Computer Vision (CV) (English) – JMU Wuerzburg, Master, Spring 2021

There are **4 home assignments** roughly corresponding to the teaching sessions. They are counting for a 10% bonus in the final grade for this course and are meant for familiarizing the students with the written exam which will contain similar tasks.
Each home assignment (HA) has 8 topics that are weighted equally. The solutions to each HA should be sent as a single *pdf* by email at radu.timofte@uni-wuerzburg.de no later than May 26 for HA1, June 10 for HA2, July 1 for HA3, and July 15 for HA4, respectively. Each student shall include with her/his email full name and the details of the study program. The Subject of the email should be of the form:
"CV 2021 HA[#home_assignment] [master_program] [lastname] [firstname]"
where [#home_assignment] should be replaced with the number of the home assignment, the [master_program] with the corresponding short abbreviated program (like XtAI, LuRI) and [lastname] and [firstname] are the student's last name and first name, respectively.
The grades on the home assignments will be provided together with the final course grade.
Plagiarism and misconduct are not tolerated!
The grading of Computer Vision (taught in English) is as follows:

- **The final CV (Eng) grade (maximum 100%)** is composed from 100% written exam.
- Up to **10% extra bonus** will be awarded for those attending the lectures.
- Up to **10% extra bonus** will be awarded for those attending the exercise sessions.
- Up to **10% extra bonus** will be awarded for those sending their solutions to the HAs.
- For **exam** the students are required to prepare the contents covered during the lectures, exercise sessions, home assignments and found on the provided slides and lab materials. Examples of potential exam topics are found in the HAs.

Home assignment II

(due June 10th, 2021)

9. There are two viewpoints/images of the same object. What kind of interest points and descriptors should we use in order to establish correspondences? What is the maximum number of pixel correspondences that are needed in order to estimate the transformation between the images if it is know to be affine? What about if the transformation is assumed to be just inplane translation and rotation? Justify your answers.

10. What is the target of histogram equalization? Are there cases when the application of histogram equalization is not desirable? Justify your answers.

11. There are two viewpoints/images of the two different objects from the same category. What kind of interest points and descriptors should we use in order to establish correspondences? What is the minimum number of pixel correspondences that are needed in order to robustly estimate an affine transformation between the objects? What about if the transformation is assumed to be just inplane translation and rotation? Justify your answers.

12. Given the following pixel values: {1, 1, 1, 5, 100, 3, 1, 1, 1}, apply Otsu algorithm step by step and report the final optimal segmentation threshold. What about the result of applying Otsu for the pixel values {3, 3, 3, 3, 0}? Justify your answers.

13. Apply K-means and show the intermediary results for unsupervised segmentation in 2 classes of the following values on a row: 1, 1, 1, 5, 100, 3, 1, 1, 1. You can consider as starting seeds 2 and 6. Now, considering the same values and seeds, how can be adapted the algorithm if we know that 5 and 1 belong to different classes. Show intermediary steps/results. Justify your answers.

14. Given the face detection task and the Viola&Jones CVPR 2001 paper. What is the difference in train and test/inference complexity and, also, in robustness and accuracy between AdaBoost and Cascaded AdaBoost detectors? Justify your answers.

15. Let it be a set of N=100 image instances of object class “car”, T=100 images without “cars” and a pool of M=100000 images. The task is to find in the “car” images from the pool images. The basic processing is the extraction of SIFT feature descriptors for detected Harris keypoints, let’s say, on average, P such points in each image. Let’s say that we have three solutions:
(a) matching each “car” labeled image with each image in the pool, through matching of the extracted SIFT descriptors in each image and thresholding a matching score;
(b) representing each image over a vocabulary of H=100 visual words and matching “car” labeled images and pool images directly through histograms;
(c) representing the images as histograms (like described for (b)),normalized to sum up to 1, then training a linear Support Vector Machine (SVM) classifier on such histograms to distinguish between the labeled images N and T, and using this classifier to score on the M pool images.

(i)What is the most efficient and what is the least efficient from the 3 solutions?
(ii)What is the solution that would have the best accuracy? What about the worst?
(iii)What would be a combination of solutions that will trade-off between accuracy and time complexity?
(iv)What is the time complexity for each of the (a),(b), and (c) solutions? Use big *O* notation and report offline/training (if any) time complexity and testing time complexity.
Justify your answers.

16. All the cars images are instances of the same car model (let’s say Toyota Solara SLE coupe), but the images are captured by different cameras, at different viewpoints, in different locations, and the cars have different colors. How can be improved the solutions described at *Task 15.* in:

(i) interest points and descriptors;
(ii) matching scheme (would RANSAC help?);
(iii) given the answers at (i) and (ii), how the time complexity would change?
Justify your answers.