9 9	**12.8** 1	10.1 13	10.8 6
01/21/16	MCCNN_Layout	H	9.16 9	5.53 6	5.63 23	**5.06** 1	3.59 8	12.6 11	9.97 10	7.53 8	8.86 12	5.79 24	23.0 13	13.6 5	15.0 10	14.7 7	5.85 2	10.4 4
01/26/16	MC-CNN-fst	H	9.69 10	7.35 13	5.07 19	7.18 12	4.71 9	16.8 14	11.2 14	7.37 7	6.97 7	2.82 5	20.7 12	17.4 12	15.4 12	15.1 9	7.90 9	12.6 11
07/03/16	LPU	H	10.5 11	11.4 15	3.18 2	8.10 15	6.08 12	20.9 15	9.84 9	6.94 6	**4.00** 1	4.04 14	33.9 27	16.9 10	15.2 11	17.8 15	9.12 10	11.6 8

Stereo reconstruction pipeline

- **Steps**
  - Calibrate cameras (+ Undistortion)
  - Rectify images
  - Compute disparity
  - Estimate depth

- **What will cause errors?**
  - Camera calibration errors
  - Poor image resolution
  - Occlusions
  - Violations of brightness constancy (specular reflections)
  - Large motions
  - Low-contrast image regions