MATHEMATIC MODELISATION OF A DEPTH CAMERA

BOYER Apolline

August 13, 2024

Contents

1 Introduction

The objective is to do a mathematic modelisation of a depth camera. This depth camera will use a binolucar camera which permits to make a stereo vision of an image. A Time-of-Flight sensor is also used.

First, we will see the modelisation of a binocular system. Then, we will see the modelisation of a Time-of-Flight sensor.

2 The Binocular System

Binocular vision, inspired by the human ability to perceive depth and the three-dimensional structure of the world, uses two cameras to capture images of the same scene from slightly different angles. By exploiting disparity, which is the difference in the position of objects in the two images, it is possible to calculate the distance of objects relative to the cameras and thereby reconstruct a three-dimensional scene.

2.1 Camera Modeling

A camera is modeled using the pinhole model. It allows the conversion from a 3D image to a 2D image.

This is done through three transformations. The first transformation consists of converting from a 3D reference frame to the camera reference frame. The second transformation allows for a projection onto the image. Finally, the third transformation converts dimensions from mm to pixels.The two last are rerouped in one tranformation.

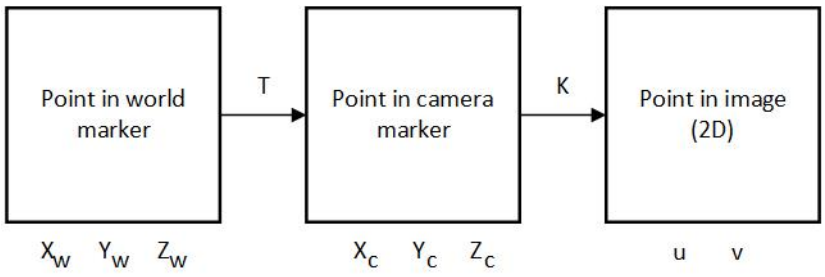

Figure 1: Schematic of different transformations

First, we take a point M in what is called the world coordinate system. M has coordinates (X_W, Y_W, Z_W). The point C is the projection of M (X_C, Y_C, Z_C) in the camera reference frame. The point m (u, v) is the projection of C in the image reference frame.

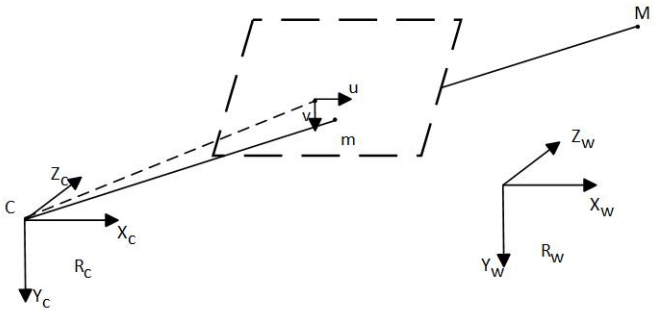

Figure 2: Schematic for transformation

First of all, we can see the transformation from M to C as a combination of rotations and translations, all in homogeneous coordinates.

The rotation matrices around the three principal axes are defined as follows:

Rotation around the X-axis (angle α) :

$$R_x(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}$$

Rotation around the Y-axis (angle β) :

$$R_y(\beta) = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix}$$

Rotation around the Z-axis (angle γ) :

$$R_z(\gamma) = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The translation vector is:

$$T(T_x, T_y, T_z) = \begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}$$

In homogeneous coordinates, rotations and translation can be combined into a single transformation matrix T.

The homogeneous transformation matrix T is obtained by combining the rotation matrices and the translation vector:

$$T = \begin{pmatrix} R & \mathbf{t} \\ 0 & 1 \end{pmatrix} \tag{1}$$

Where R is the combined rotation matrix and $\mathbf{t}$ is the translation vector.

The combined rotation matrix can be obtained by multiplying the individual rotation matrices (according to the desired order of rotations):

$$R = R_z(\gamma)R_y(\beta)R_x(\alpha)$$

Thus, the complete homogeneous transformation matrix T is:

$$T = \begin{pmatrix} \cos\beta\cos\gamma & -\cos\beta\sin\gamma & \sin\beta & T_x \\ \sin\alpha\sin\beta\cos\gamma + \cos\alpha\sin\gamma & -\sin\alpha\sin\beta\sin\gamma + \cos\alpha\cos\gamma & -\sin\alpha\cos\beta & T_y \\ -\cos\alpha\sin\beta\cos\gamma + \sin\alpha\sin\gamma & \cos\alpha\sin\beta\sin\gamma + \sin\alpha\cos\gamma & \cos\alpha\cos\beta & T_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2}$$

The matrix T corresponds to the extrinsic parameters of the camera.

The intrinsic parameters of a camera are used to transform the 3D coordinates in the camera reference frame into 2D coordinates in the image reference frame. These parameters include the focal lengths f_x and f_y, as well as the principal point coordinates c_x and c_y.

The transformation of camera coordinates (X_C, Y_C, Z_C) to image coordinates (u, v) is carried out in three steps:

1. **Perspective projection**: This transformation projects the 3D coordinates onto an image plane.
2. **Scaling**: This transformation converts the units of the image coordinates from mm to pixels.
3. **Translation**: This transformation adjusts the image coordinates to account for the principal point of the image.

The intrinsic projection matrix K is defined as follows:

$$K = \begin{pmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{3}$$

Thus, for a point C in the camera reference frame with homogeneous coordinates $(X_C, Y_C, Z_C, 1)$, the homogeneous coordinates of the projected point m in the image reference frame are obtained by:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = K \begin{pmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{pmatrix} \tag{4}$$

By combining the extrinsic and intrinsic transformations, the complete projection matrix M of the camera is:

$$M = K \cdot T = \begin{pmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{5}$$

2.2 Camera Calibration

The objective of calibration is to find the values of the elements m_{ij} of the matrix M that minimize the projection error between the 3D coordinates of the real world and the observed 2D coordinates of the image. This can be achieved using optimization algorithms such as the least squares method.

First, we need to determine the system of equations between the different parameters. We have:

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = M \begin{pmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{pmatrix}$$

where (X_W, Y_W, Z_W) are the 3D coordinates of the point in the world coordinate system, (u, v) are the 2D coordinates of the image in pixels, s is a scale factor, and M is the projection matrix.

Divide the equations by s to obtain the normalized coordinates:

$$\begin{aligned} u &= \frac{m_{11}X_W + m_{12}Y_W + m_{13}Z_W + m_{14}}{s} \\ v &= \frac{m_{21}X_W + m_{22}Y_W + m_{23}Z_W + m_{24}}{s} \\ s &= m_{31}X_W + m_{32}Y_W + m_{33}Z_W + m_{34} \end{aligned} \tag{6}$$

By substituting s into the previous equations, the normalized coordinates (u, v) become:

$$\begin{aligned} u &= \frac{m_{11}X_W + m_{12}Y_W + m_{13}Z_W + m_{14}}{m_{31}X_W + m_{32}Y_W + m_{33}Z_W + m_{34}} \\ v &= \frac{m_{21}X_W + m_{22}Y_W + m_{23}Z_W + m_{24}}{m_{31}X_W + m_{32}Y_W + m_{33}Z_W + m_{34}} \end{aligned} \tag{7}$$

These equations represent the projection of a world point into an image point using the matrix M with its elements m_{ij}, considering the factor s.

We then have the following system of equations:

$$\begin{cases} 0 = m_{11}X_W + m_{12}Y_W + m_{13}Z_W + m_{14} - u \cdot m_{31}X_W - u \cdot m_{32}Y_W - u \cdot m_{33}Z_W - u \cdot m_{34} \\ 0 = m_{21}X_W + m_{22}Y_W + m_{23}Z_W + m_{24} - v \cdot m_{31}X_W - v \cdot m_{32}Y_W - v \cdot m_{33}Z_W - v \cdot m_{34} \end{cases} \tag{8}$$

To calibrate the camera, we need to solve this system of equations for multiple points (X_W, Y_W, Z_W) and their correspondences (u, v). Using N points, we obtain a system of $2N$ linear equations, which can be expressed in matrix form to be solved using the least squares method:

$$\begin{pmatrix} X_{W1} & Y_{W1} & Z_{W1} & 1 & 0 & 0 & 0 & 0 & -u_1X_{W1} & -u_1Y_{W1} \\ -u_1Z_{W1} & & & & & & & & & \\ 0 & 0 & 0 & 0 & X_{W1} & Y_{W1} & Z_{W1} & 1 & -v_1X_{W1} & -v_1Y_{W1} \\ -v_1Z_{W1} & & & & & & & & & \\ X_{W2} & Y_{W2} & Z_{W2} & 1 & 0 & 0 & 0 & 0 & -u_2X_{W2} & -u_2Y_{W2} \\ -u_2Z_{W2} & & & & & & & & & \\ 0 & 0 & 0 & 0 & X_{W2} & Y_{W2} & Z_{W2} & 1 & -v_2X_{W2} & -v_2Y_{W2} \\ -v_2Z_{W2} & & & & & & & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & & & & & & & & & \\ X_{WN} & Y_{WN} & Z_{WN} & 1 & 0 & 0 & 0 & 0 & -u_NX_{WN} & -u_NY_{WN} \\ -u_NZ_{WN} & & & & & & & & & \\ 0 & 0 & 0 & 0 & X_{WN} & Y_{WN} & Z_{WN} & 1 & -v_NX_{WN} & -v_NY_{WN} \\ -v_NZ_{WN} & & & & & & & & & \end{pmatrix} \begin{pmatrix} m_{11} \\ m_{12} \\ m_{13} \\ m_{14} \\ m_{21} \\ m_{22} \\ m_{23} \\ m_{24} \\ m_{31} \\ m_{32} \\ m_{33} \\ m_{34} \end{pmatrix} = \begin{pmatrix} u_1m_{34} \\ v_1m_{34} \\ u_2m_{34} \\ v_2m_{34} \\ \vdots \\ u_Nm_{34} \\ v_Nm_{34} \end{pmatrix}$$

The least squares method minimizes the sum of the squares of the residuals, which allows for an optimal solution even in the presence of noise or measurement errors.

To reduce complexity, we can set $q_{ij} \cdot m_{34} = m_{ij}$.
Thus, the system of equations becomes:

$$\begin{pmatrix} X_{W1} & Y_{W1} & Z_{W1} & 1 & 0 & 0 & 0 & 0 & -u_1X_{W1} & -u_1Y_{W1} \\ -u_1Z_{W1} & & & & & & & & & \\ 0 & 0 & 0 & 0 & X_{W1} & Y_{W1} & Z_{W1} & 1 & -v_1X_{W1} & -v_1Y_{W1} \\ -v_1Z_{W1} & & & & & & & & & \\ X_{W2} & Y_{W2} & Z_{W2} & 1 & 0 & 0 & 0 & 0 & -u_2X_{W2} & -u_2Y_{W2} \\ -u_2Z_{W2} & & & & & & & & & \\ 0 & 0 & 0 & 0 & X_{W2} & Y_{W2} & Z_{W2} & 1 & -v_2X_{W2} & -v_2Y_{W2} \\ -v_2Z_{W2} & & & & & & & & & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & & & & & & & & & \\ X_{WN} & Y_{WN} & Z_{WN} & 1 & 0 & 0 & 0 & 0 & -u_NX_{WN} & -u_NY_{WN} \\ -u_NZ_{WN} & & & & & & & & & \\ 0 & 0 & 0 & 0 & X_{WN} & Y_{WN} & Z_{WN} & 1 & -v_NX_{WN} & -v_NY_{WN} \\ -v_NZ_{WN} & & & & & & & & & \end{pmatrix} \begin{pmatrix} q_{11} \\ q_{12} \\ q_{13} \\ q_{14} \\ q_{21} \\ q_{22} \\ q_{23} \\ q_{24} \\ q_{31} \\ q_{32} \\ q_{33} \end{pmatrix} = \begin{pmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ \vdots \\ u_N \\ v_N \end{pmatrix}$$

We can rewrite this problem as:

$$U = K_{2N}.Q \tag{9}$$

Additionally, by observing the relationships in the rotation matrix, we can deduce the following relationship:

$$m_{31}^2 + m_{32}^2 + m_{33}^2 = 1 \tag{10}$$

From this, we can deduce the following relationship:

$$m_{34} = \sqrt{\frac{1}{q_{31}^2 + q_{32}^2 + q_{33}^2}} \tag{11}$$

Equation (9) is then solved using the least squares method. We obtain the result:

$$Q = (K_{2N}^t.K_{2N})^{-1} K_{2N}^t.U \tag{12}$$

Once the matrix Q is determined, and using equation (11), we can then determine the elements m_{ij}.

Next, we want to calculate the intrinsic and extrinsic parameters.

From the relation (4), we obtain the following system of equations:

$$\begin{aligned}
m_{11} &= f_x R_{11} + c_x R_{31} \\
m_{12} &= f_x R_{12} + c_x R_{32} \\
m_{13} &= f_x R_{13} + c_x R_{33} \\
m_{14} &= f_x T_x + c_x T_z \\
m_{21} &= f_y R_{21} + c_y R_{31} \\
m_{22} &= f_y R_{22} + c_y R_{32} \\
m_{23} &= f_y R_{23} + c_y R_{33} \\
m_{24} &= f_y T_y + c_y T_z \\
m_{31} &= R_{31} \\
m_{32} &= R_{32} \\
m_{33} &= R_{33} \\
m_{34} &= T_z
\end{aligned} \tag{13}$$

Let $m_i = (m_{i1}, m_{i2}, m_{i3})$.

The relationships to find all the elements are:

$$
\begin{aligned}
r_3 &= m_3 \\
c_x &= m_1 \cdot m_3 \\
c_y &= m_2 \cdot m_3 \\
f_x &= -\|m_1 \wedge m_3\| \\
f_y &= \|m_2 \wedge m_3\| \\
r_1 &= \frac{1}{f_x}(m_1 - c_x m_3) \\
r_2 &= \frac{1}{f_y}(m_2 - c_y m_3) \\
t_x &= \frac{1}{f_x}(m_{14} - c_x m_{34}) \\
t_y &= \frac{1}{f_y}(m_{24} - c_y m_{34}) \\
t_z &= m_{34}
\end{aligned}
\tag{14}
$$

2.3 Stereo Vision: Using Two Cameras

2.3.1 Principle of Stereo Vision

Stereo vision is a technique that uses two cameras placed at slightly different positions to capture a scene. By exploiting the differences in perspective between the two obtained images, it is possible to reconstruct the three-dimensional structure of the scene.

The basic principle of stereo vision is based on triangulation. By comparing the positions of the same point in the two images, we can determine its depth using the geometry of the triangles formed by the optical rays and the optical axes of the cameras.

Specifically, for each corresponding point in the two images, a line of sight is drawn from the optical center of each camera to the point. The 3D position of the point is then determined by the intersection of these two lines of sight. This process is repeated for several points in the image to reconstruct the scene in three dimensions.

2.3.2 Calculating the Transformation Matrix Between the Two Images

When using two cameras to capture a scene in stereo view, it is often necessary to calculate the transformation matrix between the two images. This matrix, usually called the transformation matrix or displacement matrix, aligns the corresponding points in the two images to facilitate comparison and three-dimensional reconstruction of the scene.

The transformation matrix is generally calculated using image registration or feature matching techniques. These techniques can include algorithms such as interest point matching, texture matching, or descriptor-based methods.

Once the corresponding points have been identified in the two images, the transformation matrix can be calculated from these correspondences. This matrix is essentially a geometric transformation that expresses the displacement and orientation needed to align the two images.

Let A_i be the projection matrix that transforms coordinates from the world frame to the camera frame.

$$A_i = \begin{pmatrix} R^i_{11} & R^i_{12} & R^i_{13} & t^i_x \\ R^i_{21} & R^i_{22} & R^i_{23} & t^i_y \\ R^i_{31} & R^i_{32} & R^i_{33} & t^i_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The projection matrices of the left and right cameras are denoted as A_L and A_R, respectively. These matrices can be defined as follows:

$$P_{Ci} = A_i \cdot P_W \tag{15}$$

where P_{Ci} represents the projected coordinates in the camera frame (left or right) and P_W represents the coordinates of the point in the world frame.

We can then deduce the matrix that transforms coordinates from the right camera frame to the left camera frame:

$$\begin{gathered} P_{CR} = A_s P_{CL} \\ \text{where } A_s = A_R \cdot A_L^{-1} \end{gathered} \tag{16}$$

We have

$$A_s = \begin{pmatrix} R_{11} & R_{12} & R_{13} & t_x \\ R_{21} & R_{22} & R_{23} & t_y \\ R_{31} & R_{32} & R_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

We define the two points as follows:

$$\mathbf{P}_{CR} = \begin{pmatrix} X_R \\ Y_R \\ Z_R \end{pmatrix}$$

$$\mathbf{P}_{CL} = \begin{pmatrix} X_L \\ Y_L \\ Z_L \end{pmatrix}$$

We then have the following system of equations:

$$\begin{aligned} X_R &= R_{11}X_L + R_{12}Y_L + R_{13}Z_L + t_x \\ Y_R &= R_{21}X_L + R_{22}Y_L + R_{23}Z_L + t_y \\ Z_R &= R_{31}X_L + R_{32}Y_L + R_{33}Z_L + t_z \end{aligned} \tag{17}$$

The goal is to eliminate Z_L. By setting $x_R = \frac{X_R}{Z_R}$ and $y_R = \frac{Y_R}{Z_R}$, we develop the following:

$$\begin{aligned} x_R = \frac{X_R}{Z_R} &= \frac{R_{11}X_L + R_{12}Y_L + R_{13}Z_L + t_x}{R_{31}X_L + R_{32}Y_L + R_{33}Z_L + t_z} \\ y_R = \frac{Y_R}{Z_R} &= \frac{R_{21}X_L + R_{22}Y_L + R_{23}Z_L + t_y}{R_{31}X_L + R_{32}Y_L + R_{33}Z_L + t_z} \end{aligned}$$

From these equations, we can express Z_L as follows:

$$\begin{aligned} Z_L &= \frac{t_x - x_R t_z}{x_R R_{31} x_L + x_R R_{32} y_L + x_R R_{33} - R_{11} x_L - R_{12} y_L - R_{13}} \\ Z_L &= \frac{t_y - y_R t_z}{y_R R_{31} x_L + y_R R_{32} y_L + y_R R_{33} - R_{21} x_L - R_{22} y_L - R_{23}} \end{aligned}$$

By setting these expressions equal, we obtain the following right equation:

$$\begin{aligned} & a.x_R + b.y_R + c = 0 \\ \text{where: } & a = (R_{31}t_Y - R_{21}t_Z)\,x_L + (R_{32}t_Y - R_{22}t_Z)\,y_L + R_{33}t_Y - R_{23}t_Z \\ & b = (R_{11}t_Z - R_{31}t_X)\,x_L + (R_{12}t_Z - R_{32}t_X)\,y_L + R_{13}t_Z - R_{33}t_X \\ & c = (R_{21}t_X - R_{11}t_Y)\,x_L + (R_{22}t_X - R_{12}t_Y)\,y_L + R_{23}t_X - R_{13}t_Y \end{aligned} \tag{18}$$

This line allows for estimating the correspondence between points in the right camera frame and the left camera frame. It is known as an epipolar line.

Epipolar geometry provides a crucial framework for understanding the spatial relationships between images captured by two cameras. A key aspect of this geometry is the concept of epipolar lines, which emerge from the essential matrix E that describes the fundamental properties of the scene and the cameras. Similarly, the fundamental matrix F offers a mathematical representation of the epipolar geometry, highlighting the projective relationships between corresponding points in stereo images.

Indeed, epipolar lines are of critical importance in the process of finding correspondences between points in stereo images, a necessary step for applications such as stereo triangulation and three-dimensional reconstruction.

In the following section, we will closely examine the calculations associated with the essential matrix E and the fundamental matrix F, as well as their impact on determining epipolar lines. We will also explore the practical implications of these concepts for stereo vision, highlighting the methods and algorithms used to effectively exploit epipolar lines in practical applications.

2.3.3 Epipolar Geometry

To express the essential matrix E in terms of the rotation matrix related to A_s, we need to decompose A_s into its rotation and translation components.

The matrix A_s is defined as follows:

$$A_s = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix}$$

where R is the rotation matrix and t is the translation vector.

We can then refer back to the epipolar line equation (18). We can express the elements a, b, and c as:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = E \cdot \begin{pmatrix} x_L \\ y_L \\ 1 \end{pmatrix} \tag{19}$$

We can write E from the rotation matrix R and the translation vector t:

$$E = [t]_\times R$$

The antisymmetric matrix $[t]_\times$ is then defined as:

$$[t]_\times = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix}$$

By combining these elements, we obtain:

$$E = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix} R \tag{20}$$

We can rewrite the epipolar line as the equation:

$$P_{CR}^t E P_{CL} = 0 \tag{21}$$

Next, we want to calculate the fundamental matrix F. For this, we need to revisit the projection from the camera frame to the image frame.

$$\mathbf{p}_R = K_R \mathbf{P}_{CR}$$
$$\mathbf{p}_L = K_L \mathbf{P}_{CL}$$

By substituting the homogeneous coordinates, we get:

$$(K_R^{-1}\mathbf{p}_R)^T E (K_L^{-1}\mathbf{p}_L) = 0$$

Which expands to:

$$\mathbf{p}_R^T K_R^{-T} E K_L^{-1} \mathbf{p}_L = 0$$

We define the fundamental matrix F in terms of the essential matrix E and the intrinsic matrices:

$$F = K_R^{-T} E K_L^{-1} \tag{22}$$

The epipolar equation then becomes:

$$\mathbf{p}_R^T F \mathbf{p}_L = 0 \tag{23}$$

Thus, we have a relationship between the image coordinates of the right camera and the left camera.

2.4 Depth Calculation

Depth calculation is a crucial step in stereo vision, allowing the reconstruction of the three-dimensional structure of a scene from two-dimensional images. By exploiting the differences in the positions of corresponding points in the two images (disparity), it is possible to determine the distance of these points from the cameras.

2.4.1 Principle of Disparity

Disparity refers to the difference in position of the same point in the two images captured by stereo cameras. Specifically, if a point in the scene is projected onto the left and right images at coordinates (x_L, y_L) and (x_R, y_R) respectively, the disparity d is defined as:

$$d = x_L - x_R \tag{24}$$

Disparity is directly related to the depth of the point in the scene. Points closer to the cameras will have a larger disparity, while points further away will have a smaller disparity.

2.4.2 Relationship between Disparity and Depth

The depth Z of a point in the scene can be calculated using the disparity and the camera parameters. For horizontally aligned stereo cameras, the relationship between depth and disparity is given by:

$$Z = \frac{f \cdot B}{d} \tag{25}$$

where:

- Z is the depth of the point.
- f is the focal length of the cameras.
- B is the stereo baseline, i.e., the distance between the two cameras.
- d is the disparity of the point.

The depth calculation described above is valid in the case where the optical axes of the cameras are parallel and one of the image axes is aligned. This configuration is common in standard stereo vision systems, where the cameras are horizontally aligned and their image planes are parallel to the stereo baseline.

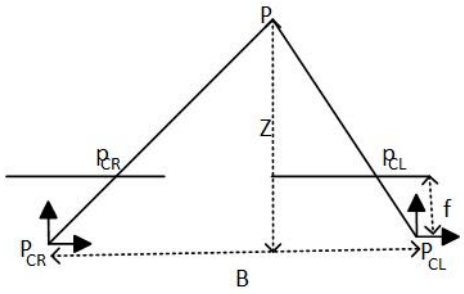

Figure 3: Schematic of stereo vision

It is important to note that deviations from these ideal conditions can introduce errors in depth calculation. For instance, optical distortions, image deformations, and camera misalignments can affect the accuracy of disparity measurements and introduce errors in depth estimates.

3 Time-of-Flight (ToF) Sensors

3.1 Principle of ToF Sensors

Time-of-Flight (ToF) sensors operate on the principle of measuring the time it takes for a light signal to travel a given distance and return to the sensor. To do this, the sensor

emits a light signal, usually a laser beam, towards the object to be measured. This signal is then reflected by the object and returned to the sensor. By measuring the time elapsed between the emission and reception of the signal, the sensor can determine the distance between itself and the object.

3.1.1 ToF Related Calculations

When the optical axes of the cameras are parallel and one of the image axes is aligned, it is possible to calculate the distance between the sensor and the object using the following equation:

$$\text{Distance} = \frac{c \times \Delta t}{2} \tag{26}$$

In this equation:

- Distance represents the distance between the sensor and the target object.
- c is the speed of light in the propagation medium (typically air).
- Δt is the time difference between the emission and reception of the light signal.

This equation allows for the calculation of the distance between the sensor and the object using the measured time of flight and the speed of light.

4 Conclusion

Consequently, the different equations necessary for a depth camera with a binocular system and a ToF sensor, are :

Projection matrix M :

$$M = K \cdot T = \begin{pmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R_{11} & R_{12} & R_{13} & T_x \\ R_{21} & R_{22} & R_{23} & T_y \\ R_{31} & R_{32} & R_{33} & T_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Least Square result :

$$Q = (K_{2N}^t . K_{2N})^{-1} K_{2N}^t . U$$

The relationships to find all the elements

$$\begin{aligned} r_3 &= m_3 \\ c_x &= m_1 \cdot m_3 \\ c_y &= m_2 \cdot m_3 \\ f_x &= -\|m_1 \wedge m_3\| \\ f_y &= \|m_2 \wedge m_3\| \\ r_1 &= \frac{1}{f_x}(m_1 - c_x m_3) \\ r_2 &= \frac{1}{f_y}\left(m_2 - c_y m_3\right) \\ t_x &= \frac{1}{f_x}(m_{14} - c_x m_{34}) \\ t_y &= \frac{1}{f_y}\left(m_{24} - c_y m_{34}\right) \\ t_z &= m_{34} \end{aligned}$$

Essential matrix

$$E = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix} R$$

Fondamentale F

$$F = K_R^{-T} E K_L^{-1}$$

Epipolar line equation

$$P_{CR}^{t} E P_{CL} = 0$$

$$\mathbf{p}_R^T F \mathbf{p}_L = 0$$

Depth equation

$$Z = \frac{f \cdot B}{d}$$

ToF sensor equation

$$\text{Distance} = \frac{c \times \Delta t}{2}$$