REVIEWER'S OPINION OF FINAL THESIS

I. IDENTIFICATION DATA

Thesis name:	**Evaluation Framework for Infant 3D Pose Extraction from RGB Images Using RGB-D Cameras and Motion Capture System**
Author's name:	**Vaculínová Noemi**
Type of thesis :	bachelor
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Cybernetics
Thesis reviewer:	Dr.-Ing. Nikolas Hesse
Reviewer's department:	Swiss Children's Rehab, University Children's Hospital Zurich

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **challenging**

Evaluation of thesis difficulty of assignment.

Capturing movements of infants with a marker-based system poses many challenges. The recording setup and protocol have to be chosen carefully to provide high data quality, and ensure the infant's safety at all times.

Satisfaction of assignment **fulfilled**

Assess that handed thesis meets assignment. Present points of assignment that fell short or were extended. Try to assess importance, impact or cause of each shortcoming.

The topic is important, since a high quality "ground truth" data set of infant motions does not exist. The thesis thoroughly investigated different data acquisition setups. The tasks proposed in the thesis guidelines were fulfilled.

Method of conception **correct**

Assess that student has chosen correct approach or solution methods.

The chosen approach was correct and design choices were motivated. Multiple preparatory experiments were conducted to determine the best solution for the experiment with real infants.

Technical level **A - excellent.**

Assess level of thesis specialty, use of knowledge gained by study and by expert literature, use of sources and data gained by experience.

Technical backgrounds, methods and experimental settings were described in detail.

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

Assess correctness of usage of formal notation. Assess typographical and language arrangement of thesis.

The language level is good. Some of the sub-sections are very short – merging some of them would help the readability. The thesis contains some typos that could have been detected using spell checking.

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

Present your opinion to student's activity when obtaining and using study materials for thesis creation. Characterize selection of sources. Assess that student used all relevant sources. Verify that all used elements are correctly distinguished from own results and thoughts. Assess that citation ethics has not been breached and that all bibliographic citations are complete and in accordance with citation convention and standards.

Most of the relevant literature is cited, some relevant work is missing, e.g., Meinecke et al., "Movement analysis in the early detection of newborns at risk for developing spasticity due to infantile cerebral palsy", Human movement science, 2006. In the Introduction, the motivation should cite recent papers to show that the topic is important right now. The two chosen papers [1, 2] are from 1995 and 1998. The citation for early detection of CP using motion analysis should be one of the main papers/books on the topic, e.g., Prechtl, "Qualitative changes of spontaneous movements in fetus and preterm infant are a marker of neurological dysfunction", Early Hum Dev, 1990, not a website of a clinical trial [3]. Citation [15] is incorrect and should be Robinette et al., "Civilian American and European Surface Anthropometry Resource (CAESAR) final report", 2002. References 19/21 and 23/24 are duplicates.

Additional commentary and evaluation

Present your opinion to achieved primary goals of thesis, e.g. level of theoretical results, level and functionality of technical or software conception, publication performance, experimental dexterity etc.

The primary goals of the thesis were achieved: an experimental protocol was developed that suits the application. The work builds a foundation on which further studies can be built. The technical approach is sound and the detailed description allows reproducibility. The interpretation of the results should have been more extensive – to me, the question remains if the author believes that the goal of creating a high-quality infant motion data set is possible, given the encountered problems. The concrete steps that would be necessary should be outlined, but the conclusion/future work section stays rather general. To summarize, the topic is important, the experiments lead in the right direction and were well designed. The thesis can be used as a foundation for further exploration of the topic.

Additional comments (ways to potentially improve the manuscript):

- A definition of terms “joints”, “pose”, “motion”, “body model” would have helped, e.g., at the start of the related work section where the terms are frequently used, rather than at the end.
- In Section 2.1, paragraphs for “direct pose estimation” and “lifting” are very short and the explanations are not easy to understand. The “model-based pose estimation” paragraph is mixing up some things, e.g., models (SMPL, SMPL-X) vs. method (SMPLify-X). It also includes and explanation of the training of the *body model*, which is not relevant for the use of body models for *pose estimation*. The process of fitting a body model to landmarks (e.g., SMPLify-X) doesn’t seem to be clear.
- The “SMIL pipeline” should have been explained more detailed. It is an important component of the evaluation, which is why the reader should be able to understand how it works.
- Sometimes the term “our” method is used. It wasn’t clear to me if this refers to the marker-based method or to the SMIL pipeline. It would be good to define this with the first use of the term.
- Some of the parameters are changed between experiments (flash time, fps). Why?
- The RGB-D vs. mocap interference experiment was performed with one camera, but later, a different one was used. The experiment should have been repeated with this camera as well.
- The first preparatory experiments led to the conclusion that the marker model is not suitable because some markers are attached at the back. Getting this information didn’t require an experiment.
- I don’t find the error unit very intuitive. I understand that from the RGB only-method, you may not get the correct size of the infant, and that a direct comparison in 3D would lead to large errors. However, transforming both (mocap and SMIL pipeline) results to a space that is very hard to interpret seems confusing to me. Why not transform the SMIL pipeline results to the 3D space of the mocap system (by scaling SMIL spine to match the length of the mocap spine)?
- If I understand the error unit correctly, a MPJPE of 1 unit would mean that the MPJPE is the length of the spine, so for an infant of 60 cm, this could be roughly 30 cm (as a rough guess for spine length). Results in Fig. 4.17 then would indicate that the **mean** error is ~24 cm? This would seem like there are big problems either in the SMIL pipeline or the mocap or in the transformation. It would have been helpful to present errors per joint, since these usually differ substantially for limbs and trunk joints.
- As mentioned above, the analysis/interpretation of the results could have been more extensive. It seems difficult to understand where the errors are coming from when no ground truth is available. Gaps in marker trajectories are mentioned – it would have been interesting to have numbers on this. How often is each marker occluded/unusable? This could also provide information on improving marker placement in the future.
- Also regarding marker errors, I was thinking that the cameras seem very far away for capturing infants. In order to properly capture infants, I think moving the cameras much closer to the infant might be one of the more important things to improve.
- Regarding evaluation metrics, the agreement of motion signals over time, e.g., using Pearson’s correlation coefficient, would have been interesting. The joint position error does not always tell the whole story, especially if only the mean is presented. For motion analysis, it is more important that the joints move in a similar manner, than being close to each other (in average). E.g., an estimated joint that is jumping/jittering in all directions within 3 cm might be worse for analysis than a relatively constant offset of 5 cm.

- The most important question remains if the mocap quality is going to be good enough, and, if not, if there are ways to make it good enough. Some illustrations of joint position values over time could have given a better impression of mocap quality (as well as SMIL pipeline results).

III. OVERALL EVALUATION, QUESTIONS FOR DEFENSE, CLASSIFICATION SUGGESTION

Summarize thesis aspects that swayed your final evaluation. Please present apt questions which student should answer during defense.

The topic is important and relevant, the assignment tasks were fulfilled, the experiments were thorough and described in detail. The student seems to have a very good understanding of the assignment.

There still were some aspects that could be improved, like citations, or the presentation and interpretation of results.

Overall, the thesis is well executed.

Questions:

- Do you think it will be possible to generate high-quality ground truth data with the presented setup/methods?
- Can you explain the unit of MPJPE? E.g., what does a MPJPE of 1.0 represent?
- Please give a short explanation of the SMIL pipeline. From your experiments, would you say that the SMIL pipeline accurately captures infant motion? Do you have ideas for improvement?
- In the evaluation of MPJPE, did you account for differences in skeleton definitions (mocap vs. SMIL)?
- What kind of evaluation metric could you use to avoid the canonical representation?

I evaluate handed thesis with classification grade **B - very good.**

Date: **6.6.2023** Signature: