Article

Preliminary Validation of a Low-Cost Motion Analysis System Based on RGB Cameras to Support the Evaluation of Postural Risk Assessment

Thomas Agostinelli [1], Andrea Generosi [1], Silvia Ceccacci [1,*], Riccardo Karim Khamaisi [2], Margherita Peruzzini [2] and Maura Mengoni [1]

[1] Department of Industrial Engineering and Mathematical Science (DIISM), Università Politecnica delle Marche, 60131 Ancona, Italy; thomas.agostinelli@gmail.com (T.A.); a.generosi@univpm.it (A.G.); m.mengoni@univpm.it (M.M.)
[2] Department of Engineering "Enzo Ferrari" (DIEF), Università degli Studi di Modena e Reggio Emilia, 41125 Modena, Italy; riccardokarim.khamaisi@unimore.it (R.K.K.); margherita.peruzzini@unimore.it (M.P.)
[*] Correspondence: s.ceccacci@univpm.it

Featured Application: We introduce a motion capture tool that uses at least one RGB-camera, exploiting an open-source deep learning model with low computational requirements, already used to im-plement mobile apps for mobility analysis. Experimental results suggest the suitability of this tool to perform posture analysis aimed at assessing the RULA score, in a more efficient way.

Abstract: This paper introduces a low-cost and low computational marker-less motion capture system based on the acquisition of frame images through standard RGB cameras. It exploits the open-source deep learning model CMU, from the tf-pose-estimation project. Its numerical accuracy and its usefulness for ergonomic assessment are evaluated by a proper experiment, designed and performed to: (1) compare the data provided by it with those collected from a motion capture golden standard system; (2) compare the RULA scores obtained with data provided by it with those obtained with data provided by the Vicon Nexus system and those estimated through video analysis, by a team of three expert ergonomists. Tests have been conducted in standardized laboratory conditions and involved a total of six subjects. Results suggest that the proposed system can predict angles with good consistency and give evidence about the tool's usefulness for ergonomist.

Keywords: motion capture; ergonomic risk assessment; industrial ergonomics; postural analysis; RULA

1. Introduction

Nowadays, reducing the risk of musculoskeletal diseases (MSDs) for workers of manufacturing industries is of paramount importance to reduce absenteeism on work, due to illnesses related to bad working conditions, and to improve process efficiency in assembly lines. One of the main goals of Industry 4.0 is to find solutions to put workers in human-suitable working conditions, improving the efficiency and productivity of the factory [1]. However, we are still far from this goal, as shown in the 2019 European Risk Observatory Report [2]: reported working-related MSDs are decreasing but remain too high (58% in 2015 against 60% in 2010), and if we consider the working-class aging, these reports can only get worse. According to the European Commission's Ageing Report 2015 [3], the employment rate of people over-55 will reach 59.8% in 2023 and 66.7% by 2060, because many born during the "baby boom" are getting older, life expectancy and retirement age are increasing, while the birth rate is decreasing. Solving this problem is of extreme importance. Working demand usually does not change with age, while this cannot be said for working capacity: in aging, physiological changes in perception, information processing and motor control reduce work capacity. The physical work capacity of a

65-year-old worker is about half that of a 25-year-old worker [4]. On the other hand, age-related changes in physiological function can be dampened by various factors including physical activity [5], so work capability is a highly individual variable.

In this context, industries will increasingly need to take the human variability into account and to predict the workers' behaviors, going behind the concept of "the worker" as a homogeneous group and monitoring the specific work-related risks more accurately, to implement more effective health and safety management systems to increase factory efficiency and safety.

To reduce ergonomic risks and promote the workers' well-being, considering the characteristics and performance of every single person, we need cost-effective, robust tools able to provide a direct monitoring of the working postures and a continuous ergonomic risk assessment along with the work activities. Moreover, we need to improve the workers' awareness about ergonomic risks and define the best strategies to prevent them. Several studies in ergonomics suggested that providing workers with ergonomic feedback can positively influence the motion of workers and decrease hazardous risk score values [6–8]. However, this goal still seems far from being achieved due to the lack of low-cost, continuous monitoring systems to be easily applied on the shop floor.

Currently, ergonomic risk assessment is mainly based on postural observation methods [9,10], such as: National Institute for Occupational Safety and Health (NIOSH) lifting equation [11], Rapid Upper Limb Assessment (RULA) [12], Rapid Entire Body Assessment (REBA) [13], Strain Index [14], and Occupational Repetitive Action (OCRA) [15]. They require the intervention of an experienced ergonomist who observed workers' actions, directly or by means of video recordings. Data collection required to compute the risk index is generally obtained through subjective observation or simple estimation of projected joint angles (e.g., elbow, shoulder, knee, trunk, and neck) by analyzing videos or pictures. This ergonomic assessment results to be costly and time-consuming [9], highly affects the intra- and inter-observer results variability [16] and may lead to low accuracy of such evaluations [17].

Several tools are available to automate the postural analysis process by calculating various risk indices, to make ergonomic assessment more efficient. They are currently embedded in the most widely used computer-aided design (CAD) packages (e.g., CATIA-DELMIA by Dassault Systèmes, Pro/ENGINEER by PTC Manikin or Tecnomatix/Jack by Siemens) and allow detailed human modeling based on digital human manikins, according to an analytical ergonomic perspective [18]. However, to perform realistic and reliable simulations, they require accurate information related to the kinematic of the worker's body (posture) [19].

Motion capture systems can be used to collect such data accurately and quantitatively. However, the most reliable systems commercially available, i.e., motion capture sensor-based (e.g., Xsense [20], Vicon Blue Trident [21]) and marker-based optical systems (e.g., Vicon Nexus [22], OptiTrack [23]) have important drawbacks, so that their use in real work environments is still scarce [9]. Indeed, they are expensive in terms of cost and setup time and have a limited application in a factory environment due to several constraints, ranging from lighting conditions to electro-magnetic interference [24]. Therefore, their use is currently limited to laboratory experimental setups [25], while they are not easy to manage on the factory shop floor. In addition, these systems can also be intrusive as they frequently require the use of wearable devices (i.e., sensors or markers) positioned on the workers' bodies according to proper specification [25] and following specific calibration procedures. These activities require the involvement of specialized professionals and are time consuming, so it is not easy to carry them out in real working conditions, on a daily basis. Moreover, marker-based optical systems greatly suffer from occlusion problems and need the installation of multiple-cameras, which could rarely be feasible in a working environment where space is always limited, and so they cannot be optimally placed.

In the last few years, to overcome these issues, several systems based on computer vision and machine learning techniques have been proposed.

The introduction on the market of low-cost body-tracking technologies, based on RGB-D cameras, such as the Microsoft Kinect®, has aroused great interest in many application fields such as: gaming and virtual reality [26], healthcare [27,28], natural user interfaces [29], education [30] and ergonomics [31–33]. Being an integrated device, it does not require calibration. Several studies evaluated its accuracy [34–36] and evaluated it in working environments and industrial contexts [37,38]. Their results suggest that Kinect may successfully be used for assessing the risk of the operational activities, where very high precision is not required, despite errors depending on the performed posture [39]. Since the acquisition is performed from a single point of view, the system suffers occlusion problems, which can induce large error values, especially in complex motions with auto-occlusion or if the sensor is not placed in front of the subject, as recommended in [36]. Using multiple Kinects can only partially solve these problems, as the quality of the depth images degrades with the number of Kinects concurrently running, due to IR emitter interference problems [40]. Moreover, RGB-D cameras are not as widely and cheaply acceptable as RGB, and their installation and calibration on the workspace is not a trivial task, because of ferromagnetic interference that can cause significant noise in the output [41].

In a working industrial environment, having motion capture working with standard RGB sensors (such as those embedded in smartphones or webcams) can represent a more viable solution. Several systems have been introduced in the last few years to enable real-time human pose estimation from video streaming provided by RGB cameras. Among them, OpenPose, developed by researchers of the Carnegie Mellon University [42,43], represents the first real-time multi-person system to jointly detect the human body, hands, face, and feet (137 keypoints estimation per person: 70-keypoints face, 25-keypoints body/foot and 2x21-keypoints hand) on a single image. It is an open-source software, based on Convolutional Neural Network (CNN) found in the OpenCV library [44] initially written in C++ and Caffe, and freely available for non-commercial use. Such a system does not seem to be significantly affected by occlusion problems, as it ensures body tracking even when several body joints and segments are temporarily occluded, so that only a portion of the body is framed in the video [18]. Several studies validated its accuracy by comparing one person's skeleton tracking results with those obtained from a Vicon system. All found a highly negligible relative limb positioning error index [45,46]. Many studies exploited OpenPose for several research purposes, including Ergonomics. In particular, several studies carried out both in a laboratory and in real life manufacturing environments, suggest that OpenPose is a helpful tool to support worker posture ergonomics assessment based on RULA, REBA, and OCRA [18,46–49].

However, deep learning algorithms used to enable people tracking using RGB images usually require hardware with high computational requirements, so it is essential to have a good CPU and GPU performance [50].

Recently, a newer open-source machine learning pose-estimation algorithm inspired from OpenPose, namely Tf-pose-estimation [51], has been released. It has been implemented using Tensorflow and introduced several variants that have some changes to the CNN structure it implements so that it enables real-time processing of multi-person skeletons also on the CPU or on low-power embedded devices. It provides several models, including a body model variant that is characterized by 18 key points and runs on mobile devices.

Given its low computational requirements, it has been used to implement mobile apps for mobility analysis (e.g., Lindera [52]) and to implement edge computing solutions for human behavior estimation [53], or for human posture recognition (e.g., yoga pose [54]). However, as far as we know, the suitability of this tool to support ergonomic risk assessment in the industrial context has not been assessed yet.

In this context, this paper introduces a new low-cost and low-computational markerless motion capture system, based on frame images from RGB cameras acquisition, and on their processing through the multi-person key points detection Tf-pose-estimation algorithm. To assess the accuracy of this tool, a comparison among the data provided

by it with those collected from a Vicon Nexus system, and those measured through a manual video analysis by a panel of three expert ergonomists, is performed. Moreover, to preliminary validate the proposed system for ergonomic assessment, RULA scores obtained with the data provided by it have been compared to (1) those measured by the expert ergonomists and (2) those obtained with data provided by the Vicon Nexus system.

2. Materials and Methods

2.1. The proposed Motion Analysis System

The motion analysis system based on RGB cameras (RGB-motion analysis system, RGB-MAS), here proposed, is conceptually based on that described in [18] and improves its features and functionalities as follows:

- New system based on the CMU model from the tf-pose-estimation project, computationally lighter than those provided by Openpose and therefore able to provide real-time processing requiring lower CPU and GPU performances.
- Addition of estimation of torso rotation relative to the pelvis and of head rotation relative to shoulders.
- Distinction between abduction, adduction, extension, and flexion categories in the calculation of the angles between body segments
- Person tracking is no longer based on K-Means clustering, as it was computationally heavy and not very accurate.
- Modification of the system architecture to ensure greater modularity and ability to work even with a single camera shot.

The main objective, using these tools, is to measure the angles between the main body segments that characterize the postures of one or more people framed by the cameras. The measurement is carried out starting from the recognition of the skeletal joints and following with the estimation of their position in a digital space. It is based on a modular software architecture (Figure 1), using algorithms and models of deep learning and computer vision, to analyze human subjects by processing videos acquired through RGB cameras.
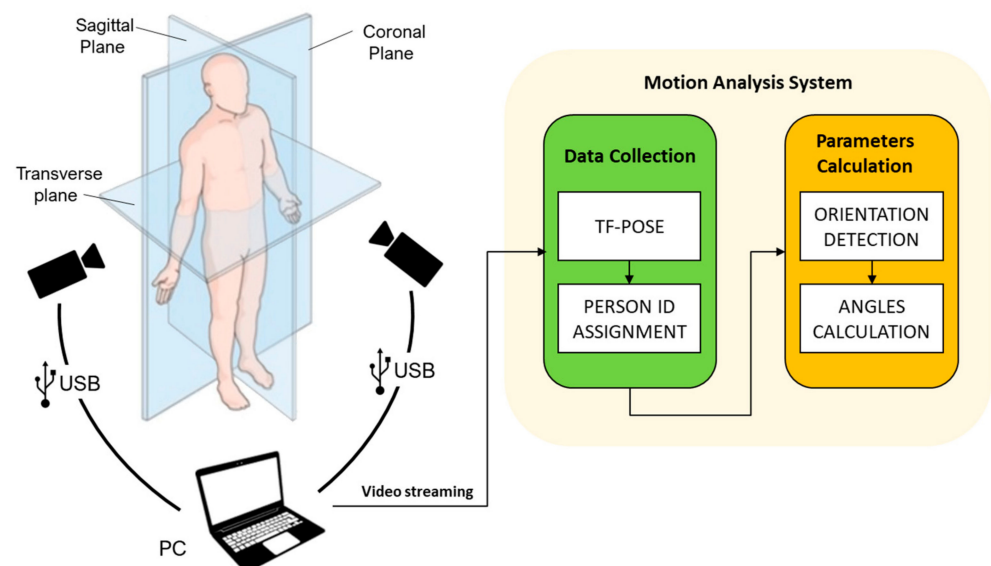
Figure 1. High-level software architecture and an example of possible hardware configuration.

The proposed system needs one or more video recordings, retrieved by pointing a camera parallel to the three anatomical planes (i.e., Sagittal Plane, Frontal Plane, and Transverse Plane) to track subjects during everyday work activities. In most cases, it is necessary to use at least two points of view, taken from the pelvis height, to achieve good accuracy: this trade-off guarantees a good compromise between the prediction accuracy and the system portability. For the calculation of the angles, the optimal prediction has been

evaluated when the camera directions are perpendicular to the person's sagittal and coronal planes. However, the system tolerates a deviation in the camera direction perpendicular to these planes in the range between $-45°$ and $+45°$. In this case, empirical tests evidenced that the system performs angle estimation with a maximum error equal to $\pm 10\%$. The accuracy of the skeletal landmark recognition tends to worsen the closer the orientation angle of the subject is to $\pm 90°$ to the reference camera, due to obvious perspective issues in a two-dimensional reference plane.

The system does not necessarily require that video recordings taken from different angles be simultaneously collected. The necessary frames can be recorded in succession, using only one camera. An important requirement is that each video must capture the subject during the entire working cycle.

Any camera with at least the following minimum requirements can be used:

- Resolution: 720 p.
- Distortion-free lenses: wide angle lenses should be avoided.

The PC(s) collects the images from the camera(s) and processes them through the Motion analysis software, which is characterized by two main modules (i.e., "Data Collection" and "Parameters Calculation" modules), which are described in detail below.
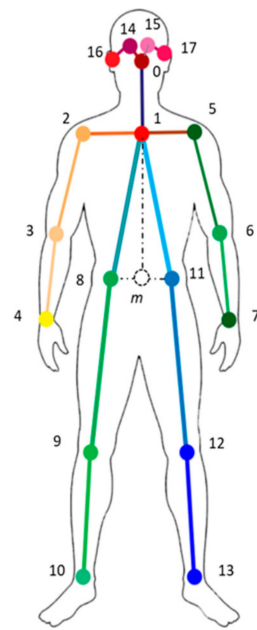
2.1.1. Data Collection

This module enables the analysis of the frames from the camera(s) to detect and track people presented in the frame. It uses models and techniques of deep learning and computer vision.

The deep learning model used to track the skeleton joints (i.e., key points) is based on the open-source project Tf-pose-estimation. This algorithm has been implemented using the C++ language and the Tensorflow framework. Tf-pose-estimation provides several models trained on many samples: CMU, mobilenet_thin, mobilenet_v2_thin, and mobilenet_v2_small. Several tests were done, and the CMU model was chosen as the optimal model for this project, looking for a compromise between accuracy and image processing time. It allows the identification of a total of 18 points of the body (Figure 2). When a video is analyzed, for any person detected in each processed frame, the system associates a skeleton and returns the following three data, for each single landmark:

- x: horizontal coordinate.
- y: vertical coordinate.
- c: confidence index in the range [0; 1].

For the subsequent parameter calculation, the system considers only the landmarks with a confidence index higher than 0.6.

To ensure univocal identification of the subjects, as long as they remain and move in the visual field of the camera(s), the system also associates a proper index to each person. A proper algorithm has been implemented to distinguish the key points belonging to different subjects that could eventually overlap each other when passing in front of the camera. It considers a spatial neighborhood for each detected subject through the key points detection model: if the subject maintains its position within that neighborhood in the subsequent frame, the identifier associated with it at the first recognition will remain the same. Collected data is then saved in a high-performance in-memory datastore (Redis), acting as a communication queue between the Data Collection and the Parameter Calculation modules.

Figure 2. tf-pose-estimation CMU model: acquired joints of body.

2.1.2. Parameter Calculation

This module, developed in Python, uses the C++ output from the Data Collection module to determine the person orientation with respect to the camera(s) and to calculate the 2D angles between the respective body segments.

The angles between the body segments are evaluated according to the predicted orientation. For each frame, they are computed from the coordinates (x, y) of the respective keypoints, by using proper algorithms. To estimate the angle between the segments i-j and j-k (considering the coordinates of the keypoints i, j and k) the following formulas are applied:

$$\theta = \left(\arccos\left(\frac{\gamma}{\delta}\right)\right) \times \left(\frac{180}{\pi}\right) \tag{1}$$

where γ is the scalar product between the vector formed by the first cathetus (i-j) and the one formed by the second cathetus (j-k) and δ is the cross product between the norms of the aforementioned vectors:

$$\gamma = \left(x_j - x_i\right) \times \left(x_k - x_j\right) + \left(y_j - y_i\right) \times \left(y_k - y_j\right) \tag{2}$$

$$\delta = \left(\left|\sqrt{\left(x_j - x_i\right)^2 + \left(y_j - y_i\right)^2}\right|\right) \times \left(\left|\sqrt{\left(x_k - x_j\right)^2 + \left(y_k - y_j\right)^2}\right|\right) \tag{3}$$

In the case of two cameras, it is necessary to consider that one camera among all will have the best view to predict some angles correctly: to this end, a special algorithm has been developed. Considering that the cameras are positioned at about the pelvis height, it performs a comparison between the distances interposed between the key points (corresponding to specific body segments) and the expected average anthropometric proportions of the human body reported in [55], which are calculated on a large dataset of individuals. In particular, it analyzes the ratio between the width of the shoulders (i.e., Euclidean distance between key points 2 and 5) and the length of the spine. The Euclidean distance between the key point 1, and the midpoint (m), between key points 8 and 11, is considered to estimate the spine length since the CMU model does not consider a pelvis key point. Based on the human body proportions reported in [55], such a ratio is estimated equal to 80%, when the subject is in front of the camera, and to 0% when the person turns at 90°. Consequently, the angle between the sagittal plane of the person and the direction of the

camera (α) can be estimated through the following equation, considering the length of the segments related to shoulder (x) and spine (l) measured in each frame:

$$\alpha = \arcsin\left(\frac{x}{l \cdot 0.8}\right) \quad (4)$$

To determine whether the camera is framing the person's right side or left side, the x coordinate of the key point 0, X_0, has been considered in relation to the average x coordinate of keypoints 1, 2, and 5, X_a. The person's side framed by the cameras is then determined according to the following algorithm:

$$X_0 - X_a \begin{cases} < 0, \text{ left} \\ > 0, \text{ right} \end{cases} \quad (5)$$

In case two video shots are available, recorded from angles settled at 90° to each other (e.g., respectively parallel to the frontal and the sagittal planes), it is then possible to refer the orientation estimation to a single reference system, using as origin the direction of the first camera (camera index 0).

The compute of each angle is then performed considering the frame coming from the camera view that better estimates it. For example, the elbow flexion/extension angle is calculated using the frame coming from the camera that has a lateral view of the person, while in case of abduction, the frame coming from the frontal camera is considered, as reported in Table 1.

Table 1. Body key points and camera views considered in the computation of the angles.

Angles	Considered Keypoints		Camera View	
	Left	Right	Frontal	Lateral
Neck flexion/extension	17-1-11	16-1-8		X
Shoulder abduction	11-5-6	8-2-3	X	
Shoulder flexion/extension	11-5-6	8-2-3		X
Elbow flexion/extension angle	5-6-7	2-3-4	X	X
Trunk flexion/extension	1-11-12	1-8-9		X
Knee bending angle	11-12-13	8-9-10		X

To evaluate which camera view is more convenient to measure the elbow flexion/extension angle, the extent of the shoulder abduction angle is considered: if it is less than 45°, the elbow flexion/extension angle is estimated considering the frames from the lateral camera. Otherwise, data provided by the frontal camera are used.

Finally, the system can estimate whether the person is rotating his/her head in respect to the shoulders or not, considering a proportion between the distance from the ear to the eyes (CMU keypoints 16-14 for the right side and 17-15 for the left side), and the length of shoulders (euclidean distance between CMU keypoints 2-5). A reference threshold for this calculated ratio is empirically estimated due to the lack of literature. Currently, this solution has been applied only when a person has an orientation between −30 and +30 degrees with respect to the camera (i.e., in a frontal position). A similar approach has been considered to detect a torso rotation, calculating the proportion between the segment from the left to the right shoulder (CMU key points 2-5), and the pelvis one (CMU key points 8-11).

2.2. Experimental Case Study

#### 2.2.1. Experimental Procedure

Tests had carried out in the "X-in the Loop Simulation Lab" (XiLab) at the University of Modena and Reggio Emilia. Participants in the experiment received informed consent prior to accessing the lab to take part in the test.

A total of 6 subjects i.e., 2 females (age: median 31 years, IQR 5.0 years; height: median 1.68 m, IQR 0.005 m; mass: median 69 kg, IQR 13.5 kg; BMI: median 24.74 kg/m^2, IQR

4.78 kg/m^2) and 4 males (age: median 30 years, IQR 4.0 years; height: median 1.86 m, IQR 0.0035 m; mass: median 78 kg, IQR 9.0 kg; BMI: median 22.31 kg/m^2, IQR 1.90 kg/m^2), were involved and asked to pose for five seconds, while recording by the two systems, in the following five different postures (chosen because they are very frequent in standard working procedures or ergonomic assessment research works), in the following order:

- T-pose: the subjects have to stand straight up, with their feet placed symmetrically and feet slightly apart and with their arms fully extended.
- Seated: subjects have to sit on a stool 70 cm in height, with the back straight, hands leaned on the knees, and feet on the floor.
- Standing Relaxed: the subjects have to stand comfortably facing straight ahead, with their feet placed symmetrically and feet slightly apart.
- Reach: the subjects must stand straight, feet well apart and with the arms stretched forward, simulating the act of grasping an object placed above their head.
- Pick up: the subjects have to pick up a box (dimensions 30.5 cm $\times$ 21 cm $\times$ 10.5 cm, weight 5 kg) from the floor, and raise it in front of them, keeping it at pelvic level.

An example of the analyzed postures can be found in Figure 3.
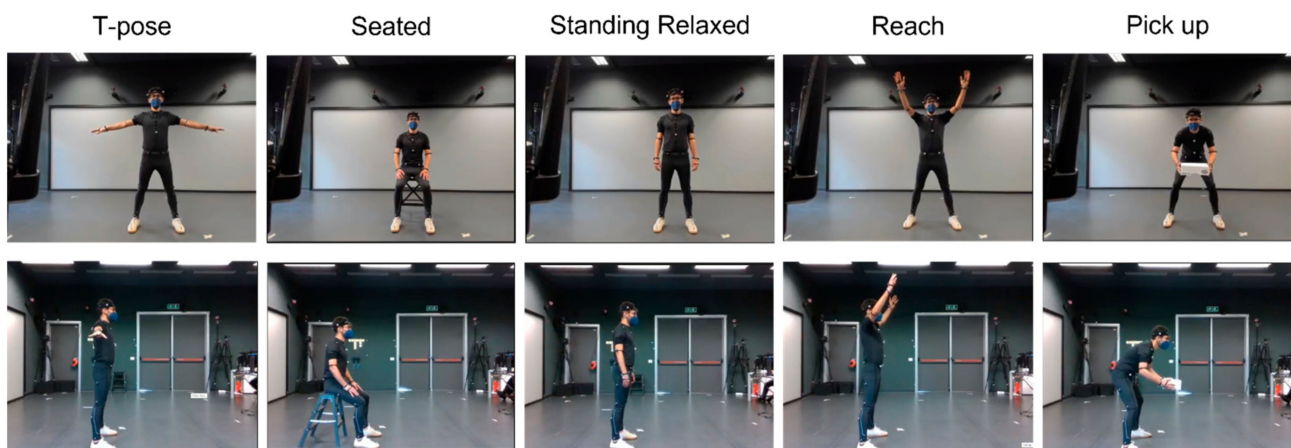
Figure 3. Postures assessed in this research paper.

Participants' postures were tracked using a Vicon Nexus system powered by 9 Vicon Bonita 10 optical cameras. Cameras were distributed in the space in the most symmetrical configuration possible (Figure 4) to cover the entire working volume. Participants had to stay in the middle of the system acquisition space. They were equipped with a total of 35 reflective markers, positioned on the whole body according to PlugInGait Full Body model specification defined in the Vicon PlugInGait documentation (Figure 5). Vicon Nexus session has been performed on a Dell Precision T1650 workstation with an Intel(R) Xeon(R) CPU E3-1290 V2 at 3.70 GHz, 4 Core(s), with 32 GB RAM, and an NVIDIA Quadro 2000 GPU, running Windows 10 Pro.

System calibration was carried out at the beginning of the experiment. Before starting the experiment, PluginGait (PiG) biomechanical models have been defined for the subjects (i.e., one for each) according to their anthropometric parameters.

The video capturing for the RGB-MAS was carried out by two Logitech BRIO 4K Ultra HD USB cameras, with a setting video streaming/recording configuration of 1080 p and 30 fps and a 52 vertical and 82 horizontal degree field of view. They were placed at 1.2 m high from the ground (the pelvis height of the subjects) and angled 90 degrees to each other. Both cameras were mounted on tripods to ensure stability.

The system works regardless of the subject's position in the camera's field of view, but the cropped image of the subject must consist of a sufficient number of pixels. This depends on the characteristics of the camera used (e.g., resolution, focal length). For example, considering the cameras chosen in this case, the system would work properly

when the user was positioned at no more than 7 m from the camera. Therefore, the first camera was placed in front of the subjects, at a distance of 2 m. The second one was at the right at a distance of 3.5 m, to ensure that the cameras correctly capture the subjects' entire body. Figure 4 shows the overall layout.
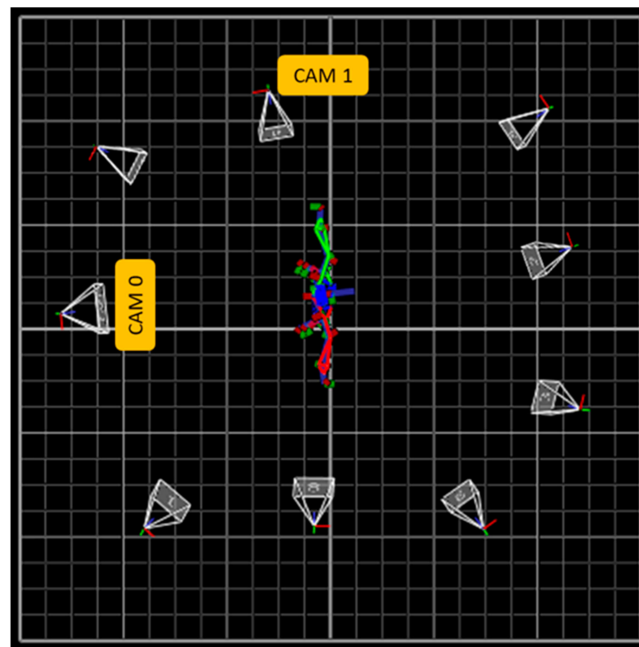

Figure 4. Experimental camera layout for RGB (those labeled in yellow) and Vicon system.

Figure 5. Markers' layout according to the PlugInGait full body model.

Stream videos were processed through a PC workstation with an Intel(R) Core (TM) i7-7700K CPU at 4.20 GHz and 32 GB RAM, and a GTX 1080 Ti GPU, running Windows 10 Pro.

Postures recording was carried out simultaneously using the two different systems to ensure numerical accuracy and to reduce inconsistencies. The camera frame rate showed to be consistent and constant along with the experiment for both systems.

2.2.2. Data Analysis

Angles extracted by Vicon PiG biomechanical model are compared with those predicted by the proposed RGB-MAS. To provide a better understanding, the considered angles respectively measured by the two systems are reported in Table 2.

Table 2. Pairs of angles that have been compared between Vicon and RGB systems.

Pairs of Angles Compared between the Two Systems.	
Vicon	**RGB-MAS**
Average between L and R neck flexion/extension	Neck flexion/extension
L/R shoulder abduction/adduction Y component	L/R shoulder abduction
L/R shoulder abduction/adduction X component	L/R shoulder flexion/extension
L/R elbow flexion/extension	L/R elbow flexion/extension
Average between L and R spine flexion/extension	Trunk flexion/extension
L/R knee flexion/extension	L/R knee bending angle

Resulted angles measured through these two systems are also compared with those manually extracted by expert ergonomists.

Shapiro–Wilk test is used to check the normality of the distribution of the error in all these analyses. Results evidence that the distributions follow a normal law for this experiment. Root mean square error (RMSE) is computed for the following condition:

$$RMSE_1 = \sqrt{\frac{\sum_{i=1}^{N}(RGB_i - MAN_i)^2}{N}} \tag{6}$$

$$RMSE_2 = \sqrt{\frac{\sum_{i=1}^{N}(VIC_i - MAN_i)^2}{N}} \tag{7}$$

$$RMSE_3 = \sqrt{\frac{\sum_{i=1}^{N}(RGB_i - VIC_i)^2}{N}} \tag{8}$$

where RGB_i is the ith angle measured by the RGB-MAS system, VIC_i the one measured by the Vicon system and, finally, MAN_i the angle measured manually.

Based on the collected angles, a Rapid Upper Limb Assessment (RULA) is performed manually, according to the procedure described in [12]. Then, RMSE is computed to compare the resulting RULA scores estimated according to the angles respectively predicted by the RGB-MAS and Vicon with that performed considering the angles estimated from the video analysis by the experts themself.

3. Results

Table 3 shows the RMSE comparison between the angles extracted from the RGB-MAS and the ones extracted by the Vicon system for a pure system-to-system analysis.

Table 3. RMSE values obtained comparing RGB-MAS angles with the Vicon ones.

	RMSE RGB-MAS vs. Vicon [°]			
	T-Pose	**Seated**	**Standing Relaxed**	**Reach**
Neck flexion/extension	6.83	16.47	19.21	9.58
Left shoulder abduction	12.66	45.16	13.07	45.46
Right shoulder abduction	11.64	50.66	7.93	43.23
Left shoulder flexion/extension	27.86	57.19	21.53	71.29
Right shoulder flexion/extension	33.73	52.90	57.88	82.93
Left elbow flexion/extension	7.13	16.90	21.15	27.05
Right elbow flexion/extension	5.46	13.19	22.11	53.30
Trunk flexion/extension	0.35	8.61	0.91	2.95
Left knee flexion/extention	2.39	46.25	7.38	24.76
Right knee flexion/extention	0.21	4.79	0.07	0.12

As it can be observed, angle predictions provided by the proposed system result in general lower accuracy, in the case of shoulder abduction and shoulder and elbow flexion/extension. The accuracy is particularly low if we consider the reach posture. This is probably because of perspective distortions.

However, the pickup posture could not be traced with the Vicon system, due to occlusion problems caused by the presence of the box that occludes some of the markers needed by it.

Table 4 allows easy comparison of the RMSE between angles respectively predicted by the RGB-MAS and the Vicon system, and those manually measured. These results suggest that the proposed system can be considered somehow feasible to support ergonomists doing their analysis.

Table 4. RMSE values obtained when comparing RGB-MAS angles with manually extracted ones.

	RMSE RGB-MAS vs. Manual [°]					RMSE Vicon vs. Manual [°]			
	T-Pose	Seated	Standing Relaxed	Reach	Pick Up	T-Pose	Seated	Standing Relaxed	Reach
Neck flexion/extension	9.19	13.10	19.63	6.87	28.05	8.26	15.47	8.64	8.32
Left shoulder abduction	6.90	44.13	7.54	49.68	8.03	7.29	7.52	13.00	38.11
Right shoulder abduction	6.67	47.00	7.82	51.65	7.55	6.35	6.54	7.09	38.08
Left shoulder flexion/extension	32.62	53.84	12.03	82.10	30.43	21.08	16.19	19.55	101.52
Right shoulder flexion/extension	50.91	50.70	70.02	71.31	28.01	23.14	15.66	21.43	110.20
Left elbow flexion/extension	3.60	15.03	14.21	36.16	15.83	6.21	7.08	15.56	16.35
Right elbow flexion/extension	2.20	9.67	18.63	25.45	10.26	5.60	17.97	10.35	39.00
Trunk flexion/extension	3.48	30.06	5.46	20.71	33.46	3.28	33.60	4.59	17.93
Left knee flexion/extention	3.81	53.12	6.82	22.73	30.10	1.98	18.85	2.68	8.39
Right knee flexion/extention	3.95	20.69	1.94	22.63	29.19	3.75	21.24	1.90	22.52

Figure 6 highlights the similarities and discrepancies between the angle predictions provided by RGB-MAS and Vicon system with respect to the manual analysis. It can be observed that the prediction provided by the RGB-MAS suffers from a wider variability compared to the reference system. As for the neck flexion/extension angle, the RGB-MAS system slightly overestimates the results compared to the Vicon system. This occurs more pronouncedly for the shoulder abduction and flexion/extension as well, especially when abduction and flexion both occur simultaneously. In Figure 7, the keypoint's locations and the skeleton are shown superimposed over the posture pictures. In particular, from the picture of the Pick Up posture, we can see that the small occlusion that caused problems to the Vicon system did not cause any consequences to the RGB-MAS system.

Moreover, high variability is also found for all the angles of the left-hand side of the body. Nevertheless, RGB-MAS accurately predicts the trunk flexion/extension and the right-hand side angles. This left-right inconsistency can be due to the lack of a left camera, and so the left-hand side angles are predicted with less confidence than their right-hand side counterparts.
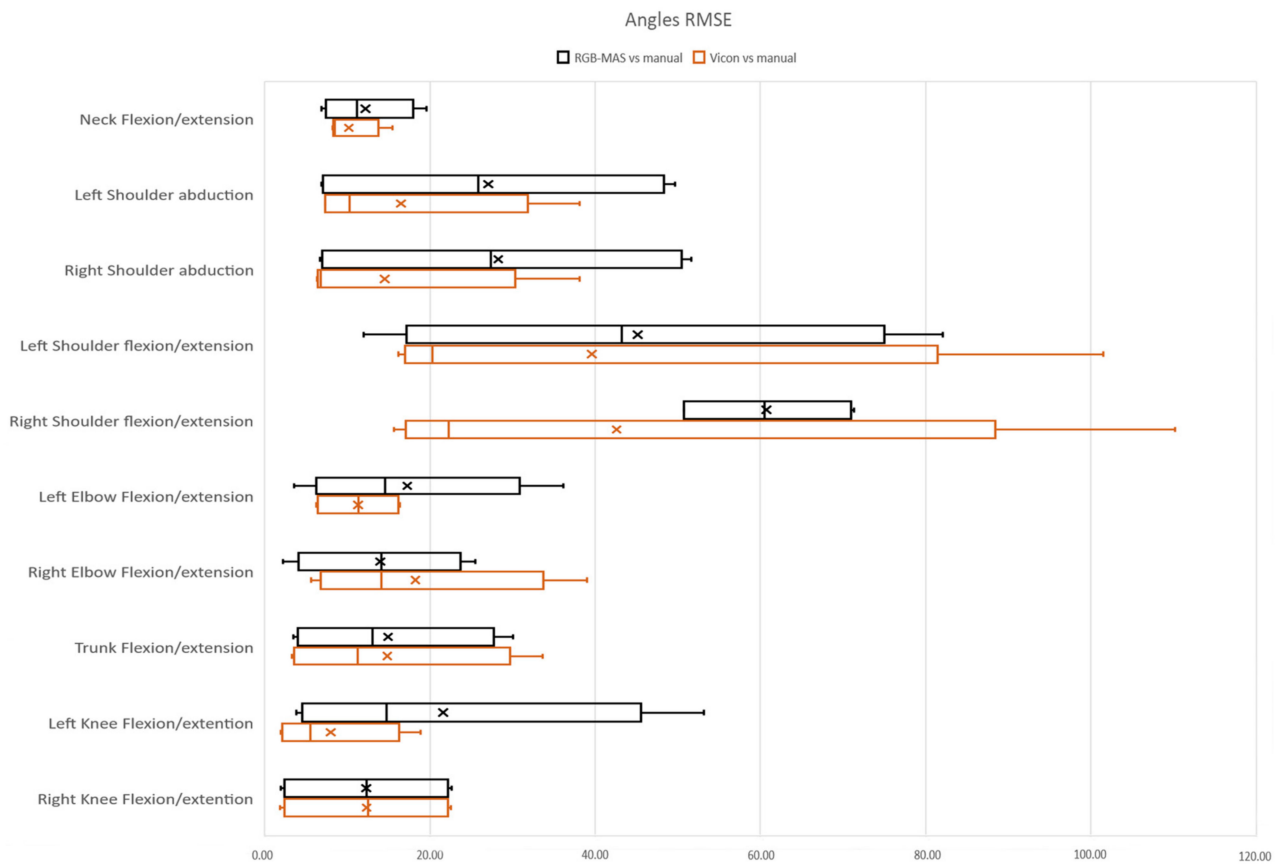
Figure 6. Graph comparing the RMSE values "RGB-MAS vs. manual" and "Vicon vs. manual".
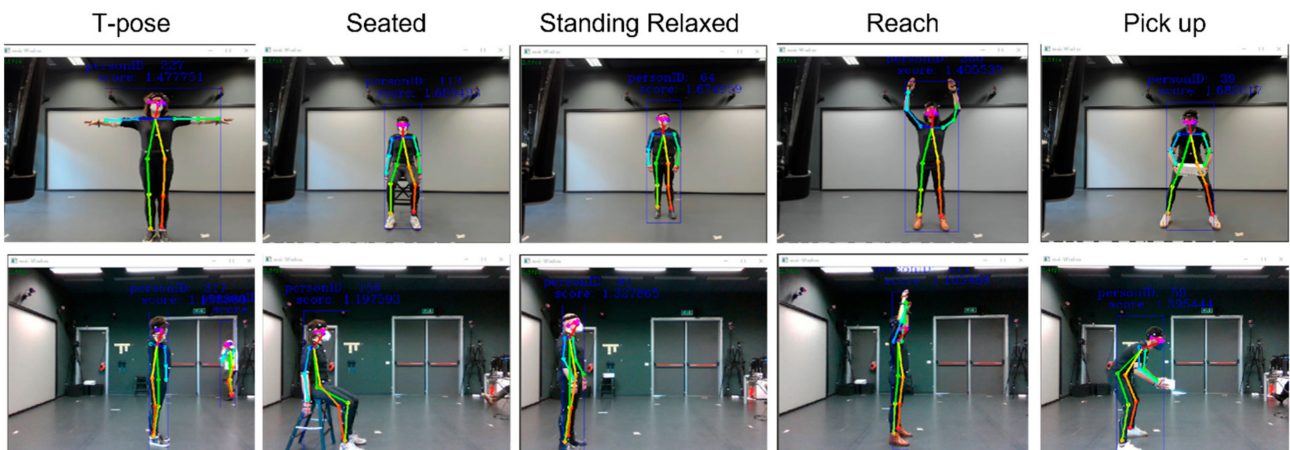

Figure 7. Body skeleton predicted by the RGB-MAS.

Table 5 shows the median RULA values obtained using each angle extraction method considered (i.e., manual measurement, RGB-MAS prediction, Vicon tracking). Table 6 shows the RMSE between the RULAs determined through manual angle measurement, and those calculated from the angles predicted by RGB-MAS and the Vicon system, respectively. The maximum RMSE between RGB-MAS and manual analysis is 1.35, while the maximum Vicon vs. manual RMSE is 1.78. Since the closer a value of RMSE is to zero, the better the prediction accuracy is, it can be observed that RGB-MAS compared to a most used manual analysis is generally able to provide a result closer to that of the Vicon systems. However, this should not be interpreted as a better performance of RGB-MAS than Vicon: the result that can be obtained, instead, is that the values provided by Vicon, in some cases,

are very different from both those of the RGB-MAS system and those estimated manually. This is because the result of the Vicon system is not affected by estimation errors due to perspective distortions, which instead occur with the other two systems. Ultimately, the RGB-MAS system can provide estimates that are very similar to those obtained from a manual extraction, although the system's accuracy is poor compared to the Vicon.

Table 5. RULA median scores for the three angle extraction methods and corresponding level of MSD risk (i.e., green = negligible risk; yellow = low risk; orange = medium risk).

| | Median | | | | | |
| | Manual | | RGB-MAS | | Vicon | |
	Left	Right	Left	Right	Left	Right
T-Pose	3.00	3.00	3.00	3.00	3.00	3.00
Relaxed	2.50	2.50	3.00	3.00	3.00	3.00
Sit	4.00	4.00	4.00	4.00	4.00	4.00
Reach	4.50	4.50	4.00	4.00	3.00	3.00
Pickup	5.50	5.50	6.00	6.00	-	-

Table 6. RMSE (+SD) values obtained when respectively comparing RGB-MAS and the Vicon RULA with the manual one.

| | RULA RMSE (+SD) | | | |
| | RGB-MAS vs. Manual | | Vicon vs. Manual | |
	Left	Right	Left	Right
T-Pose	0.00 (0.58)	1.00 (0.75)	0.41 (0.37)	0.41 (0.37)
Relaxed	0.71 (0.00)	2.45 (0.37)	0.82 (0.37)	0.82 (0.37)
Sit	0.58 (0.75)	1.41 (0.76)	0.71 (0.69)	0.82(0.58)
Reach	1.35 (1.07)	1.35 (0.82)	1.78 (0.76)	1.78 (0.76)

As can be seen from Table 5, the risk indices calculated from the angles provided by the RGB-MAS system, those evaluated from manually measured angles, and those provided by Vicon belong to the same ranges. The only exception can be found for the Reach posture, where the Vicon underestimated the scores of a whole point. Despite the overestimations in angle prediction, the RULA score evaluation made considering the extension of the angles within particular risk ranges filters out noise and slight measurement inaccuracies. This leads to RULA scores that are slightly different in numbers but can be considered almost the same in terms of risk ranges.

4. Discussion

This paper aims to introduce a novel tool to help ergonomists in ergonomic risk assessments by automatically extracting angles from video acquisition, in a quicker way than the traditional one. Its overall systematic reliability and numerical accuracy are assessed by comparing the tool's performance in ergonomics evaluation with the one obtained by standard procedures, representing the gold standard in the context.

Results suggest that it generally provides a good consistency in predicting the angles from the front camera and a slightly less accuracy with the lateral one, with a broader variability than the Vicon. However, in most cases, the average and median values are relatively close to the reference one. This apparent limitation should be analyzed in the light of the setup needed to obtain this data; in fact, by using only two cameras (instead of the nine ones the Vicon needs), we obtained reliable angles proper to compute a RULA score.

Although it has greater accuracy than the proposed tool, the Vicon requires installing a vast number of cameras (at least six), precisely positioned in the space, to completely cover the work area and ensure the absence of occlusions. In addition, such a system requires calibration and forces workers to wear markers in precise positions. However, when

performing a manual ergonomic risk assessment in a real working environment, given the constraints typically presented, an ergonomist usually can collect videos from one or two cameras at most: the proposed RGB-MAS copes with this aspect providing predicted angles even from the blind side of the subject (like a human could do, but quicker), or when the subject results partially occluded.

As proof of this, it is worth noticing that the pickup posture, initially considered for its tendency to introduce occlusion, was then discarded from the comparison just for the occlusion leading to the lack of data from the Vicon system, while no problems seemed to arise with the RGB-MAS.

In addition, the RMSE values obtained comparing the RGB-MAS RULA scores with the manual one showed tighter variability than the same values resulting from the comparison between the RULA scores estimated through the Vicon and the manual analysis. This suggests that the RGB-MAS can be helpful to fruitfully support ergonomists to estimate the RULA score on a first exploratory evaluation. The proposed system can extract angles with a numerical accuracy comparable to one of the reference systems, at least in a controlled environment such as a laboratory. The next step will be to test its methodological reliability and instrumental feasibility in a real working environment, where a Vicon-like system cannot be introduced due to its limitations (e.g., installation complexity, calibration requirements, occlusion sensitivity).

Study Limitations

This study provides the results of a first assessment of the proposed system, with the aim to measure its accuracy and to preliminary determine its utility for ergonomic assessment. Many studies should be carried out to fully understand its practical suitability to be used for ergonomic assessment in real working environments. The experiment was conducted only in the laboratory and not in a real working environment. This limits the study results. Therefore, it did not allow the researchers to evaluate the instrument's sensitivity to any changes in lighting or unexpected illumination conditions (e.g., glares or reflections). Further studies are needed to fully evaluate the implementation constraints of the proposed system in a real working environment.

In addition, the study is limited to evaluating the RULA risk index related to static postures only. Further studies will be needed to evaluate the possibility of using the proposed system for the acquisition of data necessary for other risk indexes (e.g., REBA, OCRA), also considering dynamic postures.

Another limitation is that the experiment conducted did not entirely evaluate the proposed system functionalities in conditions of severe occlusion (e.g., as could happen when the workbench partially covers the subject). Despite results evidenced that the proposed system, unlike Vicon, does not suffer from minor occlusion (i.e., due to the presence of a box during a picking operation), further studies are needed to accurately assess the sensitivity of the proposed system with different levels of occlusion.

Another limitation is the small number of subjects involved in the study. A small group of subjects was involved, with limited anthropometric variation, assuming that the tf-pose-estimation model was already trained on a large dataset. Further studies will need to confirm whether anthropometric variations affect the results (e.g., whether and how the BMI factor may affect the estimated angle accuracy).

5. Conclusions

This work proposes a valuable tool, namely RGB motion analysis system (RGB-MAS), to make a more efficient, and affordable ergonomic risk assessment. Our scope was to aid ergonomists in saving up time doing their job while maintaining highly reliable results. The lengthy part of their job is manually extracting human angles from video analysis based on video captures, by analyzing how ergonomists carry out an RULA assessment. In this context, the paper proposed a system able to speed up angle extraction and RULA calculation.

The validation in the laboratory shows the promising performance of the system, suggesting its possible suitability also in real working conditions (e.g., picking activities in the warehouse or manual tasks in the assembly lines), to enable the implementation of more effective health and safety management systems in the future, so as to improve the awareness of MSDs and to increase the efficiency and safety of the factory.

Overall, experimental results suggested that the RGB-MAS can be useful to support ergonomists to estimate the RULA score, providing results comparable to those estimated by ergonomic experts. The proposed system allows ergonomists and companies to reduce the cost necessary to perform ergonomic analysis, due to decreasing time for risk assessment. This competitive advantage makes it appealing not only to large-sized enterprises, but also to small and medium-sized enterprises, wishing to improve the working conditions of their workers. The main advantages of the proposed tool are: the ease of use, the wide range of scenarios where it can be installed, its full compatibility with every RGB commercially available camera, no-need calibration, low CPU and GPU performance requirements (i.e., it can process video recordings in a matter of seconds by using a common laptop), and low cost.

However, according to the experimental results, the increase in efficiency that the system allows comes at the expense of small errors in angle estimation and ergonomic evaluation: since the proposed system is not based on any calibration procedure and is still affected by perspective distortion problems, it obviously does not reach the accuracy of the Vicon. Nonetheless, if it is true that the Vicon system is to be considered as the absolute truth as far as accuracy is concerned, it is also true that using it in a real working environment is actually impossible, since it greatly suffers problem occlusion (even the presence of an object such as a small box can determine the loss of body tracking), and requires:

- A high amount of highly expensive cameras, placed in the space in a way that is impracticable in a real work environment.
- A preliminary calibration procedure.
- The use of wearable markers may invalidate the quality of the measurement as they are invasive.

Future studies should aim to improve the current functionalities of the proposed system. Currently, the system is incapable of automatically computing RULA scores. A spreadsheet based on the derived angles is filled to obtain them. However, it should not be difficult to implement such functionality. In particular, future studies should be focused to implement a direct stream of the angles extracted by the RGB-MAS system to a structured, ergonomic risk assessment software (e.g., Siemens Jack) to animate a virtual mannikin, again automatically obtaining RULA scores.

Moreover, the proposed system cannot predict hand- and wrist-related angles: further research might cope with this issue and try to fill the gap. For example, possible solutions can be those proposed in [56,57].

For a broader application of the proposed RGB-MAS system, other efforts should be made to improve the angles prediction accuracy.

Moreover, the main current issue is that it is not always possible to correctly predict shoulder abduction and flexion angles with non-calibrated cameras, e.g., when the arms are simultaneously showing flexion in the lateral plane and abduction in the frontal plane. This comes from the fact that, at the moment, there is no spatial correlation between the two cameras: the reference system is not the same for both, so it is not possible to determine 3D angles. Thus, another topic for future work may cover the development of a dedicated algorithm to correlate the spatial position of the cameras one to each other. In addition, such an algorithm should provide a (real-time) correction to effectively manage the inevitable perspective distortion introduced by the lenses, to improve the system accuracy. However, all of this would require the introduction of a calibration procedure that would slow down the implementation of the system in real working places.

References

1. Badri, A.; Boudreau-Trudel, B.; Souissi, A.S. Occupational health and safety in the industry 4.0 era: A cause for major concern? *Saf. Sci.* **2018**, *109*, 403–411. [CrossRef]
2. European Agency for Safety and Health at Work. Work-Related Musculoskeletal Disorders: Prevalence, Costs and Demographics in the EU. EU-OSHA. Available online: https://osha.europa.eu/en/publications/msds-facts-and-figures-overview-prevalence-costs-and-demographics-msds-europe/view (accessed on 5 July 2021).
3. European Commission. The 2015 Ageing Report: Economic and Budgetary Projections for the 28 EU Member State. Available online: https://ec.europa.eu/economy_finance/publications/european_economy/2015/pdf/ee3_en.pdf (accessed on 5 July 2021).
4. Ilmarinen, J. Physical requirements associated with the work of aging workers in the European Union. *Exp. Aging Res.* **2002**, *28*, 7–23. [CrossRef] [PubMed]
5. Kenny, G.P.; Groeller, H.; McGinn, R.; Flouris, A.D. Age, human performance, and physical employment standards. *Appl. Physiol. Nutr. Metab.* **2016**, *41*, S92–S107. [CrossRef] [PubMed]
6. Battini, D.; Persona, A.; Sgarbossa, F. Innovative real-time system to integrate ergonomic evaluations into warehouse design and management. *Comput. Ind. Eng.* **2014**, *77*, 1–10. [CrossRef]
7. Mengoni, M.; Ceccacci, S.; Generosi, A.; Leopardi, A. Spatial Augmented Reality: An application for human work in smart manufacturing environment. *Procedia Manuf.* **2018**, *17*, 476–483. [CrossRef]
8. Vignais, N.; Miezal, M.; Bleser, G.; Mura, K.; Gorecky, D.; Marin, F. Innovative system for real-time ergonomic feedback in industrial manufacturing. *Appl. Ergon.* **2013**, *44*, 566–574. [CrossRef]
9. Lowe, B.D.; Dempsey, P.G.; Jones, E.M. Ergonomics assessment methods used by ergonomics professionals. *Appl. Ergon.* **2019**, *81*, 10. [CrossRef]
10. Ceccacci, S.; Matteucci, M.; Peruzzini, M.; Mengoni, M. A multipath methodology to promote ergonomics, safety and efficiency in agile factories. *Int. J. Agil. Syst. Manag.* **2019**, *12*, 407–436. [CrossRef]
11. Snook, S.H.; Ciriello, V.M. The design of manual handling tasks: Revised tables of maximum acceptable weights and forces. *Ergonomics* **1991**, *34*, 1197–1213. [CrossRef]
12. 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]
13. Hignett, S.; McAtamney, L. Rapid entire body assessment (REBA). *Appl. Ergon.* **2000**, *31*, 201–205. [CrossRef]
14. Moore, J.S.; Garg, A. The strain index: A proposed method to analyze jobs for risk of distal upper extremity disorders. *Am. Ind. Hyg. Assoc. J.* **1995**, *56*, 443. [CrossRef]
15. Occhipinti, E. OCRA: A concise index for the assessment of exposure to repetitive movements of the upper limbs. *Ergonomics* **1998**, *41*, 1290–1311. [CrossRef] [PubMed]
16. Burdorf, A.; Derksen, J.; Naaktgeboren, B.; Van Riel, M. Measurement of trunk bending during work by direct observation and continuous measurement. *Appl. Ergon.* **1992**, *23*, 263–267. [CrossRef]

17. Fagarasanu, M.; Kumar, S. Measurement instruments and data collection: A consideration of constructs and biases in ergonomics research. *Int. J. Ind. Ergon.* **2002**, *30*, 355–369. [CrossRef]

18. Altieri, A.; Ceccacci, S.; Talipu, A.; Mengoni, M. A Low Cost Motion Analysis System Based on RGB Cameras to Support Ergonomic Risk Assessment in Real Workplaces. In Proceedings of the ASME 2020 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, American Society of Mechanical Engineers Digital Collection, St. Louis, MO, USA, 17–19 August 2020. [CrossRef]

19. De Magistris, G.; Micaelli, A.; Evrard, P.; Andriot, C.; Savin, J.; Gaudez, C.; Marsot, J. Dynamic control of DHM for ergonomic assessments. *Int. J. Ind. Ergon.* **2013**, *43*, 170–180. [CrossRef]

20. Xsense. Available online: https://www.xsens.com/motion-capture (accessed on 5 July 2021).

21. Vicon Blue Trident. Available online: https://www.vicon.com/hardware/blue-trident/ (accessed on 5 July 2021).

22. Vicon Nexus. Available online: https://www.vicon.com/software/nexus/ (accessed on 5 July 2021).

23. Optitrack. Available online: https://optitrack.com/ (accessed on 5 July 2021).

24. Manghisi, V.M.; Uva, A.E.; Fiorentino, M.; Gattullo, M.; Boccaccio, A.; Evangelista, A. Automatic Ergonomic Postural Risk Monitoring on the Factory Shopfloor-The Ergosentinel Tool. *Procedia Manuf.* **2020**, *42*, 97–103. [CrossRef]

25. Schall, M.C., Jr.; Sesek, R.F.; Cavuoto, L.A. Barriers to the Adoption of Wearable Sensors in the Workplace: A Survey of Occupational Safety and Health Professionals. *Hum. Factors.* **2018**, *60*, 351–362. [CrossRef] [PubMed]

26. Aitpayev, K.; Gaber, J. Collision Avatar (CA): Adding collision objects for human body in augmented reality using Kinect. In Proceedings of the 2012 6th International Conference on Application of Information and Communication Technologies (AICT), Tbilisi, GA, USA, 17–19 October 2012; pp. 1–4. [CrossRef]

27. Bian, Z.P.; Chau, L.P.; Magnenat-Thalmann, N. Fall detection based on skeleton extraction. In Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry, New York, NY, USA, 2–4 December 2012; pp. 91–94. [CrossRef]

28. Chang, C.Y.; Lange, B.; Zhang, M.; Koenig, S.; Requejo, P.; Somboon, N.; Rizzo, A.A. Towards pervasive physical rehabilitation using Microsoft Kinect. In Proceedings of the 2012 6th International Conference on Pervasive Computing Technologies for Healthcare (Pervasive Health) and Workshops, San Diego, CA, USA, 21–24 May 2012; pp. 159–162.

29. Farhadi-Niaki, F.; GhasemAghaei, R.; Arya, A. Empirical study of a vision-based depth-sensitive human-computer interaction system. In Proceedings of the 10th Asia Pacific Conference on Computer Human Interaction, New York, NY, USA, 28–31 August 2012; pp. 101–108. [CrossRef]

30. Villaroman, N.; Rowe, D.; Swan, B. Teaching natural user interaction using openni and the microsoft kinect sensor. In Proceedings of the 2011 Conference on Information Technology Education, New York, NY, USA, 20–22 December 2011; pp. 227–232. [CrossRef]

31. Diego-Mas, J.A.; Alcaide-Marzal, J. Using Kinect sensor in observational methods for assessing postures at work. *Appl. Ergon.* **2014**, *45*, 976–985. [CrossRef] [PubMed]

32. Manghisi, V.M.; Uva, A.E.; Fiorentino, M.; Bevilacqua, V.; Trotta, G.F.; Monno, G. Real time RULA assessment using Kinect v2 sensor. *Appl. Ergon.* **2017**, *65*, 481–491. [CrossRef]

33. Marinello, F.; Pezzuolo, A.; Simonetti, A.; Grigolato, S.; Boscaro, B.; Mologni, O.; Gasparini, F.; Cavalli, R.; Sartori, L. Tractor cabin ergonomics analyses by means of Kinect motion capture technology. *Contemp. Eng. Sci.* **2015**, *8*, 1339–1349. [CrossRef]

34. Clark, R.A.; Pua, Y.H.; Fortin, K.; Ritchie, C.; Webster, K.E.; Denehy, L.; Bryant, A.L. Validity of the Microsoft Kinect for assessment of postural control. *Gait Posture* **2012**, *36*, 372–377. [CrossRef]

35. Bonnechere, B.; Jansen, B.; Salvia, P.; Bouzahouene, H.; Omelina, L.; Moiseev, F.; Sholukha, C.J.; Rooze, M.; Van Sint Jan, S. Validity and reliability of the kinect within functional assessment activities: Comparison with standardstereo-photogrammetry. *Gait Posture* **2014**, *39*, 593–598. [CrossRef] [PubMed]

36. Plantard, P.; Auvinet, E.; Le Pierres, A.S.; Multon, F. 2015, Pose Estimation with a Kinect for Ergonomic Studies: Evaluation of the Accuracy Using a Virtual Mannequin. *Sensors* **2015**, *15*, 1785–1803. [CrossRef]

37. Patrizi, A.; Pennestrì, E.; Valentini, P.P. Comparison between low-cost marker-less and high-end marker-based motion capture systems for the computer-aided assessment of working ergonomics. *Ergonomics* **2015**, *59*, 155–162. [CrossRef] [PubMed]

38. Plantard, P.; Hubert PH, S.; Le Pierres, A.; Multon, F. Validation of an ergonomic assessment method using Kinect data in real workplace conditions. *Appl. Ergon.* **2017**, *65*, 562–569. [CrossRef] [PubMed]

39. Xu, X.; McGorry, R.W. The validity of the first and second generation Microsoft Kinect™ for identifying joint center locations during static postures. *Appl. Ergon.* **2015**, *49*, 47–54. [CrossRef] [PubMed]

40. Schroder, Y.; Scholz, A.; Berger, K.; Ruhl, K.; Guthe, S.; Magnor, M. Multiple kinect studies. *Comput. Graph.* **2011**, *2*, 6.

41. Zhang, H.; Yan, X.; Li, H. Ergonomic posture recognition using 3D view-invariant features from single ordinary camera. *Autom. Constr.* **2018**, *94*, 1–10. [CrossRef]

42. 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]

43. Cao, Z.; 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, Honolulu, HI, USA, 21–26 July 2017; pp. 7291–7299. [CrossRef]

44. Bradski, G. The OpenCV Library. *Dr. Dobb's J. Softw. Tools* **2000**, *25*, 120–123.

45. Ota, M.; Tateuchi, H.; Hashiguchi, T.; Kato, T.; Ogino, Y.; Yamagata, M.; Ichihashi, N. Verification of reliability and validity of motion analysis systems during bilateral squat using human pose tracking algorithm. *Gait Posture* **2020**, *80*, 62–67. [CrossRef] [PubMed]
46. He Ling, W.A.N.G.; Yun-Ju, L.E.E. Occupational evaluation with Rapid Entire Body Assessment (REBA) via imaging processing in field. In Proceedings of the Human Factors Society Conference, Elsinore, Denmark, 25–28 July 2019.
47. Li, L.; Martin, T.; Xu, X. A novel vision-based real-time method for evaluating postural risk factors associated with muskoskeletal disorders. *Appl. Ergon.* **2020**, *87*, 103138. [CrossRef] [PubMed]
48. MassirisFernández, M.; Fernández, J.Á.; Bajo, J.M.; Delrieux, C.A. Ergonomic risk assessment based on computer vision and machine learning. *Comput. Ind. Eng.* **2020**, *149*, 10. [CrossRef]
49. Ojelaide, A.; Paige, F. Construction worker posture estimation using OpenPose. *Constr. Res. Congr.* **2020**. [CrossRef]
50. Da Silva Neto, J.G.; Teixeira, J.M.X.N.; Teichrieb, V. Analyzing embedded pose estimation solutions for human behaviour understanding. In *Anais Estendidos do XXII Simpósio de Realidade Virtual e Aumentada*; SBC: Porto Alegre, Brasil, 2020; pp. 30–34.
51. TF-Pose. Available online: https://github.com/tryagainconcepts/tf-pose-estimation (accessed on 5 July 2021).
52. Lindera. Available online: https://www.lindera.de/technologie/ (accessed on 5 July 2021).
53. Obuchi, M.; Hoshino, Y.; Motegi, K.; Shiraishi, Y. Human Behavior Estimation by using Likelihood Field. In Proceedings of the International Conference on Mechanical, Electrical and Medical Intelligent System, Gunma, Japan, 4–6 December 2021.
54. Agrawal, Y.; Shah, Y.; Sharma, A. Implementation of Machine Learning Technique for Identification of Yoga Poses. In Proceedings of the 2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT), Gwalior, India, 10–12 April 2020; pp. 40–43. [CrossRef]
55. Contini, R. Body Segment Parameters, Part II. *Artif. Limbs* **1972**, *16*, 1–19.
56. Romero, J.; Kjellström, H.; Ek, C.H.; Kragic, D. Non-parametric hand pose estimation with object context. *Image Vis. Comput.* **2013**, *31*, 555–564. [CrossRef]
57. Wu, Z.; Hoang, D.; Lin, S.Y.; Xie, Y.; Chen, L.; Lin, Y.Y.; Fan, W. Mm-hand: 3d-aware multi-modal guided hand generative network for 3d hand pose synthesis. *arXiv* **2020**, arXiv:2010.01158.