

Article

Industrial Ergonomics Risk Analysis Based on 3D-Human Pose Estimation

Prabesh Paudel ¹, Young-Jin Kwon ², Do-Hyun Kim ² and Kyoung-Ho Choi ^{1,*}¹ Department of Electronics Engineering, Mokpo National University, Jeonnam 58854, Korea² Intelligent Robotics Research Division, Electronics and Telecommunications Research Institute, Daejeon 34129, Korea

* Correspondence: khchoi@mokpo.ac.kr

Abstract: Ergonomics is important for smooth and sustainable industrial operation. In the manufacturing industry, due to poor workstation design, workers frequently and repeatedly experience uncomfortable postures and actions (reaching above their shoulders, bending at awkward angles, bending backwards, flexing their elbows/wrists, etc.). Incorrect working postures often lead to specialized injuries, which reduce productivity and increase development costs. Therefore, examining workers' ergonomic postures becomes the basis for recognizing, correcting, and preventing bad postures in the workplace. This paper proposes a new framework to carry out risk analysis of workers' ergonomic postures through 3D human pose estimation from video/image sequences of their actions. The top-down network calculates human body joints when bending, and those angles are compared with the ground truth body bending data collected manually by expert observation. Here, we introduce the body angle reliability decision (BARD) method to calculate the most reliable body-bending angles to ensure safe working angles for workers that conform to ergonomic requirements in the manufacturing industry. We found a significant result with high accuracy in the score for ergonomics we used for this experiment. For good postures with high reliability, we have OWAS score 94%, REBA score 93%, and RULA score 93% accuracy. Similarly, for occluded postures we have OWAS score 83%, REBA score 82%, and RULA score 82%, compared with expert's occluded scores. For future study, our research can be a reference for ergonomics score analysis with 3D pose estimation of workers' postures.

Keywords: joint angles; Ovako working posture assessment system (OWAS); rapid upper limb assessment (RULA); rapid entire body assessment (REBA); pose estimation



Citation: Paudel, P.; Kwon, Y.-J.; Kim, D.-H.; Choi, K.-H. Industrial Ergonomics Risk Analysis Based on 3D-Human Pose Estimation. *Electronics* **2022**, *11*, 3403. <https://doi.org/10.3390/electronics11203403>

Academic Editor: Gemma Piella

Received: 6 September 2022

Accepted: 14 October 2022

Published: 20 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The state-of-the-art method in machine learning has achieved exceptional precision on many computer vision tasks exclusively from image learning models. There are factors associated with the work environment that can affect a worker's mental health, such as an inappropriate interaction between the type of job and the person's skills and competencies. These aspects can also influence, for example, the level of organization of the environment and the benefits that a company can offer to get the job done. Musculoskeletal disorder (MSDs) are perhaps the basic medical condition and the primary justification for nonattendance from work. MSDs are caused by musculoskeletal load built up from repeated improper postures, so workers' postures and movements provide key information in determining the likelihood of musculoskeletal injury. A recent statistical study conducted by the Bureau of Labor Statistics (BLS) showed that cases of MSDs account for 31% of all work-related injuries and illnesses [1]. Adopting ergonomically invalid or uncomfortable work postures while performing these manual activities can potentially lead to long-term MSDs. To resolve this issue of laborers in ergonomics, specialized labor are utilized to dissect the specialists working posture and the sort of hazards established with

that workplace. These manual techniques might be wrong and wasteful because of abstract inclination [2]. Since manual perception is done to minimize mishaps in various work places, we need to adjust that point of view. We zeroed in for the most part on three manual perception strategies in ergonomics mishap minimization, which are the Ovako Working Analysis System (OWAS), Rapid Limb Upper Assessment (RULA), and Rapid Entire Body Assessment (REBA). Although the previously mentioned ergonomics risk measurement tools are famous manual perception techniques, they have a few limits. Certain body points, which are hard to access physically, are not obviously characterized. RULA is an overview strategy produced for use in an ergonomics examination of work where business-related upper appendage issues are accounted. Profundity sensors can give rich data about human postures in indoor settings and are utilized in different arrangements, such as gaming in Microsoft's Xbox with Kinect [3]. Essentially REBA is a postural examination instrument, which is touchy to musculoskeletal dangers in an assortment of undertakings and evaluations of working stances found in medical examination of human body muscles [4]. The OWAS technique for instance does not give data about pose term. It does not distinguish whether arms are left or right, and, furthermore, no data are available on elbow position. Additionally, the other two strategies do not provide substantial point for upper appendage and body. Pose estimation in computer vision alludes to the task of assessing the area of key joints of human body in image or video recording. Contingent on the application, the present assessment is divided into three classes of 2D pose estimation [5], static monocular pictures, 3D pose estimation from profundity/range pictures, and 3D pose estimation straightforwardly from monocular 2D pictures, as described in article [6]. The adaptability of the human body as different degrees of freedom prompting self-impediment has kept the posture assessment of people a long way from being addressed. These methods predict the working pose but there is a need for advanced methods with deep learning in the ergonomics with the help of 2D and 3D pose estimation to minimize error in ergonomics risk evaluation. We have seen much research on ergonomics and pose estimation but no researcher has combined research on the ergonomics as we did. Hence, we came up with idea of combining the pose estimation and ergonomics manual observation to reduce risk at a workplace. In this paper, our goal is to estimate 3D human poses from single image/video and calculate the ergonomic score on three different ergonomics: OWAS, RULA, and REBA. Our objective is to digitize the ergonomics score calculation method with 3D pose estimation with higher accuracy. The 3D multi person pose estimation (3DMPPE Pose Net) [7] is used for 3D joint localization where human body joints are allocated as key points. These located key points are used to calculate the bending of joint angles. Hence, from input 2D image we have a 3D human pose joint view, which makes it much easier to calculate the joint angles. As our 3D human pose is ready to calculate the ergonomics score, we need to make sure best human pose is chosen as input data; for this, we introduce a 'Body Angle Reliability Decision' (BARD) model. That enables us to project the best input human pose, limiting the low reliability human pose for better accuracy and ergonomics score, as shown in Figure 1 below.

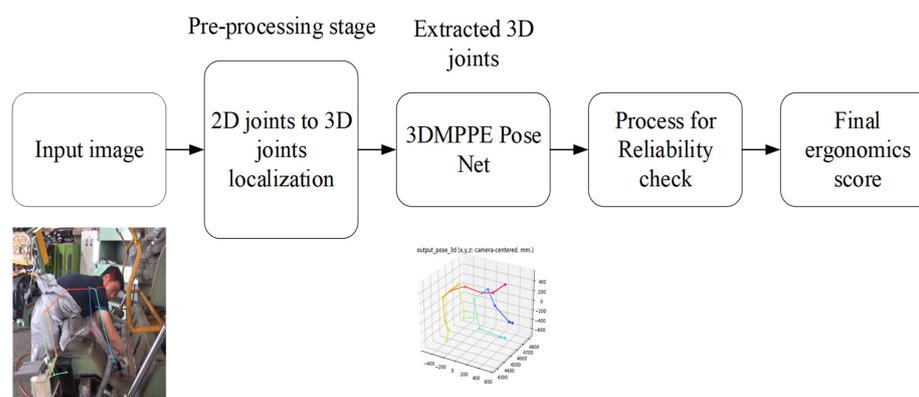


Figure 1. Block diagram of BARD network framework.

In summary, our contributions are as follows:

- We propose a framework for an automatic pose analysis of industrial workers to prevent long-term MSDs.
- We propose a novel approach of reliability decision to make sure input video sequence is appropriate for industrial workers' pose analysis.
- We present a linear model for the reliability analysis, producing an accuracy estimate for a corresponding workers' pose input.
- We aim for our research to be beneficial in ergonomics-related work to prevent a human workforce in the long term by reducing unnecessary injuries caused by bad posture working conditions.

The rest of article is organized as follows. In Section 2, related research is covered. In Section 3, the proposed approach is described in detail, including datasets preparation, pose estimation, and body angle reliability decision. Section 4 signifies all the experiments and results with datasets preparations and ergonomics score analysis. Section 5 is all about discussion and simulation results. Section 6 will be the conclusions and future work that could come from our research.

2. Related Work

This section discusses recent approaches to ergonomics score calculation: OWAS, RULA, REBA, and 3D pose estimation. Ergonomic risk was analyzed with manual expert inspection until a few years back when machine learning with computer vision revolutionized human action detection and pose estimation techniques. For our research, we selected three methods: OWAS, RULA, and REBA. The OWAS method estimates the static workload of the worker in the workplace by analyzing worker's postures during operation. It identifies four classes, which show static load risk degree [5]. Rapid Upper Limb Assessment (RULA) by Mc Atamney [3] in 2005, was used in ergonomics examinations of working environments where upper human body parts were only included for posture manual examination. Recent studies on RULA ergonomics are done referring to computer vision and machine learning [8]. In particular, the kinetic method with camera and software development kit (SDK) have been used to analyze the posture and RULA score [9–11]. REBA, as the name recommends, is an ergonomic examination instrument that is easy to use to evaluate an undertaking or a movement to check for dangers of musculoskeletal problems [2]. Similarly, we can see how the RULA score was calculated with differently adopted postures [9,10]. It provides necessary information about posture with convolutional neural network (CNN) and lower post-processing operation.

In recent years, deep learning methods [12–14] for evaluating human posture in 2D have been developed significantly. There are mainly two approaches in human pose estimation. The first is a top-down approach where bounding boxes are formed to detect human first, and the second one as bottom-up approach, which locates all human body key points in an image and then, with clustering techniques, groups them in an input image. These methods take advantage of advances in human recognition and additional person bounding box identification information. The top-down paradigm requires satisfactory performance, but at an additional cost for personal box recognition. Notable top-down approach work includes HR Net [15,16], Pose Net [17], RMPE [18], and Mask R-CNN [19]. In addition, key points localization from heat map [20–22], data augmentation [23], multi-task learning [24], handling occlusion [25–27], and pose estimation [28,29], are further top-down approaches. Deep learning has recently demonstrated its capabilities in many computer vision tasks, such as 3D evaluation of human posture. Recent advances in 3D assessment of human posture are largely due to the use of various deep neural network models. However, these rely heavily on well-annotated data from fully supervised trained model and can rarely be generalized to new scenarios representing missing templates from the training dataset, such as new camera angle and human poses. Therefore, some recent research is exploring how to use external information to increase generalizability [30]. Even though 2D human pose estimations have made significant progress as described

in [31], which focuses on human body shape, it talks a little about occlusion and invariant changes in human body appearance due to the hourglass network for providing human pose estimation. It is an improvement compared to the low-dimensional parameter model of body shape in [31]. Later, [32] showed a significant improvement for learning in spatial models with CNN incorporated into pose machine framework. With multi person pose estimation and joint localization, [32–34] gives significant improvement in human pose estimation. It still remains a challenge as some methods use camera array systems to track accurate 3D body motion [35,36] due to occlusion and unclear tracking. In addition, the effective human structure information was used in [37], and this approach was much more improved in hierarchical joint prediction [38], similarly 2D keypoints refinement on [39] and view-invariant constraint in [40].

3. Method

The goal of our framework is to analyze the ergonomics risk in work places with 3D pose estimation from 2D input image/video dataset. This section provides an overview of our framework for risk analysis and scoring of different ergonomics proposed on this research. We argue that 2D poses alone are not enough for accurate human pose estimation for action and body bending recognition. To justify this, we provide different key point features and conversion of 2D pose to 3D pose with 3DMPPE Pose Net [7] method to automate the ergonomics manual risk analysis.

The proposed architecture is described in Figure 2, having a video sequence as an input and producing an action category output based on the result of ergonomic risk analysis. As proposed in our architecture design, the first step is getting an input video from hand-collected data from the manufacturing industry. The second step is locating a worker from the input video sequence. In our implementation, Darknet-53 is used to detect a worker and the region of the human body is located. After that, the human joint localization process is done via Pose Net network. This concludes our feature extraction and pose estimation process. The fourth step is a reliability check step with our proposed body angle reliability decision (BARD) network. In the proposed BARD step, the extracted human pose is evaluated whether or not it is good enough for the ergonomic score evaluations. More specifically, the BARD produces a reliability score, which is 1 for the maximum reliability and a score of 0 for minimum reliability. If the reliability score is high enough, ergonomics score evaluation is performed. In the final step, an action category is decided according to the ergonomic score.

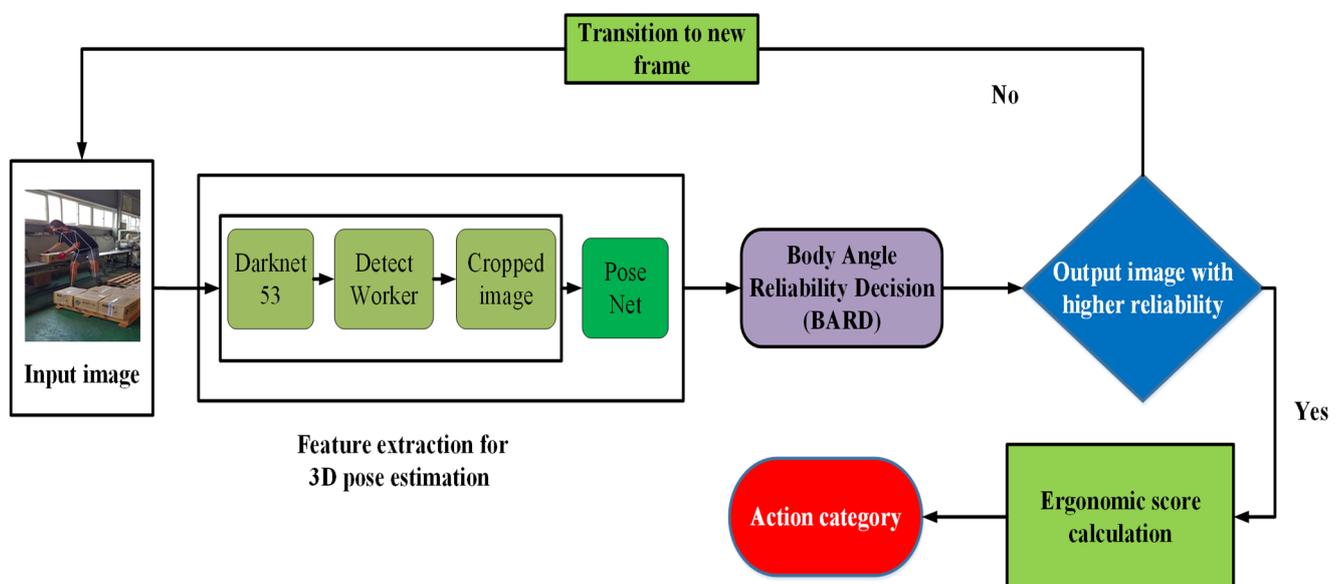


Figure 2. Pipeline of proposed architecture and BARD for workers' pose estimation and action category.

3.1. Feature Extraction and Worker Detection

We use Yolov3 [27] as a framework to locate a human worker. Yolov3 consists of two parts: a bounding box prediction and feature extraction. It predicts an object score for each box using logistic regression with width and height from an input image, based on the created bounding box and feature extracted from Darknet-53. It contains 53 different convolutional layers. This new feature extraction network is much more powerful than Darknet-19 and ResNet-101 or Resnet-152. Darknet-53 also achieved the highest measured floating-point operations per second. This means that the network topology makes better use of the GPU, making evaluation more efficient and faster. This is mainly because ResNets have too many layers and is inefficient. Thus, Darknet-53 performs on par compared to the state-of-the-art classifier with maximum speed and minimum floating-point operations. In the proposed architecture, information about only detected person, such as x, y coordinates and width and height, i.e., $(P_x, P_y, P_{width}, P_{height})$, are returned from an input image I , as described in (1),

$$(P_x, P_y, P_{width}, P_{height}) = Person[Yolov3(I)]. \quad (1)$$

3.2. Our Approach to Pose Estimation

In most of pose estimation approaches, there are two approaches and the most commonly used one is called a top-down approach, deploying a human detector estimating bounding boxes of humans. Most of detected human area is cropped and fed into the pose estimation network. The second one, bottom-up approach, localizes all human body key points in an input image first, and then groups each person using clustering techniques. In our approach we used 3DMPPE [9] for human pose extraction, but the location of a human is provided in the proposed approach using Yolov3. The pose estimation part takes the feature map from the body part and up-samples it, using a batch normalization layer [7] and three successive deconvolutional layers with ReLU activation. A 1×1 convolution is applied to the up-sampled feature map to generate a 3D heat map for each joint. For 2D image coordinate extraction soft-argmax operation is used. As shown in Figure 1, 3DMPPE was used to estimate relative root 3D pose from cropped human images. 3DMPPE uses RootNet and PoseNet to generate the 3D human pose from the 2D human pose, as described below. Please refer to [9] for further information.

In RootNet, ResNet50 is used as a backbone network to extract a feature map. Then, 1×1 convolution is used to produce a correction factor, followed by a global average pooling. Lastly, the depth value of each feature point is calculated by multiplying a value k that is calculated by using (2):

$$k = \sqrt{\alpha_x, \alpha_y \frac{A_{real}}{A_{img}}}, \quad (2)$$

where α_x , α_y , A_{real} and A_{img} are focal lengths divided by per-pixel distance factors and the areas of human in real and image spaces, respectively. In PoseNet, the depth of feature points relative to root is calculated. For training of PoseNet, L1 distance is used to minimize the distance between real 3D coordinates and the corresponding estimated coordinates.

3.3. Body Angle Reliability Decision (BARD)

In the previous section, we introduced two processes of feature extraction and 3D pose estimation for input data. This section explains how the data are selected to ensure high accuracy calculations of workers poses are obtained. Here, we proposed body angle reliability decision between camera and workers pose with three major joints from waist, arm and leg in calculating the body-bending angle. As shown in Figure 1, we can see how the x-axis and z-axis of human and camera positions align with each other to ensure maximum reliability can be measured. The main purpose of introducing BARD is to measure workers' poses accurately. Human experts are likely to use images in which

workers' poses can be seen clearly. In other words, human experts skip the images where poses are not estimated accurately. Therefore, we introduced a reliability measure to detect poorly captured angles in images. In the proposed approach, we define the reliability R with a linear function:

$$R = K \frac{\vec{x} \cdot \vec{z}}{|\vec{x}| |\vec{z}|}, \quad (3)$$

where K is a constant.

The main goal of our system is to recover the maximum likelihood value of reliability, denoted as R in above equation with K as a constant. We use it as a trained parameter, making sure all the high reliability images are taken into account for ergonomics calculation. It ensures the high reliable angles are taken for camera angle and workers' position. Similarly, as shown in Figure 1, we have three axes. The z is the optic axis and x axis denotes the line connecting the left and right shoulder points of a worker. For instance, if a worker stands right in front of the camera, the angle between z and x axis is 90 degrees. If the worker turns around 90 degrees and the camera sees exact the side view of the worker, the angle between z and x axis is 0 degrees. The BARD was calculated with the cameras z -axis coordinate and human x -axis coordinate values. The coordinates output from 3D heat maps for each joint is used to measure different bending angles between the joints. To calculate values of BARD we use (3), and the value is used to decide whether input image is appropriate to calculate workers' pose estimation. We choose the camera angle and workers' body pose and model their relationship as a linear function. As in Figure 1, we want to ensure that the 3D output model from PoseNet is of higher reliability. To block out the unnecessary low reliable human pose data, BARD model is appropriate for our research with minimal cost.

4. Experiments and Results

In this experiment, we focus on calculating the ergonomic score of workers' poses using three ergonomic score analysis methods: OWAS, RULA, and REBA. To calculate the ergonomic scores for each method and to analyze the risk of working poses, we used Pose Net model to extract the body key point features. All of the joints labelled as key points, were transformed into 3D models by Darknet and Pose Net feature extractors, and we introduced the reliability check, as explained in Figure 2.

Therefore, we showed how the poses were prepared and analyzed before calculating the final score for workers' poses. Publicly available datasets were used to train our collected dataset. The Human3.6 dataset [41] is the largest dataset for 3D single person benchmark, and consists of 15 activities for 11 different subjects, captured from four different viewpoints. In addition, datasets, such as COCO [42] and MPII [43,44], were used for training. Pycharm was used for implementation. We trained our datasets with five NVIDIA RTX 2080Ti GPUs. We present figures of simulations and tables to explain our experiment in detail. We conducted our experiment using these models and datasets to test our system output with the expert-generated ergonomic score. The extracted 2D key points features from YOLO Darknet model are fed into Pose Net model for 3D human pose estimation.

4.1. Datasets Preparation and Extraction

For the evaluation of workers' poses, we captured videos of industrial workers. Samples of captured images from video are shown in Figures 3 and 4. It consists of more than 10,000 video frames. We selected 600 images for ground truth evaluation as benchmarks. Three experts separately evaluated the same datasets giving three different scores for ground truth variability. In our experiment, we compared our system output to justify the experts' decision to ensure that our system produced the similar results. We compared the results with Cohens kappa κ [45] to compare the agreement index with the experts evaluation, where Cohens kappa κ is measured with experts' observation agreement and probable

agreement on different poses of workers' body angles. This method is helpful in comparing machine-learning predictions with manually established predictions. Many researchers have used the Cohens kappa measurement in most posture reliability studies [46–48]. If the Cohens kappa κ values are less than 0, then there is no agreement at $\kappa = 0.01 - 0.20$, there is poor agreement, $\kappa = 0.21 - 0.40$, there is fair agreement, while, $\kappa = 0.41 - 0.60$. indicates moderate agreement, $\kappa = 0.61 - 0.80$ good agreement, and the $\kappa = 0.81 - 1$ is in very good agreement [45,49]. Hence, we use the below equation to compare our system prediction with expert prediction scores of workers' postures as follows:

$$\kappa = \mathcal{P}_0 - \mathcal{P}_e / 1 - \mathcal{P}_e \tag{4}$$

where, \mathcal{P}_0 is the relative observed agreement between experts on ground truth data and \mathcal{P}_e is the probable chance of agreement. If the rates are in complete agreement, $\kappa = 1$. In case of no agreement, it would be expected by chance (as given by \mathcal{P}_e), $\kappa = 0$. For κ categories, M observations categorize and m_{ki} is the number of times rater i predicted category κ , \mathcal{P}_e is described as:

$$p_e = \frac{1}{M^2} \sum_k m_{k1} m_{k2} \tag{5}$$



Figure 3. Good posture image with high reliability.



Figure 4. Sample images of workers' different working poses.

Here, we present detailed results comparing the accuracy of ergonomics OWAS, RULA, and REBA scores with different data and methods. Tables 1–3 show the raw data for an input image and the scoring of different body parts taken for measurement. For OWAS waist, arm and leg are used and for RULA and REBA, upper and lower body parts, such as upper and lower arm, wrist, neck, trunk, leg, and waist are considered. From scoring, we showed the accuracy of our system with different datasets in Tables 4–8. The accuracy of good postures, where all of the body joints are aligned to the x axis of human pose and the z axis of camera position, is shown in Table 4. This poses shows that accuracy was high, compared with the data sets with occlusion. The occluded images have slightly less accuracy because the angle calculation from those key points are not accurate every time. As shown in Tables 4–8, we can see how the data sets are divided into different sections for reliability calculations. Some of the data sets have high reliability, while some have low reliability in terms of the positioning with camera angles. Some data sets have high reliability but have low scores because of faulty detection where reliability is high. Similarly, occlusion is a major factor whether its self-occlusion affects the reliability and ergonomics score. Figure 5 shows an example of occlusion image, and it is one reason the ergonomics accuracy score is low. Getting reliable and accurate 3D joints from a single image is an intractable problem. We have seen few methods with LSTM [42] and RNN [50] using joint inter dependence and temporal convolutional methods to generate 3D pose from 2D key point sequence. However, it is not easy to use on each frame, as it requires the estimation of all 2D key points in every frames. Assuming all the prediction error it generates with temporary non-continuous and independent results, this does not apply to most of the occlusion cases. Thus, we choose the cylinder man model and apply it on occlusion as a network in [51,52], and it generated occlusion labels for the 3D data. We have results on our own data sets in Tables 4 and 5, and we have results for before and after BARD trained with publicly available datasets, such as Human3.6M, COCO, and MPII, in Tables 6 and 7. They show the accuracy on ergonomics we obtained for our dataset. Similarly, we trained our data sets with higher HR Net [53] but the results were not satisfactory, as shown in Table 8. The higher HR Net [10] 2D extracted feature were not effective on occluded data and body key joints detection, which causes decrement of accuracy on ergonomics scores. From the results, we can see that using the higher HR Net [10] on feature extraction has lower accuracy on ergonomics accuracy before and after the application of BARD. This method achieved relatively lower accuracy, compared to 3DMPPE [10] on all three ergonomic methods we have used on our system.



Figure 5. Occluded image data with low reliability.

We found too much occlusion in the key points creates unreliable angle measurement and later affects the reliability check of the input data. Angles that are too small reliable angle and lower key points creates a huge accuracy dip in the system output. To fix this problem we separated the good posture and occluded posture datasets as input in the network. We then evaluated our approach on two method datasets. However, our key

focus was on matching our final ergonomics score with those of the experts, which are shown in Tables 9–11 respectively.

Table 1. Example of OWAS score chart for input images.

Input Image	Waist	Arm	Leg	Weight	OWAS score	Action Category
	3	3	2	1	1	1
	1	2	2	1	1	1
	1	1	2	1	1	1
	4	1	3	1	2	2

Table 2. Example of REBA score chart for input images.

Input Image	Neck Score	Trunk Score	Leg Score	U/A Score	L/A Score	Wrist Score	Action Category
	2	3	1	3	2	1	2
	1	2	1	3	1	2	1
	2	1	1	1	1	3	1
	2	4	2	1	2	1	2

Table 3. Example of RULA score chart for input images.

Input Image	Neck Score	Trunk Score	Leg Score	U/A Score	L/A Score	Wrist Score	Action Category
	3	3	1	3	3	1	2
	3	1	3	1	2	1	2
	2	1	4	3	1	1	2
	1	2	1	3	4	2	3

Table 4. Evaluation and comparison of accuracy of good posture.

Method	OWAS Accuracy	RULA Accuracy	REBA Accuracy
Before applying BARD	91%	92%	92%
After applying BARD	94%	93%	93%

Table 5. Evaluation and comparison of accuracy of occluded posture.

Method	OWAS Accuracy	RULA Accuracy	REBA Accuracy
Before applying BARD	74%	78%	81%
After applying BARD	83%	82%	82%

Table 6. Evaluation and comparison of accuracy of good posture for validation datasets.

Method	OWAS Accuracy	RULA Accuracy	REBA Accuracy
Before applying BARD	91%	92%	92%
After applying BARD	95%	94%	94%

Table 7. Evaluation and comparison of accuracy of occluded posture for validation datasets.

Method	OWAS Accuracy	RULA Accuracy	REBA Accuracy
Before applying BARD	78%	86%	86%
After applying BARD	83%	81%	81%

Table 8. Evaluation and comparison of higher HR Net feature extracted datasets.

Method	OWAS Accuracy	RULA Accuracy	REBA Accuracy
Higher HR Net	73%	75%	72%
Ours	75%	76%	74%

Table 9. OWAS score accuracy compared with expert scores with observed agreement (P_0) and Cohens kappa (κ).

OWAS Score	P_0	κ
Back	0.959	0.941
Arms	0.943	0.941
Legs	0.946	0.947
Risk	0.961	0.958

Table 10. RULA score performance with experts' data.

RULA Score	P_0	κ
Table A score	0.956	0.946
Table B score	0.947	0.947
Risk	0.931	0.941

Table 11. REBA score performance with experts' data.

RULA Score	P_0	κ
Table A score	0.954	0.961
Table B score	0.944	0.936
Risk	0.941	0.924

4.2. OWAS Score Analysis

OWAS was developed to evaluate the exposure of individual workers to ergonomic risk factor associated with both upper and lower body, such as back, arm, and leg postures [48]. It counts the score from a different position of body and gives a final score, which will determine the category of ergonomics risk level. Figure 6 elaborates different working postures used for OWAS score analysis.

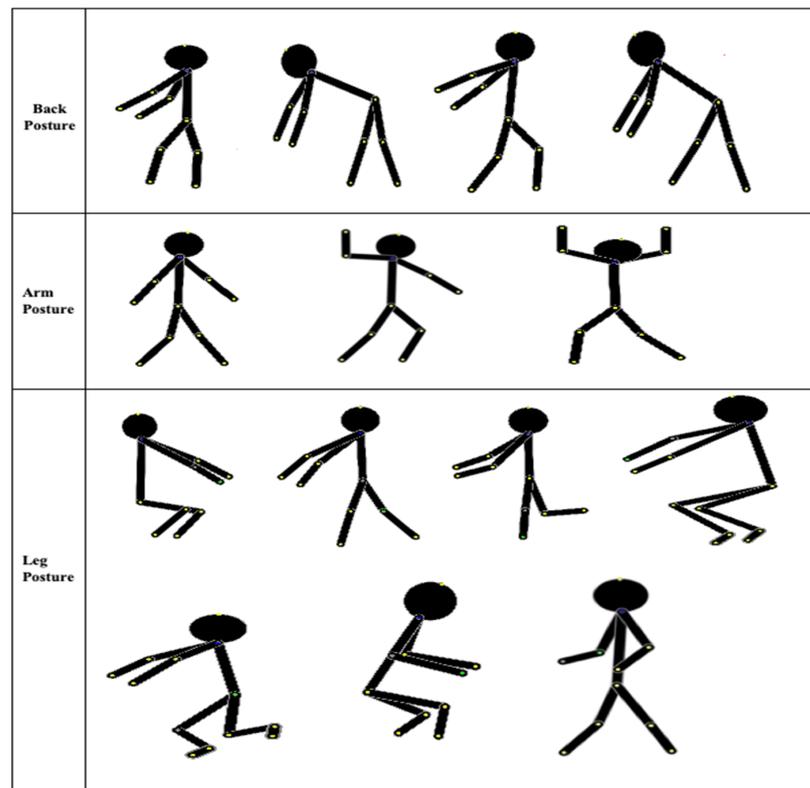


Figure 6. Standard working postures in the OWAS method.

To validate the reliability of our score, we need to match the agreement between the score from expert and our system score, so we use Cohen’s kappa method. The nearer the numbers are to 1, the more agreement there is of the calculated score. Table 9 shows the observed and probable agreement between OWAS scores computed from our estimated joint angle scores and scores from expert data. We considered the leg, arm, and waist for OWAS scoring. We had to adjust the weight to minimum.

4.3. RULA Score Analysis

RULA was developed to evaluate the workers’ ergonomics risk factor associated with the upper extremity MSD. This method also considers the load extremities on neck and trunk. For RULA, we also consider the minimum weight, force/load and muscle as static. As shown in Table 10, we calculated accuracy and matched it with experts’ scores. It is divided into three different score tables, as shown in Figure 7 below. For the Table A score, upper arm, lower arm, and wrist angles were considered. Similarly, for the Table B score, we considered neck, trunk, and leg angles. In addition, the final score was matched from Tables A and B to analyze the risk on Table C, as shown in Figure 7. The minimum RULA score is 1 and maximum is 7, which represents the ergonomics risk associated with the job.

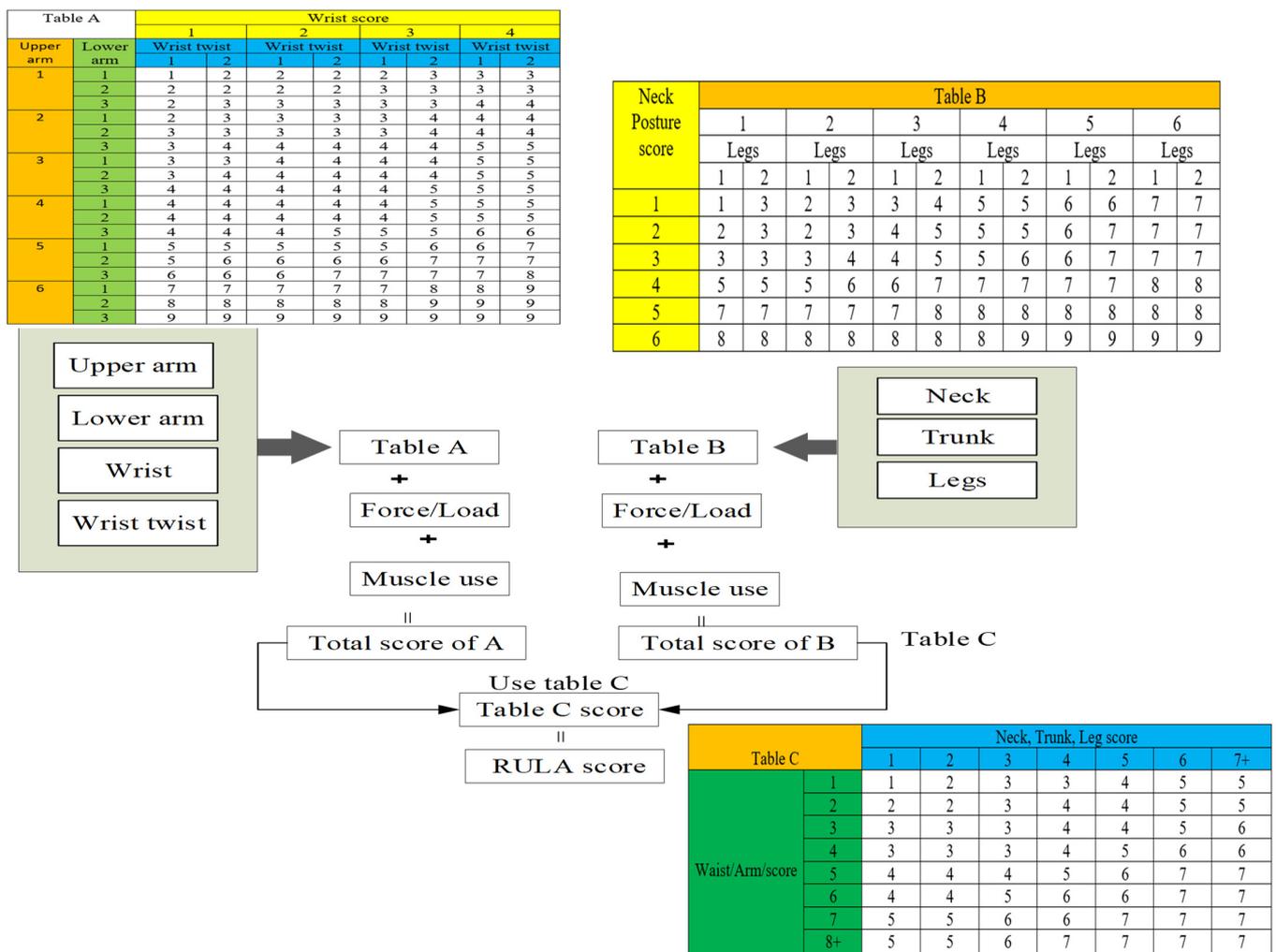


Figure 7. Standard scoring sheet for RULA.

To validate the calculated scores, we also matched the agreement between observed and calculated score and it has high agreement values, as shown on Table 10 below.

4.4. REBA Score Analysis

This is also similar to the RULA score. Only Table A and B are switched with some modifications in how the bending of body angles are considered. Addition in REBA is the leg score further, as shown on Figure 8. We followed the same protocol as was the case in the RULA table. In addition, the coupling effect is adjusted as fitted, good grip, and acceptable [2]. Our REBA score also has a higher agreement score calculated with Cohen’s kappa.

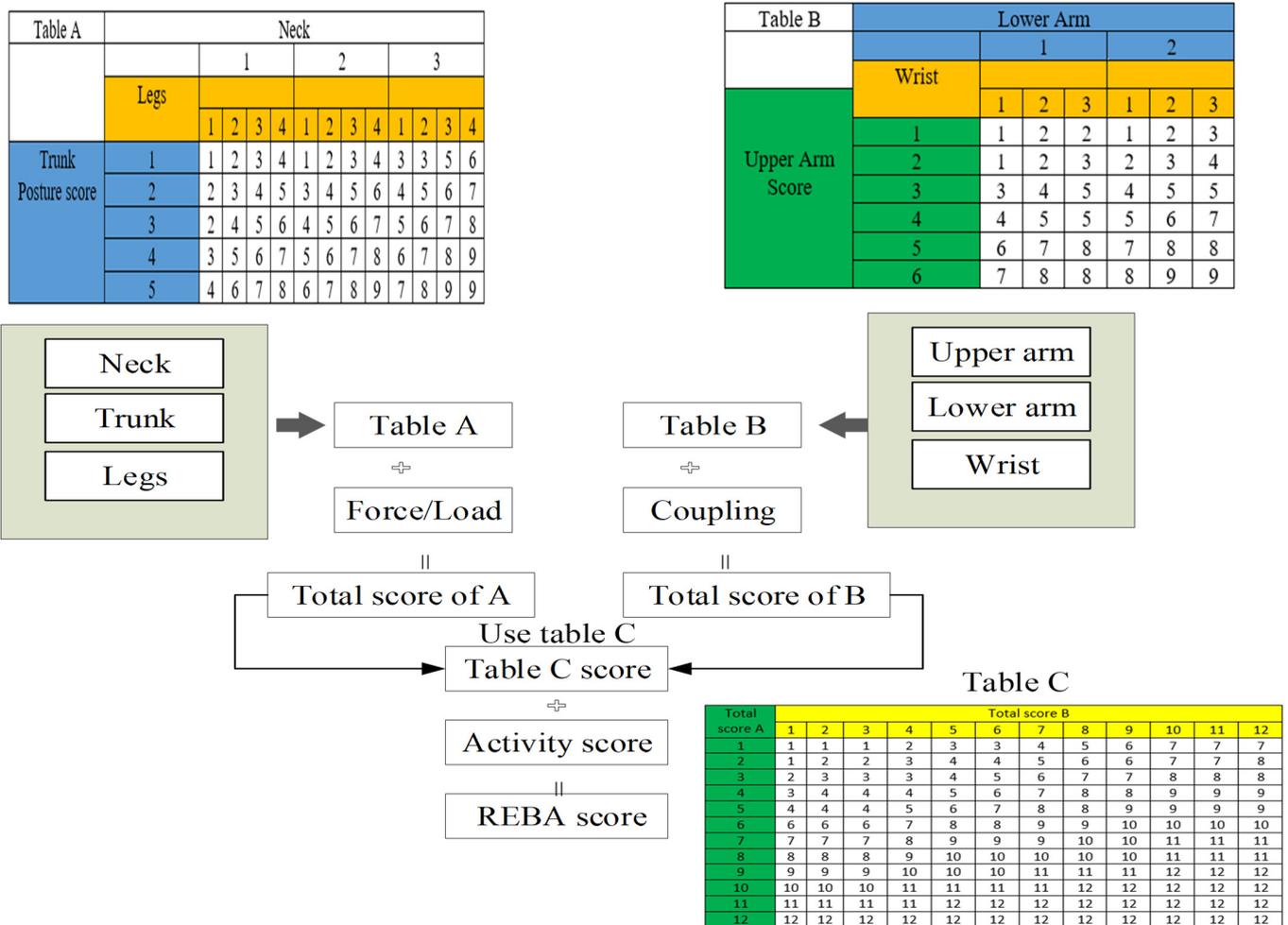


Figure 8. Standard scoring sheet for REBA.

Figure 8 shows a complete REBA scoring and risk evaluation on workers’ body. Hence, Tables 9–11 show how much accuracy our system produces while compared with the experts’ scores. These tools are used in our system to evaluate upper and lower body parts and MSD risks associated with the workers’ job or tasks.

5. Discussion and Simulation Results

When occlusion occurs, it has adverse effects on reliability, as well as on ergonomics scores. For example, self-occlusion case reliability is high but some key points are occluded, which affects the overall reliability score. To decide the best-fit model for our experiment, we modeled the initial relationship between angle and accuracy on linear and exponential functions. Based on the experiment results shown in Figure 9, we found that the best fit for the reliability function was the linear regression model. This model achieves good accuracy, suggesting that the features contain meaningful key points and bending angles of the joints. Notably 3D human pose estimation is sensitive to occlusions and joint angles. We can conclude that maximum likelihood value R is highly dependent to the x coordinate

of human body and z coordinate of camera angle, which are considered in reliability decision. Reliability is impacted from object-oriented occlusions and undetected joints on angular measurements. We also explored that the accuracy of the ergonomics score are high when angle between camera placing and workers' position is placed between 45 to 90 degree. Hence, to estimate workers' poses for high accuracy, the corresponding reliability line can be used to define the angle and its relationship with accuracy. The simulation plot in Figure 10 shows the accuracy incremental in our system after the introduction of BARD. As shown on the plot, before BARD the accuracy was low but after we removed all unreliable 2D key points and added the occlusion calculation of 2D key points it produced more accurate and stable results. We channeled the reliability threshold to 0.5 and the constant K optimum value set temporarily at 1 from the camera viewing angle, so that lower reliable input data would not be used in accuracy calculation. We also made sure the heavily occluded key points are also recovered. Without the occlusion awareness and reliability check function, such enormous key points detected were treated as the same, leading to a possible hindrance on accuracy.

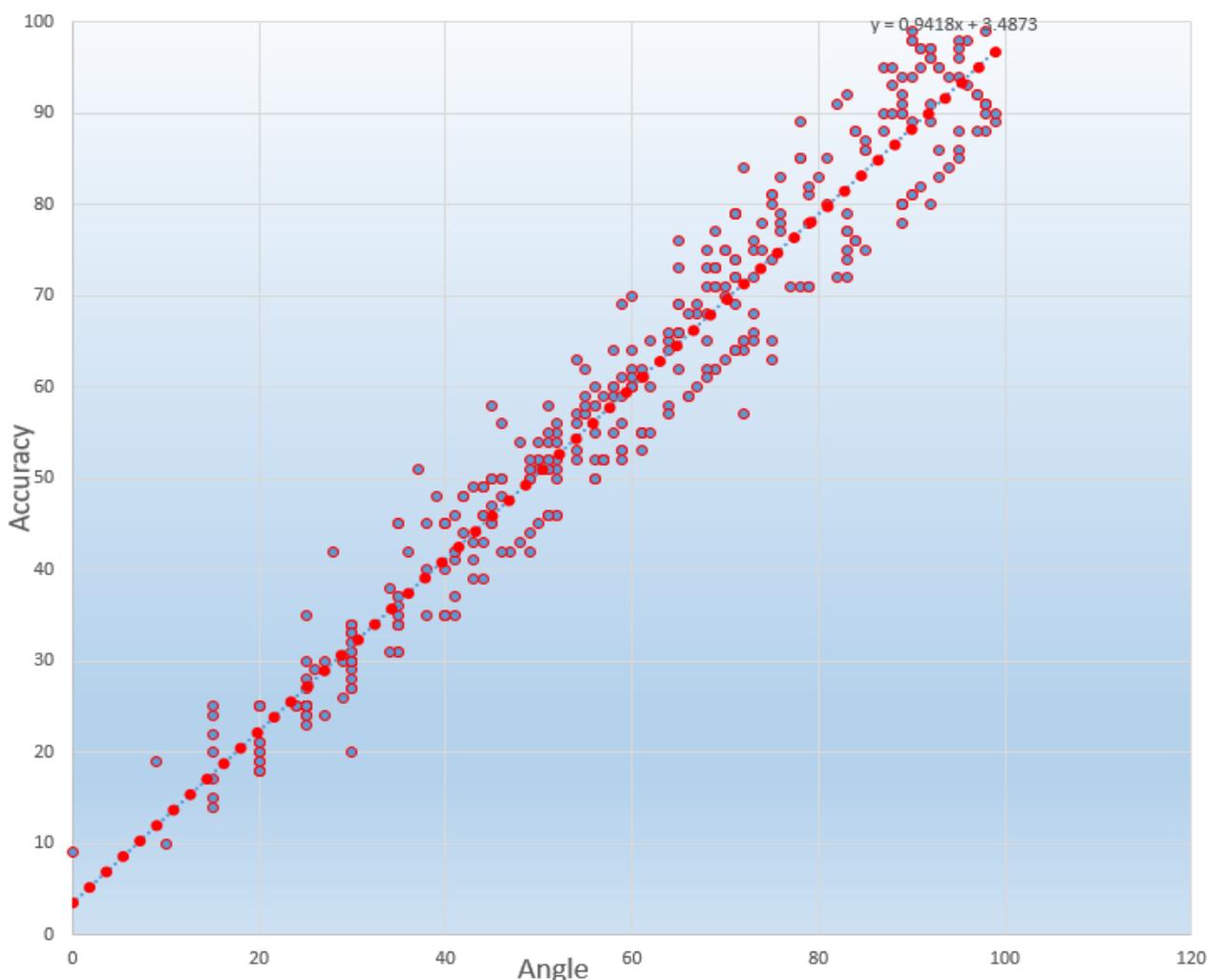


Figure 9. Modelling the reliability using linear regression on test datasets.

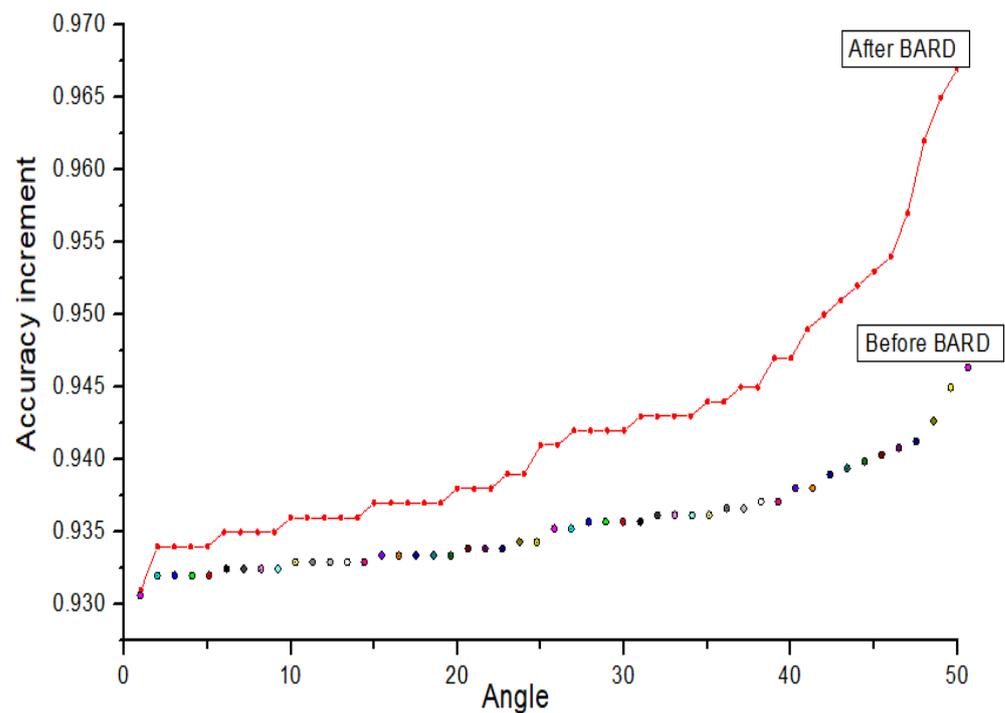


Figure 10. Accuracy incremental plot before and after introducing BARD.

6. Conclusions

In this paper, we proposed a novel ergonomics risk analysis framework for 3D human pose estimation. Our system addresses ergonomic risk with help of 3D human pose estimation, which automates the ergonomics score analysis. To improve the accuracy of ergonomics, this paper provided a 3D skeleton joint pose estimation from 2D joint pose and combined them with introduction of BARD method for reliability check of input datasets. Our research applies 3D single-person pose estimation on a single RGB image for workers' pose estimation and body joint bending angle calculation. In addition, a new dataset is captured, which will provide big advantage in future research, requiring big datasets in ergonomics. In addition, we used occlusion calculation method for estimation of workers' pose from input data image. As far as we know this will be the first piece of work to 3D human pose in ergonomics to address the industrial work risk problem. This research will lead to a new idea for automated postural ergonomics calculation contributions, combined with different complex working environments. In future work, we will focus on resolving dense occlusion problems and present a more sophisticated version of the reliability function for workers' pose estimation.

Author Contributions: Conceptualization, P.P., Y.-J.K., D.-H.K., K.-H.C.; methodology, P.P., K.-H.C.; software, P.P., K.-H.C.; formal analysis, P.P., K.-H.C.; investigation, P.P., K.-H.C.; writing, P.P., K.-H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by overseas training support funds of Mokpo National University in 2022.

Data Availability Statement: Expert Dataset is not publicly available.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bureau of Labor Statistics, US Department of Labor. Nonfatal Occupational Injuries and Illnesses Resulting in Days Away from Work in 2015. 2016. Available online: <https://www.bls.gov/news.release/pdf/osh2.pdf> (accessed on 22 February 2018).
2. Hignett, S.; McAtamney, L. Rapid Entire Body Assessment (REBA). *Appl. Ergon.* **2000**, *31*, 201–205. [CrossRef]
3. McAtamney, L.; Corlett, E.N. RULA: A survey method for the investigation of work-related upper limb disorders. *Appl. Ergon.* **1993**, *24*, 91–99. [CrossRef]
4. Wright, E.; Haslam, R. Manual handling risks and controls in a soft drinks distribution centre. *Appl. Ergon.* **1999**, *30*, 311–318. [CrossRef]
5. Liu, H.; Liu, T.; Chen, Y.; Zhang, Z.; Li, Y.-F. EHPE: Skeleton Cues-based Gaussian Coordinate Encoding for Efficient Human Pose Estimation. *IEEE Trans. Multimedia* **2022**, *2*, 1–12. [CrossRef]
6. Salvendy, G. *The Occupational Ergonomics Handbook*; Karwowski, W., William Marras, S., Eds.; CRC Press LLC: Boca Raton, FL, USA, 1999; p. 2065. ISBN 0-8493-2641-9.
7. Plantard, P.; Auvinet, E.; Pierres, A.S.; Multon, F. Pose estimation with a kinect for ergonomic studies: Evaluation of the accuracy using a virtual man-nequin. *Sensors* **2015**, *15*, 1785–1803. [CrossRef] [PubMed]
8. Liebrechts, J.; Sonne, M.; Potvin, J. Photograph-based ergonomic evaluations using the Rapid Office Strain Assessment (ROSA). *Appl. Ergon.* **2016**, *52*, 317–324. [CrossRef] [PubMed]
9. Moon, G.; Chang, J.Y.; Lee, K.M. Camera Distance-Aware Top-Down Approach for 3D Multi-Person Pose Estimation from a Single RGB Image. In Proceedings of the International Conference on Computer Vision, Souel, Korea, 27 October–3 November 2019. [CrossRef]
10. Alexander, T.; Szegedy, C. Deeppose: Human pose estimation via deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1799–1807.
11. Wei, S.-E.; Ramakrishna, V.; Kanade, T.; Sheikh, Y. Convolutional Pose Machines. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 4724–4732. [CrossRef]
12. Liu, H.; Fang, S.; Zhang, Z.; Li, D.; Lin, K.; Wang, J. MFD Net: Collaborative Poses Perception and Matrix Fisher Distribution for Head Pose Estimation. *IEEE Trans. Multimed.* **2021**, *24*, 2449–2460. [CrossRef]
13. Liu, H.; Liu, T.; Zhang, Z.; Sangaiah, A.K.; Yang, B.; Li, Y. ARHPE: Asymmetric Relation-Aware Representation Learning for Head Pose Estimation in Industrial Human–Computer Interaction. *IEEE Trans. Ind. Inform.* **2022**, *18*, 7107–7117. [CrossRef]
14. Alejandro, N.; Yang, K.; Deng, J. Stacked hourglass networks for human pose estimation. In Proceedings of the European Conference on Computer Vision 2016, Amsterdam, The Netherlands, 23–28 June 2016.
15. Su, K.; Yu, D.; Xu, Z.; Geng, X.; Wang, C. Multi-Person Pose Estimation with Enhanced Channel-Wise and Spatial Information. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hoboken, NJ, USA, 10–13 June 2019; pp. 5667–5675. [CrossRef]
16. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X.; et al. Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 3349–3364. [CrossRef] [PubMed]
17. Papandreou, G.; Zhu, T.; Kanazawa, N.; Toshev, A.; Tompson, J.; Bregler, C.; Murphy, K. Towards accurate multi-person pose estimation in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4903–4911.
18. Fang, H.-S.; Xie, S.; Tai, Y.-W.; Lu, C. RMPE: Regional Multi-person Pose Estimation. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2353–2362.
19. He, K.; Gkioxari, G.; Dollar, P.; Ross, B. Mask R-CNN. In Proceedings of the ICCV, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
20. Huang, J.; Zhu, Z.; Guo, F.; Huang, G. The Devil Is in the Details: Delving Into Unbiased Data Processing for Human Pose Estimation. In Proceedings of the CVPR 2020, online, 14–19 June 2020; pp. 5699–5708. [CrossRef]
21. Zhang, F.; Zhu, X.; Dai, H.; Ye, M.; Zhu, C. Distribution-Aware Coordinate Representation for Human Pose Estimation. In Proceedings of the CVPR 2020, Online, 14–19 June 2020; pp. 7091–7100. [CrossRef]
22. Alejandro, N.; Huang, Z.; Deng, J. “Associative Embedding: End-to-End Learning for Joint Detection and Group-Ing.” *Advances in Neural Information Processing Systems* **30** (2017). Available online: <https://proceedings.neurips.cc/paper/2017/file/8edd72158ccd2a879f79cb2538568fdc-Paper.pdf> (accessed on 6 August 2022).
23. Bin, Y.; Cao, X.; Chen, X.; Ge, Y.; Tai, Y.; Wang, C.; Li, J.; Huang, F.; Gao, C.; Sang, N. Adversarial Semantic Data Augmentation for Human Pose Estimation. In Proceedings of the ECCV 2020, Glasgow, UK, 23–28 August 2020; pp. 606–622. [CrossRef]
24. Kocabas, M.; Karagoz, S.; Akbas, E. MultiPoseNet: Fast Multi-Person Pose Estimation Using Pose Residual Network. In Proceedings of the ECCV 2018, Munich, Germany, 8–14 September 2018; pp. 437–453. [CrossRef]
25. Li, J.; Wang, C.; Zhu, H.; Mao, Y.; Fang, H.; Lu, C. Crowdpose: Efficient crowded scenes pose estimation and a new benchmark. In Proceedings of the CVPR 2019, Long Beach Convention, CA, USA, 16–20 June 2019.
26. Qiu, L.; Zhang, X.; Li, Y.; Li, G.; Wu, X.; Xiong, Z.; Han, X.; Cui, S. *Peeking into Occluded Joints: A Novel Framework for Crowd Pose Estimation*; Springer: Cham, Switzerland, 2020; pp. 488–504. [CrossRef]
27. Zhou, L.; Chen, Y.; Gao, Y.; Wang, J.; Lu, H. Occlusion-Aware Siamese Network for Human Pose Estimation. In Proceedings of the ECCV 2020, Glasgow, UK, 23–28 August 2020; pp. 396–412. [CrossRef]

28. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [[CrossRef](#)]
29. Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the ACML 2015, Lille, France, 6–11 July 2015.
30. Abobakr, A.; Nahavandi, D.; Iskander, J.; Hossny, M.; Nahavandi, S.; Smets, M. A kinect-based workplace postural analysis system using deep residual networks. In Proceedings of the ISSE 2017, Brussels, Belgium, 15 November 2017. [[CrossRef](#)]
31. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.E.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 172–186. [[CrossRef](#)] [[PubMed](#)]
32. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.E.; Sheikh, Y. Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2017, Honolulu, HI, USA, 21–26 July 2017.
33. Bogo, F.; Black, M.J.; Loper, M.; Romero, J. Detailed Full-Body Reconstructions of Moving People from Monocular RGB-D Sequences. In Proceedings of the IEEE international conference on computer vision 2015, Washington, DC, USA, 2015; pp. 2300–2308. [[CrossRef](#)]
34. Yu, T.; Zhao, J.; Zheng, Z.; Guo, K.; Dai, Q.; Li, H.; Pons-Moll, G.; Liu, Y. DoubleFusion: Real-Time Capture of Human Performances with Inner Body Shapes from a Single Depth Sensor. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 2523–2539. [[CrossRef](#)] [[PubMed](#)]
35. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common Objects in Context. In *Computer Vision—ECCV 2014. ECCV 2014. Lecture Notes in Computer Science*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer: Cham, Switzerland, 2014; Volume 8693. [[CrossRef](#)]
36. Ionescu, C.; Papava, D.; Olaru, V.; Sminchisescu, C. Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *36*, 1325–1339. [[CrossRef](#)] [[PubMed](#)]
37. Lee, K.; Lee, I.; Lee, S. Propagating LSTM: 3D Pose Estimation Based on Joint Interdependency. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 123–141. [[CrossRef](#)]
38. Xu, J.; Yu, Z.; Ni, B.; Yang, J.; Yang, X.; Zhang, W. Deep kinematics analysis for monocular 3d human pose estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020, Seattle, WA, USA, 13–19 June 2020.
39. Chen, X.; Lin, K.-Y.; Liu, W.; Qian, C.; Lin, L. Weakly-Supervised Discovery of Geometry-Aware Representation for 3D Human Pose Estimation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2019, Long Beach, CA, USA, 15–20 June 2019; pp. 10887–10896. [[CrossRef](#)]
40. Martinez, J.; Hossain, M.; Rayat, I.; Romero, J.; Little, J.J. View invariant 3D human pose estimation. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 4601–4610.
41. Andriluka, M.; Pishchulin, L.; Gehler, P.; Schiele, B. 2d human pose estimation: New benchmark and state of the art analysis. In Proceedings of the IEEE Conference on computer Vision and Pattern Recognition 2014, Columbus, OH, USA, 23–28 June 2014.
42. Pavllo, D.; Feichtenhofer, C.; Grangier, D.; Auli, M. 3d human pose estimation in video with temporal convolutions and semi-supervised training. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
43. Imtiaz Hossain, M.R.; Little, J.L. Exploiting temporal information for 3d human pose estimation. In Proceedings of the ECCV 2018, Munich, Germany, 8–14 September 2018; pp. 69–86.
44. Malik, J.; Abdelaziz, I.; Elhayek, A.; Shimada, S.; Ali, S.A.; Golyanik, V.; Theobalt, C.; Stricker, D. HandVoxNet: Deep Voxel-Based Network for 3D Hand Shape and Pose Estimation From a Single Depth Map. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 7111–7120. [[CrossRef](#)]
45. Karhu, O.; Kansil, P.; Kuorinka, I. Correcting working postures in industry: A practical method for analysis. *Appl. Ergon.* **1977**, *8*, 199–201. [[CrossRef](#)]
46. Rahman, A.; Nasrull, M.; Rani, M.R.A.; Rohani, J.M. WERA: An observational tool develop to investigate the physical risk factor associated with WMSDs. *J. Hum. Ergol.* **2011**, *40*, 19–36.
47. Rahman, M.N.A.; Rahman, S.A.A.; Ismail, A.E.; Sadikin, A. Inter-Rater Reliability of the New Observational Method for Assessing an Exposure to Risk Factors Related to Work-Related Musculoskeletal Disorders (WMSDs). *MATEC Web Conf.* **2017**, *135*, 00024. [[CrossRef](#)]
48. Cheng, B.; Xiao, B.; Wang, J.; Shi, H.; Huang, T.S.; Zhang, L. HigherHRNet: Scale-Aware Representation Learning for Bottom-Up Human Pose Estimation. In Proceedings of the CVPR 2020, online, 14–19 June 2020; pp. 5385–5394. [[CrossRef](#)]
49. Luopajarvi, T. Ergonomic analysis of workplace and postural load. Ergonomics: The physiotherapist in the workplace. In Proceedings of the AIP Conference 2017, Yogyakarta, Indonesia, 9–10 November 2017; pp. 51–78.
50. Yu, C. Occlusion-aware networks for 3d human pose estimation in video. In Proceedings of the IEEE/CVF International Conference on Computer Vision 2019, Seoul, Korea, 27 October–2 November 2019.
51. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [[CrossRef](#)]
52. Burt, S.; Punnett, L. Evaluation of interrater reliability for posture observations in a field study. *Appl. Ergon.* **1999**, *30*, 121–135. [[CrossRef](#)]
53. Martinez, J.; Hossain, R.; Romero, J.; Little, J.J. A simple yet effective baseline for 3d human pose estimation. In Proceedings of the IEEE International Conference on Computer Vision 2017, Venice, Italy, 22–29 October 2017.