3D Human Body Reconstruction for Worker Ergonomic Posture Analysis with Monocular Video Camera

W. Chu[a], S.H. Han[a], X. Luo[b], and Z. Zhu[c]

[a] Department of Building, Civil and Environmental Engineering, Concordia University, Canada
[b] Department of Architecture and Civil Engineering, City University of Hong Kong, Hong Kong
[c] Department of Civil and Environmental Engineering, University of Wisconsin - Madison, USA
E-mail: monian0627@gmail.com, sanghyeok.han@concordia.ca, xiaowluo@cityu.edu.hk, and zzhu286@wisc.edu

Abstract –

In the modular construction industry of Canada, workers experience awkward postures and motions (reaching above shoulder, back bending backward, elbow/wrist flex, etc.) due to improper workstation designs. The awkward postures often lead to worker injuries and accidents, which do not only reduce the productivity but also increases the production cost. Therefore, the ergonomic posture analysis becomes essential to identify, mitigate and prevent the awkward postures of workers when workstation designs are changed. This paper proposes a novel framework to conduct the worker ergonomic posture analysis through the 3D reconstruction of human body from the video sequences captured by a monocular camera. The framework consists of four components: tracking worker of interest; detecting worker joints and body parts; refining 2D worker pose; and generating 3D human body model. The human body model generated from the framework could be used to estimate the joint angles of the workers to identify whether their postures meet the ergonomic requirements. The proposed framework has been tested on real construction videos, and the test results showed its effectiveness.

Keywords –

Joints detection; Body parts detection; 3D reconstruction; and Ergonomic posture analysis

1 Introduction

The modular construction has gained significant interest in recent years. According to the annual report from the modular building institute, the gross revenue in the modular construction industry in 2016 was roughly $3.3 billion in North America, which was increased by more than 60% from the year of 2015 [1]. Compared with the traditional, onsite construction, the factory-controlled processes in the modular construction provides the benefits of generating less material waste and reducing potential site disturbances [2]. They mitigate the adverse weather impacts on the project and faster the construction schedule [3]. Also, the factory controlled working environments are supposed to be safer for the workers involved.

However, workers' awkward postures are often noticed in several modular construction workshops [4]. These awkward postures might be due to the improper workstation designs in the factory controlled working environments. As shown in Figure 1, the workbenches are not high enough. Then, the workers have to bend their backs forward and strain their necks in order to reach materials and tools. The foot pedals in the machines are set too close. As a result, the workers have to bend their backs backward in order to reach the pedals. Sometimes, the workers are required to lift the materials from one spot to another over their shoulders, twist their wrists and elbows, and kneel or crouch to complete their assigned tasks in the production lines.

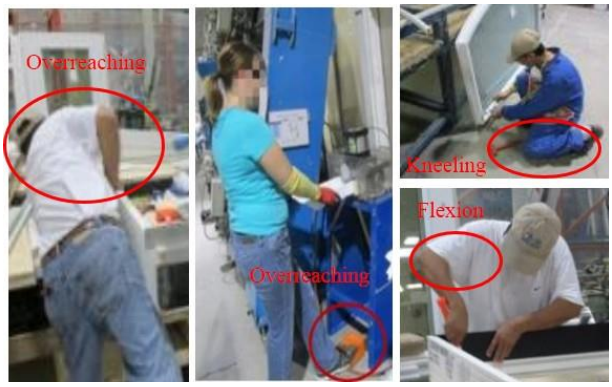

Figure 1. Awkward postures of the workers in modular production lines

The awkward postures easily lead to work-related musculoskeletal disorders [5, 6]. For example, the frequent kneeling will cause workers' pain and strain in their low backs and knees, which pose a high risk of developing muscle and joint problems. The musculoskeletal disorders hurt the health of the workers and result in the absenteeism [6]. Also, they impact the employers simultaneously. Additional time and efforts must be spent on handling the lost-time and disabling injury claims with high compensations; and the workflow in the production lines are delayed [7]. New workers need to be hired to replace the injured ones, which might not always be easy.

In order to reduce the occurrences of the work-related musculoskeletal disorders, the employers in the modular construction workshops are encouraged to conduct the Physical Demand Analysis (PDA) [8]. PDA is a systematic procedure to help the employers quantify and evaluate the physical and environmental demands of a job [9]. One important step in the Physical Demand Analysis is to measure the frequency of the body posture of a worker in a job, such as percentage of the worker's back forward or backward; and then identify any potential ergonomic risk for the worker from the measurements

Traditional measures heavily rely on direct manual observations and self-reporting [10], which are easy to implement with little costs associated in the workshops. However, the manual observations and self-reporting are subjective; and the measurement results are always error-prone [11]. Recently, the idea of attaching physical sensors or tags on the worker's body to record their motions, postures and even muscle activity in the work to indicate whether the muscle is fatigue [12, 13]. The sensors can provide the accurate measurements, but their implementation cost is high. Also, it is not widely acceptable by the workers in practice, who are not willing to wear these sensors and tags during the work [14].

An alternative solution is to use digital video cameras that could be set up in the workshops by the employers. This paper combined different computer vision techniques and proposed a novel framework that relies on the video from one monocular camera to reconstruct the 3D human body of a worker. The framework consists of four main components. First, the worker of interest in the video sequences is identified through the visual tracking. Then, the 2D joints and body parts of the worker are detected. The detected joints and body parts are combined to refine the worker's 2D pose in the video sequences. The 3D human body model is further generated by matching the model with the refined 2D pose in the video sequences. This way, the 3D posture and joint angles of the worker could be estimated for the corresponding ergonomic posture analysis.

The proposed framework has been implemented in Python 2.7 with the support of GPU (Graphic Processing Unit) computing. It was tested with the videos of two real working scenarios. The first video was recorded in the production line of the Fortis LGS Structures Inc., where a worker was cutting and transporting boards. The second one was provided by Alwasel et al. [15], where a worker was laying masonry units. The test results from both scenarios showed that the framework could generate the 3D human bodies of the workers of interest and obtain their joint angles effectively and efficiently. Moreover, the joint angle information could be further input to existing ergonomic posture analysis tools, such as 3D Static Strength Prediction Program (3DSSPP) [16], to identify whether the postures of the workers meet the ergonomic requirements.

2 Related Work

This section first provides a holistic view on the techniques available for ergonomic posture analysis and their limitations. Then, 2D and 3D human pose estimation methods in the field of computer vision are presented. The 2D pose estimation methods are reviewed, since they are the solid foundation to most of existing 3D pose generation methods.

2.1 Ergonomic Posture Analysis

Existing techniques available for ergonomic analysis can be classified into the categories of self-reporting, manual observation, sensor-based direct measurement, and vision-based analysis. Self-reporting is to collect the data from worker diaries, interviews, and web-based questionnaires [10] to conduct the ergonomic analysis. It is straightforward to implement in a wide range of workplaces and appropriate for surveying large numbers of workers at low cost. However, it was found that the self-reporting data were not always precise and/or reliable [11]. Also, the levels of comprehension and question interpretation may increase the difficulty, when adopting the self-reporting in practice [17].

Manual observation mainly relies on experienced experts to record the body postures of the workers in a workplace to conduct the ergonomic analysis. Several tools have been designed and developed to facilitate the observations and evaluation of ergonomic risk factors, such as Ovako Working Posture Analysing System (OWAS) [18]. The observation produces the minimal disturbances to the workers, which makes it applicable in various working environments. On the other hand, the manual observation results are error-prone due to the influence from the subjective judgement of the experts.

Sensor-based direct measurement is to complement or replace the self-reporting or manual observation. A wide range of direct measurement sensors have been developed, and they are directly attached to the workers to improve the measurement accuracy. For example, the Lumbar motion monitor (LMM) [12] was developed to assess the risk of the worker' low back injury. The electromyography (EMG) [13] was used to study the muscle exertions. Also, the retroreflective markers were attached on the worker bodies. This way, the 3D motion of the joints and body segments of the workers could be tracked with infrared cameras [19]. The measurements with sensors and/or markers are accurate and detailed. However, the workers complain about the physical requirement of attaching sensors and/or markers on the bodies, and not willing to wear them in practice [14].

The vision-based analysis tried to capture the joint motions of the workers and assess their body postures in a marker-less way. For example, Diego-Mas and Alcaide-Marzal [20] computerized the OWAS and processed the RGB-D data from a Microsoft Kinect camera to identify the risk level of each recorded posture. Ray and Teizer categorized the ergonomic or non-ergonomic body postures captured by a Kinect camera with a predefined set of rules [21]. Both methods solely focused on the classification of simple postures, such as lifting and crawling, in the indoor environments.

In addition to the Microsoft Kinect cameras, video cameras are also adopted. For example, in the method of Han and Lee [22], they extracted and matched the visual features of a worker in 2D video frames and then the worker's 3D skeleton can be extracted through the triangulation. This way, the unsafe actions could be detected by comparing the skeleton with pre-trained motion templates and skeleton models.

Compared with sensor-based direct measurement, the vision-based analysis does not have to physically tag workers, which makes it more acceptable in the workplaces. However, the vision-based analysis mainly relies on the data from the range or video cameras to approximate the joint motions of the workers. The accuracy and robustness of such approximation are always affected by environmental factors. Any illumination change, occlusions, and/or far shooting distance might lead to the vision-based analysis inaccurate and non-robust. So far, several methods were proposed to improve the accuracy and robustness of the vision-based ergonomic analysis.

2.2 2D Human Pose Detection

A classical method for 2D pose detection refers to the use of a pictorial structural (PS) model, in which the spatial relationships between various body parts are represented with kinematics priors. One example of the PS models is a tree-like structure, which was adopted by Lan and Huttenlocher [23] in their work of determining the human body pose. Andriluka et al. [24] combined the PS model with a strong human body part detector to make the human pose detection more generic. In addition, the mixture of the deformable parts model (DPM) was also introduced [25]. The introduction of DPM extended the application scope of the PS models, but it requires the substantial computations.

Recently, the 2D human pose detection has been significantly advanced with Deep Learning technologies. For example, DeepPose [26], the first method for human pose estimation with Deep Neural Networks (DNNs), was built on a 7-layered convolutional neural network (CNN). It formulated the pose detection as a joint regression problem and each joint could be directly regressed from a full image [26]. Pfister et al. [27] created the Flow ConvNets to detect 2D human poses, which benefitted from video temporal contexts to improve the pose estimation performance. Also, researchers developed the convolutional pose machines (CPM) [28] and stacked hourglass [29], both of which estimated the human pose without the need of an explicit human body model.

The methods described above have been used only for single-person pose estimation. They typically fail when multiple persons are captured into one image or video frame. This issue was overcome in the method of DeepCut [30], but the computational intensity is high. Its upgraded version, DeeperCut [31], was introduced to adapt to the newly proposed residual network for body part extraction. This way, the computation is reduced significantly and the robustness to the human body pose detection is maintained.

2.3 3D Human Pose Generation

The generation of 3D human pose from 2D images or videos is still one of the promising and popular research directions. Existing methods could be divided into two categories: i.e. multi-view vs. monocular view, depending on the number of video cameras adopted. Multi-view methods were inspired by human vision and infer a 3D human pose from two or more cameras [32]. The main mechanism is to obtain the 2D pose in each camera view first, and then reconstruct the 3D skeletal pose from the 2D poses [33].

Compared with the multi-view methods, it is much challenging to generate the 3D human pose with the monocular view methods. The methods tried to recover the depth information by creating a relationship between the 2D visual features (e.g. silhouette) and 3D skeletal pose [34]. This relationship could even ben learned with deep learning technologies, such as Vnect [35], which regressed 2D and 3D poses jointly through a CNN-based pose prior with Kinematic skeleton fitting. Moreover, Federica et al. [36] described how to automatically generate the 3D pose of a human body as well as its 3D shape from a single unconstrained image.

3 Research Objective

The monocular view methods for 3D human pose generation are supposed to achieve better performance in terms of the cost-effectiveness and wide-range applicability for ergonomic posture analysis. However, they have not been well studied yet. It is still necessary to improve the 3D pose generation accuracy before the methods could be adopted in practice. One important aspect for improvement is to locate the body joints in the images or videos more precisely, considering that the locations of the body joints are directly related to body angles calculation for the ergonomic analysis.

The main objective of this research is to investigate whether it is possible to improve the 3D pose generation accuracy with the integration of existing computer vision techniques. A novel framework is proposed here for the 3D reconstruction of human poses with one monocular video camera. The overview of the framework is illustrated in Figure 2. Under the framework, the worker of interest is first tracked visually in the video sequences, and represented as a rectangular bounding box. Then, the 2D joints and body parts of the worker are detected. Based on the detection results, the 2D pose of the worker is refined and the 3D human body model is further generated by matching the model with the refined 2D pose in the videos. The joint angles are computed based on the 3D joint coordinates to serve as the input for the ergonomic posture analysis. The proposed framework is expected to function in real modular construction scenarios, which could help to identify the awkward and improper postures of the workers in modular construction workshops.

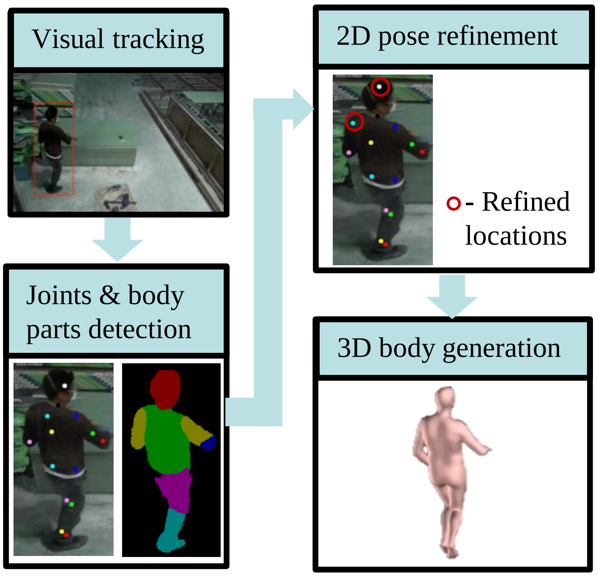

Figure 2. Overview of the proposed framework

4 Research Methodology

4.1 Visual Tracking

Visual tracking here is to locate the worker of interest along the consecutive video sequences. It could narrow down the image region with the less irrelevant background information contained for processing. Here, the CNN-based tracking algorithm *MDNet* [37] was selected due to its significant performance on the human-tracking challenge competition. It is worth noting that the tracking results need to be appropriately resized, since the accuracy of the pose estimation is easily affected by the image size. The size of 150 (width) x 350 (height) pixels was adopted in this study, based on the previous finding that the human pose detection could always perform well, when the standing height of the persons was scaled at around 340 pixels in the images [31].

4.2 Joints and Body Parts Detection

The resized visual tracking results are further processed to detect the worker’s joints and body parts. Here, the joint detection is conducted by the DeeperCut algorithm [31]. It could identify a total of 14 joints from their corresponding heat maps, where the probability of each pixel in the image region to be a joint is indicated.

As for the body parts detection, the Deeplab v2 method [38] is modified to have the method only detect 6 body parts, i.e. head, torso, upper /lower arm and upper/lower leg, instead of 24 detailed body parts shown in the Pascal-Person-Part dataset [39]. Moreover, the reliability of the body part detection results is evaluated with the heatmaps produced by the DeeperCut algorithm [31]. A detected body part is considered reliable only when it has a high probability of containing a joint (larger than 0.2 in this research study). For example, the detected head body part is reliable when its probability of containing the head joint is high. The high probability does not mean that it must include the head joint.

4.3 2D Pose Refinement

The joints detected in the previous step compose an initial pose for the 3D reconstruction later. However, this initial pose is not always accurate. This is mainly because each joint is not located perfectly at the joint detection stage. For example, it is highly possible that a left shoulder joint is located on the right shoulder area instead of the left one. Therefore, the initial pose from the joint detection needs to be refined by combining the joint and body part detection results.

The refinement first checks whether the initial joints lie in their corresponding body parts. If not, the joints are relocated in the body parts based on the confidence values of the body part regions in the heat maps. For example, if the initial head joint does not lie in the head part, then it is relocated in the head part region. The position where the highest confidence value to be the joint point in the head part area is selected. If the initial joints lie in their corresponding body parts, the refinement is then conducted on a case-by-case basis. For example, the head joint will be preferred in the top of the head part area. More details for the adjustment of the joints with the body part detection results could be found in the authors' recent work [40].

4.4 3D Human Body Reconstruction

Based on the refined 2D pose, the 3D human body of the worker of interest is reconstructed. Here, the generative Skinned Multi-Person Linear (SMPL) model [41] is adopted in the reconstruction process. The SMPL model could represent a wide variety of natural 3D human poses with body joints and shapes [41]. Following the workflow of Bogo et al. [36], these 3D poses are projected onto the camera view and compared with the refined 2D pose from the previous step. The one that optimally matches to the refined 2D pose is selected as the final reconstruction result.

The joint angles are further calculated from the reconstructed 3D human body. There are a total 14 joint angles under the consideration: clavicle (left and right), upper arm (left and right), lower arm (left and right), hand (left and right), upper leg (left and right), lower leg (left and right), and foot (left and right). Each joint angle is described both horizontally and vertically. These joint angles could be input into the 3DSSPP program [16] with other work-related information (e.g. external loads, worker's gender, age, height and weight), and assess the risk factors that may produce excessive physical loads on the worker's body.

5 Implementation and Results

5.1 Implementation and Tests

The proposed framework has been implemented in the Python 2.7 environment. It runs under the Ubuntu 16.04 LTS operation system and relies on the support of the GPU computing from an NVIDIA Titan Xp. Two real scenarios were selected to test the framework. In the first scenario, the video was collected from the production line of the Fortis LGS Structures Inc., where the worker was cutting boards in a sheathing table. The second scenario was provided by Alwasel et al. [15]. The worker in the scenario was laying concrete masonry units. He is also equipped with a motion capture suite with the attachment of 17 sensors to record his joint angles during the work.

5.2 Results

Figure 3 showed an example of the results from the first test scenario. Figure 3a illustrated the visual tracking of the worker of interest, where the red bounding box indicated the tracking result. Figure 3b indicated the detection of worker's body parts, which were represented with different colors. Figure 3c showed the refined locations of the 2D joints through the combination of the joints and body parts detection. Figure 3d was the final 3D human body reconstructed from the refined 2D joints.

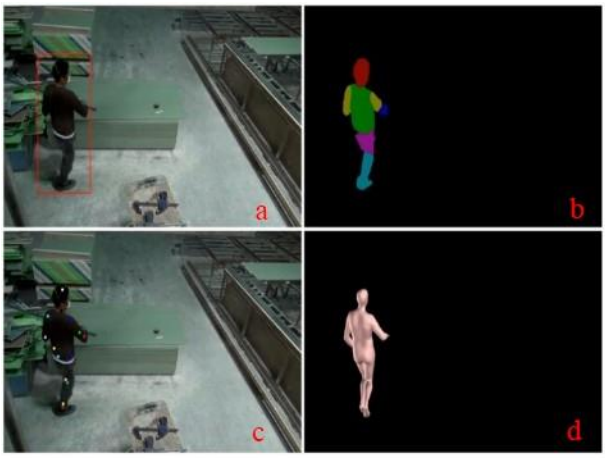

Figure 3. Results from the first test scenario

Figure 4 showed the examples of the results from the second test scenario. In the figure, the reconstructed 3D human body was placed beside the worker in the video sequences, in order to compare their visual similarity. It could be seen that the reconstructed 3D human body generally reflect the worker's posture during the work. Also, the human body could be generated, even when the worker experienced partial occlusions. In Figure 4c, both worker's head and arm were occluded, the human body was still reconstructed through the references from other visible 2D joints.

Moreover, the joint angles estimated from the 3D human body were compared with the sensory data from the motion capture suite in the second test scenario. The suit has the sampling rate of 125 Hz, and the video was captured at 25 frames per second. Therefore, the sensory data from the motion capture suite were down sampled for the frame-by-frame comparison.

Table 1 summarized the estimation error for each joint type. It could be seen that the minimum error (4.5°**)** occurs on measuring the horizontal angle of the lower arm joint. The maximum error (45.2°**)** occurs on measuring the horizontal angle of the upper leg joint. The errors for the remaining horizontal and vertical joint angles range from 10.0° to 28.0°**.** In average, the measurement error is around 17.5°**.**

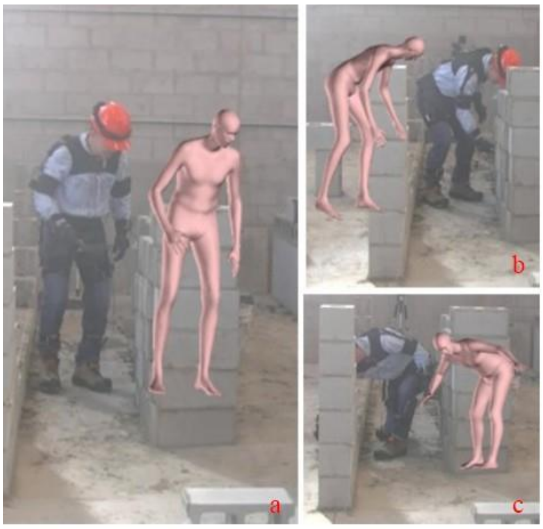

Figure 4. Results from the second test scenario

Table 1. Estimation error for each joint type

Joint Type	Horizontal Angle	Vertical Angle
Clavicle	11.7°	13.2°
Upper Arm	15.9°	10.0°
Lower Arm	4.5°	10.8°
Hand	14.2°	10.1°
Upper Leg	45.2°	20.2°
Lower Leg	14.2°	20.7°
Foot	19.4°	28.0°

6 Discussion

The errors for measuring the joint angles may come from several aspects. First, the joint definitions in the SMPL model and the measurement from the motion capture suites are not one-on-one matching. The SMPL model has defined 24 joints, while the motion capture suite measured a total of 28 joint data. Therefore, those close to the joint definitions in the SMPL model were selected for the comparison, which may introduce the measurement errors.

Also, the occlusions affect the joints and body parts detection, as well as the quality of the 3D human body generation. Therefore, they may increase the joint angle measurement errors. Figure 5 illustrated the frame-by-frame comparison of the horizontal angle measurements for the lower arm from the 3D human body and motion sensory data in the second test scenario. The worker's head and arm were severely occluded from the 234th to the 388th video frames. As a result, the difference of the joint angle measurement fluctuated significantly during the occlusion period, as shown in Figure 5.

Also, the 2D pose refinement played an important role on the quality of the 3D human body reconstruction. In order to highlight its effectiveness, the pose similarity is calculated and compared for the 3D human bodies generated with and without the refinement step. It was found that the refinement improved the horizontal and vertical measurements of the pose similarity by 7.0% and 2.1%.

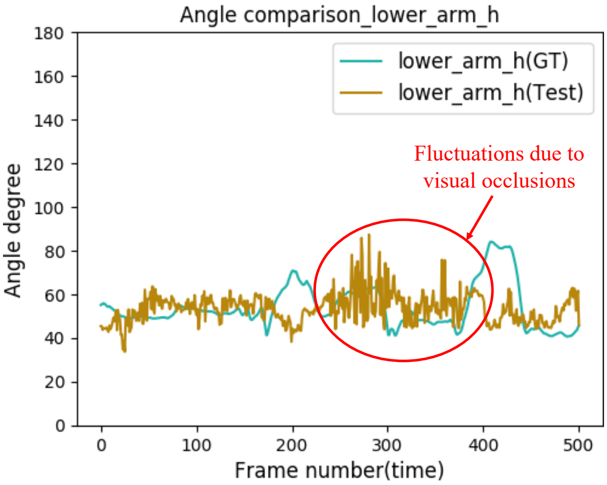

Figure 5. Comparison of the horizontal joint angle measurements for lower arm

7 Conclusions and Future Work

This paper presented an integrated framework for the 3D reconstruction of human body from the videos captured by monocular video cameras. The framework includes four main components: visual tracking, joints and body parts detection, 2D pose refinement, and 3D human body generation. The 3D human body generated from the framework could be used to estimate the body joint angles. This information could be input into the ergonomic posture analysis tool, 3DSSPP, to evaluate the risk factors that the worker may experience in a job.

The proposed framework was tested in two real scenarios. The joint angles of the 3D human body from the framework were also compared with the sensory data directly captured by a motion capture suite. The comparison results indicated that the average error for measuring joint angles was around 17.5°. The error for measuring the horizontal angle of the lower arm was as low as 4.5° and the error for measuring the horizontal angle of the upper leg reached up to 45.2°.

Future work will focus on reducing the errors of joint angle measurements. The temporal continuity in consecutive video frames and the worker's silhouettes in the video frames will be considered. They may improve the accuracy of joints and body parts detection and 3D human body reconstruction in the framework. This way, the joint angle measurement from the 3D human body could be more accurate.

Acknowledgement

This research study was financially supported by the National Science and Engineering Research Council of Canada (NSERC) and the NVIDIA GPU Grant Program. Any opinions, findings, and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies. Also, the authors would like to thank the support of Dr. C. Haas and Dr. A. Alwasel from the University of Waterloo, Dr. S. Lee from the University of Michigan and their research teams. They kindly provided the videos and the sensory data as ground truth in the second test scenarios.

References

[1] Modular Building Institute, Permanent Modular Construction: Annual Report, Online: http://www.modular.org/documents/document_publication/mbi_sage_pmc_2017_reduced.pdf, Accessed: 04/ 2018.

[2] InnovativeModular Solutions, The sustainability of modular construction: reduce. Online: https://blog.innovativemodular.com/sustainability-modular-construction-reduce, Accessed: 01/2019.

[3] Modular building insitute, What is modular construction? Online: http://www.modular.org/htmlpage.aspx?name=why_modular Accessed: 01/2019.

[4] Inyang, N.I., A framework for ergonomic assessment of residetnal construction tasks. *Ph.D. Thesis*, University of Alberta, 2013.

[5] Canadian Centre for Occupational Health and Safety, Worker-related Musculoskeletal Disorders. Online: https://www.ccohs.ca/oshanswers/diseases/rmirsi.html Accessed: 01/2019

[6] Simoneau, S., St-Vincent, M., and Chicoine, D. Worker-related musculoskeltal discorders – a better understanding for more effective prevention. Online: https://www.irsst.qc.ca/media/documents/PubIRSST/RG-126-ang.pdf Accessed: 01/2019

[7] Hinze, J. and Appelgate, L.L. Costs of construction injuries." *Journal of construction engineering and management*, 117(3): 537-550.

[8] Li, X., Fan, G., Abudan, A., Sukkarieh, M., Inyang, N., Gül, M., El-Rich, M., and Al-Hussein, M. Ergonimics and physcial demand analysis in a construction manufacturing facility. In: *Proceedings of 5th International/11th Construction Special Conference*, Vancouver, British Columbia, June 810, 2015, 231-1:10

[9] Occupational Health Clinics for Ontario Workers Inc. Physical demands analysis (PDA). Online: https://www.ohcow.on.ca/ Accessed 01/2019

[10] Dane, D., Feuerstein, M., Huang, GD., Dimberg, L, Ali, D., and Lincoln, A. "Measurement properties of a self-report index of ergonomic exposures for use in an office work environment." *Journal of Occupational and Environmental Medicine*, 2002, 44(1): 73-81.

[11] David, GC. Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. *Occupational medicine*, 2005, 55(3): 190-199.

[12] Marras, WS. and Granata, KP. Spine loading during trunk lateral bending motions. *Journal of biomechanics*, 1997, 30(7): 697-703.

[13] Ning X., Zhou J., Dai B., and Jaridi, M. The assessment of material handling strategies in dealing with sudden loading: the effects of load handling position on trunk biomechanics. *Applied ergonomics*, 2014, 45(6): 1399-1405.

[14] Yu, Y., Li, H., Yang, X., and Umer, W. Estimating construction workers' physical workload by fusing computer vision and smart insole technologies. In: *Proceedings of the 35th International Symposium on Automation and Robotics in Construction*, Berlin, Germany, July 20-25, 2018, paper-247.

[15] Alwasel, A., Sabet, A, Nahangi, M., Haas, C.T., and Abdel-Rahman, E. Identifying poses of safe and productive masons using machine learning. *Automation in Construction*, 84(2017): 345-355.

[16] Center for Ergonomics, University of Michigan, 3DSSPP Software, Online: https://c4e.engin.umich.edu/tools-services/3dsspp-software/ Accessed: 01/2019

[17] Spielholz, P., Silverstein, B., Morgan, M., Checkoway, H., and Kaufman, J. Comparison of self-report, video observation and direct measurement methods for upper extremity musculoskeletal disorder physical risk factors. *Ergonomics*, 2001, 44(6): 588-613.

[18] Karhu O, Kansi P, and Kuorinka I. Correcting working postures in industry: a practical method for analysis. *Applied ergonomics*, 1977, 8(4): 199-201.

[19] Richards, J.G. The measurement of human motion: A comparison of commercially available systems. *Human movement science*, 1999, 18(5): 589-602.

[20] Diego-Mas. J.A, and Alcaide-Marzal J. Using Kinect™ sensor in observational methods for assessing postures at work. *Applied ergonomics*, 2014, 45(4): 976-985.

[21] Ray S.J, and Teizer J. Real-time construction worker posture analysis for ergonomics training. *Advanced Engineering Informatics*, 2012, 26(2):

439-455.

[22] Han, S.U, and Lee, S.H. A vision-based motion capture and recognition framework for behavior-based safety management. *Automation in Construction*, 35(2013): 131-141.

[23] Lan, X. and Huttenlocher, D.P. Beyond trees: Common-factor models for 2d human pose recovery, In: *Proceedings of the 10th International Conference on Computer Vision*, IEEE, 2005, 1: 470-477.

[24] Andriluka, M, Roth, S., and Schiele, B. Pictorial structures revisited: People detection and articulated pose estimation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009.: 1014-1021.

[25] Felzenszwalb, P.F, Girshick, R.B., McAllester, D., and Ramanan, D., Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence*, 2010, 32(9): 1627-1645.

[26] Toshev, A., and Szegedy, C. Deeppose: Human pose estimation via deep neural networks In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014: 1653-1660.

[27] Pfister, T., Charles, J., and Zisserman, A. Flowing convnets for human pose estimation in videos. In: *Proceedings of the IEEE International Conference on Computer Vision*. 2015: 1913-1921.

[28] Wei, S.E., Ramakrishna, V., Kanade, T, and Sheikh, Y. Convolutional pose machines. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016: 4724-4732.

[29] Newell, A., Yang, K., and Deng, J. Stacked hourglass networks for human pose estimation. In: *Proceedings of the European Conference on Computer Vision. Springer*, Cham, 2016: 483-499.

[30] Pishchulin, L., Insafutdinov, E., Tang S, Andres, B., Andriluka, M., Gehler, P., and Schiele, B. Deepcut: Joint subset partition and labeling for multi person pose estimation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016: 4929-4937.

[31] Insafutdinov, E, Pishchulin, L., Andres, B., Andriluka, M., and Schiele, B. Deepercut: A deeper, stronger, and faster multi-person pose estimation model. In: *Proceedings of the European Conference on Computer Vision*. 2016: 34-50.

[32] Trucco, E, and Verri, A. *Introductory techniques for 3-D computer vision*. Prentice Hall, 1998.

[33] Hofmann, M., and Gavrila, D.M. Multi-view 3d human pose estimation combining single-frame recovery, temporal integration and model adaptation. In: *Proceedings of the IEEE Computer Vision and Pattern Recognition*, 2009, 2214-2221.

[34] Atrevi, D.F., Vivet, D., Duculty, F., Emile, B. A very simple framework for 3D human poses estimation using a single 2D image: comparison of geometric moments descriptors. *Pattern Recognition*, 2017, 71: 389-401.

[35] Mehta, D, Sridhar, S, Sotnychenko, O, Rhodin, H., Shafiei, M., Seide, H.P., Xu, W., Casas, D., and Theobalt, C. Vnect: Real-time 3d human pose estimation with a single rgb camera. *ACM Transactions on Graphics*, 2017, 36(4): 44.

[36] Bogo, F., Kanazawa, A., Lassner, C., Gelher, P., Romero, J., and Black, M.J. Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image. In: Proceedings of *European Conference on Computer Vision*. 2016: 561-578.

[37] Nam, H. and Han, B. Learning multi-domain convolutional neural networks for visual tracking. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016: 4293-4302.

[38] Chen, L.C, Papandreou, G., Kokkinos I, Murphy, K., and Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 2018, 40(4): 834-848.

[39] Chen, X., Mottaghi, R., Liu, X., Fidler, S., Urtasum, R. and Yuille, A. Detect what you can: Detecting and representing objects using holistic models and body parts. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014: 1971-1978.

[40] Chu, W. Han, S.H., and Zhu, Z. Devlopment of Human Pose using Hybrid Motion Tracking System. In: Proceedings of the 18th International Conference on Construction Applications of Virtual Reality, Nov. 22-23, 2018, Auckland, New Zealand

[41] Loper, M., Mahmood, N, Romero J, Pons-Moll, G., and Black, M.. SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics*, 2015, 34(6): 248.