Automatic assessment of posture deviations in assembly tasks

Konstantinos Papoutsakis[1,] Manolis Lourakis[1], Maria Pateraki [1,2]

[1] Institute of Computer Science,
Foundation for Research and Technology,
Heraklion, Greece
[2] National Technical University of Athens,
Greece

ABSTRACT

The aim of this study is to investigate the development and the evaluation of a computer vision-based framework to aid the automatic assessment of posture deviations in assembly tasks in realistic work environments. A posture deviation refers to a time-varying working posture performed by the worker, that deviates from ergonomically safe body postures expected in the context of particular work tasks and is known to impose increased physical strain. The estimation of their occurrences can serve as indicators, known as risk factors, for the assessment of physical ergonomics towards the prevention of physical strain and in the-long-term of work-related musculo-skeletal disorders (WMSD). Using visual information acquired by camera sensors, our goal is to estimate the full body motion of a line worker in 3D space, unobtrusively, and to perform classification of four types of posture deviations, also noted as ergonomically sub-optimal working postures that were employed by the MURI risk analysis tool. We formulate a learning-based action classification task using Deep Graph-based Neural Networks and differential temporal alignment cost as a classification measure to estimate the type and risk level of the observed posture deviation during work activities. To evaluate the efficiency of the proposed approach, a new video dataset was captured in the context of the sustAGE project, that demonstrate two different workers during car door assembly actions in a simulated production line in an actual workplace. Rich annotation data were provided by experts

in manufacturing and ergonomics. Both quantitative and qualitative evaluation of the proposed framework provide evidence for its efficiency and reliability in supporting ergonomic risk assessment and preventive actions for WMSD in real working environments.

Keywords: working postures, ergonomic risk analysis, manufacturing, occupational safety, deep learning, computer vision

INTRODUCTION

The assessment of ergonomic risks for the prevention of work-related musculo-skeletal disorders (WMSD) is considered a common and critical task related to occupational safety and well-being in work environments (Papoutsakis, et al., 2021). Especially in the manufacturing industry, labor-intensive assembly works attribute repetitive tasks, often in sustained, awkward working postures (Vieira & Kumar, 2004) that lead to increased physical discomfort and stress, according to several studies on physical ergonomics (Brito, et al., 2019) (Bao, et al., 2020). Such a dynamic working posture, noted as ergonomically sub-optimal posture, refers to a sequence of upper- or whole- body configurations or poses of a certain minimum duration that deviates from safe postures expected in the context of particular work activities. Each type of sub-optimal working posture imposes increased physical stress to body joints or parts, while assessing their occurrences serve as risk indicators, which are known as risk factors, for WMSD.

Our case study in this work addresses the car manufacturing industry and is part of the susAGE system[1], which is developed to provide a person-centered smart solution to support the employment, safety and health of ageing workers in occupational contexts. An actual car manufacturing workplace is considered that is available at the CRF-SPW Research & Innovation department of the Stellantis group in Melfi, Italy. In this context, we specifically focus on line workers that work in shifts, each on a single workstation of a simulated car door production line, as shown in Figure 1. Each worker executes a specific set of car door assembly activities, noted as task cycle, that lasts for 4 to 5 minutes and is continuously repeated during her shift.

One of the novelties of sustAGE is the adoption and integration of the *Micro-Moments* (MiMos) concept (Athanassiou, et al., 2021). MiMos are used to digitize interactions with the physical environment, repeated patterns, or events occurring in workers' daily living routines and they link with recommended actions targeted directly at the workers themselves or at their supervisors. By issuing recommendations through MiMos, the system capitalizes on the early detection and avoidance of risky and stressful conditions that affect the performance of individual

[1] http://www.sustage.eu

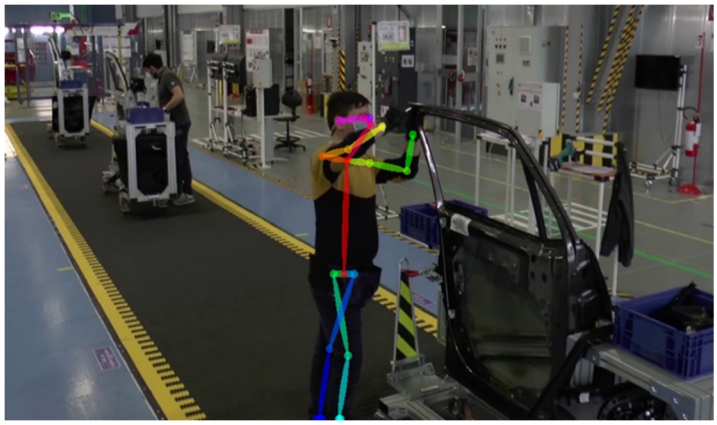

Figure 1. The car door production line at CRF that comprises three assembly workstations.

workers or worker groups. Accordingly, recommendations reach the users at the right moment and place, and proposed actions match the users' preferences and current needs. Therefore, for being able to issue recommendations and prevent risky situations in the actual work environment, it is important to monitor the workers' body motion and early detect such situations and their underlying conditions.

In this direction, we aim to detect posture deviations of increased ergonomic risk for physical strain. To achieve this, we rely on visual information acquired by low-cost cameras placed along the simulated production, as shown in Figure 1. Our goal is to estimate and track, in real-time and unobtrusively, the full human body motion in 3D space, in the presence of severe and possibly long-term occlusions, and to classify the ergonomically sub-optimal posture deviations performed during assembly activities. The module that process this information provides the sustAGE system with information on the detected events and trigger personalized recommendations to the workers for preventive actions aiming to enhance occupational safety.

The main contributions of this work regard: (a) an unobtrusive and low-cost solution for the unobtrusive, automatic classification of posture deviations during work activities using visual information. It relies on a novel combination of Graph-based Convolutional Networks (Yan, et al., 2018) and the soft-DTW method (Cuturi & Blondel, 2017) for pairwise temporal alignment of 3D skeletal data sequences. (b) A new dataset that comprises synchronized color and depth image sequences of car door assembly activities captured in an actual manufacturing environment. Annotation data is available for the assembly actions and posture deviations ergonomics according to the MURI risk analysis method (Womack, 2006).

In the following Sections, we elaborate on the proposed methodology, the visual data acquisition and annotation processes as well as the experimental evaluation of the proposed method using the new video dataset. Finally, the last section reports the main findings of this work and summarizes our next steps.

VISUAL DETECTION OF POSTURE DEVIATIONS IN ASSEMBLY TASKS

The main parts of the proposed methodology are described in the following sections. Visual information acquired using low-cost camera sensors installed in the actual workplace feeds two state-of-the-art deep-learning based methods that efficiently estimate the skeleton-based human poses in 2D and finally in 3D space, unobtrusively and in the presence of body occlusions. Then, a novel deep-learning based classification approach is proposed for classifying the observed posture deviation based on sub-optimal working postures indicated by the MURI ergonomic risk analysis method (Womack, 2006).

We are interested in modelling and classifying the set of posture deviation, as shown in Figure 2, that regard a set of time-varying, ergonomically sub-optimal working postures based on the MURI risk screening method~\cite{womack2006lean}. According to the World Class Manufacturing strategy (WCM) (Schonberger, 2008) the MURI analysis is a generic and widely-used tool for efficiency evaluation and screening of the physical ergonomics in workstations in different production contexts (Brito, et al., 2019) and especially in the automotive industry. We focus on four main types (sketches) of working posture deviations used by the MURI risk analysis method, as shown in Figure 2. Each main type is further analyzed into three postural variants that are associated with increasing levels of ergonomic risk for physical strain/stress imposed to specific body parts and joints during work activities. These variants refer to the low ('Level 3'), medium ('Level 2') and high ('Level 1') risk level, that are quantified according to specific criteria linked to the pose-based angles and positions of the body parts. The low risk variants of the postures correspond to a neutral body pose of low or no ergonomic risk for physical strain. Each working posture is realized as time-varying event; thus, a sequence of body configurations with a duration of at least 4 seconds. Therefore, we define a set of nine target classes of working postures that comprise the high-risk ('Level 1') and the medium-risk ('Level 2') variants for each of the four main types of working postures and a single class of low-risk or normal body posture for all four types.

Visual Human Pose Estimation

Given a sequence of N images that shows a line worker performing a single or multiple assembly actions for a workstation, we employ state-of-the-art deep-learning-based methods for the estimation of the skeleton-based pose (or posture) of the human body per image. Specifically, we use the popular OpenPose method (Cao, et al., 2019) to obtain a set of 2D image coordinates (x,y), that is the locations of 25 skeletal body joints according to the BODY25 pose output model[2]. The set of 2D

[2] https://cmu-perceptual-computing-lab.github.io/openpose/web/html/doc/md_doc_02_output.html

coordinates of the body joints per image is subsequently used to feed the MocapNet2 method (Qammaz & Argyros, 2021) that relies on ensembles of Deep Neural Networks to efficiently regress a view-invariant skeleton-based pose in 3D space.

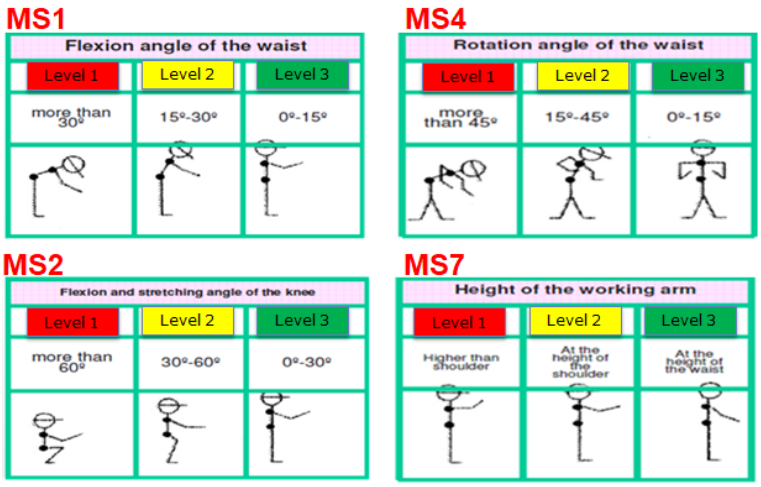

Figure 2.Four types of ergonomic working postures according to the MURI risk analysis method. Deviations per type of working postures regard three variants labeled as Level 1 (red), 2 (yellow) to 3 (green) that corresponds to high, medium and low ergonomic risk.

Moreover, the hierarchy of the 3D skeletal body model is split into the upper and the lower body parts that are estimated independently. The output of the MocapNet2 method is a Biovision Hierarchy (BVH) character animation file format (Meredith, et al., 2001) representing the estimated 3D human motion for the input image sequence. Based on this information, we extract a set of view-invariant, user-centric 3D coordinates (x,y,z) of 25 main body joints per frame with respect to the body torso (Theodorakopoulos, et al., 2014). Essentially, this information constitutes a 3D skeletal data sequence; a multi-dimensional (75D) time-series of length N. The following body joints are considered based on the body configuration used in the NTU RGB+D 120 dataset/benchmark for 3D human understanding (Liu, et al., 2019): base/middle/upper spine, neck, head, left/right shoulder, left/right elbow, left/right wrist, left/right hand, tip of left/right hand, left/right thumb, left/right hip, left/right knee, left/right ankle, left/right foot.

Visual Classification of Posture Deviations

We use the estimated 3D skeletal sequence to feed a Spatio-temporal Graph Convolutional Network model (ST-GCNs) (Yan, et al., 2018) for learning efficient discriminative representations of the spatio-temporal dynamics of the human postures and actions, as shown in Figure 3. The ST-GCN model represents the locations and the dependencies of the human skeletal joints, as graph edges in a spatial graph-based CNN for each video frame. The temporal aspect of the graph is constructed by

connecting the same joints across consecutive frames to model the spatial temporal dynamics of the human motion.

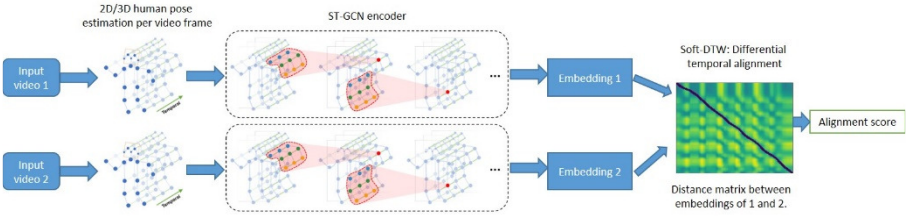

Figure 3. An overview of the proposed approach for posture deviation classification is shown.

The last layer of the ST-GCN model provides a 256-dimensional feature vector, that is considered as an encoding or embedding of the input sequence to a new feature space. Then, a SoftMax classifier is used to transform the embedding values of the network to probabilities towards the target classes. We train a ST-GCN model using 3D skeletal sequence as input and a modified SoftMax layer as output towards the set of nine target classes (see Figure 2) in order to optimize for the network weights and for learning to encode the input information of user-centric 3D skeletal poses into a new shared embedding space, as shown in Figure 3.

The next step of the proposed approach is to define a metric for the pairwise comparison of an unlabeled skeletal sequence X with one or more labeled skeletal sequences Y_i assigned to each of the target classes C_i of working postures. The metric scores will be used to classify the unlabeled sequence X to one with the lowest score; the most similar C_i class. To this end, we the soft Dynamic Time Warping approach (softDTW) (Blondel, et al., 2021) to estimate the temporal alignment cost between sequences as our classification metric, as shown in the outline of the proposed approach in Figure 3. To solve a minimal-cost temporal alignment problem between two data sequences, the Dynamic Time Warping (DTW) approach (Garreau, et al., 2014) employs dynamic programming and uses the Euclidean distance to measure the discrepancy between all pairs of time-stamped values of the data sequences (or feature vectors in case of multidimensional data). soft-DTW extends the DTW and computes the soft-minimum of all alignment costs, providing also a differentiable loss function that can be computed with quadratic time/space complexity.

VISUAL DATA ACQUISITION

Data acquisition setup

To facilitate the implementation and evaluation of the proposed methodology, we collected synchronized image and depth sequences during car-door assembly activities. Those were performed by actual line workers for a simulated production line that is available in a realistic manufacturing workplace at the CRF-SPW Research & Innovation department of the Stellantis group in Melfi, Italy. In this realistic

setting, three assembly workstations are sequentially arranged on a conveyor belt that moves at a low, constant speed realizing a continuous workflow of the car-door assembly process. A single line-worker is assigned to each workstation to perform a sequence of up to 25 assembly actions of total duration between 4-5 minutes, noted as task cycle. We follow a low-cost, unobtrusive (non-invasive) sensing approach using camera sensors for data acquisition that allows workers to perform ordinary assembly activities in the real working environment while capturing visual data without the need for the installation of special expensive equipment and wearable suits/reflectors (i.e., a motion capture system). Thus, the proposed solution is potentially applicable across the whole production line. Specifically, cameras are placed at stationary positions at each side of the production line, two cameras per side, to simultaneously monitor the workers' assembly activities from both the inner-door and outer-door working areas of the workstations. Each camera captures a stereo pair of RGB (color) image sequences and a depth-based image sequence of resolution 1080p at 30 frames per second, while visual data sequences acquired by all cameras are time-synchronized using a common reference clock.

MURI ANALYSIS
TASK DESCRIPTION: User 124_WS30_May 24 morning

Assembly front door

Level 1 =red (high risk) 3 points
Level 2 =yellow (meduium risk) 2 points
Level 3 =green (low risk) 1 point

					FLEXION ANGLE OF THE WAIST			ROTATION ANGLE OF THE WAIST			HEIGHT OF THE WORKING ARM			FLEXION AND STRETCHING ANGLE OF THE KNEE		
					3	2	1	3	2	1	3	2	1	3	2	1
N	operazione	ACTIVITY	Execution time in seconds (s) start ts	end te	> 30°	15° ÷ 30°	0° ÷ 15°	> 45°	15° ÷ 45°	0° ÷ 15°	> SHOULDER	= SHOULDER	= WAIST	> 60°	30° ÷ 60°	0° ÷ 30°
1	10	Take screwdriver on line side and 4 screws from pouch	7:47	8:00			1			1			1			1
2		Take one front left speaker from carriage and insert it correctly into retaining slots on carrier	8:00	8:08		2				1			1			1
3		Place one screw at a time on the screwdriver tip and tighten four screws on the loudspeaker as shown in the sketch.	8:08	8:36	3				2				1			1
4		Leave the screwdriver on the line side	8:36	8:45			1			1			1			1
5	20	Take the door panel from cart and place on top of door frame.	8:45	8:50		2			2				1			1
6		With both hands hook the superior part of the door panel on the door frame (inferior part of the window)	8:50	9:15			1		2			2			2	
7	30	Take the screwdriver from the line side and take N° 1 screw from the pouch, Place the screw on the screwdriver bit and fix	9:15	9:29	3				2				1			1
8		Take 1 screw from the pouch , Place the screw on the screwdriver bit Fix the panel on the window lifter area	9:29	9:45	3				2				1	3		
9		Leave the screwdriver on the line side	9:45	9:51			1			1			1			1

Figure 4. Annotation data shows the instances of posture deviations and ergonomic risk scores observed during part of a task cycle based on the MURI analysis method (Womack, 2006).

Data annotation

The data collected, formed a new video dataset with 12 task cycle executions of time-synchronized visual data; i.e. RGB and depth image sequences. Annotation data for the recorded task cycles were provided by experienced professionals in manufacturing and ergonomics. As shown in the example annotation data table in Figure 4, annotation include: (a) the temporal boundaries (start and end timestamps) and the semantic label for each assembly action (one action per row) performed by

the worker during the task cycle, (b) the instances of the target types of ergonomic working postures of interest (noted in columns) for each assembly action (row), and (c) the overall ergonomic risk score for the task cycle execution estimated according to the MURI risk analysis method, as shown in Figure 2. Annotations toward the three risk levels for each type of working posture correspond to high, medium, low risk as semantic labels, to red, yellow and green as color-coded labels, and to integers 1, 2, 3 as numerical scores, respectively. The annotated task cycles comprise 310 assembly actions, each of average duration of 13 seconds. Moreover, the annotated instances of Level 1, Level 2 and Level 3 posture deviations for all types of working postures are 38, 126 and 818, respectively.

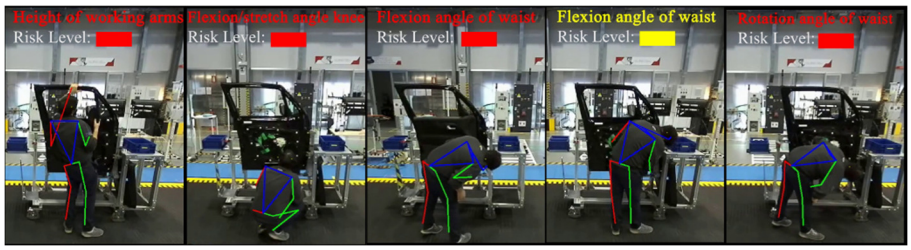

Figure 5. Qualitative results of the proposed method for the classification of the ergonomic working postures. Overlaid information regards the estimated 3D body pose and the identified types and ergonomic risk level for physical strain of the detected working postures.

EXPERIMENTAL EVALUATION

The performance of the postural classification approaches is evaluated using the videos of the segmented assembly actions. A subset of 220 and 90 videos were used as training/validation and test samples, respectively. Then, our goal is to classify each video clip against the set of target classes of ergonomic postures defined in Figure 2. Qualitative results of the proposed approach are demonstrated in Figure 5. To measure the quantitative performance of the classification task, we employ the metrics of Precision, Recall and F1 score that are commonly used in the fields of statistics, data science and information retrieval. The F1 score metric is the harmonic mean of precision (ratio of the positive occurrences towards a category that are actually correct) and recall (ratio of actual positive occurrences that were classified correctly) in the range [0, 1], where 1 indicates perfect precision and recall: $F_1 score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$.

The efficiency of the proposed deep-learning based classifier is reported in Table 1, based on the F1 scores and the average values for all types of ergonomic working posture deviations. The poor performance for the posture type of the flexion-stretching-angle-of-the-knee is mainly due to the low quality of 3D body pose estimations during this time-varying body posture as the knees are self-occluded mainly by the upper legs and possibly by the waist. The overall performance marginally exceeds 70% of correctly classified posture deviations, which is

considered satisfactory considering the challenging conditions for acquiring high quality estimation of the human motion in real-world manufacturing environments.

Table 1: Quantitative results of the deep learning-based approach for the classification of the ergonomic working postures, as shown in Figure 2.

Posture deviations	*Flexion angle of the waist*	*Rotation angle of the waist*	*Height of the working arm*	*Flexion/stretching angle of the knee*
Risk level	*L3 / L2 / L1*	*L3 / L2 / L1*	*L3 / L2 / L1*	*L3 / L2 / L1*
F1 score	*0.73/0.80/0.90*	*- /0.63/0.90*	*0.80/0.62/0.89*	*0.25/0.38/0.88*
AVG F1 score	*0.831*	*0.766*	*0.771*	*0.504*

CONCLUSION

Based on the experimental evaluation conducted using the newly compiled video dataset, the proposed vision-based approach is able to accurately capture and monitor the fine-grained body motion of the line workers in 3D space and to effectively estimate posture deviations during assembly activities, as ergonomic risk indicators for physical strain. Beyond monitoring, the proposed framework can also be used to automate the task of physical ergonomic risk assessment during computer-aided evaluation and optimization of manufacturing workflows, either at the design or the production stage, and possibly for automated computation of visual analytics during manufacturing tasks. Overall, it can be applied to highlight process variability and quality assurance for manufacturing tasks in various industrial contexts and to help industrial and lean engineers to jointly balance the workload in production lines, ensure worker safety and enhance productivity. Future goals are related to the development of an end-to-end deep neural architecture for the temporal localization and recognition of an enriched set of target posture deviations using an RGB image sequence or skeletal-based data information as input and to link our approach to an ergonomic risk index/checklist (e.g. OCRA checklist, EAWS methodology) for the automatic assessment of physical ergonomics.

ACKNOWLEDGMENTS

The authors would like to acknowledge the consortium partners Stellantis - Centro Ricerche FIAT (CRF) / SPW Research \& Innovation department in Melfi, Italy, for their valuable feedback in the design, implementation, visual data acquisition and evaluation of this study.

REFERENCES

Athanassiou, G., Pateraki, M. & Varlamis., I., (2021). Micro-moment-based Interventions for a Personalized Support of Healthy and Sustainable Ageing at Work: Development and Application of a Context-sensitive Recommendation Framework. s.l., SciTePress, pp. 409-419.

Bao, S., Howard, N. & Lin, J.-H., (2020). Are work-related musculoskeletal disorders claims related to risk factors in workplaces of the manufacturing industry?. Annals of work exposures and health, Volume 64, p. 152–164.

Brito, M. F., Ramos, A. L., Carneiro, P. & Gonçalves, M. A., (2019). Ergonomic Analysis in Lean Manufacturing and Industry 4.0—A Systematic Review. In: Lean Engineering for Global Development. Springer International Publishing, p. 95–127.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.E., & Sheikh, Y., (2019). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 43, p. 172–186.

Cuturi, M. & Blondel, M., (2017). Soft-DTW: A Differentiable Loss Function for Time-Series. In Proceedings of the 34th International Conference on Machine Learning - Volume 70 (ICML'17). JMLR.org, 894–903.

Garreau, D., Lajugie, R., Arlot, S. & Bach, F., (2014). Metric learning for temporal sequence alignment. s.l., MIT Press, p. 1817–1825.

Liu, J., Shahroudy, A., Perez, M., Wang, G., Duan, L.Y. and Kot, A.C., (2019). NTU RGB+D 120: A Large-Scale Benchmark for 3D Human Activity Understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Meredith, M., Maddock, S. & others, (2001). Motion capture file formats explained. Department of Computer Science, University of Sheffield, Volume 211, p. 241.

Papoutsakis, K. Papadopoulos T., Maniadakis M., Lourakis M., Pateraki M., & Varlamis I., (2021). Detection of physical strain and fatigue in industrial environments using visual and non-visual sensors. PETRA 2021, pp. 270-271.

Pateraki, M., Fysarakis, K., Sakkalis, V., Spanoudakis, G., Varlamis, I., Maniadakis, M., Lourakis, M., Ioannidis, S., Cummins, N., Schuller, B. & Loutsetis, E., (2020). Biosensors and Internet of Things in smart healthcare applications: Challenges and opportunities. In Wearable and Implantable Medical Devices (pp. 25-53). Academic Press.

Qammaz, A. & Argyros, A. A., (2021). Occlusion-tolerant and personalized 3D human pose estimation in RGB images. In International Conference on Pattern Recognition (ICPR), 2021.

Schonberger, R. J., (2008). World class manufacturing. Simon and Schuster.

Theodorakopoulos, I., Kastaniotis, D., Economou, G. & Fotopoulos, S., (2014). Pose-based human action recognition via sparse representation in dissimilarity space. Journal of Visual Communication and Image Representation, Volume 25.

Vieira, E. R. & Kumar, S., 2004. Working postures: a literature review. Journal of occupational rehabilitation, Volume 14, p. 143–159.

Womack, J. P., (2006). From lean tools to lean management. Lean Enterprise Institute Email Newsletter, Volume 21.

Yan, S., Xiong, Y. & Lin, D., (2018). Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition. In AAAI conference 2018.