Image Recording and Calibration[1]

(80 min lecture)

See Material in
Reinhard Klette: Concise Computer Vision
Springer-Verlag, London, 2014

ccv.wordpress.fos.auckland.ac.nz

[1]See last slide for copyright information.

Cameras in a Test Vehicle

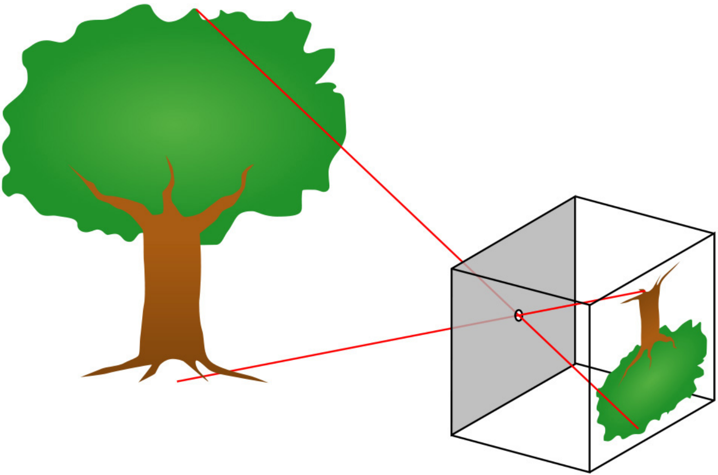
Examples of mounted cameras inside or on top of the vehicle

Agenda

1. Image Recording
2. Camera Calibration
3. Stereo Vision
4. Stereo Matching (Brief Intro)

Pinhole Camera

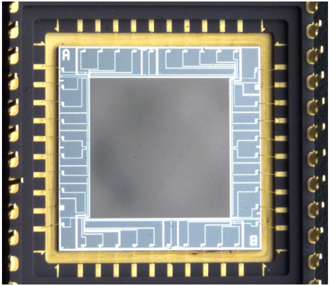
Light rays pass through the pinhole and create a top-down projection



[Image by Pbroks13 in the public domain]

Recording Today: Matrix Sensors

Digital camera uses one or several *matrix sensors* for recording an image



Edges of individual sensor cells (phototransistors) are 1.4μm to 20μm

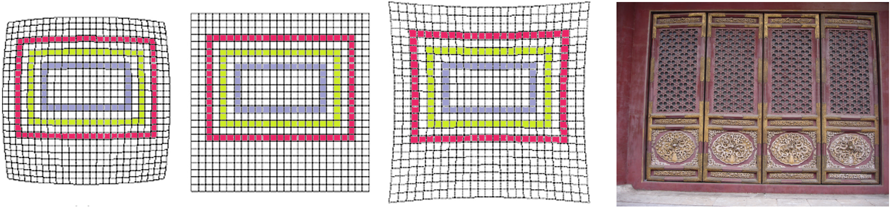
Produced in *charge-coupled device* (CCD) or *complementary metal-oxide semiconductor* (CMOS) technology

Aspect Ratio. Each phototransistor is an $a \times b$ rectangular cell
Ideally, the *aspect ratio* a/b should be equal to 1 (i.e. square cells)

Lens Distortion

Optic lenses contribute *radial lens distortion* to the projection process

Barrel transform or *pincushion transform*

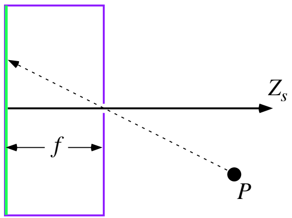


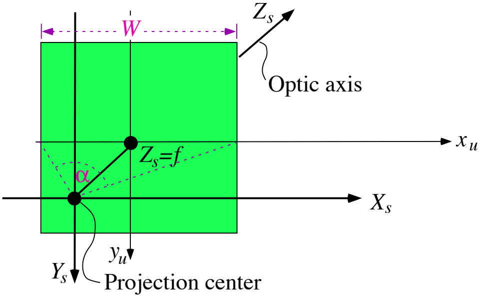
Left to right: Barrel transform, ideal rectangular image, pincushion transform, and projective and lens distortion combined in one image

Model of a Pinhole Camera

Theoretical model for light projection through a small hole

- Diameter of the hole is assumed to be "very close" to zero
- The hole is the projection center





Left: Sketch of an existing pinhole camera ("shoebox camera")

- Point P projected onto an image plane at distance f behind the hole

Right: Model of a pinhole camera,

- Image (width W, viewing angle α) between world and projection center

3D Sensor Coordinates, Image Plane, Focal Length

3D Sensor Coordinates

In figure above: Right-hand $X_s Y_s Z_s$ camera coordinate system
Subscript "s" comes from "sensor" (also, e.g., laser range-finder, or radar)

Z_s-axis points into the world; is the *optic axis*

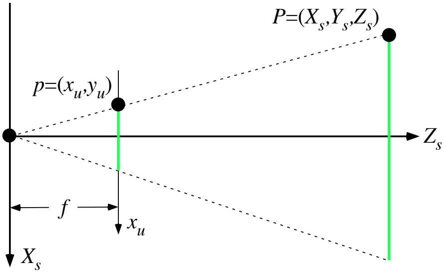
Image Plane

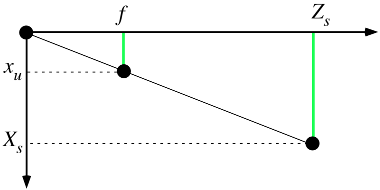
This model excludes the consideration of radial distortion
Thus: *undistorted projected points* in image plane with coordinates x_u, y_u

Focal Length

Distance f between $x_u y_u$ plane and projection center is the *focal length*

Central Projection





Left: Central projection in the X_sZ_s plane for focal length f

Right: Illustration of ray theorem for x_u to X_s and f to Z_s

Central Projection Equations

$X_s Y_s Z_s$ camera coordinates represent points in the 3D world

Visible point $P = (X_s, Y_s, Z_s)$ mapped into $p = (x_u, y_u)$ in the image plane

Ray theorem of elementary geometry

f to Z_s is the same as x_u to X_s

f to Z_s is the same as y_u to Y_s

$$x_u = \frac{fX_s}{Z_s} \qquad y_u = \frac{fY_s}{Z_s}$$

By knowing x_u and y_u we cannot recover all three values X_s, Y_s, Z_s

The Principal Point

Optic axis intersects the image somewhere close to its center

xy image coordinate system: Coordinate origin in the upper left corner

Principal Point

Intersection point (c_x, c_y) of optic axis with image plane in xy coordinates

$$(x, y) = (x_u + c_x, y_u + c_y) = (\frac{fX_s}{Z_s} + c_x, \frac{fY_s}{Z_s} + c_y)$$

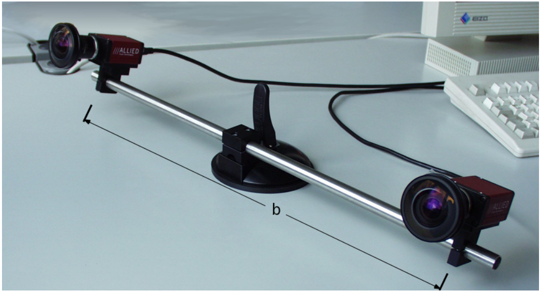
Pixel location (x, y) in 2D xy image coordinates
has 3D camera coordinates $(x - c_x, y - c_y, f)$ in $X_s Y_s Z_s$ system

Camera calibration has to provide c_x, c_y, and f (and more)

Two-Camera Systems

3D geometry of a scene can be measured by using more than one camera
Stereo vision or *binocular vision*: use of two or more cameras

Two Examples of Two-Camera Systems: For Car or Quadcopter



Left: A stereo camera rig on a suction pad with indicated base distance b

Right: Stereo camera system integrated into a quadcopter

Base Distance

Camera calibration
    needs to ensure that we have virtually two identical camera

Base distance b
    the translational distance between projection centers of both cameras

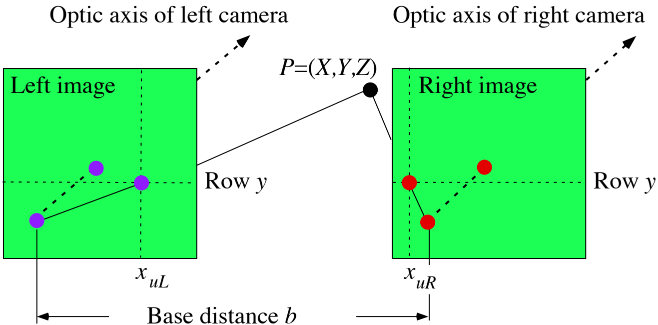
Also to be calibrated

Figure on page before:

    Suction pad: Base distance of about 500 mm
    Quadcopter: Base distance of 110 mm

Result of Camera Calibration

Two virtually-identical cameras perfectly aligned as illustrated below



We describe each camera by using the model of a pinhole camera

Canonical Stereo Geometry

$X_s Y_s Z_s$ camera coordinate system for the left camera

Projection center of the left camera is at $(0, 0, 0)$

Projection center of the right camera is at $(b, 0, 0)$

We have

1. Two coplanar images of identical size $N_{cols} \times N_{rows}$
2. Parallel optic axes
3. An identical effective focal length f
4. Collinear image rows (i.e., row y in one image is collinear with row y in the second image)

Central Projection in Both Cameras

A visible 3D point $P = (X_s, Y_s, Z_s)$ in the $X_s Y_s Z_s$ coordinate system of the left camera is mapped into undistorted image points

$$\begin{aligned} p_{uL} &= (x_{uL}, y_{uL}) = (\frac{f \cdot X_s}{Z_s}, \frac{f \cdot Y_s}{Z_s}) \\ p_{uR} &= (x_{uR}, y_{uR}) = (\frac{f \cdot (X_s - b)}{Z_s}, \frac{f \cdot Y_s}{Z_s}) \end{aligned}$$

in the left and right image plane, respectively

Those two equations are used for stereo vision:

3 input parameter x_{uL}, x_{uR}, and $y_{uL} = y_{uR}$ (same row)
3 parameters X_s, Y_s, and Z_s to be computed

Agenda

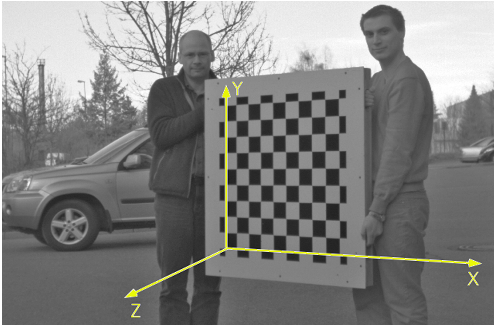
1. Image Recording
2. Camera Calibration
3. Stereo Vision
4. Stereo Matching (Brief Intro)

World Coordinates

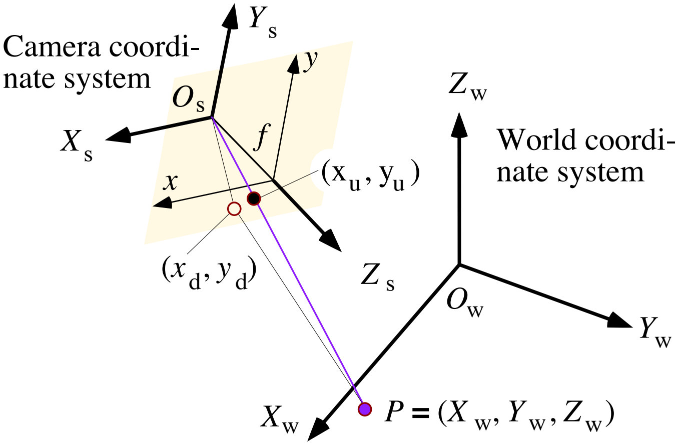
Cameras and 3D objects in the scenes are in one
$X_w Y_w Z_w$ *world coordinate system*

$X_s Y_s Z_s$ camera coordinate system needs to be related
to the chosen world coordinates

Pose (= position + direction) of world coordinate system at a particular
moment during a camera calibration procedure



Left-hand Camera and Right-hand World Coordinates



x_d and y_d are distorted (by lens distortion) coordinates in the image

Affine Transform

Affine transform in 3D space: Maps straight lines into straight lines,
does not change distance ratios between 3 collinear points
is a linear mapping defined by matrix multiplication and translation

Example: First translation $\mathbf{T} = [t_1, t_2, t_3]^\top$ then rotation

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = \mathbf{R}_1(\alpha) \cdot \mathbf{R}_2(\beta) \cdot \mathbf{R}_3(\gamma)$$

Eulerian rotation angles α, β, γ

$$\mathbf{R}_1(\alpha) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{bmatrix} \qquad \mathbf{R}_2(\beta) = \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix}$$

$$\mathbf{R}_3(\gamma) = \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Coordinate Transform

World and camera coordinates to be transformed into each other

E.g. point $P_w = (X_w, Y_w, Z_w)$ in 3D space in world coordinates
into a representation $P_s = (X_s, Y_s, Z_s)$ in camera coordinates

Coordinate $P_w = (X_w, Y_w, Z_w)$ or *vector notation* $P_w = [X_w, Y_w, Z_w]^\top$
(e.g. as measured for calibration points)

$$[X_s, Y_s, Z_s]^\top = \mathbf{R} \cdot \left[[X_w, Y_w, Z_w]^\top + \mathbf{T} \right] = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \cdot \begin{bmatrix} X_w + t_1 \\ Y_w + t_2 \\ Z_w + t_3 \end{bmatrix}$$

Rotation matrix **R** and translation vector **T** to be specified by calibration

Resulting Equational System

By multiplying the matrix and the vector we obtain that

$$
\begin{aligned}
X_s &= r_{11}(X_w + t_1) + r_{12}(Y_w + t_2) + r_{13}(Z_w + t_3) \\
Y_s &= r_{21}(X_w + t_1) + r_{22}(Y_w + t_2) + r_{23}(Z_w + t_3) \\
Z_s &= r_{31}(X_w + t_1) + r_{32}(Y_w + t_2) + r_{33}(Z_w + t_3)
\end{aligned}
$$

Projection from World Coordinates into an Image

Point $P_w = (X_w, Y_w, Z_w)$ in 3D scene is projected into a camera and visible at image point (x, y) in the xy coordinate system (we ignore radial distortion; $Z_s = f$ for image points)

$$\begin{bmatrix} x - c_x \\ y - c_y \\ f \end{bmatrix} = \begin{bmatrix} x_u \\ y_u \\ f \end{bmatrix} = f \begin{bmatrix} X_s/Z_s \\ Y_s/Z_s \\ 1 \end{bmatrix}$$

$$= f \begin{bmatrix} \frac{r_{11}(X_w+t_1)+r_{12}(Y_w+t_2)+r_{13}(Z_w+t_3)}{r_{31}(X_w+t_1)+r_{32}(Y_w+t_2)+r_{33}(Z_w+t_3)} \\ \\ \frac{r_{21}(X_w+t_1)+r_{22}(Y_w+t_2)+r_{23}(Z_w+t_3)}{r_{31}(X_w+t_1)+r_{32}(Y_w+t_2)+r_{33}(Z_w+t_3)} \\ \\ 1 \end{bmatrix}$$

$(c_x, c_y, 0)$ is the principal point in undistorted image coordinates

Intrinsic and Extrinsic Parameters

Camera calibration specifies *intrinsic* (i.e., camera-specific) and *extrinsic* parameters of a given one- or multi-camera configuration

Intrinsic (or internal) parameters are

1. focal length
2. dimensions of the sensor matrix
3. sensor cell size
4. aspect ratio of sensor height to width
5. radial distortion parameters
6. coordinates of the principal point
7. scaling factor

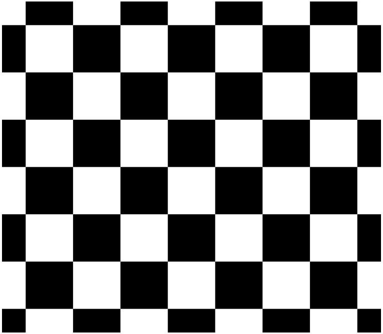
Extrinsic (or external) parameters are

> those of the applied affine transforms for
> identifying poses of cameras in a world coordinate system

A User's Perspective on Camera Calibration

Camera-producer specifies some internal parameters (e.g. the physical size of sensor cells), often not accurate enough for computer vision

Calibration Board



2D checkerboard pattern as commonly used for camera calibration
the corners of squares are the *calibration points*

A Quick Guide

Exact measurement of positions of calibration points

Used calibration points are recorded and localized in images

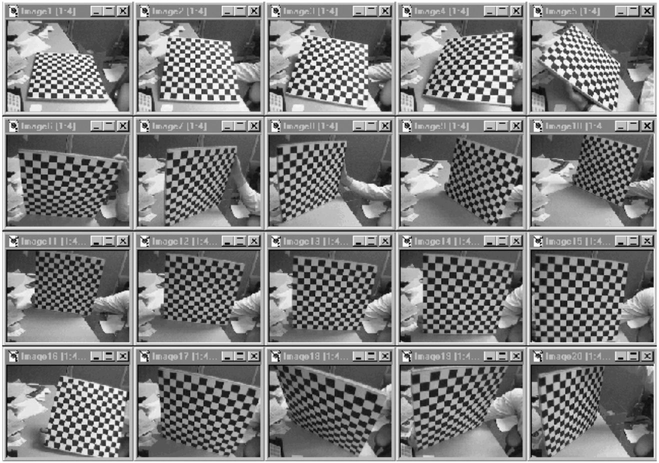
Projected positions in the image grid are compared with measured positions using a *projection model* (e.g. see Page 23 for a model example):

1. Calibration points $P_{w,i}$ are mapped into image points $p_i = (x_i, y_i)$
2. Identify correspondences between $P_{w,i}$ and p_i
3. Insert coordinates of corresponding points into model equations
4. Solve a set of non-linear equations for involved intrinsic and extrinsic parameters by some error minimisation scheme

When calibrating a multi-camera system, all cameras need to be exactly time-synchronized, especially if the calibration rig moves during the procedure

Calibration Software

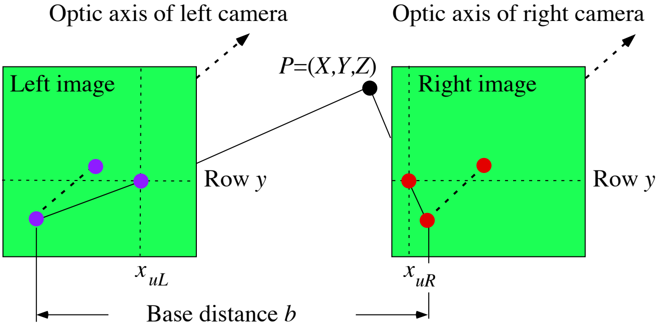
Calibration software is available online, e.g. provided by J.-Y. Bouget



See `www.vision.caltech.edu/bouguetj/calib_doc/` or the `OpenCV` library

Mapping into Canonical Stereo Geometry

Warp recorded image pairs such that it appears that they are recorded in canonical stereo geometry (CSG) by a pair of identical cameras



This process is called *geometric rectification*; it can be done by using intrinsic and extrinsic parameters as obtained by camera calibration

Example 1: Two “non-CSG” Input Images

Recorded pair of images before and after rectification
Recording: $b \approx 10$ m, different viewing directions, definitely not CSG

Example 2: Trinocular "CSG-Recording"

Three cameras installed with goal to match canonical stereo geometry

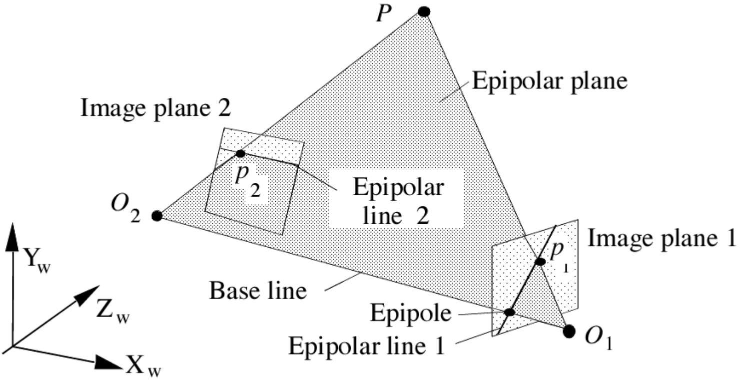
Base distances: left to middle: ≈40 cm, middle to right: ≈30 cm

Rectified middle view shows on the left barrel transform correction

Agenda

Epipolar Geometry

Two cameras in general poses
with projection centers O_1 and O_2



Point $P_w = (X_w, Y_w, Z_w)$ projected into p_1 and corresponding p_2

Simplifying the Task of Stereo Correspondence Analysis

Starting at point p_1 in Image 1,
find the *corresponding point* p_2 in Image 2

Three non-collinear points in 3D space define a plane

Points O_1, O_2 and point p_1 define an *epipolar plane*

Epipolar line = intersection of an image plane with an epipolar plane

Observation

Search for point p_2 can proceed along the epipolar line in Image 2

Canonical Epipolar Geometry

Given: Canonical stereo geometry with a left and a right camera

Visible 3D point P defines an epipolar plane which intersects both image planes in the same row

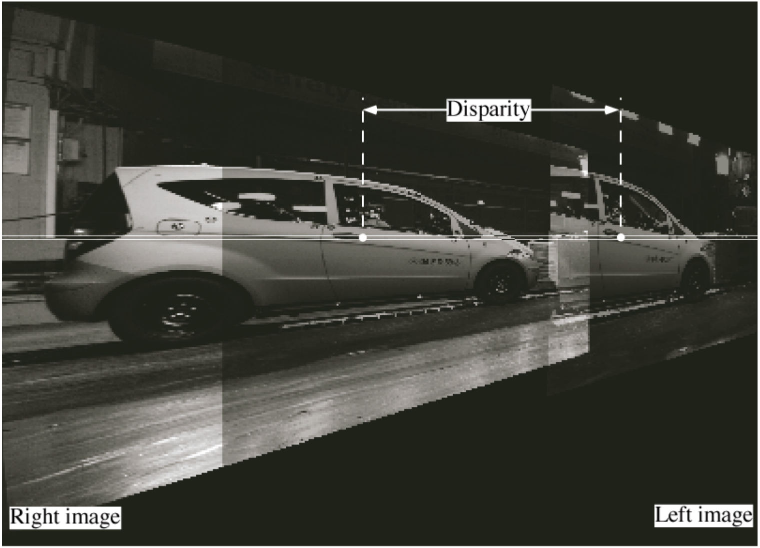
Epipolar Line = Same Image Row

In undistorted coordinates: Start at $p_1 = (x_u, y_u)$ in one camera, search for corresponding point in image row y_u in the other camera

In image coordinates: E.g., start at a pixel $p_L = (x_L, y)$ in the left image, search for corresponding pixel $p_R = (x_R, y)$ in the right image

In this case: $\quad x_R \leq x_L$

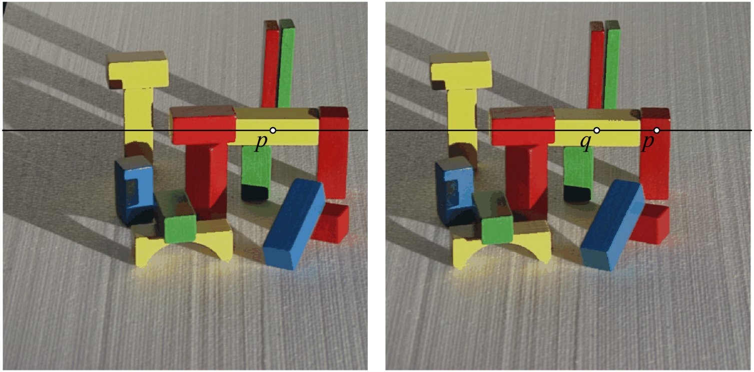
Disparity



Two corresponding points define a *disparity* $x_L - x_R$

$x_L \geq x_R$

A stereo pair recorded in the early 1990*s* with the goal to ensure CSG:



Corresponding point q to a point p in left image
is left of this point in the right image

Triangulation for Canonical Stereo Geometry

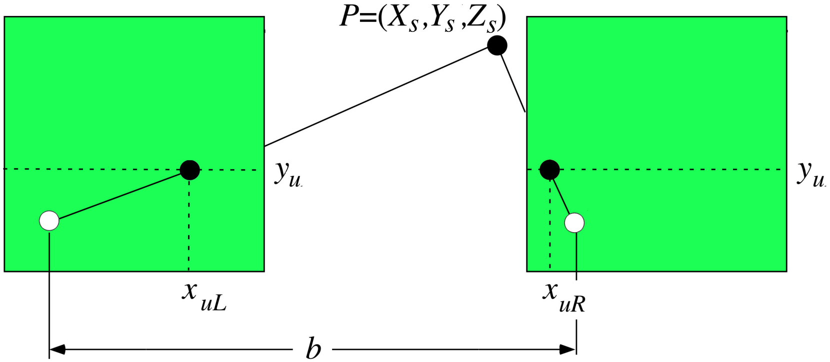
Now: We have all together for going from stereo-image input data to recovered 3D points $P = (X_s, Y_s, Z_s)$ in camera coordinates

Have: Base distance $b > 0$ and unified focal length f

Central projection equations for left and right camera

$$\begin{aligned} p_{uL} &= (x_{uL}, y_u) = (\frac{f \cdot X_s}{Z_s}, \frac{f \cdot Y_s}{Z_s}) \\ p_{uR} &= (x_{uR}, y_u) = (\frac{f \cdot (X_s - b)}{Z_s}, \frac{f \cdot Y_s}{Z_s}) \end{aligned}$$

The Triangle



Solution

Eliminate Z_s from both equations

$$Z_s = \frac{f \cdot X_s}{x_{uL}} = \frac{f \cdot (X_s - b)}{x_{uR}}$$

Solve for X_s

$$X_s = \frac{b \cdot x_{uL}}{x_{uL} - x_{uR}}$$

By using X_s we also obtain

$$Z_s = \frac{b \cdot f}{x_{uL} - x_{uR}}$$

By using Z_s we also derive that

$$Y_s = \frac{b \cdot y_u}{x_{uL} - x_{uR}}$$

with $y_u = y_{uL} = y_{uR}$

Summary

Observation

Two corresponding pixels $(x_L, y) = (x_{uL} + c_{xL}, y_u + c_{yL})$ and
$(x_R, y) = (x_{uR} + c_{xR}, y_u + c_{yR})$
identify its joint pre-image $P = (X_s, Y_s, Z_s)$ in 3D space using the triangulation formulas on the page before

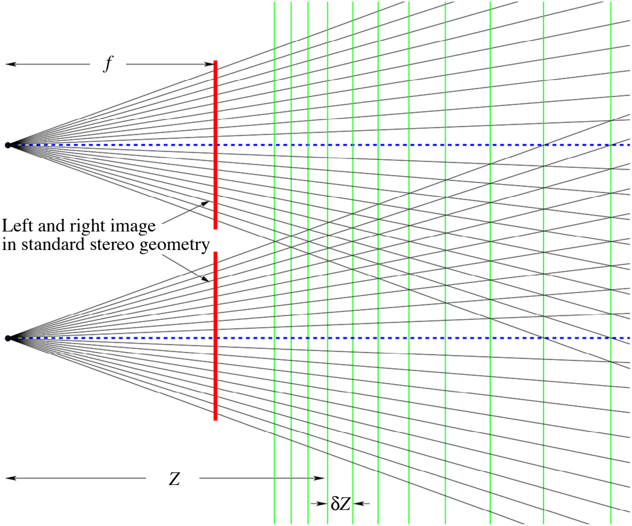
Examples

Disparity $x_{uL} - x_{uR} = 0$ then pre-image $P = (X_s, Y_s, Z_s)$ is "at infinity"

The larger the disparity $x_{uL} - x_{uR}$ the closer is P to the cameras

Integer coordinates in images, thus integer disparities $x_{uL} - x_{uR}$ as well

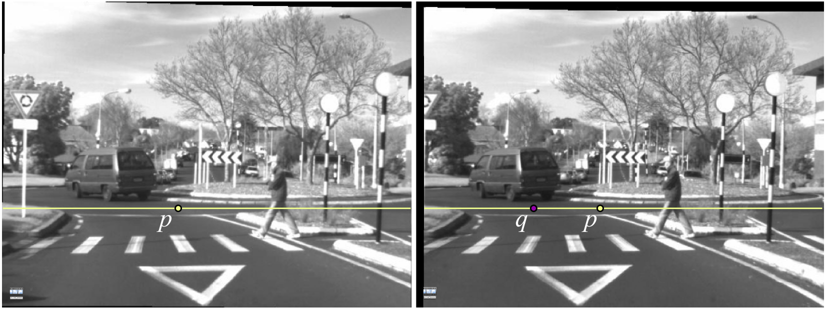
Locations of Reconstructed Points P



Potentially at points where lines cross (which pass through pixel locations)

Agenda

A Stereo Pair: Two Rectified Images



Start at p and search along the epipolar line for a corresponding q

How to detect q?

A correct q would be to the right of the shown q

A Difficult Problem and Approximate Solutions

How to detect pairs of corresponding pixels in a stereo pair?

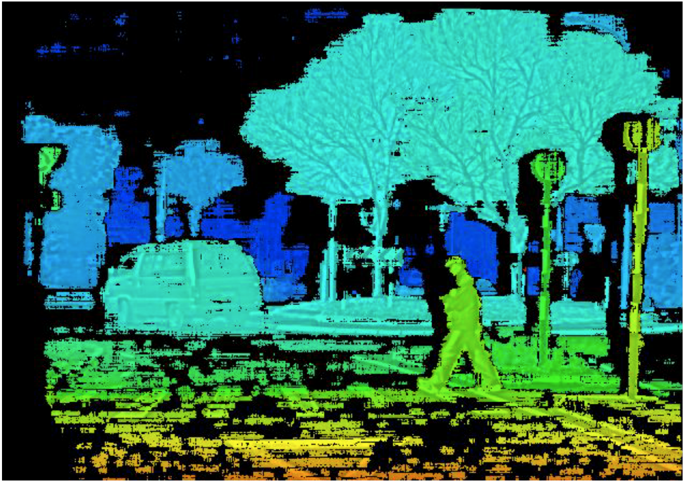
Stereo Matchers

Hundreds of techniques have been proposed for detecting "similar" pixels, typically combined with assuming constraints (e.g. adjacent pixels should have about the same disparity) for supporting an optimisation strategy

Two Examples of Stereo Matchers

Semi-global matching (SGM) and its variant, *iterative SGM* (iSGM)

SGM Disparity Map for the Shown Stereo Pair

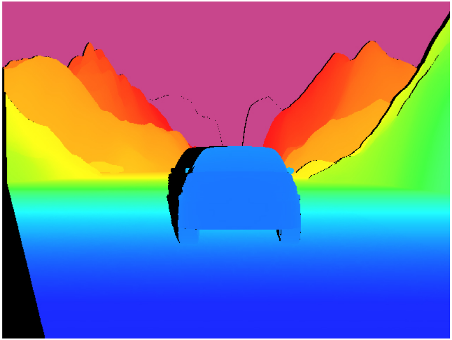


Color: Color-encoded disparity values at pixel locations

Black: Pixel location without assigned disparity value

Ground Truth for Synthetic Data

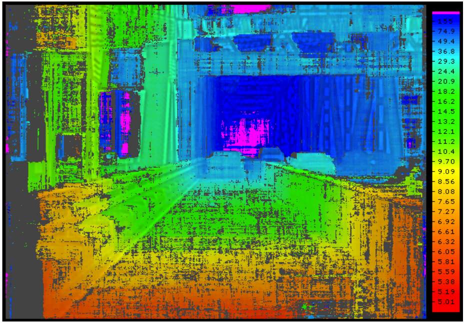
Useful for evaluating the accuracy of a stereo matcher



Left: Original rendered image from Set 2 of EISATS
Right: Disparity map, here illustrating *ground truth*
(Note: Use of a different color key)

iSGM Depth Map

Each disparity defines (via triangulation) a depth or distance value



Left: One of the two input images

Right: Depth map using the color key as shown on the right
distances are in meters

Copyright Information