Berlin University of Technology

Computer Vision & Remote Sensing

Prof. Olaf Hellwich

Final Examination
Introduction to Computer Vision
Summer term 2016

Name: . . . .   **Student ID number:** . . . .

Auxiliary resources: none

July 28, 2016

1. Pinhole Camera

4P

As is common practise, in the course we made use of the so-called Pinhole Camera Model.

a) Describe the Pinhole Camera Model. Draw a sketch and introduce variables.   2P

sketch: rays from object straight through projection center C onto image plane

All rays straight from object point through C generate image points.

b) What is the major difference between a real camera and the pinhole camera model? Despite of this difference the pinhole camera model is sufficient for many tasks. Give reasons why!   2P

A real camera has a lens bundling the rays generating a focussed image. In this way the image is brighter as it is generated by more light than the few rays going through a pinhole. However, the image is also subject to distortions.

The pinhole model is a sufficiently good geometric approximation allowing simple equations modelling the projection of object points into image points sufficiently correct.

2. Rasberry Pi Cameras for 3D Object Reconstruction

14P

In an exercise Raspberry Pi 3 single board computers equipped with Raspberry Pi cameras have been used to acquire images of a moving object simultaneously in order to compute a 3D reconstruction of the object.

a) What do you have to do in order to get a C++ program to run on a newly bought Raspberry PI which uses OpenCV library methods for processing video frames?   4P

b) Which problems occur when 3D reconstructions of moving objects are wanted? In your answer consider both a single/individual camera as well as the compound system consisting of several cameras.   2P

c) In order to obtain high-quality 3D reconstructions, cameras need to be calibrated. The camera calibration expresses the interior geometry, called interior orientation, of the camera. The most important calibration parameters are contained in the calibration matrix   5P

$$\begin{bmatrix} \Phi_x & \gamma & \delta_x \\ 0 & \Phi_y & \delta_y \\ 0 & 0 & 1 \end{bmatrix} .$$

Name and shortly explain the parameters of the calibration matrix.

d) The sketch below shows a stereo normal image pair. For a stereo normal image pair the two image planes Π_1 and Π_2 are located in the same, (common) plane, and the camera axes are parallel to each other and perpendicular to the baseline b connecting the projection centers O_1 and O_2 of the cameras. The image points $\mathbf{x}_1$ and $\mathbf{x}_2$ are used to compute the object space point $\mathbf{X}$.
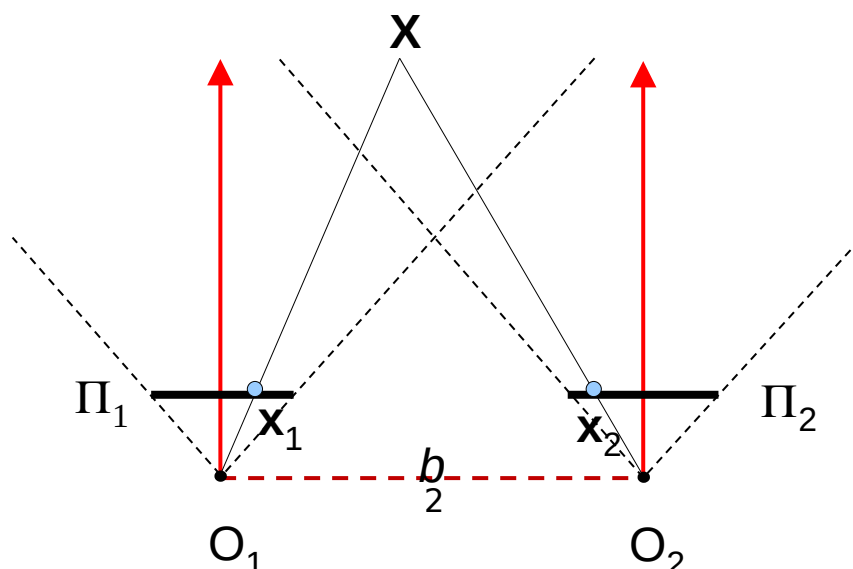
Derive a formula to compute the distance of the object space point from the baseline *b*? If necessary introduce further parameters.

$d = (b * Phi) / (x\_1 - x\_2)$

3. Vocabulary Search Trees

15P

Vocabulary Search Trees are, for instance, used to recognize locations that have been seen before.

a) Vocabulary Search Trees make use of "visual words". What is a visual word? Give a general description and an example with reference to a specific image processing method.   3P

Local group of pixels being a typical part of a (visual) "description" of an object by an image. Or: Typical part of an image of an object. Needs a suitable "parametrization".

Example: SIFT feature description (description part of SIFT rather than localization/extraction part)

b) Vocabulary Search Trees make use of k-means clustering. Describe the k-means clustering algorithm.   3P

Choose k
seed k potentional cluster centers
a) assign feature points to cluster centers
b) compute centers newly by averaging the assigned points
c) repeat until convergence

c)   The ability of the Vocabulary Search Tree method to identify objects or locations is – to some degree – based on a weighting making use of a "term frequency – inverse document frequency" (TF-IDF) scheme. Describe the TF-IDF scheme. Explain both components and their influence on the weighting of a visual word found in an image.   3P

TF: The more often a visual word occurs the more evidence is provided that the object contained is of a type having the visual word in its description.
IDF: If – however – the visual word occurs in many/all object type descriptions, it is not particularly meaningful w.r.t. the decision which object is in the image.

d) A tourist agency asks you to develop a system that allow tourists to use a   6P

smartphone image of a sight to get information about it.

Describe such a system making use of the Vocabulary Search Tree method.

In addition to the description also list the practical things (steps of work) you have to do in order to set up the system. Hint: Think in practical work in town as well as the visual location recognition software system.

- Acquire images of the town's sights
- Learning/training step:
  - extract features
  - build vocabulary search tree
    - k-meas clustering on several hierachy levels generates a tree structure
- Determine the relevance of a database image to the query image based on how similar the paths down the vocabulary tree are.
  - Each image's tree is described by a vector.
  - A database image is given a relevance score based on the normalized difference between query and database vectors.
  - Score has a TF-IDF character.
- Use the tree to find the image being most similar to a tourist's query image.
- Let the "app" give a description of the sight shown on the most similar database image.

Altogether **33 points** can be obtained.
A short and accurate style as well as a <u>clear handwriting</u> should be intended.
Pay attention to a clear and comprehensible preparation of sketches.