Evaluation of 3D Human Pose Estimation

1. Introduction

TuMeke Ergonomics offers an innovative AI-powered safety platform that revolutionizes workplace ergonomic risk assessment and mitigation, with a primary focus on preventing musculoskeletal injuries. Leveraging advanced computer vision technology, TuMeke analyzes worker movements from video footage captured by standard smartphones, generating detailed 3D models of human activity without requiring cumbersome wearable devices.[1]

Our platform excels in conducting automated ergonomic risk assessments, including the Rapid Upper Limb Assessment (RULA), Rapid Entire Body Assessment (REBA), National Institute of Occupational Safety and Health (NIOSH) Lifting Equation, and Revised Strain Index (RSI), providing faster and more efficient evaluations compared to traditional methods.[2]

Central to these automated assessments is our proprietary 3D human pose estimation technique. This sophisticated computer vision task determines the precise 3D positions and orientations of a person's body joints and bones from images or videos. To perform these assessments accurately, we calculate relative joint angles from the 3D human joints captured in the camera footage

This white paper examines the accuracy and validity of our 3D human pose estimation model. We explore the dataset used for model validation, the methodologies for joint angle calculations, and the overall performance of our system. This comprehensive analysis aims to demonstrate the robustness and reliability of TuMeke's ergonomic risk assessment technology.

2. Joint Angle Computation

Our ergonomic evaluation framework focuses on five critical joint angles essential for assessing workplace posture and movement: trunk (forward/lateral flexion and extension), neck (forward/lateral flexion and extension), shoulders (flexion and abduction), elbows (flexion), and knees (flexion). This section explores our methodologies for calculating these joint angles, providing insight into the biomechanical principles underlying our assessments.

Our local coordinate system (LCS) is aligned with the camera coordinate system. Our LCS orientation is defined as follows:

- +X axis: Oriented to the right
- +Y axis: Oriented posteriorly (towards the back)
- +Z axis: Oriented superiorly (upwards)

In the following subsections, we will delve into the specific computational methods for each joint angle, demonstrating how our advanced 3D pose estimation technology translates visual data into actionable ergonomic metrics.

2.1. Trunk

The trunk flexion angle is a crucial metric in ergonomic assessments, providing valuable insights into overall posture and potential risk factors for musculoskeletal disorders.

The trunk vector is established by connecting two key anatomical landmarks within our LCS: the mid-hip joint (pelvic center) and the mid-shoulder joint (shoulder girdle center). Given the 3D coordinates of the mid-hip joint $\mathbf{P}_{hip}$ and the mid-shoulder joint $\mathbf{P}_{shoulder}$, we can compute the trunk vector as follows:

$$V_{trunk} = P_{shoulder} - P_{hip}$$

The normalized trunk vector serves as a representation of the trunk's central axis in three-dimensional space, providing a foundation for accurate postural analysis. We normalize the trunk vector by dividing it by its magnitude:

$$V_{trunk,\ norm} = \frac{V_{trunk}}{\| V_{trunk} \|}$$

Using vector mathematics, we calculate the angle between the normalized trunk vector and the vertical Z-axis of our LCS. The vertical Z-axis vector $\mathbf{v}_z$ of the Local Coordinate System (LCS) is [0, 0, 1]. The angle between the normalized trunk vector and the vertical Z-axis can be calculated using the dot product:

$$\cos(\theta) = V_{trunk,\ norm} \cdot V_z$$

This computation yields the trunk angle, a quantitative measure that reflects the degree of trunk flexion or extension:

$$\theta = \arccos(V_{trunk, norm} \cdot V_z)$$

Here, the trunk angle θ is calculated using the cosine inverse of the dot product between the normalized trunk vector and the vertical Z-axis vector. This angle quantifies the trunk's orientation relative to the vertical axis, indicating the degree of flexion or extension.

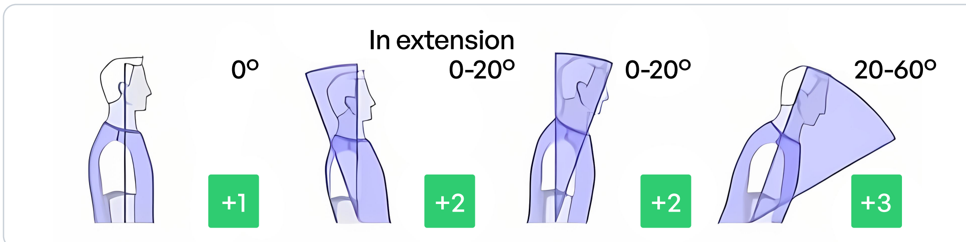

Figure 1 Computation of the trunk angle and the calculation of the REBA score.[3]

2.2 Neck

To accurately determine the neck angle, we employ a two-step process that accounts for both the neck's orientation and the trunk's position. This method ensures a precise measurement of the neck's flexion or extension relative to the shoulders.

Initial Neck Angle Calculation: We define a neck vector by connecting the thorax joint to the jaw joint in our LCS. Given the 3D coordinates of the jaw joint $\mathbf{P}_{jaw}$ and the thorax joint $\mathbf{P}_{thorax}$, we can compute the neck vector as follows:

$$V_{neck} = P_{jaw} - P_{thorax}$$

Next, we normalize the neck vector:

$$V_{neck, norm} = \frac{V_{neck}}{\| V_{neck} \|}$$

The angle between the normalized neck vector and the vertical Z-axis $\mathbf{V}_z = [0, 0, 1]$ is computed using the dot product:

$$\cos(\theta_{neck}) = V_{neck, norm} \cdot V_z$$

This provides the initial neck angle measurement:

$$\theta_{neck} = \arccos(V_{neck, norm} \cdot V_z)$$

Trunk Angle Calculation: The trunk angle is calculated using the method described in the Trunk section. Given the 3D coordinates of the mid-hip joint P_{hip} and the mid-shoulder joint $P_{shoulder}$, the trunk vector is computed as follows:

$$V_{trunk} = P_{shoulder} - P_{hip}$$

We then normalize the trunk vector:

$$V_{trunk, norm} = \frac{V_{trunk}}{\| V_{trunk} \|}$$

The angle between the normalized trunk vector and the vertical Z-axis is calculated using the dot product:

$$\cos(\theta_{trunk}) = V_{trunk, norm} \cdot V_z$$

This yields the trunk angle:

$$\theta_{trunk} = \arccos(V_{trunk, norm} \cdot V_z)$$

Relative Neck Angle Determination: To obtain the neck angle relative to the shoulders, we subtract the trunk angle from the initial neck angle:

$$\theta_{relative\ neck} = \theta_{neck} - \theta_{trunk}$$

This methodology accounts for the overall posture of the body, ensuring that the reported neck angle accurately reflects the relative position of the head to the shoulders.

Figure 2 Computation of the neck angle and the calculation of the REBA score.[3]

2.3 Shoulder

The flexion of the shoulder is a crucial measure that quantifies how much the upper arm is raised from the shoulder, relative to the body's orientation. This angle is particularly important in ergonomic assessments as it can indicate potential risk factors for shoulder strain and related musculoskeletal disorders. Our calculation method involves several steps:

1. Establishing a Body-Centric Reference Frame: We begin by creating a reference frame that aligns with the subject's body:

- We calculate an average lateral vector by averaging normalized vectors from the hips and shoulders

$$V_{\text{lateral}} = \frac{V_{\text{hip, norm}} + V_{\text{shoulder, norm}}}{2}$$

- A body vector is determined by connecting the pelvis to the thorax.

$$V_{\text{body}} = P_{\text{thorax}} - P_{\text{pelvis}}$$

- These vectors help us define a personalized coordinate system for each subject.

2. Creating a Custom Coordinate System: Using the vectors from step 1, we construct a 3D coordinate system:

- The X-axis aligns with the body vector (vertical orientation).

$$V_x = \frac{V_{\text{body}}}{\|V_{\text{body}}\|}$$

- The Z-axis aligns with the average lateral vector (side-to-side orientation).

$$V_z = \frac{V_{\text{lateral}}}{\|V_{\text{lateral}}\|}$$

- The Y-axis is derived to be perpendicular to both X and Z axes.

$$V_Y = V_Z - V_X$$

3. Analyzing Upper Arm Orientation: We then focus on the upper arm's orientation:

- The elbow-to-shoulder vector is calculated and normalized.

$$V_{\text{upper arm}} = P_{\text{shoulder}} - P_{\text{elbow}}$$

$$V_{\text{upper arm, norm}} = \frac{V_{\text{upper arm}}}{\|V_{\text{upper arm}}\|}$$

- This vector is projected onto our custom coordinate system.

$$V_{\text{upper arm, proj}} = V_{\text{upper arm, norm}} \cdot [\, V_x \;\; V_Y \;\; V_Z]$$

4. Calculating the Shoulder Angle: Finally, we determine the lateral arm angle:

- We use tangent inverse to compute the angle between the projected arm vector and the body's vertical axis.

$$\theta_{\text{shoulder}} = \arctan\left(\frac{V_{\text{upper arm, proj}} \cdot V_Y}{V_{\text{upper arm, proj}} \cdot V_X}\right)$$

- This calculation gives us the degree of upper arm raise in a way that accounts for the overall body posture.

This approach provides a more reliable assessment of arm positions in various work scenarios, contributing to more effective ergonomic evaluations and recommendations.

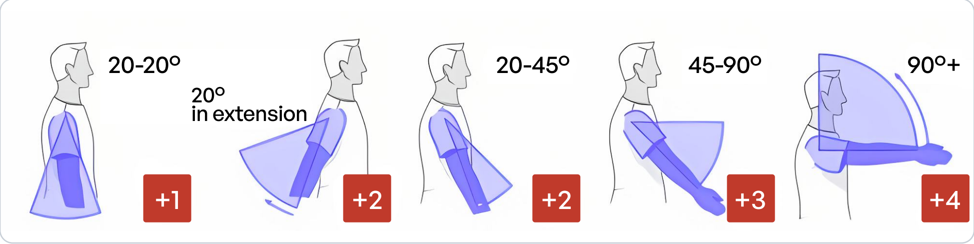

Figure 3 Computation of the shoulder angle and the calculation of the REBA score.[3]

2.4 Elbow

The elbow angle is a critical metric in ergonomic assessments, providing valuable insights into upper limb posture and potential risk factors for repetitive strain injuries. Our method for calculating the elbow angle employs a vectorbased approach, ensuring precise and consistent measurements.

We define two key vectors based on upper limb anatomical landmarks:

Shoulder-to-Elbow Vector: A normalized vector connecting the shoulder joint to the elbow joint. Given the 3D coordinates of the shoulder joint P_{shoulder} and the elbow joint P_{elbow}, we can compute the shoulder-to-elbow vector as follows:

$$V_{\text{shoulder to elbow}} = P_{\text{elbow}} - P_{\text{shoulder}}$$

Next, we normalize the shoulder-to-elbow vector:

$$V_{\text{shoulder to elbow, norm}} = \frac{V_{\text{shoulder to elbow}}}{\left\| V_{\text{shoulder to elbow}} \right\|}$$

Elbow-to-Wrist Vector: A normalized vector extending from the elbow joint to the wrist joint. Given the 3D coordinates of the elbow joint P_{elbow} and the wrist joint P_{wrist}, we can compute the elbow-to-wrist vector as follows:

$$V_{elbow\ to\ wrist} = P_{wrist} - P_{elbow}$$

We then normalize the elbow-to-wrist vector:

$$V_{elbow\ to\ wrist,\ norm} = \frac{V_{elbow\ to\ wrist}}{\left\| V_{elbow\ to\ wrist} \right\|}$$

Calculating the Elbow Angle: The elbow angle is computed by calculating the inverse cosine (arccos) of the dot product between these two normalized vectors, then converting the result from radians to degrees. The dot product of the two normalized vectors is calculated as follows:

$$\theta_{elbow} = \arccos\left(V_{shoulder\ to\ elbow,\ norm} \cdot V_{elbow\ to\ wrist,\ norm}\right)$$

This approach allows for precise quantification of upper limb positioning, which is essential for comprehensive ergonomic risk assessment and the development of targeted intervention strategies to prevent upper extremity disorders in the workplace.

Figure 4 Computation of the elbow angle and the calculation of the REBA score.[3]

2.5 Leg

The leg angle, often referred to as the knee angle, is a critical metric in biomechanical analysis, providing valuable insights into lower limb kinematics and potential risk factors for knee-related disorders. Our method for calculating the leg angle employs a vector-based approach, ensuring precise and consistent measurements across various postures and movements.

We define two key vectors based on lower limb anatomical landmarks:

Hip-to-Knee Vector: A normalized vector connecting the hip joint center to the knee joint center. Given the 3D coordinates of the hip joint P_{hip} and the knee joint P_{knee}, we can compute the hip-to-knee vector as follows:

$$V_{hip\ to\ knee} = P_{knee} - P_{hip}$$

Next, we normalize the hip-to-knee vector:

$$V_{\text{hip to knee, norm}} = \frac{V_{\text{hip to knee}}}{\left\| V_{\text{hip to knee}} \right\|}$$

Knee-to-Ankle Vector: A normalized vector extending from the knee joint center to the ankle joint center. Given the 3D coordinates of the knee joint P_{knee} and the ankle joint P_{ankle}, we can compute the knee-to-ankle vector as follows:

$$V_{\text{knee to ankle}} = P_{\text{ankle}} - P_{\text{knee}}$$

We then normalize the knee-to-ankle vector:

$$V_{\text{knee to ankle, norm}} = \frac{V_{\text{knee to ankle}}}{\left\| V_{\text{knee to ankle}} \right\|}$$

Calculating the Leg Angle: The leg angle is computed by calculating the inverse cosine (arccos) of the dot product between these two normalized vectors, then converting the result from radians to degrees. The dot product of the two normalized vectors is calculated as follows:

$$\theta_{\text{leg}} = \arccos(V_{\text{hip to knee, norm}} \cdot V_{\text{knee to ankle, norm}})$$

This approach allows for precise quantification of lower limb positioning, which is essential for comprehensive ergonomic risk assessment and the development of targeted intervention strategies to prevent knee-related disorders in the workplace.

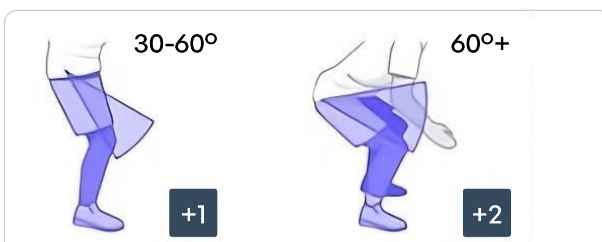

Figure 5 Computation of the leg angle and the calculation of the REBA score.[3]

3. Dataset

The Human3.6M dataset was chosen for its comprehensive and highly accurate annotations in controlled settings. This dataset enables us to rigorously evaluate our model's performance across a wide range of everyday activities. In the following sections, we will explore the specifics of the dataset, highlighting its unique features and explaining how it contributes to the robustness and versatility of our 3D pose estimation technology.

3.1 Human3.6M Dataset

Human3.6M[4] is renowned as one of the largest and most comprehensive datasets for 3D human pose estimation. It boasts an impressive collection of 3.6 million 3D human poses with corresponding images, captured in a controlled laboratory setting. The dataset features 11 professional actors (6 male and 5 female) performing 17 everyday activities, ensuring a wide range of human motions and poses.

What makes Human3.6M particularly valuable is its high-precision 3D joint positions, captured using a sophisticated motion capture system. The motion capture system employed in the Human3.6M dataset is centered around the Vicon T40, a high-resolution camera setup that captures detailed 3D human poses. Here's how the system works:

3.1.1 Equipment and Setup

- **Vicon T40 Cameras:** The dataset employs 10 Vicon T40 cameras, each with a resolution of 4 megapixels, capturing at a frequency of 200Hz. These cameras are strategically placed around the capture area to maximize the visibility and coverage of the subjects.
- **Basler piA1000 Cameras:** In addition to the Vicon system, 4 Basler piA1000 digital video cameras with a resolution of 1000x1000 pixels and a frequency of 50Hz are used to capture high-resolution video data.
- **Mesa SR4000 Time-of-Flight Sensor:** IA TOF sensor with a resolution of 176x144 pixels and a frequency of 25 Hz is included to provide depth information.
- **Human Solutions Vitus LC3 Body Scanner:** This 3D laser scanner captures accurate volumetric models of the subjects, with a point density of 7 dots/cm^2 and a tolerance of less than 1mm.

3.1.2 Technology and Methodology

- **Reflective Markers:** Small reflective markers are attached to key points on the subject's body. These markers reflect light back to the cameras, which track their positions over time.
- **Tracking and Labeling:** The system maintains the label identity of each marker and propagates it through time from an initial pose, which can be labeled either manually or automatically.
- **Fitting Process:** Using the positions and identities of the body markers, along with proprietary human motion models, the system infers accurate pose parameters. This process involves fitting a 3D skeleton model to the tracked marker data, allowing for precise reconstruction of human poses.

3.1.3 Data Collection and Synchronization

- **Capture Area:** The designated laboratory area is approximately 6m x 5m, providing an effective capture space of about 4m x 3m where subjects are fully visible to all cameras.
- **Synchronization:** Hardware and software synchronization ensure that all sensors and cameras capture data in unison, allowing for precise alignment of the video, depth, and motion capture data.

The scale of this dataset and the accuracy of its 3D annotations have established Human3.6M as a standard benchmark for evaluating 3D human pose estimation algorithms in controlled environments.

4. Results

This section presents the accuracy evaluation of our 3D pose estimation model using the Human3.6M dataset. Accuracy is quantified using three statistical measures: the mean absolute error (MAE) of the differences between the predicted and ground truth joint angles, the standard deviation of these differences, and the 90th percentile error. These metrics provide a comprehensive overview of the model's performance and reliability.

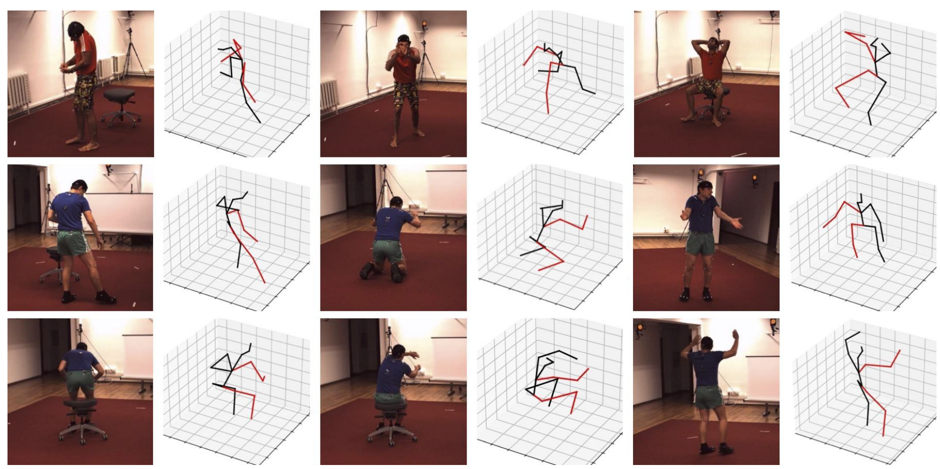

Figure 6 Sample images and their 3D ground-truth annotations from Human3.6M dataset

4.1 Metrics

Mean Absolute Error (MAE): The mean absolute error represents the average magnitude of the absolute errors between the predicted and ground truth joint angles. It provides a straightforward indication of the typical error the model makes when estimating joint angles. A lower MAE value indicates that the model's predictions are, on average, closer to the ground truth, which implies higher accuracy.

Standard Deviation: The standard deviation measures the dispersion or variability of the prediction errors. It quantifies how much the individual errors deviate from the mean error. A lower standard deviation suggests that the prediction errors are consistently close to the mean, indicating that the model's performance is stable and reliable. Conversely, a higher standard deviation would indicate more variability in the prediction errors, suggesting that the model's accuracy fluctuates more widely across different instances.

90th Percentile Error: The 90th percentile error represents the angle error below which 90% of the absolute differences between the predicted and ground truth joint angles fall. This metric provides an indication of the upper bound of the model's error distribution, showcasing its reliability in most cases. A lower 90th percentile error indicates that the model's predictions are accurate for the majority of instances.

4.1.1 Importance of Standard Deviation on Differences

Calculating the standard deviation on the differences between the ground truth and predicted values, rather than on the raw predictions, provides critical insights into the consistency of the model's errors. This approach helps in understanding whether the model's errors are consistently small or if there are occasional large errors. By focusing on the differences:

- **Error Consistency:** It becomes easier to identify if the model has a consistent error pattern across different joint angles and frames. Consistent errors with low standard deviation are preferable as they are predictable and can potentially be corrected.
- **Model Reliability:** Assessing the variability in errors helps in evaluating the model's reliability. A model with low variability in errors is considered more reliable as it performs consistently across different scenarios.

4.2 Results

The following table (Table 1) shows the mean absolute error (MAE), standard deviation, and 90th percentile error metrics for various joint angles in degrees for the Human3.6M dataset. The 90 %ile error indicates that 90% of the time, the prediction error for each joint angle is below the listed value, showcasing the reliability and precision of our 3D pose estimation model.

Joint Angle	Mean Absolute Error (degrees)	Standard Deviation (degrees)	Angle Error (Degrees) at 90%ile
Trunk	1.827	1.833	4.419
Left Shoulder	3.412	5.451	6.792
Right Shoulder	3.742	5.44	7.635
Left Elbow	4.267	5.869	9.121
Right Elbow	4.394	6.071	9.346
Left Leg	3.167	3.271	7.853
Right Leg	3.241	3.25	8.150

Table 1 Mean absolute error, standard deviation, and 90th percentile error of joint angle errors for Human3.6M dataset.

By utilizing the mean absolute error, standard deviation, and 90th percentile error metrics, we can comprehensively evaluate the performance of our 3D pose estimation model, ensuring it not only produces accurate predictions on average but also maintains consistency and reliability across different poses and frames. This multi-metric approach helps in identifying specific areas where the model excels and where further improvements may be needed.

5. Conclusion

The evaluation of our 3D pose estimation model underscores its accuracy and reliability, particularly when tested against the comprehensive Human3.6M dataset. The combination of low mean absolute errors, low standard deviation degrees, and low 90th percentile errors across multiple joint angles highlights the robustness of our approach. This high level of precision ensures that TuMeke Ergonomics can confidently provide actionable insights for preventing musculoskeletal disorders.

As we continue to refine our technology, our commitment to advancing ergonomic risk assessment remains steadfast. By integrating cutting-edge computer vision and AI with a deep understanding of human biomechanics, TuMeke is setting new industry standards in the quest to create safer workplaces. The results presented in this white paper affirm that our 3D pose estimation model is not only a powerful tool for current ergonomic challenges but also a foundation for future innovations in workplace safety and health.

This analysis reaffirms our mission to eliminate workplace musculoskeletal injuries, offering our clients a reliable, data-driven solution that continuously evolves to meet the demands of diverse and dynamic work environments.

References

1 https://www.tumeke.io/updates/step-by-step-conducting-your-first-ergonomic-risk-assessment-with-tumeke

2 https://www.tumeke.io/industry/manufacturing

3 https://www.tumeke.io/updates/reba-the-rapid-entire-body-assessment-comprehensive-overview

4 http://vision.imar.ro/human3.6m/description.php