Technical University of Cluj - Napoca
Computer Science Department

Image Processing

(Year III, 2-nd semester, English-class)

Lecture 13: Stereovision

STEREOVISION

Goal

The fundamental equations of the *pinhole camera model* are [Trucco98]:

$$\begin{cases} x = f \cdot \dfrac{X_C}{Z_C} \\ y = f \cdot \dfrac{Y_C}{Z_C} \end{cases}$$

$P(X_C,Y_C,Z_C)$ 3D point in the camera coordinate system
$p(x,y,-f)$ its projection on the image plane

Knowing the image coordinates (*x,y*) we cannot infer the depth (Z), only the projection equations

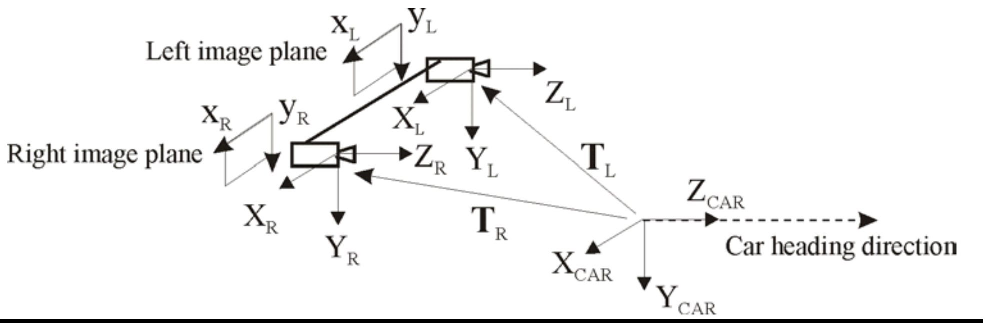
Measure depth (Z) ⇒ at least two cameras (stereo-system)

Stereo camera configurations

- Canonic (parallel axes) – theoretical model (impossible to obtain in practice) ⇒ image rectification
- Coplanar axes (but unparallel)
- General configuration

STEREOVISION

The model of a stereovision system



Parameters	Type of parameters	Set of parameters
Internal parameters of the stereo system	Intrinsic parameters set / camera	Principal points: $PP_L\ (x_{0L}, y_{0L})$, $PP_R\ (x_{0R}, y_{0R})$ Focal distance: $f_L\ (f_{XL},f_{YL})$, $f_R\ (f_{XR},f_{YR})$ Radial distortion: (k_1^L, k_2^L), (k_1^R, k_2^R) Tangential distortion: (p_1^L, p_2^L), (p_1^R, p_2^R)
	Relative extrinsic parameters set / stereo system	$T_{REL} = R_{CL}^T(T_{CR} - T_{CL})$ $R_{REL} = R_{CL}^T R_{CR}$
External parameters of the stereo system	Absolute extrinsic parameters set / camera	-For cameras: Translation vectors: T_{CL}, T_{CR} Rotation matrices: R_{CL}, R_{CR} Rotation vectors: r_{CL}, r_{CR} -For the stereo rig: Translation vector: $T_{C\text{-}rig} = T_{CL}$ Rotation matrix: $R_{C\text{-}rig} = R_{CL}$ Rotation vector: $r_{C\text{-}rig} = r_{CL}$

STEREOVISION

The stereovision process

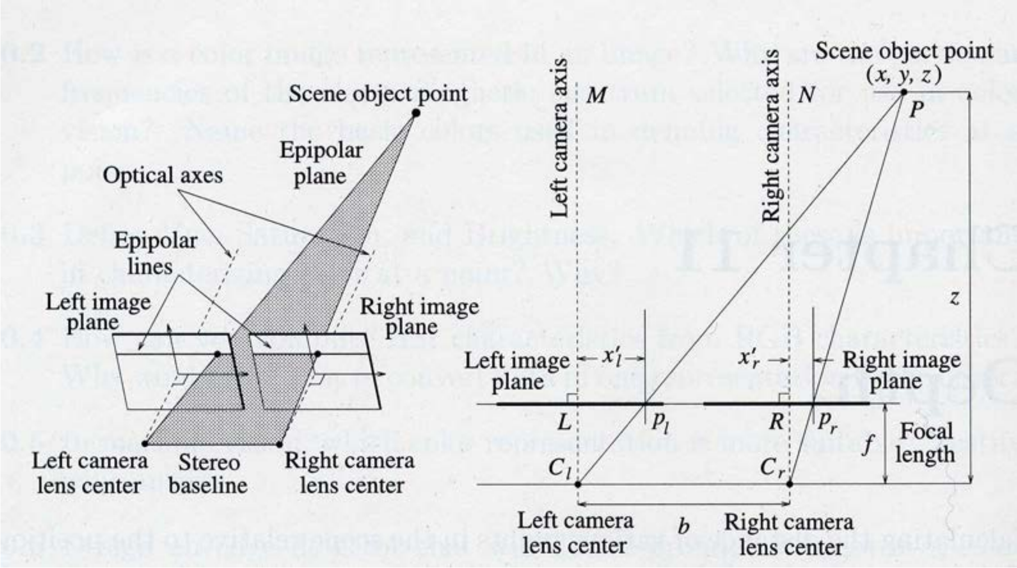
1. **Stereo-rig setup**
2. **Camera and stereo-rig calibration**
3. **Synchronous acquisition of image pairs**
4. **Image rectification (if the canonic approach is used)**
5. **Feature extraction and selection (edges or pixels)**
6. **Correspondence searching between left and right image features**
7. **3D reconstruction in camera /world coordinate system**

References:

E. Trucoo, A. Verri, "Introductory techniques for 3D Computer Vision", Prentice Hall, 1998

STEREOVISION

The canonical model



Assumptions

- Image planes are coplanar ⇒ optical axes are parallel
- Horizontal image axes are collinear
- Epipolar lines – horizontal
- $v_{0L} = v_{0R} \Rightarrow y_L = y_R$

Depth estimation (canonical config.)

$$x_l^{'} = f_X \cdot \frac{X_1}{Z_1}$$

$$x_r^{'} = f_X \cdot \frac{X_2}{Z_2}$$

$$d = x_l^{'} - x_r^{'} = f_X \cdot \left(\frac{X_1}{Z_1} - \frac{X_2}{Z_2} \right) = f_X \cdot \frac{X_1 - X_2}{Z} = f_X \frac{b}{Z}$$

$$Z = \frac{f_X \cdot b}{d}$$

$X_1 = x_l^{'} * Z / f_x$; $Y_1 = y_l * Z / f_y$

Depth estimation (coplanar config.)

Coplanar but non-parallel optical axes: θ angle

$$Z = \frac{f_X \cdot b}{d + f_X \cdot \tan(\theta)}$$

STEREOVISION

Range Resolution

Often it's important to know the minimal change in range that stereo can differentiate, that is, the *range resolution* of the method. Range resolution is a function of the range itself. At closer ranges, the resolution is much better than at farther ranges.

Range resolution is governed by the following equation, [Konolige1999]:

$|\Delta Z| = (Z^2/fb)\Delta d \quad (Z=fb/d;\ \Delta Z/\Delta d = -fb/d^2;\ \Delta Z/\Delta d = -Z^2/fb;\ |\Delta Z| = (Z^2/fb)\Delta d;)$

The range resolution **ΔZ** is the smallest change in range **Z** that is discernible by the stereo geometry, given a change in disparity of **Δd**. The range resolution goes up (gets worse) by the square of the range.

The baseline **b** and focal **f** length both have an inverse influence on the resolution, so that larger baselines and focal lengths (telephoto) make the range resolution better.

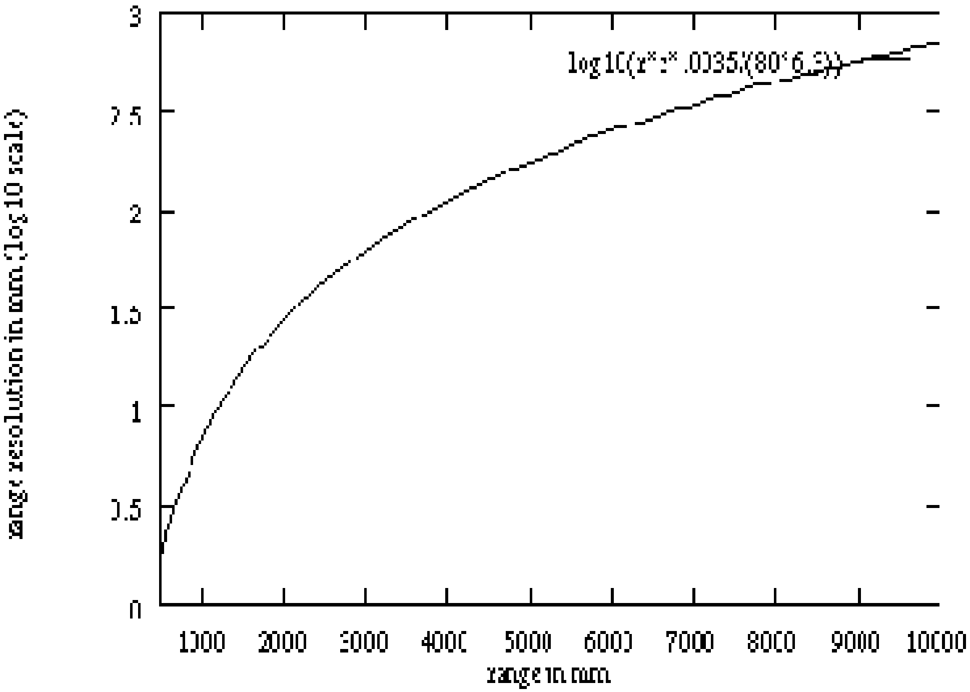
Finally, the pixel size has a direct influence, so that smaller pixel sizes give better resolution.

Typically, stereo algorithms can report disparities with sub-pixel precision, which also increases range resolution.

STEREOVISION

Range Resolution

The figure plots range resolution as a function of range for the STH-V1 stereo head, given a baseline of 8 cm and 6.3 mm lenses. The Stereo Engine interpolates disparities to 1/4 pixel, so **Δd** is 1/4 * 14 um = 3.5 um.

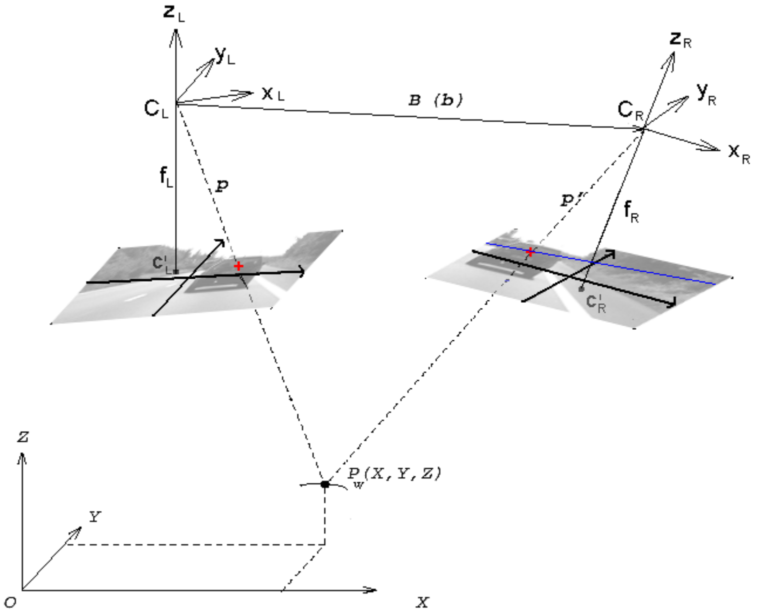


The range resolution is plotted on a log10 scale to show more detail at closer ranges.

At 1 meter, the range resolution is about 10 mm. At 4 meters, it has grown to 100 mm; by 10 meters, it is almost a meter.

STEREOVISION

3D stereo-reconstruction for GENERAL CAMERA CONFIGURATION



Intrinsic parameters

Focal lengths

- Left camera: f_L (f_{XL}, f_{YL})
- Right camera: f_R (f_{XR}, f_{YR})

Principal points

- Left camera: (x_{CL},y_{CL}).
- Right camera: (x_{CR},y_{CR}).

Extrinsic parameters

$$\mathbf{T}_{CL} = \begin{vmatrix} \mathrm{X}_{\mathrm{CL}} \\ \mathrm{Y}_{\mathrm{CL}} \\ \mathrm{Z}_{\mathrm{CL}} \end{vmatrix} \qquad \mathbf{R}_{CL} = \begin{vmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{vmatrix}$$

$$\mathbf{T}_{CR} = \begin{vmatrix} \mathrm{X}_{\mathrm{CR}} \\ \mathrm{Y}_{\mathrm{CR}} \\ \mathrm{Z}_{\mathrm{CR}} \end{vmatrix} \qquad \mathbf{R}_{CR} = \begin{vmatrix} r'_{11} & r'_{12} & r'_{13} \\ r'_{21} & r'_{22} & r'_{23} \\ r'_{31} & r'_{32} & r'_{33} \end{vmatrix}$$

STEREOVISION

3D stereo-reconstruction problem := mapping 2D image pairs ($p_L(x_L,y_L)$, $p_R(x_R,y_R)$) into a unique 3D point P_W:

$$\mathbf{P}_W = \begin{vmatrix} X_W \\ Y_W \\ Z_W \end{vmatrix}$$

The solution:

$$\mathbf{P}_W = \mu \mathbf{R}_L * \begin{vmatrix} x_L - x_{CL} \\ y_L - y_{CL} \\ -f_L \end{vmatrix} + \mathbf{T}_{CL} \quad \Leftrightarrow \quad \begin{vmatrix} X_W \\ Y_W \\ Z_W \end{vmatrix} = \mu \mathbf{R}_L * \begin{vmatrix} x_L - x_{CL} \\ y_L - y_{CL} \\ -f_L \end{vmatrix} + \begin{vmatrix} X_{CL} \\ Y_{CL} \\ Z_{CL} \end{vmatrix} \qquad (1)$$

Where μ is a scaling factor depending on Z and can be expressed from the 3-rd equation of the following system:

$$\begin{vmatrix} x_L - x_{CL} \\ y_L - y_{CL} \\ -f_L \end{vmatrix} = \mu^{-1} \mathbf{R}_L^T * \begin{vmatrix} X_W - X_{CL} \\ Y_W - Y_{CL} \\ Z_W - Z_{CL} \end{vmatrix} \quad \Leftrightarrow \quad \begin{vmatrix} x_L - x_{CL} \\ y_L - y_{CL} \\ -f_L \end{vmatrix} = \mu^{-1} \begin{vmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{vmatrix} * \begin{vmatrix} X_W - X_{CL} \\ Y_W - Y_{CL} \\ Z_W - Z_{CL} \end{vmatrix} \qquad (2)$$

$$\mu^{-1} = \frac{-f_L}{r_{13}(X_W - X_{CL}) + r_{23}(Y_W - Y_{CL}) + r_{33}(Z_W - Z_{CL})}$$

STEREOVISION

The 3D reconstruction solution:

Replacing μ^1 in the first two equations of (2), the following relations are obtained

The projection equations of P in the left camera coordinate system:

$$\begin{cases} x_L - x_{CL} = -f_L \dfrac{r_{11}(X_W - X_{CL}) + r_{21}(Y_W - Y_{CL}) + r_{31}(Z_W - Z_{CL})}{r_{13}(X_W - X_{CL}) + r_{23}(Y_W - Y_{CL}) + r_{33}(Z_W - Z_{CL})} \\ y_L - y_{CL} = -f_L \dfrac{r_{12}(X_W - X_{CL}) + r_{22}(Y_W - Y_{CL}) + r_{32}(Z_W - Z_{CL})}{r_{13}(X_W - X_{CL}) + r_{23}(Y_W - Y_{CL}) + r_{33}(Z_W - Z_{CL})} \end{cases} \tag{3}$$

The projection equations of P in the right camera coordinate system:

$$\begin{cases} x_R - x_{CR} = -f_R \dfrac{r'_{11}(X_W - X_{CR}) + r'_{21}(Y_W - Y_{CR}) + r'_{31}(Z_W - Z_{CR})}{r'_{13}(X_W - X_{CR}) + r'_{23}(Y_W - Y_{CR}) + r'_{33}(Z_W - Z_{CR})} \\ y_R - y_{CR} = -f_R \dfrac{r'_{12}(X_W - X_{CR}) + r'_{22}(Y_W - Y_{CR}) + r'_{32}(Z_W - Z_{CR})}{r'_{13}(X_W - X_{CR}) + r'_{23}(Y_W - Y_{CR}) + r'_{33}(Z_W - Z_{CR})} \end{cases} \tag{4}$$

Notation:

$$\begin{cases} x'_L = x_L - x_{CL} \\ y'_L = y_L - y_{CL} \\ x'_R = x_R - x_{CR} \\ y'_R = y_R - y_{CR} \end{cases}$$

STEREOVISION

The 3D reconstruction solution:

$$(3) + (4) \Rightarrow \qquad \mathbf{A} * \begin{vmatrix} X_W \\ Y_W \\ Z_W \end{vmatrix} = \mathbf{B} \quad \Leftrightarrow \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{vmatrix} * \begin{vmatrix} X_W \\ Y_W \\ Z_W \end{vmatrix} = \begin{vmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{vmatrix} \qquad (5)$$

where:

$$\mathbf{A} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{vmatrix} = \begin{vmatrix} r_{13}x'_L + r_{11} f_L & r_{23}x'_L + r_{21} f_L & r_{33}x'_L + r_{31} f_L \\ r_{13}y'_L + r_{12} f_L & r_{23}y'_L + r_{22} f_L & r_{33}y'_L + r_{32} f_L \\ r'_{13}x'_R + r'_{11} f_R & r'_{23}x'_R + r'_{21} f_R & r'_{33}x'_R + r'_{31} f_R \\ r'_{13}y'_R + r'_{12} f_R & r'_{23}y'_R + r'_{22} f_R & r'_{33}y'_R + r'_{32} f_R \end{vmatrix}$$

$$\mathbf{B} = \begin{vmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{vmatrix} = \begin{vmatrix} a_{11}X_{CL} + a_{12}Y_{CL} + a_{13}Z_{CL} \\ a_{21}X_{CL} + a_{22}Y_{CL} + a_{23}Z_{CL} \\ a_{31}X_{CR} + a_{32}Y_{CR} + a_{33}Z_{CR} \\ a_{41}X_{CR} + a_{42}Y_{CR} + a_{43}Z_{CR} \end{vmatrix}$$

The algebraic solution for system (5) can be obtained using the least squares method:

$$\mathbf{X} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{B}$$

Geometric meaning of the solution → the middle of the shortest segment connecting the two projection lines

The stereo-correlation problem

For an given left image point p_L a right image point p_R must be find so that the pair (p_L, p_R) represents the projections of the same 3D point P on the two image planes.

There are two major approaches:

-edge based reconstruction (sparse reconstruction) : only the highly relevant features like as edges are used

-dense reconstruction: all correlated pixels are reconstructed

The search space reduction is essential for real time software or hardware implementations. Is based on epipolar geometry constraints (epipolar lines). Has benefic effects also on reconstruction accuracy through diminishing of the false positive correspondences.

STEREOVISION

Basics of epipolar geometry

- Epipole (baseline intersection with the image plane) ⇒ one epipole / each image (is the intersection of all epipolar lines of the image)
- Epipolar plane ⇒ one plane for each 3D point
- Epipolar line ⇒ one line for each 3D point

Epipolar geometry constraint

- For every image point (x_L, y_L) there is a unique epipolar line (e_R) in the right image which will contain the corresponding right projection point (x_R, y_R) and vice-versa

Computing the epipolar lines [Trucco] :

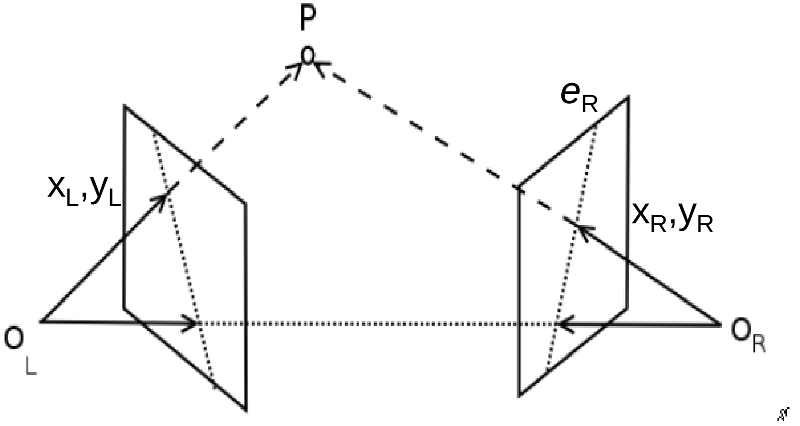
$$a * x + b * y + c = 0$$

$$\begin{vmatrix} a_R \\ b_R \\ c_R \end{vmatrix} = \mathbf{F} * \mathbf{P}_L \quad ; \quad \begin{vmatrix} a_L \\ b_L \\ c_L \end{vmatrix} = \mathbf{F}^T * \mathbf{P}_R$$

where:

F – fundamental matrix

$$\mathbf{P}_L = \begin{vmatrix} x_L \\ y_L \\ 1 \end{vmatrix} \quad ; \quad \mathbf{P}_R = \begin{vmatrix} x_R \\ y_R \\ 1 \end{vmatrix}$$



STEREOVISION

Computing the fundamental (F) and essential (E) matrices

$$\mathbf{F} = \left(\mathbf{A}_R^{-1}\right)^T * \mathbf{E} * \mathbf{A}_L^{-1}$$

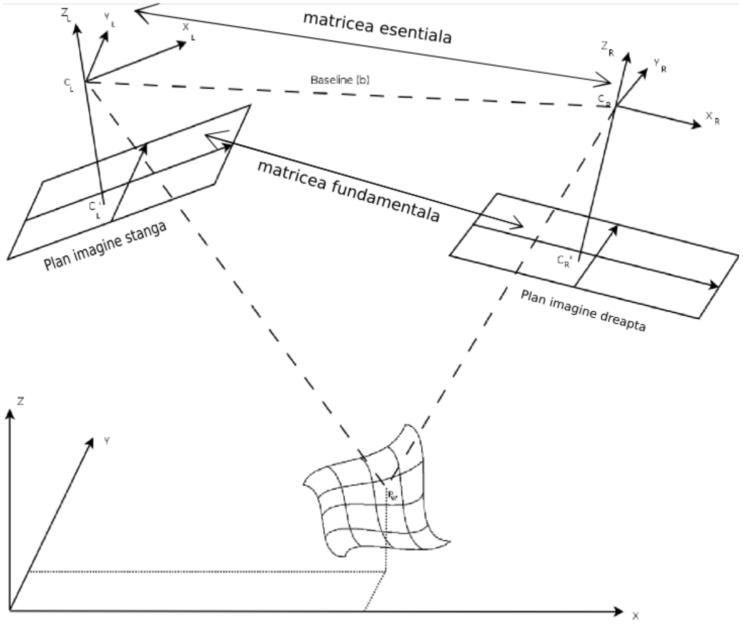
where:

$$\mathbf{E} = \mathbf{R}_{LR} * \mathbf{S}$$

$$\mathbf{R}_{LR} = \mathbf{R}_{CW-R}^T * \mathbf{R}_{CW-L}$$

$$\mathbf{T}_{LR} = \mathbf{T}_{CW-R} - \mathbf{T}_{CW-L} = \begin{bmatrix} T_X & T_Y & T_Z \end{bmatrix}^T$$

$$\mathbf{S} = \begin{bmatrix} 0 & -T_z & T_y \\ T_z & 0 & -T_x \\ -T_y & T_x & 0 \end{bmatrix}$$



$\mathbf{E}$ – essential matrix

$\mathbf{R}_{LR}$ – relative left-to-right rotation matrix

$\mathbf{T}_{LR}$ – relative left-to-right rotation matrix

A_L, A_R – the internal matrices of the two cameras

STEREOVISION

The stereo-correlation

a. *Features selection*

- - low level features: pixels.
- high level features: edge segments

b. *Features matching*

- use of epipolar geometry constraints (epipolar lines) for search space reduction

- for a left image point, a right correspondent point must be chosen out of a set of candidates. - the correlation function is the measure used for discrimination.

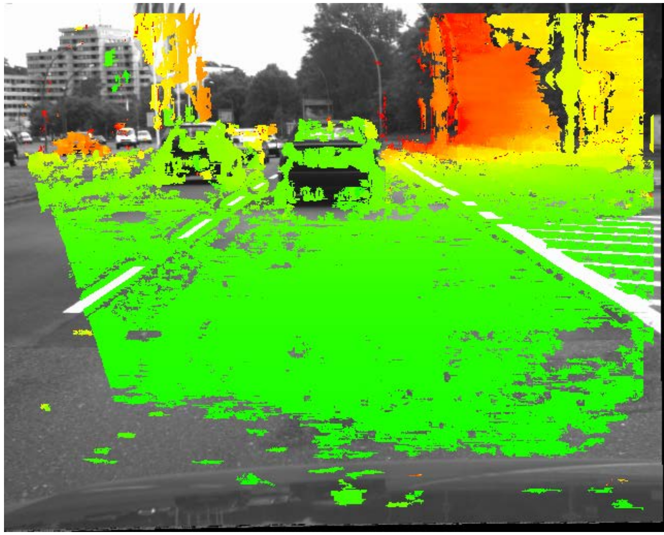
c. *Increasing the resolution of the correlation*

- compute the disparity with sub-pixel accuracy ⇒ far range & high accuracy stereo-vision

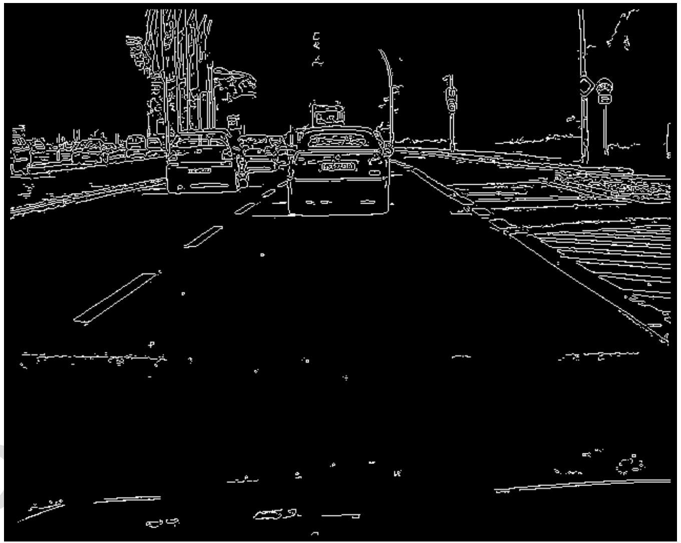
STEREOVISION

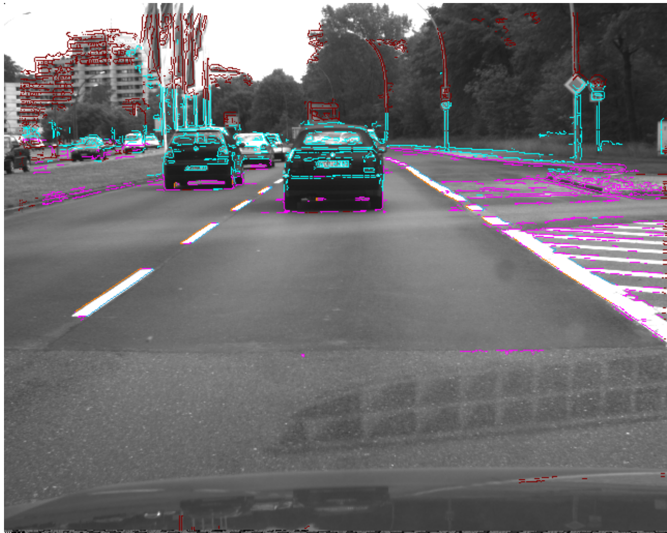
Features selection

Low level features - each image pixel is reconstructed (dense stereo)



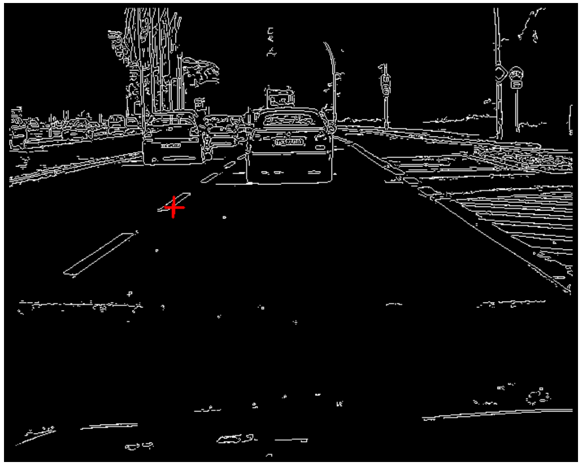
High level features – only edge pixels are reconstructed (edge based stereo)



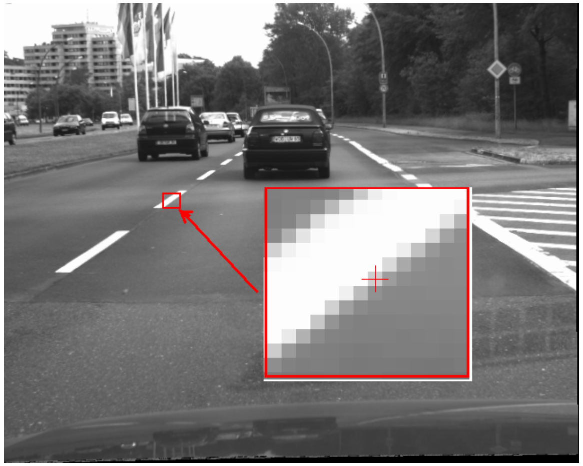


STEREOVISION

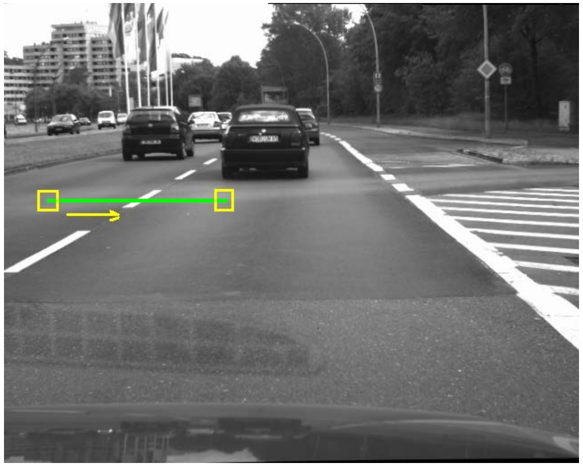
Features matching



Edge feature in left image

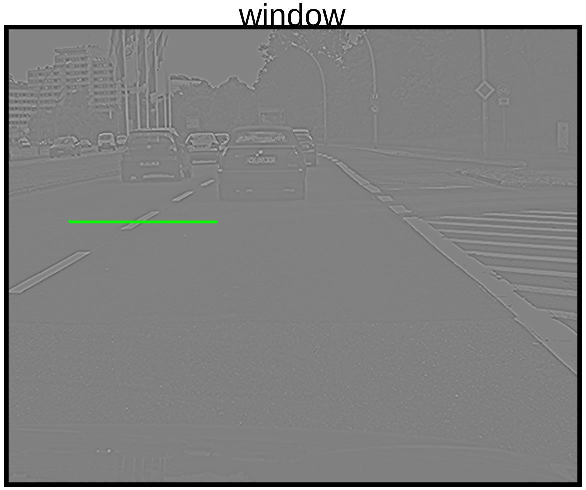


Grayscale left image correlation window



The correlation window slides along the epipolar line in the right image window

LoG left image correlation window

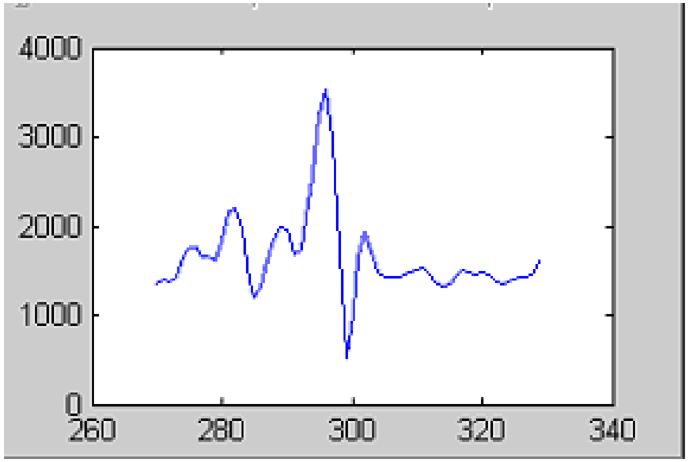


STEREOVISION

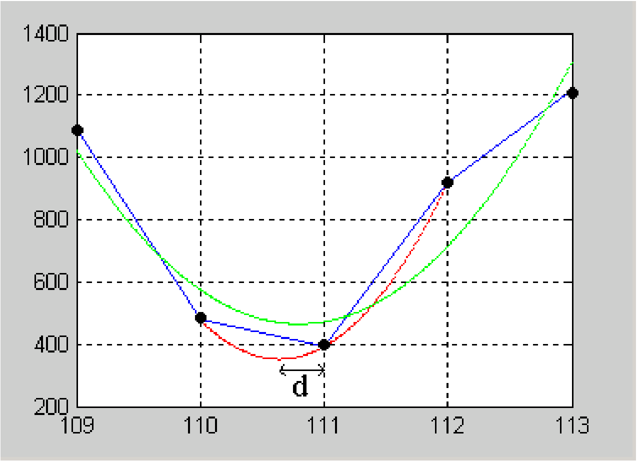
The correlation function

Any distance measure function SAD, SSD, normalized correlation

$$SAD(x_R, y_R) = \sum_{i=-\frac{w}{2}}^{\frac{w}{2}} \sum_{j=-\frac{w}{2}}^{\frac{w}{2}} \left| I_L(x_L + i, y_L + j) - I_R(x_R + i, y_R + j) \right|$$



Global minima of the correlation function



Detection of the sub-pixel position of global minima of the correlation function using a parabolic interpolator