Slimane LARABI

Computer Vision

From Bidimensional Images to Three Dimensional Scene

January 02, 2025

Springer Nature

Chapter 1

Uncalibrated stereo cameras

1.1 Introduction

Uncalibrated stereo vision is valuable in computer vision because it enables 3D scene reconstruction without requiring prior knowledge of the camera's intrinsic or extrinsic parameters. This flexibility makes it practical for real-world applications where pre-calibrating cameras is difficult or infeasible. Uncalibrated stereo vision offers several advantages, including:

- Saving time and effort and money in case where pre-calibration is impractical or unnecessary, therefore there is no need to use equipment designed calibration patterns for calibration.
- Suitable for dynamic environments where cameras are moved, and calibration cannot be performed beforehand.
- Insuring scalability for systems with multiple cameras, uncalibrated setups avoid the logistical challenges of calibrating every camera.
- Estimating projective or affine transformations without knowing the exact camera parameters for Structure-from-Motion (SfM).
- Getting $3D$ reconstruction using uncalibrated cameras is valuable especially for monitoring large areas with pan-tilt-zoom cameras by mobile robot or drones that change position constantly.

In this chapter we study how can we estimate 3D structure of a static scene from two different views without camera calibration. We study how to use the epipolar geometry in order to perform the stereo matching of key points. This is done by estimating the fundamental matrix and finding the correspondences and computing the depth map.

1.2 Epipolar Geometry of an uncalibrated stereo

We studied in the previous chapter how to calibrate a camera and therefore we can estimate the intrinsic parameters. We suppose than the position and orientation (extrinsic parameters) of the camera with respect to an external frame coordinates are unknown. Our aim is to reconstruct the 3D structure using two or more images taken by a camera at different view points. To do this, we need to extract keypoints from these images and to compute the stereo correspondences.

Figure 1.1 illustrates such case where two images taken by a camera at different view points in indoor scene. In order to match key points located into the two images, we need to constraint the search space of the matches from $2D$ to $1D$ (image plane to line). To do this, we will use the epipolar constraint.

1.2.1 Epipolar Geometry: Definitions

The **epipoles** are defined as the image point of pinhole of one camera as viewed by the other camera. In the figure 1.2, e_l, image of O_r and e_r, image of O_l, are the two epipoles, they are unique for a given stereo pair.

The **Epipolar plane** is the plane associated to a scene point P: is formed by projection centers O_l and O_r, epipoles e_l and e_r and scene point P. Every scene point P lies on a unique epipolar plane.

Fig. 1.1 Two images acquired of indoor scene.

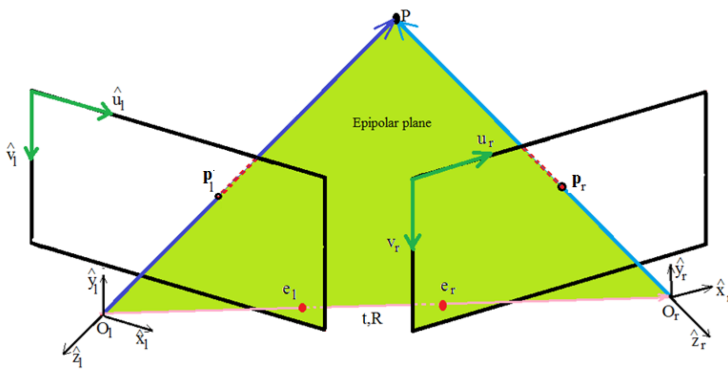

Fig. 1.2 The epipolar plane in green color. O_l, O_r are the projection centers, e_l, e_r are the epipoles and P_l, P_r are images of the point P.

1.2.2 Epipolar constraint

We note $\overrightarrow{O_l p_l}$ the vector with coordinates (x_l, y_l, z_l) with respect to the left camera coordinate frame. Let $\overrightarrow{n}$ be the vector normal to the epipolar plane (see figure 1.3). We can write using the vector product :

$$\overrightarrow{n} = \overrightarrow{t} \times \overrightarrow{O_l p_l} \tag{1.1}$$

where $\overrightarrow{t}$ is the translation vector from O_l to O_r.
Applying the epipolar constraint we obtain:

$$\overrightarrow{O_l p_l}.\overrightarrow{n} = \overrightarrow{O_l p_l}.(\overrightarrow{t} \times \overrightarrow{O_l p_l}) \tag{1.2}$$

As the angle between $\overrightarrow{O_l p_l}$ and $(\overrightarrow{n})$ is equal to $\Pi/2$, then the dot product $\overrightarrow{O_l p_l}.\overrightarrow{n}$ is equal to zero.

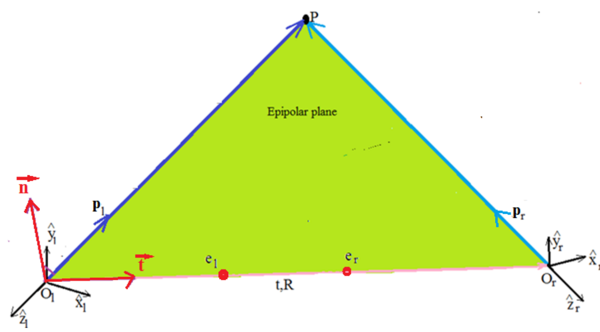

Fig. 1.3 The epipolar constraint.

We note (t_x, t_y, t_z) the coordinates of the vector of $\overrightarrow{t}$, the position of the right camera in the left camera frame, Writing the expression $(\overrightarrow{t} \times \overrightarrow{O_l p_l})$ in matrix form, we obtain:

$$\begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \times \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} = \begin{bmatrix} t_y z_l - t_z y_l \\ t_z x_l - t_x z_l \\ t_x y_l - t_y x_l \end{bmatrix} \tag{1.3}$$

we obtain then for $\overrightarrow{O_l p_l}.(\overrightarrow{t} \times \overrightarrow{O_l p_l})$:

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} \begin{bmatrix} t_y z_l - t_z y_l \\ t_z x_l - t_x z_l \\ t_x y_l - t_y x_l \end{bmatrix} = 0 \tag{1.4}$$

We can write this equation as follow:

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 - t_x & \\ -t_y & t_x & 0 \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} = 0 \tag{1.5}$$

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} T \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} = 0 \tag{1.6}$$

We note $R_{3\times 3}$: the orientation of the left camera in the right camera's frame.

$$\begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{1.7}$$

Combining these two last equations, we obtain:

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} T (\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}) = 0 \tag{1.8}$$

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} T \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} + \begin{bmatrix} x_l & y_l & z_l \end{bmatrix} T \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} = 0 \tag{1.9}$$

As:

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} T \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} = 0 \tag{1.10}$$

then:

$$\begin{bmatrix} x_l & y_l & z_l \end{bmatrix} TR \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = 0 \tag{1.11}$$

We define the essential matrix $E = TR$ [9], which relates the 3D coordinates of the point P with respect to the right and left coordinate frames. We can then express the epipolar constraint, which describes the relationship between the left and right image points using the essential matrix. This matrix encodes the relative rotation and translation between the two cameras, up to a scale factor.

$$p_l E p_r \tag{1.12}$$

where

$$E = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \tag{1.13}$$

1.3 The Fundamental Matrix *F*

We study in this section how to compute the fundamental matrix F which encapsulates both intrinsic and extrinsic parameters. The fundamental matrix is then used in uncalibrated stereo systems where intrinsic parameters are unknown, .

1.3.1 Decomposition of the Essential matrix *E*

It is possible to decompose R and T from E using SVD Decomposition.

To find E, we write:

$$p_l^T E p_r = 0 \tag{1.14}$$

where p_l, p_r are assumed to correspond to the same 3D point.

Using the equations of perspective projection of each camera, we can write:

$$z_l \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = \begin{bmatrix} z_l u_l \\ z_l v_l \\ z_l \end{bmatrix} \begin{bmatrix} f_x^l x_l + z_l O_x^l \\ f_y^l y_l + z_l O_y^l \\ z_l \end{bmatrix} = \begin{bmatrix} f_x^l & 0 & O_x^l \\ 0 & f_y^l & O_y^l \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} \tag{1.15}$$

The matrix K_l of calibration of the left camera is assumed known:

$$K_l = \begin{bmatrix} f_x^l & 0 & O_x^l \\ 0 & f_y^l & O_y^l \\ 0 & 0 & 1 \end{bmatrix} \tag{1.16}$$

We can then write:

$$z_l \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = K_l p_l \tag{1.17}$$

If we multiply the two members of the equation (1.17) by the term K_l^{-1}, we obtain:

$$K_l^{-1} z_l \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = K_l^{-1} K_l p_l \tag{1.18}$$

We obtain then:

$$p_l = K_l^{-1} z_l \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} \tag{1.19}$$

If we take the transpose of the two members we obtain:

$$p_l^T = \begin{bmatrix} u_l & v_l & 1 \end{bmatrix} z_l (K_l^{-1})^T \tag{1.20}$$

In the other side, we have for the right camera:

$$z_r \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = \begin{bmatrix} f_x^r & 0 & O_x^r \\ 0 & f_y^r & O_y^r \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} \tag{1.21}$$

If we multiply the two members of the previous equations by K_r^{-1}, where K_r is the matrix calibration of the right camera, we obtain:

$$p_r = K_r^{-1} z_r \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} \tag{1.22}$$

We replace p_r and p_l^T in the equation$p_l^T E p_r = 0$ by their expressions we obtain:

$$\begin{bmatrix} u_l & v_l & 1 \end{bmatrix} z_l (K_l^{-1})^T \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} K_r^{-1} z_r \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = 0 \tag{1.23}$$

As z_l, z_r are non zero, we obtain:

$$p_l^T F p_r = 0 \tag{1.24}$$

where:

$$F = (K_l^{-1})^T \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} K_r^{-1} \tag{1.25}$$

The F is called Fundamental matrix [10]. Once F is computed, we can retrieve the matrix E as follow:

$$F = (K_l^{-1})^T E K_r^{-1} \tag{1.26}$$

$$(K_l)^T F K_r = E \tag{1.27}$$

1.3.2 Estimation of the Fundamental Matrix

The first step is to find a set of corresponding features (at least 8) in left and right images (using SIFT for example). Figure 1.4 shows some matched keypoints.

For each stereo correspondence (i), write the epipolar constraint:

$$\begin{bmatrix} u_l^i & v_l^i & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u_r^i \\ v_r^i \\ 1 \end{bmatrix} = 0 \tag{1.28}$$

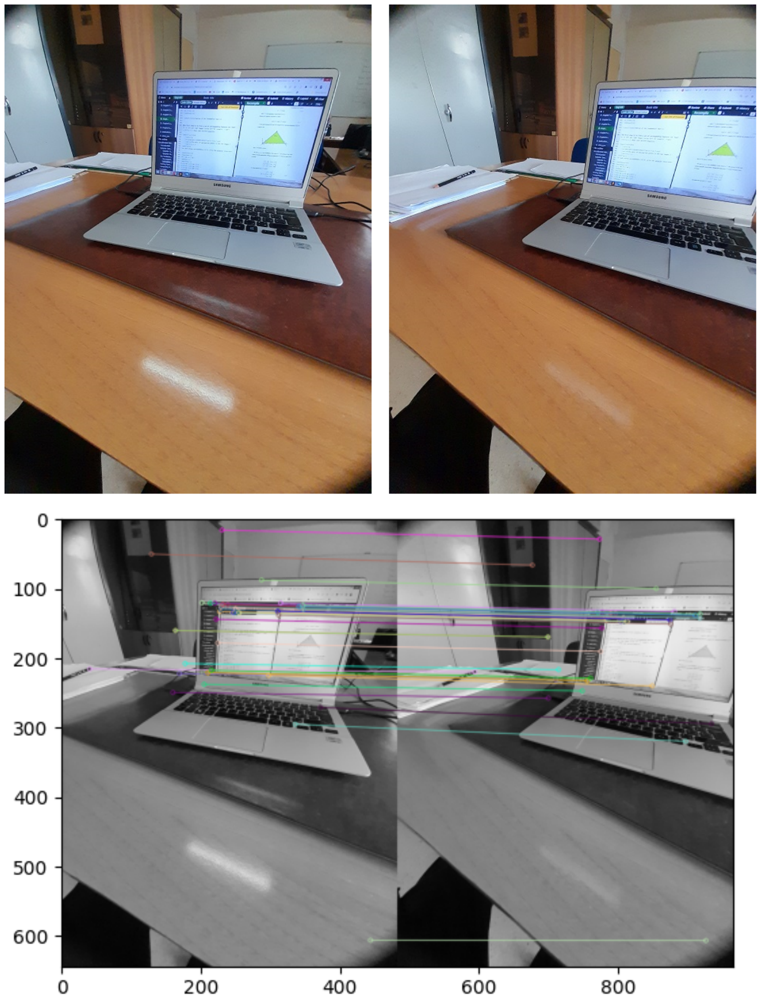

Fig. 1.4 The pairs of images and the matched key points in the two images.

We obtain for all stereo correspondences the linear system:

$$\begin{pmatrix} u_l^1u_r^1 & u_l^1v_r^1 & u_l^1 & u_r^1v_l^1 & v_r^1v_l^1 & v_l^1 & u_r^1 & v_r^1 & 1 \\ u_l^2u_r^2 & u_l^2v_r^2 & u_l^2 & u_r^2v_l^2 & v_r^2v_l^2 & v_l^2 & u_r^2 & v_r^2 & 1 \\ -- & -- & -- & -- & -- & -- & -- & -- & -- \\ u_l^iu_r^i & u_l^iv_r^i & u_l^i & u_r^iv_l^i & v_r^iv_l^i & v_l^i & u_r^i & v_r^i & 1 \\ -- & -- & -- & -- & -- & -- & -- & -- & -- \\ u_l^nu_r^n & u_l^nv_r^n & u_l^n & u_r^nv_l^n & v_r^nv_l^n & v_l^n & u_r^n & v_r^n & 1 \end{pmatrix} \begin{pmatrix} f_{11} \\ f_{12} \\ f_{13} \\ --- \\ f_{31} \\ f_{32} \\ f_{33} \end{pmatrix} = 0 \qquad (1.29)$$

Then we have the expression $AF = 0$, where F is the fundamental matrix, A is the matrix composed by elements related to the coordinates of matched image points.

Note that the fundamental matrix F and kF describe the same epipolar geometry, then F is defined only up to scale. In order to find the eight parameters, we can set $||F||^2 = 1$.

This is the same problem like solving Projection matrix during camera calibration, or Homography matrix for image stitching. Then, we compute the matrix F which verify:

$$\min_F ||AF||^2, with||F||^2 = 1 \qquad (1.30)$$

The next step is the computation of the Essential matrix E, where: $E = K_l^T F K_r$, then we extract the rotation matrix R and translation vector t_x from E using SVD Decomposition $(E = RT)$.

1.4 Finding Stereo Correspondences Using Epipolar Constraint

Epipolar line is the intersection of image plane and epipolar plane. At each scene point, there are two corresponding epipolar line, one each on the two image planes (see figure 1.5). Given one point on the left image, the corresponding point on the right image must lie on the epipolar line. The search space of correspondences is then reduced to one dimensional space.

How to compute the epipolar line?

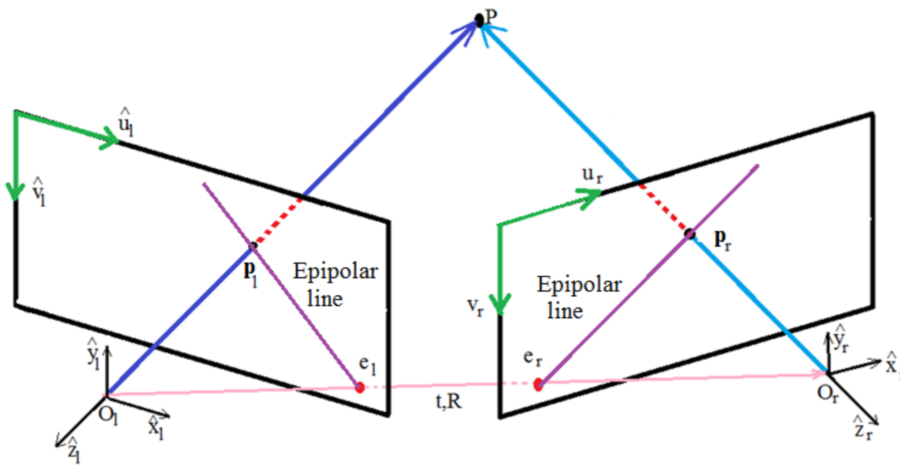

Fig. 1.5 The epipolar line defined as the intersection of the plane $(O_l p_l O_r)$ and the right image plane.

Given the Fundamental matrix F and point $p_l(u_l, v_l)$ on left image. The matched point $p_r(u_r, v_r)$ belong to the epipolar line located in the right image.

We give in follow, how to express this line using the fundamental matrix and the 2D coordinates of the left point.

If we expand the equation:

$$\begin{bmatrix} u_l & v_l & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = 0 \tag{1.31}$$

we obtain then:

$$\begin{bmatrix} u_l f_{11} + v_l f_{21} + f_{31} & u_l f_{12} + v_l f_{22} + f_{32} & u_l f_{13} + v_l f_{23} + f_{33} \end{bmatrix} \begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = 0 \tag{1.32}$$

$$u_r(u_l f_{11} + v_l f_{21} + f_{31}) + v_r(u_l f_{12} + v_l f_{22} + f_{32}) + (u_l f_{13} + v_l f_{23} + f_{33}) = 0 \quad (1.33)$$

This equation may be written as follow:

$v_r = au_r + b$ where the parameters a, b are function of the matrix F and u_l, vl.

1.5 Computing Depth with Unknown External Parameters

We assume that the two cameras are calibrated and then the two internal matrices are available. We can write the expressions of the two dimensional coordinates of image points on the two image planes as follow:

$$\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = \begin{bmatrix} f_x^l & 0 & O_x^l & 0 \\ 0 & f_y^l & O_y^l & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ z_l \\ 1 \end{bmatrix} \quad (1.34)$$

$$\begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = \begin{bmatrix} f_x^r & 0 & O_x^r & 0 \\ 0 & f_y^r & O_y^r & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix} \quad (1.35)$$

where $(x_l, y_l, z_l), (x_r, y_r, z_r)$ are the coordinates of the point P with respect to the left and right 3D camera coordinates.

$$\begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix} \quad (1.36)$$

If we replace $(x_r, y_l, z_l, 1)$ in the equation giving the (u_l, v_l), we obtain the:

$$\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = \begin{bmatrix} f_x^l & 0 & O_x^l & 0 \\ 0 & f_y^l & O_y^l & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix} \tag{1.37}$$

For the right camera, we have:

$$\begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = \begin{bmatrix} f_x^r & 0 & O_x^r & 0 \\ 0 & f_y^r & O_y^r & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix} \tag{1.38}$$

We will write these two last equation as follow:

$$\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ z_l \\ 1 \end{bmatrix} \tag{1.39}$$

$$\begin{bmatrix} u_r \\ v_r \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \\ 1 \end{bmatrix} \tag{1.40}$$

Applying the cross product for the left camera we obtain:

$$\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} \times \begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} = 0 \tag{1.41}$$

$$\begin{bmatrix} u_l \\ v_l \\ 1 \end{bmatrix} \times \begin{bmatrix} P_{l1}\tilde{p_r} \\ p_{l2}\tilde{p_r} \\ p_{l3}\tilde{p_r} \\ p_{l4}\tilde{p_r} \end{bmatrix} = 0 \tag{1.42}$$

where P_{li} is the i^{th} line of the matrix P and $\tilde{p_r} = (x_r, y_r, z_r, 1)^T$. This is equivalent to the following expression:

$$\begin{bmatrix} v_l p_{l3}\tilde{p_r} - p_{l2}\tilde{p_r} \\ u_l p_{l3}\tilde{p_r} - p_{l1}\tilde{p_r} \\ u_l p_{l2}\tilde{p_r} - p_{l2}\tilde{p_r} \end{bmatrix} = 0 \tag{1.43}$$

$$\begin{bmatrix} v_l p_{l3} - p_{l2} \\ u_l p_{l3} - p_{l1} \\ u_l p_{l2} - p_{l2} \end{bmatrix} \tilde{p_r} = 0 \tag{1.44}$$

Applying the cross product for the right camera we obtain the similar results:

$$\begin{bmatrix} v_l m_{l3} - m_{l2} \\ u_l m_{l3} - m_{l1} \\ u_l m_{l2} - m_{l2} \end{bmatrix} \tilde{p_r} = 0 \tag{1.45}$$

If we rearrange the terms of the two last equations, we obtain:

$$\begin{bmatrix} u_r m_{31} - m_{11} & u_r m_{32} - m_{12} & u_r m_{33} - m_{13} \\ v_r m_{31} - m_{21} & v_r m_{32} - m_{22} & v_r m_{33} - m_{23} \\ u_l p_{31} - p_{11} & u_l p_{32} - p_{12} & u_l p_{33} - p_{13} \\ v_r p_{31} - p_{21} & v_r p_{32} - p_{22} & v_r p_{33} - p_{23} \end{bmatrix} \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = \begin{bmatrix} m_{34} - m_{14} \\ m_{34} - m_{24} \\ p_{34} - p_{14} \\ p_{34} - p_{24} \end{bmatrix} \tag{1.46}$$

We have: $Ap = b$ where A, b are known. The solution to this problem give the 3D unknown coordinates (x_r, y_r, z_r).

1.6 Examples

1.6.1 Example 1

Applying the following pseudo code onto the images of figure 1.6, we obtain the results shown in figure 1.7.

Fig. 1.6 The stereo pair of images.

- 1. Read left and right images,
- 2. Find the keypoints and descriptors with SIFT,
- 3. Match the SIFT descriptors using the KNN technique,
- 4. Select only inlier points,
- 5. Find epipolar lines corresponding to points in right image (second image) and drawing its lines on left image,
- 6. Find epipolar lines corresponding to points in left image (first image) and drawing its lines on right image.

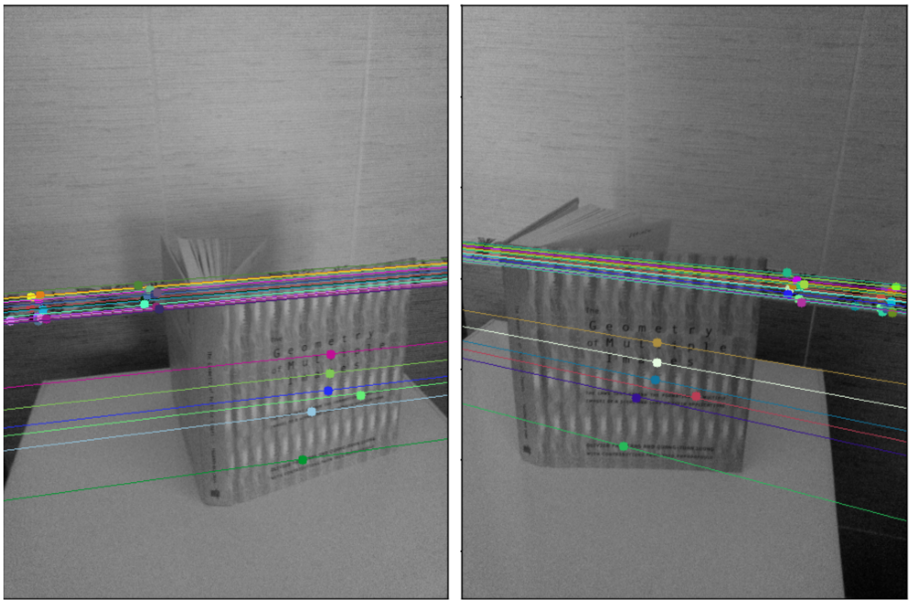

Fig. 1.7 The located epipolar lines in both images.

1.6.2 Example 2: Computation of the Fundamental Matrix

We apply for a second stereo images pair (see figure 1.8) the same steps described in the pseudo algorithm of the example 1. In addition, the good matched serve to compute the fundamental matrix. Once this matrix is computed, we select from the pair of matched points those giving the best matrix. The epipolar lines are draw and illustrated by figure 1.9.

1.6.3 Example 3: Depth Map Computation after Images Rectification using the Fundamental Matrix

In order to compute a depth map of the scene corresponding to the images stereo of Figure 1.8, we extract first SIFT features as illustrated by figure 1.10 (for one image). The features are used to compute the fundamental matrix, then the two images are

Fig. 1.8 The second stereo images.

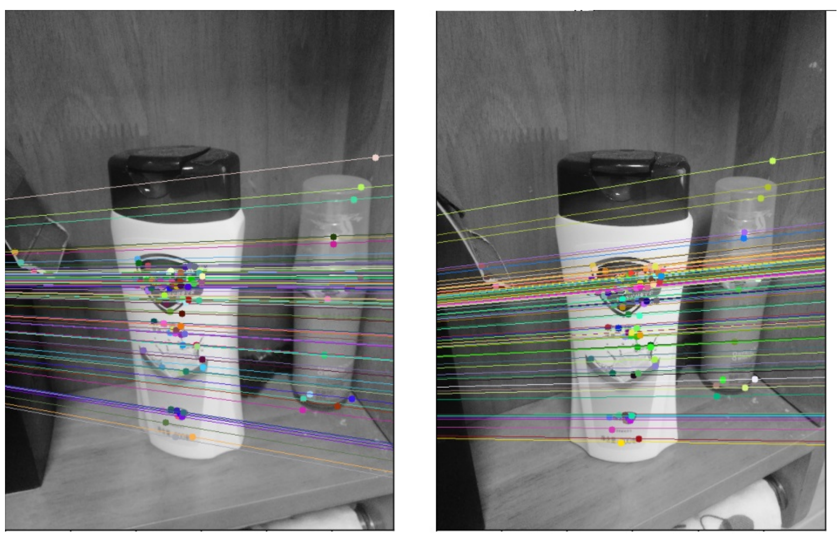

Fig. 1.9 The located epipolar lines in both images.

rectified such that their epipolar lines became horizontal as shown by figure 1.11. At the end, we get a simple stereo system and the depth map is computed as explained in the chapter 3 and shown by figure 1.12.

Fig. 1.10 The set of SIFT features located on one image.

1.7 Conclusion

In this chapter we provided a comprehensive framework for understanding how uncalibrated stereo systems can achieve effective 3D scene reconstruction through epipolar geometry.

We explored the fundamental concepts that enable 3D reconstruction from an uncalibrated stereo system, focusing on the role of epipolar geometry.

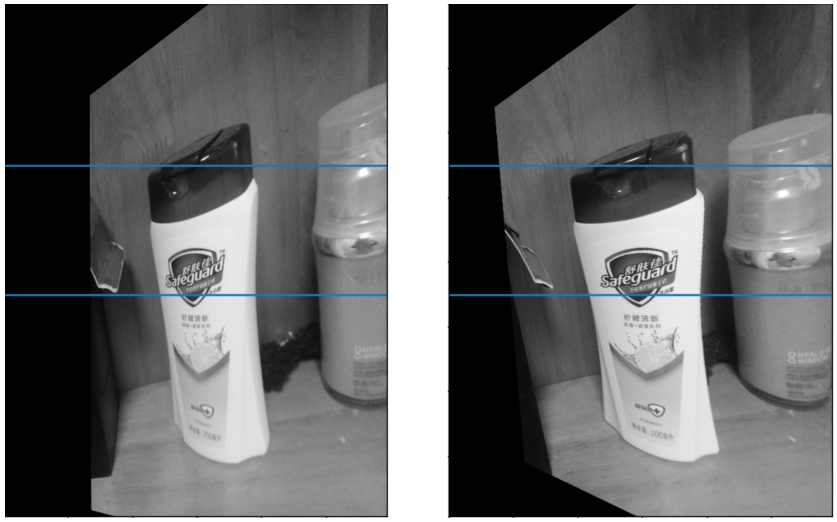

Fig. 1.11 The located epipolar lines in both images after images rectification.

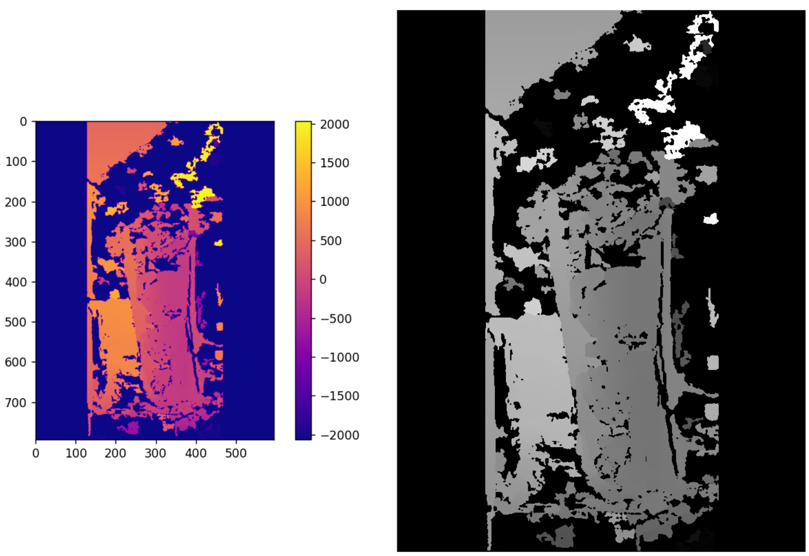

Fig. 1.12 The computed depth map.

The fundamental matrix was introduced as a key tool in encoding this geometric relationship, allowing us to map points in one image to corresponding epipolar lines in the other, regardless of the cameras' intrinsic parameters.

Using the epipolar constraint, we then addressed the challenge of finding stereo correspondences, highlighting how the constraint simplifies the search for matching points by reducing it from a 2D to a 1D problem.

We discussed also methods for computing depth information in scenarios where external parameters are unknown.

References

1. Seymourt Papert. The Summer Vision Project. Artificial Intelligence Group, MIT, 1966. https://dspace.mit.edu/handle/1721.1/11589
2. Bob Sumner. Augmented Creativity, TEDxZurich, 19th of January, 2015. https://www.youtube.com/watch?v=AJJOWemfOYI
3. Sachiko Iwase and Hideo Saito. Tracking soccer players based on homography among multiple views", Proc. SPIE 5150, Visual Communications and Image Processing 2003, (23 June 2003); https://doi.org/10.1117/12.502967
4. Martin A. Fischler, Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, Volume 24, Issue 6, pp 381–395, https://doi.org/10.1145/358669.358692
5. "The Eye of a Robot: Studies in Machine Vision at MIT" and "TX-O Computer". Courtesy of MIT Museum.
6. Richard Hartley and Andrew Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, 2000.
7. D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003), volume 1, pages 195-202, Madison, WI, June 2003.
8. Zhengyou Zhang. A Flexible New Technique for Camera Calibration. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 22, NO. 11, NOVEMBER 2000
9. A computer algorithm for reconstructing a scene from two projections. Nature volume 293, pages 133–135 (1981)
10. OD Faugeras, QT Luong, SJ Maybank. Camera self-calibration: Theory and experiments. Second European Conference on Computer Vision ECCV'92, 1992
11. Olsson, Carl; Enqvist, Olof, Stable Structure from Motion for Unordered Image Collections Scandinavian Conference on Image Analysis, 2011.
12. Marc Pollefeys and Luc Van Gool. 3-D Modeling from Images. COMMUNICATIONS OF THE ACM July 2002/Vol. 45, No. 7.
13. Peng-Shuai Wang, Yang Liu, Yu-Xiao Guo, Xin ; O-CNN: Octree-based Convolutional Neural Networks for 3D Shape Analysis , July 2017. ACM Transactions on Graphics, 36(4):1-11
14. M. Michalkiewicz, J. K. Pontes, D. Jack, M. Baktashmotlagh, A. Eriksson. Deep Level Sets: Implicit Surface Representations for 3D Shape Inference. 2019. arXiv:1901.06802 [cs.CV]. https://arxiv.org/pdf/1901.06802.pdf

15. A. Podlozhnyuk, S. Pirker, C. Kloss. Efficient implementation of superquadric particles in Discrete Element Method within an open-source framework. Computational Particle Mechanics 4(1), 2016. DOI: 10.1007/s40571-016-0131-6
16. M.A. Turk et al. Face recognition using eigenfaces,CVPR 1991.
17. S.K.Nayar et al., Real-Time 100 Object Recognition System, ICRA, 1996
18. C. B. Choy, D. Xu, J. Gwak, K. Chen, and S. Savarese. 3dr2n2: A unified approach for single and multi-view 3d object reconstruction. In European conference on computer vision, pages 628–644. Springer, 2016
19. R. Girdhar, D. F. Fouhey, M. Rodriguez, and A. Gupta. Learning a predictable and generative vector representation for objects. In European Conference on Computer Vision, pages 484–499. Springer, 2016
20. D. J. Rezende, S. A. Eslami, S. Mohamed, P. Battaglia, M. Jaderberg, and N. Heess. Unsupervised learning of 3d structure from images. In Advances in Neural Information Processing Systems, pages 4996–5004, 2016
21. S. R. Richter and S. Roth. Matryoshka networks: Predicting 3d geometry via nested shape layers. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1936–1944, 2018
22. C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition., 1(2):4, 2017.
23. Y. Liao, S. Donne, and A. Geiger. Deep marching cubes: ´ Learning explicit surface representations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2916–2925, 2018
24. H. Fan, H. Su, and L. J. Guibas. A point set generation network for 3d object reconstruction from a single image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition., volume 2, page 6, 2017
25. Mateusz Michalkiewicz et al. Deep Level Sets: Implicit Surface Representations for 3D Shape Inference. arXiv:1901.06802v1 [cs.CV], 21 Jan 2019.
26. Roberts, Lawrence G. 1963. Machine perception of three-dimensional solids. Outstanding PhD dissertations in the computer sciences. Garland Publishing, New York. 1963.
27. Adolfo Guzman-Arénas, Computer Recognition of Three-Dimensional Objects In a Visual Scene. PhD Dissertations in the computer sciences, MIT 1968.
28. M.B. Clowes. On seeing things. Artificial Intelligence, Volume 2, Issue 1, Spring 1971, Pages 79-116.
29. David L. Waltz. GENERATING SEMANTIC DESCRIPTIONS FROM DRAWINGS OF SCENES WITH SHADOWS. MIT Artificial Intelligence Laboratory. Technical Report 271, November 1972.

30. Kemp, M. Julesz's joyfulness. Nature 396, 419 (1998). https://doi.org/10.1038/24753
31. Julesz B. Foundations of Cyclopean Perception. Chicago University Press; Chicago, IL, USA: 1971.
32. David Marr. Vision, A Computational Investigation into the Human Representation and Processing of Visual Information. Originally published: San Francisco : W. H. Freeman, c1982.
33. Arthur Coste. Affine Transformation, Landmarks registration, Non linear Warping. https://www.sci.utah.edu/ãcoste/uou/Image/project3/ArthurCOSTE_Project3.pdf October 2012.
34. Sid Yingze Bao, Mohit Bagra, Yu-Wei Chao, Silvio Savarese. Semantic Structure From Motion with Points, Regions, and Objects. CVPR 2012.
35. Marco Crocco, Cosimo Rubino, Alessio Del Bue. Structure from Motion with Objects, CVPR 2016.
36. D.P. Robertson and R. Cipolla. Structure from Motion. In Varga, M., editors, Practical Image Processing and Computer Vision, John Wiley, 2009.
37. C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. International Journal of Computer Vision, 9(2):137–154, 1992.
38. P. F. Sturm andW. Triggs. A factorization based algorithm for multi-image projective structure and motion. In European Conference on Computer Vision (ECCV'96), pages 709–720, 1996.
39. F. Schaffalitzky, A. Zisserman, R. I. Hartley, and P. H. S. Torr. A six point solution for structure and motion. In European Conference on Computer Vision (ECCV'00), pages 632–648, 2000.
40. W. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle adjustment: A modern synthesis. In W. Triggs, A. Zisserman, and R Szeliski, editors, Vision Algorithms: Theory and Practice, LNCS, pages 298–375. Springer Verlag, 2000.
41. D. C. Brown. The bundle adjustment - progress and prospects. International Archives of Photogrammetry, 21(3), 1976.