ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Computer Vision
AS 2016

Prof. Andreas Geiger *Prof. Luc Van Gool* *Prof. Vittorio Ferrari*

Final Exam

28 January 2017

First and Last name: ____________________

ETH number: ____________________

Signature: ____________________

General Remarks

- Remove all material from your desk which is not allowed by examination regulations. The following materials are allowed for this exam:
  - exam questionnaire & blank paper (both provided by us)
  - ruler/square & pen (pencil and red color pens are not allowed)
- Check that your exam questionnaire is complete.
- Fill in your first and last name and your ETH number and sign the exam. Place your student ID in front of you.
- You have **2** hours for this exam.
- **The answers can be written in the exam sheet**. If you answer each question on a separate sheet, put your name and ETH number on top of each sheet.
- Please do not use a pencil or red color pen to write your answers.
- You may provide at most one valid answer per question. Invalid solutions must be canceled out clearly.

	Topic	Max. Points	Points Achieved	Visum
1	Transformations	12		
2	Homogeneous Coordinates	13		
3	Optical Flow	5		
4	Camera Calibration	11		
5	Epipolar Geometry	12		
6	Shape-from-X	15		
7	Object Class Recognition	12		
Total		80		

Grade: ....

Question 1: Transformations (12 pts.)

Suppose we take images of a planar shape. We saw three cases, for which the deformations between views could be described as similarity, affine, or projective transformations.

a) Describe the conditions regarding the camera and the viewpoint (with respect to the planar shape) for each of the three cases. **6 pts.**

b) What is the number of degrees of freedoms (independent parameters) for these transformation groups? **3 pts.**

c) Under which of these groups are SIFT descriptors invariant? **3 pts.**

Question 2: Homogeneous Coordinates (13 pts.)

a) How many degrees of freedom does a line have in 3D? Explain your answer. **3 pts.**

b) Let $\mathbf{l} = (0, \frac{4}{7}, -\frac{8}{7})^T$ and $\mathbf{l}' = (-\frac{2}{3}, \frac{2}{3}, \frac{2}{3})^T$ denote two 2D lines in homogeneous coordinates. Calculate the non-homogeneous 2D point $\mathbf{p} \in \mathbb{R}^2$ where the two lines intersect. **5 pts.**

c) Let $\mathbf{q} = (a, b, 0)^T$ denote a 2D point in homogeneous representation. Determine the set of 2D lines which intersect $\mathbf{q}$. What do all these lines have in common? **5 pts.**

Question 3: Optical Flow (5 pts.)

The optical flow algorithm that we saw (Horn & Schunck version) corresponds to which of the following descriptions? (each time pick one option on this page)

a) It extracts ... **2 pts.**

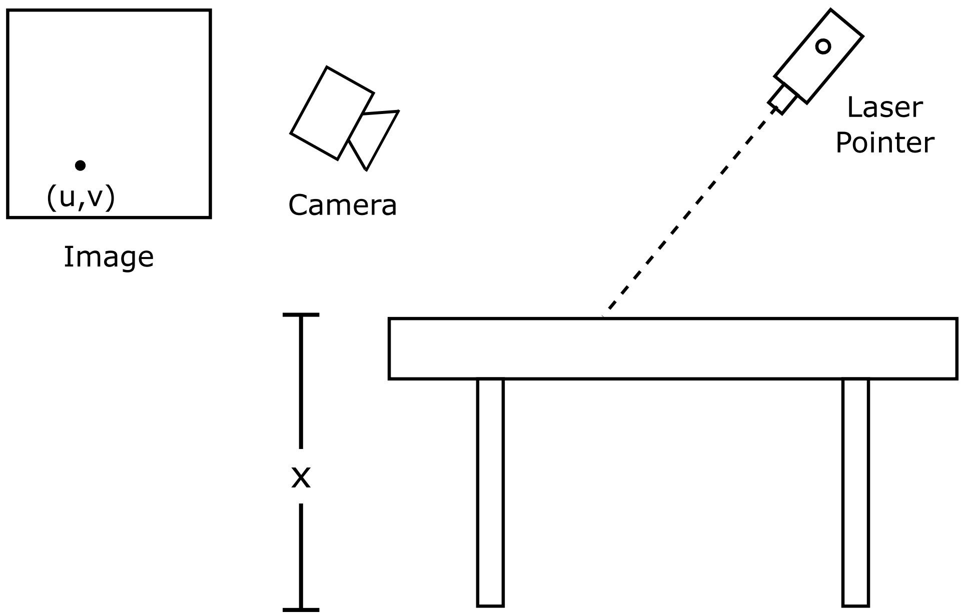
- ○ a 3D motion vector for every pixel.
- ○ a global 3D motion vector for an object that is being tracked.
- ○ a 2D translation vector for every pixel.
- ○ the 2D projection of the 3D motion (rotation and translation) of an object that is being tracked.

b) Its implementation is an example of ... **2 pts.**

- ○ regularization.
- ○ generalization.
- ○ Bayesian inference.
- ○ particle filtering.

Question 4: Camera Calibration (11 pts.)

There is a table whose height, x, can be adjusted. A laser pointer shines a narrow light beam at the table, which is seen by a pinhole-model camera (see figure below). The laser pointer and the camera are both rigidly mounted but their 3D position and orientation with respect to the table are unknown.



The homogeneous 3D coordinates of the point (in the world frame) where the laser beam is projected on the table, $\mathbf{Y}$, can be expressed as

$$\begin{bmatrix} Y_x \\ Y_y \\ Y_z \\ Y_w \end{bmatrix} = \mathbf{Y} = \mathrm{P}_x \begin{bmatrix} x \\ 1 \end{bmatrix} .$$

The matrix P_x is a projectivity that represents the intersection of a plane (the table) and a 3D line (the laser beam). Its parameters depend on the frame of reference of the pointer and the table, but for this problem they are unimportant.

a) By concatenating the pinhole-camera matrix P_c and the matrix P_x we can arrive to a linear expression that relates the height of the camera, x, and the camera coordinates $(u,\ v,\ 1)^T$. Write down that matrix equation. What are the dimensions of the combined matrices? **5 pts.**

b) How many degrees of freedom does this model (the combined matrices) have? **3 pts.**

c) How many laser pointer measurement at different heights are required to fully calibrate the camera? **3 pts.**

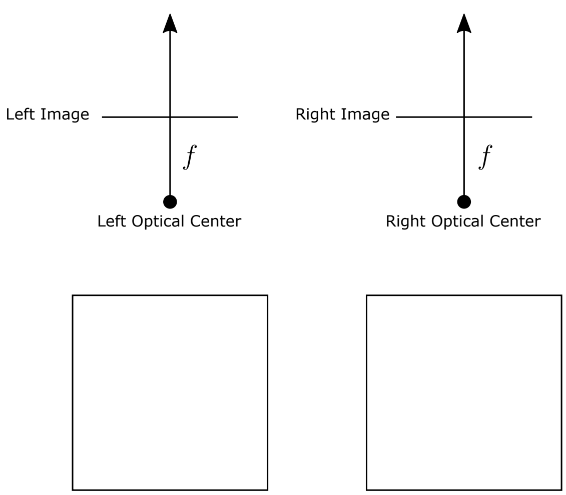
Question 5: Epipolar Geometry (12 pts.)

a) Give the compact equation that summarizes the epipolar relation between pairs of corresponding points, and in which the fundamental matrix plays a central role. **3 pts.**

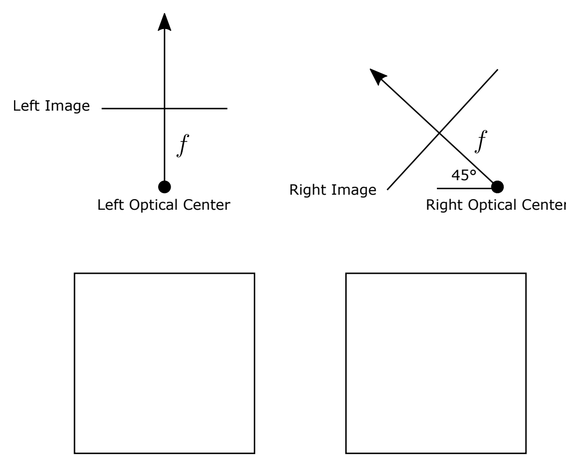
b) What are the dimensions of the fundamental matrix? (i.e., $F = \mathbb{R}^{M \times N}$; $M = ?$, $N = ?$) What are the degrees of freedom of the fundamental matrix? **2 pts.**

c) How many genuine point correspondences does one minimally need to *linearly* extract the fundamental matrix from only those correspondences? Explain. **3 pts.**

d) Below are two different top view configurations of stereo cameras (left side: rectified views, right side: non-rectified views). For each, draw (onto this page) how the epipolar lines would appear in both images. **4 pts.**



(a) Rectified Views



(b) Non-rectified Views

Figure 1: Top view configurations of stereo cameras. Draw your solution onto this page.

Question 6: Shape-from-X (15 pts.)

a) Describe the properties and relationship between the 3D convex hull, the visual hull and the photo hull. **8 pts.**

b) Name one algorithm for computing the visual hull and one algorithm for computing the photo hull. **2 pts.**

c) Describe the relationship between image irradiance, albedo, surface normal and light source direction in shape from shading for the lambertian case. How many observations are required if only the surface albedo and the normal direction are unknown? **5 pts.**

Question 7: Object Class Recognition (12 pts.)

a) What is the main advantage of using object proposals instead of sliding windows in the context of object detection? **3 pts.**

b) You have the choice between two classifiers: (1) a support-vector-machine (SVM) on top of a histogram of oriented gradients representation; (2) a convolutional neural network. What is the main advantage of using (2) over (1)? **3 pts.**

c) You are given a classification problem where each sample is represented by 1M features. Each feature is slow to compute. After training, which of the following methods typically runs faster (at test time): a linear SVM or AdaBoost? Explain your answer. **3 pts.**

d) Assume a visual word codebook containing the following 3 words: "O", "X", "+". Now consider bag-of-word representations of the two shapes A="O X +" and B="X + O". How well will the bag-of-words representation differentiate between shape A and B? Explain your answer. **3 pts.**