# 3D Vision for Multiple and Moving Cameras (2024/25)

# Lab Evaluation Script

To carry out this exam, which takes place during two sessions, develop your code in separate 'Surname_Name_X.m' script files, where 'X' stands for the section number. Once you have finished, compress the generated '.m' files in a single file "Surname_Name.zip', connect to the Moodle delivery system and submit it to the link enabled for it.

To answer the sections/subsections of this exam (both textual or numerical answers and sample images), do it in a file 'Surname_Name.pdf', clearly indicating the section you are responding to and what image is represented (include them in real size, do not use subplot). This is the file that will be evaluated. The script files will only be evaluated in case of doubt, and only those that do not throw errors when executed individually (make sure of this fact putting `clear all` at the beginning of all and including all the required functions in the zip file).

**ASSIGNMENT OVERVIEW**

The objective of this evaluation assignment is to check the degree of assimilation of the practical procedures presented during the laboratory sessions. For this purpose, we propose to resolve step by step a practical problem of similar difficulty to that developed during those sessions.

The final aim of the proposed problem is to obtain a 3D reconstruction of a 3D object/scene of your choice. Which requires to select and calibrate a camera, to select an adequate object/scene according to some indications, to select an adequate number of views from the object, to extract and match feature points between views, to compute the fundamental matrix between views, to obtain a 3D points cloud reconstruction, and finally to represent object geometric elements over this points cloud.

**Section 1: Obtention of the intrinsic parameters of a camera**

The objective of this section is to calibrate a camera of your choice. As this will be the camera used throughout the rest of the evaluation, please be sure that your results are, at least, coherent.

First, select the camera you are going to work with. Remember that its internal parameters should be fixed, i.e., no changes in the focal distance nor changes in the aspect ratio should be performed after calibration.

In order to use a flexible method to calibrate a camera, we replace the use of a physical checkerboard with the use of your laptop or desktop screen. We provide a square checkerboard image, 1080 pixels size, which you should open to simulate a physical checkerboard enlarging it to cover all the screen.

1. Use the code developed during the assignments of 3DVMMC Unit I to calibrate your camera, using the checkerboard image provided. Include in your exam report the following data:
   - Size, in millimeters, of the checkerboard in your screen.
   - The set of images of the screen checkerboard that you have used for calibration.
   - The resolution of these captured images (in pixels).
   - The obtained matrix of internal parameters, $\mathbf{A}$.

Additionally, in the light of the computed $\mathbf{A}$ matrix, include a **detailed, reasoned and quantified** analysis of the following three aspects:

- Up to what extent are the pixels of your camera square?
- Which is the degree of coincidence between the principal point and the center of the image plane?

- Up to what extent are the axes of the image plane orthogonal?

2. In order to assess that you have really assimilated the concepts behind the Zhang's calibration method, create/select your own physical calibration pattern and repeat the previous exercise for this pattern: include in your report the corresponding captured images, the obtained $\mathbf{A}'$ matrix and the corresponding quantitative analysis of the obtained parameters as for the previous exercise. Finally, comment on the theoretical and practical relationship between $\mathbf{A}$ and $\mathbf{A}'$. You might decide either to create a physical pattern, to use a piece of wall or a picture or pattern instead, and you should decide how many points (and where) you should fix and measure in the pattern in order to get reliable results. The originality and rationale behind these decisions will mainly weight the evaluation of this subsection.

*Note 1: For Sections 2 and 3 (point matches and reconstruction), you may use the same scene and images. Figure 1 shows several viewpoints of an example scene.*

**Figure 1 Different viewpoints of a scene for point-matching and reconstruction**

*Note 2: For computational reasons, you may need to reduce the size of your images. If you do so, you must modify the camera parameters obtained in Section 1 accordingly. For example: if you have an 800x600 camera and you reduce your images to a 400x300 resolution (1/2-scaling), the coefficients in your intrinsic matrix must be scaled accordingly (except the scale coefficient of the matrix: $K_{3,3}$).*

**Section 2: Finding local matches between several views of an object.**

In this section, you are asked to use the camera that you have calibrated in Section 1 to capture several views of an object or a scene. Then, for pairs of views, you are asked to detect, describe, and match key-points using the toolboxes studied in Unit II lab sessions. You are asked to select a detector, descriptor and matching functions from the toolboxes, providing a rationale for the selection (based according to qualitative and quantitative indicators). We suggest adopting the following strategy:

1. **Object/Scene capture.** Prepare a scene composed of one or more objects. Move around the scene and capture different views of it with your—already calibrated, camera. Note that the more views you capture, the better will be the 3D point cloud reconstruction that can be obtained from them, but also the harder will be to effectively match feature points. The captured images will compose your dataset.

   Suggestions: Include textured 3D-objects (to ease feature detection) with sharp and rectilinear contours (to ease tasks in section 3). We also suggest capturing several views of the scene varying the angle and the distance of the camera with respect to the object/scene. Moreover, a nice experiment could be challenging the detectors/descriptors, by altering additional scene capture conditions (e.g., light, shadows, or occlusions,).

2. **Detection, description and matching of feature points.** Select some pairs of views in your dataset, you do not have to be exhaustive, i.e., not every possible pair needs to be studied, but a representative set should be selected. Extract and describe feature points for each view of the pair, and match points between the views. Estimate the homography transformation between the views for each combination

and each selected pair of views and evaluate its correctness according to the correspondences.

**Qualitative and quantitative evaluation.** Estimate the fundamental matrix between pair of views for every selected pair of views and the selected detector/descriptor/matching functions/parameters (functions setup). To this aim, we suggest the use of MatLab's **estimateFundamentalMatrix** function.

In order to select the best pairs of views and functions combination you may qualitatively evaluate the quality of the estimated fundamental matrix through the vgg_gui_F.m function of the vgg-mvg toolbox. Furthermore, **points for this section will be given if you also propose a strategy for evaluating quantitatively the quality of the estimated fundamental matrix**.

3. **Selection.** Choose the best four pairs of views and functions setups according to the results obtained in the previous stages.

   Include your reflections, numerical and graphical results in your report—use at most three pages for this section, exceeding this limit has a negative effect in the section score. Your report may contain in three separate subsections:

   a. A mosaic representation of your scenario, including all the captured views (see MatLab's **montage** function). Also include a brief description of the challenging factors that it contains.
   b. For some of the explored pair of views and functions setups:
      - A brief argumentation on the selection of the pairs of views and functions setups that determine the evaluated setups.
      - Numerical data in the shape of tables or graphs comparing the different setups.
      - A discussion on the performance of the explored setups according to the complexity and the challenging factors of the captured scene, the individual capabilities of the selected methods and their parameters.
      - A discussion on the quality of the estimated homography and fundamental matrix according to the correspondences and the scene nature.
   c. Only for the best combination:
      - The pair of images with the correspondences overlaid on them.
      - The estimated homography.
      - The warped images.
      - The estimated Fundamental matrix.
      - Screen captures of the vgg_gui_F.m GUI.

**Section 3: 3D reconstruction and calibration**

In this section, you will use the intrinsic parameters of the camera, obtained in Section 1 and the feature point matches between images, obtained in Section 2, to obtain a 3D reconstruction of your object/scene (a 3D point cloud). To achieve this goal, you will use the strategies that you have used and developed during the lab sessions of Unit III of the course. We suggest that you follow this strategy:

1. Compute consistent point matches among *N* views. To that end, you can use the **n_view_matching** function that is provided (you are also allowed to modify this function or use other strategy to match feature points among all views). As an input to this function, you must provide interest points and descriptors in the same format that was used in the lab sessions of the 2nd Unit of the course. Use the results in Section 2 to decide which detector/descriptor to use and the separation between consecutive images of the scene. In the report:

   a. Provide the images that you have used for the N-view point matching, indicating the detected interest points in each of them.

2. Compute the Fundamental matrix and an initial projective reconstruction from 2 of the cameras. In the report:
   a. Provide the images that you have used for the estimation of the Fundamental matrix, indicating the detected interest points and point matches.
   b. Provide the mean re-projection error and the reprojection error histogram.

3. Improve this initial reconstruction by means of a Projective Bundle Adjustment, using a higher number (maybe all) of your images. In the report:
   a. Provide the mean re-projection error and the reprojection error histogram at two points: (i) after the resectioning step, and (ii) after the Projective Bundle Adjustment step.
   b. Comment on the justification of the different re-projection error values in 2.b and the two steps of 3.a.

4. Re-compute the Fundamental matrix between two of the cameras, using the projection matrices obtained after the Projective Bundle Adjustment step. For this purpose, you can use the **vgg_F_from_P** function from the vgg-mvg toolbox.

5. Use the properties of the Essential matrix (between two cameras) to obtain a Euclidean reconstruction of the scene (use the re-projected points obtained after the Projective Bundle Adjustment step). Remember that you have already computed the intrinsic parameters of your camera from Section 1.

6. Obtain the Euclidean projection matrices for all the cameras by means of resectioning techniques. In the report:
   a. Provide the mean re-projection error and the reprojection error histogram (aggregated for all cameras)
   b. Provide illustrative results (several viewpoints and the 3D Matlab figure) of your 3D point cloud reconstruction. In the report, include figures that include all cameras and the scene, and others with scene alone (for better interpretation).
   c. **Extra**: Provide illustrative results (several viewpoints and the 3D Matlab figure) of an "improved" 3D point cloud. You can improve your point cloud with strategies such as: line segments that connect points that are joined by straight lines in your object/scene, "painting" your point cloud with RGB values of the pixels, cluster points from different objects and "paint" them with different colors, etc.