



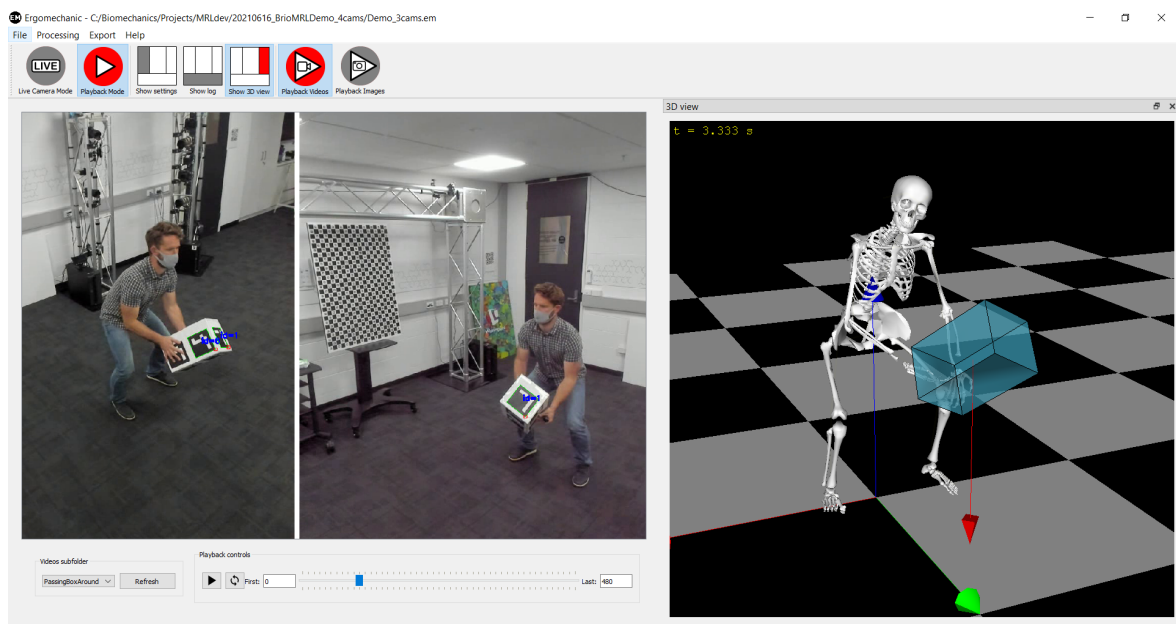
# Mechanical digital twinning of the human body in the workplace for reduced injury risk and improved health

**S. M. Harrison**  and **R. C. Z. Cohen** 

CSIRO Data61, Clayton VIC  
Email: [Simon.Harrison@csiro.au](mailto:Simon.Harrison@csiro.au)

**Abstract:** A digital twin (DT) is a virtual model that acts as a mirror of its physical counterpart that may be solely a replication of certain features of the physical twin, or it may also include the ability to simulate potential future states. The fusion of data from the real world with the virtual model requires the use of sensors and processing of sensor data for inclusion in the digital representation. Simulation capabilities of the DT must be of sufficient detail to accurately forecast the effect of proposed system changes. Usability of the DT for operational tasks depends on the quality of the aforementioned requirements plus the clarity in which the state of the DT is presented to the user.

Here we describe development of a human digital twin (HDT) pipeline for predicting injury risk during workplace activities. The system comprises a set of cameras (typically cheap webcams), a GPU enabled computer, and an in-house developed software package called “Ergomechanic”. Figure 1 shows an example visualisation of the HDT for a participant performing a manual handling task. The software combines markerless motion capture (MMC) with biomechanical modelling to calculate the pose of each worker’s body and the loading upon major body components. Extremes in movement speeds and loading can be predicted and related to injury risk. Metrics defining injury risk can be improved and refined by wider spread use of the technology and correlation with recorded injuries. Example uses of the system to measure workplace activities and body loading are described. Future extensions of the system are discussed, such as viewing outputs in Augmented Reality (AR).



**Figure 1.** Example interface for the Human Digital Twin (HDT) using the Ergomechanic software

**Keywords:** Digital twin, ergonomics, injury, biomechanics, motion capture

## 1. INTRODUCTION

A digital twin (DT) is a virtual model that acts as a mirror of its physical counterpart (Barricelli et al., 2019). This may be solely a replication of certain features of the physical twin, or it may also include the ability to simulate potential future states. Ideally, all physical features should be measured and provided to the DT in real time, but this is not possible in most situations because the process of measurement is too slow. Instead, the resolution (or number) of measurements must be chosen pragmatically, with DT updates limited by measurement bandwidth and latency. These choices may be to only sample a few coarse measurements quickly, to sample many fine measurements slowly, or some mixture of both. Additionally, bandwidth can be used more efficiently when measurement systems are designed to only update changed measurements if suitable (Zipper et al., 2018). The design of the DT and the choices for updating the current state must be referenced back to the intended use of the DT and the needs of that use.

Musculoskeletal injury (MSI) is the most frequent cause for workers compensation claims in Australia (Safe Work Australia, 2019). More than 60,000 claims relating to MSI were made in 2018-19, costing more than \$700M. Injuries to the back and shoulder are most common (Safe Work Australia, 2016). However, it is not clear how to reduce its incidence because the identification of activities that could lead to MSI is not currently possible in most scenarios. Rather than a once-off incident (or an acute injury), most MSIs are an overuse (chronic) injury. They result from an accumulation of undetectable small-scale damage to the muscles, tendons, joints and/or bones, which eventually results in an injury. The time frame over which this damage occurs varies substantially between people, if, in fact, injury occurs at all. As a result, overuse injuries cannot be reliably predicted with current technology.

A human-body digital twin (HDT) can enable improved access to important data, and the ability to assess and optimise future interventions. A HDT can be as simple as a dashboard showing real-time data from a step-counter, GPS device, or heart-rate sensor; and as complex as a 3D representation of the body movements, internal stresses, and/or biochemical states. Traditionally the development of an HDT involved marker-based motion capture in a gait laboratory, which is typically far removed from the real work environment, wearable devices that can be limited in accuracy and can limit movement, or manual estimates of body position (Beyrami et al., 2021). Use of devices with infra-red cameras have been trialled but typically produce less than desirable accuracy measures (Pfister et al., 2014). Recently the fields of camera calibration, computer vision, and biomechanics have been combined to produce markerless motion capture (MMC), which can measure human movement without the need for placing any objects on the body. A HDT for monitoring injury risk in the workplace can be built from the combination of a MMC and biomechanical simulation. This HDT can be used to calculate the number of repetitions of body movements, the speeds and accelerations of movement, and the resulting loading on the internal body structures. Others have trialled a similar framework with deep learning algorithms, but without the custom camera calibration and skeleton modelling presented here, the results were often sub-optimal (Kim et al., 2021).

Here we present a novel prototype system for a HDT that is focused on health and safety in the workplace. The current prototype comprises:

1. Numerous cameras, typically webcams. Here we show examples with 3 Logitech BRIO cameras.
2. A GPU-enabled laptop or desktop computer. A Dell Precision 5540 laptop with Intel® Core™ Processor i9 and Nvidia Quadro T2000 with 4GB GDDR5 is used here.
3. Custom software, called “Ergomechanic” (Cohen and Harrison, 2021; Harrison et al., 2021), which processes incoming data and displays the current state of the HDT.

The software was built in a modular fashion (Cohen and Harrison, 2021) using CSIRO’s Workspace platform (Cleary et al., 2019). A description of use of the prototype is presented by Watkins et al. (2021).

## 2. SYSTEM METHODS AND DEMONSTRATION

The HDT is populated, analysed, and reported on using the following processes:

1. Camera calibration: Prior to use of the system the intrinsic and extrinsic properties of the cameras are calibrated.
2. Capture: When the system is activated video streams from the cameras are buffered in RAM and stored locally.
3. Image processing: A deep learning (DL) algorithm to identify joint positions of any people in view. A second algorithm determines if any objects from the environment are imposing a force onto the body, e.g. are being held by the hands.
4. Reconstruction: The 2D joint data are triangulated into 3D using the camera calibration parameters.

5. Quality assessment and data clean up: Quality of data is assessed for each video stream using the projection error as a quality metric. Smoothing may be applied to noisy data.
6. Model fitting: A generic rigid body musculoskeletal model is scaled to the subject in an initial recording using the 3D joint trajectory data. This scaling includes the lengths of body segments and their inertial properties (masses, centre of masses and moments of inertia).
7. Simulation: The loading on the body joints (and muscles if required) is calculated from gravity, inertial and external forces.
8. Analysis: the injury risk can be assessed by comparing the joint and muscle loading to known limits. Human performance may be assessed using the magnitude of movement and speed of movement.
9. Reporting: pre-identified outputs can be summarised on a dashboard. This could be as simple as an injury risk shown by a traffic light colour (green: low risk, orange: medium risk, red: high risk), real-time graphing of numerical results, and 3D visualisations of movements and loading.

Each of these steps are explained in more detail below.

### **2.1. Camera calibration**

The camera calibration enables the mapping of pixels in each video stream to a point in the 3D world. The intrinsic and extrinsic properties of each camera are calibrated using a ChArUco board (Figure 2a). The size of the squares is known. Every second square contains an ArUco marker (Garrido-Jurado et al., 2014), which can be automatically be identified and is non-symmetric. Therefore edges of each marker can be automatically labelled and are of known size. Dozens of photographs are taken, each with the ChArUco board in different positions in the capture area (Figure 2b, c). The calibration procedure, which leverages functions from OpenCV, finds the best fit of calibration parameters of each camera that maps the position of the ArUco corners in the images to their 3D position. Figure 2d shows a visualisation of a typical set up for four cameras in which rays from each camera are shown as pink lines. The position and orientation of each camera (the extrinsic properties) are indicated by small XYZ axes and the camera name. The origin of the 3D coordinate system is shown as a large set of XYZ axes at the edge of the ChArUco board. The calculated calibration parameters are used later to project body joints visible in the video streams back into their 3D position.

### **2.2. Video capture**

The Ergomechanic user interface has a button that triggers start and end of video recording using the webcams with a prescribed framerate. An initial test is performed to check that the required framerate is achieved. Images from each camera are buffered in RAM and stored locally on the hard disk once the end of recording is triggered. The examples show one able bodied male trial participant who gave informed consent to a procedure approved by the CSIRO ethics board.

### **2.3. Image processing**

Human bodies and objects are identified in each video frame, from each camera, using machine learning approaches. Body pose can be characterised by the position of each major joint (e.g. shoulder, elbow, knee etc). Here we use deep learning libraries such as OpenPose (Cao et al., 2018) or Detectron2 (Wu et al., 2019) to identify the joints in each image. Items of interest such as a box being picked up by the participant can be similarly identified using DL approaches. Here we simply use ArUco markers in known locations on the box to identify its position. Each of the joints and ArUco marker corners are treated as keypoints. Keypoints are a composite data type have a 2D position in each camera (with coordinates  $u, v$ ) and a corresponding 3D position (with coordinates  $x, y, z$ ), as described in the next section.

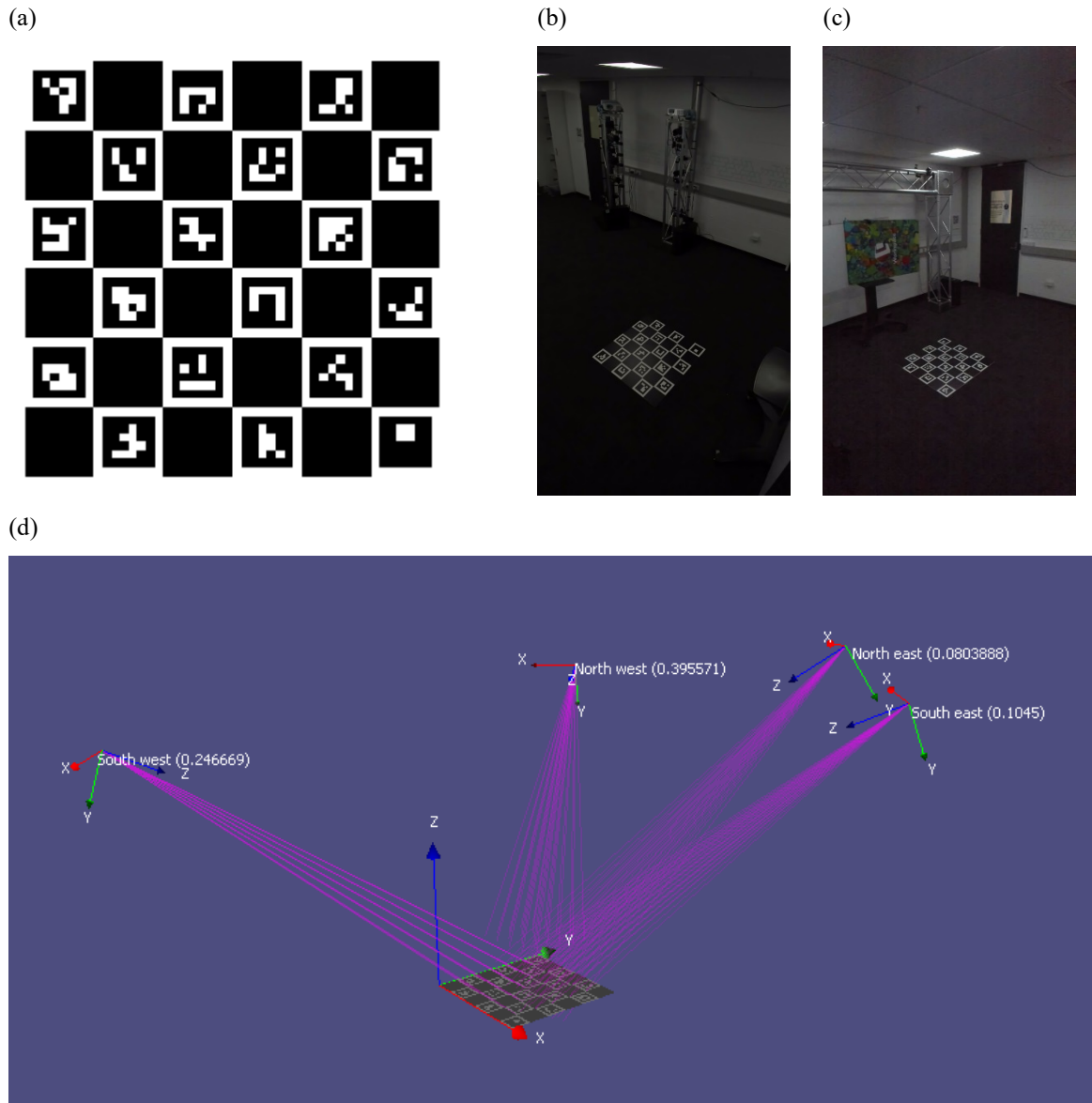
### **2.4. Reconstruction**

Keypoints identified in the Image processing step (Section 2.3) are triangulated into 3D space using the camera calibration parameters (Harrison et al., 2021). This is a fitting exercise that solves for a position in space that minimises the disparity between the original keypoint and the current 3D point transformed back into the camera 2D space.

### **2.5. Quality assessment and data clean up**

Projection error can be calculated by transforming the triangulated point (in  $x, y, z$  coordinates) back into the  $u, v$  coordinates for each camera and calculating the difference in pixel locations. For keypoints that are visible in more than two cameras there is an ability to exclude the data for one or more cameras if their inclusion substantially increases projection error. The image processing algorithms can sometimes give erroneous

outputs such as misidentifying a left leg as a right leg in one video frame. The same approach of checking projection error can be used to rectify left-right identification issues.



**Figure 1.** Calibration procedure: (a) Example ChArUco calibration board, (b) and (c) calibration images with the ChArUco board placed in the view of all cameras, (d) a visualization of the 3D camera calibration showing each of the four cameras and their local coordinate systems, the calibration board, the world coordinate system, and rays (pink lines) from the cameras to the calibration board.

## 2.6. Model fitting

A musculoskeletal model can be fit to the skeleton joint data using an inverse kinematics approach. We use operations in the Bio-Mechanic plugin (Harrison and Cohen, 2019), also implemented in Workspace. Each joint  $i$  in the human structural model has between 0 and 6 degrees of freedom (DoF). The total number of DoF for the whole body is  $N$ . The value of each DoF at any time instance is stored in a single vector,  $\mathbf{q}$ , for the whole body.

These DoF are either rotational or translational, and the direction of the movement is determined by an axis and an origin that are defined in the local frame of the body segment. For instance, the knees and elbows are often treated as a one DoF joint in which the DoF is the flexion angle. The shoulders and hips are treated as three DoF joints that can rotate about all three axes. The root joint can rotate and translate relative to the origin and so this interaction is treated as a six DoF joint.

Harrison and Cohen, Mechanical digital twinning of the human body in the workplace for reduced injury risk and improved health

The position and orientation of the  $i$ th body segment relative to its parent body ( $k$ th body segment) in the  $i$ th coordinate frame is expressed as a 4x4 matrix,  $\mathbf{T}_{i,k(j)}$ , which is dependent only on the degrees of freedom:

$$\mathbf{T}_{i,k(j)} = \begin{bmatrix} \mathbf{R}(\mathbf{q}) & \mathbf{u}(\mathbf{q}) \\ 0 & 1 \end{bmatrix} \quad (1)$$

The position of the  $i$ th body segment in WC (0<sup>th</sup> frame) can be calculated recursively using the relative transforms from each body segment from the root joint to the  $i$ th segment:

$$\mathbf{T}_{i,0(0)} = \prod_{j=1}^i \begin{bmatrix} \mathbf{R}_j & \mathbf{u}_j \\ 0 & 1 \end{bmatrix} \quad (2)$$

For each time increment of the video data, the centre of the joints (specified by the data in  $\mathbf{T}_{i,0(0)}$ ) can be compared to the project 3D data from the MMC process. An optimisation procedure is used to find the set of DoF,  $\mathbf{q}$ , that produce the smallest error in joint centre position. The Dlib library is used currently to perform this optimisation (King, 2009). Typically, each of the DoF are temporally smoothed once the inverse kinematics process is complete.

## 2.7. Simulation

Loading on each body joint is calculated using an inverse dynamics approach (Legnani et al., 1996b, 1996a). The following equation, analogous to Newton's second law of motion (Force = mass  $\times$  acceleration) is used to calculate muscle forces:

$$\mathbf{F}_E + \mathbf{F}_G + \mathbf{F}_M = \mathbf{M}\ddot{\mathbf{q}} + \mathbf{D}(\mathbf{q}, \dot{\mathbf{q}}) \quad (3)$$

Where  $\mathbf{F}_E$  is the vector of external forces and torques,  $\mathbf{F}_G$  is the vector of gravity forces,  $\mathbf{F}_M$  is the vector of muscle forces and torques,  $\mathbf{M}$  is the mass/moment of inertia matrix for the system,  $\ddot{\mathbf{q}}$  is a vector of the accelerations of the degrees of freedom, and  $\mathbf{D}(\mathbf{q}, \dot{\mathbf{q}})$  is a vector of all Coriolis and centripetal forces (which depend on position and speeds only).

All but  $\mathbf{F}_M$  are known if we know the mass of the box and the person.

$$\mathbf{F}_M = \mathbf{M}\ddot{\mathbf{q}} + \mathbf{D}(\mathbf{q}, \dot{\mathbf{q}}) - \mathbf{F}_E - \mathbf{F}_G \quad (4)$$

The masses, centre of masses, and moments of inertia of the body segments, used to populate  $\mathbf{M}$  in Equation (4), are calculated from a published anthropometric regression equation that is based on population averages (Dumas et al., 2007).

## 2.8. Analysis and validation

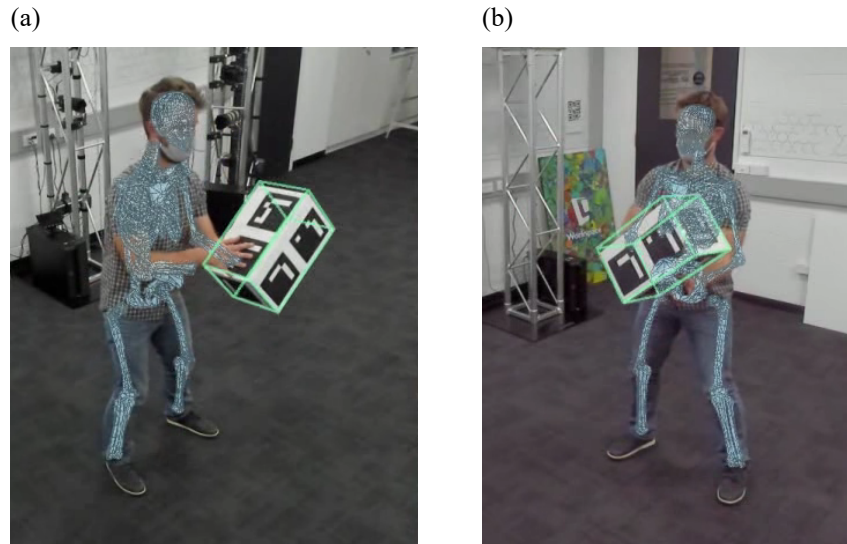
Muscle forces are compared against allowable limits to assess the risk of injury. These limits are not well known as widespread measurements of the type described here have not yet been performed or correlated against injury occurrences. The NIOSH lifting equation (Waters et al., 1994) can be used as a guide to acceptable lower back loading. Many studies with non-invasive, continuous monitoring systems are needed to quantify suitable metrics to use as indicators of injury risk.

Model validation can be performed by projecting the 3D skeleton and environment objects back onto the video images. Figure 3 shows images from two different cameras at the same time increment. Overlaid is the model 3D skeleton (shown in Figure 1) reprojected back into the 2D image using the camera calibration parameters from Section 2.1. Agreement and differences between the physical items and their DTs can be visually identified in a qualitative evaluation procedure. Benchmarking of the system against incumbents such as marker-based systems is currently being performed and will be the subject of future publications.

## 2.9. Reporting

Multiple levels of detail can be used in the reporting of the system's data. First, an instantaneous measure of load on the lower back can be used to guide a traffic-light-type heads up display. Low load is equivalent to a green coloured output. Excessively high load is indicated by a red light. Moderate levels of load are indicated by an amber coloured light. As indicated in Section 2.8, the criteria for these outputs are not yet well established. Until they are precisely established, the criteria must be conservative in order to reduce injury risk. Second, a plot of body loading can be presented to the worker to show the cumulative load on the body. It is likely that many injuries result from a high number of cycles of moderate magnitude forces instead of a few cycles of very high magnitude loads. This type of reporting is less suitable for routine checks during a workday, but may be useful for periodically reviewing longer times, such as at the end of a work week. Third, a 3D

visualisation of body, environment, and objects can be displayed (Figure 1). This can show, interactively, how small or large changes to body position affect the loading through the joints and muscles. This type of visualisation could be used in a training environment to explain the necessity for correcting body posture. It could also be used AR headsets to enable team leaders to quickly identify team members moving in a possibly injurious manner.



**Figure 2.** Reprojection of skeleton and box mesh back onto video stream data. This can be used to evaluate the MMC process and in AR to show key metrics overlaid on participants while they perform movements

### 3. FUTURE DIRECTIONS

As this work progresses there will be enhancements to the digital twinning process in terms of data collection, simulation, analysis, and reporting. Improvements will include:

- **Data collection:** Improvements to data collection will involve the use of cameras with higher frame rates and resolution. This will enable the measurement of faster movements and smaller areas of the body such as the hands and fingers. Second, DL algorithms will be improved to provide quicker results for the higher resolution images and to more types of data from the images. Third, other types of sensors can be incorporated into the pipeline, such as inertial measurement units (IMUs), muscle activity sensors (electromyography), heart rate sensors (as a measurement of work effort, and force measuring devices such as ground-mounted force plates).
- **Finding the best problem-solution fit use cases:** the system is novel and highly suitable for situations where multiple cameras can be used to evaluate human movement. It is not as suitable for scenarios where people move over large distances (and wearable sensors would be more appropriate).
- **Simulation:** The simulation component will be extended to include 3D representations of the bone, muscle, tendon and joint components (Harrison et al., 2014) and their biological changes with time and changed levels of loading. This will enable a more precise determination of injury causes and specific sites of injury progression.
- **Analysis:** As large data volumes are captured using the described method, they can be correlated with the frequency, severity, and type of injury. It is likely that predictions of injury risk will require machine learning algorithms to link the cumulative loading on the body to previously recorded injuries and determine a risk metric from the real-time data.
- **Reporting:** real-time data will be available in AR to enable managers to identify elevated injury risk for their team members. Injury metrics will be improved in accuracy as injuries are reported and correlated with the recorded data.

### 4. CONCLUSIONS

We present a system for developing and updating a mechanical DT of the human body that is intended for use in the workplace. The combination of video camera and computer hardware with software employing deep learning and biomechanical modelling algorithms enables the system to scale a musculoskeletal model to an

Harrison and Cohen, Mechanical digital twinning of the human body in the workplace for reduced injury risk and improved health

individual employee. As they move and interact with the environment their motions are calculated and the forces inside their body are predicted. Body forces are compared to known injury risk metrics. As the system is used more there will be a larger database from which to improve injury risk metrics. The system is capable of displaying the state of a DT using a range of methods, from a simple traffic light system to a detailed 3D visualisation of the skeleton and the distribution of body loads. Future work will concentrate on expanding the accuracy of the system, the types of sensors used, the resolution of the body model and the types of reporting methods.

## REFERENCES

- Barricelli, B.R., Casiraghi, E., Fogli, D., 2019. A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. *IEEE Access* 7, 167653–167671.
- Beyrami, S., Sahlabadi, A.S., Talebolhagh, S., Ramezanifar, S., Sahihazar, Z.M., 2021. Evaluation of Compressive and Shear Forces Exerted on the Lower Back in Manual Load Handling Tasks among Young Workers of Selected Block Maker Using 3DSSPP. *International Journal of Occupational Hygiene* 13, 49–63.
- Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., Sheikh, Y., 2018. OpenPose realtime multi-person 2D pose estimation using Part Affinity Fields, in: *ArXiv Preprint ArXiv1812.08008*.
- Cleary, P.W., Thomas, D., Hetherington, L., Bolger, M., Hilton, J.E., Watkins, D., 2019. Workspace: A workflow platform for supporting development and deployment of modelling and simulation. *Math. Comput. Simul.*
- Cohen, R.C.Z., Harrison, S.M., 2021. Ergomechanic: a markerless motion capture and ergonomic assessment tool developed using the Workspace workflow engine. Presented at the 24th International Congress on Modelling and Simulation (MODSIM2021), Sydney, Australia.
- Dumas, R., Chèze, L., Verriest, J.-P., 2007. Adjustments to McConville et al. and Young et al. body segment inertial parameters. *J. Biomech.* 40, 543–553.
- Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F.J., Marín-Jiménez, M.J., 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognit.* 47, 2280–2292.
- Harrison, S.M., Chris Whitton, R., Kawcak, C.E., Stover, S.M., Pandy, M.G., 2014. Evaluation of a subject-specific finite-element model of the equine metacarpophalangeal joint under physiological load. *J. Biomech.* 47, 65–73.
- Harrison, S.M., Cohen, R.C., 2019. The Bio-Mechanic Workspace plugin: enabling biomechanical measurement and simulation across workflows and software applications. Presented at the 23rd International Congress on Modelling and Simulation, Canberra ACT Australia, pp. 442–448.
- Harrison, S.M., Cohen, R.C.Z., Cleary, P.W., 2021. Whole body movement modelling in realistic environments for understanding performance and injury, in: *Digital Human Modelling and The Digital Twin Patient*. Elsevier.
- Kim, W., Sung, J., Saakes, D., Huang, C., Xiong, S., 2021. Ergonomic postural assessment using a new open-source human pose estimation technology (OpenPose). *International Journal of Industrial Ergonomics* 84, 103164.
- King, D.E., 2009. Dlib-ml: A machine learning toolkit. *J. Mach. Learn. Res.* 10, 1755–1758.
- Legnani, G., Casalo, F., Righettini, P., Zappa, B., 1996a. A homogeneous matrix approach to 3D kinematics and dynamics—II. Applications to chains of rigid bodies and serial manipulators. *Mech. Mach. Theory* 31, 589–605.
- Legnani, G., Casolo, F., Righettini, P., Zappa, B., 1996b. A homogeneous matrix approach to 3D kinematics and dynamics—I. Theory. *Mech. Mach. Theory* 31, 573–587.
- Pfister, A., West, A.M., Bronner, S., Noah, J.A., 2014. Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis. *Journal of Medical Engineering & Technology* 38, 274–280.
- Safe Work Australia, 2019. *Australian Workers Compensation Statistics 2018-19*.
- Safe Work Australia, 2016. *Statistics on work-related musculoskeletal disorders*.
- Waters, T.R., Putz-Anderson, V., Garg, A., 1994. *Applications manual for the revised NIOSH lifting equation*.
- Watkins, D., Bolger, M., Hetherington, L., Zhang, Z., Xie, P., Harrison, S.M., Cohen, R.C.Z., 2021. Using Workspace to Implement Digital Twins in the Mixed Reality Lab. Presented at the 24th International Congress on Modelling and Simulation (MODSIM2021), Sydney, Australia.
- Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., Girshick, R., 2019. Detectron2.
- Zipper, H., Auris, F., Strahilov, A., Paul, M., 2018. Keeping the digital twin up-to-date — Process monitoring to identify changes in a plant, in: *2018 IEEE International Conference on Industrial Technology (ICIT)*. Presented at the 2018 IEEE International Conference on Industrial Technology (ICIT), pp. 1592–1597.