Upper Body Pose Estimation Utilizing Kinematic Constraints from Physical Human-Robot Interaction

Richardo Khonasty, Marc G. Carmichael, Dikai Liu and Kenneth J. Waldron
University of Technology Sydney, Australia
Richardo.Khonasty@student.uts.edu.au

Abstract

In physical Human-Robot Interaction (pHRI), knowing the pose of the operator is beneficial and may allow the robot to better accommodate the human operator. Due to a large redundancy in the human body, determining the pose of the human operator is difficult to achieve in unstructured environments especially in human-robot collaborative operations where the robot often occludes the human from vision-based sensors. This work presents an upper body pose estimation method based on exploiting known positions of the human operator's hands while performing a task with the robot. Upper body pose is estimated using upper limb kinematic models alongside sensor information and model approximations to produce solutions that are biomechanically feasible. The pose estimation method was compared to upper body poses obtained using a motion capture system. It was shown to be able to perform robustly with varying amounts of available information. This approach is well suited in applications where robots are controlled using well-defined interfaces such as handlebars, operating in unstructured environments.

1 Introduction

The field of robotics and intelligent machines is undergoing a paradigm shift that sees more systems interacting physically with human operators. Combining the strength of the robot with the adaptability of the human operator allows for completion of tasks that would be difficult for each to handle alone. One task that benefits from human-robot collaboration is in abrasive blasting where the robot is able to provide assistance to augment the human operator's strength during this particularly laborious activity.

For physical Human-Robot Interaction (pHRI), it is advantageous for the robot to know as much information about the operator as possible. Important information is the pose of the human operator so that they can be accommodated by the robot. One benefit can be seen in collision avoidance algorithms, by accounting for a safe margin of the robot manipulable area around the human operator [Wang *et al.*, 2013; Liu and Narayanan, 2013]. Knowing the pose of the operator may also allow models of the human body to be used to estimate fatigue [Silva *et al.*, 2011; Karg *et al.*, 2014] and the operator's capability to contribute towards tasks, allowing the robot to adapt its assistance accordingly [Carmichael and Liu, 2013], [Carmichael and Liu, 2011].

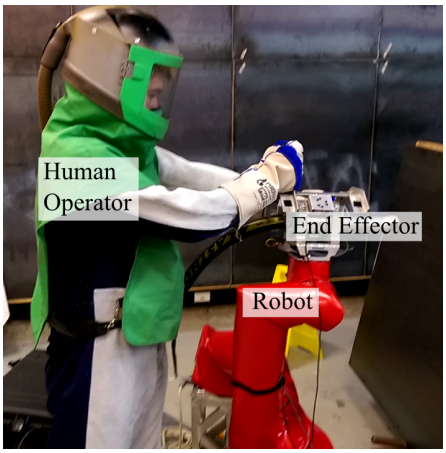


Figure 1: An example of a collaborative task where the operator controls the robot using specified handle positions

The problem of identifying the pose of a human has long been investigated [Schwarz *et al.*, 2012], [Menychtas *et al.*, 2016]. Due to the large number of degree of freedom (DOF) that the human body inherently possess, it is difficult to precisely determine the pose that the human upper limb would take given a specific hand position. Humans are also naturally varied, resulting in individuals performing the same task in a different manner. Due to the close proximity between the operator

and the robot, the robot may occlude some information from sensors such as a camera.

Robots designed for physical interaction often contain a specially designed point of interaction. A common example is handles for the operator to grip during a collaborative task. The design of the interaction point is often regulated through standards [Standards Australia, 2016]. For example a robot may require enabling switches to be pressed with both hands before it will operate, such as the robot in Figure 1. This mode of interaction constrains the operator's hands during collaboration between the human and the robot. This kinematic constraint provides knowledge of the operator's hand positions which can be utilized in the estimation of the full upper body pose of the human operator.

In this work we present a method utilizing biomechanical upper limb models of the left and right arms to estimate the pose of a human operator's upper body as they interact with a robot. The upper body refers to both the upper limbs and the torso of the user. The presented method exploits information about the kinematic constraint that occurs during interaction with the robot, namely the known position and orientation of the operator's hands. When available, sensory information and model approximations are combined to assist in the redundancy resolution in the operator's upper body. The use of biomechanical models to represent the upper limbs limit solutions to biomechanically feasible limb configurations. Experiments compared the results from the proposed method against a ground truth obtained using a commercial motion capture system.

2 Existing Human Tracking Methods

An example of an existing human tracking method is the Motion Capture (MoCap). It is a form of marker-based tracking, which measures human body pose by tracking the positions of markers or sensors located on the bodies of human subjects [Moeslund *et al.*, 2006]. These systems utilize multiple cameras fixed in the environment to triangulate and track the 3D positions of markers worn on the body. Its use is often limited to clinical and research applications, as systems are typically expensive and require well structured indoor environments. Inertial sensing MoCap systems require subjects to wear sensors, usually consisting of 3-axis accelerometers, gyroscopes and magnetometers, to estimate the motion of the user's body segments without the use of cameras. MoCap provides fast and accurate tracking of the human body and is often used as a baseline for measuring the human body pose [Shi *et al.*, 2014]. However due to its inherent restrictions as well as the need for subjects to wear markers/sensors it is unsuitable in practical or industrial applications.

Methods for classifying human subjects using RGB and depth images have recently become popular due to their capability to be used in a wide variety of fields with sensors such as the Kinect camera. These methods attempt to classify a skeleton based on the sensor data [Schwarz *et al.*, 2012][Lo Presti and La Cascia, 2016]. Similar to motion capture systems, skeleton tracking methods work best when the object being tracked is able to be viewed clearly. Occlusions affect the estimate greatly which makes it difficult to realize these systems in dynamic environments [Moeslund *et al.*, 2006]. Such occlusions can be common in pHRI due to the operator's close proximity with the robot.

Model-based approaches have been developed to predict the pose of the human body. These models estimate the pose that the human body would assume based on various concepts. The model presented in [Kim *et al.*, 2012] resolves the upper limb redundancy by maximizing the manipulability of the hand for motion towards the head. Others attempt to determine the natural arm configuration through minimization of the total work done during specific movements [Kang *et al.*, 2005]. These models often assume a free unconstrained case where the hands are free to be manipulated and with the body at a known position and orientation.

In this paper the interaction of the human and robot is exploited, namely through prescribed interaction points on the robot. The interaction points are in the form of handles where the operator's hands should be fixed relative to the robot at all times during the operation. Through these handles, information regarding the position and orientation of the hand is known. Even with the poses of the left and right hand known, the large redundancy in the human body allows for a wide range of upper limb poses to be adopted. Factors such as the position and orientation of the torso will change depending on the person. Additional sensor information and model approximations are integrated to resolve the missing information required for an upper body pose estimate.

3 Upper Body Kinematic Model

The human upper-body is often modeled as a pair of 7-DOF upper limbs connected to a rigid torso, as shown in Figure 2. Rotation of the upper arm is represented using a 3-DOF spherical joint coinciding with the glenohumeral joint of the shoulder. Elbow motion is represented using a 1-DOF hinge joint connecting the upper arm and the forearm. Forearm and wrist motions were represented by another 3-DOF spherical joint coinciding with center of the wrist.

Human shoulder range of motion (ROM) is complex and measurements of isolated articulations are not adequate to provide a general representation of the complex shoulder ROM. The shoulder is often simplified as a spherical joint capable of only rotational movement

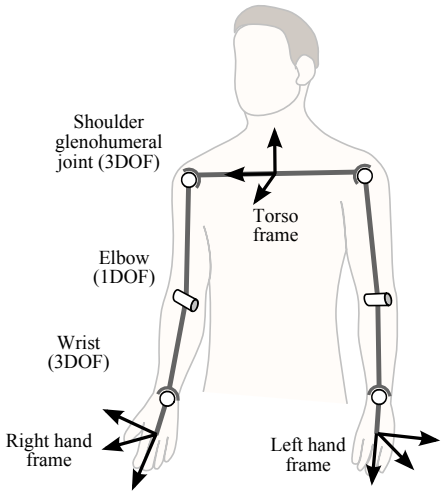


Figure 2: Kinematics of the upper body.

about the center of the glenohumeral joint. In reality the shoulder is capable of translation due to sternoclavicular articulations. To account for the translation of the shoulder, the Klopčar-Lenarčič model [Klopčar and Lenarčič, 2005] was used. In this model, the motion of the sternoclavicular articulations are represented as two rotational and one translational joint.

Kinematics of the upper body utilized in this work are modeled as a left and right upper limb as shown in Figure 3. Each limb is modeled using nine joints. Attached to each limb is a rigid left or right torso half which are initially separated. Both left and right shoulder and elbow kinematics are based on the upper limb model presented in [Klopčar and Lenarčič, 2005; Lenarčič and Umek, 1994]. Added to both left and right upper limb models are three additional joints which represents forearm pronation, wrist deviation, and wrist extension. In total the kinematic model used to represent the upper body has 18 joints.

Biomechanically feasible configurations of the human body are limited by the ROM of each joint. Anthropometric measurements of human joint ROM are widely available in the literature, however measurements are typically with respect to articulation about a single axis or within a plane of motion (e.g. abduction or flexion). In this work the equations presented in [Klopčar and Lenarčič, 2005] are used to represent the human upper limb ROM. This work presented a series of coupled equations to define the joint limits in the upper limb and achieve a realistic representation of the complex shoulder ROM. This model was used previously to generate a representation of biomechanically feasible human shoulder ROM for the purpose of optimizing the design of robotic exoskeletons [Carmichael *et al.*, 2015].

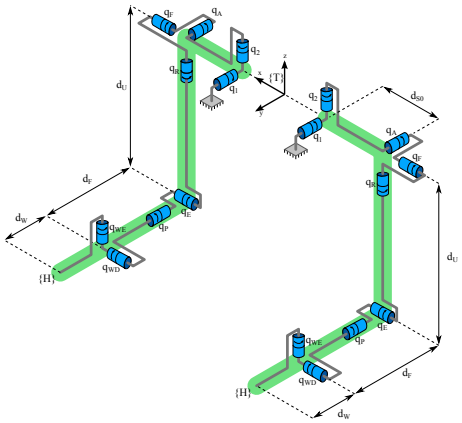


Figure 3: Upper-body kinematic model

4 Upper Body Pose Estimation Method

The method initializes by prescribing the two arm models to the respective hand position and orientation. The pose of the left hand in the global frame is denoted by ${}^{G}\mathbf{T}_{LH} \in SE3$ with ${}^{G}\mathbf{R}_{LH} \in SO3$ defining the orientation and ${}^{G}\mathbf{p}_{LH} \in \mathbb{R}^{3\times 1}$ position. Likewise the right hand in the global frame is denoted by ${}^{G}\mathbf{T}_{RH} \in SE3$ with ${}^{G}\mathbf{R}_{RH} \in SO3$ and ${}^{G}\mathbf{p}_{RH} \in \mathbb{R}^{3\times 1}$. This method works under the assumption that during the human-robot interaction the position and orientation of the operator's hands are the same as the handles on the robot. Therefore from the robot's forward kinematics the pose of the operator's hands are always known. The upper body pose estimation is conducted in two modes as shown in Figure 4. The mode that is used depends on whether the upper body kinematic chain from left hand to right hand is closed to within a set threshold. If the kinematic chain is closed (Mode 1) the upper limb pose estimates are resolved using available sensor information or model approximation. If the kinematic chain is open (Mode 2) the two torso halves are brought together to close the chain. Both modes use the same joint step calculation algorithm to update joint values for the left and right arm models.

4.1 Joint Step Calculation

A variation of the velocity level saturation in the null space method [Flacco *et al.*, 2015] was used to calculate each joint step for the arm models. This method accommodates the integration of a task priority structure to incorporate various sensor information or model approximations. In Algorithm 1, sensor information or model approximations are implemented as tasks. In this structure, the tasks of higher priority precede those of lower priority. For each task vector $\dot{x}_k$, an accompanying Jacobian J_k is required to calculate the new joint step (Equation 1). The same algorithm is used in both

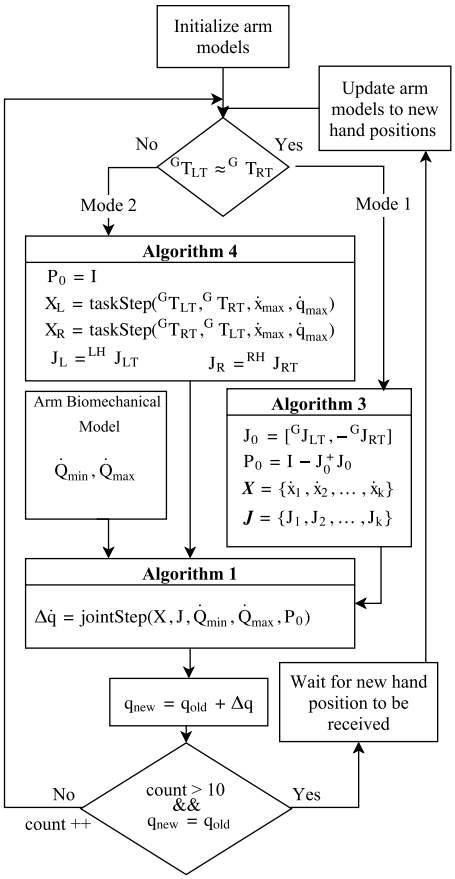


Figure 4: Flowchart of the proposed pose estimation method

modes, with the only variation being the task and Jacobian structure being used as the input.

$$X = \{\dot{x}_1, \dot{x}_2, \ldots, \dot{x}_k\} \quad (1)$$
$$J = \{J_1, J_2, \ldots, J_k\}$$

In Algorithm 1, joint steps Δq for the arm model are calculated. The symbol $^{+}$ denotes a Moore-Penrose pseudo-inversion of the matrix. The limits to each joint are represented by the vectors, $\dot{Q}_{min}, \dot{Q}_{max}$. For each task $\dot{x}_i$ for $i = 1 \rightarrow k$, a contribution to the joint step $\dot{q}_k$ is calculated. This calculation takes into consideration the previous task through the projection matrix P_{k-1}. This prevents the lower priority task from interfering with the previously calculated task. If any joint is found to be outside of the limits the *mostCriticalJoint* function is called. This function determines how much each joint exceeded the limits. The joint that exceeds its limit by the largest margin is saturated by setting it to the current limit. The value of the joint availability matrix W corresponding to the most critical joint is set to 0 and a new $\dot{q}_k$ is calculated. The algorithm is repeated until the calculated $\dot{q}_k$ satisfies all joint limits or until all joints are saturated. If all joints are saturated, the algorithm calculates the closest approximation for the new joint commands. To maintain the closed kinematic chain in Mode 1 an initial projection matrix P_0 is set to the self motion projection. In Mode 2 this projection is set to identity.

Algorithm 1 Joint step calculation with task priority

Function $\Delta q = \textbf{jointStep}(X, J, \dot{Q}_{min}, \dot{Q}_{max}, P_0)$
for k = 1 to m **do**
  $W_k = I, \dot{q}_{N,k} = 0, \bar{P}_k = P_{k-1}$
  while solution_found == false **do**
    $\tilde{P}_k = (I - (J_k\bar{P}_k))^{+}((I - W_k)P_{k-1})^{+}$
    $\dot{q}_k = \dot{q}_{k-1} + (J_k\bar{P}_k)^{+}(\dot{x}_k - J_k\dot{q}_{k-1}) + \tilde{P}_k\dot{q}_{N,k}$
    if $any(\dot{q}_k < \dot{Q}_{min} \| \dot{q}_k < \dot{Q}_{max})$ **then**
      $j = mostCriticalJoint(\dot{q}_k, \dot{Q}_{min}, \dot{Q}_{max})$
      $W_k(j,j) = 0$
      $\dot{q}_{N,k}(j) = \begin{cases} \dot{Q}_{min} & \dot{q}_k(j) < \dot{Q}_{min} \\ \dot{Q}_{max} & \dot{q}_k(j) > \dot{Q}_{max} \end{cases}$
      $\bar{P}_k = (I - ((I - W_k)P_{k-1}^{+})P_{k-1}$
      if $W_k == zero$ **then**
        $\tilde{P}_k = (I - (J_k\bar{P}_k)^{+}((I - W_k)P_{k-1})^{+}$
        $\dot{q}_k = \dot{q}_{k-1} + (J_k\bar{P}_k)^{+}(\dot{x}_k - J_k\dot{q}_{k-1}) + \tilde{P}_k\dot{q}_{N,k}$
        solution_found = true
      end if
    else
      solution_found = true
    end if
  end while
  $P_k = P_{k-1} - (J_kP_{k-1})^{+}(J_kP_{k-1})$
end for
$\Delta q = \dot{q}_k$
End Function

4.2 Mode 1 - Kinematic Chain Closed

In Mode 1 sensor information and/or model approximations are utilized to provide a better estimate for the pose of the upper body. These additional information are often represented as Cartesian space tasks to move certain parts of the upper limb to a certain positions or orientations. In this mode each tasks are implemented while keeping the left and right torso halves together through the self motion projection P_0. Examples of these tasks include moving the torso to a specified position or maintaining an upright body position.

The *taskStep* function (Algorigthm 2) is used to calculate a required Cartesian step to bring the current pose (T_C) towards the desired pose (T_D). This function calculates a displacement vector which is used in Mode 1 and 2. In the task step function P_C and P_D are the position vectors of T_C and T_D respectively. R_C and R_D are the

Algorithm 2 Task step calculation

Function $X = \text{taskStep}({}^{G}\mathbf{T}_C, {}^{G}\mathbf{T}_D, \dot{x}_{max}, \dot{q}_{max})$
$\dot{x} = P_D - P_C$
if $\|\dot{x}\| > \dot{x}_{max}$ **then**
  $\dot{x} = \dot{x}_{max} \times \frac{\dot{x}}{\|\dot{x}\|}$
end if
$\dot{R} = R'_D \times R_C$
Let θ = magnitude of the axis angle representation of $\dot{R}$
and e = the rotation vector
$w = e \cdot \theta$
if $\|w\| > \dot{q}_{max}$ **then**
  $w = \dot{q}_{max} \times \frac{w}{\|w\|}$
end if
$X = \begin{bmatrix} \dot{x} \\ w \end{bmatrix}$

rotation matrices of T_C and T_D respectively. Using this function both the required position and angular step, $\dot{x}$ and w are calculated within the distance and angular constraint, $\dot{x}_{max}, \dot{q}_{max}$.

$$\begin{aligned} {}^{G}\dot{x}_{LT} &= {}^{G}J_{LT}\dot{q}_L \\ {}^{G}\dot{x}_{RT} &= {}^{G}J_{RT}\dot{q}_R \end{aligned} \tag{2}$$

In Equation 2, $\dot{x}_{LT}$ and $\dot{x}_{RT}$ represent the velocities of the torso halves. ${}^{G}J_{LT}$ and ${}^{G}J_{RT}$ respectively represent the Jacobian of the left and right torso in the global frame with $\dot{q}_L$ and $\dot{q}_R$ being the current joint values for the left and right arm. For the torso to maintain a closed kinematic chain ${}^{G}\dot{x}_{LT}$ must be equal to ${}^{G}\dot{x}_{RT}$ (Equation 3).

$$\begin{aligned} {}^{G}J_{LT}\dot{q}_L &= {}^{G}J_{RT}\dot{q}_R \\ \begin{bmatrix} {}^{G}J_{LT}, & -{}^{G}J_{RT} \end{bmatrix} \begin{bmatrix} \dot{q}_L \\ \dot{q}_R \end{bmatrix} &= 0 \\ J_0\dot{q} &= 0 \end{aligned} \tag{3}$$

J_0 is a Jacobian whose null-space spans the self-motions of the closed kinematic chain. To accommodate concurrent tasks, the projection P_0 is used such that the torso movement on one side is able to be replicated by the other (Equation 4).

$$P_0 = I - J_0^{+} J_0 \tag{4}$$

Algorithm 3 shows the task array $X\{k\}$ alongside its corresponding Jacobian array $J\{k\}$ consisting of $J_1, J_2, \ldots, J_k$ to be used in Algorithm 1 where:

Algorithm 3 Redundancy Resolution With Kinematic Chain Closed

$J_0 = [J_{LT}, -J_{RT}]$
$P_0 = I - J_0^{\#} J_0$
$X = \{\dot{x}_1, \dot{x}_2, \ldots, \dot{x}_k\}$
$J = \{J_1, J_2, \ldots, J_k\}$
$\Delta q = \text{jointStep}(X, J, \dot{Q}_{min}, \dot{Q}_{max}, P_0)$
$q = q + \Delta q$

$$\dot{x}_k = \underbrace{\begin{bmatrix} J_{k,L} & 0 \\ 0 & J_{k,R} \end{bmatrix}}_{J_k} \begin{bmatrix} q_L \\ q_R \end{bmatrix} \tag{5}$$

In Equation 5 $J_{k,L}$ and $J_{k,R}$ correspond to the Jacobian of the left and right arm model for a certain task $\dot{x}_k$.

Due to the large number of DOFs in a human body, there are a large number of feasible upper body poses for a given left and right hand pose. To obtain an estimate of the upper body pose, additional information is required for the resolution of the redundancy in the upper body model. In this work, the additional information is split into two categories, sensor information or model approximations. The various additional information are taken into consideration in Algorithm 3.

Depending on the setup of the robot or sensors, it may be possible to obtain information about the upper body that aids in the redundancy resolution. For example a camera placed behind the user could estimate the position of the user's head [Ohayon and Rivlin, 2006] or a laser scanner placed at leg or torso height could be used to determine the user's 2-D position. The additional information obtained from sensors could be added into the pose estimation model as additional tasks to obtain a better estimate of the upper body. However, it may not always be feasible to rely on sensor data. Human operators may be outside of the range of the sensor or information could be lost. When there is no sensor information available, model approximations could be utilized as a substitute.

In common collaborative tasks, it may be appropriate to assume the operator is standing in a relatively upright position, adding a 2-DOF constraint via the torso's roll and pitch when sensory data of the torso is not available. We may also approximate the position of the elbow through the use of kinematic models [Kim *et al.*, 2012]. Other approximations include the height of the user, which could be measured before operation is started and could be implemented as a another constraint to the position of the torso.

4.3 Mode 2 - Kinematic Chain Open

In the event that the two torso halves becomes separated, the pose estimation switches into Mode 2. The two torso

halves are often found to be separated when new hand positions are prescribed or when the arm models are in a singular configuration. While the two torso halves are separated, the kinematic chain of the full upper body model is broken. In this mode the task structure $\boldsymbol{X}$ and the Jacobian structure $\boldsymbol{J}$ are prepared such that the two torso halves are brought together to close the kinematic chain. Starting with the current configuration of the left and right arms, $\mathbf{q}_L$ and $\mathbf{q}_R$, the poses of the two torso halves are calculated, ${}^{LH}\mathbf{T}_{LT}$ and ${}^{RH}\mathbf{T}_{RT}$. Using ${}^{LH}\mathbf{T}_{LT}$ and ${}^{RH}\mathbf{T}_{RT}$, the pose of the torso in the global frame ${}^{G}\mathbf{T}_{LT}$ and ${}^{G}\mathbf{T}_{RT}$ can be calculated using Equation 6.

$$
\begin{aligned}
{}^{G}\mathbf{T}_{LT} &= {}^{G}\mathbf{T}_{LH} \cdot {}^{LH}\mathbf{T}_{LT} \qquad (6)\\
{}^{G}\mathbf{T}_{RT} &= {}^{G}\mathbf{T}_{RH} \cdot {}^{RH}\mathbf{T}_{RT}
\end{aligned}
$$

Where:

$$
{}^{LH}\mathbf{T}_{LT} = \begin{bmatrix} ({}^{LT}\mathbf{R}_{LH})^T & -({}^{LT}\mathbf{R}_{LH})^T \cdot {}^{LT}\mathbf{p}_{LH}) \\ 0 \quad 0 \quad 0 & 1 \end{bmatrix}
$$

$$
{}^{RH}\mathbf{T}_{RT} = \begin{bmatrix} ({}^{RT}\mathbf{R}_{RH})^T & -({}^{RT}\mathbf{R}_{RH})^T \cdot {}^{RT}\mathbf{p}_{RH}) \\ 0 \quad 0 \quad 0 & 1 \end{bmatrix}
$$

A displacement vector X is calculated using the taskStep function (Algorithm 2). Utilizing this function a small step is calculated to move the left torso half towards the right torso half, X_L, and the right torso half towards the left, X_R. A Jacobian that moves the torso frame relative to its respective hand ${}^{LH}J_{LT}$ or ${}^{RH}J_{RT}$ alongside a vector of the minimum and maximum joint limits $\dot{Q}_{min}\&\dot{Q}_{max}$, and the displacement vector, X_L or X_R is then used as an input to the jointStep function (Algorithm 1). Once the two torso frames are within a desirable threshold, the kinematic chain is considered closed and the method switches to Mode 1 to integrate sensor information and model approximations.

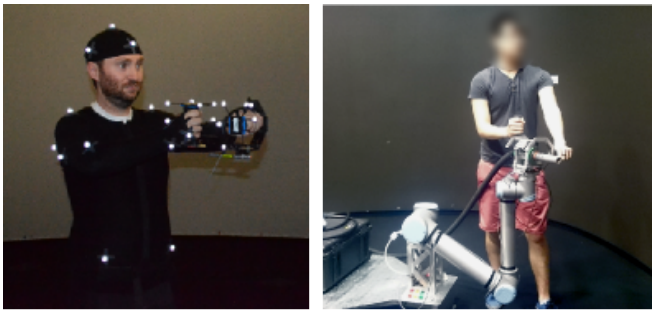
Algorithm 4 Moving the left torso half towards the right to close the kinematic chain

$P_0 = I$
X_l= taskStep$({}^{G}\mathbf{T}_{LT}, {}^{G}\mathbf{T}_{RT}, \dot{x}_{max}, \dot{q}_{max})$
$J_l = {}^{LH}\ J_{LT}$
Δq_l = jointStep$(X_l, J_l, \dot{Q}_{l,min}, \dot{Q}_{l,max}, P_0)$
$q_l = q_l + \Delta q_l$

At time $t = 0$ where no previous information is known about the user, each arm model needs to be initialized. Several joint value conditions were tested to determine the capability and robustness of the arm models. Three conditions were tested which were to set the joint values to be random, zero or to the middle of the joint range. The arm model had difficulty closing the kinematic chain within an iteration constraint when the joint values were randomly generated with it having 74.93% of success. Having the model initialize to zero provided the best result at 97.43% for the data collected for the investigation. Initializing the model to the middle of the range had a success rate of 90.97%. Thus for this work the model is initialized with the joints set to zero.

5 Experimental Investigation

To analyze the performance the pose estimation method presented in this paper, several recordings were conducted using a motion capture system. The data was collected from 3 subjects wearing the motion capture suit for marker based tracking to obtain positions and orientations of key body segments. The suit worn utilized reflective markers (Figure 6), recorded using the Motive software and 12 Optitrack cameras running at 120Hz. To simulate a collaborative operation with a robot, two motion-tracked handles were held by the subjects during the recording.



(a) (b)

Figure 6: a) The motion capture suit and end effector used in Motion 3. b) the collaborative task being simulated in Motion 3

Three sets of motion were recorded for each individual, which are:

- Motion 1: Subject moving to match set poses
- Motion 2: Subject may freely move
- Motion 3: Simulated human robot collaboration

Motion 1 involved the subject performing a set of poses one after the other. Each pose did not involve much torso movement with a higher emphasis on arm movement (Figure 5). In Motion 2, each subject were asked to perform random motions. This involved movement of the whole body and the subject was not restricted to a stationary torso. Motion 3 was chosen to simulate a collaborative task where the two handles were connected as

Figure 5: Poses set for Motion 1

shown in Figure 6. Each subject was asked to trace the outline of a box placed 2 meters in front of the subject using a laser pointer attached to the handles. The handle assembly restricted the relative hand orientations.

From the data collected, several scenarios were used to simulate estimating upper body pose using various combinations of available information. The sets of information that are included in this investigation are the measured torso position and orientation, measured elbow positions, elbow position approximated by a model, and setting the torso height and keeping the torso level. These scenarios are shown with the task priority in Table 1. The pose estimation for each scenario is compared to the actual torso position and orientation obtained using the motion capture system.

Table 1: List of scenarios used in Mode 1. Numbers represent the task priority

	Sensing			Model		
Scenario	T.Pos	T.Ori	E.Pos	E.Mdl	T.Hgt	T.Lvl
1	1	1	2	-	-	-
2	1	-	2	-	-	-
3	1	-	-	2	-	-
4	1	-	-	-	-	-
5	-	-	2	-	1	1
6	-	-	-	2	1	1
7	-	-	-	-	1	1

Scenario 1 represents a system where all information required to define the upper limb pose is available to be used. In Scenarios 2 to 4 the torso position was used as the highest priority task. The second priority task in Scenario 2 is the position of the elbow. In Scenarios 3 and 6 the position of the elbow is approximated using the method presented in [Kim *et al.*, 2012]. In Scenarios 5 to 7 there is very little information available to be used. In these scenarios the highest priority task is to keep the torso at a certain height based on the subject's height whilst keeping the torso level. The torso is kept level by specifying the roll and pitch of the torso.

6 Results & Discussion

The error in the estimated torso position and orientation for each scenario is calculated by comparing results of the motion capture value recorded. The position error reflects the distance between the estimated torso frame to the recorded torso frame. The orientation error is calculated by using an angle axis representation of the difference between the estimated and the recorded torso frame.

Figure 7 shows that from the recorded data, the average error in the torso position and orientation is generally lower in Scenario 1. The maximum average position error calculated was below 6cm (Subject 2 performing Motion 2). For most of the cases in this scenario, the pose estimation method was able to perform similarly to the motion capture recording. Some differences between the motion capture recording and the estimated upper body pose is attributed to differences in model kinematics between the two. A higher standard error was found for Subject 1 performing Motion 1 and Subject 2 performing Motion 2.

In Scenarios 2 to 4, torso orientation measurements are not utilized in the estimation. As expected, the error in the torso orientation increases when compared to the results obtained in Scenario 1. For these scenarios it was thought that prescribing the elbow position would limit the orientation that the torso could achieve. From

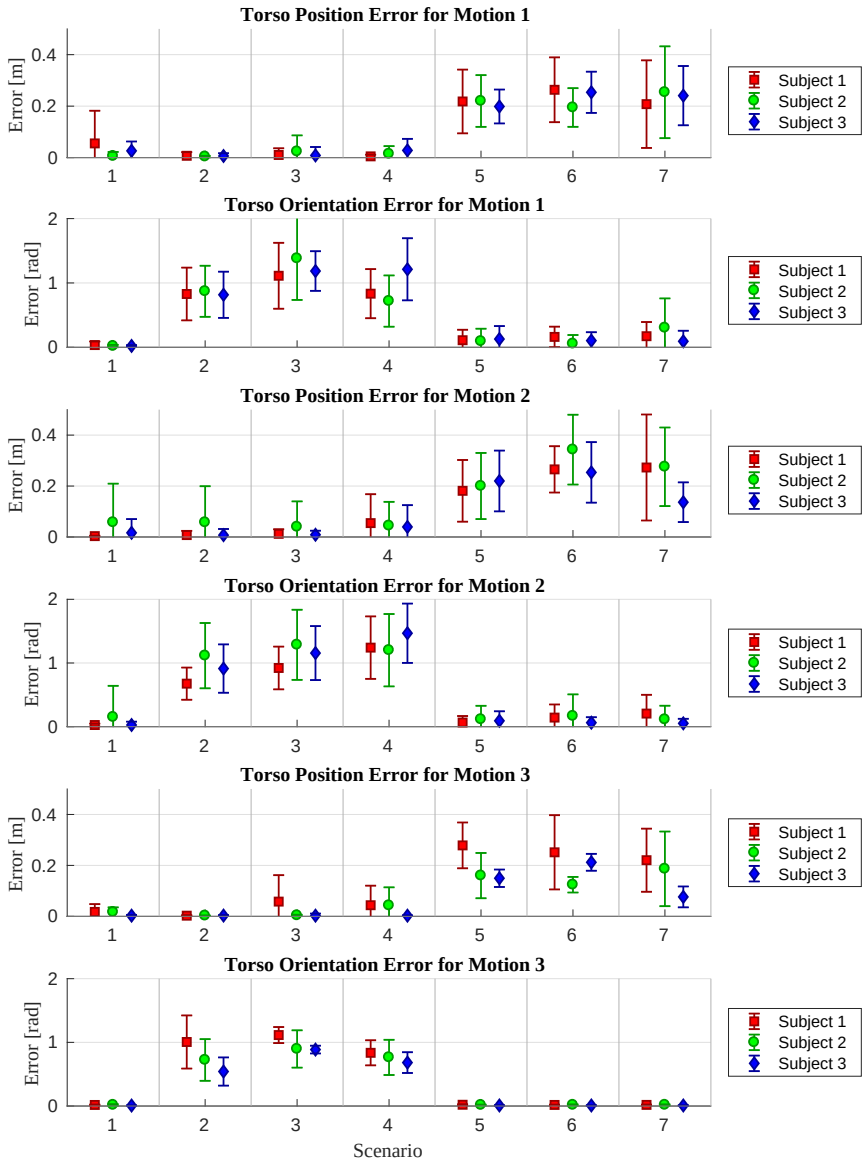


Figure 7: Average torso position and orientation error for all subject performing all three motions

the result this trend was not able to be seen as when comparing Scenario 2 to Scenario 4 (having elbow information compared to having no elbow information) there is very little change in the average orientation error and the standard error.

In Scenarios 5 to 7, torso position measurements are

not utilized in the estimation, and instead the approximated height of the subject is utilized. To assist the model in estimating the current configuration of the subject, a task that was implemented in these modes is to keep the torso upright by specifying a required rotation in the torso's roll and pitch. This resulted in a lower torso orientation error when compared to the results obtained in Modes 2-4. For all subjects in Modes 1, 5, 6 and 7, the orientation error in Motion 3 is reduced to almost zero. This is likely because the abrasive blasting task conducted for this motion required almost no change in the orientation of the user's torso with the task conducted mainly using the subject's arms.

As highlighted in Figure 7 depending on the information being used in Mode 1, the error in the estimate provided changes. In the calculations conducted in the investigation, it was assumed that the information was available at all times and the information obtained had no error. In reality, the measurement may not be free of error and available at all times. Despite being used with little sensory information, the proposed method was able to produce a pose estimation that was physically realizable due to the use of the biomechanical model integrated in the framework.

One advantage that this method provides is the lack of dependence on a single source of information or model approximation. The different scenarios highlighted in this paper are but a subset of a large number of combinations which could be altered depending on the necessity of the upper limb model. Sources of information may not be available at all times. An information source that may not be consistent could be given a lower priority. Utilizing this method, the information used between one time step to the next could be altered depending on the best available combination. Future work should consider the minimum number of tasks required by the method to produce an accurate pose estimation of the torso and upper limb, and the potential for improvements through dynamic task prioritization.

7 Conclusions

In this paper, a method for upper body pose estimation was presented. The estimation method exploits information obtained from the physical interaction between a human and robot, specifically the position and orientation of the human hands. The upper body pose of the operator is estimated through a task priority of available information or model-based approximations. From the results it was shown that the estimated torso position and orientation could be determined in different modes. The different modes utilized various amount of information and resulted in differing levels of accuracy. Utilizing this method, the upper body pose of the robot's operator is able to be estimated which will improve performance of operations requiring human-robot collaboration.

Acknowledgements

This work is supported in part by the Australian Research Council (ARC) Linkage Project (LP140100950), Burwell Technologies, an Australian Government Research Training Program (RTP) Scholarship and the Centre for Autonomous Systems (CAS) at the University of Technology Sydney. The authors would also like to thank the UTS Data Arena for facilitating the motion capture experiments.

References

[Carmichael and Liu, 2011] Marc G. Carmichael and Dikai Liu. Towards using musculoskeletal models for intelligent control of physically assistive robots. In *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, pages 8162 –8165, Sept 2011.

[Carmichael and Liu, 2013] Marc G. Carmichael and Dikai Liu. Estimating physical assistance need using a musculoskeletal model. *IEEE Transactions on Biomedical Engineering*, 60(7):1912–1919, 2013.

[Carmichael *et al.*, 2015] Marc G. Carmichael, Richardo Khonasty, and Dikai Liu. A multi-stage design framework for the development of task-specific robotic exoskeletons. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2015.

[Flacco *et al.*, 2015] F. Flacco, A. De Luca, and O. Khatib. Control of redundant robots under hard joint constraints: Saturation in the null space. *IEEE Transactions on Robotics*, 31(3):637–654, June 2015.

[Kang *et al.*, 2005] Tao Kang, Jiping He, and Stephen I Helms Tillery. Determining natural arm configuration along a reaching trajectory. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, 167(3):352–361, 2005.

[Karg *et al.*, 2014] Michelle Karg, Gentiane Venture, Jesse Hoey, and Dana Kulic. Human movement analysis as a measure for fatigue: A hidden markov-based approach. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2014.

[Kim *et al.*, 2012] Hyunchul Kim, Levi Makaio Miller, Nancy Byl, Gary M. Abrams, and Jacob Rosen. Redundancy resolution of the human arm and an upper limb exoskeleton. *IEEE Transactions on Biomedical Engineering*, 59(6):1770–1779, 2012.

[Klopčar and Lenarčič, 2005] Nives Klopčar and Jadran Lenarčič. Kinematic model for determination of human arm reachable workspace. *Meccanica*, 2005.

[Lenarčič and Umek, 1994] J. Lenarčič and A. Umek. Simple model of human arm reachable workspace. *Systems, Man and Cybernetics, IEEE Transactions on*, 24(8):1239 –1246, Aug 1994.

[Liu and Narayanan, 2013] Fan Liu and Ajit Narayanan. A human-inspired collision avoidance method for multi-robot and mobile autonomous robots. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2013.

[Lo Presti and La Cascia, 2016] Liliana Lo Presti and Marco La Cascia. 3D skeleton-based human action classification: A survey. *Pattern Recognition*, 2016.

[Menychtas *et al.*, 2016] Dimitrios Menychtas, Stephanie Carey, Rajiv Dubeh, and Derek Lura. A Robotic Human Body Model with Joint Limits for Simulation of Upper limb Prosthesis Users. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016.

[Moeslund *et al.*, 2006] Thomas B. Moeslund, Adrian Hilton, and Volker Kruger. A survey of advances in vision-based human motion capture and analysis, 2006.

[Ohayon and Rivlin, 2006] Shay Ohayon and Ehud Rivlin. Robust 3D Head Tracking Using Camera Pose Estimation. *ICPR 2006, 18th International Conference on Pattern Recognition*, 2006.

[Schwarz *et al.*, 2012] Loren Arthur Schwarz, Artashes Mkhitaryan, Diana Mateus, and Nassir Navab. Human skeleton tracking from depth data using geodesic distances and optical flow. In *Image and Vision Computing*, 2012.

[Shi *et al.*, 2014] Guangtian Shi, Yongsheng Wang, and Shuai Li. Human Motion Capture System and its Sensor Analysis. *Sensors & Transducers*, 172(6):206–212, 2014.

[Silva *et al.*, 2011] Miguel T. Silva, Andr F. Pereira, and Jorge M. Martins. An efficient muscle fatigue model for forward and inverse dynamic analysis of human movements. In *Procedia IUTAM*, 2011.

[Standards Australia, 2016] Standards Australia. Safety of machinery - Electrical equipment of machines. Technical report, Australian Standards IED 60204-1, 2016.

[Wang *et al.*, 2013] Lihui Wang, Bernard Schmidt, and Andrew Y C Nee. Vision-guided active collision avoidance for human-robot collaborations. *Manufacturing Letters*, 2013.