Computer-Vision Based Rapid Entire Body Analysis (REBA) Estimation

Chao FAN[1], Qipei MEI[2], Qiuling YANG[3], Xinming LI[4*]

[1] *Ph.D. Student, Department of Mechanical Engineering, University of Alberta*
[2] *Assistant Professor, Department of Civil & Environmental Engineering, University of Alberta*
[3] *Postdoctoral Fellow, Department of Mechanical Engineering, University of Alberta*
[4] *Assistant Professor, Department of Mechanical Engineering, University of Alberta*
[*] *Corresponding Author's E-mail: xinming.li@ualberta.ca*

ABSTRACT
Although much attention has been paid to the safety risk of construction sites and ergonomic risk assessment of workers, the automation of ergonomic risk assessment has not been significantly developed. This article presents a non-intrusive, automated ergonomic risk assessment approach based on computer vision, machine learning, and Rapid Entire Body Assessment (REBA). The method is called Computer-Vison Based Rapid Entire Body Analysis Estimation (CVRE). This approach is expected to realize automated monitoring and early-stage warning of ergonomic risks by automating the procedure of calculating REBA scores for construction site workers. This method consists of machine learning-based key joints and joint angles estimation of human bodies and computer-vision-based automated risk estimation. With the extensive development of machine learning and computer vision, researchers have been paying attention to assessing ergonomic risks with machine learning techniques. The proposed method has been further validated using the experimental data obtained by a motion capture system.

KEYWORDS
Construction safety; Computer vision; Ergonomic risks assessment; Machine learning; REBA

INTRODUCTION
Ergonomic risk assessment aims to identify potential ergonomic risks when workers carry out functional or physical tasks and then mitigate the risk accordingly. REBA is a widely used and effective ergonomic risk rapid detection method (Li et al., 2017, 2019). REBA relies on information about the body's key joints and the angles of the multiple bones. Therefore, after the traditional REBA has been widely used in the past few years, many equipment-based methods of human joint localization and bone angle measurement have been used to assess ergonomic risks (Antwi-Afari et al., 2020; de Freitas et al., 2019; Muñoz et al., 2020; Nath et al., 2018). Wearable sensors and motion capture systems are the most widely used devices for acquiring human movement data, including bone-joint angles. These high-precision sensors are typically composed of an accelerometer, a gyroscope, and a magnetometer (Poitras et al., 2019). Body motion data can be obtained by fusing the data collected by these sensors. Although wearable sensors are portable, they are still intrusive. Having wearable sensors on workers' bodies can hamper their ability to perform physical or functional tasks. Working with sensors on their bodies for an extended period

can make them feel more fatigued. Until wearable sensor technology has been developed to the level that the intrusion becomes minimal or negligible, workers on construction sites won't be able to use them daily (Ceseracciu et al., 2014).

On the other hand, motion capture systems can avoid intrusion and interference to workers when collecting motion data. In terms of accuracy, motion capture systems are regarded as the best body-motion-data acquisition method (Ceseracciu et al., 2014). However, motion capture systems have a lot more strict requirements on the data acquisition environment, and most of these systems are set up in controlled environments such as laboratories (Poitras et al., 2019).

Computer vision is a sub-area of artificial intelligence that enables computers to extract meaningful information from digital videos and images. This study proposes an automated ergonomic risk assessment framework based on REBA and computer vision. If the framework can achieve the desired results in a single-person image or video application scenario, it will achieve the same results in a multi-person application scenario. The system applies 3D joint estimation and ergonomic risk assessment to every worker in the same image separately. In other words, CVRE treats detected people as people in different photos. Hence, validation for single-person scenarios is sufficient to prove if the approach will work for multi-worker scenarios or not. The proposed framework consists of three modules. The first module is an image or video-based human body key joints estimation module, and it is capable of estimating the 3D poses of Multiple people in the same image. The second module is a data processing module based on the joint angle requirement of REBA score calculations. The third module is the ergonomic risk assessment module utilizing REBA.

METHODS

The CVRE is designed to automate the ergonomic risk assessment of workers in video frames obtained from monocular cameras deployed on construction sites. The system is designed in a modular fashion based on the functionalities needed. The human body key joints estimation module detects every human's presence in a video frame with YOLOv5. It provides bounding boxes of those detected humans, then the output of YOLOv5 is fed into 3D Multi-person Pose Estimation (3DMPPE) PoseNet (Moon et al., 2019) for human joints estimation. The estimated coordinates of those detected human joints are processed before propagating to the last module since the joint coordinates obtained from PoseNet contain irrelevant data for REBA calculations. Thus, a data processing module is required. The ergonomic risks assessment module takes the processed data and performs joint angle calculations for every worker detected. The results were compared to corresponding joint angle data from a motion capture system, specifically a three-dimensional Vicon motion analysis system. The accuracy of our proposed approach was calculated by comparing the estimation of angles between bones to the ground-truth values extracted from the Vicon motion analysis system.

Data collection

Three-dimensional coordinates of key joints were acquired with the Vicon motion capture system. Eight Vicon motion capture cameras were installed in a laboratory with approximately 30 cubic meters of capture volume, and two video cameras were also connected to the Vicon motion analysis system. The motion capture cameras were placed about 2.2 meters above the ground, and the motion capture system's reference origin was placed on the ground at the center of the capture space. The sampling rate of the Vicon motion capture system was set to 100 Hz, while the video camera captured only 50 frames per second at the same time. Synchronous capturing of markers' trajectories and pictures of subjects ensures the input consistency for joint angle comparison in the later stage. Because motion capture systems are considered the gold standard for motion quantification in terms of accuracy, marker trajectory data from eight motion capture cameras were used to extract joint angles. The proposed method took the video frames from the video cameras connected to the Vicon motion analysis system. It carried out ergonomic risk assessments for the subjects, including estimating their joint angles. Thirty-eight retro-reflective markers based on the Plugin Gait model are placed on the represented subjects. The motion capture system acquired the joint angles by sampling signals reflected from those markers. The markers were twelve millimeters in diameter. Because the Plugin Gait model does not directly fit the needs of our proposed approach, post-processing is necessary to convert the Plugin Gait data further to be compared with the CVRE.

Human body key-joints estimation

PoseNet of 3D Multi-person Pose Estimation is one of the state-of-the-art human key points inference algorithms (Moon et al., 2019). Workers' key joints were estimated using PoseNet. Mainly, PoseNet uses the output from human detection algorithms as its input. We chose YOLOv5, one of the state-of-the-art algorithms of its kind, as the human detection algorithm and fed the output data to the PoseNet module. The PoseNet is based on its joint estimation neural network, and it is suitable for a variety of publicly available annotated human joint data sets. The human body key joints estimation module has also been tested on various images of workers on construction sites. The results were not ideal because the photos were from cameras too far from the workers for the human detection model to work (Fang, Ding, et al., 2020; Fang, Love, et al., 2020; Kim et al., 2021). Hence, improvement needs to be made. Before application to construction sites, neural network models for YOLOv5 and PoseNet need to be trained with construction site datasets or simulated datasets. Figure 1 shows some graphical output of this module (Vicon, 2022).

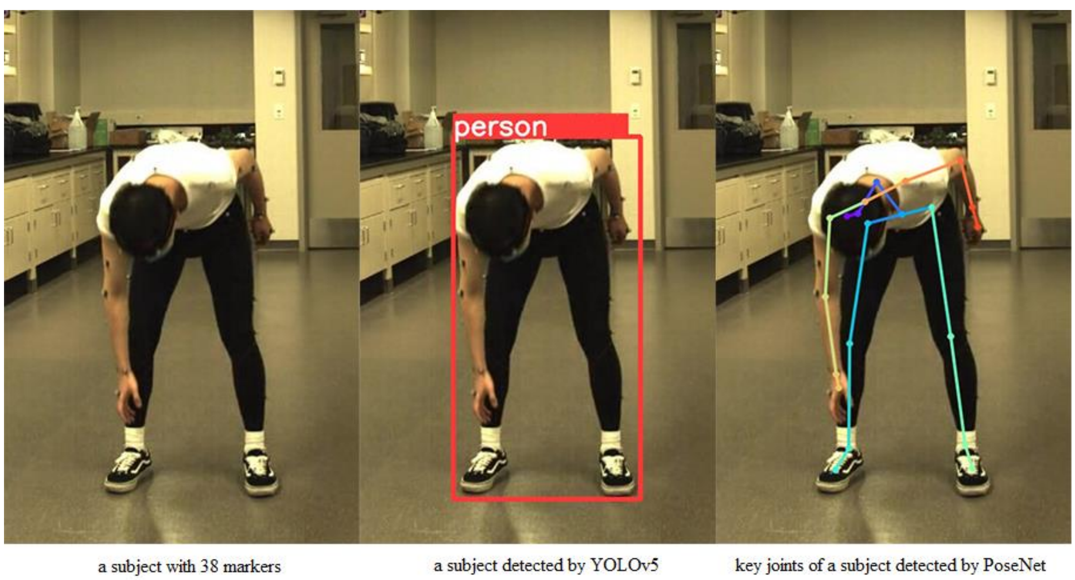

Figure 1. a subject with 38 retro-reflective markers was detected by YOLOv5 and then processed by PoseNet

Data Processing
The PoseNet estimation of the human joints from video frames contained some superfluous data to our system. The processed data is the minimum needed for angle calculation in the ergonomic risk assessment module.

Ergonomic risk assessment
The ergonomic risk assessment module estimates ergonomic risk levels for every worker detected in video frames. Ergonomic risk levels are obtained from this module by first calculating the skeletal angles based on the processed data from the data processing module and then by using these angles to estimate the REBA score (Chu et al., 2020). Thus, real-time ergonomic risk assessment is realized for each detected worker. The CVRE requires sixteen 3D joint coordinates, and they are head, nose, left shoulder, left elbow, left wrist, left hand, right shoulder, right elbow, right wrist, right hand, left hip, left knee, left ankle, right hip, right knee, and right ankle.

Validation
According to the information available, there is some research related to automatic single-person REBA estimations (Chu et al., 2020; Martinez et al., 2017; MassirisFernández et al., 2020; Pavlakos et al., 2017; Seo et al., 2016). Thus, we fill the gap by proposing a multi-human ergonomics risk assessment system that quantifies movement in monocular camera frames and utilizes computer vision and REBA. Vicon system data was obtained to compare the highly accurate motion capture system data to the data obtained with our system. For the Vicon motion capture system to have the same data samples corresponding to the video camera frames, joint data captured by the Vicon system was used every two samples. The video cameras capture 50 frames per second, while the Vicon motion capture cameras have a 100 Hz sampling rate. The Vicon system tracked the thirty-eight retro-reflective markers attached to the subjects. Because the markers can only be attached to the body surface of the subjects, the trajectories of the markers captured by Vicon do not coincide with the trajectories of the subjects' joints. Thus, the CVRE included an algorithm to estimate subjects' 3D joint coordinates required by REBA using the marker coordinates obtained from the Vicon Plugin gait model. The algorithm takes the Vicon joint coordinates to calculate their joint coordinates in the CVRE configuration. For example, Vicon head coordinates in CVRE configuration could be obtained using coordinates of the right forehead and left-back head markers in Vicon. The 3D joint coordinates calculated by this algorithm and the 3D joint coordinates estimated by our approach were fed to the same skeletal angle calculation algorithms in CVRE. The CVRE was validated by comparing its skeletal angles with the same angles obtained from the Vicon motion capture system. Difference and standard deviation (SD) of corresponding skeletal angles were obtained. Since REBA scores are calculated based on ranges of skeletal angles, they don't represent the accuracy of the skeletal angle estimations. Therefore, the skeletal angles were compared instead of REBA scores in the validation process. Figures 2, 3, and 4 are reconstructed 3D skeletons from the Vicon motion capture system and CVRE data using the ergonomic risk assessment module.

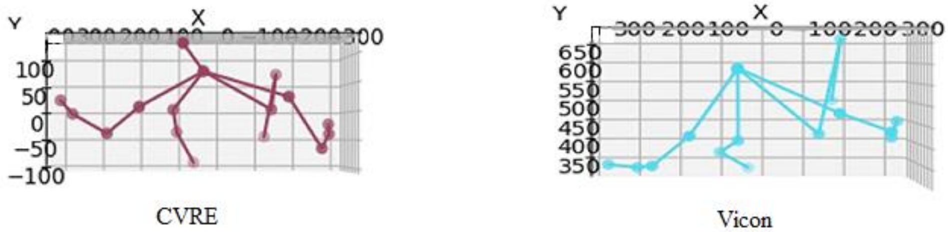

Figure 2. top view of reconstructed skeleton for a simulated construction task

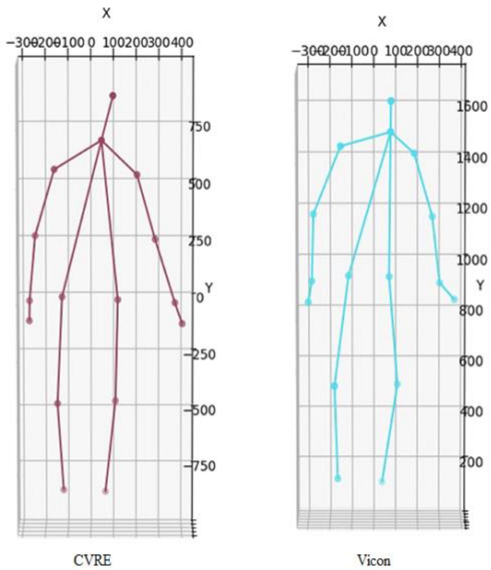

Figure 3. front view of reconstructed skeleton for a simulated construction task

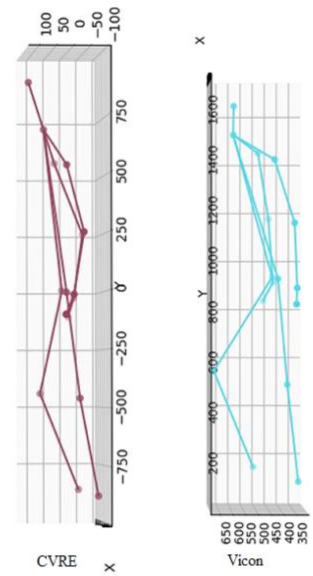

Figure 4. side view of reconstructed skeleton for a simulated construction task

RESULTS AND DISCUSSION

A total of 8 experiment trials involving 500 frames were extracted from the video clips recording simulated tasks, and the tasks were actual construction tasks observed in construction site videos. The experiments were two trials of every task, including walking, dragging, bending, and hammering. The skeletal angles of the subjects in these 500 frames were all estimated by CVRE. The actual skeleton angles of the experiment subjects were obtained using the same algorithm applied to the marker trajectories. The result of the neck, trunk, upper arm, lower arm, leg, and wrist angles between bones from both CVRE and Vicon motion analysis systems were compared. Figure 5 shows the neck and trunk angles from CVRE and Vicon motion analysis systems. As it shows in the figure, the skeletal neck angle and trunk angle from Vicon in 85 consecutive frames follow some periodic patterns. However, some frames have skeletal angles not following the pattern. Those random angles are caused by gap-filling in the Vicon motion analysis system or marker slipping due to subjects' movement. There is hardly perfect gap-filling because filled gaps between adjacent frames are predictions, not actual values. Although the angle values obtained by CVRE are different from those of the Vicon motion analysis system, they follow a similar pattern as the pattern of angles obtained by Vicon. In addition, the neck and trunk angles of each frame obtained by CVRE were only a few degrees higher than those obtained by Vicon.

The difference between joint angles obtained by CVRE and Vicon was calculated for all the frames. Then, the mean and SD of the differences were calculated for every simulated task. Table 1 shows the mean and SD of neck angles, trunk angles, upper arm angles, lower arm angles, leg angles, and wrist angles for each simulated task. These results indicate that the differences between the joint angles obtained by CVRE and Vicon are less than ten degrees for most skeletal angles. The wrist angles are not close to those obtained from the Vicon motion analysis system. They demonstrate that skeletal angles of small-size body segments might not be detected at all. The skeletal angles of large body segments estimated by our method are close to their actual values.

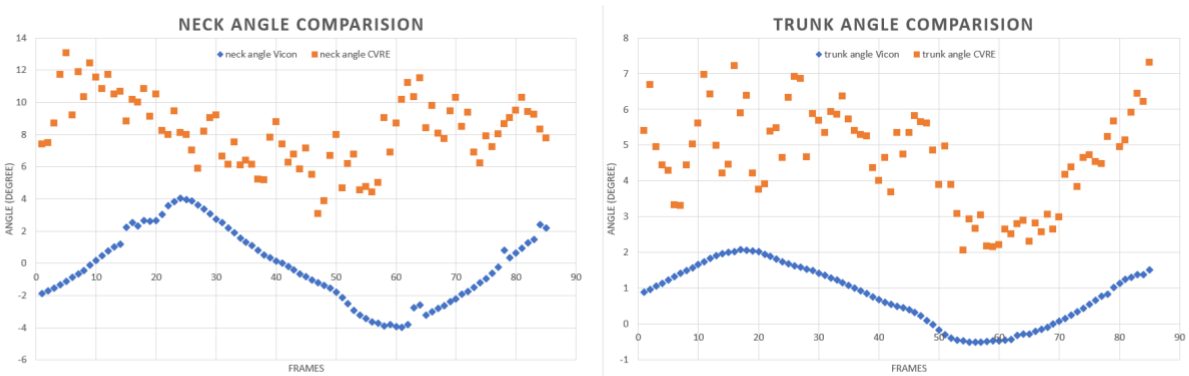

Figure 5. Neck angle and Trunk angle of 85 frames for a large object dragging task

Given by Figure 1 to Figure 4 and Table 1, it is evident that the skeletal angles estimated by our approach have differences that need to be reduced. The main reason for these errors is that our method relies on artificial neural networks, which require extensive and accurate s network models trained with enough and accurately labeled data that will have an extremely low probability of fault diagnosis. The proposed approach utilizes a neural network model trained with two publicly available human key points detection datasets, Human36M and MPII. Wrist angle is not compared due to the limitation of the Human36M dataset; the dataset does not include any key point annotation for hands. Hence, future work could focus on producing a comprehensive dataset with enough motion quantification information for ergonomic risk assessment.

Furthermore, our neural network model can be further trained with the dataset needed for REBA score estimations. In addition, the accuracy of CVRE could also be further improved by having a more accurate Vicon-CVRE coordinates conversion algorithm. A custom Vicon human key joints model corresponding to CVRE joints will be developed for future studies. The Vicon motion capture system can give more accurate joint coordinates in CVRE configuration by utilizing the custom human key joints model. Hence, the differences between CVRE and Vicon motion capture system coordinates can be reduced.

Table 1. Mean and standard deviation of differences between CVRE and Vicon angles for four simulated construction tasks

Simulated task	Neck angle		Trunk angle		Leg angle		Upper arm angle		Lower arm angle	
(degree)	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
dragging	8.38	2.92	3.84	0.99	7.44	7.60	10.22	5.00	4.99	3.44
walking	6.53	2.26	2.77	0.91	7.21	7.31	8.65	3.23	5.06	3.13
bending	8.27	2.70	4.76	1.12	8.21	7.44	7.69	3.77	5.21	3.41
hammering	6.71	3.01	3.55	1.07	7.84	6.81	8.63	4.36	4.77	4.18

CONCLUSION

A computer vision and REBA-based ergonomic risk assessment framework is developed. The framework could be used for construction site workers and workers from various industries. We demonstrated that the approach has significant accuracy in human key joint coordinates estimation. The proposed framework has three advantages over traditional human body movement quantification methods. Firstly, it's minimally intrusive to the workers because it neither requires workers to wear sensors nor requires special clothing with markers. Secondly, very little equipment is needed, and even the less expensive ordinary monocular camera can implement the functions of this framework. Finally, different from some existing methods of ergonomic risk assessment based on computer vision, it can simultaneously assess the ergonomic risks of multiple people in the same image frame. Most importantly, the proposed method can maintain the high accuracy of ergonomic risk assessment while possessing these advantages.

The employment of this system is expected to reduce work-related musculoskeletal disorders and other ergonomic risks and provide timely feedback and information on any modifications or other immediate actions that may be required. Also, the system will improve workers' productivity by giving them feedbacks to reduce fatigue (Konstantinidis et al., 2021). For example, a worker could be notified when carrying out a task with a high REBA score for 3 minutes.

However, this research still has limitations. Since the Human36M does not include annotations for hands, the neural network model cannot estimate wrist angles. Thus, the wrist angle reconstructed from CVRE's results in Figure 1 is almost zero, while the Vicon motion capture system shows the wrist angle is more than 40 degrees. Other limitations include limited trial subjects, a limited number of experimental trials, and the experiment was not on a real construction site with a corresponding camera setup. In addition, the accuracy of CVRE could be improved by building a custom Vicon motion capture model for CVRE joints configuration and then comparing the coordinates with the coordinates obtained with CVRE.

ACKNOWLEDGMENTS

This study was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) through Alliance Grant with Alberta Innovates (File No. ALLRP 561120 - 20).

REFERENCES

Antwi-Afari, M. F., Li, H., Umer, W., Yu, Y., & Xing, X. (2020). Construction activity recognition and ergonomic risk assessment using a wearable insole pressure system. Journal of Construction Engineering and Management, 146(7), 04020077.

Ceseracciu, E., Sawacha, Z., & Cobelli, C. (2014). Comparison of markerless and marker-based motion capture technologies through simultaneous data collection during gait: proof of concept. PloS one, 9(3), e87640.

Chu, W., Han, S., Luo, X., & Zhu, Z. (2020). Monocular vision–based framework for biomechanical analysis or ergonomic posture assessment in modular construction. Journal of Computing in Civil Engineering, 34(4), 04020018.

de Freitas, P. V. A., Mendes, P. R. C., Busson, A. J. G., Guedes, Á. L. V., Silva, G. L. F. D., de Paiva, A. C., & Colcher, S. (2019, October). An ergonomic evaluation method using a mobile depth sensor and pose estimation. In Proceedings of the 25th Brazillian Symposium on Multimedia and the Web (pp. 445-452).

Fang, W., Ding, L., Love, P. E., Luo, H., Li, H., Pena-Mora, F., ... & Zhou, C. (2020). Computer vision applications in construction safety assurance. Automation in Construction, 110, 103013.

Fang, W., Love, P. E., Luo, H., & Ding, L. (2020). Computer vision for behaviour-based safety in construction: A review and future directions. Advanced Engineering Informatics, 43, 100980.

Kim, W., Sung, J., Saakes, D., Huang, C., & Xiong, S. (2021). Ergonomic postural assessment using a new open-source human pose estimation technology (OpenPose). International Journal of Industrial Ergonomics, 84, 103164.

Konstantinidis, D., Dimitropoulos, K., & Daras, P. (2021, June). Towards Real-Time Generalized Ergonomic Risk Assessment for the Prevention of Musculoskeletal Disorders. In The 14th PErvasive Technologies Related to Assistive Environments Conference (pp. 466-472).

Li, X., Han, S., Gül, M., & Al-Hussein, M. (2019). Automated post-3D visualization ergonomic analysis system for rapid workplace design in modular construction. Automation in Construction, 98, 160-174.

Li, X., Han, S., Gül, M., & Al-Hussein, M. (2017). Automated Ergonomic Risk Assessment based on 3D Visualization. In ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction (Vol. 34). IAARC Publications.

MassirisFernández, M., Fernández, J. Á., Bajo, J. M., & Delrieux, C. A. (2020). Ergonomic risk assessment based on computer vision and machine learning. Computers & Industrial Engineering, 149, 106816.

Martinez, J., Hossain, R., Romero, J., & Little, J. J. (2017). A simple yet effective baseline for 3d human pose estimation. In Proceedings of the IEEE international conference on computer vision (pp. 2640-2649).

Moon, G., Chang, J. Y., & Lee, K. M. (2019). Camera distance-aware top-down approach for 3d multi-person pose estimation from a single rgb image. In Proceedings of the ieee/cvf international conference on computer vision (pp. 10133-10142).

Muñoz, A., Martí, A., Mahiques, X., Gracia, L., Solanes, J. E., & Tornero, J. (2020). Camera 3D positioning mixed reality-based interface to improve worker safety, ergonomics and productivity. CIRP Journal of Manufacturing Science and Technology, 28, 24-37.

Nath, N. D., Chaspari, T., & Behzadan, A. H. (2018). Automated ergonomic risk monitoring using body-mounted sensors and machine learning. Advanced Engineering Informatics, 38, 514-526.

Pavlakos, G., Zhou, X., Derpanis, K. G., & Daniilidis, K. (2017). Coarse-to-fine volumetric prediction for single-image 3D human pose. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7025-7034).

Poitras, I., Dupuis, F., Bielmann, M., Campeau-Lecours, A., Mercier, C., Bouyer, L. J., & Roy, J. S. (2019). Validity and reliability of wearable sensors for joint angle estimation: A systematic review. Sensors, 19(7), 1555.

Seo, J., Yin, K., & Lee, S. (2016, January). Automated postural ergonomic assessment using a computer vision-based posture classification. In Construction research congress 2016 (pp. 809-818).

Vicon, (2022, March 10). Vero: Compact super wide camera by Vicon. Vicon. Retrieved March 17, 2022, from https://www.vicon.com/hardware/cameras/vero/