Master in Computer Vision Barcelona

UAB UOC UPC upf.

Module: **M4. 3D Vision** **Final exam**
Date: February 20, 2020
Teachers: Antonio Agudo, Josep Ramon Casas, Pedro Cavestany, Gloria Haro, David Reixach, Javier Ruiz, Federico Sukno
Time: 2h

- Books, lecture notes, calculators, phones, etc. are not allowed.
- All sheets of paper should have your name.
- Answer each problem in a separate sheet of paper.
- All results should be demonstrated or justified.

Problem 1 *1.8 Points*

a) (0.1p) In 2D geometry, name two invariants under similarity, affine and projective transformations, in this order.

b) (0.2p) What makes the points at ℓ_∞ under affine transformation different from the points at ℓ_∞ under projective transformation?

c) (0.5p) Decompose a general 3D homography as a product of three transformations: from projective space to affine space, from affine space to similarity space and from similarity space to euclidean space. Tipify the submatrices that comprise these three transformations. In the context of 3D reconstruction, what is the meaning of each one of these submatrices?

d) (1p) Describe in your own words the Gold Standard Algorithm for 3D homography estimation.

Problem 2 *1.75 Points*

a) (0.25p) Explain how a conic is represented in projective geometry. What is the conic equation?

b) (0.25p) Explain the concept of the image of the absolute conic and how it is related to the internal parameters of the camera.

c) (0.25p) Enumerate three different practical applications that use the image of the absolute conic.

d) (0.75p) Consider the algebraic method for camera resectioning seen in class that assumes some 3D to 2D point correspondences are known. Explain the resectioning problem, derive the optimization problem we need to solve and justify the minimum amount of correspondences we need to solve it.

e) (0.25p) State the differences between the camera resectioning problem and the Perspective-n-Point problem in terms of unknown and available data.

Problem 3 — *1.80 Points*

We want to compute the Fundamental Matrix F between two images I and I' of 100×100 pixels capturing the same secene from different viewpoints taken by the same camera. We compute two possible correspondences between the images as $p_1 = [2, 0]^T$, $p'_1 = [2, 8]^T$ and $p_2 = [10, 1]^T$, $p'_2 = [12, 2]^T$ and two epipolar lines in image U' as $l'_1 \equiv 2x - 7y + 52 = 0$ and $l'_2 \equiv 5x - 7y + 4 = 0$ (corresponding to point p_1 and p_2 respectively). Answer the following questions (consider the image coordinates origin at the bottom-left and positive going up and right):

a) Enumerate briefly the steps to compute matrix F using the 8-point algorithm.

b) Write the first two rows of matrix W that allows us to estimate the Fundamental Matrix F (expressed as a vector f) with a homogeneous system $Wf = 0$.

c) Propose a transformation to the pixel coordinates of p_1,p_2 and another for p'_1,p'_2 to reduce the numerical unstability of the 8-point algorithm.

d) Can any of the two correspondences be considered as an outlier?

e) Find the epipole e' in image I'.

f) Justify if the epipole is inside or outside of image I'.

g) Are the two images I and I' rectified? Why?

h) Would it be possible to compute the Essential Matrix E in this configuration?

Problem 4 — *1 Point*

a) (0.25p) Describe the triangulation problem, what are the unknowns and the available data.

b) (0.25p) How does the angle between visual rays affect the uncertainty in the estimation of the 3D point by triangulation given two views?

c) (0.5p) Enumerate the main steps of the view morphing technique proposed by Seitz and Dyer in 1996 in order to generate a new intermediate view given two images from two general points of views. Is it possible to generate a new view at any location of the virtual camera? Do the cameras need to be calibrated? Justify your answers.

Problem 5 — *0.90 Points*

Regarding the 3D shape recovery from a set of images:

a) (0.2p) Using voxel-based methods the problem is generally ill-posed. Explain how this affects these methods by specifying the assumptions or limitations on **two** of them.

b) (0.5p) Formulate the factorization method for projective reconstruction from two or more views given by the paper: *Sturm, P., & Triggs, B. (1996, April). A factorization based algorithm for multi-image projective structure and motion. In European conference on computer vision (pp. 709-720). Springer, Berlin, Heidelberg.* **Detail each step of the algorithm.**

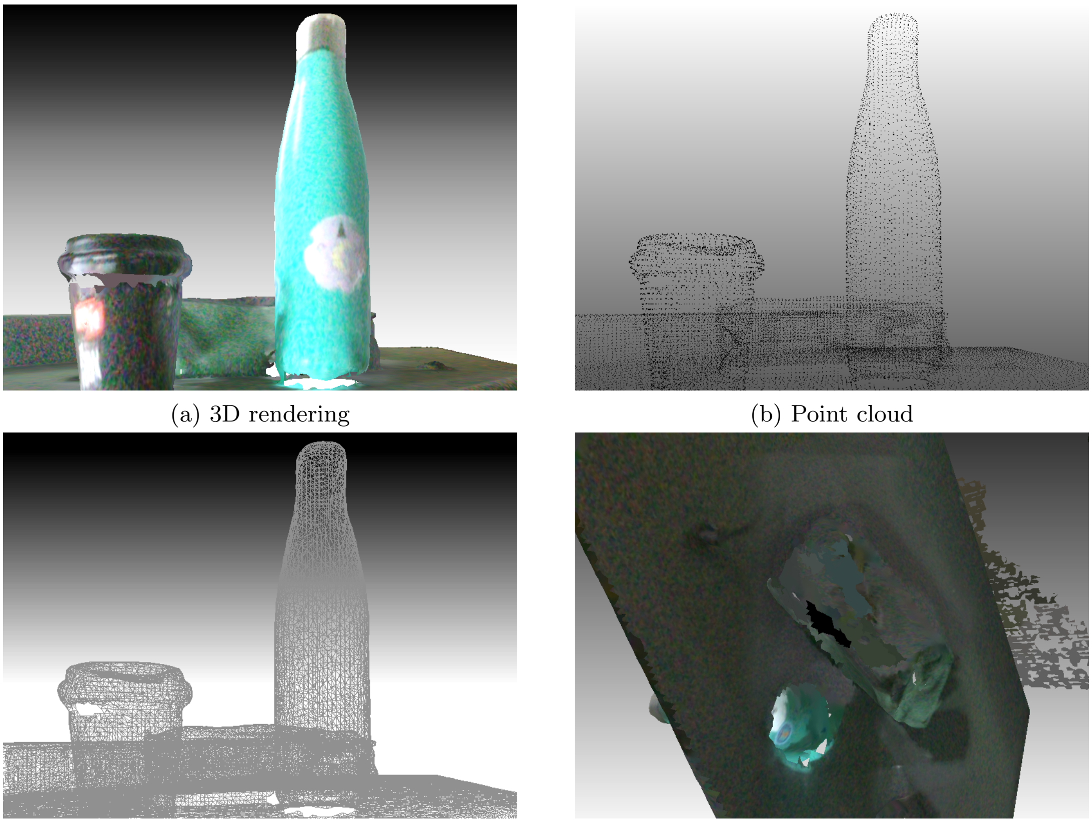
c) (0.2p) Explain the projective ambiguity inherent to image calibration and briefly describe a solution to this problem.

Problem 6 — *0.45 Points*

Formulate the structure-from-motion problem from 2D point tracks in a monocular video (indicating the corresponding size of every matrix) in terms of a measurement matrix $\mathbf{M}$, assuming an orthographic camera for: 1) a rigid shape, 2) a non-rigid one. To compute shape and motion by factorization, which rank do we have to enforce in every case and how can this be done?

Problem 7 *1 Point*

We captured a 360^o 3D scene using an RGBD sensor and turning the sensor around some objects laying on a table. The double nature (photometry+geometry) of the captured data is shown in the images of Fig. 1.



(a) 3D rendering (b) Point cloud

(c) Mesh (d) Bottom view of the 3D scene

Figure 1: RGBD capture of a 3D scene

a) Images in Fig. 1 (a), (c), and (d) show some parts of the geometry not being captured. Which are those parts? And, what may be the reasons why they have not been correctly captured?

b) Why should we avoid having any motion in the objects of a scene captured this way?

c) Does this mean that we cannot capture dynamic 360^o 3D scenes (i.e. RGBXYZT)? How can we capture them?

d) Even if we avoid the problems mentioned in the previous questions to capture a 3D dynamic scene, what would we always miss from a dynamic scene capture with an RGBD sensor?

e) If we want to get the "whole geometry" of a dynamic scene, what solution would you propose?

Problem 8 *0.40 Points*

Organized data structures may help in vision analysis tasks for pointclouds. Graphs and trees have been proposed to analyze (segment, detect, classify) point clouds.

a) Explain how adjacency, hierachy, captured primitives (visual or geometric features) and visual boundaries play a role for analysis tasks within the edge and node elements of graphs and trees.

b) What is the reason why graph and tree structures, specially for point clouds, are expected to perform better than raw data?

Problem 9 *0.90 Points*

Let $\mathcal{G} = \{V, E\}$ be a graph with vertices V and edges E. To facilitate the implementaiton of convolutions on the graph, we can use the spectrum of the graph, which can be obtained by eigen-decomposition of the Laplacian operator Δ such that $\Delta = \Phi^T \Lambda \Phi$. Assuming that we wish to convolve a function $f : V \to \mathbb{R}$ (e.g. a scalar function over the graph vertices) with a filter g (also defined on the graph domain).

a) Indicate how can we map function f onto the spectral domain (i.e. in analogy to the Fourier Transform).

b) Idem (a) for the inverse mapping (back from the spectrum to the graph domain, in analogy to the inverse Fourier Transform).

c) Provide the mathematical expression to compute $f * g$ (convolution) using the spectral domain.

d) Indicate one dissadvantage of this straight-forward use of the spectral domain as building block for a graph convolutional network.