Master in Computer Vision
Barcelona
UAB UOC UPC upf.

Module: **M6. 3D Vision** **Final exam**
Date: April 30, 2015
Teachers: Coloma Ballester, Josep Ramon Casas, Francesc Moreno, Gloria Haro, Javier Ruiz
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 Point*

(a) *(0.25 points)* How do we represent a homography in the 2D projective space?

(b) *(0.25 points)* How does a 2D homography act on points and lines?

(c) *(0.5 points)* Enumerate the different situations where two images are related by a homography.

Problem 2 *1.75 Points*

(a) Consider two image views of a plane objet. Let $\mathbf{x}_i$ in $\mathbb{P}^2$, $i = 1, \ldots, n$, be a set of points on the first image and let $\mathbf{x}'_i$ in $\mathbb{P}^2$, $i = 1, \ldots, n$, be a set of points on the second image such as, in pairs, they correspond: $\mathbf{x}_i \longleftrightarrow \mathbf{x}'_i$, $\forall i = 1, \ldots, n$.

(i) *(0.25 points)* How many corresponding pairs of points in general position do you need to compute the 2D homography H such that $\mathbf{x}'_i = H\mathbf{x}_i$, $\forall i = 1, \ldots, n$? (Recall that general position means that no three points are collinear).

(ii) *(0.75 points)* Describe the Direct Linear Transformation (DLT) algorithm to compute H.

(b) *(0.75 points)* Consider an image of a 3D scene containing plane objects. Explain the method of affine rectification via the vanishing line.

Problem 3 *0.75 Point*

(a) *(0.25 points)* Consider the pinhole camera model and consider a reference frame where its origin is the center of projection of the camera and the image plane is given by $Z = f$ where f is the focal length of the camera. Define camera centre, principal axis and principal point of the camera.

(b) *(0.5 points)* What is the general form of a finite projective camera matrix P ?

Problem 4 *0.5 Points*

Let H be a 3×3 homography that maps the points (u, v) of an image to points (x, y) on a 3D plane. Assuming that this matrix is known, and that we also know the 3×3 camera calibration matrix K, show how to compute the pose of the camera (rotation matrix R and translation vector t) with respect to the 3D plane.

Note: There is no need to show how to orthonormalize the matrix R. Assume we use a general routine $R_{ortho} = \texttt{orthonormalize}(R)$.

Problem 5 *2 points*

We plan to estimate the fundamental matrix F between two images I and I' taken by the same camera using the 8-point algorithm.

(a) *(0.5 points)* Given two correspondences $p_1 = (10, 5), p_1' = (12, 5)$, and $p_2 = (20, 30), p_2' = (25, 30)$ write the first 2 rows of the matrix W that allows us to estimate the fundamental matrix F (expressed as a vector column f) with a homogeneous system ($Wf = 0$).

If the singular value decomposition of the matrix W can be expressed as:

$$W = UD \begin{bmatrix} 1 & 0 & -1 & 1 & -1 & 0 & -1 & 1 & 1 \\ 0 & 0 & 0 & 1 & -1 & 0 & 1 & 1 & 0 \\ -1 & 1 & 0 & 1 & -1 & 1 & 1 & -1 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ -1 & 1 & 0 & 0 & 0 & -1 & 1 & -1 & -1 \\ 1 & 0 & -1 & -1 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 1 & 1 & 1 \\ -1 & 1 & 0 & -1 & 1 & 1 & 1 & -1 & 0 \end{bmatrix}^T$$

(b) *(0.25 points)* Obtain a first approximation of the fundamental matrix.

Suppose now that the singular value decomposition of the fundamental matrix obtained in the previous question is

$$F = \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}^T$$

(c) *(0.25 points)* Obtain a second approximation of the fundamental matrix that ensures all the epipolar lines cross at the same point (epipole).

(d) *(0.5 points)* Taken into account the fundamental matrix obtained above, are the two images I and I' rectified?

(e) *(0.5 points)* Enumerate three main disadvantages (or problems) with the 8-point algorithm and briefly comment a possible solution to them.

Problem 6 *1 Point*

(a) *(0.25 points)* Describe the triangulation problem, i.e. what are the unknowns and the known data.

(b) *(0.5 points)* Which is the energy that the geometric triangulation method minimizes and how do we use its solution to solve the triangulation problem? Limit the problem to the two-view case and define every variable that you use in the energy.

(c) *(0.25 points)* Describe the main steps of a local algorithm for stereo matching given a pair of rectified images.

Problem 7 *1 Point*

Consider the factorization method of Sturm and Triggs 1996 and assume we have three points seen by two cameras.

(a) *(0.5 points)* Write the measurement matrix and define its elements. Show how it relates to the 3D points and the camera projection matrices.

(b) *(0.5 points)* How do we extract a projective reconstruction once we have the Singular Value Decomposition of the measurement matrix?

Problem 8 *0.5 Points*

Describe how to estimate the image of the absolute conic given three vanishing points from three orthogonal scene directions and assuming that the camera has zero skew and square pixels.

Problem 9 *1 Point*

3D sensors

Rendering is the process of synthetically generating a raster image from a 3D model of a scene. Rendering, usually implemented in a graphics pipeline, tries to mimic the equivalent physical process of image acquisition through an RGB capture device. The synthesis is *computed* from the information in the scene model about (1) scene geometry, (2) camera viewpoint, (3) texture, (4) lighting and (5) shading.

In general, and only from the raster image resulting of a rendering process, it is not possible to recover the 5 items of information listed above. However, RGB+Depth (RGBD) capture devices perform an operation which, in some way, can be considered inverse to the rendering operation in Computer Graphics. Let's analyze this statement more closely.

(a) Could you provide a generic specification of the exact information recovered in an RGBD sensor?

(b) In commercial depth sensors (Kinect, Asus Xtion), how is the D component of this information recovered?

(c) How can we recover scene geometry (1) from the RGBD data? Would we do we need additionally?

(d) Is it possible to recover camera viewpoint (2) from RGBD data?

(e) Is there any way to recover reflectivity and lighting properties (3, 4, and 5) from RGBD data?

(f) Reason whether it would be possible to recover (3, 4 and 5) with any other imaging sensor device (active or passive, even if based in other imaging wavelengths)

Problem 10 *0,5 Points*

Meshing

We discussed the advantages of the new trend of directly processing the 3D raw data (point clouds) produced by 3D scanners. For image/scene analysis applications, what are the advantages and disadvantages of this strategy regarding actual XYZ values and connectivity when we compare it with working with a meshed input? Otherwise, what would be the main motivation for meshing?