Synthetic Ground Truth Generation for Testing, Technology Evaluation and Verification (SyntTEV)

Robert Manthey[1], Rico Thomanek[3],
Christian Roschke[3], Marc Ritter[2], and Danny Kowerko[1]

[1] Junior Professorship Media Computing
Faculty of Computer Science
Technical University of Chemnitz
D-09107 Chemnitz, Germany
{firstname.lastname}@informatik.tu-chemnitz.de

[2] Faculty Applied Computer Sciences & Biosciences
University of Applied Sciences
Technikumplatz 17, D-09648 Mittweida
ritter@hs-mittweida.de

[3] Faculty Media Sciences
University of Applied Sciences
Technikumplatz 17, D-09648 Mittweida
{roschke,rthomane}@hs-mittweida.de

Nowadays, several computer devices are used to visually detect objects, people and activities. Their quality and performance depends on limited datasets created and annotated by error-prone and expensive human handwork. But to reach high quality for complex detection tasks extensive datasets with errorless annotations are needed. To overcome this dilemma we create a system for automatic generation of synthetic ground truth data to allow learning of complex detection tasks as well as testing, verification and evaluation.

Dataset generation, system evaluation, human detection, human activity recognition, usability testing.

1. INTRODUCTION

Today, many systems like mobile phones[1] and autonomous driving vehicles[2] use computers to detect humans. Also industrial quality inspection devices as well as assistance systems for hindered [1] or old people use computers to detect humans and try to interpret their behaviour [2] to improve the quality of Human-Computer Interaction. This ability, their correctness and the quality heavily based on datasets showing the expected behaviour in several different facets captured in sufficient scale [3]. But common datasets from real world contain only limited amounts of facets, like a small number of perspectives, a small amount of textures at the surface of objects, few repetitions of similar person activities or same resolutions. Creating a new dataset from real world or extending one is expensive, time consuming and needs error-prone human work for annotation.

Based on the experience of previous works with synthetic data to analyse different software and hardware systems [4, 5], a system was created for automatic synthesising of humanoids, objects, scene environments and activities to produce scenarios of arbitrary structure forming new datasets with exact and well-defined ground truth. Due to the programmable nature of the system shown in Fig. 1, the variability and complexity of the facets are almost unlimited and only restricted by the amount of storage and time of the synthesis. Additionally, the content of the datasets may show hard to observe human behaviour, dangerous activities as well as accidents. Datasets containing these providing the capability to train assistance systems being able to detecting these occurrences.

On the other side it is possible to realize extensive quality checks, usability tests and performance evaluations on existing systems.

[1] https://www.androidauthority.com/facial-recognition-technology-explained-800421/
[2] https://www.engadget.com/2016/11/19/tesla-self-driving-demo-shows-car-view/

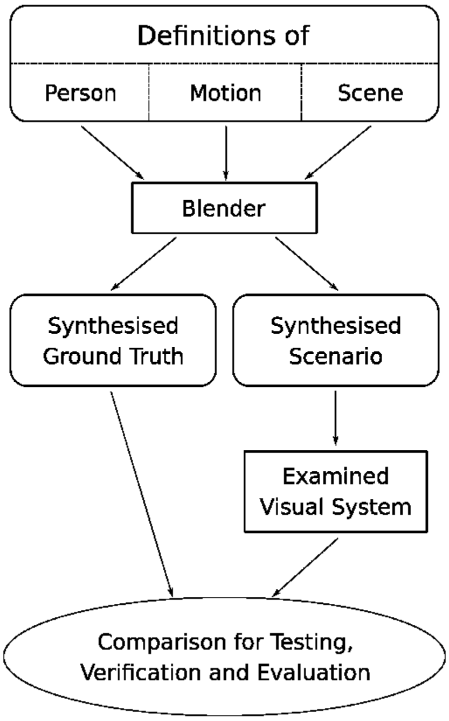

Figure 1: *The definitions of persons, there motion and the visible scene be processed by Blender to synthesise the scenario and the corresponding information about the properties of all objects inside this scenario representing the ground truth. The following systems process the scenario and allow the comparison there results with the ground truth to evaluate quality, correctness, performance etc.*

2. SYSTEM ARCHITECTURE AND TECHNICAL REALISATION

To generate the synthetic ground truth we use the 3D modelling tool MakeHuman[3] to create templates of photo realistic humanoids with different definitions of age, gender, height, width, muscularity, hairstyle and clothing. The anatomical structure of humans is approximated by 163 bone-like elements as shown in Fig. 2, permitting natural movements. Depending on the expected field of application, the test cases or by random selection the values of each definition is set, for instance dark clothing in case of testing the limitations of person detection systems in dark environments.

The definitions of motion be based on motion captures of humans and be modified and combined in well-defined ways to form the requested activity.

With the professional, open source 3D computer graphics software Blender[4] we define the scenes and set up the environment properties like light sources as well as cameras with their resolution and field of view, as shown in Fig.3.

The humanoids, their activities and the scene are combined to create the test scenario being rendered by Blender, as in Fig. 4. Modification of the inputs or their combination made it possible to generate nearly infinite numbers of different test cases being synthesized and overcome the limitations of real world datasets. At the same time a self-developed python-based program extract the exact and correct data from the internal storage of Blender to get the ground truth of the scenario. Therefore no further annotations are needed.

The synthesised scenario is applicable to examine visual systems like the OpenPose[5] pose detection system as shown in Fig. 5. Comparing the ground truth with these results made it possible to test the applicability of the system, to verify the correctness and to evaluate the performance.

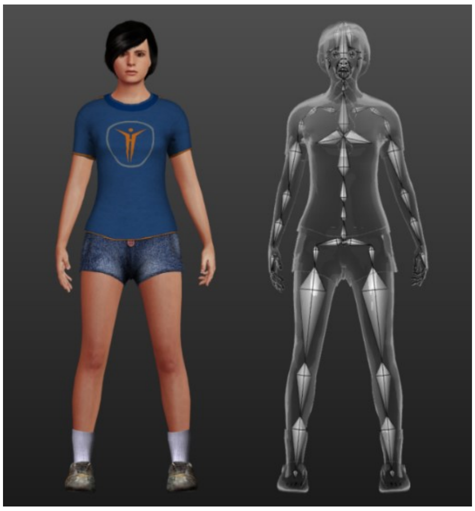

Figure 2: *Sample of a MakeHuman-model with surface (left) and corresponding bone structure (right).*

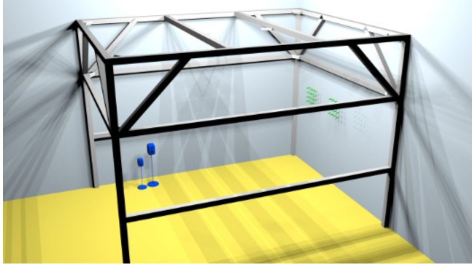

Figure 3: *Sample scene modelling parts of our laboratory with eight optical stereo sensors in the upper corners and two on the top beams, three arrays of microphones at the right wall and two loudspeakers.*

[3] https://www.makehuman.org
[4] https://www.blender.org

[5] https://github.com/CMU-Perceptual-Computing-Lab/openpose

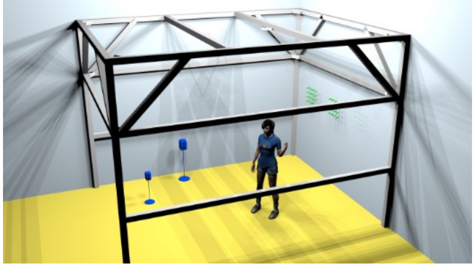

Figure 4: *Synthesised scenario based on the data shown in Figure 2 and 3 combined with a predefined left-arm beckon pose.*

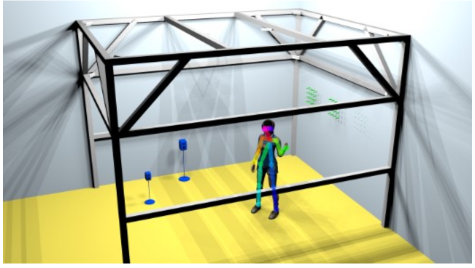

Figure 5: *Exemplary visual result of the examined pose analysing system OpenPose, showing the detected extremities of the person.*

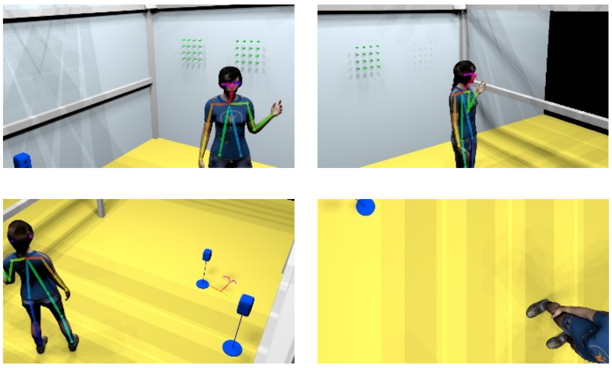

Figure 6: *Selected images from four sensors processed by OpenPose. The upper two illustrate good detection results of frontal and side-view. The lower left illustrate a similar result for the person, but with a false detection at the shadow between the two loudspeakers. The lower right show no results, indicating the limitation of the pose detection when looking downward.*

3. SYSTEM APPLICATION AND EXPLORATORY ALGORITHM PROCESSING

In order to explorate the usefulness of the presented system we synthesise a scenario based on a model of our laboratory and a humanoid with beckon pose. The images of all ten sensors are captured and processed by OpenPose. The results are embedded into the images as overlay containing colored lines at the position of detection, as shown in Fig. 6.

As clearly recognisable the humanoid is well detected for views from all sides, but not from above. The detection of the nodes show good correlation to the ground truth from the scenario.

4. SUMMARY & FUTURE WORK

We create a system for the generation of datasets based on well-defined synthetic scenarios. We show the structure of our system with objects, humanoids and activities as well as a short empirical exploration to show the usefulness and the potential of the system and the synthesised datasets. Future developments will increase the quantity of fundamental building block to simplify the definitions and include the synthesis of sound propagation.

Acknowledgments. This work was partially accomplished within the project localizeIT (funding code 03IPT608X) funded by the Federal Ministry of Education and Research (BMBF, Germany) in the program of Entrepreneurial Regions InnoProfile-Transfer.

5. REFERENCES

[1] Kozar, T., Rudolf, A., Cupar, A., Jevsnik, S., Stjepanovic, Z. (2014) Designing an Adaptive 3D Body Model Suitable for People with Limited Body Abilities. Journal of Textile Science & Engineering, OMICS International, 2014, p. 1–14

[2] Hägele, M., Schaaf, W., Helms, E. (2002) Robot assistants at manual workplaces: Effective co-operation and safety aspects. Proceedings of the 33rd International Symposium on Robotics (ISR 2002), Stockholm, 2002, pp. 1–6

[3] Bilenko, M., Kamath, B., Mooney, R.J. (2006) Adaptive Blocking: Learning to Scale Up Record Linkage. Proceedings of the Sixth IEEE International Conference on Data Mining (ICDM-06), Hong Kong, pp. 87–96

[4] Manthey, R., Conrad, S., Ritter, M. (2016) A Framework for Generation of Testsets for Recent Multimedia Workflows, Universal Access in Human-Computer Interaction, Human Computer Interaction International. Toronto, 2016, pp. 460–467, Springer

[5] Manthey, R., Ritter, M., Heinzig, M., Kowerko, D. (2017) An Exploratory Comparison of the Visual Quality of Virtual Reality Systems Based on Device-Independent Testsets, Proceedings of Human Computer Interaction International 2017, Vancouver, 2017, pp. 130–140, Springer