THESIS REVIEWER'S REPORT

I. IDENTIFICATION DATA

Thesis title:	**USAC 2021 with Spatial Coherence Modeling Robust to Degenerate Data**
Author's name:	**Maksym Ivashechkin**
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Cybernetics
Thesis reviewer:	Radim Šára
Reviewer's department:	Cybernetics

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **challenging**

How demanding was the assigned project?

The problem of wide-baseline inter-image matching has been researched for decades and the literature is quite extensive and broad in topics, statistical models, optimization, and algorithmic approaches. Moreover, it is not easy to exceed or even reach state of the art results in this area. On the other hand, the student benefited from working in a group that has been doing its own cutting-edge research in the area for a long time; hence he received a perfect guidance.

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 student accumulated the required knowledge very well, including many side results. He understood all core aspects of RANSAC-type robust parametric regression in the domain of image matching (specifically, regressing homography matrix in P^2, fundamental matrix, essential matrix and perspective projection matrix). He has integrated state of the art methods into a common framework, implemented improvements in several of the core components, parallelized the algorithm, performed an extensive experimental evaluation on several large publicly available datasets, and published the results at a premier peer-reviewed computer vision conference. In my opinion the assignment was more than fulfilled.

Methodology **outstanding**

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

Methodology applied meets the standards of scientific work. In my view the most important concept is that of independence in the set of inliers and non-parametric model cross-validation. These constitute contributions beyond RANSAC, specifically in designing efficient MCMC proposals and in genetic programming methods.

Technical level **B - very good.**

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?

All methods are presented in a sufficient detail. Technical level is mostly sound, see Detailed Comments below. Here, I only have a few minor comments:

- Methods that are used and extended should be described in a self-contained way. This concerns especially MAGSAC, DEGENSAC, the SO(3) method in Sec. 4.2.1. and some additional methods, as commented below.
- An illustration would help in places, for instance when describing the reason why symmetric epipolar distance provides independent information over Samson error in rejecting potential mismatches (Sec 7.1.1).
- Page 15, 2nd paragraph: *However, there is no guarantee that these 5 inliers are not dependent* should read *[...] these five inliers are not independent*.
- Page 15, last but one paragraph: I believe rows 5-12 are stacked to B, not 4-12 as written there.

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?

English usage is generally good, with only minor mistakes. The text is mostly easy to follow, except where the methods are not explained in self-contained way and in words rather in mathematical language of symbols and formulas. See Detailed Comments below.

Some basic rules of technical style are broken, specifically,

- Numbered references to sections, figures, appendices, etc. should be written capitalized and without an article, e.g. Figure 1.1, Section 3.4, Table 1.3, Appendix A, etc.
- Displayed formulas are part of a sentence, including punctuation (commas, periods etc.), see e.g. (3.1) on p. 15 or the unnumbered formula on p. 21 or (4.4).
- If a sentence contains non-reference numbers up to ten, all are written in words rather than numerals.
- NaN is not a standard mathematical way of indicating the fact that the value of the ratio 0/0 is undefined, see (7.3).

Selection of sources, citation correctness **C - good.**

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?

Referencing is generally good, bibliographic sources are covered well. But bibliographic references are often incomplete or formatted in a non-standard way. For instance, publication type and publisher are missing in [11,24,34,38,42] and elsewhere, abbreviations such as RANSAC are lowercase, e.g. in [11] and elsewhere, volume and page information is missing e.g. in [24,26] and elsewhere, conference names are sometimes abbreviated and sometimes not, volume/issue/page information is often in non-standard format.

Detailed Comments

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.

This is a mixed list of comments, criticisms and suggestions. It follows the page order. Items emphasized in bold face should be discussed at the defence.

1. I would suggest distinguishing points (i.e. image keypoints, positions in R^2) and geometric primitives that are required by the minimal solver (matches in this case). RANSAC primarily works with primitives; properties of the points have only a supporting role.
2. The concept of independent inliers requires existence of a metric because it is based on proximity. Not all problems solvable by RANSAC allow existence of such metric. This limits the applicability of the concept.
3. Chirality mentioned in Point 4 at Page 6 is somewhat difficult to consider a dependence. The word dependence evokes a statistical dependence. Chirality (and possibly other problem-dependent constraints) should be distinguished from proximity-based dependence as a separate mechanism for filtering inliers, even if there was no algorithmic difference in the end.
4. **I do not know what cumulative binomial distribution is (p.7). Was it meant binomial cumulative distribution function? I am missing a formula; it would disambiguate the meaning.**
5. The concept of good structure (p. 8) is left undefined, just mentioned in the passing. All concepts should be defined and explained in a thesis, the space is not limited.
6. **The description of deliberate generation of bad samples is unclear (p. 8). It seems to be an important concept, can you explain how is it done and when exactly is it used?**
7. The description of the DEGENSAC estimation (p. 11) is somewhat unpedagogical. It is not sufficiently clear that the three correspondences required to estimate epipolar homography are used *together* with the fundamental matrix estimate. Otherwise, the presentation is confusing, as four correspondences are ordinarily needed for a unique estimate of homography. Second, one should prove that the homography estimated this way is supported by a scene plane which induces some additional correspondences besides the three.
8. I am not convinced by the proposed method of fundamental matrix estimation when images are related by homography (due to a planar scene or pure rotation of the camera), as proposed in Section 3.3. The author proposes a method that uses the homography H and camera calibration K. The homography is calibrated by K, and then decomposed to rotation R and translation t. A rank-2 matrix F is then constructed from K, R, t using a standard formula for F. The F is then output from the algorithm as if it was the fundamental matrix of the image pair. But this is misleading because there are infinitely many other matrices F' that satisfy all the homographic correspondences.

This way only a representative of this set is computed. What is the purpose of such procedure? I believe outputting the decomposition K, R, t makes more sense than outputting F.

9. **The estimation of focal length in calibration matrix described on p. 15 is left unexplained. Apparently some iterative procedure is used. How does it work?**
10. P. 15, last but one paragraph claims: *Nevertheless, DEGENSAC+ does not completely rely on approximated instrinsic matrices […]*. But this is in contradiction with Sec. 3.3.1 which does require calibration.
11. The derivation in Sec. 3.3.1 is slightly incorrect. The result is correct but X in formula (3.2) is not a 3D object point but the map by [I,-C] of its homogeneous representation.
12. The pure rotation detection of Sec. 3.3.1 could work without the knowledge of calibration matrix K. I believe the story could go this way: Let H be an inter-image homography induced by camera rotation. Then $\boldsymbol{H} \cong \boldsymbol{K_2 R_2 R_1^{-1} K_1^{-1}}$. When $\boldsymbol{K_1} \cong \boldsymbol{K_2}$ then $\boldsymbol{H}$ is similar to rotation, hence it must have the same singular values and their multiplicities as a rotation matrix.
13. The last sentence on p. 18 says that fundamental matrix has eight degrees of freedom. That is not correct. I believe the discussion is not about the rank of fundamental matrix but about the matrix above the last sentence on that page.
14. **I am not convinced by the proposed procedure for fundamental matrix estimation by Gaussian elimination, as described in Sec. 4.1.1. I believe it would fail in the case of no rotation between the cameras. In that case $\boldsymbol{f_9 = 0}$ and the assumption $\boldsymbol{f_9 = 1}$ on p. 19 does not hold. One could possibly remedy this by introducing artificial image homographies, re-mapping the corresponding points, estimating F as described, and then correcting the resulting fundamental matrix back. Alternatively, solve the problem twice, once with $\boldsymbol{f_9 = 1}$ and once with $\boldsymbol{f_9 = 0}$.**
15. The PnP solver in Sec. 4.1.2 assumes finite points (with unit homogeneous coordinate). Would the approach hold with (some) infinite points, too?
16. I did not fully verify this but I believe the rotation estimation in SO(3) described as the second method in 4.2.1 could be solved elegantly, efficiently and accurately by the quaternion method from R. Horaud and F. Dornaika. Hand-eye calibration. The International Journal of Robotics Research, 14(3):195-210, 1995.
17. I object to the proposition that the matrix from the 8-point algorithm should be left uncorrected (Sec 4.2.2 and elsewhere). Of course the residual is better because the uncorrected matrix is an affine model that has more degrees of freedom (a greater flexibility), which means it can explain data better. But it is not a fundamental matrix; hence it does not correspond to any relevant geometry. Using the uncorrected matrix for finding RANSAC inliers is perhaps thinkable but a very thorough analysis and verification should be done, as this is rather controversial, especially if the sample is not an all-inlier one. It could lead to overestimating the set of inliers.
18. **I am not clear about the LO+ procedure (p. 24). Why is it iterative, with repeated sampling? Is the goal achieving robustness at a lower computational cost than, for instance, a one-pass robust M-estimator? If so, how does it fare in terms of robustness, compared to M-estimator or other robust regression methods?**
19. What is meant by *inner threshold* on p. 26? Is it $\boldsymbol{K_{max}}$?
20. Sec 5.4 is hard to follow. I guess the purpose is to design an adaptive proposal (sampling) procedure in which promising inliers are sampled more often. A more formal description would be necessary. The presentation is ambiguous, for instance what exactly is meant by *The indices of inliers and outliers are randomly shuffled*? Are they shuffled within each subset or not? What is meant by *quasi-random sampling in chunks*? As a comment, from MCMC it is known that adaptive proposals are rather hard to design since they can hinder or even stop exploration. That might be the reason for mixed results reported in Sec. 9.7
21. **What kind of point normalization is performed in the least-squares procedure of Sec. 6.3?**
22. As far as I understood, the iterative LSQ procedure reclassifies inliers in each iteration (least trimmed squares). Such procedure is weakly robust. Why a more robust M-estimator is not used?
23. What is meant by image dimension in Sec 7.1?
24. The chirality test (7.1) removes *some* incorrect inliers.
25. Triangulated point clipping as described in Sec. 7.1.2 is a standard procedure but it should be used with care. The default setting should definitely be `off`.
26. The proof that there is half-homography (Sec. 7.2.1) should not be difficult. Homography matrix normalized to unit determinant is an element of the SL(4) special linear group. This Lie group is connected and I believe the existence of the square root follows from the connectedness.

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.

Contributions described in this thesis have been accepted by the Computer Vision community in a peer-review process with ICCV, which is a premier CV conference.

In summary, the thesis is high above the average quality. The detailed review is meant as a feedback, even if some of the comments are in the form of a question. Assuming the student can answer the core questions from the Detailed Comments section and summarize his own contributions to this work, I propose to grade the thesis **A - excellent.**

Date: **3.9.2021** Signature: