Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 100 (2021) 631–636

www.elsevier.com/locate/procedia

31st CIRP Design Conference 2021 (CIRP Design 2021)

How to use human pose estimation to measure the hand-arm motion in craft application with no influence on the natural user behavior

Sebastian Helmstetter [a], Johannes Sänger [a], René Germann [a], Sven Matthiesen [a]*

[a]*IPEK – Institute of Product Engineering, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany*

* Corresponding author. Tel.: +49 721 608 47156; fax: +49 721 608 46051. *E-mail address:* Sven.Matthiesen@kit.edu

Abstract

The interaction between human and machine plays an important role in the design and optimization of human-machine systems. This interaction is characterized by human motion using the technical system. Especially in the field of hand and power tool applications, the motion capture should be performed under the real working condition and without influencing the user. There are already motion tracking systems that allow capturing the motion during the interaction, but there is no mobile motion capture system that allows an individual analysis of the user for biomechanical analysis in the normal work process without influencing him. Therefore, requirements for a motion capture system are derived and a system is presented that meets these requirements. This system consists of two cameras and is based on the pose estimation algorithm OpenPose. The comparison of the presented system and the state-of-the-art system Xsens is performed and based on the measured elbow angle and the wrist position. The results show a very good correspondence between the curve characteristic of the elbow angle and the wrist position of both systems. However, inexplicable values shifting at two different levels still occur, which need to be investigated further. Overall, the presented system shows great potential in terms of mobility and flexibility of the presented system with some weaknesses in the data processing and efficiency. By addressing these weaknesses, the presented system can be used in the analysis and optimization of human-machine systems.

Keywords: human-machine system; user centered design; motion capturing; OpenPose; human factor; power tool

1. Introduction

Human-machine systems are characterized by the purposeful interaction of humans with technical systems for the best possible interactive fulfillment of a given or self-selected human activity [1]. To analyze and optimize human interaction with the technical system, studies often use human motion capture systems.

An example of human-machine systems with a strong physical interaction between the user and the technical systems is the usage of hand and power tools. An early step in the design process of these systems is the definition of the design aims for optimal interaction between the user and the technical system. Therefore, the product designer observes the users of these products in their working environment during regular applications. Measurement technology as motion capture systems helps to quantify their observation. [2–6]

In the state of art and the state of research there exist different technologies for motion capturing, e.g.: IMU-based motion trackers [7], infrared camera systems with body markers [8], marker-less pose-estimation, and silhouette tracking [9]. All of these systems have disadvantages in using them in a field study [10]. Regular camera systems using six or more cameras. At a construction site or in a workshop, there is often not enough space to install all of them. Furthermore, the effort for the installation of these systems is very high. Motion

10.1016/j.procir.2021.05.135

capture systems that are working with markers or IMU trackers also can influence the user in his natural interaction with the hand or power tool. There are a lot of problems to place the marker because of the working cloth and self-protection equipment of the users at the correct places on the body. This harms the analysis of the interaction in the system and the derived requirements for the target system. What is still needed, is a mobile and flexible motion capture system, for small spaces that allows an individual analysis of the user's motion without influencing his regular working process by the measurement equipment. A marker-less pose-estimation-based motion capture system promises the best chances for measurements without influence on the user. However, existing systems consist of many complex aligned cameras, which make mobile use in field studies almost impossible. A solution for a simple motion capture system offers the Kinect camera system from Microsoft. It is working with a stereo camera system and is design as a consumer product which detects gestures for controlling video games. The advantage of this system is that it is very easy to use and mobile. However, previous studies have shown that the Kinect system is not suitable for carrying out biomechanical analyses of human movement [8]. In particular, a non-constant measurement frequency and a too low recording rate of maximum 30 Hz is one of the problems why the existing system cannot be used. The gab at the state of research leads to the research question:

Is it possible to use pose estimation with a simple and mobile stereo camera system and pose estimation in a motion capture system for biomechanical analysis of users in craft applications without affecting the user in the real working environment?

In the following, a self-designed motion capture system based on the human pose estimation algorithm OpenPose [11] is presented, to solve this research question. The further study compares in the craft application of hammering nails the new motion capture system to the IMU-based commercial motion capture system Xsens [7], which is used at IPEK – Institute for Product Engineering to measure the user's motion in hand and power tool application so far.

2. Material and Methods

2.1. Requirement to the Motion Capture System

For the success of the Motion Capture System, which can improve the mobile and flexible measurement of human motion by using hand or power tools in a rough real working environment with low influence on the user, a bunch of requirements has to be fulfilled, that differentiates this system from other. The following requirements are selected to validate the aims for that system:

1. QUALITY
   The human motion capture system has to measure the motion of the human hand-arm system during hand or power tools application with a similar quality of measurement data as commercial human motion capture systems. To validate the quality of the new system, the motion of one body part – e.g. the wrist-relative to the shoulder – has to be the same values in both systems. That allows a validation even if the origin and orientation of one coordinate system are different from the other.
2. MOBILITY
   The human motion capture equipment needs less than 5 components of collecting data which can be moved and installed by one person.
3. USAGE UNDER REAL WORKING CONDITIONS
   The human motion capture system should be adaptable to different working places (big areas to small spaces), different working positions (from the bottom under the ceiling) and, rough working conditions.
4. LOW INFLUENCE
   The human motion capture system has no physical contact with the user and no component on the tool or in the working area that forces the user to change his regular practice.
5. SYNCHRONIZATION WITH OTHER MEASUREMENT EQUIPMENT
   The human motion capture system has to measure by a constant and adaptable measuring frequency. A possibility is needed to synchronize the system to other measurement equipment.

2.2. Technical Approach for the Human Pose Estimation Based Motion Capture System

2.2.1. Mobile, Synchronized Stereo Camera System

To realize a three-dimensional measurement of human motion, at least two cameras are needed, which capture the action from different angles. The image capturing of these cameras have to be exactly synchronized, to get good results for the triangulation of the 3D skeleton model in the further process. Furthermore, the recording frequency of the cameras corresponds to the recording frequency of the whole motion capture system. For the demonstrator, we used two cameras "acA1300-200uc" from Basler [12] with a USB interface for image data transfer and an analog I/O interface for the hardware trigger signal. This camera type can be triggered with a recording frequency up to 160 Hz and with reduced image size up to 220 Hz. For the first tests, we use a lens with a focal length of 8 mm that allows capturing a working area with a width of 1913 mm and a height of 1533 mm in a distance of 2.5 m (see Fig. 1).

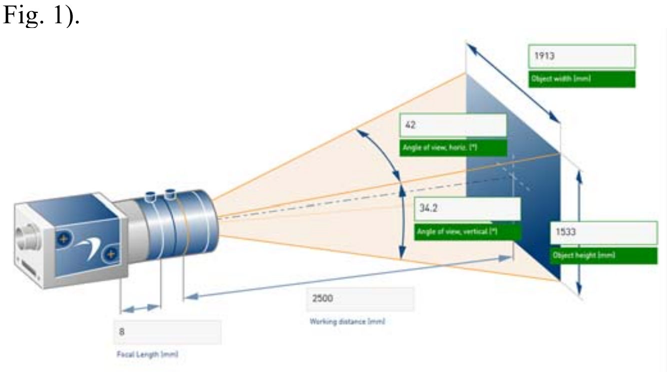

Fig. 1: Observation space of a Basler acA1300-200uc camera with a 8mm focal length (from Basler Lens Selector)

Furthermore, an external trigger signal and a frame grabber are needed. The frame grabber compromise and save the camera data of both camera for further evaluation. We use a Lenovo P50 with 25 GB RAM for grabbing the camera date with a C++ code based on Basler Pylon with a recording frequency up to 150 Hz. The analog signal for the hardware trigger of the cameras is an analog or with pulse wide modulation formed single with an amplitude of 10 V. We tested different signal generators. For the benchmark study, we used the mobile measuring system Gantner QBrixx [13] to generate the signal for the hardware trigger of the cameras. This solution allows synchronizing the camera trigger with other measurement equipment like force or acceleration sensors that are also often used for analyzing the interaction in human-machine systems.

2.2.2. Calibration of Camera System

For the triangulation of the two 2D camera images to a 3D model, calibration of the camera system is needed every time the relative position or rotation of the two cameras to each other is changed. For the calibration of the cameras, the stereo camera calibration toolbox from Matlab is used. Therefore, a bunch of images of a chessboard in different positions is synchrony captured (see Fig. 2). The evaluation of that image data with the stereo camera calibration toolbox results in a stereo parameter set, that can be used for the triangulation of the 2D human body models analyzed with OpenPose to a 3D human skeleton model.

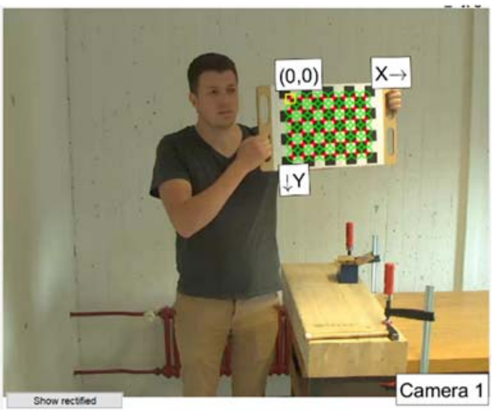

Fig. 2: Calibration of the stereo camera system by a black and white chessboard

2.2.3. Human Pose Estimation with OpenPose

The human pose estimation algorithm OpenPose detects humans in 2D RGB images and calculates the position of their joints in the picture. For the human motion system, the output of the joint coordinates is used in the Javascript Object Notation (JSON) data format. For every recorded frame, there exists one JSON file, with the model of the human skeleton at the time. All JSON files of one recording are imported into Matlab for both camera perspectives and combined to a time sequence of the human motion.

2.2.4. Triangulation of 3D Human Skeleton

To get a 3D model of the recorded human motion, the data of the human joint position that is calculated by the OpenPose algorithm is triangulate by the Matlab function *triangulate ()* using the stereo parameter set from the camera calibration in 2.2.2. By that triangulation, the 2D coordinates of the joint positions which are given by OpenPose in pixels are transformed to 3D position data in millimeters. The origin of the global coordinate system that is used for the 3D joint positions is located in the sensor of that camera which is given as “camera 1” in the stereo parameter set. The x-axis and y-axis are orientated of the orientation by the chessboard (see Fig. 2).

The time sequence of the 3D joint positions for the recorded human motion are saved in a Matlab struct file, that is standardized at IPEK for recorded measurement data. However, any other data structure can be used to save the 3D joint positions.

2.3. Study Design

After the implementation of a first pose estimation based human motion capture system, the system is applied in a bunch of field and laboratory studies with power tool application. It is compared to other measurement systems and the procedure is optimized. Afterward, the pose estimation based human motion capture system is benchmarked to the commercial IMU-based motion capture system Xsens. For the study, 6 subjects were hammering nails for 2 min in a beam of wood. More details on the study design can be found in Helmstetter et al [14].

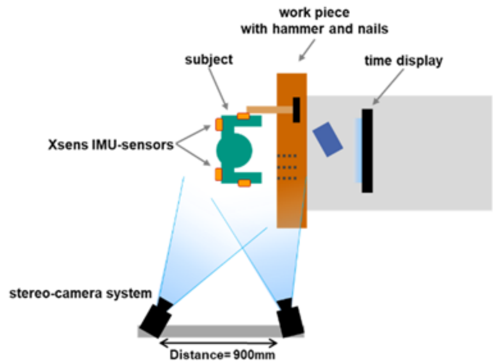

Fig. 3: Setup of the validation study for the pose estimation based motion capture system in comparison to the commercial system Xsens

The motion of their hand-arm system was measure by the pose estimation based motion capture system and the IMU-based Xsens in parallel. The setup is shown in Fig. 3. Both measurement systems were triggered by the same start signal and both data were recorded with a frequency of 60 Hz. In the study presented here, the subjects were equipped with both motion capturing systems. The equipment with the IMU sensors required measurement of the subject's body measurements. Besides, a calibration must be carried out with the sensors applied to the subject´s body. The applied IMU sensors are worn by the test person during the study and thus influence his natural movements. The subjects are not influenced by the OpenPose measuring system. After the cameras have been calibrated, motion capturing can begin. While the experiment is being carried out, the subject is not influenced by the OpenPose measurement system.

2.4. Evaluation Methods

To compare the two motion capturing systems, specific values are chosen for comparison. These values are selected so that they are independent of the respective coordinate system used in the different motion capturing systems which makes it easier to compare them.

Joint angles are suitable values for that comparison because they are not dependent on the coordinate systems. Furthermore, Xsens provide an export of the relevant joint angles for hand arm systems such as the joint angle of the elbow. In the following, only the joint angles that are assigned to flexion and extension of the right elbow are evaluated.

Based on the joint position data provided by OpenPose each joint location of the human-arm model can be described. The joint angle can be calculated from this position data. E.g. the joint angle of the elbow can be calculated by determining the angle between the forearm and upper arm. The forearm can be described by the connection of the elbow joint and the wrist joint. The upper-arm can be described by the connection between the elbow joint and the shoulder joint.

The second comparison variable is a relative reference position. It is used because the absolute position data of the motion capturing systems are not comparable to each other. Therefore, a reference position is chosen. The reference position serves as the starting point for further relative position calculations, which always relate to the reference position.

For the investigation in this study, the movement of the right wrist is considered. The reference position is selected at one second after the start of the measurement. The change of the current right wrist position to the reference position of the wrist at time t = 1s is referred to as the Delta Position. The delta position is calculated for both the Xsens and the OpenPose data.

3. Results

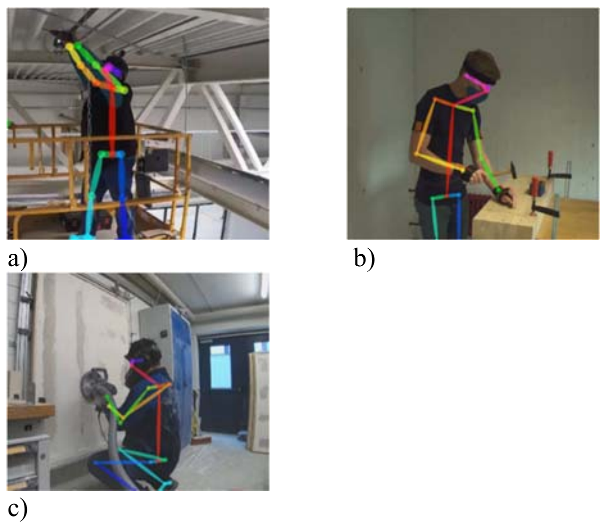

Fig. 4: Application cases of OpenPose in different environments (installation work (a), hammering (b), drywall grinding (c))

Fig. 4 shows different application cases of the motion capturing system OpenPose. The tests show the mobility and flexibility of the new motion capture system. The system was still applied at field study, as e.g. to measure the working pose of an electrician at a mobile lifting platform (Fig. 4a). Furthermore, it is used for laboratory studies with a drywall grinder (Fig. 4c) and the mentioned study of hammering (Fig. 4b).

Fig. 5 shows the current position of the right wrists of subject three concerning the recorded reference position of the wrist at the time of one second. Consequently, the delta position represents the length of the vector from the reference position to the position at time t. There is a very similar characteristic in the course of the two courses of OpenPose and Xsens.

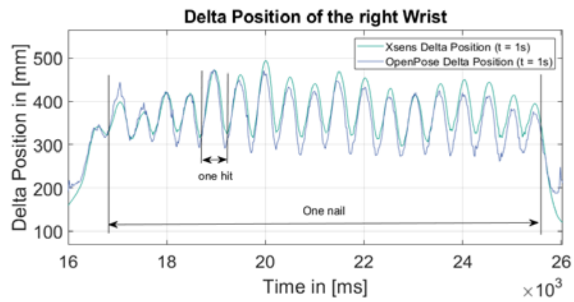

Fig. 5: Delta Position of the right wrist during the hammering process for one nail

The correspondence of the calculated position changes during the individual hits to the nails can be seen. Fig. 5 shows the hits the subject is needed to hammer one nail. Each oscillation around the value of 300-360 mm can be recognized as one individual hit. The difference between both measurement systems is about 10 mm to 20 mm.

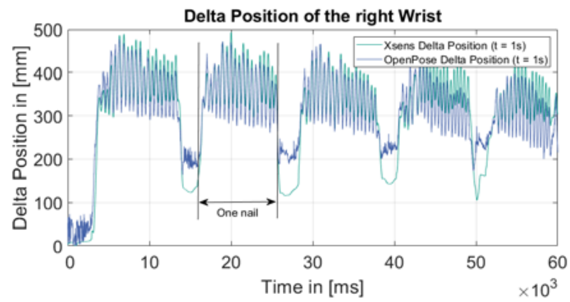

Fig. 6: Delta Position of the right wrist during the whole study period for hammering 5 nails

Several hits form a plateau in which a nail is completely driven in are shown in Fig. 6. The subject grabs a new nail between the plateaus of the oscillation four times. In these areas, the calculated position from the different measuring systems deviates more from each other than in the areas of the plateaus. Fig. 7 shows the joint angle of the right elbow of subject three. This angle describes the flexion/extension of the elbow. Areas in which the joint angle fluctuates around a value can be seen. This value is approx. 70 degrees for the Xsens measurement data. The joint angle calculated from OpenPose is slightly below, approx. 40 degrees. The oscillation range is comparable.

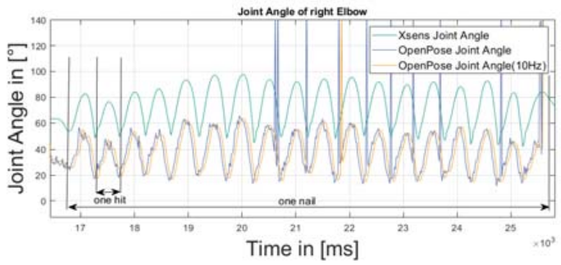

Fig. 7: Joint angle of the right elbow during the hammering process for one nail

Fig. 7 shows the joint angle of the right elbow of subject three. This angle describes the flexion/extension of the elbow. Areas in which the joint angle fluctuates around a value can be seen. This value is approx. 70 degrees for the Xsens measurement data. The joint angle calculated from OpenPose is slightly below, approx. 40 degrees. The oscillation range is comparable.

It is especially evident between the hammering of one nail that the calculated elbow angle from the OpenPose position data shifts on a higher level (see Fig. 8).

In the first tryout to solve the problem, the data were filtered by a 10 Hz low pass filter to reduce the influences from outside the human-controlled movement [15]. It reduces the singular outliers but the shifts in the areas between the hammering of one nail remain.

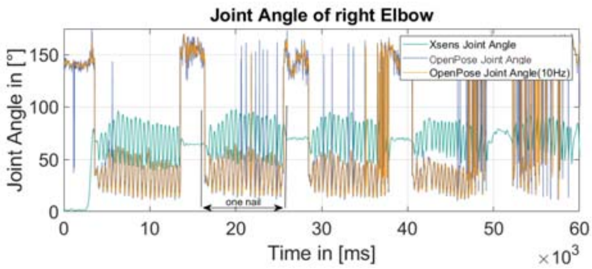

Fig. 8: Joint angle of the right elbow during the whole study period for hammering 5 nails

4. Discussion

The discussion focuses on the requirements that are made at the beginning of the design process for the image-based motion capturing system with no influence on the user working process in craft applications.

QUALITY

In general, the image-based motion capturing system can achieve good results for the joint position tracking as shown in Fig. 5. However, there are also differences in der determination of the position, such as between the plateaus of the nail impacts (see Fig. 6). At the current state, it is not identified where this constant deviation comes from. The might be a failure of the pose estimation algorithm, the triangulation of the 2D joint data, or a failure at the Xsens system. For a detailed investigation, a study would have to be conducted in which the exact movement of the human is known.

A good match of the systems can also be seen in the detection of the joint angle. The signals have the same characteristic but an offset of about 25° can be mentioned. By computing the elbow angle between the forearm and the upper arm, it is excluded that the offset comes from the different coordinate systems, but maybe from the pose estimation algorithm or the calibration of the Xsens system, too.

Furthermore, there are still effects in the evaluation process of the image data, which lead to a shift in the joint angle to two different levels. A possible cause could be the camera positioning and calibration. This should be examined more closely in further investigations. However, the accuracy of the presented motion capture system is equal to that of Xsens and is completely sufficient for user studies, if the problems of offsets and shifts are solved.

MOBILITY

The camera system is mobile and can be set up in different places, as shown in Figure 1. No special infrastructure is required for the installation of the cameras. Therefore, it is suitable for construction sites, as the environments and operations conditions are very different from each other.

The system can be operated by one person during the study and is easy to handle. To capture the data just two cameras, a signal generator for the trigger signal and a laptop for frame grabbing are necessary. Furthermore, the systems can be prepared and calibrated without the subjects. That makes studies with a lot of subjects more effective, than using a measurement system, where the user's body dimensions must be measured and markers or trackers must be attached to his body.

USAGE UNDER REAL WORKING CONDITIONS

As mentioned above the image-based motion capturing system can be used in different application cases. Fig. 4 shows the wide possible use of this imaged based motion capture systems. The subject can work in a rough working environment without an application having to be simulated in a laboratory. Furthermore, the measurement system is easily adaptable to individual requirements, e.g. increasing the measurement accuracy by adding further cameras or an adjustment of the measurement frequency. However, in applications with a large range of motion of the user, the field of view of the cameras limits the movement of the user.

LOW INFLUENCE

There is no influence on the subject based on the measurement equipment. Furthermore, the user is not impaired by the measurement technique when performing the working task compared to other systems

In comparison with the Xsens, it is not necessary to wear IMU sensors or to take measurements of the body parts of the subjects. Furthermore, the system is also working with no passive markers like Vicon. So, the user is neither handicapped by the measurement equipment nor can the system be damaged during the working process. [7, 8].

SYNCHRONIZATION WITH OTHER MEASUREMENT EQUIPMENT

The synchronization of other measurement systems is possible by using the same signal that is used for triggering the cameras or the measurement recording systems itself if it is used to generate the trigger signal for the cameras.

Overall, the presented pose estimation based motion capture system working with an open-source pose estimation algorithm and low-cost physical equipment fulfills the requirements that have been made. It also shows great potential regarding the possibilities to define individual raw data filters and data smoothing.

It offers the product developer of human-machine systems a great benefit for the biomechanical analysis of the complex interaction between the user and a technical system due to its mobile usage in the real working environment and the low influence of the measurement technique on the natural behavior of the users, as shown by the fulfillment of the requirements defined at the beginning. The product developer receives an important advantage for the improvement of the user-centered product development.

However, there are limitations on the measurement equipment, such as the limited camera recording capacity (maximal recording-frequency of 150 Hz with a maximal recording time of 1 minute), depending on the computer and algorithm that are used for the image grabbing. Also, the pose estimation computing is not optimized. So it needs about ten times the length of the recording time in our study. The results show that there are still challenges with the camera orientation and calibration as well as the triangulation of the images, which lead to the shifts of the joint angles of the wrist. Moreover, there were also challenges with low contrast values in the image data. This can be improved by illuminate the scenery well and use clothes with a good contrast to the background.

5. Conclusion

The state of research shows the need for a mobile and flexible human motion capture system for small spaces that can be used to record the motion of the hand-arm system of users in craft applications with no influence on the user's regular working process. Commercial motion capture systems are designed for laboratory application or they are using IMU-sensors that can influence during real application. Furthermore, the IMU-sensors are sensitive to metallic objects and can be influenced by metal or magnetic surroundings and affect the measurement accuracy. Therefore, a mobile human capture system is designed that is specialized in requirements for measuring users in craft applications. The human motion capture system is based on a stereo camera system and OpenPose, an open source human pose estimation algorithm. The evaluation of the first prototype of that system fulfills the requirements and a comparison to the commercial motion capture system Xsens shows a high potential for the concept of the new system. Although the camera system is not influenced by the surroundings there are limitation to the movement of the user in relation to the measurement systems due to the line-of-sight of the cameras. Even if some problems with the pose estimation and the triangulation have to be fixed, the new system offers a chance to optimize the efficiency and quality of human motion capturing for the user centered design.

References

[1] Johannsen, G., 1993. *Mensch-Maschine-Systeme*. Springer, Berlin.

[2] Matthiesen, S., Germann, R., Schmidt, S., Hoelz, K. *et al.*, 2016. Prozessmodell zur anwendungsorientierten Entwicklung von Power-Tools: Technische Unterstützungssysteme, die die Menschen wirklich wollen *2*.

[3] Biryukova, E.V., Bril, B., 2008. Organization of goal-directed action at a high level of motor skill: the case of stone knapping in India: The case of stone knapping in India. Motor Control *12*, p. 181.

[4] Germann, R., Jahnke, B., Matthiesen, S., 2019. Objective usability evaluation of drywall screwdriver under consideration of the user experience. Appl Ergon *75*, p. 170.

[5] Lee, C.-C., Nelson, J.E., Davis, K.G., Marras, W.S., 1997. An ergonomic comparison of industrial spray paint guns *19*, p. 425.

[6] Matthiesen, S., Mangold, S., Bruchmueller, T., Stelzer, D. *et al.*, 2015. The influence of the user on the power tool functionality - a force sensing handle for a hammer drill, in *Proceedings of the 13th International Conference on Hand-Arm Vibration*, Beijing University Health Science Center Capital Medical University, Beijing, China.

[7] Paulich, M., Schepers, M., Rudigkeit, N., Bellusci, G. *Xsens MTw Awinda: Miniature WirelessInertial-Magnetic Motion Tracker for HighlyAccurate 3D Kinematic Applications*.

[8] Pfister, A., West, A.M., Bronner, S., Noah, J.A., 2014. Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis. J Med Eng Technol *38*, p. 274.

[9] Corazza, S., Mündermann, L., Gambaretto, E., Ferrigno, G. *et al.*, 2010. Markerless Motion Capture through Visual Hull, Articulated ICP and Subject Specific Model Generation *87*, p. 156.

[10] van der Kruk, E., Reijne, M.M., 2018. Accuracy of human motion capture systems for sport applications; state-of-the-art review. Eur J Sport Sci *18*, p. 806.

[11] Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., Sheikh, Y., 2018. *OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields*.

[12] Basler AG. acA1300-200uc Basler ace USB 3.0 camera.

[13] Gantner Instruments GmbH. Gantner Q.brixx.

[14] Helmstetter, S., Germann, R., Abbes, M., Matthiesen, S., 2020. Human factors for the evaluation of the user expertise in the usage of power tools, in *Proceedings of the 31st Symposium Design for X (DFX2020)*, The Design Society, p. 21.

[15] Kern, T.A., 2009. *Entwicklung Haptischer Geräte*. Springer Berlin Heidelberg, Berlin, Heidelberg.