SCIENTIFIC REPORT

on project implementation status between July – December 2022

Project PN-III-P4-PCE-2021-1959

Acronym: FIGHS

Fine-Grained Three-dimensional Human Sensing

`Simion Stoilow' Institute of Mathematics of the Romanian Academy

Principal Investigator: Prof. Cristian Sminchisescu, Ph.D.

Scientific activities and results:

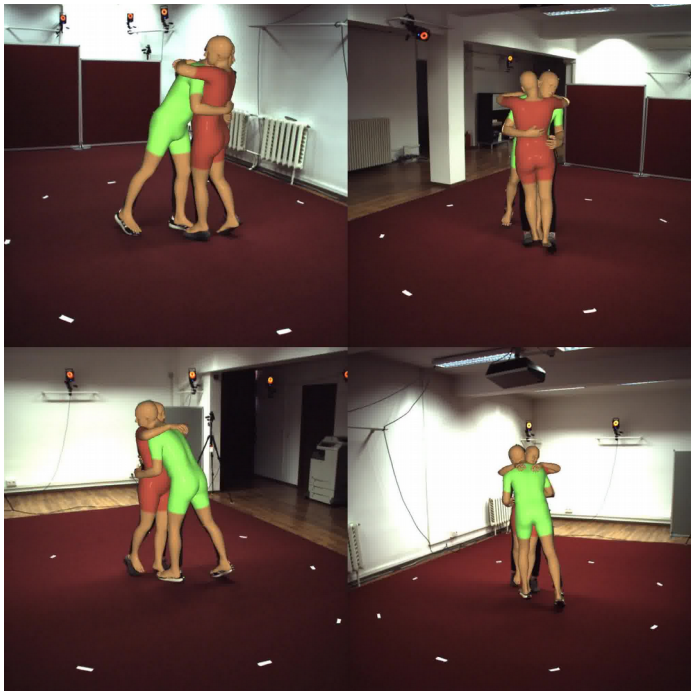
Three-dimensional understanding of human interactions is fundamental for detailed 3d scene analysis and modeling of human behavior. Most of the existing models focus on analyzing a single isolated person, and those that process multiple people focus on resolving the association between multiple people, rather than understanding the interactions between them. This leads to estimates that, even when impressive in terms of plausible posture and shape, miss the essence of the event at a close scrutiny when, for example, two reconstructions fail to capture contact during a handshake, a tap on the shoulder, or a hug. Such interactions are particularly difficult to resolve because their effects are compounded: on the one hand, uncertainty about body depth and shape could lead to compensation by pushing the limbs forward or further from their ground-truth position, when 3D inference is made from the monocular images; on the other hand, partial occlusion and relatively limited detail (resolution) for contact areas in images, typical of many human interactions, can make visual evidence inconclusive. Thus, these models lead to incorrect, unrealistic 3d estimates that lose sight of the subtle aspects of human contact and are of little use in understanding human behavior from images.

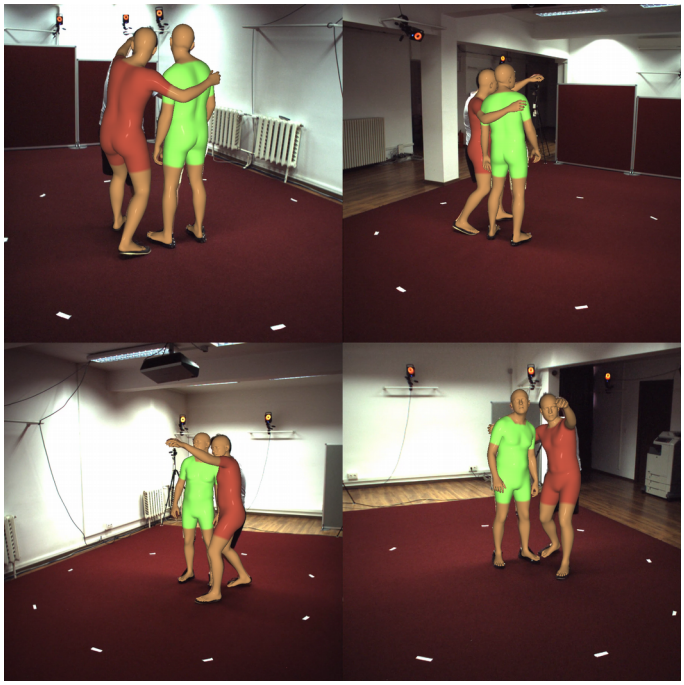
In an article submitted for publication at IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), the following contributions, which are aligned with the FIGHS objectives, are investigated and proposed: (1) models for interaction signature estimation that include contact detection, contact segmentation, and signature prediction 3d of contact; (2) demonstrating the utility of such models to produce augmented loss functions to ensure the consistency of the representation of human contact in the process of 3d reconstruction of persons; (3) proposing a methodology for recovering the ground-truth posture and shape of people in a controlled environment; (4) presentation of baseline models and algorithms to illustrate how human contact estimation supports the superior 3d reconstruction of persons when essential interactions between them are captured. Existing datasets that use multiple popular formats (GHUM and SMPLX, 3d Human3.6 joints) were used for training, while currently in the project the data collection activity in real environments has been postponed until next year, according to of the updated implementation plan în 2022.

The article proposes a first set of methodological elements to approach the reconstruction of interacting persons in a more principled manner, based on recognition, segmentation, mapping and 3d reconstruction. More specifically, the problem of providing truthful 3d reconstructions of interacting people is decomposed into (a) contact detection, (b) binary segmentation of contact regions on corresponding surfaces associated with interacting people; (c) contact signature prediction to produce estimates of the potential many-to-many correspondence map between regions in contact; and (d) 3d reconstruction using augmented loss functions constructed using additional surface contact constraints, when a contact signature exists. With the help of an extensive set of experiments, all system components have been evaluated and quantitative and qualitative comparisons have been provided showing how the proposed approach can realistically capture 3D human interactions.

The datasets used were also extended by annotating the entire temporal period of physical contact in each video sequence. Also, the proposed methodology to obtain the ground-truth posture and shape of interacting people goes beyond the simple extrapolation of 3d marker positions to the parameters of a body shape model, because: (a) only one subject in each video is motion tracked with the help of the

VICON motion capture system and (b) the recorded subjects are not only close to each other, but are, most of the time, even in physical contact. By leveraging information from motion detection sensors, from several RGB cameras and respectively from a 3d scanner, but also from contact annotations, a priori information about body posture and physical constraints, 3d reconstructions with similar correctness to those resulting from ground-truth representations have been achieved (see fig. below).





Moreover, in addition to common 3d sequences, the article makes public the ground-truth motion sequences in both GHUM and SMPLX formats and makes available to the research community an evaluation server with a hidden test set, along with a public benchmark, with the aim of advancing the state of knowledge in the 3d reconstruction of interacting persons.

Objectives

The objectives of the stage were fully achieved as follows. The setup of the data collection was carried out in the context of the augmentation of some existing data sets, as previously mentioned. This solution was chosen in the context of the change in the project calendar and related activities, whereby the acquisition of 3D motion capture equipment for natural scenes was moved to the 2nd stage (year 2023). We specify, however, that the methodology used for the organization and setup of the data is fundamentally the same, the differences occurring exclusively with regard to the technical method of data acquisition, which is obviously specific to the

equipment for capturing human movement in 3d nature scenes that will be purchased in the next stage.

In this context, as stated above, the existing databases were used to evaluate the state-of-the-art methodology of 3d estimation of human interaction and, as previously mentioned, this methodology was extended by proposing a first set of methodological elements that assume a principled approach to the 3d reconstruction of interacting people, based on recognition, segmentation, mapping and 3d reconstruction. All these developments are aligned with the objectives of the project and will constitute the starting basis of the methodology for estimating human interactions in 3d scenes in the wild to be developed in the next stages of the project, once the acquisition of the specialized equipment for capturing human movement is completed . From this point of view, we consider that the objective of this stage has also been fully achieved.

Progress summary

1. the available data collection setup
2. evaluation and further development of the 3d estimation methodology of human interactions (including the extension of some existing datasets)
3. submission of an article describing the achieved progress for publication at IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)

Conclusions:

The project activities were carried out according to the newly established calendar and were reoriented towards the use of existing databases for the continuation of research in the field of 3D scene analysis with multiple people who interact with each other. As a concrete result, an article was submitted for publication to IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), the journal with one of the highest impact factors in the field of computer science (current IF: 24,314).