# Evaluation of Upper Body Postural Assessment of Forklift Driving using a Single Depth Camera

Veeresh ELANGO [a,b,1], Simona PETRAVIC [b] and Lars HANSON [a,b]
[a] *University of Skövde, School of Engineering Science, 541 28 Skövde, Sweden*
[b] *Scania CV, Södertälje, Sweden*

## Abstract

Observational postural assessment methods which are commonly used in industry are time consuming and have issues of inter- and intra-rater reliability. Computer vision (CV) based methods have been proposed, but they have mainly been tested inside lab environments. This study aims to develop and evaluate an upper body postural assessment system in a real industry environment using a single depth camera and OpenPose for the task of forklift driving. The results were compared with XSens, an Inertial Measurement Unit (IMU) based system. Data from three forklift drivers performing seven indoor and outdoor tasks were recorded with a depth camera and XSens sensors. The data were then analyzed with OpenPose with additional custom processing. The angles calculated by the computer vision system showed small errors compared to the XSens system and generally followed the trend of the XSens system joint angle values. However, the results after applying ergonomic thresholds were vastly different and the two systems rarely agreed. These findings suggest that the CV system needs further study to improve the robustness on self-occlusion and angle calculations. Also, XSens needs further study to assess its consistency and reliability in industrial environments.

**Keywords:** upper body postural assessment, forklift driving, depth camera, OpenPose.

## Introduction

Occupational ergonomics is a discipline which attempts to adapt the job to the worker with the goal of promoting worker health, safety, comfort, and productivity (Frederick et al., 1984). The promotion and maintenance of good occupational ergonomics reduces the risk of chronic injuries and illnesses, and consequently decreases the cost incurred by lost productivity, sick leave, and medical care (Kok et al., 2020). Forklift driving in industry is one of the work tasks linked to increased risk of upper body problems (Flodin et al., 2018; Viruet et al., 2008). The risk factors are commonly identified and managed using observational methods based on direct on-site observation or watching recorded videos of workers performing their usual tasks. Examples of such methods for postural assessment are RULA (McAtamney

& Nigel Corlett, 1993), REBA (Hignett & McAtamney, 2000), and RAMP (Lind et al., 2014). Apart from requiring a time-consuming manual analysis, the major disadvantage of observational ergonomic assessments is that they are heavily subjective which leads to considerable intra- and inter-rater variability (Burdorf et al., 1992; Fagarasanu & Kumar, 2002).

Industry 4.0 has led to the incorporation of motion capture (MoCap) solutions in industrial applications for improving processes and productivity, with a focus on workers' health and safety (Menolotto et al., 2020). In line with that, low-cost MoCap systems based on computer vision (CV) have been researched to solve the aforementioned posture assessment problems. One such MoCap solution is OpenPose, a real-time multi-person 2D pose estimation system, which has been used for postural assessments with both 2D (Lin et al., 2022) and 3D approaches (Kim et al., 2021). These studies have shown that OpenPose has acceptable accuracy for posture assessment with joint angle errors of 8.0° and 8.4° respectively.

Previous studies on CV solutions for postural assessments were mainly constrained to lab simulations. Factors such as spatial constraints, limited field of view, (self-)occlusions, and external interference (light conditions, vibrations) could deteriorate the performance of the aforementioned solutions (Menolotto et al., 2020), and these factors are often present in industry environments such as forklift driving. To the best of the authors' knowledge, no studies have been performed to evaluate postural assessment in forklift driving in industry using a CV system.

Therefore, the aim of this paper is to develop and evaluate an upper body postural assessment system for forklift driving in a real industry environment using a single depth camera and OpenPose.

## Methods

### *Computer Vision System Setup*

A single depth camera (Intel RealSense D455) was rotated vertically and positioned in front of the driver at an angle of 45° above their head, to obtain the best view on the operator's upper body posture, as shown in Figure 1. Forklift driving has low possibility of external occlusions between the driver and camera. To calculate all angles needed for upper body posture analysis including neck and trunk twist, 3D representation of the posture is required. Calibration of multiple cameras is complex and time-consuming, and the setup requires space which is limited inside a forklift. Hence, a single depth camera was selected for the study.

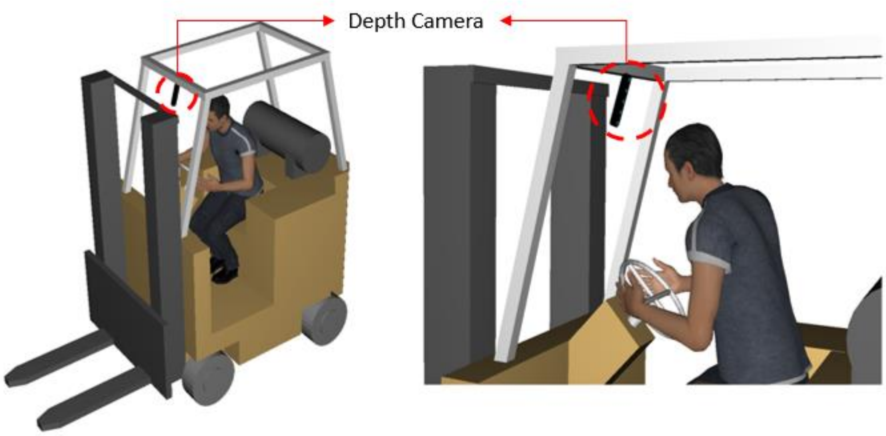


Figure 1. Depth camera placement inside forklift.

To capture data from the depth camera, a Python script was developed which records RGB images and depth maps at the rate of 15 frames per second with resolution of 640 x 360 (width x height) pixels. The depth maps were aligned with the RGB images using methods in the Pyrealsense2 library, an opensource library developed and released by the camera manufacturer. The depth data may contain some missing values where data is not available. This may happen because of occlusions between the two depth sensors or between the depth and color stream which do not see exactly the same image, under- or over-exposure, objects being too close to the camera (less than the minimum depth range), etc. In order to correct missing depth data in the resulting depth images, a hole filling filter was applied which uses the value from the neighboring pixel closest to the sensor. Python 3.8 environment was used to execute the code.

*Inertial Measurement Unit (IMU) System Setup*

XSens MTw Awinda, an IMU-based MoCap system, was chosen as the reference system for this study because it has shown acceptable performance for ergonomic analysis (Kim et al., 2020). 11 inertial sensors were used to record joint angles from the upper part of the body as shown in Figure 2. XSens MVN Analyze Pro 2021.2 software was used for calibration and data collection.

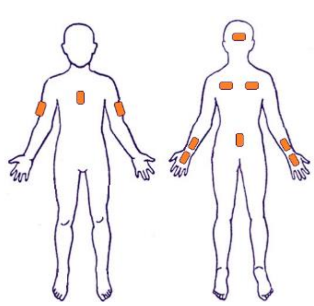

Figure 2. Placement of XSens sensors on the upper part of the body.

### *Participants*

One female and two male forklift operators participated in the study. Their average age was 30 ± 11, their average height was 171 ± 4 cm, and they have been working as a forklift operator for 4 ± 1 years on average. The participants took part in the study voluntarily and they gave consent for processing and using the recorded photos and videos for research purposes. The forklifts used by the participants in the experiments were Linde E30, E25, and E20.

### Ergonomic Assessment Method

The ergonomic assessment thresholds selected for this study were the upper body joint angle thresholds from RAMP II (Lind et al., 2019). This method considers the duration of the postures such that the longer the duration, the higher the risk for the operator. The resolution of the method is at the level of seconds.

### *Experimental Procedure*

The forklift operator wore an XSens T-shirt and Velcro straps on the upper part of the body where 11 inertial measurement units were attached in accordance with the manufacturer's instructions (Figure 2). The XSens calibration for initializing the setup was performed using the XSens MVN software set up in vehicle mode. The data was recorded at a frequency of 60 Hz. Calibration was repeated until the estimated quality was 'good'.

Like the XSens system, the CV system needed to be calibrated to obtain the baseline position of the forklift driver. Thus, the participant was asked to sit inside the forklift and remain static in their usual sitting posture with hands placed on the thighs and looking straight ahead for 30 seconds. The laptops connected to both the systems and the XSens Awinda's recording and docking station were placed inside the forklift for the duration of the task performed.

### *Dataset*

The three forklift drivers performed seven different forklift driving tasks within the logistics area of the automotive industry, both in indoor and outdoor conditions. In total, 65.17 minutes of data were processed. The XSens and CV system data were split into seven parts corresponding to each task.

### *Data Processing*

The CV system dataset consisted of RGB images and their corresponding depth maps. Both were rotated to counteract the rotated position of the camera, and then the RGB images were processed with OpenPose to obtain the 2D keypoints of eyes, nose, ears, shoulders, neck, elbow, wrists, hips, knees and feet of the driver. The resulting 2D keypoints and the corresponding depth values from the depth maps were deprojected into the camera coordinate system using the intrinsic camera parameters to obtain 3D keypoints.

For the calibration of the CV system, the coordinates of the hips and the offset values for three joint angles (head flexion and extension, trunk flexion and extension, and trunk lateral bending) were calculated from the 3D keypoints. The offset values were set as the average of all angle values calculated from 15 frames selected from the collected calibration data. The reason for this is that these joint angles, by definition, are non-zero in the baseline position. Regarding the hips, they are prone to occlusions due to arm movement while steering the vehicle but remain mostly static during driving. Hence, the hip coordinates were calculated at the beginning and fixed throughout the task.

After calibration, the joint angles were calculated similar to other publications (Kim et al., 2021; Van Crombrugge et al., 2022) for the remaining frames using the fixed hip coordinates, followed by the application of the previously calculated offset values. The joint angle values from XSens were exported from the MVN Analyze Pro software. The XSens C1 Head Flexion/Extension angle was also offset to counterbalance the non-zero angle in the baseline position. The joint angles and the corresponding XSens angles for comparison are defined in Table 1.

Table 1. Joint angle equations, were $v\_(i,j)$ is a vector pointing from joint $i$ to joint $j$, $vertical$ is a line pointing up (opposite of the $y$-axis), and $proj(v\_(i,j),n)$ is an orthogonal projection of vector $m$ on $n$.

| Joint Angle | Equation | XSens Angle |
|---|---|---|
| Head lateral bending | $angle(proj(v\_(left\ eye, right\ eye), v\_(neck, left\ eye) \times v\_(neck,\ right\ eye)), proj(v\_(left\ shoulder, right\ shoulder), v\_(neck, left\ eye) \times v\_(neck,\ right\ eye)))$ | C1 Head Lateral Bending |
| Head flexion/ extension | $angle(v\_(midhip, neck), v\_(neck, nose)) - offset$ | C1 Head Flexion/ Extension + 10° |
| Head twist | $90° - angle(proj(v\_(neck, nose), xz - plane), proj(v\_(left\ shoulder, right\ shoulder), xz - plane))$ | C1 Head Axial Rotation |
| Trunk lateral bending | $angle(proj(v\_(midhip, neck), xy - plane), vertical) - offset$ | Vertical_T8 Lateral Bending |
| Trunk flexion/ extension | $angle(v\_(midhip, neck), vertical) - offset$ | Vertical_T8 Flexion/ Extension |

| | | |
|---|---|---|
| Trunk twist | $angle(proj(v\_(left\ shoulder, right\ shoulder), xz - plane), proj(v\_(left\ hip, right\ hip), xz - plane))$ | Vertical_T8 Axial Bending |

The 2D keypoints of the shoulders were prone to self-occlusions by the head while performing twisting or bending motions. A self-occlusion correction script was created to assess if the head occluding a shoulder based on the depth of the shoulders, nose, and ears. The occlusion was corrected with the depth of the other (non-occluded) shoulder. In addition, most of the time the head was occluding the neck keypoint because of the perspective from the top, hence it was calculated as the midpoint of the shoulders. The outliers which were further than 3 standard deviations from the median of each angle were removed and Gaussian smoothing with variance of 2 was applied on the resulting joint angles. The XSens data was down sampled to 15 samples per second using Fourier method and synchronized with the CV data by manual inspection.

The CV system results may contain missing values. This can happen because the CV system could not detect some of the keypoints needed for angle calculation or because the depth camera could not estimate the depth at some of the required joint locations. These missing value frames (MVF) were removed, and they were not used for further comparison.

To apply RAMP for upper body ergonomic assessment, the average of each joint angle was calculated for every 15 frames on both CV and XSens data. Then, the RAMP thresholds were applied to generate binary data for each joint angle which states if the average angle, in that second is above or below the threshold. To evaluate the performance after applying RAMP thresholds, the precision, recall and specificity (Hossin & Sulaiman, 2015) were calculated as mentioned in Table 2.

Table 2. Evaluation Metrics

| Evaluation Metric | Formula |
|---|---|
| Precision | $TP/(TP + FP)$ |
| Recall | $TP/(TP + FN)$ |
| Specificity | $TN/(TN + FP)$ |

Note: TP-True Positive, FP-False Positive, TN-True Negative

## Results

The root mean square error (RMSE) was chosen to assess the closeness of the joint angles between XSens and the CV system and it was calculated for every frame per task. The distribution of RMSE of each joint angle across tasks is shown in Figure 3. The overall mean of RMSE of all joint angles was 7.1 ± 3.2 °.

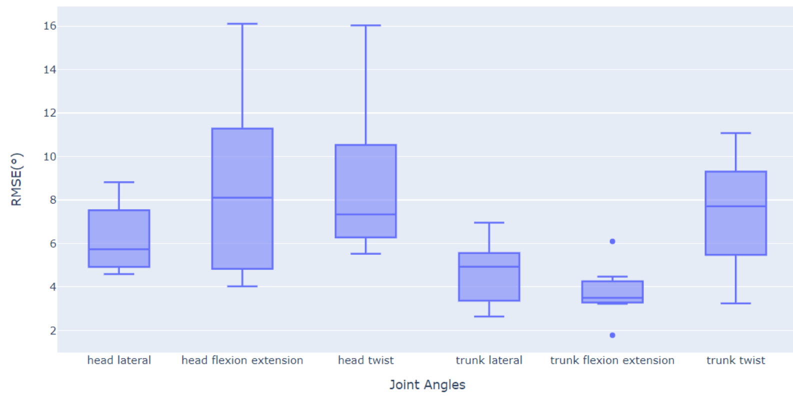


Figure 3. Distribution of RMSE (°) of joint angles between XSens and CV results across tasks.

In addition, for every joint angle, the ratio of MVF and total frames of the task was calculated and averaged across tasks as shown in Table 3.

Table 3. Frames with missing values as a percentage of the total number of frames.

| Joint Angle | Mean (%) | Standard Deviation (%) |
|---|---|---|
| Head lateral | 13.27 | 11.48 |
| Head flexion/extension | 3.55 | 2.49 |
| Head twist | 5.06 | 3.61 |
| Trunk lateral | 0.72 | 1.08 |
| Trunk flexion/extension | 0.72 | 1.08 |
| Trunk twist | 2.40 | 1.52 |

Table 4. Comparison of CV and XSens system after applying RAMP thresholds.

| Joint Angle | Precision | Recall | Specificity | Mean Absolute Difference (°) of FN CV and Threshold | FN Count (secs) |
|---|---|---|---|---|---|
| Head lateral | 0.00 | 0.00 | 0.99 | 8.6 ± 1.2 | 674 |
| Head flexion | 0.00 | 0.00 | 0.92 | 18.0 ± 2.5 | 3 |
| Head extension | 0.50 | 0.24 | 1.00 | 5.3 ± 3.9 | 38 |
| Head twist | 0.27 | 0.75 | 0.96 | 7.1 ± 5.9 | 20 |
| Trunk lateral - moderate | 0.39 | 0.70 | 0.85 | 3.6 ± 2.8 | 139 |
| Trunk lateral - considerable | | 0.00 | 1.00 | 14.2 ± 0.8 | 2 |
| Trunk flexion - moderate | 0.82 | 0.61 | 1.00 | 6.6 ± 3.5 | 30 |
| Trunk flexion - considerable | | | 1.00 | | 0 |
| Trunk extension | | 0.00 | 1.00 | 7.5 ± 7.5 | 24 |
| Trunk twist | | | 1.00 | | 0 |

The postures where the angle is greater than the RAMP thresholds are considered bad postures. The seconds where the XSens found bad postures and the CV system missed are false negatives (FN). These FN seconds were investigated to understand how close the CV system angles are to the threshold as shown in Table 4.

## Discussion and Conclusions

The CV system was robust to dynamic light conditions and the mean RMSE of all joint angles across different tasks is in line with earlier studies (Kim et al., 2021; Lin et al., 2022). Figure 3 shows that in comparison with other joint angles, the head flexion/extension angle had higher variance across tasks. The angle difference in two tasks suddenly increased after a period of time, which might be caused by a change of hip position. Also, on isolated occasions the head flexion angle difference was unusually high, which needs further inspection as it is inconsistent with the general behavior. Furthermore, the head twist formula estimates higher twist angles when the nose is moving along with other motions, which causes overestimation in specific cases. In similar terms, the back twist is estimated from the shoulders which are prone to more variations due to arm movements compared to the XSens sensors placed directly on the back. The results in Table 3 show that the angle with most MVF is head lateral bending. This can be attributed to at least one of the eyes often going out of the camera view when performing head twist or flexion.

From the binary data generated after applying the RAMP thresholds summarized in Table 4, it can be seen that the system can identify many bad postures for head twist and trunk lateral – moderate, but it also falsely classifies many postures as bad due to overestimation and aforementioned issues. Identification of bad postures is more precise in trunk flexion – moderate, but many bad postures were also missed. This is due to self-occlusion of the shoulders which directly affects the neck keypoint and results in underestimation of the angle after correcting the occlusions. Head flexion and extension are affected in a similar way. In the case of head lateral bending, the system performs poorly as the calculated angles are overall very small compared to XSens. Trunk lateral – considerable, suffered from underestimation due to the operator going out of the camera field of view, where the CV system predicted the neck and shoulder keypoints closer to the center of the image and resulted in classification as moderate instead of considerable lateral bending. For trunk extension, apart from one task, the RMSE between the systems was low, indicating that the CV values were close to XSens, but just under the threshold. No bad postures were identified by both systems in the cases of trunk flexion - considerable and twist.

Table 4 also shows that in most of the FN seconds the difference between the CV system and the respective RAMP angle threshold was close to the average error of the system. This indicates that the CV system would find more of the bad postures identified by XSens if the thresholds applied to the CV data are adjusted using methods like receiver operating characteristic (ROC) curve (Metz, 1978). However, a suitable balance between the precision and recall has to be found along with ergonomists. This behavior can be expected not only when using RAMP, but also any other ergonomic method that applies hard angle thresholds.

Also, issues with the XSens data were identified which brings up the question of its validity in certain moments. For example, calibration needed several attempts to achieve good quality similar to (Zelck et al., 2022). Even then, visual inspection showed issues where the XSens representation of the body did not correspond to the operator’s static pose. This also existed in the recordings where certain postures recorded by XSens were incorrect which was identified by visual inspection. Lastly, sudden changes in angle and glitches were discovered with no apparent motion performed by the operator that can be seen in the video. However, these problems were not noticed during simulations in the lab. The reason might be magnetic disturbances in the industry environment, vibrations, or motions that caused the sensors to move from their original position.

Based on the findings discussed above, this study was unable to confirm the applicability of the proposed CV system directly for RAMP ergonomic assessment in industry, even though previous studies from lab environments suggested otherwise (Kim et al., 2021; Lin et al., 2022; Massiris Fernández et al., 2020). Even though the CV system closely followed the trend of the XSens angles, the performance of the CV system after applying RAMP thresholds resulted in vastly different outcomes because of self-occlusion, camera perspective and field of view, angle calculation methods and the hard angle thresholds. However, there were inconsistencies with the XSens system which may have affected the overall results.

Future work should investigate the reliability of XSens data in order to determine the limitations of the CV system. The digital human modeling (DHM) technique could be used to model the data from both the systems for comparison which might reduce the noise and maintain appropriate postures. In addition, further research should be done into alternative camera positions, handling of self-occlusions and utilization of more cameras within the surrounding environment. Lastly, threshold adjustments can be made in consultation with ergonomists, considering that the calculated angles overall closely match XSens. The findings of this study suggest that obtaining results through CV methods may be more repeatable, faster and efficient compared to manual assessment of occupational ergonomics. This can catalyze the identification and reduction of injuries and illnesses as well as promote worker's safety and comfort.

## Acknowledgments

This research was funded by Scania CV AB. Thanks for the support from Juan Luis Jiménez Sánchez, Martin Sandberg, Malin Baresso and all the forklift drivers who participated in this study.

## References

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. https://doi.org/10.1016/0003-6870(92)90154-n

Fagarasanu, M., & Kumar, S. (2002). Measurement instruments and data collection: A consideration of constructs and biases in ergonomics research. *International Journal of Industrial Ergonomics*, *30*(6), 355–369. https://doi.org/10.1016/S0169-8141(02)00101-4

Flodin, U., Rolander, B., Löfgren, H., Krapi, B., Nyqvist, F., & Wåhlin, C. (2018). Risk factors for neck pain among forklift truck operators: A retrospective cohort study. *BMC Musculoskeletal Disorders*, *19*(1), 44. https://doi.org/10.1186/s12891-018-1956-3

Frederick, L., Habes, D., & Schloemer, J. (1984). An Introduction to the Principles of Occupational Ergonomics. *Occupational Health Nursing*, *32*(12), 643–645. https://doi.org/10.1177/216507998403201206

Hignett, S., & McAtamney, L. (2000). Rapid Entire Body Assessment (REBA). *Applied Ergonomics*, *31*(2), 201–205. https://doi.org/10.1016/S0003-6870(99)00039-3

Hossin, M., & Sulaiman, M. N. (2015). A Review On Evaluation Metrics For Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, *5*(2), 1–11.

Kim, W., Huang, C., Yun, D., Saakes, D., & Xiong, S. (2020). Comparison of Joint Angle Measurements from Three Types of Motion Capture Systems for Ergonomic Postural Assessment. In W. Karwowski, R. S. Goonetilleke, S. Xiong, R. H. M. Goossens, & A. Murata (Eds.), *Advances in Physical, Social & Occupational Ergonomics* (pp. 3–11). Springer International Publishing. https://doi.org/10.1007/978-3-030-51549-2_1

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. https://doi.org/10.1016/j.ergon.2021.103164

Kok, J., Vroonhof, P., & Snijders, J. (2020). *Work-related musculoskeletal disorders: Prevalence, costs and demographics in the EU*. European Agency for Safety and Health at Work. https://data.europa.eu/doi/10.2802/66947

Lin, P.-C., Chen, Y.-J., Chen, W.-S., & Lee, Y.-J. (2022). Automatic real-time occupational posture evaluation and select corresponding ergonomic assessments. *Scientific Reports*, *12*(1), 2139. https://doi.org/10.1038/s41598-022-05812-9

Lind, C., Forsman, M., & Rose, L. (2019). Development and evaluation of RAMP II - a practitioner's tool for assessing musculoskeletal disorder risk factors in industrial manual handling. *Ergonomics*, *63,* 1–35. https://doi.org/10.1080/00140139.2019.1710576

Lind, C., Rose, L., Franzon, H., & Nord-Nilsson, L. (2014). *RAMP: risk management assessment tool for manual handling proactively*. 4.

Massiris Fernández, 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. https://doi.org/10.1016/j.cie.2020.106816

McAtamney, L., & Nigel Corlett, E. (1993). RULA: A survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, *24*(2), 91–99. https://doi.org/10.1016/0003-6870(93)90080-S

Menolotto, M., Komaris, D.-S., Tedesco, S., O'Flynn, B., & Walsh, M. (2020). Motion Capture Technology in Industrial Applications: A Systematic Review. *Sensors*, *20*(19), 5687. https://doi.org/10.3390/s20195687

Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, *8*(4), 283–298. https://doi.org/10.1016/S0001-2998(78)80014-2

Van Crombrugge, I., Sels, S., Ribbens, B., Steenackers, G., Penne, R., & Vanlanduit, S. (2022). Accuracy Assessment of Joint Angles Estimated from 2D and 3D Camera Measurements. *Sensors*, *22*(5), 1729. https://doi.org/10.3390/s22051729

Viruet, H. B., Genaidy, A., Shell, R., Salem, S., & Karwowski, W. (2008). Effect of forklift operation on lower back pain: An evidence-based approach. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *18*(2), 125–151. https://doi.org/10.1002/hfm.20105

Zelck, S., Verwulgen, S., Denteneer, L., Vanden Bossche, H., & Scataglini, S. (2022). Combining a Wearable IMU Mocap System with REBA and RULA for Ergonomic Assessment of Container Lashing Teams. In N. L. Black, W. P. Neumann, & I. Noy (Eds.), *Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021)* (pp. 462–465). Springer International Publishing. https://doi.org/10.1007/978-3-030-74614-8_59