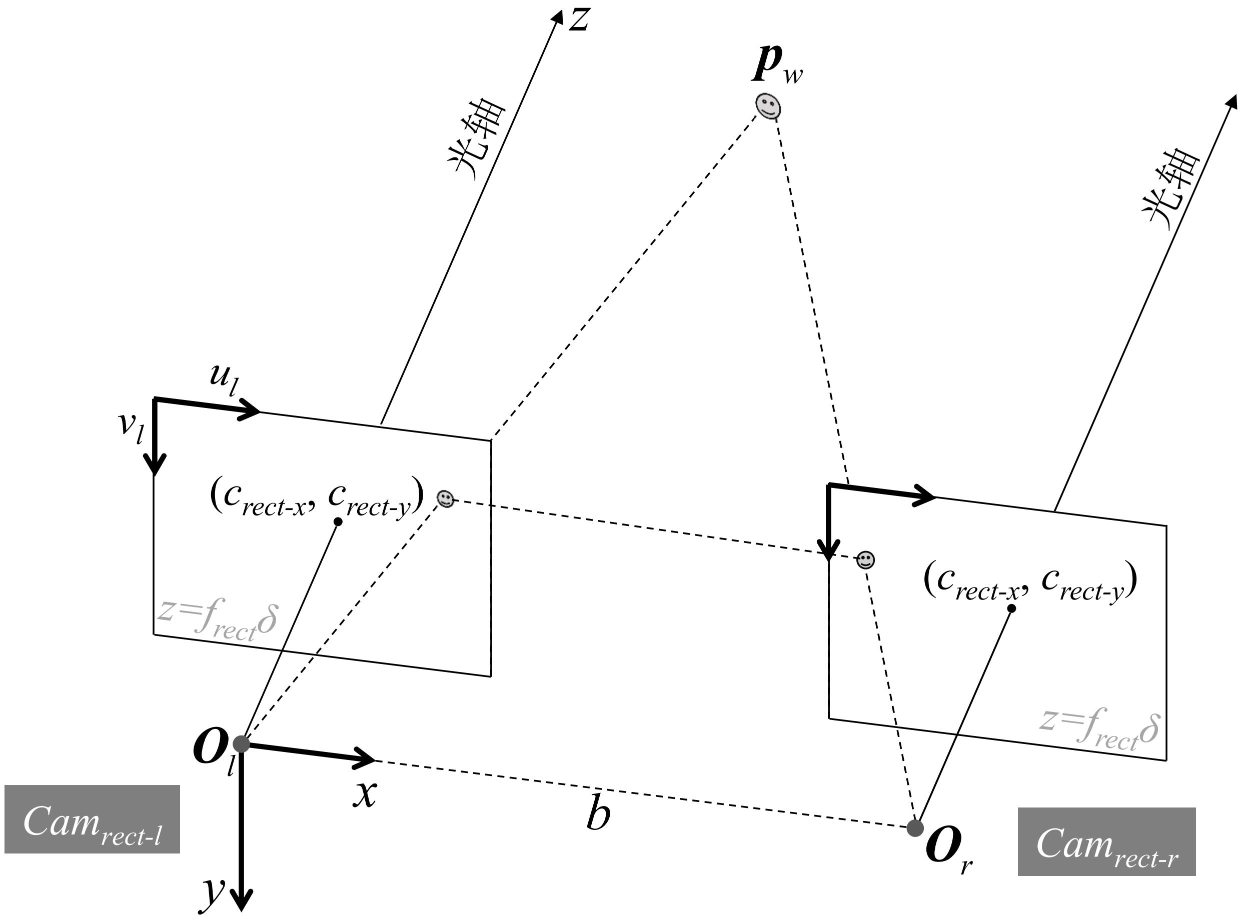
Chapter 17
Binocular Stereo Vision

Lin ZHANG
School of Computer Science and Technology
Tongji University

Outline

- Rectified binocular system
- Calibration of a physical binocular system
- Stereo matching and disparity map computation
- 3D reconstruction based on the disparity map

Rectified binocular system



- f_{rect} is the focal length (unit: pixel) of the intrinsics matrix
- δ is the physical length of each pixel (unit: *mm/pixel*)
- $(c_{rect\text{-}x}, c_{rect\text{-}y})$ is the position of the principal point on the imaging plane (unit: pixel)
- b (unit: *mm*) is the distance between the two camera centers

Rectified binocular system

- Rectified binocular system is a "virtual" system; it can simplify the depth estimation
- By moving $\boldsymbol{O}_l$-xyz $b(mm)$ along the x-axis, we get the $\boldsymbol{O}_r$-xyz coordinate system
- $Cam_{rect\text{-}l}$ and $Cam_{rect\text{-}r}$ have the same intrinsics matrix,

$$\boldsymbol{K}_{rect}=\begin{bmatrix} f_{rect} & 0 & c_{rect-x} \\ 0 & f_{rect} & c_{rect-y} \\ 0 & 0 & 1 \end{bmatrix}$$

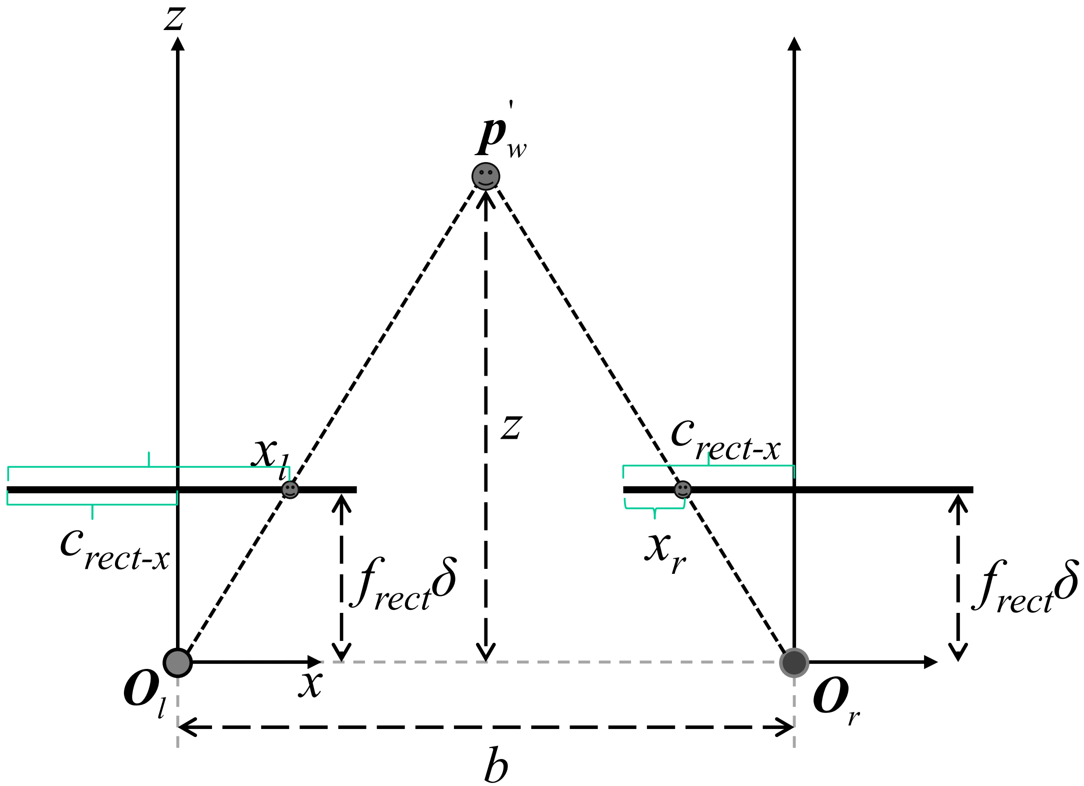
$$\Downarrow$$

The two images $\boldsymbol{I}_{rect\text{-}l}$ and $\boldsymbol{I}_{rect\text{-}r}$ of a rectified binocular system are row-aligned. That means, if $\boldsymbol{u}_l(u_l, v_l)$ and $\boldsymbol{u}_r(u_r, v_r)$ are the images of the same spatial point on $\boldsymbol{I}_{rect\text{-}l}$ and $\boldsymbol{I}_{rect\text{-}r}$, we should have $v_l=v_r$

Rectified binocular system

- Depth calculation under the rectified binocular system
  - With the rectified binocular system, given a spatial point $\boldsymbol{p}_w$, it is easy to obtain its depth w.r.t to the left camera, i.e., $\boldsymbol{p}_w$'s z-value with respect to the coordinate system $\boldsymbol{O}_l$-xyz
  - Suppose that $\boldsymbol{p}_w$'s image on $\boldsymbol{I}_{rect\text{-}l}$ is (x_l, y), and its image on $\boldsymbol{I}_{rect\text{-}r}$ is (x_r, y)
  - Projecting all the elements on the plane $\boldsymbol{O}_l$-xz. On $\boldsymbol{O}_l$-xz, the projection of $\boldsymbol{p}_w$ is $\boldsymbol{p}_w^{'}$. Obviously, $\boldsymbol{p}_w^{'}$'s z-value w.r.t to the coordinate system $\boldsymbol{O}_l$-xz is $\boldsymbol{p}_w$'s z-value with respect to the coordinate system $\boldsymbol{O}_l$-xyz

Rectified binocular system



The physical distance between two image pixels (x_l, y) and (x_r, y) is,

$$b-\left[\left(x_l-c_{rect-x}\right)+\left(c_{rect-x}-x_r\right)\right]\delta=b-\left(x_l-x_r\right)\delta$$

Then we have,

$$\frac{z-f_{rect}\delta}{b-\left(x_l-x_r\right)\delta}=\frac{z}{b} \Rightarrow z=\frac{f_{rect}b}{\boxed{x_l-x_r}}$$

Disparity of $\boldsymbol{p}_w$

$\Downarrow$

To get the depth map associated to $\boldsymbol{I}_{rect\text{-}l}$, we need to get the disparity map of $\boldsymbol{I}_{rect\text{-}l}$

With the rectified binocular system, to compute the depth of a 3D point with respect to the camera, only its disparity needs to be known

Rectified binocular system

As we have said, the rectified binocular system is a "virtual" system

Given a real physical binocular system, is there any way to get its associated rectified version?

We need to perform calibration to the physical binocular system; with the extrinsics and the intrinsics known, its rectified version can be derived

Outline

- Rectified binocular system
- Calibration of a physical binocular system
- Stereo matching and disparity map computation
- 3D reconstruction based on the disparity map

Calibration of a physical binocular system

Consider a physical binocular system. The left camera is cam_l, and its camera coordinate system is C_l; the right camera is cam_r, and its camera coordinate system is C_r.

The extrinsics of such a physical binocular system are,

$$\boldsymbol{R} \in \mathbb{R}^{3\times3} \left(\boldsymbol{R}\boldsymbol{R}^T = \boldsymbol{I}, \det(\boldsymbol{R}) = 1 \right), \boldsymbol{t} \in \mathbb{R}^{3\times1}$$

For a 3D point, if its coordinate w.r.t. to C_l is $\boldsymbol{p}_l$ and its coordinate w.r.t. C_r is $\boldsymbol{p}_r$, then $\boldsymbol{p}_l$ and $\boldsymbol{p}_r$ should satisfy,

$$\boldsymbol{p}_r = \boldsymbol{R}\boldsymbol{p}_l + \boldsymbol{t}$$

By calibration, our purpose is to get $\boldsymbol{R}$ and $\boldsymbol{t}$

Calibration of a physical binocular system

For binocular system calibration, we also need to resort to the checkerboard

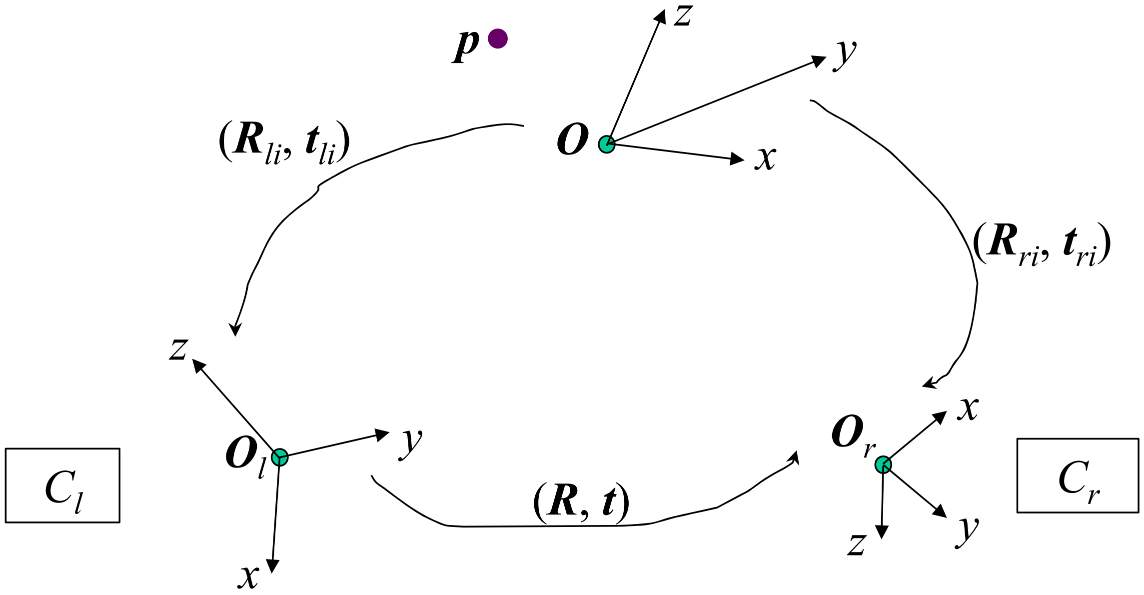
Collect m pairs of binocular images

$$\{\boldsymbol{I}_{li}, \boldsymbol{I}_{ri}\}_{i=1}^{m}$$

the left image　　　　the right image

A pair of binocular images

Calibration of a physical binocular system



$$\boldsymbol{p}_l = \boldsymbol{R}_{li}\boldsymbol{p} + \boldsymbol{t}_{li} \quad (1)$$

$$\Downarrow$$

$$\boldsymbol{p} = \boldsymbol{R}_{li}^{-1}\left(\boldsymbol{p}_l - \boldsymbol{t}_{li}\right) \quad (4)$$

$$\boldsymbol{p}_r = \boldsymbol{R}_{ri}\boldsymbol{p} + \boldsymbol{t}_{ri} \quad (2)$$

$$\boldsymbol{p}_r = \boldsymbol{R}\boldsymbol{p}_l + \boldsymbol{t} \quad (3)$$

$$\begin{aligned}\boldsymbol{p}_r &= \boldsymbol{R}_{ri}\left(\boldsymbol{R}_{li}^{-1}\left(\boldsymbol{p}_l - \boldsymbol{t}_{li}\right)\right) + \boldsymbol{t}_{ri} \\ &= \boldsymbol{R}_{ri}\boldsymbol{R}_{li}^{-1}\boldsymbol{p}_l + \left(\boldsymbol{t}_{ri} - \boldsymbol{R}_{ri}\boldsymbol{R}_{li}^{-1}\boldsymbol{t}_{li}\right) \quad (5)\end{aligned}$$

By comparing (3) and (5), we have

$$\boldsymbol{R} = \boldsymbol{R}_{ri}\boldsymbol{R}_{li}^{-1}, \quad \boldsymbol{t} = \boldsymbol{t}_{ri} - \boldsymbol{R}_{ri}\boldsymbol{R}_{li}^{-1}\boldsymbol{t}_{li}$$

$$\Downarrow$$

$$\boxed{\boldsymbol{R}_{ri} = \boldsymbol{R}\boldsymbol{R}_{li}, \quad \boldsymbol{t}_{ri} = \boldsymbol{R}\boldsymbol{t}_{li} + \boldsymbol{t} \quad (6)}$$

Calibration of a physical binocular system

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\left\{\left\|\boldsymbol{K}_l\mathcal{D}_l\left\{\frac{1}{z_{clij}}\left[\boldsymbol{R}_{li}\ \ \boldsymbol{t}_{li}\right]_{3\times4}\boldsymbol{p}_j\right\}-\boldsymbol{u}_{lij}\right\|_2^2+\left\|\boldsymbol{K}_r\mathcal{D}_r\left\{\frac{1}{z_{crij}}\left[\boldsymbol{R}_{ri}\ \ \boldsymbol{t}_{ri}\right]_{3\times4}\boldsymbol{p}_j\right\}-\boldsymbol{u}_{rij}\right\|_2^2\right\}$$

⬇

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\left\{\left\|\boldsymbol{K}_l\mathcal{D}_l\left\{\frac{1}{z_{clij}}\left[\mathcal{R}(\boldsymbol{d}_{li})\ \ \boldsymbol{t}_{li}\right]_{3\times4}\boldsymbol{p}_j\right\}-\boldsymbol{u}_{lij}\right\|_2^2+\left\|\boldsymbol{K}_r\mathcal{D}_r\left\{\frac{1}{z_{crij}}\left[\mathcal{R}(\boldsymbol{d})\mathcal{R}(\boldsymbol{d}_{li})\ \ \mathcal{R}(\boldsymbol{d})\boldsymbol{t}_{li}+\boldsymbol{t}\right]_{3\times4}\boldsymbol{p}_j\right\}-\boldsymbol{u}_{rij}\right\|_2^2\right\}$$

where $\boldsymbol{\theta} = \left(\boldsymbol{d}, \boldsymbol{t}, \{\boldsymbol{d}_{li}\}_{i=1}^{m}, \{\boldsymbol{t}_{li}\}_{i=1}^{m}\right)$ are the parameters that need to be optimized

$\boldsymbol{u}_{lij}$ is the projection of $\boldsymbol{p}_j$ onto the left image of the ith binocular image pair

$\boldsymbol{u}_{rij}$ is the projection of $\boldsymbol{p}_j$ onto the right image of the ith binocular image pair

$\left(\mathcal{R}(\boldsymbol{d}_{li}), \boldsymbol{t}_{li}\right)$ are the extrinsics of the left camera w.r.t. to the world coordinate system when taking the ith binocular image pair

Calibration of a physical binocular system

$$\boldsymbol{f}_{ij}(\boldsymbol{\theta}) \triangleq \begin{pmatrix} \boldsymbol{K}_l \mathcal{D}_l \left\{ \frac{1}{z_{clij}} \left[\mathcal{R}(\boldsymbol{d}_{li}) \;\; \boldsymbol{t}_{li} \right]_{3\times4} \boldsymbol{p}_j \right\} - \boldsymbol{u}_{lij} \\ \boldsymbol{K}_r \mathcal{D}_r \left\{ \frac{1}{z_{crij}} \left[\mathcal{R}(\boldsymbol{d}) \mathcal{R}(\boldsymbol{d}_{li}) \;\; \mathcal{R}(\boldsymbol{d}) \mathbf{t}_{li} + \boldsymbol{t} \right]_{3\times4} \boldsymbol{p}_j \right\} - \boldsymbol{u}_{rij} \end{pmatrix}_{4\times1}$$

$$\Downarrow$$

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \left(\frac{1}{2} \boldsymbol{f}^T(\boldsymbol{\theta}) \boldsymbol{f}(\boldsymbol{\theta}) \right)$$

which is a typical nonlinear least-squares problem and can be used by using techniques introduced in Chapter 9

Outline

- Rectified binocular system
- Calibration of a physical binocular system
- Stereo matching and disparity map computation
- 3D reconstruction based on the disparity map

Stereo matching and disparity map computation

When the extrinsics and the intrinsics of a physical binocular system are known, its associated rectified binocular system can be constructed

To get the corresponding depth map to the left rectified image $\boldsymbol{I}_{rect\text{-}l}$, we need to at first extract its disparity map

⬇

We need to know the matching relationships among pixels on $\boldsymbol{I}_{rect\text{-}l}$ and $\boldsymbol{I}_{rect\text{-}r}$

For a point $\boldsymbol{u}_l(x_l, y_0)$ on $\boldsymbol{I}_{rect\text{-}l}$, how to find the matching pixel $\boldsymbol{u}_r$ on $\boldsymbol{I}_{rect\text{-}r}$?

Since $\boldsymbol{I}_{rect\text{-}l}$ and $\boldsymbol{I}_{rect\text{-}r}$ are row-aligned, $\boldsymbol{u}_r$ can only happen on the y_0 row of $\boldsymbol{I}_{rect\text{-}r}$

Stereo matching and disparity map computation

On $\boldsymbol{I}_{rect\text{-}l}$, take a patch $\boldsymbol{A}$ round $\boldsymbol{u}_l$. Traverse each position of y_0 row on $\boldsymbol{I}_{rect\text{-}r}$. For the position (x_i, y_0) on $\boldsymbol{I}_{rect\text{-}r}$, take a patch $\boldsymbol{B}_i$ around (x_i, y_0)

Figuring out the position i^* whose local patch has the least distance to $\boldsymbol{A}$,

$$i^* = \underset{i=1,2,\ldots,n}{\arg\min}\, d\left(\boldsymbol{A}, \boldsymbol{B}_i\right)$$

$\left(x_{i^*}, y_0\right)$ is $\boldsymbol{u}_r$ we want to find, i.e., $\boldsymbol{u}_r = \left(x_{i^*}, y_0\right)$

Note: in implementation, usually the matching is not conducted on the raw images but on the partial derivative images, for example, the results of Sobel operators; there are also some algorithms to postprocess the raw disparity map to make it more robust, such as the weighted least-squares method[1]

[1] MIN D, CHOI S, LU J, et al. Fast global image smoothing based on weighted least squares[J]. IEEE Trans. Image Processing, 2014, 23(12): 5638-5653.

Stereo matching and disparity map computation

The raw disparity map

The disparity map with weighted least-squares postprocessing

Outline

- Rectified binocular system
- Calibration of a physical binocular system
- Stereo matching and disparity map computation
- 3D reconstruction based on the disparity map

3D reconstruction based on the disparity map

With the rectified binocular system, suppose $\boldsymbol{I}_{rect\text{-}l}$ is its left image and $\boldsymbol{u}(x, y)$ is one pixel on $\boldsymbol{I}_{rect\text{-}l}$ and its disparity is known as d. The 3D point in space corresponding to $\boldsymbol{u}$ is $\boldsymbol{p}$. Now let's compute the coordinate of $\boldsymbol{p}$ w.r.t. the coordinate system $C_{rect\text{-}l}$

Suppose $\boldsymbol{x}_n=(x_n, y_n, 1)^T$ is the point on the normalized retinal plane of $Cam_{rect\text{-}l}$ corresponding to $\boldsymbol{u}$. We can have

$$\begin{bmatrix} f_{rect} & 0 & c_{rect-x} \\ 0 & f_{rect} & c_{rect-y} \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_n \\ y_n \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \Rightarrow \begin{pmatrix} x_n \\ y_n \\ 1 \end{pmatrix} = \begin{pmatrix} \dfrac{x-c_{rect-x}}{f_{rect}} \\ \dfrac{y-c_{rect-y}}{f_{rect}} \\ 1 \end{pmatrix}$$

$\boldsymbol{p}$'s depth value (z value) w.r.t. $C_{rect\text{-}l}$ is $z = \dfrac{f_{rect}}{d} b$

$\Downarrow$

3D reconstruction based on the disparity map

$\boldsymbol{p}$'s coordinate w.r.t. $C_{rect\text{-}l}$ is,

$$z\boldsymbol{x}_n = \frac{f_{rect} b}{d} \cdot \begin{pmatrix} \dfrac{x - c_{rect-x}}{f_{rect}} \\ \dfrac{y - c_{rect-y}}{f_{rect}} \\ 1 \end{pmatrix} = \begin{pmatrix} \left(x - c_{rect-x}\right)\dfrac{b}{d} \\ \left(y - c_{rect-y}\right)\dfrac{b}{d} \\ f_{rect}\dfrac{b}{d} \end{pmatrix}$$

The homogeneous form of $\boldsymbol{p}$'s position w.r.t. $C_{rect\text{-}l}$ is,

$$\left(\left(x - c_{rect-x}\right)\frac{b}{d}, \left(y - c_{rect-y}\right)\frac{b}{d}, f_{rect}\frac{b}{d}, 1\right)^T \quad \text{or} \quad \xrightarrow{\times \frac{d}{b}} \quad \left(\left(x - c_{rect-x}\right), \left(y - c_{rect-y}\right), f_{rect}, \frac{d}{b}\right)^T$$

Notice that,

$$\begin{pmatrix} x - c_{rect-x} \\ y - c_{rect-y} \\ f_{rect} \\ \dfrac{d}{b} \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 & -c_{rect-x} \\ 0 & 1 & 0 & -c_{rect-y} \\ 0 & 0 & 0 & f_{rect} \\ 0 & 0 & \dfrac{1}{b} & 0 \end{bmatrix} \begin{pmatrix} x \\ y \\ d \\ 1 \end{pmatrix}$$

Denote $\boldsymbol{Q} \triangleq \begin{bmatrix} 1 & 0 & 0 & -c_{rect-x} \\ 0 & 1 & 0 & -c_{rect-y} \\ 0 & 0 & 0 & f_{rect} \\ 0 & 0 & \dfrac{1}{b} & 0 \end{bmatrix}$ $\Rightarrow$

3D reconstruction based on the disparity map

$$\boldsymbol{Q}\begin{pmatrix} x \\ y \\ d \\ 1 \end{pmatrix} \triangleq \begin{pmatrix} x_h \\ y_h \\ z_h \\ w_h \end{pmatrix}$$

is the 3D point w.r.t $C_{rect\text{-}l}$ corresponding to the pixel $\boldsymbol{u}$ of the left rectified image $\boldsymbol{I}_{rect\text{-}l}$

Its normal 3D coordinate is,

$$\begin{pmatrix} \frac{x_h}{w_h} \\ \frac{y_h}{w_h} \\ \frac{z_h}{w_h} \end{pmatrix}$$

For any pixel $\boldsymbol{u}(x, y)$ on the left image $\boldsymbol{I}_{rect\text{-}l}$ of the rectified binocular system, if know its disparity d, then, the spatial point $\boldsymbol{p}$ (corresponding to $\boldsymbol{u}$)'s coordinate w.r.t. the camera coordinate system of the left rectified camera is,

$$\boldsymbol{Q}(x, y, d, 1)^T$$

3D reconstruction based on the disparity map

An example

Original image pair

Rectified image pair (they are row-aligned)

raw depth map

depth map with weighted least-squares postprocessing

3D reconstruction based on the disparity map

An example

The RGB-D point cloud


Q+A