Image Processing and Computer Vision (Module 1)

Last update: 03 March 2024

Academic Year 2023 – 2024
Alma Mater Studiorum · University of Bologna

Contents

1 Image acquisition and formation

1.1 Pinhole camera

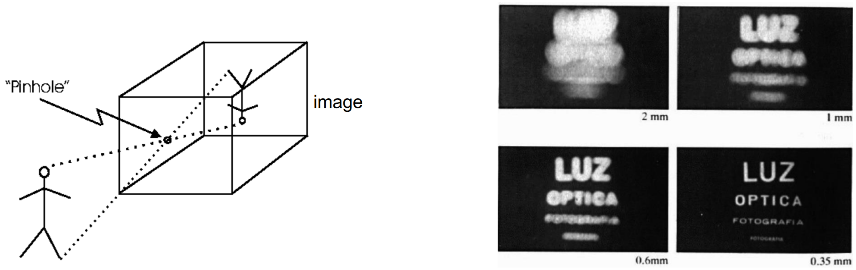
Imaging device Gathers the light reflected by 3D objects in a scene and creates a 2D representation of them. Imaging device

Computer vision Infer knowledge of the 3D scene from 2D digital images. Computer vision

Pinhole camera Imaging device where the light passes through a small pinhole and hits the image plane. Geometrically, the image is obtained by drawing straight rays from the scene to the image plane passing through the pinhole. Pinhole camera

Remark. Larger aperture size of the pinhole results in blurry images (circle of confusion), while smaller aperture results in sharper images but requires longer exposure time (as less light passes through).

Remark. The pinhole camera is a good approximation of the geometry of the image formation mechanism of modern imaging devices.



(a) Pinhole camera model

(b) Images with varying pinhole aperture size

1.2 Perspective projection

Geometric model of a pinhole camera. Perspective projection

Scene point M (the object in the real world).

Image point m (the object in the image).

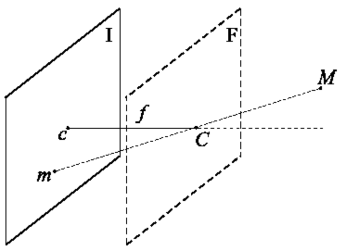
Image plane I.

Optical center C (the pinhole).

Image center/piercing point c (intersection between the optical axis – the line orthogonal to I passing through C – and I).

Focal length f.

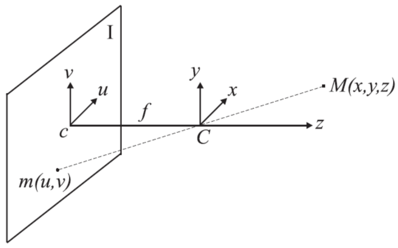
Focal plane F.



- u and v are the horizontal and vertical axis of the image plane, respectively.
- x and y are the horizontal and vertical axis of the 3D reference system, respectively, and form the **camera reference system**.

Camera reference system

Remark. For the perspective model, the coordinate systems (U, V) and (X, Y) must be parallel.



Scene–image mapping The equations to map scene points into image points are the following:

Scene–image mapping

$$u = x\frac{f}{z} \qquad v = y\frac{f}{z}$$

Proof. This is the consequence of the triangle similarity theorems.

$$\frac{u}{x} = -\frac{f}{z} \iff u = -x\frac{f}{z}$$
$$\frac{v}{y} = -\frac{f}{z} \iff v = -y\frac{f}{z}$$

The minus is needed as the axes are inverted

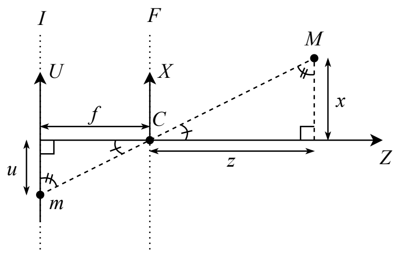


Figure 1.2: Visualization of the horizontal axis. The same holds on the vertical axis.

By inverting the axis horizontally and vertically (i.e. inverting the sign), the image plane can be adjusted to have the same orientation of the scene:

$$u = x\frac{f}{z} \qquad v = y\frac{f}{z}$$

□

Remark. The image coordinates are a scaled version of the scene coordinates. The scaling is inversely proportioned with respect to the depth.

- The farther the point, the smaller the coordinates.
- The larger the focal length, the bigger the object is in the image.

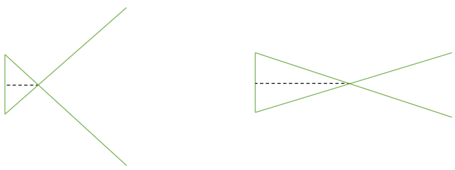


Figure 1.3: Coordinate space by varying focal length

Remark. The perspective projection mapping is not a bijection:

- A scene point is mapped into a unique image point.
- An image point is mapped onto a 3D line.

Therefore, reconstructing the 3D structure of a single image is an ill-posed problem (i.e. it has multiple solutions).

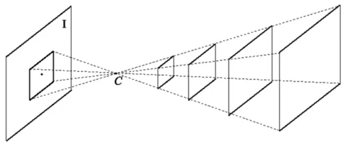


Figure 1.4: Projection from scene and image points

1.2.1 Stereo geometry

Stereo vision

Stereo vision Use multiple images to triangulate the 3D position of an object.

Stereo correspondence

Stereo correspondence Given a point L in an image, find the corresponding point R in another image.

Without any assumptions, an oracle is needed to determine the correspondences.

Standard stereo geometry

Standard stereo geometry Given two reference images, the following assumptions must hold:

- The X, Y, Z axes are parallel.
- The cameras that took the two images have the same focal length f (coplanar image planes) and the images have been taken at the same time.
- There is a horizontal translation b between the two cameras (baseline).
- The disparity d is the difference of the U coordinates of the object in the left and right image.

Fundamental relationship in stereo vision

Theorem 1.2.1 (Fundamental relationship in stereo vision)**.** If the assumptions above hold, the following equation holds:

$$z = b\frac{f}{d}$$

Proof. Let $P_L = \begin{pmatrix} x_L & y & z \end{pmatrix}$ and $P_R = \begin{pmatrix} x_R & y & z \end{pmatrix}$ be the coordinates of the object P with respect to the left and right camera reference system, respectively. Let $p_L = \begin{pmatrix} u_L & v \end{pmatrix}$ and $p_R = \begin{pmatrix} u_R & v \end{pmatrix}$ be the coordinates of the object P in the left and right image plane, respectively.

By assumption, we have that $P_L - P_R = \begin{pmatrix} b & 0 & 0 \end{pmatrix}$, where b is the baseline.

By the perspective projection equation, we have that:

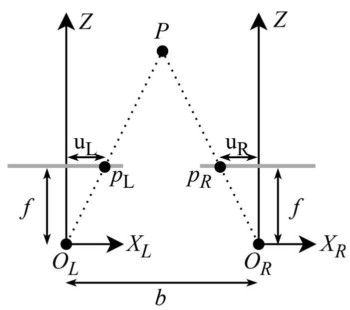
$$u_L = x_L\frac{f}{z} \qquad u_R = x_R\frac{f}{z}$$

Disparity is computed as follows:

$$d = u_L - u_R = x_L\frac{f}{z} - x_R\frac{f}{z} = b\frac{f}{z}$$

We can therefore obtain the Z coordinate of P as:

$$z = b\frac{f}{d}$$



Note: the Y/V axes are not in figure.

$\square$

Remark. Disparity and depth are inversely proportional: the disparity of two points decreases if the points are farther in depth.

Stereo matching

Stereo matching If the assumptions for standard stereo geometry hold, to find the object corresponding to p_L in another image, it is sufficient to search along the horizontal axis of p_L looking for the same colors or patterns.

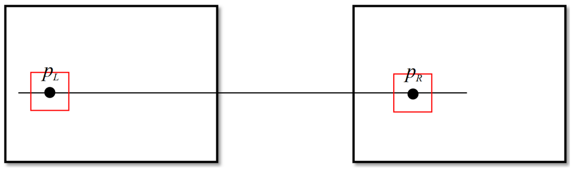


Figure 1.5: Example of stereo matching

Epipolar geometry

Epipolar geometry Approach applied when the two cameras are no longer aligned according to the standard stereo geometry assumption. Still, the focal lengths and the roto-translation between the two cameras must be known.

Given two images, we can project the epipolar line related to the point p_L in the left plane onto the right plane to reduce the problem of correspondence search to a single dimension.

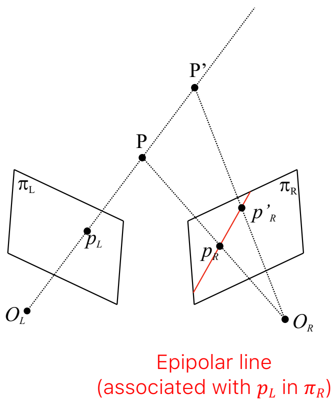
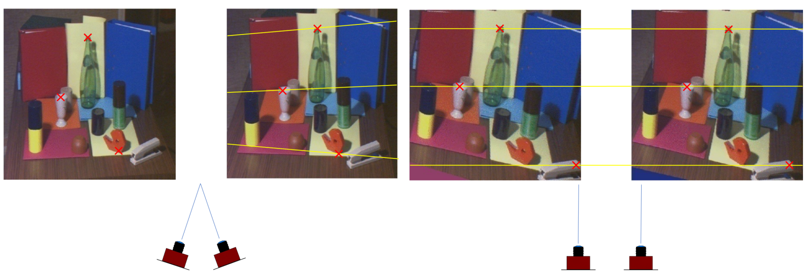


Figure 1.6: Example of epipolar geometry

Remark. It is nearly impossible to project horizontal epipolar lines and searching through oblique lines is awkward and computationally less efficient than straight lines.

Rectification

Rectification Transformation applied to convert epipolar geometry to a standard stereo geometry.



(a) Images before rectification (b) Images after rectification

1.2.2 Ratios and parallelism

Given a 3D line of length L lying in a plane parallel to the image plane at distance z, then its length l in the image plane is:

$$l = L\frac{f}{z}$$

In all the other cases (i.e. when the line is not parallel to the image plane), the ratios of lengths and the parallelism of lines are not preserved.

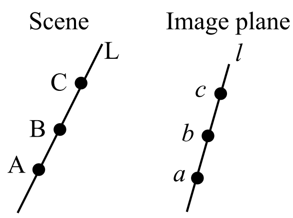


Figure 1.8: Example of not preserved ratios. It holds that $\frac{\overline{AB}}{\overline{BC}} \neq \frac{\overline{ab}}{\overline{bc}}$.

Vanishing point

Vanishing point Intersection point of lines that are parallel in the scene but not in the image plane.

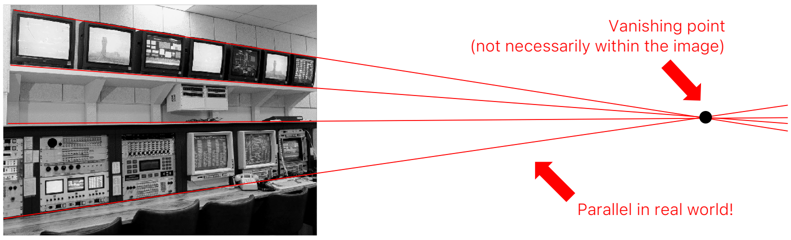


Figure 1.9: Example of vanishing point

1.3 Lens

Depth of field (DOF)

Depth of field (DOF) Distance at which a scene point is on focus (i.e. when all its light rays gathered by the imaging device hit the image plane at the same point).

Remark. Because of the small size of the aperture, a pinhole camera has infinite depth of field but requires a long exposure time making it only suitable for static scenes.

Lens

Lens A lens gathers more light from the scene point and focuses it on a single image point.

This allows for a smaller exposure time but limits the depth of field (i.e. only a limited range of distances in the image can be on focus at the same time).

Thin lens equation

Thin lens equation $\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$