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.
- Answer each question on a separate sheet. Put your name and ETH number on top of each sheet. Only write on the question sheet if explicitly stated.
- 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.**

ANSWER ANSWER

Similarity: orthogonal views of the planar at all times; affine: general camera motion relative to the planar shape, but always keeping sufficient distance from the planar shape; projective: all cameras motions are allowed, also when moving close to the planar shape

ANSWER ANSWER

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

ANSWER ANSWER

Similarity: 4, affine: 6, projective: 8

ANSWER ANSWER

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

ANSWER ANSWER

SIFT descriptors are invariant under similarities only, not affinities, not projectivities.

ANSWER ANSWER

Question 2: Homogeneous Coordinates (13 pts.)

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

ANSWER ANSWER

4. Direction could be specified by 2 angles, origin point could be defined as a point on a plane (so 2 coordinates need to be specified).

ANSWER ANSWER

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.**

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

$\mathbf{p} = (3, 2)^T$. *Can be calculated either via the cross product and normalization or by converting the lines to the representation* $y = ax + b$ *and solving the equations (graphically or numerically).*

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

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.**

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

The set comprises all lines with tangent vector $(a, b)^T$ *and normal vector* $(b, -a)^T$. *All lines are parallel with respect to each other.*

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

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.

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

2D translation vector for every pixel

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

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

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

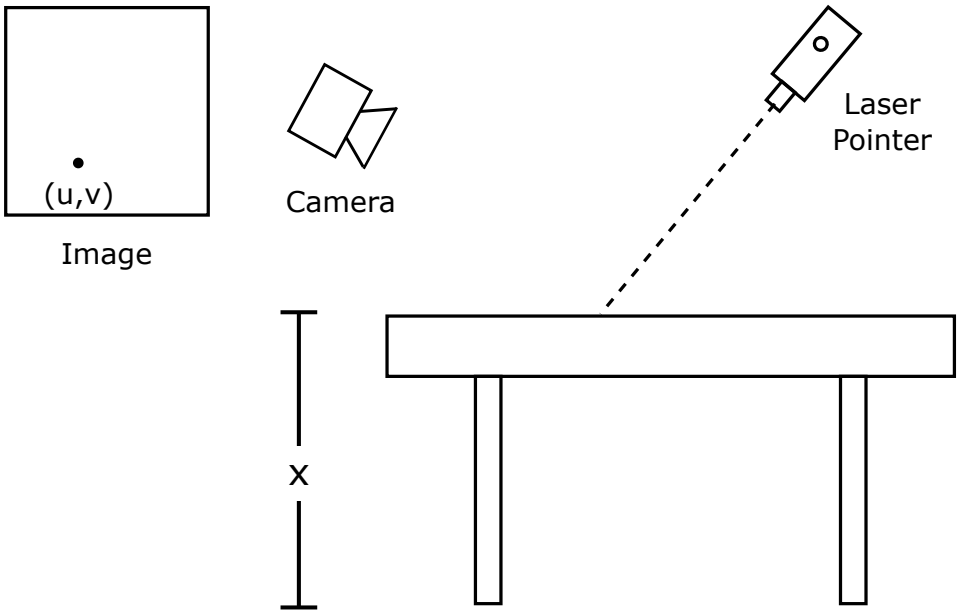
ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

Regularization

ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER ANSWER

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.

ANSWER ANSWER

Projecting the point Y using the classical 3x4 camera matrix yields P_τ. The final equation of the rig can be expressed as

$$\begin{bmatrix} \lambda u \\ \lambda v \\ \lambda \end{bmatrix} = \mathrm{P}_\tau \begin{bmatrix} x \\ 1 \end{bmatrix} = \begin{bmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \\ p_{3,1} & p_{3,2} \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix}$$

ANSWER ANSWER

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

ANSWER ANSWER

There are 5 DOF, one per matrix element up to scale.

ANSWER ANSWER

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

ANSWER ANSWER

Each measurement yields two constraints, so we need 3 such measurements.

ANSWER ANSWER

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.**

ANSWER ANSWER

$p'^t F p = 0$
p and p' are 3-vectors of homogeneous coordinates of corresponding points p and p'

ANSWER ANSWER

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.**

ANSWER ANSWER

F is a 3×3 matrix with 7 degrees of freedom.

ANSWER ANSWER

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

ANSWER ANSWER

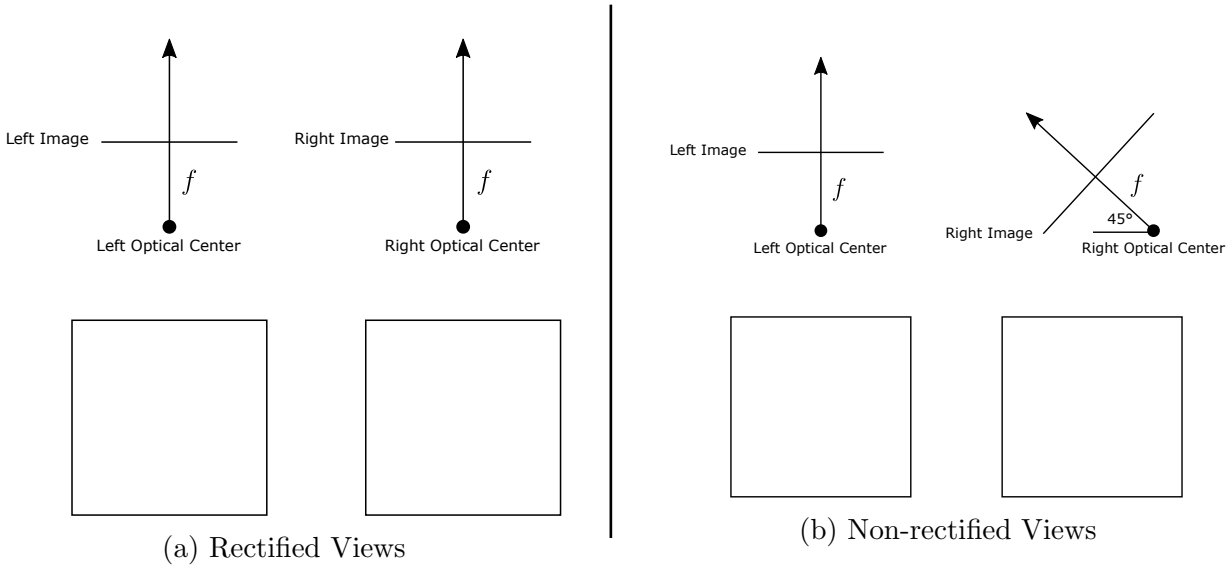
8, one equation per correspondence, not 9 because of homogeneous system of eqs.

ANSWER ANSWER

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.**

ANSWER ANSWER

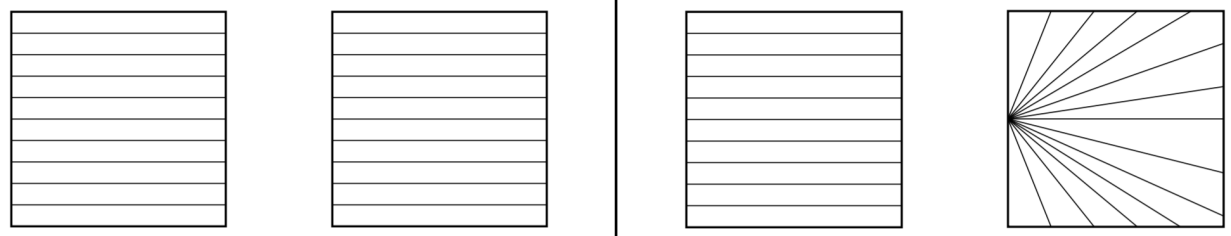
ANSWER ANSWER



(a) Rectified Views

(b) Non-rectified Views

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



(a) Rectified Views

(b) Non-rectified Views

Figure 2: Top view configurations of stereo cameras. Solution. **Rectified Views:** Epipolar lines should appear parallel in both images and to each other. **Non-rectified Views:** Epipolar lines should appear parallel in the left image and a pencil of lines with an epipole on the left side of the image for the right one.

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.**

ANSWER ANSWER

The 3D convex hull of a set of 3D points is the smallest convex set which contains all 3D points. The visual hull is the smallest volume which can be obtained by intersecting the generalized cones generated by all possible 2D shape projections of a 3D object. In 3D, the visual hull is a subset of the convex hull. The photo hull is the union of all photo-consistent volumes and the tightest possible bound on the true scene. It is a subset of the visual hull and - in contrast to the 3D convex hull and the visual hull depends on the texture of the 3D object.

ANSWER ANSWER

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

ANSWER ANSWER

Visual hull: voxel-based, marching intersections, exact polyhedral
Photo hull: voxel coloring, space carving

ANSWER ANSWER

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.**

ANSWER ANSWER

$I = c\rho\,\mathbf{n}^T\mathbf{s}$ *with image irradiance* I*, surface albedo* ρ*, surface normal* $\mathbf{n}$ *and light source direction* $\mathbf{s}$ *(the other parameters discussed in the lecture are absorbed into the constant* c *here, but of course the full equation is equally valid if properly described). If surface albedo and normal direction are unknown, 3 observations are required (3 unknowns).*

ANSWER ANSWER

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.**

ANSWER ANSWER

Object proposals cover most objects in an image using much fewer windows (about 1k vs 1M). Hence, object proposals enable using more powerful features and classifiers, which typically have much higher computational requirements.

ANSWER ANSWER

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.**

ANSWER ANSWER

CNNs learn to classify images (or windows) starting from pixels. They learn their own feature representation tailored to the task at hand. This results in better performance compared to using hand crafted features (including HOGs).

ANSWER ANSWER

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.**

ANSWER ANSWER

AdaBoost. It can choose a small subset of informative features to compute at test time.

ANSWER ANSWER

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.**

ANSWER ANSWER

It will not differentiate. The BoW historgrams for A and B will be identical (i.e. [1 1 1] in both cases).

ANSWER ANSWER