Motion Analysis of work conditions using commercial depth cameras in real industrial conditions

Dr. Pierre Plantard, 14C Rue du Patis Tatelin, 35700 Rennes, France, +33 681318282, pierre.plantard88@gmail.com, SATT Ouest Valorisation
Dr. Hubert P. H. Shum, Ellison Building, Faculty of Engineering and Environment, Newcastle upon Tyne, NE1 8ST, UK, +44 1912437607, hubert.shum@northumbria.ac.uk, Northumbria University
Pr. Franck Multon, M2S Lab., Avenue Robert Schuman, 35170 Bruz, France, +33 631646357, fmulton@irisa.fr, University Rennes, Inria

Abstract
Measuring human motion activity in real work condition is challenging as the environment is not controlled, while the worker should perform his/her task without perturbation. Since the early 2010's affordable and easy-to-use depth cameras, such as the Microsoft Kinect system, have been applied for in-home entertainment for the general public. In this chapter, we evaluate such a system for the use in motion analysis in work conditions, and propose software algorithms to enhance the tracking accuracy. Firstly, we highlighted the high performance of the system when used under the recommended setup without occlusions. However, when the position/orientation of the sensor changes, occlusions may occur and the performance of the system may decrease, making it difficult to be used in real work conditions. Secondly, we propose a software algorithm to adapt the system to challenging conditions with occlusions to enhance the robustness and accuracy. Thirdly, we show that real work condition assessment using such an adapted system leads to similar results comparing with those performed manually by ergonomists. These results show that such adapted systems could be used to support the ergonomists work by providing them with reproducible and objective information about the human movement. It consequently saves ergonomists time and effort and allows them to focus on high-level analysis and actions.

Keywords
Depth cameras, Digital Human Model, ergonomic assessment, motion capture, sensor validation.

Introduction

The Objective evaluation of postures is a key factor when considering the demanding nature of work in the manufacturing industry. Several methods, such as the famous Rapid Upper-Limb Assessment (RULA) (McAtamney & Corlett, 1993) require collecting joint angles in order to compute scores that can help ergonomists to quantify potential pain caused by a posture at work. Although it is easy to be carries-out in real conditions, manual collection of the joint angles is a tedious task and the results may differ from one experimenter to another (Burdorf, Derksen, Naaktgeboren & Van Riel, 1992; Fagarasanu & Kumar, 2002). Measurement based on motion capture systems aims at providing relevant and reproducible data and facilitating such a process for ergonomists. However, most accurate systems, such as the opto-electronic devices, cannot be used easily in real working conditions with a cluttered environment. It is also problematic to stop the working process to install, calibrate and use such systems. Alternatives consist of using wearable devices, such as the Inertial Measurement Units (IMU), which are promising means to capture workers' motion in cluttered environments such as workplaces (David, 2005; Li & Buckle, 1999). However, they require to place sensors on the body and generally need frequent calibrations to compensate measurement drift, leading to the potential perturbation of the working process in production lines. Hence, the ideal system should be markerless, with no calibration, providing reliable 3D data, and robust to cluttered environments. Video-based systems look promising but the current technology requires several cameras. Moreover, it requires heavy calibration to segment relevant data and to reconstruct the 3D position of feature points.

In the early 2010’s, depth cameras have been widely used in video games for the general public audience. Their main advantage is the ability to capture 3D information without wearable sensors/markers nor calibration. However, they have been designed to work in indoor living rooms with open spaces. Machine learning (Shotton et al., 2011) is used to segment the depth images and estimate the 3D location of joint centers of a simplified skeleton. Hence, Microsoft Kinect and its competitors have been widely used in many application domains (far beyond video games) to capture the 3D motion of users. In particular, several works have been carried-out in medicine in order to support the rehabilitation of patients at home or in clinics (Bonnechère et al., 2014; Clark et al., 2012). These previous works demonstrate that Kinect could be used in clinics to evaluate rehabilitation protocols for the upper-limbs and could be adapted to analyse treadmill walking (Auvinet, Meunier & Multon 2012; Auvinet, Meunier & Multon, 2015; Galna et al., 2014).

All these previous works used the Kinect in situations that are similar to those recommended by the provider, i.e. without occlusions, and the sensor is placed in front of the user with simple planar motion. The real working condition in the manufacturing industry are totally different, which may cause the system to fail to track the human body. In the next section, we propose to clarify to what extent Kinect can provide relevant information and its limits for joint motion capture.

Several authors proposed methods to enhance the quality of MS Kinect data, especially in complex situations with occlusions and non-recommended use. We report an overview of these correction methods in the section “Correction of Kinect data”. We then focus on the use of such a system for the evaluation of postures at work to fill-in ergonomic grids, such as RULA.

Recently, Microsoft decided to stop the development of the Kinect, but several alternatives exist, and the information reported in this chapter is also applicable to other systems based on depth cameras.

The Validity of Kinect sensor for ergonomic assessment

Several papers have been published about the accuracy of joint kinematics evaluation using depth images and tracking algorithms. For clinical gait analysis, a recent review paper (Springer & Yogev Seligmann, 2016) demonstrated good accuracy to measure spatiotemporal aspects of gait if specific methods (Auvinet et al., 2012) are developed, such as estimating relevant gait events or global clinical indexes, such as asymmetry indexes. However, they reported weak joint angles estimation, which is a limitation for many other applications.

Mentiplay et al. (2018) reported variable reliability of the Kinect V2 in estimating trunk, hip and knee joint angles during squats and drop vertical jumps. It demonstrated good to excellent inter-session reliability but sometimes with poor reliability. For postural control, with low joint ranges and almost planar motions, even Kinect V1 seems to offer good inter-trial reliability and concurrent reliability with a reference motion capture system (Clark et al. 2012). Most of the previous works consider standardized situations, with the recommended sensor placement in front of the subject, and almost planar motions (Bonnechère et al., 2014; Kurillo, Chen, Bajcsy & Han, 2013). But it has been shown that the error is dependent on the performed postures (Xu & McGorry, 2015).

In real work conditions, with cluttered environments, it seems difficult to place the sensor at the recommended place, and to limit the movement to almost planar ones. Ergonomic assessment should be performed on-site, with 3D complex movements. Previous works generally concluded 2cm mean error in the estimation of a joint center located on the upper-limbs. However, in complex and cluttered environments, this value may reach very high values, as suggested in previous works when partial occlusions occur. Dutta (2012) has shown that the depth image offers good reliability, which means that the main errors occur afterwards, when tracking a human body in these images.

To better understand the actual accuracy and limitation of such a system, we carried out a comprehensive study based on manipulating virtual mannequins and virtual sensor placement, enabling us to tests more than 500.000 joint and sensor configurations (Plantard, Auvinet, Le Pierres & Multon, 2015). In this study, an anthropometric 3D mesh representing the surface of the human body is generated using MakeHuman[1] software, and a skeleton is associated with the mesh to control its pose. The resulted mesh is then transmitted to the Microsoft Kinect pipelines to be processed as when using real depth images. It is then possible to compare joint angles estimated by this system to those actually used to deform the virtual human mesh. The virtual human pose was defined by three parameters: azimuth

(0° to 110° with a 10° step), elevation (−45° to 45° with a 10° step), and depth (associated with the elbow flexion ranging from 0° to 110° with a 10° step). If the hand position is fixed, the elbow is still free to swivel about a circular arc whose normal vector is parallel to the axis from the shoulder to the hand. Hence, we have chosen to sample the swivel angle in three main values: 0°, 90°, and 135° which corresponds to the main kinds of grips one can see in industrial work. It produced 4752 various poses, combined with various sensor position/orientation around the character: azimuth (−50° to 50° with a 10° step) and elevation (−50° to 50° with a 10° step). In these examples, natural auto-occlusions occur, and we demonstrated that it is the main reason why joint position estimation deviates from real values.

Shoulder, elbow and wrist position estimation error were (0.019 ± 0.009 m), 0.018 ± 0.023 m and 0.024 ± 0.038 m. Consequently, mean joint angles error was 4.5° ± 8.9° and 12.6° ± 17.2° for the shoulder and elbow respectively. However, peak error could reach more than 40°, which leads to very important errors when using ergonomic assessment grids such as RULA. All these peaks correspond to occlusions, which is a strong limitation for using such a device in real working condition. These errors can reach higher values (such as 56° and 41° for the shoulder and the elbow respectively) when the sensor was not placed at the recommended position. This is especially true when azimuth increases in both directions (sensor is placed more or less on the side of the subject), and less when elevation changes. An example of error estimation is given in Figure 1. Complete results can be found in (Plantard et al., 2015).

[1] MakeHuman: www.makehuman.org

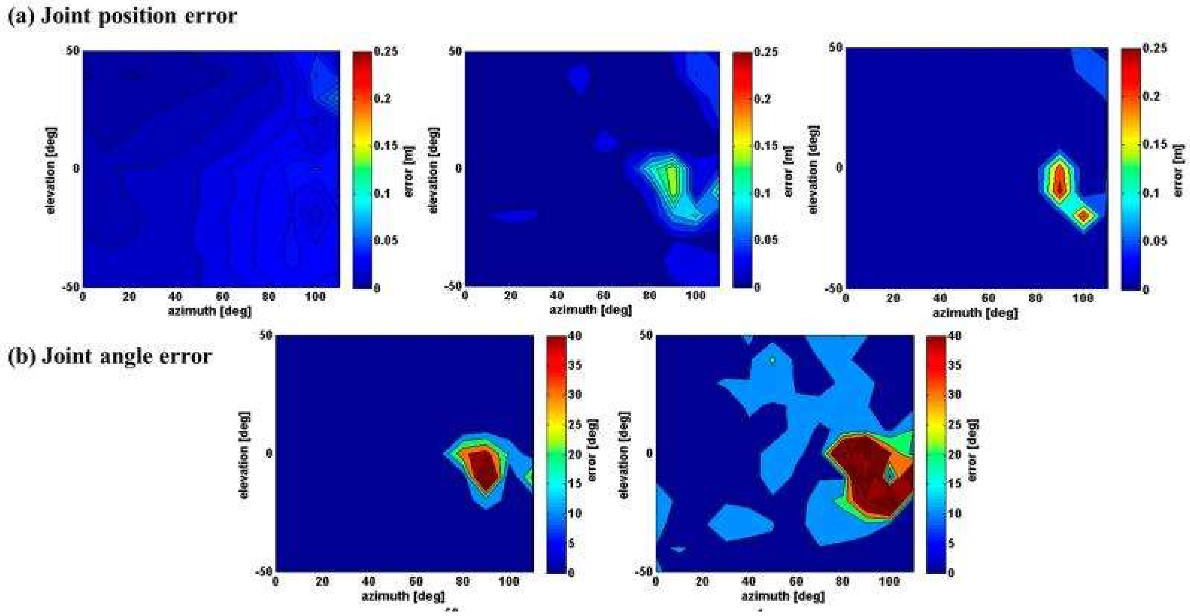

Figure 1: Accuracy of the Kinect measurement of the 135° swivel angle poses relative to azimuth and elevation pose parameters and with a zero elbow flexion (a) Error distribution of the shoulder (left), elbow (center) and wrist (right) joint positions estimated; (b) Error distribution of the shoulder (left) and elbow (right) joint angles calculated. Adapted from Plantard et al. (2015).

As a conclusion, this comprehensive study reveals that joint position and angle estimation could be used in ergonomic assessment grid, such as RULA, but only if no occlusion occurs, and for the recommended sensor placement. As this is not the case in real working condition, where the sensor cannot be placed as recommended, and occlusion occurs, reconstruction of the joint angles have to be performed before conducting ergonomic studies.

Correction of Kinect data

The previous section shows that depth images contain relevant information (Dutta, 2012) but tracking the human joints introduces errors (Plantard et al., 2015). Hence, most Kinect data correction methods focused on tracking the human body in depth images. As the method implemented by Microsoft (Shotton et al., 2012) is based on machine learning, its accuracy mainly depends on the database of poses used to train the system. Hence, it does not actually consist in a tracking system in which patterns are recognized and tracked for each frame. Alternatives exist, such as the method developed by Wei, Zhang & Chai (2012), It has been extended to multi-Kinect recently (Shuai, Li, Guo, Prabhakaran & Chai, 2017): minimizing the fitting error between an ellipsoid-based skeleton and the input point cloud data captured by multiple depth cameras. However, such a method requires calibration consisting of aligning an ellipsoid skeleton to depth images, which might be complex in real work condition with only few times dedicated to the experiment.

An alternative consists of using the proposed reconstructed skeleton (Shotton et al., 2012) as an input and try to filter or correct errors. Most of the previous works make use of pose priors modelled from a motion capture dataset. Unreliable body parts could be replaced using knowledge extracted from these pose priors aiming at retrieving more reliable poses. Regression-based approaches (Shen et al., 2012; Shen et al., 2014; Shen, Lei, Zeng & Zhang, 2015) consists in learning a regression function mapping an initially estimated skeleton to a corrected one by considering the systematic bias existing in the estimation step. However as these methods deal with each joint independently, irregular corrected poses may occur. Moreover, the bias may not be totally systematic and cannot be approximated by linear statistical models.

Another solution based on pose priors consists in searching in a database poses that are similar to the one observed with the Kinect. Hence, unreliable parts could be replaced by a relevant combination of these similar poses (Shum, Ho, Jiang, & Takagi, 2013; Ye, Wang, Yang, Ren, & Pollefeys, 2011). It ensures regular poses but does not guarantee that the resulting pose actually corresponds to the one performed by the subject. Indeed, it can only predict poses within the prerecorded pose space: if no pose with arms above the head is present in the database, it is not possible to correct unreliable arms' positions this way. Moreover, as each frame is processed independently, similar poses are searched in the database without taking past information into account.

Because we have no control over the experimental set-up in real working conditions, regression-based approaches seem difficult to apply: occlusions and noise may totally change from one situation to the other. We consequently proposed to adapt the search&replace approach in order to take continuity into account (Plantard, Shum & Multon, 2017). In previous works, continuity cannot be ensured because the optimization process used to select similar poses and to combine them restarts from scratch for each frame. Let us consider joint j estimated as reliable at frame T, and then becomes unreliable at frame T+dt. At frame T, previous methods will search for similar poses in which the configuration is similar for such a joint. It will then select and combine a set of similar poses/candidates based on this joint configuration. At frame T+dt, this joint is considered as unreliable, and is not used to select potential candidates any more. Consequently, selection may choose completely different candidates, leading to a discontinuity. Already considering candidates that were used at frame T would help to find only those that ensure continuity.

To achieve this goal, we introduced a new data structure, named "Filtered Pose Graph" (FPG) to encode the space of possible poses/candidates. FPG consists of a graph in which edges are representative poses, and edges connect two of them is we consider that it would not lead to discontinuities. A representative pose is a pose that is sufficiently different from all the other poses already stored in the graph, so that it brings new information for future combination. Indeed, a database composed of several very similar poses would guide the composition process to the same type of results, without variability. For example, let us consider a database in which thousands of low position of the arms are captured and only one above the head. If the combination process is searching for 30 similar poses for a pose in which the arm is in the upper position, it will retrieve 29 in low positions and only one in a high position. It will have little chance to simulate a pose with arms up. To tackle this problem, we perform a two-steps process: 1) filter each clip to eliminate similar poses within a clip, and 2) filter the remaining poses in-between clips, to eliminate remaining similar poses that have been captured in several clips.

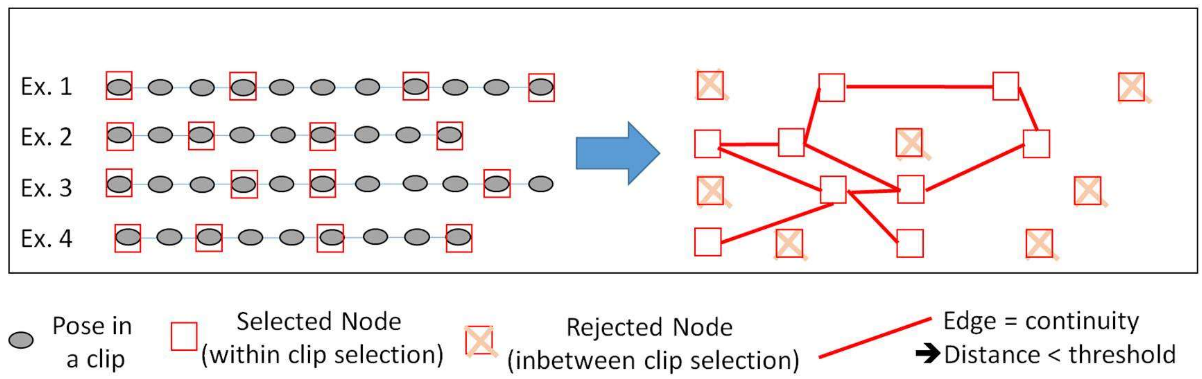

Figure 2 depicts an example with 4 clips, in which red boxes are poses considered as not similar to the previous ones, within each clip, which corresponds to the first step. On the right of

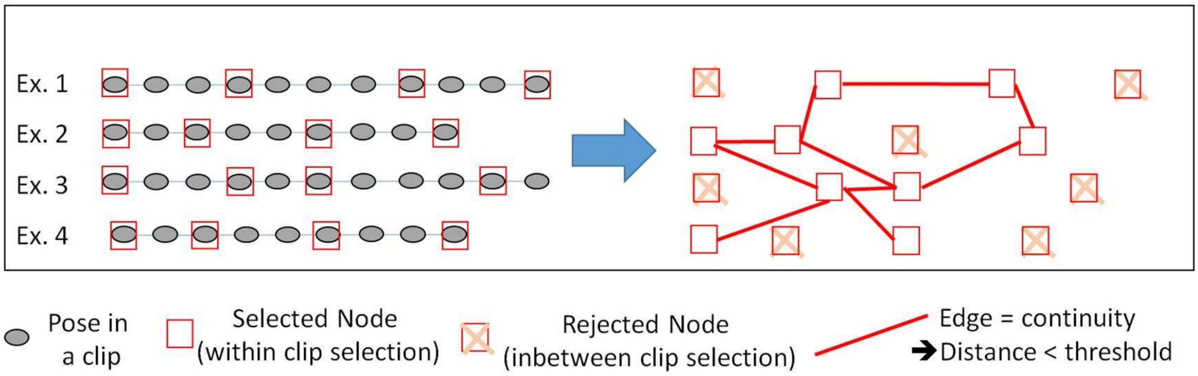

Figure 2, poses that are too similar to those selected in other clips are eliminated, corresponding the second step; the remaining poses are depicted as blue boxes.

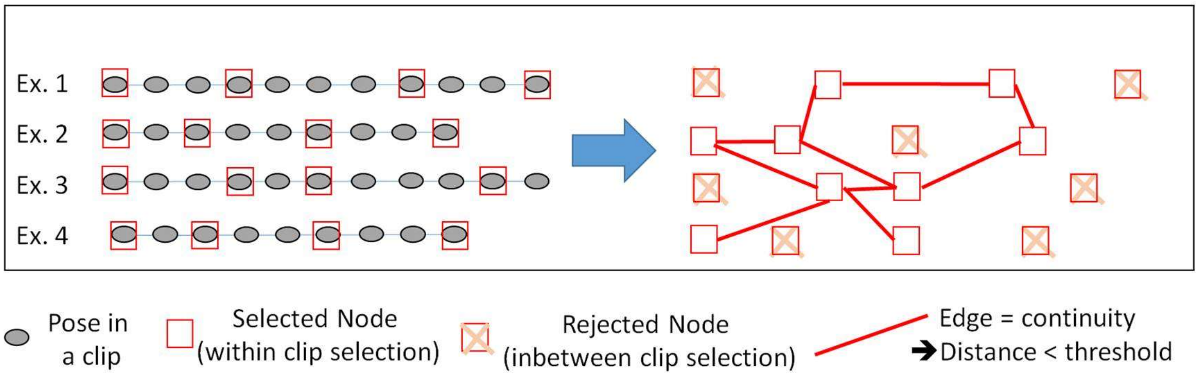

Figure 2: filtering the clips (Ex. 1 to 4) to eliminate too similar poses and build the Filtered Pose Graph. Left: Pruning similar poses within each clip (poses surrounded by squares are kept, remaining is eliminated). Right: Pruning similar resulting red nodes to eliminate too similar poses, and building edges between remaining nodes if the distance between them is below a threshold (no discontinuity).

We then build a graph in which nodes are these poses (blue boxes) and edges are created if the distance between two nodes is below a given threshold, assuming that it is possible to transit between them without discontinuity. For more details, please refer to (Plantard et al., 2017).

Selecting potential candidates in the Filtered Pose Graph based on the pose delivered by the Kinect involves defining a dedicated metric. This metric aims at comparing the input Kinect pose and all the potential candidates (Filtered Nodes) to find good candidates before reconstruction.

During the pose correction process, the FPG is used to preselect a subspace of poses that are supposed to be in continuity with the previous pose. Let us consider a set of N nodes S_T selected at frame T (depicted as red boxes in the FPG in the left part of

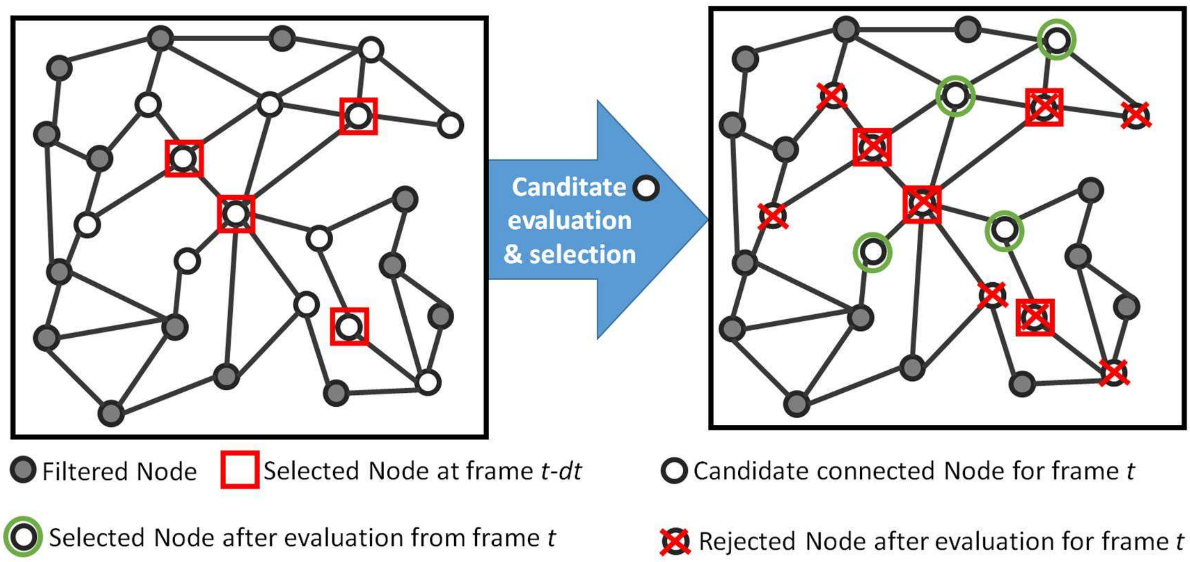

Figure 3). To ensure continuity at frame T+dt, we preselect poses that are directly connected to those in S_T (depicted as white circles in

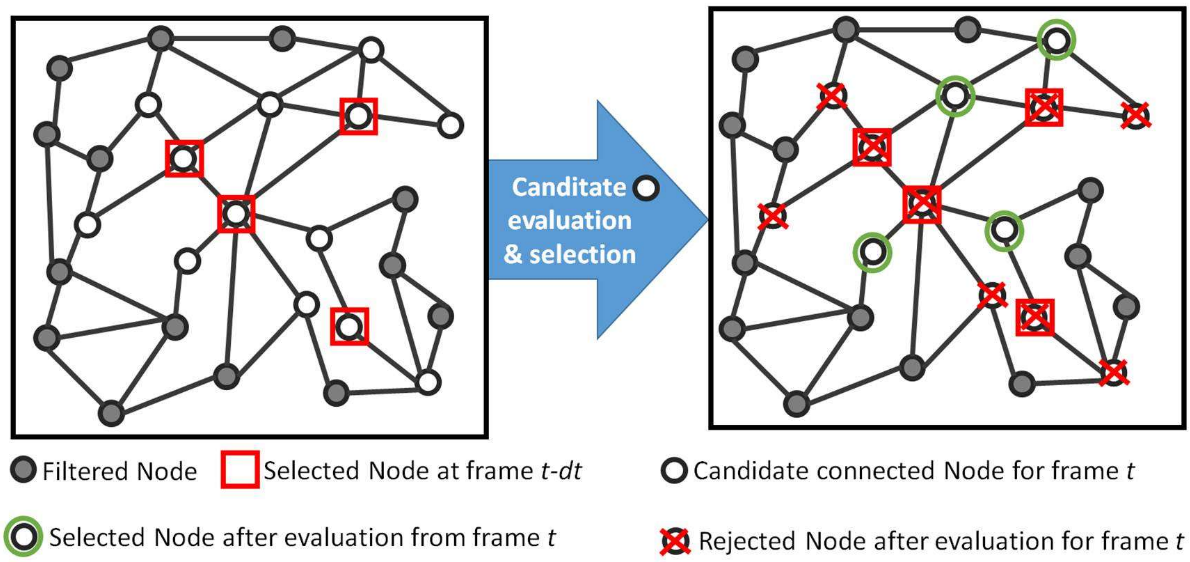

Figure 3). The number of poses in this preselection is greater than N, the fixed number of poses we wish to use for correction. We then evaluate each preselected pose to select the N best ones, i.e. best compromise between continuity for unreliable joints, and similarity with the current pose delivered by the Kinect for reliable joints. The N selected poses are depicted with green circles in

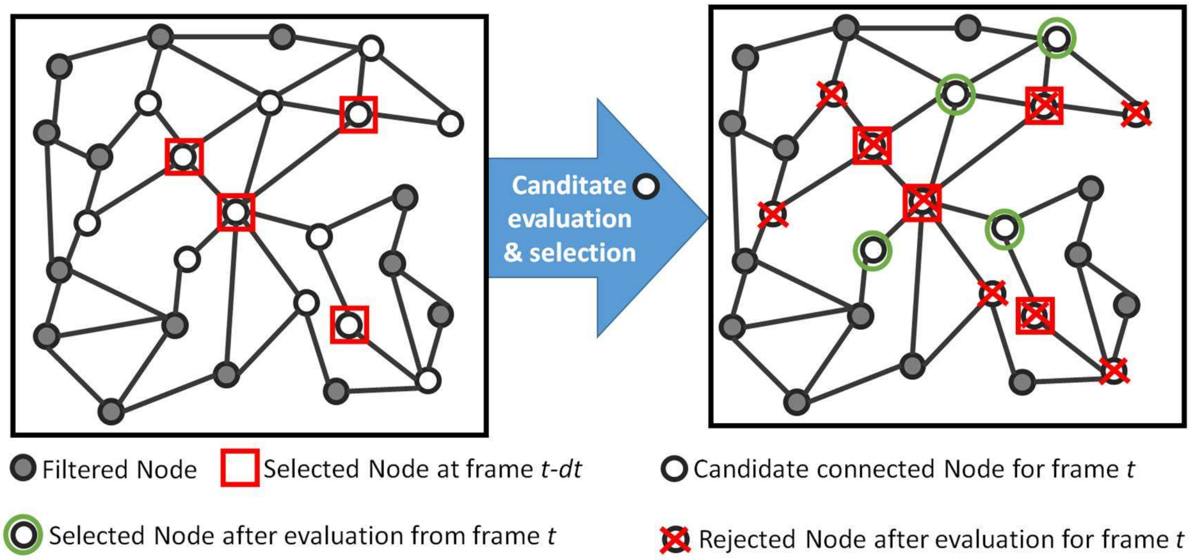

Figure 3 and grouped in S_{T+dt}. The process is repeated for each new frame.

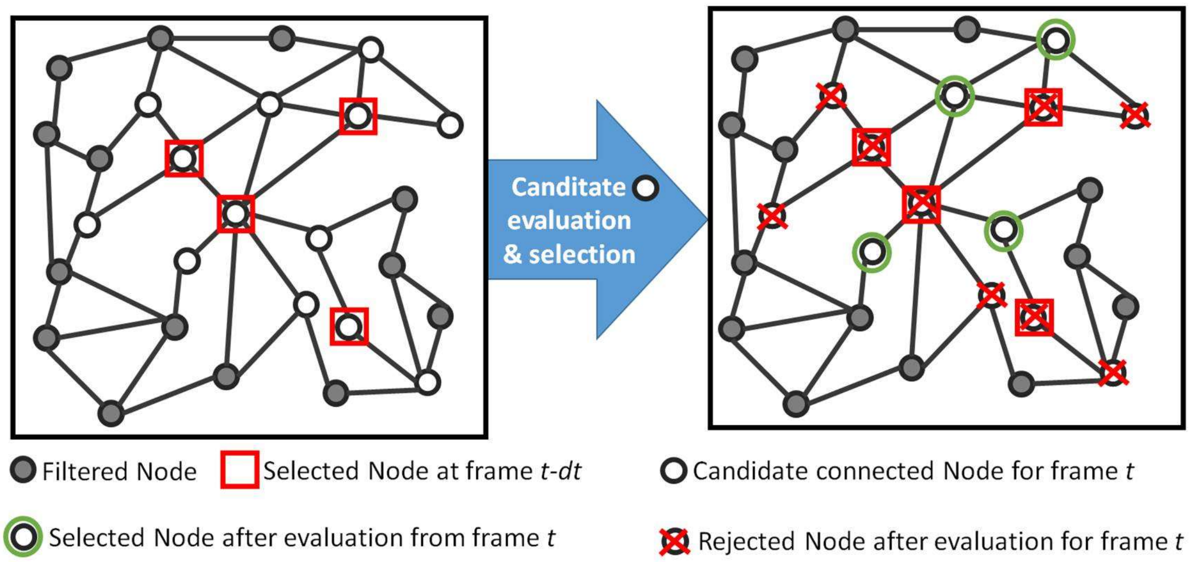

Figure 3: Candidate selection in the Filtered Pose Graph. Adapted from Plantard et al. (2017). Left: Filtered Pose Graph with selected nodes for the previous frame *t-dt* (squares) and all the connected nodes (blank circles) that can be candidates for the current frame *t*. Right: selection of the double circled nodes that best fit the current constraints at the frame *I (*rejected nodes are depicted with a cross).

Using S_{T+dt}, an optimization framework is used to search for the best combination to reconstruct the actual pose performed by the user. This is formulated as an energy minimization process with four energy terms: 1) similarity to reliable joints measured at frame T+dt by the Kinect, 2) ensure fixed bone length constraint, 3) style preservation compared to the closest neighbours in the database, and 4) minimize discontinuity with the previously corrected pose at frame T. The resulting optimized pose if filtered using a physical model, as recommended in (Shum et al., 2013).

Compared to a reference motion capture system, results demonstrated the relevance of such a correction process, especially when large errors occur. We evaluated the means, standard deviations (σ), min and max of the error between the reference (using a Vicon motion capture system) and the corrected Kinect joint positions in standardized scenarios with occlusions. With the correction, there were significantly fewer cases where the error was greater than or equal to 0.2m: 19.1+/-27.3% of error >0.2m for the Kinect alone, and 3.3+/-3.9% only after correction. Compared to other optimization-based correction approaches (Shum et al., 2013), our method also overperformed, supporting the hypothesis that the Filtered Pose Graph used to preselect pose candidates actually enhance the performance of the reconstruction method. We have also shown that the histogram of errors is significantly shifted on low values after correction compared to no-correction and previous works (Shum et al., 2013).

Evaluation in real work conditions

All the previous results have been carried-out in standardized laboratory condition. Even if occlusions have been introduced in this condition, real work condition is far more complex to deal with: bad sensor placement and significant increase of occlusions, due to the cluttered environment. Kinect has already been considered as a promising tool to evaluate ergonomics on-site, but only with simulated postures (Diego-Mas & Alcaide-Marzal, 2014), with very simple and inaccurate posture representation, and without any joint angles computation (Patrizi, Pennestrì & Valentini, 2015). Consequently, we carried-out a specific experiment in a real manufacturer (Faurecia [2]car industry) with 7 male workers,

[2] Faurecia: www.faurecia.com

in 5 different workstations (shown in

Figure 4).

Figure 4: Illustration of the 5 workstations assessed, from the Kinect point of view. Adapted from (Plantard, Shum, Le Pierres & Multon, 2017).

Finally, 22 motion capture sessions were performed in an assembly plant of car seats. The workers performed their routine work tasks, without any kind of perturbation: no wearable sensors, no calibration. We asked two ergonomists in Faurecia to carry-out a RULA assessment of these motions. The RULA scores computed using the corrected Kinect data were compared to those obtained by two human observers, similarly to previous works (Diego-Mas & Alcaide-Marzal, 2014). Instead of selecting the worst-case postures for RULA assessment as usual, the experts performed the RULA assessment with recorded Kinect colour sequences sampled at 0.2 Hz. A total of 300 different images were consequently assessed by the two experts. The experts independently assessed each body part required by the RULA method. The scores provided by the two experts may be slightly different due to inter-examiner variability. In such a case, the score returned by the method was assumed to be correct if it was in-between the results of the two experts. RMSE between the scores delivered by the experts and the method was calculated using the most different expert's score. We compared the

RULA scores computed with Kinect data to those obtained by the experts. Po and the strength of agreement on a sample-to-sample basis as expressed by unweighted Cohen's kappa (k) were computed, as proposed by (Diego-Mas & Alcaide-Marzal, 2014).

Results are reported in

	RMSE (RULA Score)	P_0	Kappa (k)
RULA Grand Score Right	0.59	0.73	0.60
RULA Grand Score Left	0.57	0.74	0.61
Score A Right (upper body)	0.67	0.71	0.55
Score A Left (upper body)	0.56	0.77	0.66
Score B (neck, trunk and legs)	0.84	0.62	0.46

Table 1. The agreement found for the RULA grand scores remain higher than 70%. The kappa index showed a strength of agreement from moderate to substantial according to the scale of (Landis & Koch, 1977).

	RMSE (RULA Score)	P_0	Kappa (k)
RULA Grand Score Right	0.59	0.73	0.60
RULA Grand Score Left	0.57	0.74	0.61
Score A Right (upper body)	0.67	0.71	0.55
Score A Left (upper body)	0.56	0.77	0.66
Score B (neck, trunk and legs)	0.84	0.62	0.46

Table 1: RMSE expressed in RULA score, P o and Cohen's kappa index, between RULA scores computed using the Kinect data and expert observations in real work conditions. Adapted from (Plantard, Shum, Le Pierres & Multon, 2017).

The results showed substantial agreement: the method correctly assessed the RULA grand score 73% and 74% for the right and left body part respectively. However, the reference data was provided by experts' evaluations, where posture could be difficult to be correctly assessed with a unique 2D picture. Let us recall that ergonomic experts used to have this limited information to perform their assessment. Further investigation with more objective reference motion capture system would be required to accurately quantify the relevance of the system (Patrizi et al., 2015). More details are given in (Plantard, Shum, Le Pierres & Multon, 2017).

As in previous studies (Dockrell et al., 2012), we also noticed inter-experts variability when assessing the same posture. In this work, our method estimated RULA scores within or very close to the range of values returned by the two experts. Involving more experts would lead to slightly increased variability, which would also lead to improved results with our method. Based on these results, we can conclude that the method could assist the ergonomists as another expert that could complement their observations at 30 Hz. Using such sampling compared to analysing sparse selected images offers many advantages: avoid subjective selection of key images, ensure reproducibility of the measurement, and provide new information such as the time spent above or with a given score, as shown in

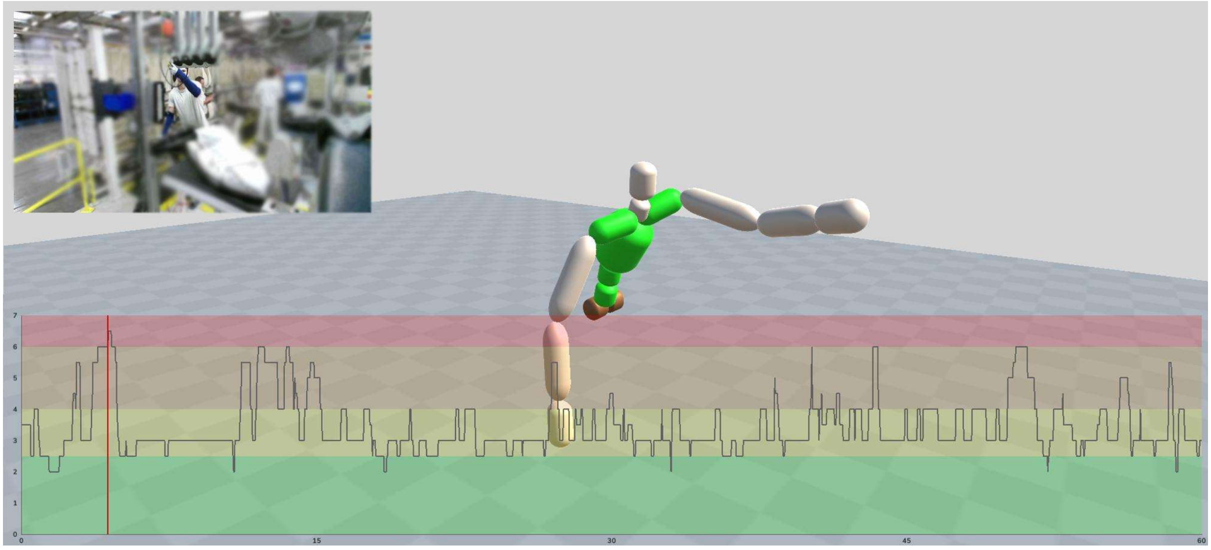

Figure 5.

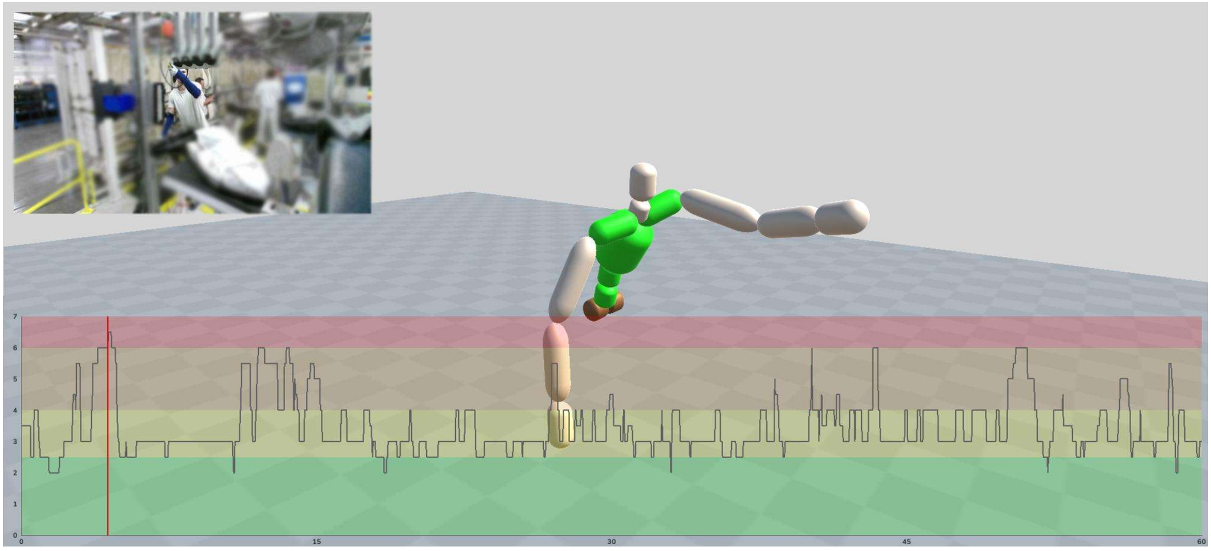

Figure 5: an example of the interface used to give feedback to the ergonomist, with 30Hz sampling RULA information. The left-upper part of the picture is an RGB image captured by the Kinect, leading to the reconstructed and corrected mannequin in the middle. The bottom is the continuous RULA score in colour bars depicting the RULA possible scores, from 1 to 7.

Physical Modelling of Human Motion Data

While the approach explained in the past sections can reconstruct human motion suffered from occlusion, it is based on a kinematics system with kinematics constraints and objective functions. In other words, the generated postures may or may not be physically correct. Artifacts such as inconsistent body part accelerations and overlapping of body parts may occur. In this section, we discuss approaches for further improving the posture reconstructed.

Early research in this area focuses on implementing filters or constraints to improve the naturalness of human motion generated by simulation systems or captured from motion capture systems. Constraints based on the zero moment point can be effectively applied to model and improve locomotion such as walking and running (Shin, Kovar & Gleicher, 2003). Using the source motion as a reference, by applying such constraints, an optimization process can be used to synthesize the physically correct version of the motion. Momentum can be used when editing and synthesizing high dynamic full-body motion that involves large body movement (Abe, Lui & Popović, 2004). Combining kinematics and physical constraints result in high-quality human motion synthesis (Tak & Ko, 2005). The multi-level optimization that considers different optimization criteria in each level improve the optimization process and avoid over-constraining the optimization system (Gall, Rosenhahn, Brox & Seidel, 2010).

Physically correct human motion can also be generated using a pure dynamic simulation approach. By optimizing for the physical parameters of a simple proportional-derivative (PD) controller, full body locomotion can be created (Yin, Loken & van de Panne, 2007; Shum, Komura, Shiratori & Takagi, 2010). Kinematic constraints such as the stepping location can be integrated in order to simulate more complex locomotion behaviour (Coros, Beaudoin, Yin & van de Panne, 2008). To simulate complicated movement that involves multiple points of contact with the floor such as rolling, a sample-based contact model is proposed to model the contact points in order to optimize for the dynamic movement (Liu et al., 2010). Since the optimization space is not smooth and optimization may easily get stuck, continuous optimization method can start with an easy version of the optimization problem and gradually increase the difficulty level is proposed (Yin, Coros, Beaudoin & van de Panne, 2008). On top of considering the skeletal dynamics, surface dynamics can be used to simulate the physically-correct deformable surface of a moving character (Liu, Yan & Guo, 2013). Due to the high dimensionality of the physical control problem, deep reinforcement learning is proposed to model the complex control space of the character movement (Naoya, Shum, Yang & Morishima, 2015; Peng, Abbeel, Levine & van de Panne, 2018).

We also implement a physical simulation engine to further optimize the reconstructed human motion and ensure physical correctness of the motion, allowing higher quality visualization of the movement. Instead of using internal control torque such as (Yin, Loken & van de Panne, 2007; Coros, Beaudoin, Yin & van de Panne, 2008) that is more suitable for robotics controls, we propose to utilize external control torque that helps to simplify the physical system (Shum & Ho, 2012). This essentially means that the control torque does not come from the ground fiction but is given by the system. During each time step, the system employs a PD controller to trace the reconstructed posture. One particularly important design here is that we control each body joint using 3-dimensional control force and 1-dimensional control torque. The forces are used to drag the body parts to the target location, while the torque is applied around the bone axis to control the orientation of the body part. We demonstrate that this setup can effectively resolve body parts penetration and inconsistent accelerations (Shum et al., 2013; Plantard, Shum & Multon, 2017). Such a system is applied as a post-process for better posture visualization and understanding.

Conclusion

In this chapter, we have shown that cheap depth cameras are promising means to capture the motion of workers in real working conditions. It offers motion capture facilities without requiring the workers to wear sensors or markers, and to calibrate the system regularly. Hence, workers are free to work naturally in a real production condition, without any impact on productivity.

However, it obviously has some limitations. Firstly, the field of view is limited and it cannot be used to follow a worker performing many displacements. Some works have proposed to use multiple depth cameras with a simple registration process to partly overcome this limitation (Auvinet et al., 2012). Secondly, as any video-based system, occlusions occur, leading to uncertainties in the measurements. We have shown that it is one of the major causes of errors. Several works have been proposed to overcome this limitation using data priors or new tracking approaches. However, these methods are limited when long occlusions occur. While it takes further research to formally resolve this, our proposed approach offers good results for many real cases. Thirdly, these systems cannot measure frequency and force adjustment classically used in assessment methods such as RULA. Fourthly, current pose estimation methods fail to measure the reliable motion of the head and the hands, which is key information in an ergonomic analysis. Further works are needed to enhance tracking or couple depth sensors with other devices.

Microsoft has stopped the production of Kinect in 2017, but there exist many other solutions to use cheap depth cameras with software toolbox for tracking human body, such as ORRBEC[3] ou Intel D435[4]. Hence, the research and exploitation of such a sensor in ergonomics remain very active and promising. As this sensor produces 30Hz measurements as other motion capture systems, it offers a cheap mean to explore real-time feedback to the user (Vignais et al., 2013) by using virtual reality or augmented reality. It opens new possibilities to train workers, and help to design industrial processes at the early stage of the design.

References

Auvinet, E., Meunier, J., & Multon, F. (2012). Multiple depth cameras calibration and body volume reconstruction for gait analysis. *International Conference on Information Science, Signal Processing and their Applications (ISSPA),* 478-483.

Auvinet, E., Multon, F, & Meunier, J. (2015). New Lower-Limb Gait Asymmetry Indices Based on a Depth Camera. *Sensors, 15*, 4605-4623.

Bonnechère, B., Jansen, B., Salvia, P., Bouzahouene, H., Omelina, L. et al. (2014). Determination of the precision and accuracy of morphological measurements using the Kinect sensor: comparison with standard stereophotogrammetry. *Ergonomics, 57(4),* 622-631.

Burdorf, A., Derksen, J., Naaktgeboren, B., & Van Riel, M. (1992). Measurement of trunk bending during work by direct observation and continuous measurement. *Applied Ergonomics, 23(4),* 263-267.

Clark, R.A., Pua, Y.H., Fortin, K., Ritchie C., Webster K.E. et al. (2012). Validity of the microsoft kinect for assessment of postural control. *Gait Posture, 36(3),* 372–377.

David, G.C. (2005) Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. *Occupationnal Medicine, 55(3)*, 190-199.

Diego-Mas, J.A. & Alcaide-Marzal, J. (2014). Using kinect sensor in observational methods for assessing postures at work. *Applied Ergonomics, 45(4),* 976–985.

Dockrell, S., O'Grady, E., Bennett, K., Mullarkey, C., Mc Connell, R. et al. (2012). An investigation of the reliability of Rapid Upper Limb Assessment (RULA) as a method of assessment of children's computing posture. *Applied Ergonomics, 43,* 632-636.

Dutta, T. (2012). Evaluation of the kinect sensor for 3-d kinematic measurement in the workplace. *Applied Ergonomics, 43(4),* 645–649.

Fagarasanu, M. and Kumar, S. (2000) Measurement instruments and data collection: a consideration of constructs and biases in ergonomics research. *Int. Journal of Industrial Ergonomics, 30,* 355-369.

Galna, B., Barry, G., Jackson, D., Mhiripiri, D., Olivier, P., & Rochester, L. (2014). Accuracy of the Microsoft kinect sensor for measuring movement in people with parkinson's disease. *Gait & Posture, 39(4),* 1062–1068.

[3] ORBBEC: orbbec3d.com

[4] Intel D435: click.intel.com/intelr-realsensetm-depth-camera-d435.html

Kurillo, G., Chen, A., Bajcsy, R., & Han, J.J. (2013). Evaluation of upper extremity reachable workspace using Kinect camera. *Technology and Health Care, 21(6),* 641-656.
Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33(1),* 159-174.
Li, G., & Buckle, P. (1999). Current techniques for assessing physical exposure to work-related musculoskeletal risks, with emphasis on posture-based methods. *Ergonomics, 42(5),* 674-695.
McAtamney, L. & Corlett, E.N. (1993). Rula: a survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics, 24(2),* 91–99.
Mentiplay, B.F., Hasanki, K., Perraton, L., Charlton, P.C, et al. (2018). Three-dimensional assessment of squats and drop jump using the Microsoft Xbox One Kinect: Reliability and validity. *Journal of Sports Sciences, 36(19),* 2202-2209.
Patrizi, A., Pennestrì, E., & Valentini, P.P. (2015). Comparison between low-cost marker-less and high-end marker-based motion capture systems for the computer-aided assessment of working ergonomics. *Ergonomics, 59(1),* 155-162.
Plantard, P., Auvinet, E., Le Pierres, A.S., & Multon, F. (2015). Pose estimation with a kinect for ergonomic studies: evaluation of the accuracy using a virtual mannequin. *Sensors, 15,* 1785–1803.
Plantard, P., Shum, H.P.H., & Multon, F. (2017). Filtered pose graph for efficient kinect pose Reconstruction. *Multimedia Tools Applications, 76,* 4291–4312.
Plantard, P., Shum, H.P.H., Le Pierres, A.S., & Multon, F. (2017). Validation of an ergonomic assessment method using Kinect data in real workplace conditions. *Applied Ergonomics, 65,* 562-569.
Shen, W., Deng, K., Bai, X., Leyvand, T., Guo, B., & Tu, Z. (2012). Exemplar-based human action pose correction and tagging*. In: Proceeding of the IEEE Computer Vision and Pattern Recognition,* 1784–1791.
Shen, W., Deng, K., Bai, X., Leyvand, T., Guo, B., & Tu, Z. (2014). Exemplar-based human action pose correction*. In: Proceeding of the IEEE Trans. Cybernetics.*
Shen W., Lei R., Zeng D., & Zhang Z. (2015). Regularity Guaranteed Human Pose Correction. In: Cremers D., Reid I., Saito H., & Yang M. H. (eds) Computer Vision -- ACCV 2014. ACCV 2014. Lecture Notes in Computer Science. 9004.
Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M. et al. (2011). Real-time human pose recognition in parts from single depth images. *In: Proceeding of the IEEE conference on computer vision and pattern recognition,* 1297–1304.
Shuai, L., Li, C., Guo, X., Prabhakaran, B., & Chai, J. (2017) Motion Capture with Ellipsoidal Skeleton using Multiple Depth Cameras. *In: Proceeding of the IEEE Transactions on Visualization and Computer Graphics, 23(2),* 1085-1098.
Shum H.P.H., Ho, E.S.L., Jiang, Y., & Takagi, S. (2013). Real-time posture reconstruction for microsoft kinect. *In: Proceeding of the IEEE Transactions on Cybernetics, 43(5)*, 1357–1369.
Springer, S., & Yogev Seligmann, G. (2016). Validity of the Kinect for Gait Assessment: A Focused Review. *Sensors, 16*, 194-207.
Vignais, N., Miezal, M., Bleser, G., Mura, K., Gorecky, D. & Marin, F. (2013). Innovative system for real-time ergonomic feedback in industrial manufacturing. *Applied Ergonomics, 44(4),* 566–574.
Wei, X., Zhang, P., & Chai, J. (2012). Accurate realtime full-body motion capture using a single depth camera. *ACM Transactions on Graphics, 31(6),* 188:1–188:12.
Xu, X., & McGorry, R.W. (2015). The validity of the first and second generation Microsoft Kinect™ for identifying joint center locations during static postures. *Applied Ergonomics, 49,* 47-54.

Ye, M., Wang, X., Yang, R., Ren, L., & Pollefeys, M. (2011). Accurate 3d pose estimation from a single depth image. *In: Proceeding of the International Conference Computer Vision.*
Shin, H.J., Kovar, L. Gleicher, M. (2003). Physical Touch-Up of Human Motions. *In: Proceeding of the Pacific Conference on Computer Graphics and Applications*, 194-203.
Tak, S. & Ko, H.S. (2005). A physically-based motion retargeting filter. *ACM Transactions on Graphics (TOG) 24(1),* 98-117.
Gall, J., Rosenhahn B., Brox, T. & Seidel, H.P. (2010). Optimization and filtering for human motion capture. International journal of computer vision 87, 75.
Abe, Y., Liu, K.C. & Popović, Z. (2004). Momentum-based parameterization of dynamic character motion. *In: Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation*, 173-182.
Yin, K.K., Loken, K. & van de Panne, M. (2007). Simbicon: Simple biped locomotion control. *In: ACM Transactions on Graphics (TOG), 26(3),* 105.
Shum, H.P.H., Komura, T., Shiratori, T. & Takagi, S. (2010). Physically-based Character Control in Low Dimensional Space. *In: Proceedings of the Third International Conference on Motion in Games,* 23-34.
Coros, S., Beaudoin, P., Yin, K.K. & van de Pann, M. (2008). Synthesis of constrained walking skills. *ACM Transactions on Graphics (TOG), 27(5),* 113.
Liu, L., Yin, K.K., van de Panne, M., Shao, T. & Xu, W. (2010). Sampling-based contact-rich motion control. *ACM Transactions on Graphics (TOG) 29(4),* 128.
Yin K.K., Coros, S., Beaudoin, P. & van de Panne, M. (2008). Continuation methods for adapting simulated skills. *ACM Transactions on Graphics (TOG), 27(3),* 81.
Liu, L., Yin K.K., Wang, B & Guo, B. (2013). Simulation and control of skeleton-driven soft body characters. *ACM Transactions on Graphics (TOG), 32(6),* 215.
Naoya, I., Shum H.P.H., Yang, L. & Morishima, S. (2015). Multi-layer Lattice Model for Real-Time Dynamic Character Deformation. *Computer Graphics Forum, 34(7),* 99-109.
Peng X.B., Abbeel, P., Levine, S. & van de Panne, M. (2018). DeepMimic: Example-Guided Deep Reinforcement Learning of Physics-Based Character Skills. *ACM Transactions on Graphics (TOG) 37(4),* 143.
Shum, H.P.H. & Ho, E.S.L. (2012). Real-time physical modelling of character movements with microsoft kinect. *In: Proceedings of the 18th ACM symposium on Virtual reality software and technology,* 17-24.