Master COSI, CIMET, MLDM and 3DMT - Computer Vision course

Exam March 2017 - 3h without documents

(6 parts with a total of 14 questions accounting for 29 points, the exam will be scored for 24 points)

First Name : ______________________________

Family Name : ______________________________

Part 1 *(8 points)*: **Polar Rectification**

In the following exercises we will study the rectification of stereo image pairs. Rectification refers to the process of warping the images such that the epipolar lines are parallel to the horizontal axis and corresponding epipolar lines between the two images are aligned. We will in particular consider the method proposed by Pollefeys et al. [1] in their paper "A simple and efficient rectification method for general motion". To get started, read the paper thoroughly.

Figure 1: Example pair of stereo images.

Question 1 *(1 point)***:** What is the advantage of having epipolar lines parallel to the horizontal axis and aligned in both images?

Solution : limit the stereo matching step to a 1D search for corresponding points (see lecture 6)

Question 2 *(2 points)***:** What is the input to the algorithm described in the paper and what is its output?

Solution:

> Input: pair of images. The only required info is the oriented fundamental matrix (see equation (1) which describes the transformation that exists between the two images.
>
> Output rectified pair of images

See also http://www.cse.psu.edu/~rtc12/CSE486/lecture20_6pp.pdf

Question 3 *(4 points)***:** Summarize the proposed rectification method by outlining its main concepts

Solution:

-step 1: determining the common region from the extremal epipolar lines that touch the outer image corners (see fig.4)

- step 2: determining the distance between epipolar lines (see fig. 5)
- step 3: constructing the rectified image (non linear warping) line by line (see fig. 6) using the homography transformation described by equation (3).

Question 4 *(1 point)***:** In which situation does the proposed rectification method have a significant advantage over other rectification approaches?

Solution:

In some cases traditional rectification schemes these approaches yield very large images (i.e. when the epipole is close to the image or when the epipoles are located in the images this would have to results in infinitely large images) or cannot rectify at all as they encounter problem. These problems are mainly due to the fact that the matching ambiguity is not reduced to half epipolar lines. Here the approach proposed preserve the length of the epipolar lines and by determining the width independently for every half epipolar line (see fig. 1).

Part 2 *(4 points)***: Camera calibration**

In order to identify and locate points in an image, the camera which produces the image has to be calibrated. Therefore, the intrinsic and extrinsic parameters of the camera (system) have to be determined.
Consider a single camera C with the following intrinsic parameters:

- f the focal length
- kx and ky as the resolution of the camera in x and y direction
- hx and hy as the x and y coordinates of the camera

and the following extrinsic parameters:

- R the rotation matrix of the camera to the world coordinate system
- t the translation vector between the optical center of the camera and the center of the world coordinate system.

In homogeneous coordinates the projection p of a point P onto the image plane is then computed with a projection matrix M as:

p =M P = K (R t) P

with the intrinsic camera calibration matrix K defined as

$$\boldsymbol{K} = \begin{pmatrix} fk_x & 0 & h_x \\ 0 & fk_y & h_y \\ 0 & 0 & 1 \end{pmatrix}$$

Question 1 *(1 point)***:** Which parameters can be calibrated given a plane calibration pattern?

Solution: the XY coordinates of a 3D point P can be estimated if the intrinsic (5 unknown parameters) and extrinsic (6 unknown parameters) parameters are calibrated.

Question 2 *(2 points)***:** How can they be estimated?

Solution: Using a planar calibration model (e.g. a chessboard pattern) proposed by Zhengyou Zhang or the two-step algorithm of Roger Y. Tsai.

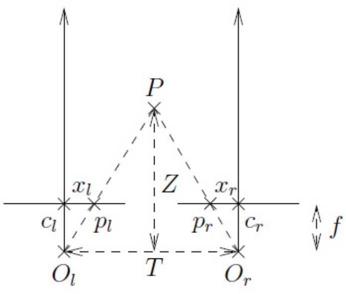
Question 3 *(1 point)***:** Which parameters cannot be calibrated? How can this problem be tackled?

Solution: the Z coordinate cannot be estimated. This problem can be tackled by moving the camera in another position (to simulate a stereo setup).

Part 3 *(4 points)***: Accuracy of stereo systems**

Consider the stereo system as displayed below. Or and Ol denote the optical centers of both camera systems. T is the baseline of the stereo system, i.e., the distance between the optical centers. P is a point which is projected onto its image point pr in the right camera and pl in the left camera. xr and xl denote the distance between the projections of P and the principal points of the cameras cr and cl.

The depth Z of P is computed as Z = f * T/d where the disparity d = xr - xl.



Question 1 *(2 points)***:** Estimate the accuracy of this stereo system assuming that the only source of noise is the localization of corresponding points in the two images. Discuss the dependence of the error in depth estimation as a function of the baseline width and the focal length.

Solution

The two cameras of the stereo system are parallel and at the same distance along the Z axis, so the epipolar lines on the two images are aligned (no need to rectify the two images). Then the two points xr and xl which are localized on the same epipolar line are necessary at the same distance along the Z axis.

Therefore, the error in depth estimation (δ Z) depends only of the error in disparity estimation (δ d) as:

$$\delta Z = f * T / (\delta d)$$

Question 2 *(2 points)*: Using your solution to the previous question, estimate the accuracy with which features should be localized in the two images in order to reconstruct depth with a relative error smaller than 1%.

Solution

The relative accuracy of x and y coordinates (between left and right images) is defined by:

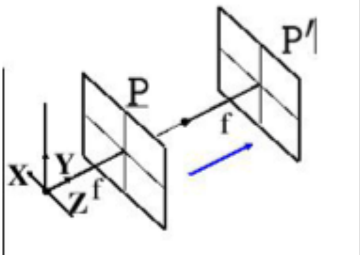
$$\delta d = f * T / (\delta Z)$$

with (δ Z) = 1%

then δ d = 100 *f *T

Part 4 *(4 points)*: **Epipolar geometry**

Question 1 *(2 points)*: Estimate the essential matrix between two consecutive images for a forward translating camera (see image below) . What is the equation of the epipolar line for the point p=[x y 1] ?



$$\begin{pmatrix} 0 & -K & j \\ K & 0 & -i \\ -j & i & 0 \end{pmatrix}$$

Solution:

In general $E = t_S R$

where t_S *is a skew-symmetric matrix related to translation vector.*

For a forward translating camera (see figure), we have:

$$R = I, \text{ and } t_S = \begin{bmatrix} 0 & -t_Z & 0 \\ t_Z & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Therefore

$$E = \begin{bmatrix} 0 & -t_Z & 0 \\ t_Z & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

From $l = E\,p$ the epipolar line for a point $p = [x\ y\ 1]$ is:

$$l = \begin{bmatrix} 0 & -t_Z & 0 \\ t_Z & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} -t_Z\, y \\ t_Z\, x \\ 0 \end{bmatrix}$$

Question 2 *(2 points)***:** Assume a pair of parallel cameras with their optical axes perpendicular to the baseline. How do the epipolar lines look like? Where are the epipoles for this type of camera system?

Solution:
The epipolar lines are parallel to each other (also, in the direction of the image y axis). If the optical centers of the camera are on the same height, the corresponding epipolar lines will be on the same height - one line for both images. The epipoles are in the infinity.

Part 5 *(6 points)***: Knowledge questions**

True/False Questions. The following statements are either true or false. Circle the correct answer to the left of the question. If you circle an incorrect answer, you will be penalized one point. If you do not circle an answer, you will not be penalized.

Questions 1 *(3 points)***:**

True False It is impossible to estimate the intrinsic parameters of the camera given a single image even with prior information about the scene.

Solution: False. Single view metrology provides us with methods for estimating the intrinsic parameters of the camera from a single image.

True False Assume that we identified two perpendicular planes in an image of the 3D world. We estimate two vanishing points (one on each plane) in the image from a pair of parallel lines on each of these planes. Using these two vanishing points and no other information from the scene, we can in general estimate the camera parameters.

Solution: False. We need three vanishing points to setup three scalar equations to recover $\omega = (KK^T)^{-1}$.

True False The vanishing point associated with a line in 3D space (when viewed through a pinhole camera) can never be a point at infinity.

Multiple Choice Questions

Question 2 *(3 points)***:**

A good corner detector identifies unique key points to characterize the image. Which of the following small window of pixels would yield the best key point? **Circle the most correct answer**.

A. A window containing a homogeneous set of pixels
B. A window containing a single line
C. A window containing intersecting lines
D. A window containing a pair of parallel lines

How many degrees of freedom are in the camera matrix K if skewness may occur, it has square pixels, and that the origin of coordinates in the image plane may not be at the principal point. **Circle the most correct answer**.

A. 2
B. 4
C. 5
D. 8
E. 11

Which of the following property applies if two cameras are rectified? **Circle all that apply**.

A. Epipolar lines are vertical
B. The cameras can differ by an arbitrary rotation with no translation
C. Epipolar lines will intersect at infinity
D. Fundamental matrix reduces to an identity matrix
E. None of the above

Part 6 *(3 points)***: Depth perception**

Let us consider the following monocular or binocular cues

A. Relative Size: If two objects are roughly the same size, the object that looks the largest will be judged as being the closest to the observer

B. Relative Height: This is depth perception in which objects higher in our field of vision are perceived as being further away

C. Relative Clarity: This is depth perception when hazy objects appear farther away than sharp, clear objects

D. Interposition: This is depth perception when one object partially blocks our view of another, we perceive the object that is blocking the other as being closer

E. Linear perspective: This is depth perception when, the more the parallel lines converge, the greater their perceived distance

F. Retinal disparity: This is the way that your left and right eye view slightly different images

G. Convergence: In order to perceive depth properly, your eyes must move slightly inward or converge. In so doing, people are able to determine if objects are close to them or far away.

Which of these visual cues induces depth perception in the following images? **Circle all that apply**.

C

D

A. Relative Size
B. Relative Height
C. Relative Clarity
D. Interposition
E. Linear perspective
F. Retinal disparity
G. Convergence
A. Relative Size
B. Relative Height
C. Relative Clarity
D. Interposition
E. Linear perspective
F. Retinal disparity
G. Convergence

E

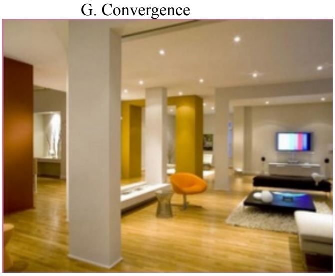
A. Relative Size
B. Relative Height
C. Relative Clarity
D. Interposition
E. Linear perspective
F. Retinal disparity
G. Convergence

A

A. Relative Size
B. Relative Height
C. Relative Clarity
D. Interposition
E. Linear perspective
F. Retinal disparity
G. Convergence

~~F~~
G

A. Relative Size
B. Relative Height
C. Relative Clarity
D. Interposition
E. Linear perspective
F. Retinal disparity
G. Convergence



A. Relative Size
B. Relative Height
C. Relative Clarity
D. Interposition
E. Linear perspective

F. Retinal disparity
G. Convergence