Computers & Industrial Engineering 149 (2020) 106816

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Ergonomic risk assessment based on computer vision and machine learning

Manlio Massiris Fernández [a,b,*], J. Álvaro Fernández [c], Juan M. Bajo [a], Claudio A. Delrieux [a,b]

[a] Department of Electrical and Computer Engineering, Universidad Nacional del Sur, Av. San Andrés, 800, 8000, Bahía Blanca, Argentina
[b] National Council of Scientific and Technical Research of Argentina - CONICET, Av. La Carrindanga, km 7, 8000, Bahía Blanca, Argentina
[c] Department of Electrical, Electronic and Automation Engineering, School of Industrial Engineering, Universidad de Extremadura, Av. Elvas, n/n, 06006, Badajoz, Spain

ARTICLE INFO

Keywords:
Ergonomic risk assessment
Occupational health and safety
Work-related musculoskeletal disorders
Computer vision
Neural networks

ABSTRACT

We develop a novel method that performs accurate ergonomic risk assessment, automatically computing Rapid Upper Limb Assessment (RULA) scores from snapshots or digital video using computer vision and machine learning techniques. Our method overcomes the limitations in recent developments based on computer vision or in wearable measurement sensors, being able to perform unsupervised assessment handling multiple workers simultaneously, even under sub-optimal viewing conditions (*e.g.*, poor illumination, occlusions, and unstable camera views). The processing workflow uses open-source neural networks to detect the workers' skeletons, after which their body-joint positions and angles are inferred, with which RULA scores are computed. The method was tested with computer-generated, controlled real-world image datasets, and with freely available videos taken in outdoor working scenarios. The computed RULA scores were in close agreement with the assessments of seven specialists in the field, achieving a Cohen's κ over 0.6 in most real-world experiments.

1. Introduction

Work-related musculoskeletal disorders (WMSDs) refer to a wide range of occupational health issues. Known and presumed causes of WMSDs include specific aspects of the working environment, the type of task performed, and the occupational human body postures. WMSDs may result in the production of inflammation or degeneration of functional body structures, such as nerves, tendons, ligaments, and muscles (Ha et al., 2009). Occupational Health and Safety (OHS) programs identify these sources of harm as ergonomic risk factors, which are the focus of Ergonomics programs (Luttmann et al., 2003). Ergonomic interventions aim to detect and assess the imbalance between workplace requirements and workers' physical abilities to prevent WMSDs. Mechanical overload, frequent task repetition, and high exposure times to non-ergonomic body postures are acknowledged risk factors, responsible for a relevant fraction of upper limb and lumbar occupational diseases in many industrialized countries (Ha et al., 2009). The so-called response capability of human tissues should not be surpassed neither in application time nor in the magnitude of applied force (Helliwell, 2004; Luttmann et al., 2003).

WMSDs are nowadays the leading cause of sick leaves, work-related disabilities, and an overall productivity loss in developed countries. In the European Union (EU), they involve over 50% of work-related diseases, being responsible for more than 40% of all economic losses from OHS issues (Bevan, 2015). These disorders are responsible for more than 30% of all diseases and non-fatal accidents in the United States of America (USA) (Bureau of Labor Statistics, 2016; Gerr et al., 2013). As a response, the National Institute for Occupational Safety and Health (NIOSH) of the USA proposed in 1997 a series of steps to be included in every Ergonomics program to prevent WMSDs. An effective Ergonomics program should include an evidence-based ergonomic risk assessment (RA) for recognizing and rectifying ergonomic deficiencies in an open range of workplace situations (Cohen, 1997). As a result of this proposal, a wealth of methods and tools have been designed in the last decades for ergonomic RA. These can be classified as follows (Battini et al., 2014; David, 2005; Vignais et al., 2017):

- ***Self-assessment***, where workers assess themselves using specifically designed forms.
- ***Human observation***, where qualified personnel collects a subjective estimation of workers' body-joint angles through on-site observations and/or off-line video analysis.
- ***Direct measurement***, where anthropometric tools and devices are attached to a worker's body to automatically collect data for ergonomic analysis.

* Corresponding author at: Department of Electrical and Computer Engineering, Universidad Nacional del Sur, Av. San Andrés, 800, 8000, Bahía Blanca, Argentina.
E-mail addresses: manlio.massiris@uns.edu.ar (M. Massiris Fernández), jalvarof@unex.es (J.Á. Fernández), juan.bajo@uns.edu.ar (J.M. Bajo), cad@uns.edu.ar (C.A. Delrieux).

https://doi.org/10.1016/j.cie.2020.106816
Received 12 December 2019; Received in revised form 16 June 2020; Accepted 2 September 2020
Available online 11 September 2020
0360-8352/© 2020 Published by Elsevier Ltd.

- ***Computer-based assessment***, where human body models are automatically estimated from camera takes by tailored computer vision (CV) applications, thus providing systematic and objective model-based ergonomic measurements.

The most prominent ergonomic RA methods have been thoroughly discussed in *e.g.*, Chiasson et al. (2012), Joshi and Deshpande (2019), Kong et al. (2017), and Roman-Liu (2014). Though practical to use, both self-assessment and human observation approaches are invariably affected by subjective biases. Currently, the most frequently used methods for ergonomic RA in industrial workplaces follow the human observation approach. However, even experienced ergonomists often err on subjective category decisions mainly due to workplace suboptimal visual conditions such as poor illumination, occlusions, and inadequate take angles in videos or pictures (Plantard et al., 2017). In addition, direct measurement and expert-based observation methods are constrained by economic costs derived from the amount of time needed to perform the assessment and the technical knowledge required from the analysts (Nath et al., 2017).

Over the last few years, several disruptive technologies related to the Industry 4.0 (I4.0) paradigm are enabling fruitful applications in ergonomics (Kadir et al., 2019; Rauch et al., 2019). More specifically, the impact of automated data collection and analysis is shaping a new group of data-driven applications where technological advances in hardware sensors and machine learning (ML) open novel roads for Ergonomics. Among other technologies, we highlight the use of on-body inertial and electro-goniometer sensor networks (Vignais et al., 2013), inclinometers and accelerometers (Jayaram et al., 2006; Nath et al., 2017), and CV systems such as color and depth (RGB-D) devices (Diego-Mas & Alcaide-Marzal, 2014; Krüger & Nguyen, 2015; Plantard et al., 2017; Xu et al., 2017), stereo cameras (Liu et al., 2016a), and cleverly used plain RGB color cameras (Fığlalı et al., 2015; Yan et al., 2017; Zhang et al., 2018).

On the one hand, direct measurement and wearable-based technologies (*e.g.*, accelerometers and inclinometers) are intrusive devices that may influence or limit the free development of work activities (Yu et al., 2019; Zhang et al., 2018). On the other hand, recent CV- and ML-based methods enable an accurate human posture identification and analysis that may assist in observation-based ergonomic RA methods of choice (Battini et al., 2014; Vignais et al., 2017; Yan et al., 2017).

However, CV-based approaches have not yet met a proper handling of the complexities associated with real-world operational environments (*e.g.*, occlusions, uneven illumination, and varying poses and object scale) (Fang et al., 2020; Liu et al., 2016b). Recent proposals (*e.g.*, Fang et al., 2020; Plantard et al., 2017; Seo et al., 2019, and Li et al., 2018) require the worker to be constrained to a limited range of movements, and typical situations like outdoor light changes or variations in camera viewpoint still hamper the accuracy and precision of the results. In Yan et al. (2017) and Zhang et al. (2018), plain RGB cameras are used together with ML-based models for providing automatic ergonomic RA data following the Ovako Working Analysis System (OWAS) from a video stream. These proposals are based on view-invariant feature detection, and thus are more robust to illumination and camera view variations, but are still restricted to single worker assessment, and brittle with respect to occlusions and varying poses.

The main contributions of the present paper are:

- A new Computer-Vision-based methodology able to provide robust and consistent ergonomic assessment. which overcomes the limitations of current approaches.
- Empirical criteria to validate the detection and pose estimation confidence.
- Experimental evidence with public-domain videos showing that our solution is robust under uneven illumination conditions, moving camera viewpoints, occlusions, and worker movements.
- Statistical evaluation of the automatically computed RULA scores against specialists' assessments.
- An inexpensive and easy to replicate deploy, based on the use of low-cost off-the-shelf RGB devices, together with recently proposed open-source Convolutional Neural Network (CNN) architectures.
- A CV-based approach that can be easily reused in many other contexts.

To the best of our knowledge, this proposal is the first to be able to provide automatic ergonomic assessment in a robust and flexible manner even in varying real world conditions.

In the next section, we introduce the background of our work, including RULA ergonomic RA, the OpenPose CNN, the proposed methodology to obtain RULA scores from OpenPose outputs, and the validation methods employed. Section 3 presents the proposed experimental setup used for testing the methodology, including synthetic and real-world datasets generated in controlled laboratory conditions. Section 4 shows the results obtained by our methodology on video takes from actual working scenarios, which are contrasted with the assessment of expert ergonomists. The validity and significance of these results are further discussed in Section 5. Finally, we briefly outline the conclusion and propose venues for future work in Section 6.

2. Materials and methods

2.1. RULA

The RULA method introduced in McAtamney and Corlett (1993) was designed to provide a so-called grand score, *i.e.*, a global numerical score representing the level of postural load of a worker's musculoskeletal system, due to the relative positions of its body parts. This RULA grand score is obtained from a modest-sized, hierarchical combination scheme of individual body part scores, where each body part independently assesses specific joint angles into some levels, which are determined by angular thresholds (see Fig. 1). These body part scores are then combined into intermediate RULA scores A and B, which respectively summarizes the stress of the upper limbs, and the stress of the rest of the body. Finally, the RULA grand score combines both intermediate scores, and their force and muscle stress factors, into a single value. This RULA grand score relates to four action levels, which gradually indicate the need for further investigation or the application of ergonomic changes in the workplace. Nowadays, RULA is a widely accepted ergonomic RA tool conducted through observation by researchers and practitioners in the field (Joshi & Deshpande, 2019; Plantard et al., 2017; Savino et al., 2017).

RULA scores are usually obtained from direct human observation, and more recently, from snapshots or takes from video recordings. To perform RULA, the analyst must select the most critical body poses that are either performed for a long duration, or are repetitive, or specifically known to be harmful, or represent a crucial deviation in comparison to safe and neutral positions. For this reason, the first step in RULA is to observe the different tasks performed by the worker. As a practical rule, several work cycles need to be observed prior to a critical posture selection. If the work cycle is very long (or there is no work cycle), the analyst often chooses to take snapshots at regular time intervals (Chiasson et al., 2012; Li et al., 2018; Roman-Liu, 2014). Our method is designed to automatically compute RULA scores from single snapshots or video sequences. The assessment is based on the following assumptions: in a single snapshot, the image is representative of a potentially critical posture, whereas in a video take, frames are taken at regular intervals and are representative as a whole of a single work task or activity.

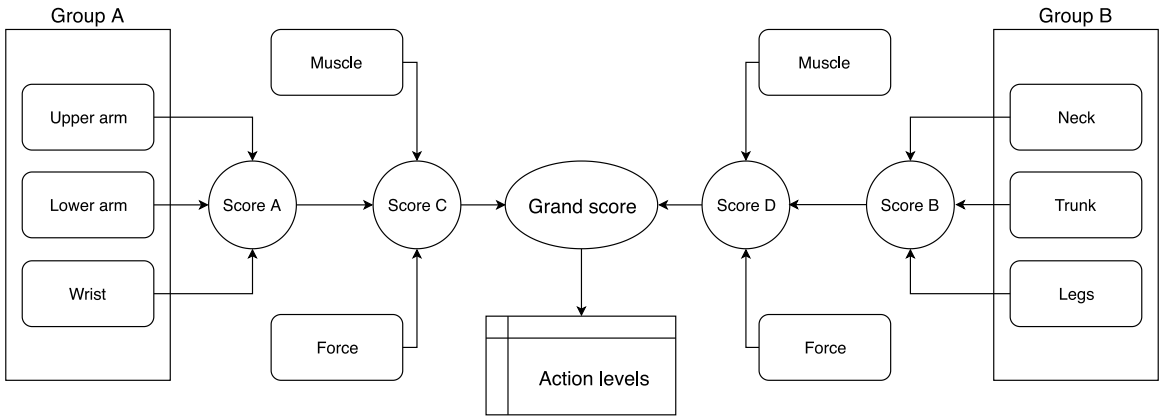


Fig. 1. RULA hierarchical scheme.

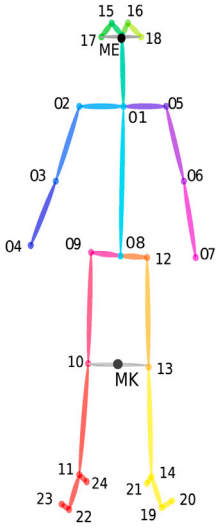


Fig. 2. OpenPose 25 skeleton and body joints.

Table 1
RULA joint angles from OpenPose 25 skeleton data.

Angle name	Acronym	Involved joints[a]
Left elbow	EL	∠04, 03, 02
Right elbow	ER	∠05, 06, 07
Left shoulder	SL	∠03, 02, 09
Right shoulder	SR	∠06, 05, 12
Left shoulder 2	SL2	∠03, 02, 01
Right shoulder 2	SR2	∠06, 05, 01
Left knee	KL	∠09, 10, 11
Right knee	KR	∠12, 13, 14
Left foot	FL	∠10, 11, 22
Right foot	FR	∠13, 14, 19
Neck twisting	NT	∠00, 01, 02
Neck bending left	NB	∠17, 01, 02
Neck bending right	NB2	∠18, 01, 05
Neck flexion	NF	∠ME, 01, 08
Trunk twisting right	TT	∠02, 08, 09
Trunk twisting left	TT2	∠05, 08, 12
Trunk Bending	TB	∠09, 08, 01
Trunk flexion	TF	∠MK, 08, 01

[a]Joints numbered according to Fig. 2 except for ME and MK (see text).

Algorithm 1 Confidence preserving average of two joints. Undetected joints encoded as [0,0,0].

1: INPUTS: J_1, J_2
2: **if** $\min(J_1, J_2) == [0, 0, 0]$ **then**
3: $J_0 \leftarrow \max(J_1, J_2)$
4: **else**
5: $J_0 \leftarrow \frac{1}{2}(J_1 + J_2)$
6: **end if**
7: OUTPUT: J_0

2.2. OpenPose

Human pose estimation is a long-standing problem in CV. It may be defined as the task of localizing human body joints (*e.g.*, knees, elbows, shoulders) in digital images, and as a second step, the search for specific poses which coincide with the observed joints in the space of possible articulated poses. The use of ML tools such as CNNs has increased the robustness of these methods. Today, state-of-the-art open-source software tools such as OpenPose (Cao et al., 2017) allow real-time detection of joints and limbs from digital images and videos. Focused on multi-person pose estimation, OpenPose uses a bottom up approach which takes an entire image as input for a two-branch CNN, to jointly predict confidence maps for body part detection, and part affinity fields for parts association. Given an input image, the network delivers a list of detected bodies, each with an associated skeleton of priorly defined joints. We used the 25-joint OpenPose model (OpenPose 25), where skeleton joints are listed following the specific order depicted in Fig. 2. For every detected joint, the model gives a vector with its relative position in the image, and the confidence of the estimation, which ranges from 0 (null) to 1 (full) estimation confidence. From this information, we calculate an overall skeleton detection confidence as the average of the joint estimation confidences, which will be useful for filtering out noisy or spurious detections.

2.3. Joint angle estimation

The body-joint angles established in Table 1 are used to estimate RULA scores automatically. For example, the left elbow angle (EL) is obtained from the observed positions of the left shoulder, elbow and wrist, corresponding respectively to OpenPose 25 skeleton joints #04, 03, and 02, as shown in Fig. 2. Two specific angles from Table 1 utilize fiducial points computed from OpenPose 25 joints: ME (mid-ears) represents a point halfway between both ears, and MK (mid-knees) denotes a point between the knees (see Algorithm 1). We define these fiducial points for computing two angles that are critical for RULA assessment, i.e., Neck Flexion (NF) and Trunk Flexion (TF), which are commonly difficult or impossible to compute otherwise since the required joints are frequently occluded (see Sections 3.2 and 3.3).

2.4. RULA scores calculation

RULA requires applying joint angle thresholds per skeleton to compute scores. These thresholds are clear-cut for some RULA joint angles

Table 2
RULA scores from skeleton joint angles.

RULA score	Formulæ[a]	Score
Upper arm position	$\overline{[SL, SR]}$	+1 (−20°, 20°) +2 (−∞, −20°) +2 (20°, 45°) +3 (45°, 90°) +4 (90°, ∞)
Upper arm abduction	max(SL2, SR2)	+1 (110°, ∞)
Lower arm position	$\overline{[EL, ER]}$	+1 (60°, 100°) +2 (−∞, 60°) +2 (100°, ∞)
Arm working outside of the body	max(SL, SR)	+1 (30°, ∞)
Neck angle	NF	+1 (0°, 10°) +2 (10°, 20°) +3 (20°, ∞) +4 (−∞, 0°)
Neck twisting angle	abs(90°−NT)	+1 (5°, ∞)
Neck bending angle	abs(65°-$\overline{[NB, NB2]}$)	+1 (5°, ∞)
Trunk flexion angle (TFA)	TF	+1 ≈ 0° +2 (1°, 20°) +3 (20°, 60°) +4 (60°, ∞)
Trunk twisting angle (TTA)	max(TT, TT2)	+1 (100°, ∞)
Trunk bending angle	TB	+1 (−∞, 85°) +1 (95°, ∞)

[a]Angle abbreviations are given in Table 1.

(*e.g.*, elbows and knees), but not explicitly delimited for others. The latter group includes shoulder abduction, arm working outside of the body, neck twisting, neck bending, trunk twisting and trunk bending angles (Battini et al., 2014; Plantard et al., 2017). Therefore, to obtain these body part or local RULA scores, we define the thresholds shown in the third column of Table 2. To provide a complete RULA grand score, the analysts manually set wrist twisting and position scores, legs balance score, and force and muscle use scores.

Specific thresholds and formulæ are derived from multiple experiments, as discussed in Section 3. In short, the provided formulæ try to compensate for possible occlusions on each side of the body, by using either the average of left and right body part angles, fiducial mid-points, or maximum values. Furthermore, our method utilizes two thresholds (size and confidence) for discarding unwanted skeleton data which may be false positives. The size threshold considers a minimum skeleton area in proportion to the image resolution, whereas the confidence threshold is applied to discard skeletons whose joints are either occluded or detected with low confidence.

2.5. Evaluation methods

As mentioned in Cao et al. (2017) and Zhang et al. (2018), common image conditions such as uneven illumination, viewpoint variation, crowding, and occlusions may cause relevant information losses. To study the applicability of our CV-based method given these factors, we designed several simulated and controlled settings to quantify the effects of self-occlusion, to compare the quality of the joint angle evaluation in simulated and real takes, and to determine suitable joint angle thresholds for RULA. Finally, we validated the method in actual outdoor working situations under the technical supervision of seven experienced ergonomists, who also evaluated the associated RULA scores. The validation methods involved three levels of comparison:

1. Skeleton and joint detection confidences by viewpoint, following Golabchi et al. (2015) and Plantard et al. (2015);
2. Angle comparison between lab-controlled and simulated viewpoints, similarly to Li et al. (2018), and
3. RULA score agreement across the proposed method and observations from experienced ergonomists, in accordance with Lee et al. (2017), Levanon et al. (2014), and Plantard et al. (2017).

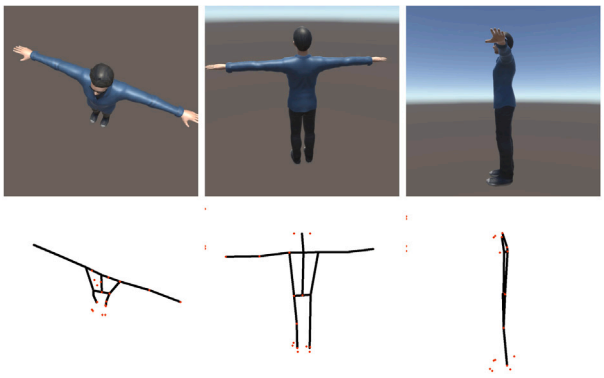


Fig. 3. Images from the artificial character dataset.

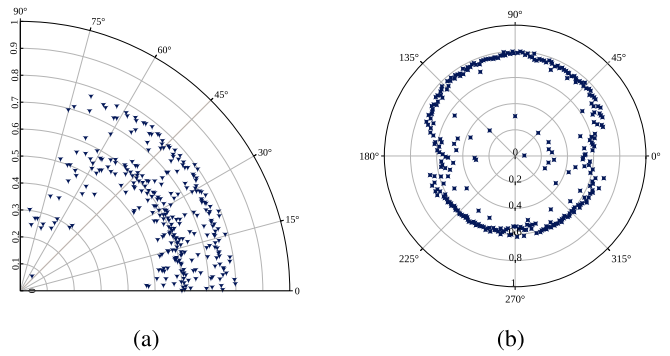
First, in Sections 3.1 and 3.2, we tested the proposed methodology against self-occlusions. We used a simulated 3D model as a reference for joints and body positions to provide controlled camera viewpoints. The generated dataset contains 300 random camera viewpoint samples following Hammersley's method, which ensures angular equidistribution between samples (Cui & Freeden, 1997). In our case, radii were chosen to capture in the foreground the 3D artificial character from each possible viewpoint, where the hemisphere center coincides with the model centroid (see Fig. 3).

Second, in Section 3.3, we extended the test by contrasting a controlled image sequence where one of the authors with no prior injuries (MM[1]) performs regulated trunk twisting and trunk flexion actions. These sequences were taken as models for generating two further datasets, in which an artificial character performs the same movements and is taken from 300 different camera viewpoints (see Figs. 6 and 7).

Finally, in Section 4, we validated our method in five videos of actual working scenarios by testing the agreement on sample-to-sample data, as expressed by the unweighted Cohen's kappa (κ) among the RULA risk ratings of the proposed method and the assisted assessments of seven experienced ergonomists. As mentioned in Li et al. (2018), since RULA is mostly conducted through manual observation, experts usually analyze just a few body postures that they judge to be critical. For this reason, the validation procedure of the proposed methodology was made with the advice of experts, in accordance with Lee et al. (2017) and Plantard et al. (2017).

These experiments were aimed to test our method with work tasks footage taken at different actual outdoor workplaces. Five videos were selected that contained one or more of the aforementioned challenging viewing conditions (*e.g.*, uneven illumination, moving camera viewpoints, occlusions). We taped two different jobs (*wall plastering* and *hammering*) following the advice of the ergonomists and within the conditions of the Ethics Committee of our Institution. We also selected three publicly available videos of different actual job performances, even though these videos were not specifically recorded for ergonomic assessments. For each video, the ergonomists selected the frames according to the method described in Section 2.1, *i.e.*, manual selection of snapshots or frames taken at regular intervals. These frames were then independently evaluated by the specialists and with our method. Finally, the assessment agreement was calculated as in Lee et al. (2017) and Plantard et al. (2017). Wrist and leg scores, and muscle and force scores, were jointly assessed by the ergonomists. Additionally, a size and confidence thresholds of 0.35 and 0.2 respectively were applied in these videos.

[1] The sequence was performed following ergonomic advice from the specialists and the corresponding Ethics protocol.



(a) (b)

Fig. 4. Skeleton detection confidence vs. camera (a) latitude and (b) azimuth.

Table 3
Skeleton confidences for viewpoint latitude ranges.

Ranges	0°–45°	45°–75°	75°–90°
mean	0.660	0.588	0.360
std	0.082	0.159	0.169
min	0.469	0.070	0.250
max	0.799	0.782	0.696

Table 4
Skeleton confidences for viewpoint azimuth ranges.

Ranges	30°–150°	210°–330°	Other
mean	0.740	0.589	0.576
std	0.088	0.041	0.129
min	0.250	0.374	0.070
max	0.799	0.653	0.750

3. Controlled environment testing

3.1. Skeleton detection confidence test

The first experiment was carried out to study the detection confidence, p_k, of every joint of a single skeleton detected by OpenPose 25 as a function of the camera viewpoint. The generated dataset (see Fig. 3) was fed to OpenPose 25 to obtain a total of 300 skeletons according to Fig. 2. The two graphs in Fig. 4 show the obtained average detection confidence values, $\bar{p}$, for the whole dataset versus viewpoint latitude (or vertical view angle) and azimuth (or horizontal view angle). As a result, camera latitudes (Fig. 4(a)) were grouped into three categories: 0°–45°, 45°–75°, and 75°–90°. Viewpoint angles within the first category provided good skeleton acquisition confidences, which slightly diminished in the second category, and much worsen in the last. Thus, camera latitude should be limited preferably to 45°, or at most to less than 75°, to avoid confidence loss due to self-occlusions (see Table 3).

In a similar vein, azimuth angles were also grouped into three categories: 30°–150°, 210°–330°, and other angles (see Fig. 4(b)). The first category represents frontal viewpoints of the 3D model, where both facial, feet, trunk, and limb joints are almost never self-occluded. In the second category (rear viewpoints), self-occlusion of toes, nose, and eyes reduces the maximum $\bar{p}$ (see Table 4). Finally, lateral viewpoints of the subject are included in the third category, providing high variability in skeleton confidences. Given the precise posture symmetry of the 3D model (Fig. 3), lateral self-occlusion affects about half of the considered body joints. Both frontal and rear views of the subject provide the best skeleton confidences required for RULA assessment, since self-occluded joints are either very few (*e.g.*, ankles in frontal view), or the least relevant for the method (*e.g.*, eyes and nose in rear view). Thus, using frontal (90° ± 45°) or rear (270° ± 45°) camera view azimuths should provide suitable body joints for the proposed method.

3.2. Shoulder and neck angle tests

In RULA, joint body scores take different values that are aggregated using tables, following the schema presented in Fig. 1. In this experiment, we tested several methods for computing RULA scores from the simulated dataset used in the previous Subsection. Our aim here was to find the best way to compute each score correctly.

According to Table 2, since both upper arm, neck, and trunk can take the maximum body part score (+4), they should be treated as most relevant when computing their related angles. However, RULA does not include a clear identification of the anthropometric points involved in their angular measurement (Plantard et al., 2017). For the neck score, we considered the angle formed by the mid-hip and neck joints, and three possible options for the third point, namely the nose, the midpoint between ears, and the midpoint between the eyes. Likewise, we selected again mid-hip and neck joints for computing the trunk score, and evaluated the midpoints between knees, ankles, and heels as candidates for the third point. Finally, the upper arm scores were evaluated from the angles formed by both hip, shoulder, and elbow joints of each side of the body.

Fig. 5 illustrates individual joint confidence (p_k) graphs with variable viewpoint azimuth for nose, eyes, ears and shoulders. For the nose, Fig. 5(a) shows quite different detection confidences (p_0) for frontal and rear views of the subject. In particular, the azimuth interval (225°–315°) shows null confidence due to the self-occlusion of the head itself. Fig. 5(b) combines left and right eye confidences (p_{16} and p_{15}). The right eye is correctly detected in the azimuth interval (315°–135°), while the left eye presents a similar behavior in the range (45°–225°). Thus, there is no possibility to infer the position of any eye in the range (225°–315°), which entirely coincides with the nose rear blind spot. Combined confidence detection of the ears (p_{18} and p_{17}) are shown in Fig. 5(c). In this case, lateral occlusions are responsible for other blind spots, *i.e.*, both ears cannot be simultaneously detected with sufficient confidence from almost any viewpoint. The absence of a combined blind spot interval for the ears is the reason which prompted us to define a ME (mid-ear) point for obtaining the neck score of Table 2 by applying Algorithm 1 over both ear points.

Finally, the obtained shoulder joint confidences, p_{02} and p_{05}, show narrow lateral blind spots located about 0° and 180° for left and right shoulder, respectively (see Fig. 5(d)). Self-occlusion explains these lateral blind spots, which were also detected in knee joints confidences (p_{10} and p_{13}). As presented in Table 2, we overcome this issue for shoulder angles using the average of left and right shoulder angles (SL and SR), whereas to compute the trunk flexion (TA) angle, a mid-knee (MK) fiducial point is defined from both knee joints using Algorithm 1.

3.3. Trunk twisting and trunk flexion threshold tests

As mentioned in Section 2.4, RULA applies thresholds to rate joint angles. Although these thresholds are unequivocal in some angles, they are defined with uncertainties in others. Therefore, in Table 2 we propose thresholds based on several controlled laboratory tests. An example of these experiments is the per-frame variation analysis of trunk twisting angle (TTA) and trunk flexion angle (TFA) in controlled video takes, in order to set suitable thresholds for their associated RULA scores (see Figs. 6 and 7). The trunk twisting pose graph in Fig. 6(b) plots TFA (orange) and TTA (blue) versus time. Noticeably, TTA shows two peaks indicating that the twist movement was executed twice (in this case, first to the left and then to the right). Moreover, critical postures of this action were unanimously identified by the specialists at these peak frames. On the other hand, TFA experimented almost null variation during this action, as expected. Since RULA rates trunk twisting postures with a +1 score, we selected a TTA threshold of 100° for computing its RULA score (see Table 2).

In the trunk flexion action sequence (from a standing position to a crouch, and then back), the right side of the subject's body was

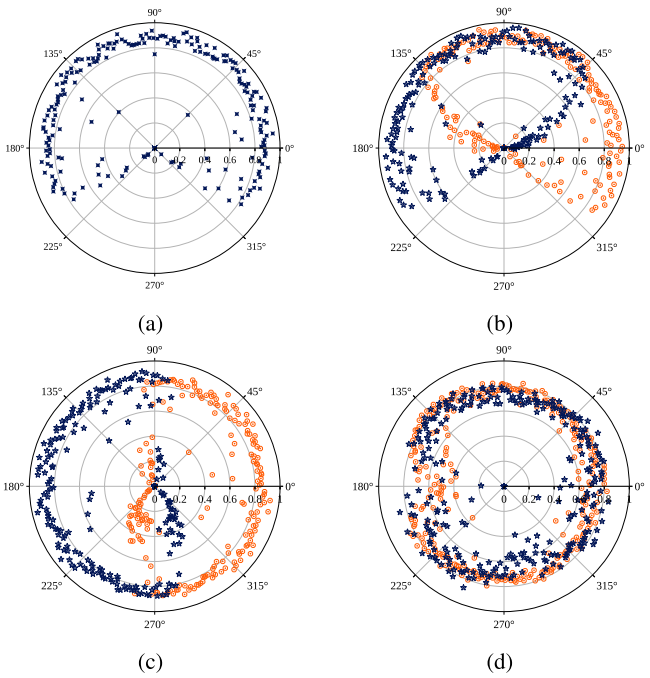


Fig. 5. Joint detection confidence as a function of camera azimuth for (a) nose, (b) left and right eyes, (c) left and right ears, and (d) left and right shoulders.

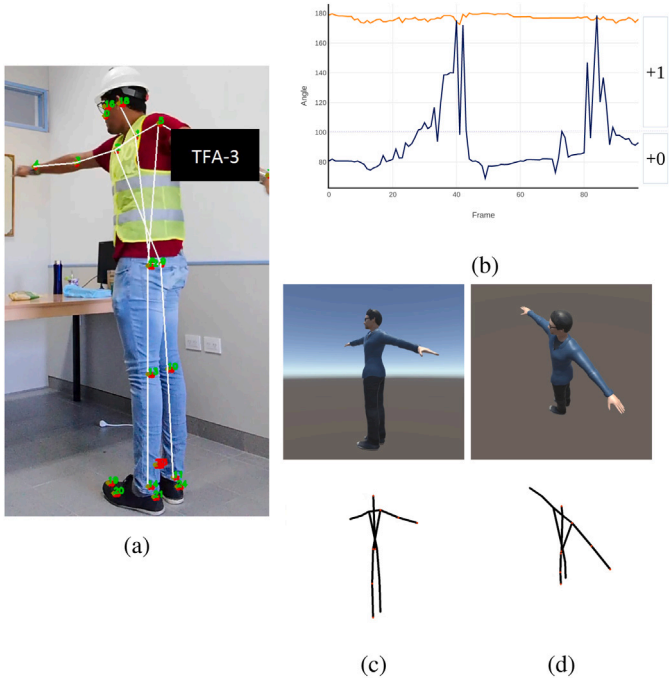


Fig. 6. Trunk twisting lab test: (a) OpenPose 25 skeleton overlaid on a critical posture frame, (b) angular time plot across the twisting action, and (c, d) samples of the 3D rendered dataset simulation from (a) and computed skeletons.

occluded (see Fig. 7(a)). However, our method was consistently capable of inferring the occluded joints required for computing RULA local scores with sufficient accuracy. The time plot of Fig. 7(b) enabled a quick way for selecting a critical pose frame from the TFA flat central band (orange). Moreover, the critical pose depicted in Fig. 7(a), for which TFA takes the minimum value in the whole sequence, was considered as representative of the action by all of the ergonomists to compute its RULA scores (TFA +4, TTA 0). This plot also shows that both TFA and TTA decreased during the trunk flexion action. On the one hand, a frame by frame analysis of this plot allowed us to set the required thresholds for computing the four trunk flexion RULA score levels, also shown in the plot. On the other hand, the previously selected threshold for TTA prevented the method from obtaining an incorrect RULA trunk twisting score in any frame of the sequence.

Finally, to check the correctness of our procedure, we portrayed both previous trunk twisting and trunk flexion critical poses to our artificial 3D character to generate two additional synthetic datasets (see Figs. 6(c)–6(d), and Figs. 7(c)–7(d)), each from 300 random viewpoints. From these datasets, we were able to test the validity of our results.

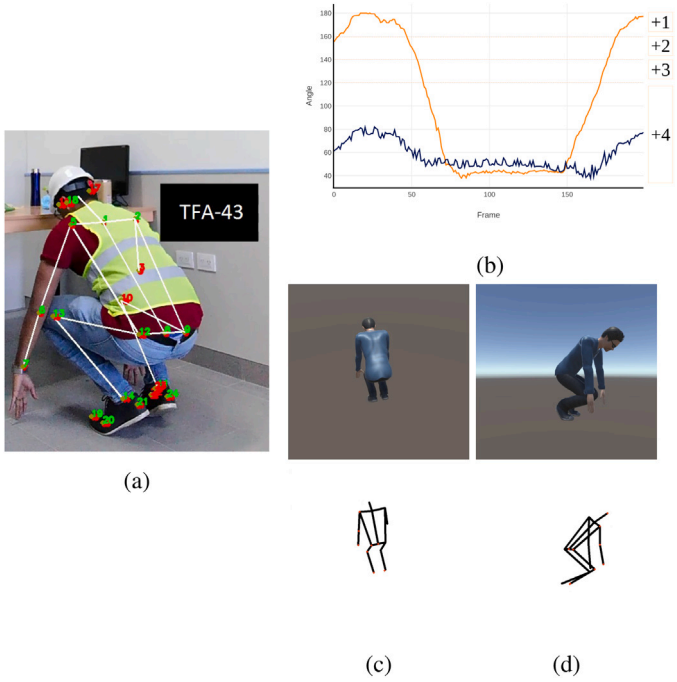


Fig. 7. Trunk flexion lab test: (a) OpenPose 25 skeleton overlaid on a critical posture frame, (b) angular time plot across the flexion action, and (c, d) samples of the 3D rendered dataset simulation of (a) and computed skeletons.

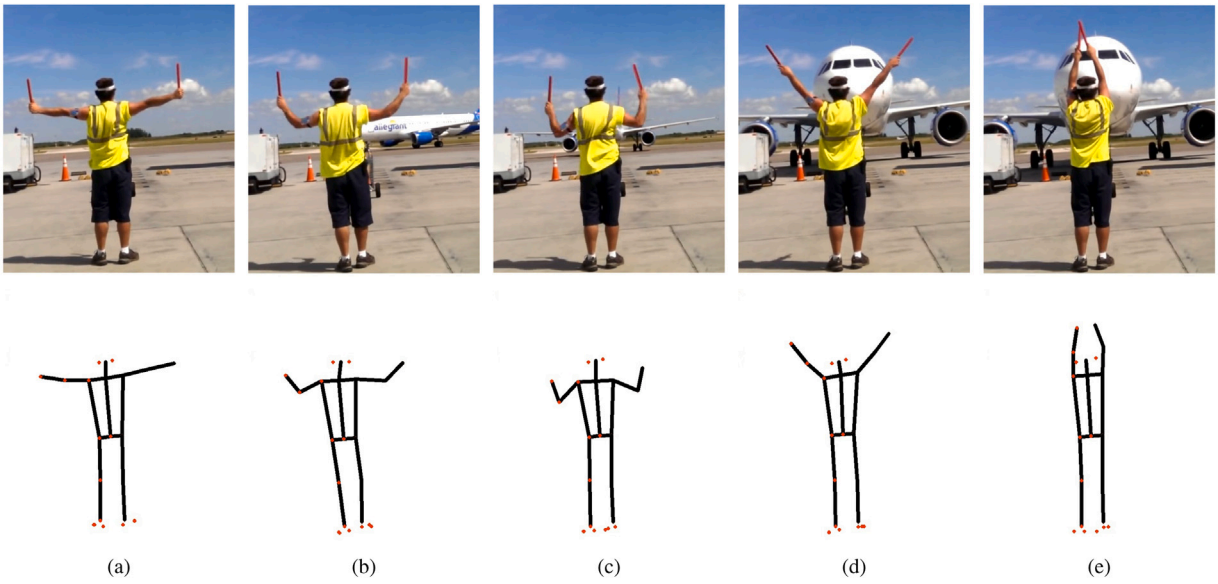


Fig. 8. Samples from Marshall signs job sequence.

4. Results

The following experiments aimed to test our method with videos of actual workers completing jobs in different outdoor working scenarios. First, we analyzed a video of an operator performing Marshall signs to an airplane from a fixed rear camera viewpoint (see Fig. 8). In this sequence, frames were taken periodically to study the estimation accuracy of our method in assessing upper limb angles from a rear viewpoint in absence of trunk bending and twisting postures. This study was promoted by the ergonomists given the relevance of these angles in RULA. Moreover, a temporal analysis similar to that of Section 3.3 allowed the statistical characterization of normal body-joint angle distributions (*i.e.*, mean ± standard deviation). According to Table 2, our method provided averaged shoulder angles (SL, SR) of $83.41° \pm 26.8°$, and averaged elbow angles (EL, ER) of $120.79° \pm 49.17°$, yielding on average RULA upper and lower arm position scores of +3 and +2, respectively. As discussed in Section 3.2, the rear viewpoint is a blind spot for various OpenPose joints that can be used for computing the RULA neck score. This was also the case for this sequence, where even the worker's left ear was self-occluded in some frames (*e.g.*, in Fig. 8(e)). However, the use of the mid-ear (ME) point allowed us to obtain accurate neck angles for each shot, of $6.72° \pm 4.34°$, thus yielding a RULA neck score of +1 for most of this sequence.

Next, we applied our method in the videos of *wall plastering* and *hammering* work activities, both taken from a hand-held smartphone. In these cases, the ergonomists had great assessment difficulties due to repeated occlusions and the non-static camera view. For instance, the plasterer's legs in Fig. 9 are always occluded by a wheelbarrow, while

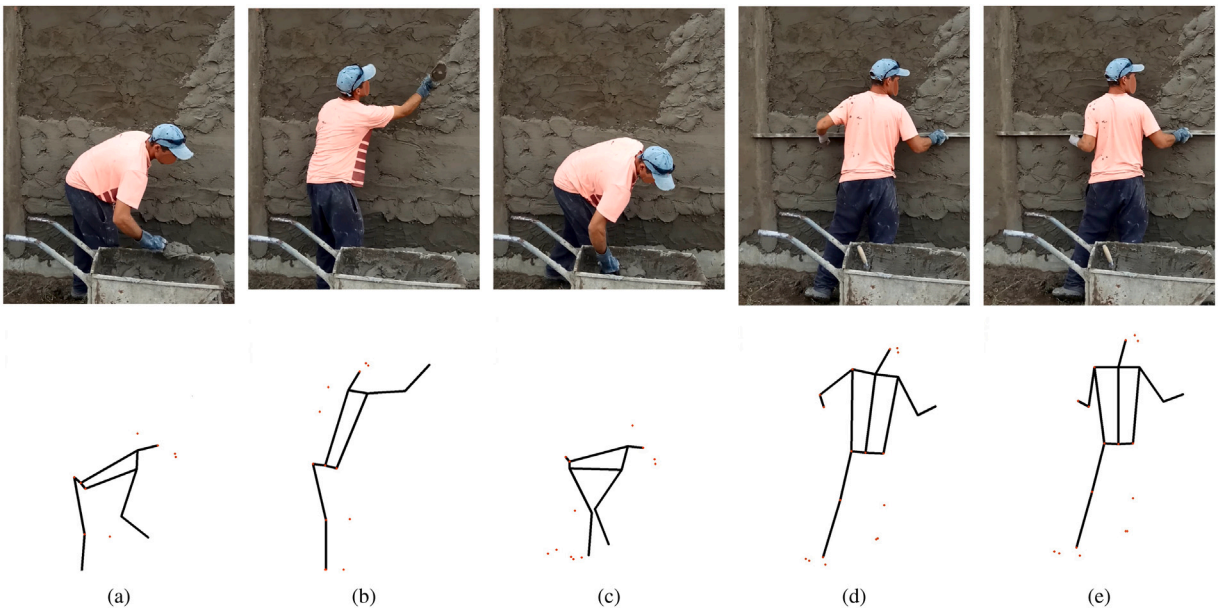


Fig. 9. Samples from Wall plastering job sequence.

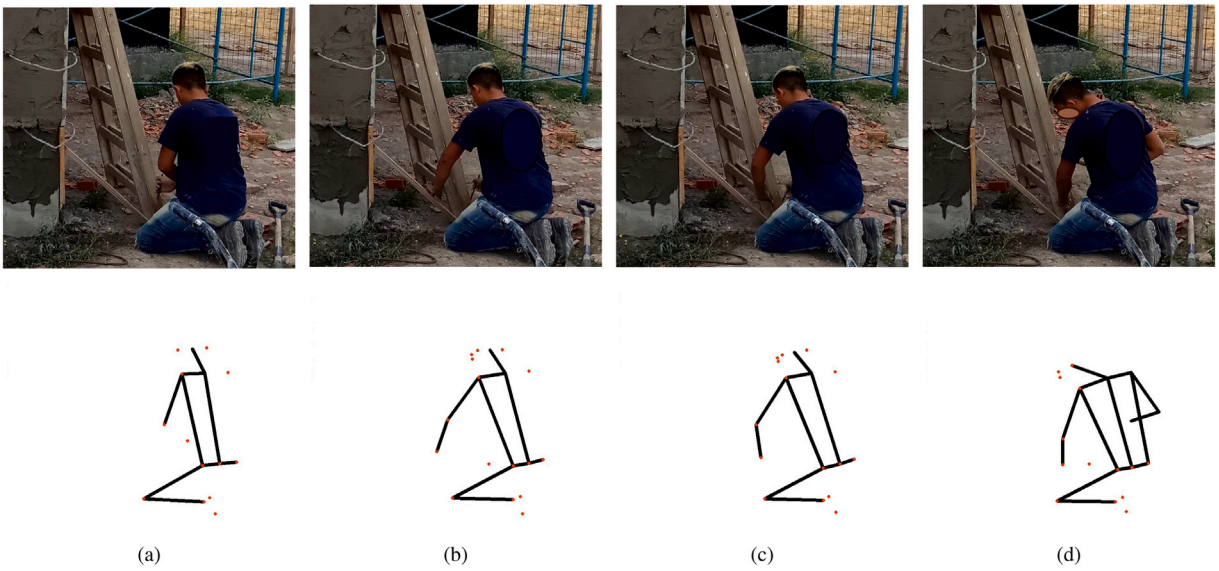


Fig. 10. Samples from Hammering job sequence.

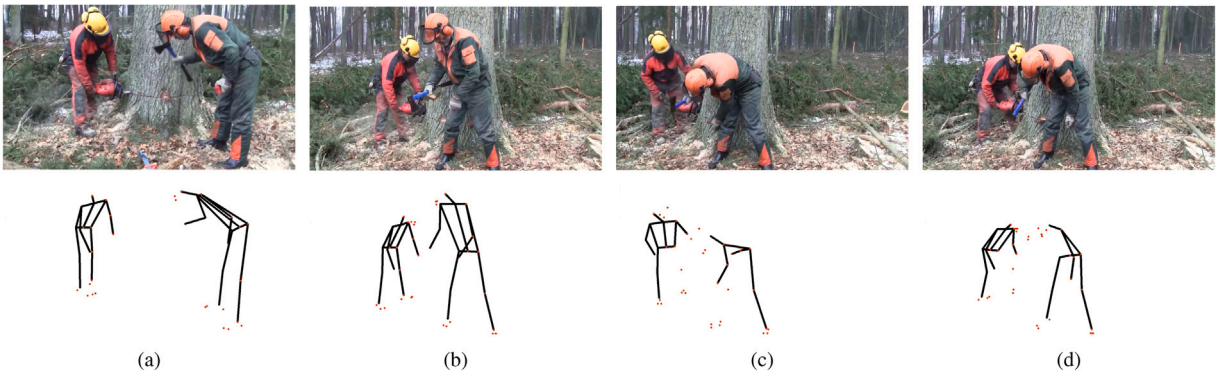


Fig. 11. Samples from Tree cutting job sequence.

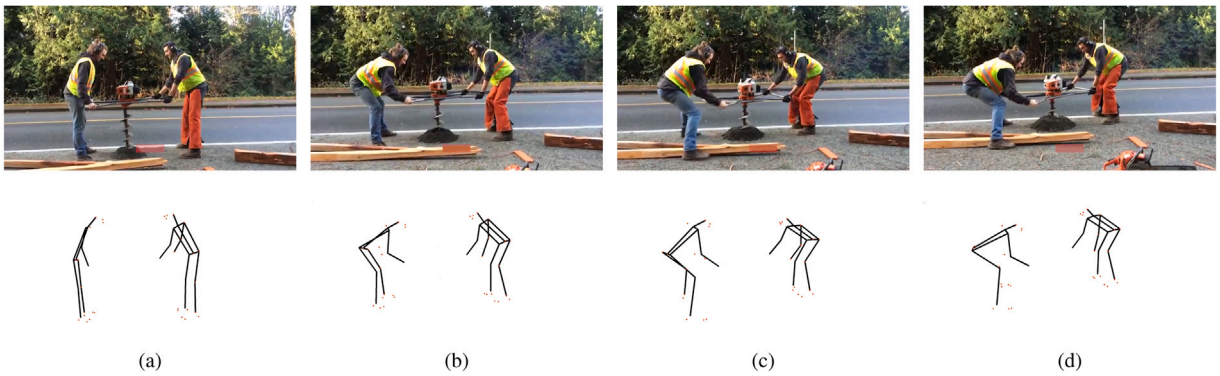


Fig. 12. Samples from Drilling job sequence.

Table 5
RULA scores and consensus with experts' assessments obtained in outdoors job videos (in *Wall plastering* and *Hammering*, two workers were assessed, obtaining the same scores).

Video name	Trunk	Neck	Score A	Upper arm	Lower arm	Score B	Grand score	Cohen's κ
Drilling	5	3	5	3	3	7	7	0.61
Tree cutting	4	4	5	3	2	7	7	1.0
Wall plastering 1	3	4	5	3	2	7	7	0.66
Wall plastering 2	3	4	5	3	2	7	7	0.66
Marshall signs	2	1	5	4	3	4	5	0.65
Hammering 1	5	4	5	3	3	7	7	0.81
Hammering 2	5	4	5	3	3	7	7	0.81
Cohen's κ	0.44	0.68	0.66	1.0	0.61	0.58	1.0	

both neck and mid-hip joints are never detected with poor confidence. However, the occasional inference of the occluded knee joints obtained from OpenPose 25 allowed our method to use the mid-knee (MK) point on those frames, and to compute a trunk flexion (TF) angle of $39.13° \pm 30.03°$ for the whole sequence In addition, the method provided left (SL) and right (SR) shoulder angle averages of $43.46° \pm 24.86°$ and $27.85° \pm 12.02°$, respectively. These results produced a suitable upper arm position RULA score of +2 for most of the sequence, even under left arm self-occlusion in some takes (Figs. 9(a)–9(c)).

Likewise, in the *hammering* job video, the right side of the subject is always occluded (see Fig. 10). The right shoulder joint is detected in all frames with good confidence. However, when the right elbow is neither visible nor inferred, the method cannot calculate SR or ER. Nevertheless, our method was occasionally capable of inferring the locations of these occluded joints, thus enabling RULA grand score computation in these few frames. For instance, in the frame shown in Fig. 10(d), the worker's right elbow is clearly visible, and the right wrist could be inferred with good confidence. This enabled the computation of RULA local scores using SR and ER angles from Table 1. Under these sub-optimal conditions, our method generated a combined shoulder angle measurement of $36.80° \pm 13.12°$ for this video, thus yielding in most instances an upper arm position RULA score of +2.

Next, we applied our method in the *tree cutting* and *drilling* job videos, where it consistently provided two groups of skeleton joints for every frame of each sequence. This allowed us to simultaneously compute individual RULA scores for each worker. In the *tree cutting* sequence (Fig. 11), two workers carry out coordinated work activities around a tree, whereas in the *drilling* video (Fig. 12), two operators perform the same activity in a synchronized fashion to keep a heavy drill balanced. In the *tree cutting* sequence, the ergonomists selected by consensus the key-frame shown in Fig. 11(b) to analyze individual trunk angles for each worker. This choice was mainly based on the absence of self-occlusions (Fig. 11(a)), occlusions from working tools (Fig. 11(c)), and occlusions from the other worker (Fig. 11(d)), thus providing better conditions for assessing RULA scores. In the selected frame, the proposed method computed trunk flexion angles (TF) of about 29° and 18° for the left and right worker, respectively. It is worth noting that the method was also able to handle fully occluded ears (inferred at the safety earmuffs), providing accurate RULA neck scores (+4) from ME points despite this challenging situation.

Finally, in the *drilling* job sequence (Fig. 12), the specialists were most interested in examining RULA trunk score variations throughout the sequence. Our method accurately measured worker-averaged TFs from about 14° to 86°, providing valuable data for the ergonomists to rapidly detect snapshots with critical postures. Table 5 shows the RULA scores obtained in the five videos, and the Cohen's κ statistic among the outcomes of the proposed method and experts' assessments, as computed by individual score, and by individual video.

5. Discussion

The results shown in the previous section, following the different experimental settings detailed in Section 3, provide positive evidence regarding the feasibility of the method. First, joint detection confidence and ergonomic angle inference were evaluated as a function of the observer's viewpoint. The use of virtual 3D characters provided a suitable and systematic way of generating ground truth datasets, otherwise difficult and expensive to obtain (Li et al., 2018; Plantard et al., 2015). Latitudes less than 45° were chosen as most adequate viewpoints to keep body self-occlusions to a minimum, while azimuth selection is shown to be important only at joint level. Shoulder, neck, and trunk angle calculations were also thoroughly tested with ground truth datasets built from simulated 3D models and controlled experiments with actual people. These body-joint angles are of utmost importance in RULA. The use of two fiducial points (ME and MK) is shown to facilitate robust measurements of neck and trunk angles under self-occlusion conditions.

Experimental results also allowed to estimate suitable body-joint angular ranges and thresholds, as needed for computing RULA scores. Furthermore, video-based ergonomic evaluations of the method in actual outdoor working scenarios were carried out under the supervision of seven experienced ergonomists. Our method provided simultaneous robust RULA scores for more than one worker in video takes with varying illumination conditions, occlusions, and unstable camera viewpoints. The obtained results, shown in Table 5, were highly coincident pursuant to experts' assessments. These results show that reasonable variations in camera view do not influence the results in real working conditions significantly, as predicted by the experimental tests implemented in Section 3.

The scores of upper arms had the best agreement among the specialists and our proposal. This score is the most significant in the RULA assessment, and therefore this agreement provides significant backing to our system. On the other hand, the trunk scores had the least agreement (though 0.44 is considered moderate agreement in the literature). However, this coefficient is similar to the agreement among the experts

themselves, and therefore can be attributed at least partially to inter-subjective variation instead of variance in our system. Finally, there is full agreement in the RULA grand score. These results demonstrate the usefulness of CV in general, and the proposed method in particular, to perform ergonomic RA from pictures taken outdoors or under sub-optimal observation conditions. It is worth noticing that small discrepancies in RULA local scores are usually of little significance with respect to the RULA grand score. This is due to the hierarchical design of RULA, which uses weighted score aggregation as a method for reducing the impact of noise through its different stages (Plantard et al., 2017). This characteristic is undoubtedly favorable for our method, since body-joint location and projection errors are assimilated as noise.

Two arguably potential weaknesses should be pointed out. Skeleton detection biases in some cases may lead to relevant angle measurement deviations. For this reason, our method incorporates skeleton-based size and confidence thresholds to discard potentially defective detections. Improved results are generally obtained from higher resolution images, and at the same time body-joint confidence data may be used for building statistical location distributions around their given estimates. The second limitation stems from the fact that angular measurements are not computed from 3D body-joint estimates, but from 2D projections, which may raise projective distortions. This phenomenon can be controlled choosing the largest possible focal distance, and with adequate camera views in which the required body-joint angles are not occluded. Our method discourages the use of large vertical viewing angles, or camera views in which the workers are taken laterally most of the times. These two issues may still lead to mistakes and sometimes systematic biases, that may weaken the significance of a RULA estimation from a single snapshot. However, when many frames are jointly analyzed, and when the working posture to be measured is non-static as viewed from the camera, the accumulated information from these frames can be used to strengthen both the measurement significance and accuracy.

6. Conclusion and future work

We presented a CV-based method to perform ergonomic RA using off-the-shelf RGB cameras that automatically calculates RULA scores from digital images or video sequences. The development of the method is based on multiple tests and experiments, some of which have been discussed in detail. The method can simultaneously analyze multiple worker postures, either individually or collectively, under challenging acquisition conditions which are typically found in actual working scenarios, especially in outdoor workplaces. To the best of our knowledge, this is the first CV-based ergonomic RA method able to produce consistent and accurate results under these conditions, which makes it operational in-the-open.

The evolution of Deep Learning technologies is expected to provide even more accurate and flexible open-source CV libraries, which may enhance the features of our proposal in the near future and to mitigate the potential drawbacks mentioned above. As noted in Xu and McGorry (2015), more improvements will also be necessary to meet more challenging RA-related tasks. For instance, it is still unknown if gender, somatotypes, or ethnicity exert a significant influence in joint detection precision. For this reason, a prior anthropometric feature evaluation might be required to devise a model that can adapt itself to these and other particular conditions. Also, in addition to RULA, there are other RA proposals that seem valuable to incorporate, such as the Rapid Entire Body Assessment (REBA) and the modified RULA for computer workers (MRULA), as both of these methods are built on RULA principles. Finally, in the context of I4.0, real-time body posture acquisition is valuable to perform data-fusion with other information sources, specifically in the discrete and process manufacturing industries, to extract latent information about working conditions and logistic optimization. Also, these CV-based technologies can be used to classify human activities in addition to assess the postures' and movements' safeness. This combined assessments will open new ergonomic venues for analysis and interpretation, for instance per activity, and possibly per time and location. At the same time, CV-based object detection can be articulated together with ergonomic analyses, in order to automatically supervise and rate the proper use of personal protective equipment.

CRediT authorship contribution statement

Manlio Massiris Fernández: Conceptualization, Methodology, Software, Data curation, Validation, Writing - original draft. **J. Álvaro Fernández:** Investigation, Methodology, Formal analysis, Visualization, Writing - original draft, Writing - review & editing, Supervision, Resources. **Juan M. Bajo:** Visualization, Investigation, Methodology, Software, Validation, Data curation. **Claudio A. Delrieux:** Investigation, Methodology, Formal analysis, Visualization, Writing - original draft, Writing - review & editing, Supervision, Resources, Funding acquisition.

Acknowledgments

The authors wish to thank all the workers and experts who helped us in the development of this research, with a special mention to Mrs. Mercedes del Castillo, Mr. Guillermo Dominella, Mr. Pablo García, Mrs. Beatriz Camacho and Mrs. Eugenia Pavón, of the Occupational Health and Safety advisors at the Universidad Nacional del Sur (https://uns.edu.ar), and Mr. Carlos Segui, Mr. Víctor Delgado, Mr. Guido de la Mano, and Mr. Alberto Carnevali from the Bahía Blanca Harbor Management Consortium (https://puertobahiablanca.com).

This research was supported by the National Council of Scientific and Technical Research of Argentina - CONICET; the Spanish Regional Government of Extremadura (Junta de Extremadura) through the European Regional Development Fund (code GR18135); and the Universidad Nacional del Sur (grant code 24/K083).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cie.2020.106816.

References

Battini, D., Persona, A., & Sgarbossa, F. (2014). Innovative real-time system to integrate ergonomic evaluations into warehouse design and management. *Computers & Industrial Engineering*, *77*, 1–10.

Bevan, S. (2015). Economic impact of musculoskeletal disorders (MSDs) on work in Europe. *Best Practice & Research Clinical Rheumatology*, *29*(3), 356–373.

Bureau of Labor Statistics (2016). Economic news release. nonfatal occupational injuries and illnesses requiring days away from work. Washington, DC, USA: NIOSH, USDL-16-2130.

Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. In *2017 IEEE conference on computer vision and pattern recognition* (pp. 1302–1310).

Chiasson, M.-È., Imbeau, D., Aubry, K., & Delisle, A. (2012). Comparing the results of eight methods used to evaluate risk factors associated with musculoskeletal disorders. *International Journal of Industrial Ergonomics*, *42*(5), 478–488.

Cohen, A. L. (1997). *Elements of ergonomics programs: A primer based on workplace evaluations of musculoskeletal disorders*. Department of Health and Human Services (NIOSH), DIANE Publishing.

Cui, J., & Freeden, W. (1997). Equidistribution on the sphere. *SIAM Journal on Scientific Computing*, *18*(2), 595–609.

David, G. C. (2005). Ergonomic methods for assessing exposure to risk factors for work-related musculoskeletal disorders. *Occupational Medicine*, *55*(3), 190–199.

Diego-Mas, J. A., & Alcaide-Marzal, J. (2014). Using Kinect™ sensor in observational methods for assessing postures at work. *Applied Ergonomics*, *45*(4), 976–985.

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

Fığlalı, N., Cihan, A., Esen, H., Fığlalı, A., Çeşmeci, D., Güllü, M. K., & Yılmaz, M. K. (2015). Image processing-aided working posture analysis: I-OWAS. *Computers & Industrial Engineering*, *85*, 384–394.

Gerr, F., Fethke, N. B., Merlino, L., Anton, D., Rosecrance, J., Jones, M. P., Marcus, M., & Meyers, A. R. (2013). A prospective study of musculoskeletal outcomes among manufacturing workers. *Human Factors*, *56*(1), 112–130.

Golabchi, A., Han, S., Seo, J., Han, S., Lee, S., & Al-Hussein, M. (2015). An automated biomechanical simulation approach to ergonomic job analysis for workplace design. *Journal of Construction Engineering and Management*, *141*(8), 1–12.

Ha, C., Roquelaure, Y., Leclerc, A., Touranchet, A., Goldberg, M., & Imbernon, E. (2009). The French musculoskeletal disorders surveillance program: Pays de la Loire network. *Occupational and Environmental Medicine*, *66*(7), 471–479.

Helliwell, P. (2004). Repetitive strain injury. *Postgraduate Medical Journal*, *80*(946), 438–443.

Jayaram, U., Jayaram, S., Shaikh, I., Kim, Y., & Palmer, C. (2006). Introducing quantitative analysis methods into virtual environments for real-time and continuous ergonomic evaluations. *Computers in Industry*, *57*(3), 283–296.

Joshi, M., & Deshpande, V. (2019). A systematic review of comparative studies on ergonomic assessment techniques. *International Journal of Industrial Ergonomics*, *74*(October), 1–15.

Kadir, B. A., Broberg, O., & da Conceição, C. S. (2019). Current research and future perspectives on human factors and ergonomics in Industry 4.0. *Computers & Industrial Engineering*, *137*, 1–12.

Kong, Y.-K., yong Lee, S., Lee, K.-S., & Kim, D.-M. (2017). Comparisons of ergonomic evaluation tools (ALLA, RULA, REBA and OWAS) for farm work. *International Journal of Occupational Safety and Ergonomics*, *24*(2), 218–223.

Krüger, J., & Nguyen, T. D. (2015). Automated vision-based live ergonomics analysis in assembly operations. *CIRP Annals*, *64*(1), 9–12.

Lee, W., Seto, E., Lin, K. Y., & Migliaccio, G. C. (2017). An evaluation of wearable sensors and their placements for analyzing construction worker's trunk posture in laboratory conditions. *Applied Ergonomics*, *65*, 424–436.

Levanon, Y., Lerman, Y., Gefen, A., & Ratzon, N. Z. (2014). Validity of the modified RULA for computer workers and reliability of one observation compared to six. *Ergonomics*, *57*(12), 1856–1863.

Li, X., Han, S., Gül, M., Al-Hussein, M., & El-Rich, M. (2018). 3D visualization-based ergonomic risk assessment and work modification framework and its validation for a lifting task. *Journal of Construction Engineering and Management*, *144*(1), 1–13.

Liu, M., Han, S., & Lee, S. (2016). Tracking-based 3D human skeleton extraction from stereo video camera toward an on-site safety and ergonomic analysis. *Construction Innovation*, *16*(3), 348–367.

Liu, M., Han, S., & Lee, S. (2016). Tracking-based 3D human skeleton extraction from stereo video camera toward an on-site safety and ergonomic analysis. *Construction Innovation*, *16*(3), 348–367.

Luttmann, A., Jäger, M., Griefahn, B., Caffier, G., Liebers, F., & Steinberg, U. (2003). *Protecting worker's health series, no. 5*, *Preventing musculoskeletal disorders in the workplace*. Geneva, Switzerland: World Health Organization.

McAtamney, L., & Corlett, E. N. (1993). RULA: a survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, *24*, 91–99.

Nath, N. D., Akhavian, R., & Behzadan, A. H. (2017). Ergonomic analysis of construction worker's body postures using wearable mobile sensors. *Applied Ergonomics*, *62*, 107–117.

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*, *15*, 1785–1803.

Plantard, P., Shum, H. P., Pierres, A.-S. L., & Multon, F. (2017). Validation of an ergonomic assessment method using Kinect data in real workplace conditions. *Applied Ergonomics*, *65*, 562–569.

Rauch, E., Linder, C., & Dallasega, P. (2019). Anthropocentric perspective of production before and within Industry 4.0. *Computers & Industrial Engineering*, *105644*, 1–15.

Roman-Liu, D. (2014). Comparison of concepts in easy-to-use methods for MSD risk assessment. *Applied Ergonomics*, *45*(3), 420–427.

Savino, M. M., Battini, D., & Riccio, C. (2017). Visual management and artificial intelligence integrated in a new fuzzy-based full body postural assessment. *Computers & Industrial Engineering*, *111*, 596–608.

Seo, J., Alwasel, A., Lee, S., Abdel-Rahman, E. M., & Haas, C. (2019). A comparative study of in-field motion capture approaches for body kinematics measurement in construction. *Robotica*, *37*(5), 928–946.

Vignais, N., Bernard, F., Touvenot, G., & Sagot, J.-C. (2017). Physical risk factors identification based on body sensor network combined to videotaping. *Applied Ergonomics*, *65*, 410–417.

Vignais, N., Miezal, M., Bleser, G., Mura, K., Gorecky, D., & Marin, F. (2013). Innovative system for real-time ergonomic feedback in industrial manufacturing. *Applied Ergonomics*, *44*, 566–574.

Xu, X., & McGorry, R. W. (2015). The validity of the first and second generation Microsoft Kinect for identifying joint center locations during static postures. *Applied Ergonomics*, *49*, 47–54.

Xu, X., Robertson, M., Chen, K. B., hua Lin, J., & McGorry, R. W. (2017). Using the Microsoft Kinect™ to assess 3-D shoulder kinematics during computer use. *Applied Ergonomics*, *65*, 418–423.

Yan, X., Li, H., Wang, C., Seo, J., Zhang, H., & Wang, H. (2017). Development of ergonomic posture recognition technique based on 2D ordinary camera for construction hazard prevention through view-invariant features in 2D skeleton motion. *Advanced Engineering Informatics*, *34*, 152–163.

Yu, Y., Yang, X., Li, H., Luo, X., Guo, H., & Fang, Q. (2019). Joint-level vision-based ergonomic assessment tool for construction workers. *Journal of construction engineering and management*, *145*(5), 1–13.

Zhang, H., Yan, X., & Li, H. (2018). Ergonomic posture recognition using 3D view-invariant features from single ordinary camera. *Automation in Construction*, *94*, 1–10.