Sensors and Sensing
Cameras

Todor Stoyanov

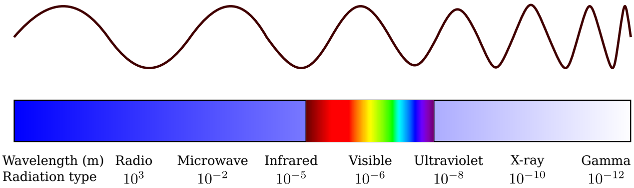
Centre for Applied Autonomous Sensor Systems
Örebro University
Sweden

Outline

Outline

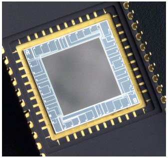
Electromagnetic Spectrum

- Cameras are passive devices that measure electromagnetic radiation, reflected by objects in the environment.
- Conventional cameras detect light in the visible range of the electromagnetic spectrum: wavelengths between 430nm-790nm.
- ... but plenty of cameras built for other ranges: e.g., infrared, thermal, UV.



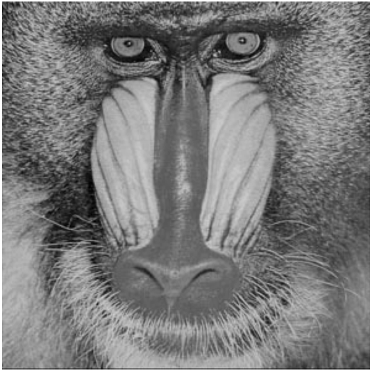
CCD Sensors, Optics and Shutters

- Modern cameras consist of a light sensitive element, a lens and (optionally) a shutter.
- The light sensor is usually implemented as a Charge Coupled Device (CCD) or a CMOS chip.
- Lenses focus light onto the CCD array.
- Mechanical shutters can be used to only expose the chip for a short period of time.
- Electronic shutters are often used instead.



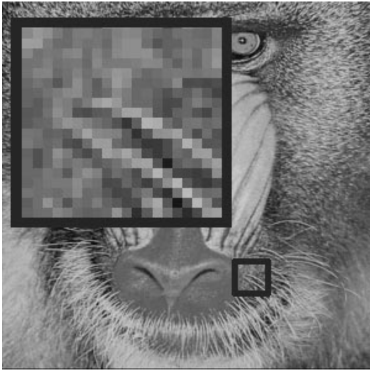
Digital Images

- A digital image I can be thought of as a function $f(x, y) : X \times Y \rightarrow Z$, where $X = [0, P_x] \in \mathbb{N}$ and $Y = [0, P_y] \in \mathbb{N}$ are pixel coordinates in the image plane.
- Depending on the type of image Z can be:
  - Binary image if $Z = \{0, 1\}$
  - Gray scale image if $Z \subset \mathbb{R}$
  - Color image if $Z \subset \mathbb{R}^3$
- Each pixel in the image corresponds to a single cell of the CCD array.

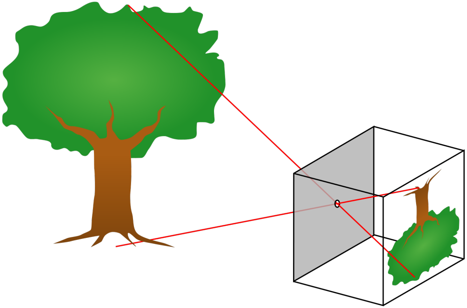


Digital Images

- A digital image I can be thought of as a function $f(x, y) : X \times Y \rightarrow Z$, where $X = [0, P_x] \in \mathbb{N}$ and $Y = [0, P_y] \in \mathbb{N}$ are pixel coordinates in the image plane.
- Depending on the type of image Z can be:
  - Binary image if $Z = \{0, 1\}$
  - Gray scale image if $Z \subset \mathbb{R}$
  - Color image if $Z \subset \mathbb{R}^3$
- Each pixel in the image corresponds to a single cell of the CCD array.

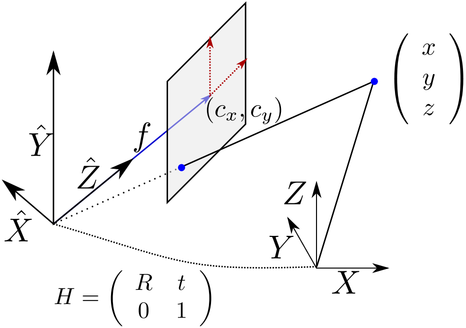


The Pinhole Camera



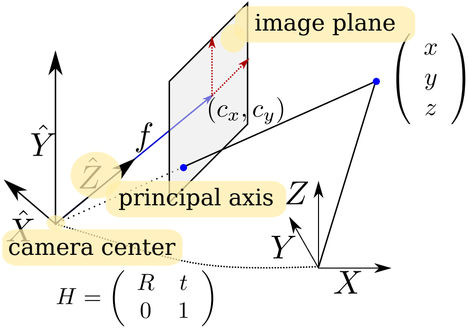
- A pinhole camera is a simple camera without a lens that projects light directly on an image plane.
- The pinhole camera model can be extended to model complex cameras
- Some definitions:

The Pinhole Camera



- The pinhole camera model can be extended to model complex cameras
- Some definitions:

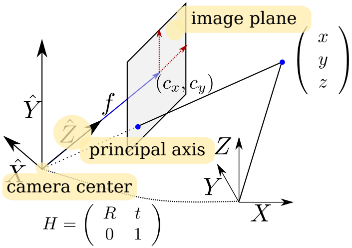
The Pinhole Camera



- The pinhole camera model can be extended to model complex cameras
- Some definitions:

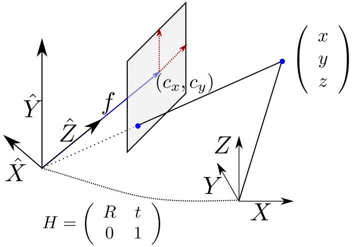
The Pinhole Camera

- Camera center: the focal point of all rays converging to the camera
- Principal axis: by definition this is the $\hat{Z}$ axis pointing out of the camera center
- Image plane: the CCD plane where the image is acquired
- Focal length f: the vector pointing from the camera center to the image plane



The Pinhole Camera

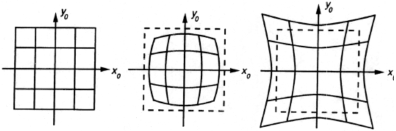
- Given a point $s = (x, y, z, 1)^T$ in world coordinate frame (homogeneous coordinates), we can obtain the projection of the point to a corresponding pixel (x_u, y_u) on the image plane as:



$$\begin{pmatrix} x_u \\ y_u \\ 1 \end{pmatrix} \sim \begin{pmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & cy & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} H \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (1)$$

$$= KHs \quad (2)$$

Lens Distortion



No Distortion Barrel Distortion Pincushion Distortion [1]

- In practice, adding a lens to the system adds several different types of distortion
  - radial distortion is due to imperfections of the lens curvature
  - tangential distortion is due to imperfect alignment of the lens center and the principle axis
- other types of distortion are more difficult to model.

[1] http://www.uni-koeln.de/~al001/radcor_files/hs100.htm

Modeling Distortion

Going back to Eq. 1, consider the projection of $s = (x, y, z, 1)^T$ onto the image plane:

$$s' = \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} x/z \\ y/z \\ 1 \end{pmatrix} \tag{3}$$

We define $r = \sqrt{x'^2 + y'^2}$ as the radius of the projected point, relative to the principal point. The undistorted point $\hat{s}$ can then be computed as:

$$\hat{s} = \begin{pmatrix} x'(1 + k_1 r^2 + k_2 r^4) + 2p_1 x'y' + p_2(r^2 + 2x'^2) \\ y'(1 + k_1 r^2 + k_2 r^4) + p_1(r^2 + 2y'^2) + 2p_2 x'y' \\ 1 \end{pmatrix} \tag{4}$$

The undistorted pixel coordinates of s are then obtained as:

$$\begin{pmatrix} x_u \\ y_u \\ 1 \end{pmatrix} = K\hat{s} \tag{5}$$

Electronic Shutter Distortions

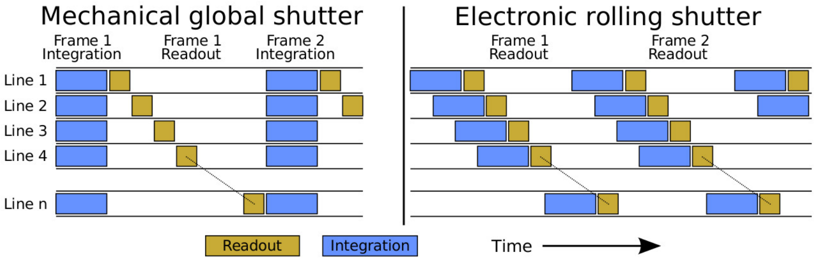


Figure: Image from [1]

Electronic Shutter Distortions

(a) Original (with rolling shutter)

(b) Rectified (no rolling shutter)

Figure: http://www.cvl.isy.liu.se/education/tutorials/rolling-shutter-tutorial/

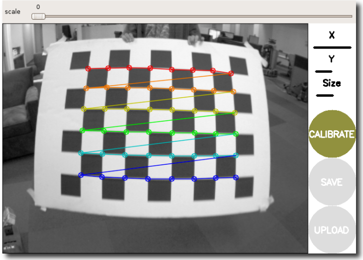
Motion Blur

Figure: Image from wikipedia

Outline

2D/3D Calibration Patterns

- In order to determine the focal length and center offsets for the camera matrix K, the radial distortion coefficients k_1, k_2 and the tangential distortions p_1, p_2 cameras are calibrated.
- Calibration from a natural scene is not easily done, so we use calibration patterns
  - 2D patterns: known pattern printed on a plane
  - rarely 3D pattern: known 3D geometry

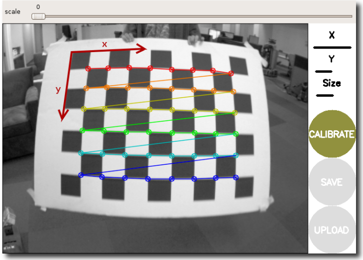


2

[2] ROS camera calibration tutorial

Chessboard calibration: basics

- The number of chessboard squares and their size is known in advance.
- We fix the world reference frame to the top-left corner of the board.
- All points lie on a plane in world frame, with $z = 0$. This means we can drop one column of rotation coefficients from H



Note: the following slides follow the derivations as shown here[3].

[3] http://ais.informatik.uni-freiburg.de/teaching/ws10/robotics2/pdfs/rob2-10-camera-calibration.pdf

Chessboard calibration: formulation

For a point $(x, y, z, 1)^T$ we have :

$$\begin{pmatrix} x_u \\ y_u \\ 1 \end{pmatrix} = K \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \tag{6}$$

Chessboard calibration: formulation

For a point $(x, y, z, 1)^T$ we have, we set z=0, thus :

$$\begin{pmatrix} x_u \\ y_u \\ 1 \end{pmatrix} = K \begin{pmatrix} r_{11} & r_{12} & 0 & t_1 \\ r_{21} & r_{22} & 0 & t_2 \\ r_{31} & r_{32} & 0 & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 0 \\ 1 \end{pmatrix} \tag{7}$$

Chessboard calibration: formulation

For a point $(x, y, z, 1)^T$ we have, we set z=0, thus :

$$\begin{pmatrix} x_u \\ y_u \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & cy \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & t_1 \\ r_{21} & r_{22} & t_2 \\ r_{31} & r_{32} & t_3 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{8}$$

$$= H \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{9}$$

Chessboard calibration: formulation

For a point $(x, y, z, 1)^T$ we have, we set z=0, thus :

$$\begin{pmatrix} x_u \\ y_u \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & cy \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & t_1 \\ r_{21} & r_{22} & t_2 \\ r_{31} & r_{32} & t_3 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (8)$$

$$= H \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (9)$$

H is called the homography matrix. Let:

$$H = (h_1, h_2, h_3) = K(r_1, r_2, t) \quad (10)$$

Chessboard calibration: formulation

Knowing

$$(h_1, h_2, h_3) = K(r_1, r_2, t) \tag{11}$$

we have that $r_1 = K^{-1}h_1$ and $r_2 = K^{-1}h_2$. We also know that r_1 and r_2 are columns from a rotation matrix, thus they form an orthonormal basis, i.e. $r_1^T r_2 = 0$ and $r_1^T r_1 = r_2^T r_2 = 1$. Therefore:

$$\begin{aligned} r_1^T r_2 &= 0 && (12)\\ h_1^T K^{-T} K^{-1} h_2 &= 0 && (13) \end{aligned}$$

and

$$\begin{aligned} r_1^T r_1 &= r_2^T r_2 && (14)\\ h_1^T K^{-T} K^{-1} h_1 &= h_2^T K^{-T} K^{-1} h_2 && (15)\\ h_1^T K^{-T} K^{-1} h_1 - h_2^T K^{-T} K^{-1} h_2 &= 0 && (16) \end{aligned}$$

Chessboard calibration: solving the problem

Re-formulating equations 13 and 16 and unwrapping the coefficients of $B = K^{-T}K^{-1}$ as $b = (b_{11}, b_{12}, b_{13}, b_{22}, b_{23}, b_{33})$ (B is symmetric) we can formulate

$$Vb = 0 \tag{17}$$

where V holds the coefficients from H as in equations 13 and 16.

- As we know the relative positions of the points on the pattern, we can obtain V for an image and solve for b.
- The parameters of K can be obtained by Cholesky factorization of $B = LL^T$
- Measurements are noisy, so we instead solve a least squares problem to minimize Vb

Chessboard calibration: distortion parameters

- The previous derivations solve the problem for the camera matrix K, but ignore distortion.
- In order to solve for distortion, we need to formulate the re-projection error.
- The resulting non-linear optimization problem is usually solved in batch by using the Levenberg-Marquardt method and linearization around the solution for K at every iteration.
- Fortunately, you don't typically have to solve the optimization problem yourself. Just use one of the many toolboxes for calibration.

Chessboard calibration: distortion parameters

- The previous derivations solve the problem for the camera matrix K, but ignore distortion.
- In order to solve for distortion, we need to formulate the re-projection error.
- The resulting non-linear optimization problem is usually solved in batch by using the Levenberg-Marquardt method and linearization around the solution for K at every iteration.
- Fortunately, you don't typically have to solve the optimization problem yourself. Just use one of the many toolboxes for calibration.

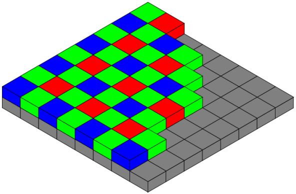
Outline

Color Cameras

- The world is colorful!
- Color images are obtained by adding a filter designed for specific wavelengths to each pixel of the CCD array
- The filter pattern is called a Bayer filter.
- Typically, we have red, green and blue sensitive pixels.
- Raw pixel values often come as a stream and camera drivers perform de-bayering to obtain the corresponding RGB vector.

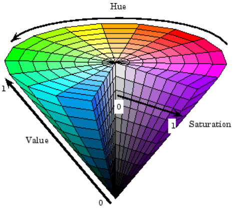
Color Cameras

- The world is colorful!
- Color images are obtained by adding a filter designed for specific wavelengths to each pixel of the CCD array
- The filter pattern is called a Bayer filter.
- Typically, we have red, green and blue sensitive pixels.
- Raw pixel values often come as a stream and camera drivers perform de-bayering to obtain the corresponding RGB vector.



Color Spaces

- Several different models for colors.
- RGB systems encode values in equal sized intervals for red, green and blue.
- HSV space encodes the color hue, saturation and value of every pixel with
- Different color spaces have advantages in different operations.
- Color spaces are not equivalent, but we can convert between representations



$$H \in [0°, 360°)$$

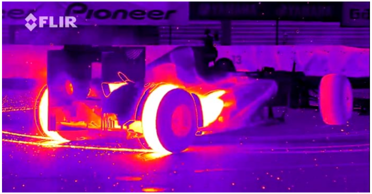
$$S \in [0, 1]$$

Infrared and Thermal Cameras

- Cameras sensitive to other parts of the EM spectrum
- IR cameras measure reflected IR light.
- Often used with an IR diode light source for night-time security applications.
- Thermal cameras can detect emitted IR light.

Infrared and Thermal Cameras

- Cameras sensitive to other parts of the EM spectrum
- IR cameras measure reflected IR light.
- Often used with an IR diode light source for night-time security applications.
- Thermal cameras can detect emitted IR light.

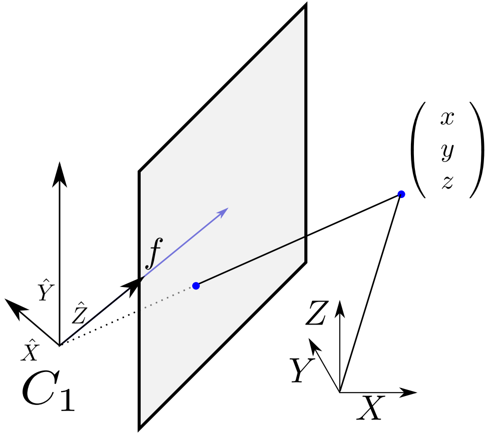


Outline

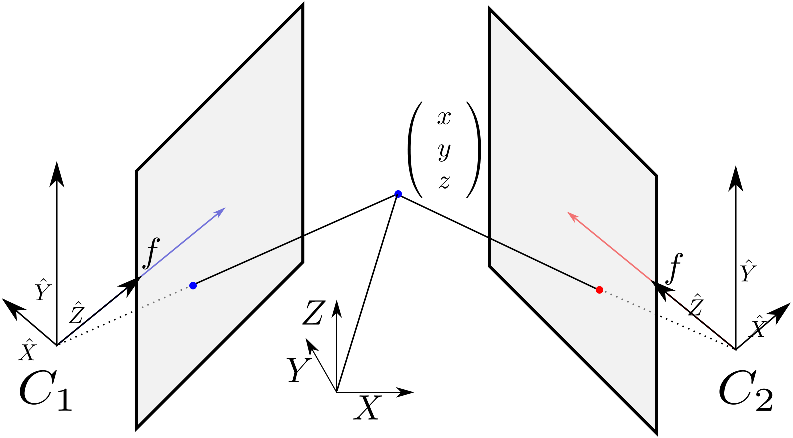
Stereo Cameras Basics

- Stereo cameras have been in use in robotics since its early days.
- The main principle of operation is to use two (or more) cameras with precisely known relative offsets.
- Points in the images are identified and matched.
- 3D distance to matching points can be determined using triangulation.

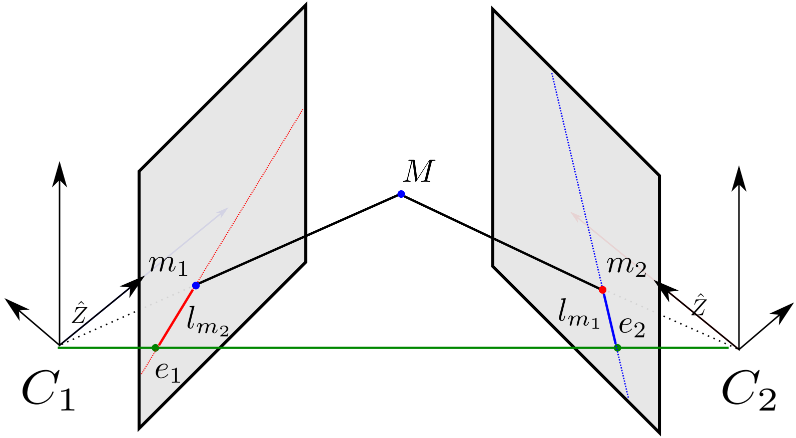
Epipolar Geometry



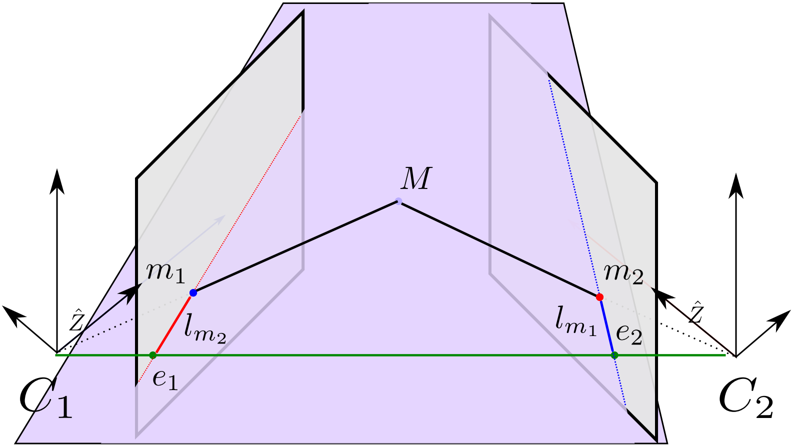
Epipolar Geometry



Epipolar Geometry

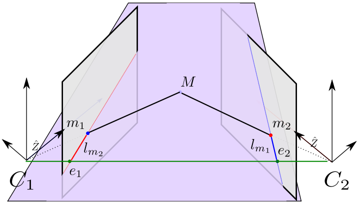


Epipolar Geometry



Epipolar Geometry

- The line l_{m_1} is the projection of the line between M and C_1 onto I_2.
- Any point m_1 on I_1 defines a line l_{m_1}: an epipolar line.
- The epipolar lines define a plane on which m_1, m_2 and M lie.
- The line between e_1 and e_2 links the two camera centers. All epipolar lines pass through e_1 and e_2.
- $l_{m_1} = Em_1$ and $l_{m_2} = E^T m_2$



Epipolar Geometry

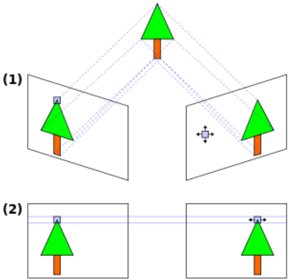
- The matrix E is determined by the rotation R and translation t between the two camera centers C_1 and C_2
- A stereo camera pair is completely described by a fundamental matrix F, such that:
  - K_1 and K_2 are the camera matrices as described previous lecture.
  - $F = K_1^{-T} E K_2^{-1}$
  - if m_1 and m_2 are the projections of M onto C_1 and C_2, then:

$$m_1^T F m_2 = 0 \tag{18}$$

- Knowing F and the pixel coordinates of a corresponding point m_1,m_2 we can find the depth along Z in each of the cameras C_1,C_2

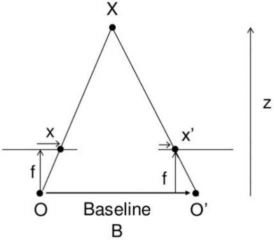
Disparity and Rectification

- The above setup can be used for triangulation of corresponding points.
- For practical reasons stereo cameras are built with only a horizontal displacement between C_1 and C_2.
- Even in cases where the cameras are not perfectly aligned, images are usually rectified to be aligned
- Advantage: searching for correspondences can be done on a line, instead of a plane.
- Using the pixel distance $d = x - x'$, we get the depth $Z = (x - x')Bf$

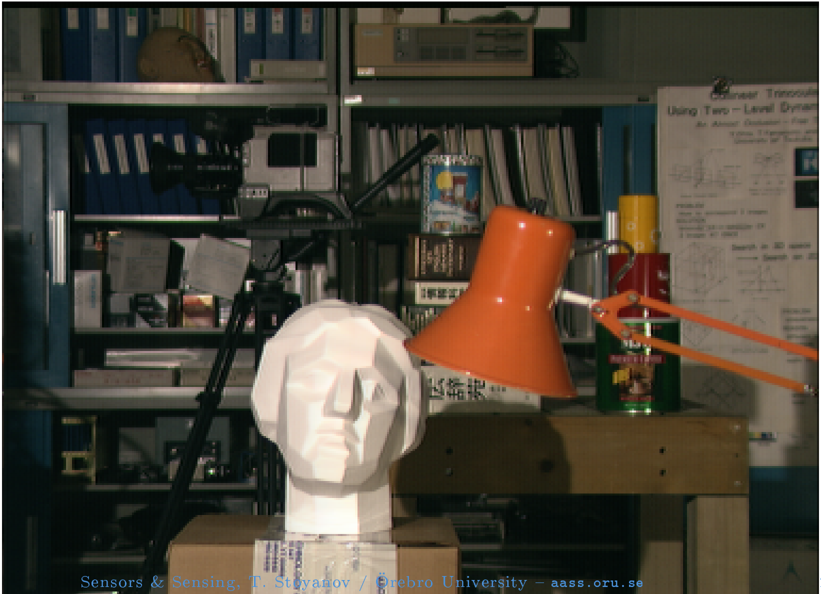


Disparity and Rectification

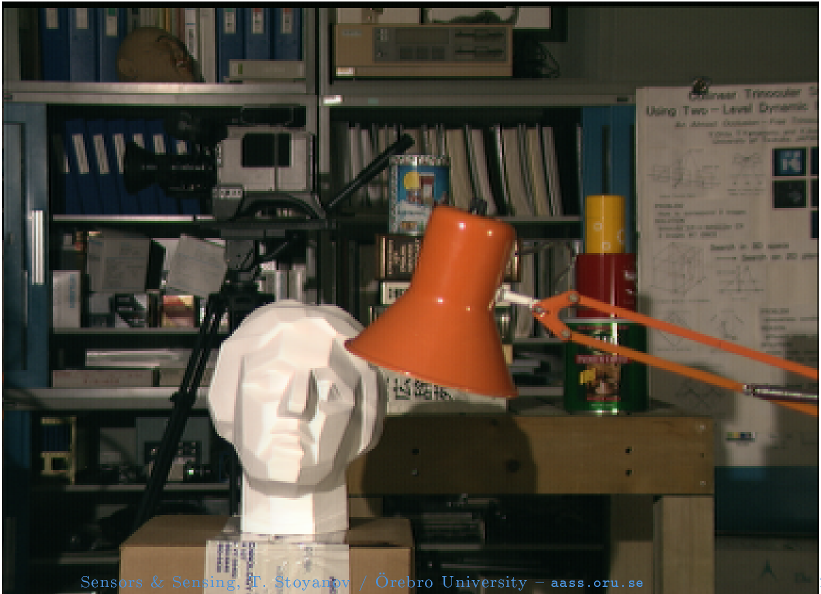
- The above setup can be used for triangulation of corresponding points.
- For practical reasons stereo cameras are built with only a horizontal displacement between C_1 and C_2.
- Even in cases where the cameras are not perfectly aligned, images are usually rectified to be aligned
- Advantage: searching for correspondences can be done on a line, instead of a plane.
- Using the pixel distance $d = x - x'$, we get the depth $Z = (x - x')Bf$



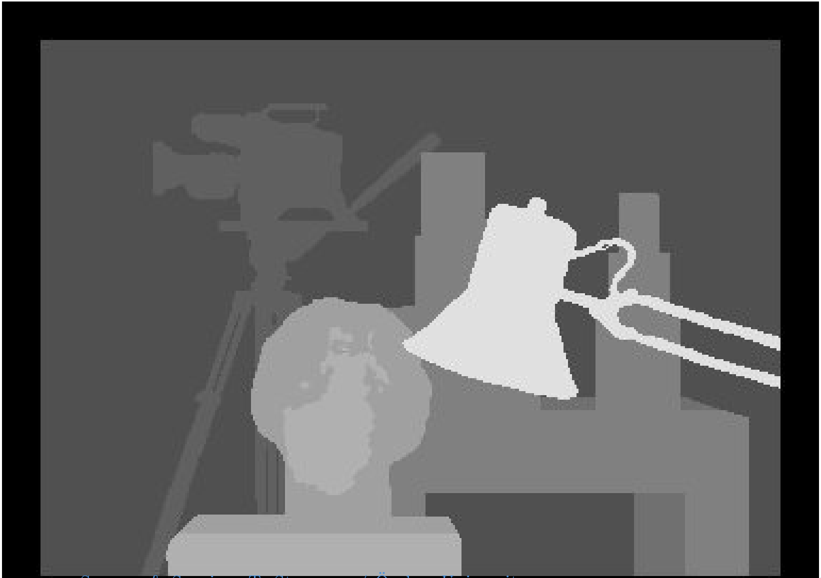
Disparity and Rectification



Disparity and Rectification

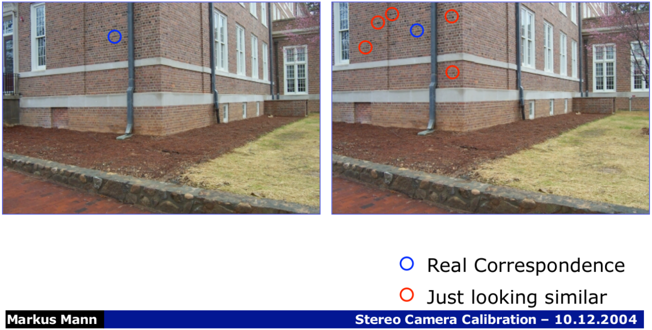


Disparity and Rectification



Identifying Matching Points

- Stereo vision algorithms rely on matching corresponding points between two images.
- This causes ambiguities in some cases:



Identifying Matching Points

- Stereo vision algorithms rely on matching corresponding points between two images.
- This causes ambiguities in some cases:
- Two basic classes of approaches:
- Intensity-based:
- Feature-based:

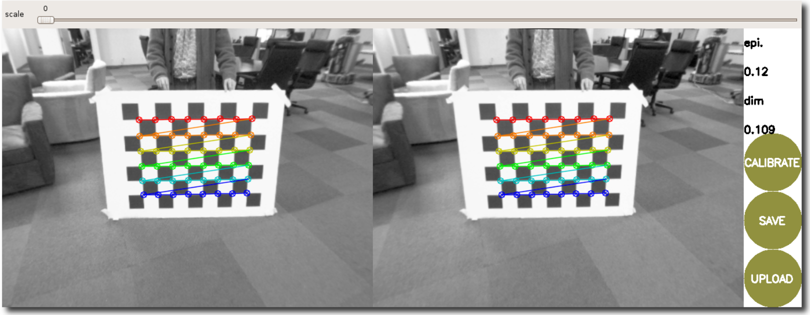
Identifying Matching Points

- Stereo vision algorithms rely on matching corresponding points between two images.
- This causes ambiguities in some cases:
- Two basic classes of approaches:
- Intensity-based:
  - Basic idea: match the information in the intensity values of each line.
  - At each pixel minimize the difference in intensities.
- Feature-based:

Identifying Matching Points

- Stereo vision algorithms rely on matching corresponding points between two images.
- This causes ambiguities in some cases:
- Two basic classes of approaches:
- Intensity-based:
- Feature-based:
  - Basic idea: extract information using a group of pixel values and their spatial distribution.
  - Commonly used are simple features, e.g. edge and corner detectors (Canny, Harris, LoG filters).

Stereo Calibration



- Procedure for calibrating is similar as for a single camera, relying on a number of known points
- The Fundamental matrix F is rank 2, i.e., there are 7 DoFs that need to be optimized.
- Typical algorithms are the 7-point and 8-point correspondence algorithms.

Limitations

- Lack of texture.
- Lightning conditions.
- Processing becomes more costly for higher resolutions.
- Need to verify depth maps using more complicated algorithms.
- Quantization effects due to discretisation of pixels.

References

Per-Erik Forssén and Erik Ringaby.
Rectifying rolling shutter video from hand-held devices.
In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pages 507–514. IEEE, 2010.

Sensors and Sensing
Cameras

Todor Stoyanov

Centre for Applied Autonomous Sensor Systems
Örebro University
Sweden