RGB-D ergonomic assessment system of adopted working postures

Paper Seminar Report

EN 601.656 Computer Integrated Surgery II

Eric Han | ehan12@jhu.edu

Project Summary

Providing a healthy and safe work environment for surgeons is of utmost importance to hospitals. However, being forced in static positions for prolonged periods of time is something that surgeons cannot seem to avoid. This is why it is crucial to develop a method to evaluate surgical ergonomics in the workplace and inform surgeons of any potentially dangerous postures before any sort of work-related injury occurs. To accomplish this, we are using a RGB camera with a depth sensor to monitor a person and converting this data into a skeleton framework where we can accurately use the locations of joints in 3D space to return an assessment score based on the RULA (Rapid Upper Limb Assessment) and ROSA (Rapid Office Strain Assessment) evaluation.

Paper

The paper selected is :

Ahmed Abobakr, Darius Nahavandi, Mohammed Hossny, Julie Iskander, Mohammed Attia, Saeid Nahavandi, Marty Smets, RGB-D ergonomic assessment system of adopted working postures, Applied Ergonomics, Volume 80, 2019, Pages 75-88, ISSN 0003-6870, https://doi.org/10.1016/j.apergo.2019.05.004.

The purpose of the study outlined in the paper was to present a semi-automated ergonomic assessment system of adopted working postures in a holistic and budget friendly manner by relying on deep learning neural network techniques. In this paper, they discuss the proposed methods using the deep ConvNet models and the results obtained from these experiments as well as the limitations of their model.

Summary and Key Result

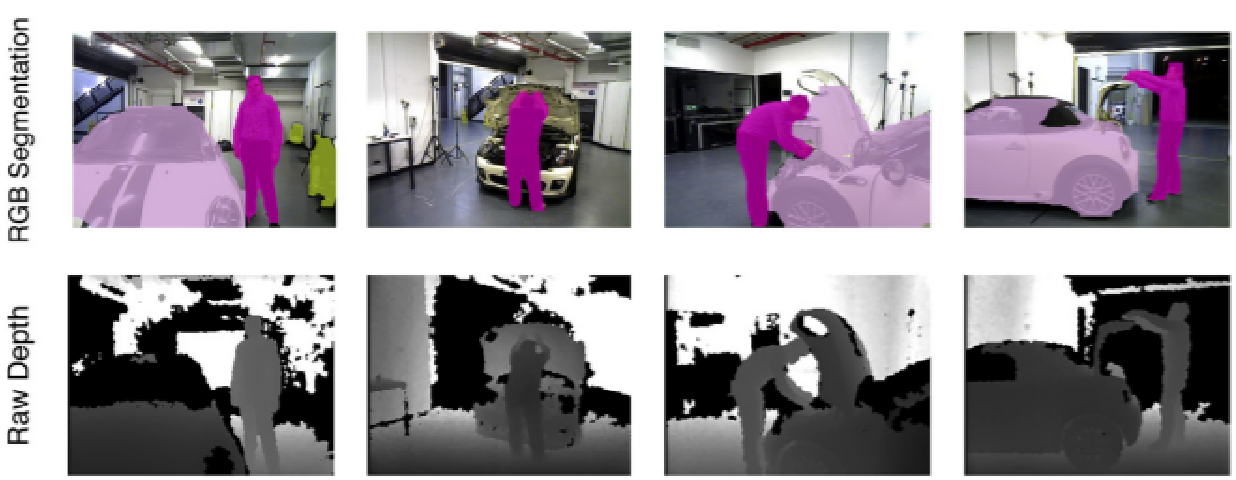


Figure 1 : RGB and depth image pairs

This paper implements a vision based ergonomic posture assessment system comprising two cascaded ConvNets, one of which is an object instance segmentation network and the other a holistic posture analysis network. The segmentation network segments the RGB portion of the image, creating a mask that is placed on the depth portion of the image, shown in figure 1. Then, the ConvNet directly maps the segmented depth image to human body joints and predicts joint angles. Finally, the estimated joint angles are used to compute the RULA score. This is possible because the models are trained on a large amount of highly varied synthetic training images with ground truth joint angles that have been biomechanically modeled using a novel inverse kinematics step.

Introduction and Background

Musculoskeletal disorders commonly occur throughout labor intensive workplaces and are a common concern for employers. A statistical study performed by the Bureau of Labor Statistics showed that musculoskeletal disorders account for 31% of all work-related injuries and illness cases (Bureau of Labor Statistic, 2016). The Rapid Upper Limb Assessment (RULA) (McAtamney and Corlett, 1993) is one of the most popular ergonomic assessment tools in the industry (Plantard et al. Multon; Liebregts et al., 2016) because it is simple, easy to compute and does not require prior knowledge in biomechanics or ergonomics. Recent studies have used the Kinect camera along with its software development kit to analyze the adopted posture and evaluate the RULA score (Plantard et al., 2015; Liebregts et al., 2016; Plantard et al., 2017; Manghisi et al., 2017; Abobakr et al., 2017a). However, the limitations to this are that it relies on local body part detectors and may produce unrealistic skeletons in cases of occlusions due to cluttered environments (Abobakr et al., 2018; Plantard et al., 2017) and it has difficulty when tracking self-occluded postures that have arms crossing, trunk bending, trunk lateral flexion and trunk rotation (Manghisi et al., 2017).

Dataset

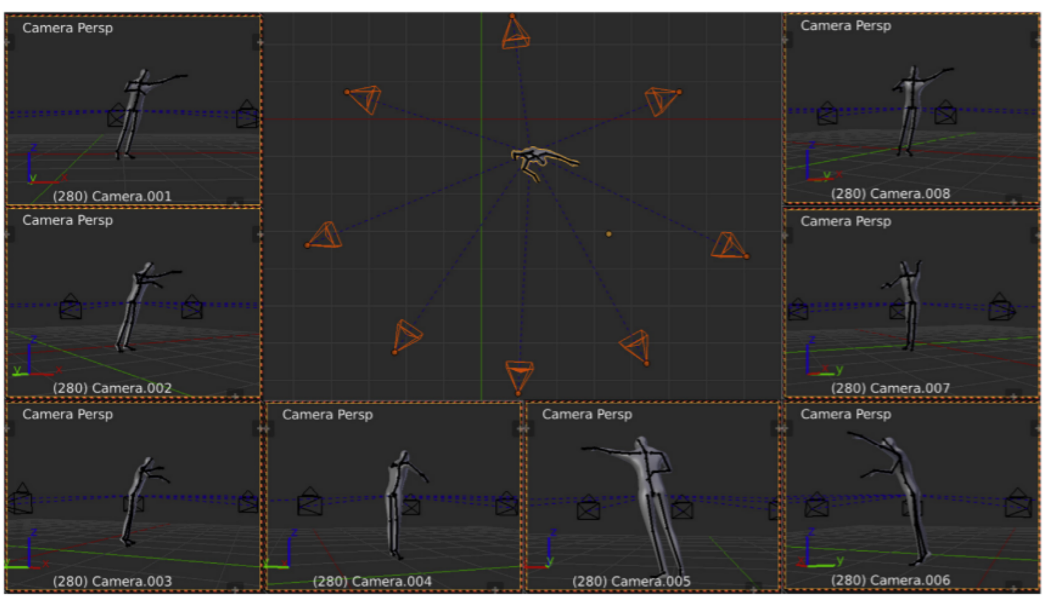


Figure 2 : Synthetic Data Generation

Collecting a labelled training dataset of postures for workers of different anthropometric measures is infeasible as it is impossible to cover all rendering scenarios and simulations in real work environments. Another issue is that manually labeling all joint angles is very expensive, difficult, and error prone. Therefore, the dataset is automatically generated from a computer that uses the MakeHuman software to create virtual human models. These models are animated using retargeting postural information from motion capture (mocap) sequences found from the Carnegie Mellon University mocap database, shown in figure 2. This creates a dataset of 3650 sparse postures. However, since the generated dataset has no noise patterns whereas real-life depth cameras would include noise, BlenSor was used to add artificial noise to the data using a realistic and statistically verified noise model (Gschwandtner et al., 2011). This creates a synthetic dataset of 350K images that was split into 280K images for training and 70K images for validation.

Workflow

To obtain the reference joints angles for training, an inverse kinematics step was performed since the original marker positions were only cartesian coordinates. Each virtual model had virtual markers placed on their body corresponding to the mocap markers. Then, the skeletal model is animated by minimizing the error between the corresponding marker positions in the skeletal model and in the captured data. The model is constrained by different joint angle constraints since each joint can only have a limited range of motion. This is done through solving a weighted least-squares problem using a quadratic programming solver with a convergence criterion of 10^-4 and a limit of 1000 iterations, which is implemented in OpenSim platform (Delp et al., 2007; Seth et al., 2011; Reinbolt et al., 2011). The minimization function is

$$\sum_{i \in m} w_i \, ||x_i^{exp} - x_i(q)||^2,$$

where m is the set of markers, w is a weighting factor, q represents the required coordinates, x and x(q) are the i th marker position in the captured marker trajectory and on the model, respectively. This kinematic modeling of mocap sequences allows obtaining kinematically plausible joint angles using OpenSim. The generated synthetic depth images and corresponding joint angles make up the dataset for training the deep ConvNet regression model.

The ergonomic posture assessment task is solved as a regression problem. The input is a depth image of the posture and the output is the joint angles vector required for computing the RULA score. Thus, given a dataset D of N samples, where each element in the database is a pair consisting of an input depth image and reference vector of 15 joint angles, the network can approximate a function that maps unseen input images of working postures to joint angles. The result is a posture vector that can be used to compute the RULA score. The grand RULA score is computed using the angular thresholds and adjustment parameters defined in the standard RULA worksheet (McAtamney and Corlett, 1993), shown in figure 3 below.

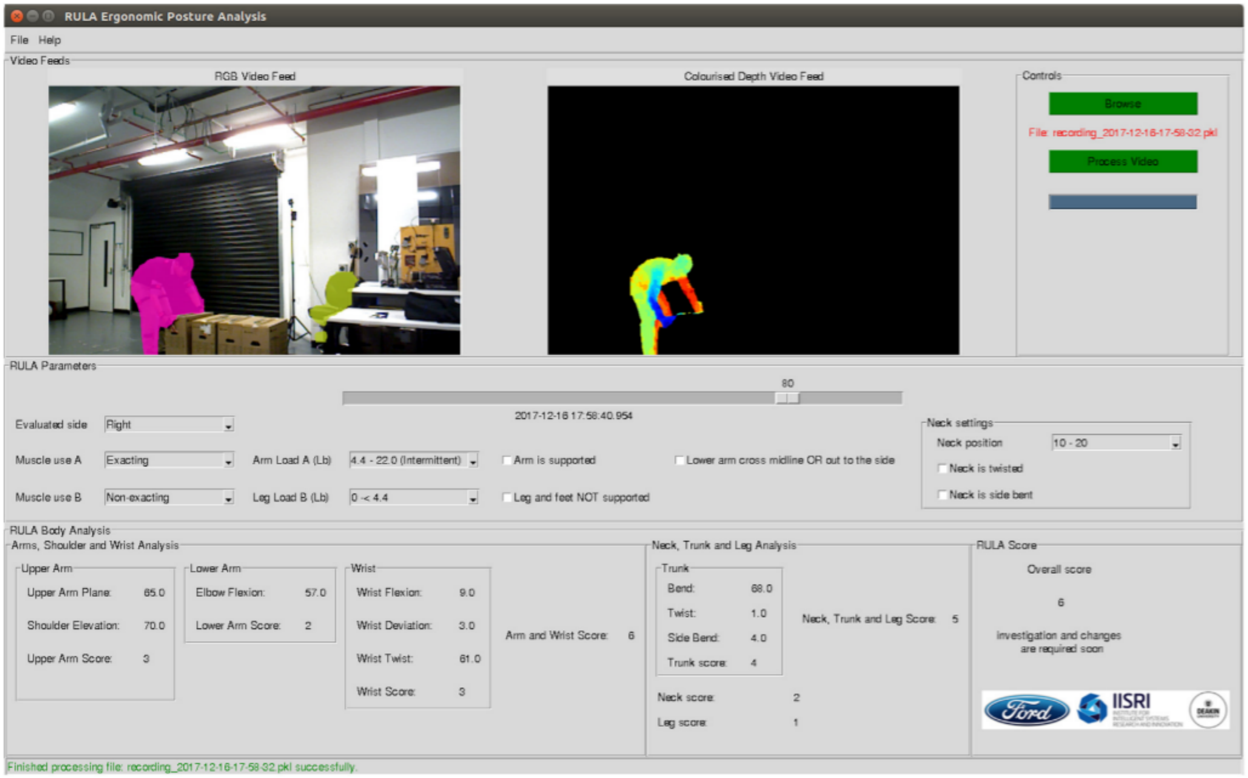


Figure 3 : Obtaining joint angles

Results

This model was applied on a real dataset of 24K postures for 6 subjects of different body shapes while doing a set of manual tasks, as shown in figure 4. The reference angles were generated from recorded mocap sequences in real conditions. The resulting angle calculations for joints are shown in the table directly below.

Table 8
Prediction errors on real data.

Joint name	MAE (deg.)	Scaled MAE (%)	RMSE (deg.)	Scaled RMSE (%)
Trunk rotation	3.23 ± 3.34	1.79 ± 1.86	4.64	2.58
Trunk twist	3.13 ± 3.05	1.74 ± 1.70	4.37	2.43
Trunk bend	2.30 ± 2.09	1.16 ± 1.28	3.10	1.72
L. Elevation	4.13 ± 4.70	1.88 ± 2.14	6.26	2.84
R. Elevation	4.04 ± 4.34	1.83 ± 1.98	5.93	2.70
L. Shoulder	4.19 ± 4.32	2.33 ± 2.40	6.02	3.34
R. Shoulder	4.27 ± 4.65	2.40 ± 2.59	6.31	3.51
L. Elbow	4.14 ± 4.54	3.18 ± 3.49	6.14	4.72
R. Elbow	4.19 ± 4.97	3.22 ± 3.82	6.50	5.00
L. Wrist flexion	2.59 ± 2.46	1.85 ± 1.76	3.58	2.56
R. Wrist flexion	2.76 ± 2.72	1.97 ± 1.94	3.87	2.77
L. Wrist deviation	1.06 ± 1.06	3.03 ± 3.04	1.50	4.29
R. Wrist deviation	1.15 ± 1.22	3.29 ± 3.49	1.68	4.79
L. Wrist twist	3.05 ± 2.75	1.70 ± 1.53	4.11	2.28
R. Wrist twist	3.58 ± 3.21	1.99 ± 1.79	4.81	2.67
Average	3.19 ± 1.57	2.23 ± 1.12	4.27 ± 2.32	2.94 ± 1.64

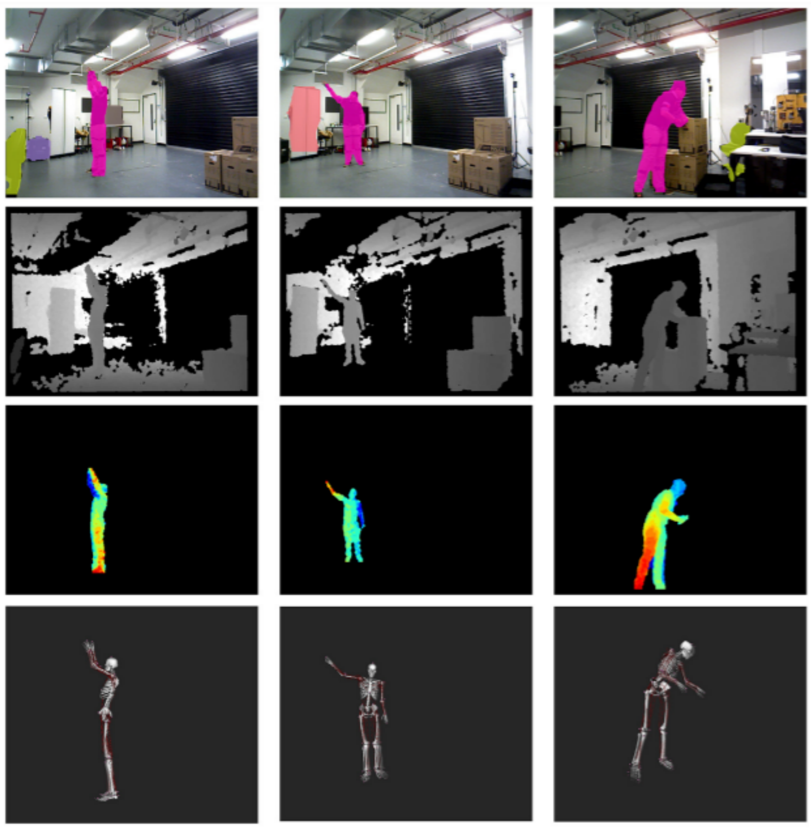
The results demonstrate low prediction error rates for most of the joints and difficulties in estimating elbow and wrist joint configurations. These low error rates are unlikely to change the

RULA scores, as RULA is based on large angular thresholds. The resulting variation between predicted RULA scores and scores using reference joint angles is shown below.

Table 9
The effect joint angle errors on RULA postural scores.

RULA Score	RMSE	Accuracy P_0	kappa (k)
Upper arm Right	0.29	0.92	0.88
Upper arm Left	0.32	0.90	0.86
Lower arm Right	0.22	0.95	0.82
Lower arm Left	0.20	0.96	0.84
Wrist score Right	0.50	0.78	0.67
Wrist score Left	0.50	0.78	0.67
Score A (arm and wrist) Right	0.39	0.86	0.78
Score A (arm and wrist) Left	0.41	0.84	0.76
Score B (neck, trunk and legs)	0.64	0.82	0.63
RULA Grand Score Right	0.49	0.86	0.66
RULA Grand Score Left	0.51	0.85	0.67

The model achieved a joint angle MAE error of 3.19 ± 1.5° and RMSE error of 4.27 ± 2.32° and an average RULA grand score prediction agreement of 89% over both right and left body sides, with a substantial Kappa index level of 0.71.



(a) RULA Grand Score Right: 4
RULA Grand Score Left: 4
Score A Right (arm and wrist): 6
Score A Left (arm and wrist): 5
Score B (neck, trunk and legs): 2
Action: 2, further investigation is needed and changes maybe required

(b) RULA Grand Score Right: 4
RULA Grand Score Left: 2
Score A Right (arm and wrist): 6
Score A Left (arm and wrist): 2
Score B (neck, trunk and legs): 2
Action: 2, further investigation is needed and changes maybe required

(c) RULA Grand Score Right: 4
RULA Grand Score Left: 5
Score A Right (arm and wrist): 2
Score A Left (arm and wrist): 4
Score B (neck, trunk and legs): 5
Action: 3, investigation and changes are required soon

Figure 4 : Model used on real dataset

Conclusion

This paper proposed a semi-automatic ergonomic assessment model using RGB-D cameras and a deep-learning network. It is composed of a segmentation model that detects and segments the person in the scene and a neural network that is trained to estimate body joint angles from a single depth image. The reference joint angles are obtained using a biomechanical model while the prediction model is trained using synthetic depth images. The corresponding RULA score has a prediction accuracy of 89%, which is more accurate and reliable than pre-existing models.

References

Ahmed Abobakr, Darius Nahavandi, Mohammed Hossny, Julie Iskander, Mohammed Attia, Saeid Nahavandi, Marty Smets, RGB-D ergonomic assessment system of adopted working postures, Applied Ergonomics, Volume 80, 2019, Pages 75-88, ISSN 0003-6870, https://doi.org/10.1016/j.apergo.2019.05.004.