Projects

within the framework of the course:

image processing & computer vision

121091

date: 13 March 2018

F. van der Heijden

mailto:F.vanderHeijden@utwente.nl

Introduction

The project is the concluding part of the course in which students show their academic competence and skills with respect to image processing and computer vision.

Groups

- Groups of two persons.
- Working individually is allowed, but cannot be used as an excuse for delivering half of the work.
- **Working together does not mean "full task division":**
    - Use your partner for sparring.
    - Each student should be able to defend all parts of the project.

Subjects

	Title	Possible ingredients
1	Virtual advertising	• line/corner/edge detection • camera calibration • 3D vision • perspective projection • augmented reality
2	Man over board	• camera calibration • image stabilization • 3D vision • tracking and prediction • point detection/localization
3	3D face reconstruction from 3 images	• stereo camera calibration • image rectification • stereo matching • 3D point clouds • 3D meshes • Iterated Closest Point registration
4	3D tracking of facial point features	• camera calibration • anatomical landmark detection • tracking and prediction

Deadline report:

April 22, 2018 (end of 3rd quartile)

Repository of resources

Resources (templates for report, data, and additional software) will be made available via:

link to ipcv project repository (This link expires on 1 June)

Report

- Usage of one of the templates (word, latex) provided in the repository is **mandatory**.
- Final format: pdf. (**don't** send word documents).
- Maximum 7 pages (without the references and the annexes)
- Describe your algorithm such that the principle of operation becomes clear. For that, use 'mathematical style' pseudo code. See: https://en.wikipedia.org/wiki/Pseudocode and pseudo-code examples provided in https://en.wikipedia.org/wiki/Category:Articles_with_example_pseudocode. The report itself does **not** contain Matlab code (you should add your listing as an annex).

Structure of the report:

1. Intro:
    - Context
    - Statement of the problem with research question and sub-questions.
2. Methods and Materials
    - Materials used
    - Analysis of the problem and the strategy for the solution (i.e. the explanation of the algorithm)
    - Description of the experimental set-up including the evaluation method
3. Results
    - Here, you give the results of what is described in Section 2: tables, images and/or graphs.
    - The accompanying text of these tables, images and/or graphs clarifies how they are related to the methods described in Section 2.
    - If you have any remarkable observations, mention and describe them here, but without an interpretation, explanation, or meaning.
    - Do not introduce new methods in this section.
    - Do not give an interpretation or a judgement of these results.
4. Discussion/Conclusion
    - Give a statement about the research question.
    - Give an interpretation and judgement of the results.
    - If you had any remarkable observations, discuss them here to give them meaning (interpretation, explanation, implication).
    - Describe limitations of the study.
    - If applicable, compare your results with results from literature.
    - Conclusion: describe the overall implication of the results to the original problem statement (or research questions).

   Literature list
   Matlab listings

Supervision

- During the scheduled lecture hours.

Grading

After the report is finished, you make an appointment for a discussion with F. van der Heijden and/or M. Abayazid via a doodle that will published on Blackboard. The intention of the discussion is to evaluate the student's understanding of the subject. As a consequence, the following questions can be expected:

- Topics that were addressed in the exercises.
- Topics that are related to the project.

The final mark is given directly after the discussion. The grading is based on the reports of the short exercises, the written exam, the project, on the understanding of the topics, and on the student's convincingness and professionalism as shown during the discussion. The final mark is the result of the overall impression of the teacher.

Literature

Syllabi will be published on Blackboard. These syllabi cover the following topics:

1. Math for computer vision
2. Camera models and camera calibration
3. 3D Surface reconstruction from stereo images
4. Key point detection and matching
5. 3D estimation and visual navigation

The syllabi contain ample examples that demonstrate Matlab functions for 3D Computer Vision

Project 1: Virtual advertising

Many sport events are broadcast on television with an overlay providing an extra advertising space. An example is shown in the adjoining image which shows advertising for 'AFAB' as an overly next to the goal in a football match.

The placement and projection of such a virtual advertisement is based on a couple of aspects:
1. The 3D localization of the side lines of the game with respect to a chosen world coordinate system.
2. The position and orientation of the camera with respect to this world coordinate system.
3. The intrinsic parameters of the camera.

There is no specific calibration object in the scene. Instead, objects with known length, like the court lines of the sport and goal posts, etc, should be used. In computer vision, this falls under the umbrella of auto-calibration (self-calibration). See, for instance, [1].

The assignment

The problem statement is how to find and to test a suitable method to accomplish virtual advertising. For that purpose, you select 4 videos clips of a play in which the court lines (field lines) are sufficiently visible during the panning and rotation of the camera. High quality videos are recommended. Make sure that also some 3D reference objects with known sizes are visible. It is advised not to select clips in which the cameras zooms in or out. Make sure that in the video at least two parallel lines are visible in two orthogonal directions (so, at least 2 times 2 lines should be visible).

Subtasks might be:
- Location of calibration points, e.g. the crossings of some court lines.
- Intrinsic camera calibration using the known 3D positions of reference objects. It is allowed to ignore nonlinear lens deformations
- External camera calibration: the pose of the camera relative to the 3D reference objects.
- Tracking of 2D points and/or lines in the movie.
- Based on these tracked points and/or lines, camera pose tracking during the movie.
- Projection of a virtual banner which has a rectangular shape in the real world and which is located near a court line.
- Creation of demo movies of the clips which show the found court lines and the banner.

You are allowed to adapt your line finding algorithm to the type of sport and the colours that have been used. For instance, a football play might need a different strategy than the volleyball play.

Grading:

This project has different levels of challenge:
a) The lines are detected correctly, and a banner is placed horizontally flattened on the surface of the field.
b) The lines are detected correctly, and a banner is placed near a court line and vertically aligned with respect to the surface of the field. One of the vanishing points of the projected banner is correct.
c) As in b), but also camera calibration has been correctly implemented. The prespective projection of the rectangular banner is correct. That is, in the videos there would be no geometrical difference between a virtual banner with a give size and a real banner with that size.
d) As in c), but now the banner is placed at an angle of 60 degrees w.r.t. the field. A shadow of the banner is also projected on the field as in the figure.

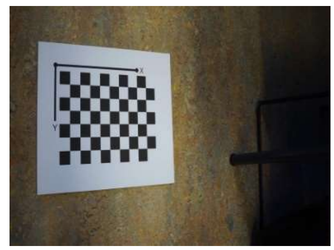
At least a) should have been implemented for a sufficient grade.

[1] Z. Zhang, "Camera Calibration," in *Emerging Topics in Computer Vision*, G. Medioni and S. B. Kang, Eds., ed: Prentice Hall Professional Technical Reference, 2004, pp. 4-43.

Project 2: Man over board

Finding someone lost at sea is a difficult task. Even when a “man over board” situation is timely detected, and the person is spotted nearby a vessel, keeping track of that person is difficult.

In this project, the possibility of tracking such a person in a wavy sea must be examined. Videos are available of a test buoy that is a stand-in for someone at sea. The first task of analysing such a video would be the detection of the object. Once detected and localized in the video, the next task is to keep track of the object so that a rescue vessel can approach it. The third task is to estimate the distance of the test buoy to the camera, which would be useful information of a rescue operation in a real situation.



Assignment: This project focuses on the second and the third task. Given the position of the test buoy in the first frame (by manual pinpointing), track the buoy in consecutive frames, and estimate its distance in meters.

Hints:
For tracking the object, different choices must be made.

- As a pre-processing step, consider stabilizing the images first. There are camera rotations along three orthogonal axes, i.e. tilting (rotations around the optical axis), panning (rotations around the vertical axis), and pitching (rotations up and down). To compensate, you need reference points/lines in the images. For which of these rotations is such a reference available? In which order of rotations do you apply these compensations? With which geometrical transformations do you compensate these rotations?
- Having estimated the position of the object in one frame, the localization of the object in the next frame starts with defining a search area, i.e. a region of interest (ROI), based on the previous estimate. What shape, size and position should this search area have?
- In some frames the object is not visible due to the heavy waves. In these cases, we should move on to following frames until the object is detected again. To do that, you need motion models.
- There are three different sources of motion in the image data: the camera motion, the travelling sea waves, and the motion of the buoy. We may safely assume that the buoy doesn’t move fast. Relative to the sea waves, it is almost static. The motion of the camera consists of rotations of which the effect is (perhaps partly) compensated by the pre-processing step. There is also a linear motion, e.g. up and down, but at a larger distance the effect of that in the image is similar to rotations. The motion of the travelling sea waves is more complex due to the perspective projection. Waves nearby seems to travel much faster than waves at a larger distance. However, at a larger distance, the optical flow within a small ROI, can be modelled as constant. The effect of rotations of the camera within this ROI can also be approximated as a uniform optical flow field. You can perhaps measure this optical flow field vector by means of a key point detection and matching, or otherwise by template matching using this blockwise motion model.
- The ROI shows travelling waves together with the buoy. The latter only if the buoy is visible. How do we detect whether the buoy is visible? Knowing the uniform optical flow of the sea waves, you might be able to apply motion compensation and image subtraction to nullify the waves.

For the estimation of the distance, a geometrical model should be applied incorporating:

- The average height of the camera above sea level. This height is given: 2.5 m.
- A spherical model of the earth surface, or an approximation of that.
- A perspective projection model of the camera. The intrinsic parameters of the camera should be obtained from calibration. For that, images of a chessboard are available. See the example. The size of a square is 40 x 40 mm^2. These images are suitable for usage in the calibration app of matlab.

Grading: the grade depends on the quality (robustness, accuracy, etc) and elegance of the solution, the thoroughness of the evaluation, and the quality of the report.

Project 3: 3D face reconstruction

The 3D geometry of an object can be defined in a computer model by describing the 3D surface of that object. A possible representation of that surface is by means of a 3D *point cloud*. That is a collection of 3D points that are represented by their 3D coordinates in a given coordinate system.

However, such a 3D point cloud is limited since it doesn't provide the connectedness between the points. A richer representation is a so-called 3D *surface mesh*. This is a collection of 3D points that form the corners of planar polygons. The 3D points are called *vertices*. The polygons are the so-called *faces*. Often, triangles are used for the polygon since a triangle is the only planar polygon that has no constraints on the 3D positions of the corners. This is in contrast with, for instance, a quadrangle.

The usual way to represent a 3D surface mesh is by enumerating the list of vertices, and then by defining each face by tabulating the three numbers of its vertices.

Stereo vision uses two images of the same object to find the 3D positions of points on the surface of that object. This can provide a partial surface mesh of the object. It is partial since only surface patches that are visible in both images can be reconstructed. To have a more complete surface mesh, the results of other pairs of images, taken from other points of view, must be merged.

Assignment: Three subjects are imaged from three different points of view. These three images are indicated by 'left', 'middle', and 'right'. Each subject is imaged with five different facial expressions. The adjoining figure provides an example. Also given are two sets of calibration images. A checker board has been imaged with different poses. The squares on the checker board has a size of 10x10 mm^2.

Assignment: Design a method for the creation of the 3D surface meshes of the faces of subjects. This is useful, for instance, for face recognition, but also for medical applications. An example is the grading of facial paralyse due to some neural disorder. .

Subtasks:
- The intrinsic and the external parameters of the camera should be obtained from camera calibration for which the checker board images are available.
- Non-linear lens deformation should be compensated for, and stereo rectification is needed to facilitate the dense stereo matching.
- Maybe a global colour normalization to make sure that the so-called 'Constant Brightness Assumption' holds true. A normalization could be applied with respect to mean and standard deviation of the colour channels.
- Stereo matching provides so-called disparity maps which is the input for depth map estimation.
- We don't want to have the background within the 3D surface mesh. So, you might want to develop and to use a background detection (maybe edge detection in combination with morphological operations).
- You also may want to create a map which indicates "unreliable disparities", so that unreliable depth can be removed from the mesh.
- To merge two 3D surface meshes to one mesh, you may want to use the "ICP algorithm (iterated closest points). Matlab implementations can be found on the internet.
- A rough assessment of the quality of the mesh is appreciated. How to quantify quality? Discuss your results.

Project 4: 3D tracking of facial features

Motivation: Patients undergoing surgery and/or radiotherapy in the oral region, especially the tongue, have the risk of limited tongue mobility with serious deterioration of oral functions, such as speech, food transport, swallowing, and mastication. The mobility of the tongue is expressed in the so-called 'Range of Motion'. To study the statistical correlation between the range of motion and a given treatment, this range of motion should be measured in a population of patients. This is to be done before and after the treatment.

Statement of the problem: To measure the range of motion, patients are asked to move their tongue to standardized, extreme positions, left, right, forward, downward and upward. A triple camera system is used to measure the 3D positions of the tongue tip. Of course, to compare these positions, pre- and post-treatment, the positions should be expressed with reference to a fixed, well-defined coordinate system that is attached to the head. How to derive these measurements from the recorded triple videos?

Assignment: Develop and test a method that is able to track the tip of the tongue and some facial landmarks so as to find the 3D positions of these points. Using these facial landmarks, define a 3D reference coordinate system that is attached to the face. Express the 3D position of the tongue tip in that coordinate system. Evaluate the method.

Tips:
- Images of checker board (size of the squares is 10x10 mm^2) are available to calibrate the cameras.
- Easy to interpret facial features, such as the nasal point, the subnasale (lowest point of the nose in the midsagittal plane), and the corners of the eyes, etc. are preferred.
- To initiate the system, it is allowed to manually pinpoint these points, including the tip of the tongue, in the first frame (e.g. Matlab function `getpts`).
- Maybe it is possible to enrich the description of some of these points by locally reconstructing the 3D surface surrounding these points. The computer vision library of Matlab contains several functions for this 3D reconstruction. See the syllabi.
- If at a certain frame, the method fails to track one or more points accurately, then automatic detection of this situation, so that the tracking can be stopped and after manually correction be continued, is not as serious as that the tracking just continues with erroneous results.
- Since we have three cameras, the system is partly redundant. This can be used to make the system more robust, but also to assess the system.