Exercise 3D Maschine Vision
Sample Solution

Prof. Dr.-Ing. Volker Willert

Technical University of Applied Sciences Würzburg-Schweinfurt

Sheet 1

In this exercise we cover the *introductory chapter*, as well as *3D camera systems* and basics of *stereo vision*. The questions are small-part and can be seen as examples of potential exam problems.

Task 1.1: Terms & Applications

1.1a)

Which application fields of 3D machine vision deal only with 3D reconstruction of 3D pose and/or 3D motion of rigid bodies without creating a 3D reconstruction of the environment?

- ☐ Visual SLAM
- × Visual odometry
- ☐ Multi-View Reconstruction
- ☐ Visual Servoing
- ☐ Structure from Motion
- × 3D Tracking
- ☐ Hand-Eye Calibration
- ☐ Optical Flow
- ☐ Ray Tracing

1.1b)

What is the difference between depth images and volume images?

Answer: Depth images only measure the depth profile of visible surfaces in a camera's field of view. Volume images measure, for example, the density of a volume and are usually displayed as a set of 2D sectional images (e.g. magnetic resonance imaging).

1.1c)

Name three real-world challenges that 3D vision algorithms must deal with?

Answer: Cast shadows, moving objects, transparent objects

1.1d)

What is the baseline? Why does the choice of baseline affect the choice of algorithms for 3D reconstruction from images?

Answer: The baseline is the distance between the projection centres of two cameras (or a moving camera at different points in time). The smaller the baseline, the smaller the disparity between the pixel correspondences. If the disparity is in the 0-1 pixel range, then differential methods (e.g. subpixel correlation) can be used to calculate the disparity. With a multiscale approach, the range can be increased to disparities of up to approx. 10 pixels.

Task 1.2: 3D Camera Techniques

1.2a)

When recording point clouds with 3D cameras, what conditions must be met to avoid large errors in depth measurement?

- [x] Objects do not move
- [ ] Light sources are moving
- [x] Camera system does not move
- [ ] Illuminance remains constant
- [ ] For a time-of-flight camera, the exposure time must be less than the run time
- [x] When configuring a stereo system, the ratio of baseline and disparity must be matched to the selected depth range

1.2b)

Explain the difference between the measurement principle of a stereo system consisting of two cameras and a projector compared to a stereo system with one camera and one projector? What is the minimum requirement for the respective principle for a measurement to be possible?

Answer: With two cameras and one projector, the projector is only used to project a pattern with unique features onto the scene. The relative pose between the projector and the cameras does not need to be known. With one camera and one projector, the relative pose between the projector and the camera must be known because the triangulation between the emitted beams of the projector and the received beams of the camera is calculated here.

1.2c)

What is the advantage of using amplitude modulated light compared to pulsed light in time-of-flight cameras? Is there also a disadvantage of amplitude modulated light?

Answer: With amplitude-modulated light, the depth can be calculated via the phase shift. Therefore, no precise time measurement is required, as is the case with pulsed light. However, the depth, which can be clearly calculated via the phase shift, is limited by the period duration of the modulation. The shorter the period duration, the more robustly the depth can be determined, but the smaller the maximum measurable depth range.

1.2d)

What systematic errors can be compensated for in a 3D camera by appropriate calibration?

- ☐ errors because of occlusion
- ☐ errors because of multipath reflections
- ☐ errors due to object motion
- × errors because of temperature changes
- × errors due to variations in the strength of the reflected light
- × error due to the selected measurement time duration
- × constant offsets in depth
- ☐ errors due to interference light influence
- ☐ errors due to camera noise

Task 1.3: Stereo Vision

1.3a)

What is special about the stereo configuration of two cameras? What are the advantages?

Answer: The image planes of the cameras are aligned parallel to each other in a common plane. This means that corresponding points can be found along horizontal scan lines with the same line index of the image matrix.

1.3b)

From the equation of depth reconstruction $Z = c\frac{b}{d}$, which holds for a calibrated stereo system, calculate the

1. depth resolution $\frac{\partial Z(Z)}{\partial d}$ and
2. the sensitivity of the disparity $\frac{\partial d(Z)}{\partial Z}$.

as a function of absolute depth Z.

Answer: Depth resolution results from the following calculation:
1. $\frac{\partial Z}{\partial d} = -c\frac{b}{d^2}$,
2. $d = c\frac{b}{Z}$,
from 2. in 1. it follows: $\partial Z = -\frac{Z^2}{cb}\partial d$

The sensitivity of disparity is given by: $\frac{\partial d}{\partial Z} = -c\frac{b}{Z^2}$

1.3c)

Calculate the distance D of a 3D point as a function of the quantities x, y, c, b and d. If you cannot assign the letters to the characteristic values, then refer to the lecture notes.

Answer: The distance is given by: $D = \sqrt{X^2 + Y^2 + Z^2}$. With $Z = c\frac{b}{d}$, $X = x\frac{b}{d}$ and $Y = y\frac{b}{d}$, the distance is: $D = \frac{b}{d}\sqrt{x^2 + y^2 + c^2}$.

1.3d)

Calculate the change in parallax in px/m (pixels per meter) for a stereo system with a baseline of $200mm$, a focal length of $f = 100mm$, and a pixel width of $10\mu m$ for distances $10m$ and $100m$, taking into account the approximation for far-field images: $c \approx f$.

Answer: Distance 10m: $\frac{\partial d}{\partial Z} = -100mm\frac{200mm}{(10m)^2} = -0.1m\frac{0.2m}{100m^2} = -200\frac{\mu m}{m} = 20\frac{px}{m}$ if the pixel size is $10\mu m$.
Distance 100m: $\frac{\partial d}{\partial Z} = -100mm\frac{200mm}{(100m)^2} = -0.1m\frac{0.2m}{10000m^2} = -2\frac{\mu m}{m} = 0.2\frac{px}{m}$ if the pixel size is $10\mu m$.

1.3e)

Compare the accuracy of the angular measurement of triangulation with stereo vision by calculating the angular change for a measurement accuracy of $\pm 1px$ in disparity a pixel width of $10\mu m$ and a camera constant of $c = 3mm$. Assume a disparity reference value of $100px$.

Answer: The relationship between camera constant c, disparity d and angle γ is given as follows: $\tan\gamma = \frac{c}{d}$. For a camera constant of $c = 3mm$ and a pixel size of $10\mu m$, changing the disparity by $\pm 1px = \pm 10\mu m$ by the disparity $d = 1mm = 100px$ results in the following angular deviations $\Delta\gamma$:

$$d = 10mm + 1\mu m \quad \rightarrow \quad \Delta\gamma = \tan^{-1}(\tfrac{3mm}{1mm}) - \tan^{-1}(\tfrac{3mm}{1mm+10\mu m}) = 71.565051 - 71.393336 = 0.1717°$$

$$d = 10mm - 1\mu m \quad \rightarrow \quad \Delta\gamma = \tan^{-1}(\tfrac{3mm}{1mm-10\mu m}) - \tan^{-1}(\tfrac{3mm}{1mm}) = 0.1721° \,.$$

This means that 1px error in the disparity measurement corresponds to approx. $0.2°$ angular deviation (tenth degree). As disparities can be measured to 0.1 pixel accuracy, the angular resolution of stereo systems is in the hundredths of a degree range.