CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

THESIS REVIEWER'S REPORT

I. IDENTIFICATION DATA

Thesis title:	**Learning and Crafting for the Wide Multiple Baseline Stereo**
Author's name:	**Dmytro Mishkin**
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Electrical Engineering
Thesis reviewer:	Professor Andrea Vedaldi
Reviewer's department:	Engineering Science, University of Oxford

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **challenging**

How demanding was the assigned project?

The thesis looks at the problem of matching two or more images. This is a fundamental problem in computer vision and has been researched for more than thirty years. The aim of the thesis is not only to improve on this prior work, but also to systematize and assess it thoroughly in terms of real-world impact.

Fulfilment of assignment **fulfilled**

How well does the thesis fulfil the assigned task? Have the primary goals been achieved? Which assigned tasks have been incompletely covered, and which parts of the thesis are overextended? Justify your answer.

The project represents a very significant intellectual achievement. The work is very substantial, and so is the publication record of the candidate, much of which is a direct byproduct of this work. The quality and academic impact of the material are excellent.

Methodology **outstanding**

Comment on the correctness of the approach and/or the solution methods.

The thesis presents several nice insights in the design and learning of feature detectors and descriptors. It also studies the problem of combining these components in end-to-end system and the significant challenge of evaluating fairly and systematically such complex systems. The candidate clearly demonstrates a mastery of this area.

Technical level **A - excellent.**

Is the thesis technically sound? How well did the student employ expertise in the field of his/her field of study? Does the student explain clearly what he/she has done?

The thesis is technically sound. Besides several nice technical contribution, particularly impressive is the quality of the empirical evaluation, which is of extraordinary calibre.

Formal and language level, scope of thesis **B - very good.**

Are formalisms and notations used properly? Is the thesis organized in a logical way? Is the thesis sufficiently extensive? Is the thesis well-presented? Is the language clear and understandable? Is the English satisfactory?

The thesis generally reads well and is clear. The structure is almost everywhere appropriate and the language is used correctly.

Selection of sources, citation correctness **A - excellent.**

Does the thesis make adequate reference to earlier work on the topic? Was the selection of sources adequate? Is the student's original work clearly distinguished from earlier work in the field? Do the bibliographic citations meet the standards?

The thesis contains a large bibliography. This is not just a list of references; instead, most of these works are discussed critically and, in many cases, assessed empirically by the candidate.

Additional commentary and evaluation (optional)

Comment on the overall quality of the thesis, its novelty and its impact on the field, its strengths and weaknesses, the utility of the solution that is presented, the theoretical/formal level, the student's skillfulness, etc.

The thesis investigates the problem of matching image pairs and sequences. This is a fundamental task in computer vision, with notable applications in image retrieval and 3D reconstruction.

Chapter 1 provides an effective introduction to the thesis, including motivating convincingly the work carried out.

Chapter 2 introduces the concept of multiple-wide baseline matching, a difficult setup in which images to be matched differs substantially in two or more factor of variations; it also introduces a new benchmark for this problem and uses it to illustrate how challenging this task is for standard approaches.

Chapter 3 introduces HardNet, a novel trainable local patch descriptor. The key technical innovation is in a new loss formulation (the `hard' loss) which is shown to be more effective than other forms of training that use batch-wise sampling of contrastive patch pairs or triplets. The chapter includes an extensive evaluation of the method, as well as a thorough discussion of alternative design choices, which are also evaluated.

Chapter 4 investigates the problem of learning an affine feature detector. After demonstrating empirically the importance of affine normalization on the potential matching performance of descriptors, it proposes a simple network architecture that, via a form of self-training, can learn to normalize patches in such a way that they perform optimally with respect to a particular descriptor, thus encouraging invariance as well as detection of informative regions. An important innovation is the use of the HardNegC loss, which performs not unlike others in terms of final accuracy but is significantly more stable in training.

Chapter 5 introduces MODS, a matching method that is based on the generation of distorted views of an image. Differently from prior works such as A-SIFT, MODS can automatically tune the number of views generated, thus remaining fast for easy matching cases, and only using more intensive computations for harder problems. It also introduces a modified second nearest neighbor test (FGINN) which works better for the case in which the data has been augmented with many similar copies of the same image, due to view synthesis. MODS also uses a combination of multiple features (detectors and descriptors).

Chapter 6 introduces a new extensive framework for benchmarking local features. The framework is open source and has been used to organize a challenge. The chapter extensively studies the combination of detectors, descriptors and matching methods. It investigates not only several such combinations, but for each it performs a thorough validation of all the relevant hyper-parameters. The result is a comprehensive evaluation and critical comparison of dozens of methods that appeared in the literature. As a result of this analysis, many new conclusions are drawn and new best practices are identified, also pointing out to inconsistencies and weaknesses in prior evaluations.

Chapter 7 summaries the findings and proposes a few future venues for investigation.

III. OVERALL EVALUATION, QUESTIONS FOR THE PRESENTATION AND DEFENSE OF THE THESIS, SUGGESTED GRADE

Summarize your opinion on the thesis and explain your final grading. Pose questions that should be answered during the presentation and defense of the student's work.

CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

THESIS REVIEWER'S REPORT

The thesis represents a very substantial contribution, and one of the most significant ones in the past several years in the specific areas of image matching.

There are a small number of technical clarifications that I would ask during the defense of the thesis, such as how the new variant of the second nearest neighbor test in Chapter 5 works. A mor significant question is to summarize in a few words what are the practical recommendation and conclusions that emerge from the thesis, especially from Chapter 6: in the end, what features/matchers should we use, why, and when?

The grade that I award for the thesis is **A - excellent.**

Date: **10.5.2021** Signature: Andrea Vedaldi