Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 41 (2016) 759 – 764

www.elsevier.com/locate/procedia

48th CIRP Conference on MANUFACTURING SYSTEMS - CIRP CMS 2015

On the use of Multi-Depth-Camera based Motion Tracking Systems in Production Planning Environments

Florian Geiselhart[a,]*, Michael Otto[b], Enrico Rukzio[a]

[a]*Institute of Media Informatics, Ulm University, James-Franck-Ring, 89081 Ulm, Germany*
[b]*Daimler AG, Wilhelm Runge Str 11, 89081 Ulm, Germany*

* Corresponding author. Tel.: +49 731 31306; fax: +49 731 31303. *E-mail address:* florian.geiselhart@uni-ulm.de

Abstract

In order to keep up with changing market demands like highly individualized products, automotive OEMs are facing these new requirements with mass-customization and more efficient digital methods. With the ongoing virtualization of production planning and verification, new technologies are necessary to facilitate interaction with virtual environments for planning experts.
While marker-based motion capture systems have been used for tracking of worker movements and interaction with those digital models until now, with the advent of markerless tracking technologies like the Kinect depth cameras, those solutions attract more interest because of their reduced cost, ease of use and the absence of cumbersome suits and markers.
However, when using depth cameras, which have been primarily designed as short-range gaming devices, different issues arise when dealing with large-scale setups common in shop floor and production planning environments. Similar to traditional motion capture systems, it is necessary to deal with occlusions, camera registration and calibration and fusion of multi-depth-camera data to obtain a sufficiently large and robust tracking space.
For this purpose, we propose such a novel multi-depth-camera approach, which is able to register itself only by observing the worker's trajectory. It is then able to reliably track the worker movements across a large space and also optimize the skeletal tracking. Through its distributed and scalable approach, the multi-depth camera system can be set up flexible for versatile tracking scenarios in practical production planning use. Based on sensor hardware specific quality criteria, it is able to fuse incoming tracking data and to provide a coherent view of the tracking space even in difficult tracking situations.

Peer-review under responsibility of the scientific committee of 48th CIRP Conference on MANUFACTURING SYSTEMS - CIRP CMS 2015

Keywords: Planning; Virtual Reality; Object recognition

1. Introduction

With the trend towards bigger product variety [1] and mass-customization [2], the OEMs face the fact of rising demands on manufacturing technologies and production planning [3]. Nowadays, in automotive sector customers expect highly individualized cars for their personal needs with various options on equipment. Therefore the external product variety is continuously increasing, due to more body types, power trains and optional equipment range (compare [4]). Existing factories have to become more versatile and flexible, since mixed-model production has to take over new car bodies on the same production line at higher volume flexibility. Assuming typical car life cycles to stay constant at approximately 7 years for each model, this automatically leads to more frequent and shorter ramp-up phases, higher demands on planning and thus to higher costs in general.

In order to overcome these negative effects, digital models and tools are a promising approach to achieve higher production- and volume-flexibility and more efficient production planning [5]. In the context of smart factories, digitalized tools and methods will replace traditional processes and methods, since information on products (e.g. CAD data), processes (e.g. former production process) and resources (e.g. factory and tool layouts) are digitally available. Several tools already exist to digitally support the planner's

Peer-review under responsibility of the scientific committee of 48th CIRP Conference on MANUFACTURING SYSTEMS - CIRP CMS 2015
doi:10.1016/j.procir.2015.12.088

work, e.g. in car sequencing procedures, process planning and layout planning. Howard et al. found that per auto maker, one can find up to 200 individual IT-systems in a traditional assembly plant [6].

However, these tools are still using mostly desktop-based interaction concepts, which require a high amount of initial training for the operators, and are not directly usable by the workers, who in turn tend to hold key knowledge useful for process optimization, especially in collaborative workshop situations like the "continuous improvements process" (CIP) or "integrated production planning" workshops.

Furthermore, these tools also do not directly reduce the need for physical prototypes, which still are major cost factors in planning and optimization.

1.1. Vision for mixed-reality, collaborative workshops in Production Planning

To overcome these challenges, two promising current interaction technologies are virtual reality (VR) and augmented reality (AR), which allow for a more direct and natural involvement with the virtual representation of the processes and their components. However, interaction models for such methods are new and unproven and thus error-prone, currently rather cancelling out the advantages that were originally hoped for. Aurich et al. discuss advantages of virtual reality for CIP workshop situations [7].

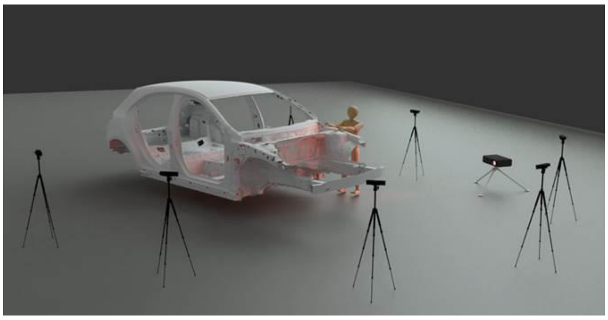

Fig. 1.: Rendering of virtual workshop situation with multiple-depth-camera based setup.

However, in the future, workshop managers will be able to choose the optimal degree between reality and virtuality, depending on the validation demand, available digital data quality and existing hardware. Blending real hardware components with virtual ones, a real, virtual reality or mixed reality scene can be created. Fig. 1 depicts a rendering of such a mixed reality scenario.

Every interaction by workshop participants ideally directly influences both the digital model and the reality simultaneously and synchronously, as presented by Ishii in 2012 [8]. Through this, direct and natural interaction is enabled and allows every participant of a collaborative workshop to modify models without specialized training.

To achieve the vision of such a virtual integrated workshop situation, two technical challenges have to be mastered – direct natural input, and VR-enabled output visualization. While the deployed visualization technology depends on the chosen degree of virtuality, a real-time tracking of human activity is a crucial enabling technology for direct and natural input. Gestures, viewpoint control and activity recognition are based on robust, real-time and large-scale tracking possibilities. The system presented in the following chapters offers all these features and will be described along exemplary use cases typical to the production planning domain.

In the remainder of this paper we first state possible requirements from the application domain in the context of this use cases. Afterwards, an affordable, fully functional technical solution is presented, which is then discussed with regard to the use cases. In the end, we conclude with an overall assessment and outlook on further optimizations.

2. Use Cases and Requirements on Motion Capture in Production Planning

2.1. State of the Art in Motion Capture Technology

The primary goal of motion capture technologies has always been to record human poses and motion, targeting a wide range of different application fields. Starting with use in the film industry, more and more analytical tasks have also been carried out using such systems, e.g. in medical or computer science. Until the recent advancements in 3D imaging, motion capture systems were typically built upon marker-based methods, where a large number of cameras are pointed towards the capture volume. The actors within the volume wear specialized suits, which carry spherical, retroreflective markers attached to characteristic body locations. Motion capture software then solves the measured location of those markers to a digital skeletal model [9], which can be used as an input for diverse applications.

With the advent of cheap and markerless 3D-imaging based tracking devices like the Microsoft Kinect around 2008, body-tracking-based interaction soon also found its way into various applications ranging from the original domain of casual gaming to research and industrial use despite their lower performance when compared to traditional MoCap systems. Again, the output of these tracking devices is a digital skeletal model, but besides this it is also possible to capture additional 3-dimensional features of the scene like geometry or overlayed color imagery, which also helped to enable many new applications.

Besides implicit, analytical tracking applications, skeletal tracking can also be employed to enable explicit interaction in augmented and virtual environments. This encompasses a wide range of interaction from gestural control of virtual parts projected onto a scene, up to fully immersive VR environments, where e.g. assembly operations can be carried out without any real prototypes from a first-person perspective. Numerous examples of works from the HCI field like [10] or [11] show off the capabilities, which also can be exploited for interaction in planning environments.

2.2. Motion Capture in Process Optimization Use Cases

In practice, production planning is using tracking and motion capture systems in order to optimize processes, products and ergonomics of workplaces.

Virtual interdisciplinary workshops prevalently use optical, large-scale tracking systems with marker suits to map human motion on an avatar or digital human model (DHM). Operators set up the virtual environments with the 3D geometry of the product and resources. By virtually simulating planned process, problems in planning can be identified.

One typical use case for motion capture systems is **process optimization** which is a crucial task of production planning. Having worked out drafts of process descriptions, interdisciplinary workshops are held to optimize processes. Process specialists are observing workers with hands-on experience, who are carrying out work tasks either on physical products or in virtual environments. By observing the worker's movements in the virtual scene, manufacturing values can be reduced and work load for mixed-model lines can be balanced. Unwanted dependencies between workers, unnecessary walking paths and the need for specialized handling tools are assessed in parallel. All motion data can be recorded for documentation purposes.

Furthermore, physical objects or mock-ups are additionally tracked for digital **buildability** checks. The tracked objects are used as interaction representations for the virtual vehicle parts. Virtual assembly parts are then checked for geometric consistency, thus ensuring that all of them can be assembled [5]. Having enough space in the static assembly position does not automatically imply that it can be built by the worker due to collisions in assembly paths of the part. Accessibility for tools and hands of workers are often not respected by product engineers in advance.

Additionally, motion capture systems are used to ensure the design principle of **visibility**. View-point verifications are carried out virtually by using DHMs. Critical mounting points are assessed related to their visibility, because task execution quality can then be directly checked by the worker, e.g. if the screw is correctly tightened. Concurrent Engineering enables production specialists still to influence the product according to the production-oriented optimizations (see [12]).

2.3. Motion Capture in Ergonomics Analysis Use Cases

Besides process optimizations, ***ergonomic analysis*** of workplaces is an important subject for production planning. Motion capture systems are widely used in order to virtually support evaluation and optimization. Literature presents various methods on ergonomic analysis, like EAWS[13], NIOSH or OCRA which are all in automotive company use.

In 2011 the ErgoToolkit presented and integrated an ergonomic analysis method into virtual manufacturing software [14], based on digital human models (DHM), but lacked the possibility of easy and interactive posture adaption. Postures had to be refined by manual work. DHMs now can be accurately and interactively controlled by motion capture systems. Workload and ergonomic assessments of manual assembly tasks therefore can be carried out in real-time as well as in a recorded, post-hoc manner. For this, spatial parameters like posture and temporal parameters like repetitive actions can be assessed, such as presented by Martin et al. in [15] or in [16].

Gathering the skeletal tracking data of the worker during assembly tasks, distinct activities can be recognized based on the skeletal configuration and temporal aspects using different techniques like machine-learning e.g. using SVM classifiers [17] or rule-based inference [18]. Ongoing research on automatic assessment of motions using biomechanical human model like DYNAMICUS [19] is facilitating process planning based on MTM-Universal Analyzing System, since motion data can automatically be segmented, classified and assigned to work instructions. Task execution times, task orders and ergonomic assessments can be derived automatically and thus also assessed and analyzed.

2.4. Requirements on the Use of Depth-Camera-based Motion Capturing in Production Planning

When trying to enable a widespread use of multi-depth-camera based motion tracking systems in the previously described use cases, a range of goals must be met, which are currently not or only partially satisfied with traditional motion capture systems. Ideally, virtual production planning workshops with motion capture systems are quick to set up, offer markerless tracking systems and make virtual assessments as easy as physical ones and do not need permanent support of virtual tool operators.

2.4.1. Precision

Most of the traditional motion capture systems are designed and optimized for sub-millimeter precision on a marker level, whereas for ergonomic and process optimizations this degree of accuracy is mostly not necessary. Embracing this aspect not only helps to cut costs, but also enables easier setup procedures like the one proposed with the multi-depth-camera implementation in this paper.

2.4.2. Ease of use

Another advantage with depth-camera motion tracking systems is the markerless approach. While this approach also results in lower (but still sufficient) tracking precision, it eliminates the need for putting on and wearing a marker suit, making it possible to capture workers in their regular clothes. This not only reduces setup time per actor, but also improves social acceptance, as several workers could get embarrassed when wearing tight tracking suits in front of the other workshop participants. Furthermore, motions may differ from the actual assembly motions when wearing working clothes.

Furthermore, for traditional motion capture systems and the downstream processing pipeline, considerable expert knowledge and also time is necessary to setup and maintain the capture system as well as the tool platform for specific tasks. To allow also non-experts to setup and use such a system, a plug-and-play operation is desirable, without complicated setup or adjustment procedures. Besides reduced time and costs, this also increases the portability of the system.

2.4.3. Costs

Initial costs for motion capturing systems are usually very high, and not always directly related to the achievable precision and quality (see Thewlis [20]). As traditional motion capture facilities also tend to be stationary because of their extensive setup procedures, it is necessary to maintain multiple motion capture areas, thus further increasing costs.

With a multi-depth-camera system, costs can be cut to a fraction of a traditional motion capture system for a comparable capture volume. Furthermore, due to the short setup time the system is more portable and can easily be moved between different departments or plants, thus eliminating the costs for the purchase of multiple systems.

Based on this main requirements, the system presented in this paper may help to spread motion-tracking based production planning methods further within the manufacturing industry, and allow for a more frequent use of this methods while reducing costs and demands on expert staff.

3. Implementing a Multi-Depth Camera Tracking System

A multi-depth-camera system has different advantages both over a traditional MoCap system and also over using single depth cameras. Having already lined out possible drawbacks of motion capture systems in the previous chapter, using single depth cameras also entails several drawbacks. For example, they are prone to occlusion effects through the workers' body, or parts and other structures within the capture area. Additionally, with a single depth camera, the tracking area is restricted to the field of view of this camera, which is usually quite limited.

Using multiple depth cameras, it is possible to overcome these effects and to achieve a more robust tracking with extended range at a fraction of the price of common MoCap systems as described in the following sections.

3.1. Depth Camera Technology

Different types of depth cameras have emerged over the last years, most notably Structured Light (SL)-based and Time-Of-Flight (ToF)-based cameras. The Kinect v1, which made depth sensing affordable not only for gaming but to a whole range of communities across industries and research, was based on SL technology. The new revision of the Kinect (v2) is now based on ToF technology that delivers more robust sensing and increased effective resolution at still affordable prices. The ToF principle works by emitting modulated light pulses, whose delay is measured through the phase difference when the pulses return to a sensing chip within the camera [21]. Based on the known speed of light and the delay, a distance for each pixel can be calculated, resulting in a 3D image of the scene in front of the camera. This allows e.g. for easy segmentation of body silhouettes for further processing like skeletal tracking [22].

However, when using multiple ToF cameras, different challenges have to be faced in order to successfully implement a multi-depth-camera system:

- **System Architecture**
  (data collection and sensor control)
- **Temporal Synchronization**
  (of unsynchronized, arbitrarily delayed data)
- **Interference**
  (between multiple cameras illuminating the scene)
- **Registration**
  (of multiple cameras to a common coordinate system)
- **Fusion**
  (of skeletons tracked by multiple cameras)

In the following, we present a solution, which will address all of those topics, and thus enable and support the use of a multi-depth-camera system for production planning use cases amongst others. While the first two challenges (architecture and synchronization) can be solved with common, well-established methods from computer science like lightweight, and distributed service-oriented architectures or network-based time synchronization (NTP), the latter three require additional attention.

3.2. Interference handling

In earlier active depth sensing devices based on SL technology, interference between multiple cameras was a major drawback, although different approaches from research provided potential solutions, e.g. by Butler et al. [23].

However, through the use of the Kinect v2 with its ToF technology, interference between the cameras is already at an acceptable level to successfully run skeletal tracking applications, because ToF uses a modulated light emitter, which allows for better coexistence of multiple devices [24].

3.3. Registration

Regarding registration, earlier methods (e.g. in [25]) until now used mostly solutions relying on external helpers and tools such as checkerboard patterns for calibrations, which also made an explicit, time-consuming and error-prone registration step necessary. Other methods proposed by literature are ICP based using the 3D point cloud [26] and also need additional time and computing power to run.

Contrary to this, we propose an implicit process for registration based on the users' skeletons moving in the capture volume. By running an ICP-like approach on the skeletal joints being tracked by the Kinect v2 cameras, it is possible to start registration as soon as an user enters the capture volume, and to have a first, coarse registration between any pair of cameras after the first frame in which the skeleton was captured by both cameras. With further frames being collected, the computed registration is also being iteratively refined. In practice, this means that there are no additional setup steps necessary besides placing the cameras and their computing nodes to deploy a functional tracking setup.

3.4. Fusion

Having transformed single sensor skeletal data into one common coordinate space, this data from different sources can be combined in a coherent view. As multiple cameras may provide redundant information of a certain user within the tracking space, also with different tracking quality parameters, it is necessary to decide which input should have more influence on the overall tracking output. Methods for this task range from simple best-skeleton-or joint-counting [27] approaches over weighted approaches [28] up to specialized fusion methods including several weighting factors (see [29], [30] or [25]). For our solution, we implemented a combination of different approaches, using quality rating algorithms on a joint basis as well as different correction methods to acquire stable and robust tracking data also in difficult tracking situations.

3.5. Putting it all together

By solving the challenges imposed by using multiple Kinect v2 sensors, it is now possible to implement an integrated suite of software components, which together form an easy-to-use environment for Kinect-based motion tracking and analysis. In our case, different distributed components have been created:

- **Sensor Server:** Enables other components to query tracking data from single sensor via REST-based web interface
- **Fusion Server:** Handles registration and fusion between multiple sensors
- **Visualizer / Recorder:** 3D visualization of tracking data from the sensors or fusion service, also able to record all tracking data to a file.

The components allow for an easy data exchange and integration with other components through open, REST/JSON-based web interfaces which can be accessed with almost any common programming environment. This allows for simplified implementation of third-party solutions based on tracking data. An additional, mDNS-based discovery mechanism also greatly simplifies networking setup, further reducing the knowledge for system setup.

4. Discussion

4.1. Suitability for production planning use cases

With regard to the use cases presented in chapter 2, the implemented system is able to meet the derived requirements to a high degree.

Comparison measurements and evaluation with a marker-based motion capture system serving as ground truth showed that the spatial tracking precision with a setup consisting of six Kinect v2 depth cameras is able to reach around 6 cm of mean euclidian distance error between the tracked joints of the depth camera setup and the ground truth. For most of the common production planning use cases, this can be considered sufficient, especially as the error is also partially originating from differing skeletal models and inferred joint positions through the markerless tracking approach.

Besides this, there are no limitations on the sensor positions and orientations, or amount of fused cameras, also fostered by the modular and distributed approach. Practical setups showed up good results, using 6 inward-looking sensors for human posture analysis and 10 or more sensors for large-scale tracking areas.

The implemented system is also cost-efficient, with around 500 € for a single tracking node and around 4.000 € for a six-camera setup. Traditional marker-based motion capture setups for a comparable capture volume start at around 20.000 €.

Furthermore, through the use of skeleton-based camera registration process and sensor network discovery services, the distributed system is practically setup-free, thus reducing the deployment barriers. After positioning the cameras freely around the area, extrinsic registrations are calculated on-the-fly, just by moving within the capture volume. This allows for a fast setup and relocalization of the system without the need for additional experts or maintenance personnel.

4.2. Extending tracking application areas in production planning

When looking at possible new and previously already mocap-based applications, it is still necessary to implement suitable interfaces to use tracking data from the presented system as an input for recognition and interaction systems.

This however also offers possibilites to use this data for further use cases, e.g. for an interactive, on-site walking path analysis or fully immersive VR environments which allow collaborative interaction on purely virtual models and environments without the need for any physical prototypes.

With more widespread use, these virtual environment methods need to prove long-term, if they can offer at least the same amount of involvement in the production preparation workshops, support planners at least to identify the same amount of problems and also offer the possibility to solve them.

5. Conclusion and Future Work

We presented a solution for a multi-depth-camera based motion tracking system to support production and process planning, which is affordable and easy to setup. This system can help in fulfilling different requirements for digitally supported process engineering from activity recognition to gestural interaction with 3D models.

In the next step, it is possible to enrich skeletal tracking data with data acquired from additional (also non-optical) sensors, which is especially helpful for improved activity recognition (like it is currently executed in the European project INTERACT [31]), but can also be used for easier interaction e.g. with virtual entities through physical “proxies”.

Furthermore, it is possible to use the registered depth data to build a fused point cloud, which not only can be used for motion tracking, but also for object segmentation and recognition, e.g. to track tools or parts across the working space. Besides this, the use of this system is not restricted to production planning use cases, but can be employed almost everywhere where a large-scale tracking of people is necessary.

Particularly in the automotive industry, there are several additional use-cases, which will directly benefit from this contribution, e.g. markerless virtual training for workers, after-sales disassembly routines and marketing applications.

Acknowledgements

The work presented in this paper was funded through the European research project INTERACT-FP7, Grant Agreement Nr. 611007.

References

[1] ElMaraghy, H.A., ElMaraghy, W.H., 2014. Variety, Complexity and Value Creation, in "Enabling Manufacturing Competitiveness and Economic Sustainability" (M. F. Zaeh) Springer International Publishing, Cham, pp. 1–7.

[2] Rungtusanatham, M.J., Salvador, F., 2008. From Mass Production to Mass Customization: Hindrance Factors, Structural Inertia, and Transition Hazard. *Prod. Oper. Manag.*, 17 (3), pp. 385–396.

[3] Weyand, L., Bley, H., 2008. Future Challenges in Process/Resource Planning Due to Increasing (Product) Variants. "Manufacturing Systems and Technologies for the New Frontier" (M. Mitsuishi et al.) Springer London, London, pp. 235–238.

[4] Pil, F.K., Holweg, M., 2004. Linking product variety to order-fulfillment strategies. *Interfaces*, 34 (5), pp. 394–403.

[5] Weber, J., 2009. Automotive development processes, *Springer*, 303.

[6] Mickey Howard, M.H., 2004. Investigating the intangible: lessons learnt from research into automotive inter-organisational IT systems. *Int. J. Automot. Technol. Manag. - Int J Automot Tech Manag*, 4 (4).

[7] Aurich, J.C. et al., 2009. Improvement of manufacturing processes with virtual reality-based CIP workshops. *Int. J. Prod. Res.*, 47 (19), pp. 5297–5309.

[8] Ishii, H. et al., 2012. Radical atoms: beyond tangible bits, toward transformable materials. *Interactions*, 19 (1), pp. 38–51.

[9] Dai, H. et al., 2010. Skeletal Animation Based on BVH Motion Data. "2010 2nd International Conference on Information Engineering and Computer Science (ICIECS)" 2nd International Conference on Information Engineering and Computer Science (ICIECS). pp. 1–4.

[10] Using Multiple Kinects to Build Larger Multi-Surface Environments, http://ase.cpsc.ucalgary.ca/uploads/Publications/2013/Using%20Multiple%20Kinects%20to%20Build%20Larger%20Multi-Surface%20Environments.pdf (accessed January 23, 2014)

[11] Suarez, J., Murphy, R.R., 2012. Hand gesture recognition with depth images: A review. "2012 IEEE RO-MAN", IEEE RO-MAN. pp. 411–417.

[12] Kusiak, A., 2012. Concurrent Engineering: Design of Assemblies for Schedulability. *Control Dyn. Syst. V47 Manuf. Autom. Syst. Tech. Technol. Adv. Theory Appl.*, p. 221.

[13] Schaub, K. et al., 2013. The European Assembly Worksheet. *Theor. Issues Ergon. Sci.*, 14 (6), pp. 616–639.

[14] Alexopoulos, K. et al., 2013. ErgoToolkit: an ergonomic analysis tool in a virtual manufacturing environment. *Int. J. Comput. Integr. Manuf.*, 26 (5), pp. 440–452.

[15] Martin, C.C. et al., 2012. A real-time ergonomic monitoring system using the Microsoft Kinect. "2012 IEEE Systems and Information Design Symposium (SIEDS)", pp. 50–55.

[16] Diego-Mas, J.A., Alcaide-Marzal, J., 2014. Using Kinect^{TM} sensor in observational methods for assessing postures at work. *Appl. Ergon.*, 45 (4), pp. 976–985.

[17] Koppula, H.S. et al., 2013. Learning human activities and object affordances from RGB-D videos. *Int. J. Robot. Res.*, 32 (8), pp. 951–970.

[18] Hachaj, T., Ogiela, M.R., 2013. Rule-based approach to recognizing human body poses and gestures in real time. *Multimed. Syst.*, 20 (1), pp. 81–99.

[19] Härtel, T. et al., 2011. Capturing and Assessment of Human Motion during Manual Assembly Operation. "First International Symposium on Digital Human Modeling", Lyon, France.

[20] Thewlis, D. et al., 2013. Next generation low-cost motion capture systems can provide comparable spatial accuracy to high-end systems. *J. Appl. Biomech.*, 29 (1), pp. 112–117.

[21] Sell, J., O'Connor, P., 2014. The Xbox One System on a Chip and Kinect Sensor. *Micro IEEE*, 34 (2), pp. 44–53.

[22] Shotton, J. et al., 2012. Efficient Human Pose Estimation from Single Depth Images. *Trans PAMI*.

[23] Butler, D.A. et al., 2012. Shake'n'sense: reducing interference for overlapping structured light depth cameras. "Proceedings of the SIGCHI Conference on Human Factors in Computing Systems". New York, NY, USA: ACM, pp. 1933–1936.

[24] Kim, Y.M. et al., 2008. Design and calibration of a multi-view TOF sensor fusion system. "IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops" (CVPRW), pp. 1–7.

[25] Zhang, L. et al., 2012. Real-time human motion tracking using multiple depth cameras. "IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)", pp. 2389–2395.

[26] Yamazoe, H. et al., 2012. Easy depth sensor calibration. "2012 21st International Conference on Pattern Recognition (ICPR)", pp. 465–468.

[27] Caon, M. et al., 2011. Context-Aware 3D Gesture Interaction Based on Multiple Kinects. AMBIENT 2011, The First International Conference on Ambient Computing, Applications, Services and Technologies. pp. 7–12.

[28] Maimone, A., Fuchs, H., 2011. Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras. "2011 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR)", pp. 137–146.

[29] Asteriadis, S. et al., 2013. Estimating Human Motion from Multiple Kinect Sensors. "Proceedings of the 6th International Conference on Computer Vision / Computer Graphics Collaboration Techniques and Applications". New York, NY, USA: ACM, pp. 3:1–3:6.

[30] Obdrzalek, S. et al., 2012. Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population. "2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)", pp. 1188–1193.

[31] INTERACT | Interactive Manual Assembly Operations for the Human-Centered Workplaces of the Futue, http://www.interact-fp7.eu/