Validation of a Human Pose Tracking Algorithm for Measuring Upper Limb Joints: Comparison With Photography-based Goniometry

Qingtang Zhu (zhuqingt@mail.sysu.edu.cn)  
First Affiliated Hospital of Sun Yat-sen University  

Jingyuan Fan  
First Affiliated Hospital of Sun Yat-sen University  

Fanbin Gu  
First Affiliated Hospital of Sun Yat-sen University  

Lulu Lv  
First Affiliated Hospital of Sun Yat-sen University  

Zhejin Zhang  
Guangdong AICH Technology Co.Ltd  

Changbing Zhu  
Guangdong AICH Technology Co.Ltd  

Jian Qi  
First Affiliated Hospital of Sun Yat-sen University  

Honggang Wang  
First Affiliated Hospital of Sun Yat-sen University  

Xiaolin Liu  
First Affiliated Hospital of Sun Yat-sen University  

Jiantao Yang  
First Affiliated Hospital of Sun Yat-sen University  

Research Article

Keywords: Pose estimation, range of motion, goniometry, automatic measurement, photography, reliability

Posted Date: January 13th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1117513/v1

License: ☑️ ① This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

**Background:** Range of motion (ROM) measurements are essential for diagnosing and evaluating upper extremity conditions. Clinical goniometry is the most commonly used methods but it is time-consuming and skill-demanding. Recent advances in human tracking algorithm suggest potential for automatic angle measuring from RGB images. It provides an attractive alternative for at-distance measuring. However, the reliability of this method has not been fully established. The purpose of this study is to evaluate if the results of algorithm are as reliable as human raters in upper limb movements.

**Methods:** Thirty healthy young adults (20 males, 10 females) participated in this study. Participants were asked to performed a 6-motion task including movement of shoulder, elbow and wrist. Images of movements were capture by commercial digital camera. Each movement was measured by a pose tracking algorithm and compared with the surgeon-measurement results. The mean differences between the two measurements were compared. Pearson correlation coefficients were used to determine the relationship. Reliability was investigated by the intra-class correlation coefficients.

**Results:** Comparing this algorithm-based method with manual measurement, the mean differences were less than 3 degrees in 5 motions (shoulder abduction: 0.51; shoulder elevation: 2.87; elbow flexion:0.38; elbow extension:0.65; wrist extension: 0.78) except wrist flexion. All the intra-class correlation coefficients were larger than 0.60. The Pearson coefficients also showed high correlations between the two measurements (p<0.001).

**Conclusions:** Our results indicated that pose estimation is a reliable method to measure the shoulder and elbow angles, supporting RGB images for measuring joint ROM. Our results proved the possibility that patients can assess their ROM by photos taken by a digital camera.

**Trial registration:** This study was registered in the Clinical Trials Center of The First Affiliated Hospital, Sun Yat-sen University (2021-387).

**Background**

Joint range of motion (ROM) is a measure of interest in clinical practice as it is significant for the diagnosis, functional assessment and treatment evaluation of the upper extremity. It is reported that measurement of ROM is required in more than 80% commonly used function assessment scales for the shoulder and elbow(1). Conventionally, the measurement of ROM was performed by manual goniometry(2). The goniometer is low-cost and portable, but its reliability highly depends on the rater's experience(3). Moreover, for the procedure is demanding and time-consuming, it is almost impossible to be frequently performed in daily clinical settings.

In addition, with the rapid development of telemedicine, how to determine the joint movement at-distance has peaked the interests of many researches(4–9). Typical motion capture system could provide accurate kinematics measurement(10, 11) but requires large space for data collection, which makes it costly, not portable, and thus impractical for home-use. Advances in smartphone technology, specifically the build-in sensors and high-resolution cameras, provides a potential platform for joint measurement. The number of mobile application used for clinical assessment have considerably increased in recent years(12). There are two main groups of these applications using the embedded sensor and images taken by phone camera. Compared with the sensor-based method, the photographic-based method provides recordable, documented information, easier procedure to follow.
and contactless measuring process, so it is well accepted by the patients(13). However, the existing application still need raters to mark on the photos(14), which means it could not actually reduce the workload of therapist nor the subjectivity of results.

Therefore, an object, accurate and automatic method is desired. Recent advances in human tracking algorithm offers a new option for this task. These algorithms can detect joint points and calculate joint angles from RGB images, providing an attractive alternative for at-distance measuring(11, 15). In this study, we employed OpenPose, one of the most widely used method proposed by Cao et al(16) to estimate joint angles from RGB images. Previous articles have evaluated the reliability of OpenPose-based system in gait analysis(17) and Parkinson rating(18, 19). Ota et al. also compared OpenPose and VICON (a 3D motion capture system) in measuring lower limb joint angle and found significant associations of the two methods(11). However, the utility of OpenPose in the assessment of upper limb angle remains unclear. Herein we constructed a measuring setup based on this algorithm, using RGB images to measure upper limb movements. This study evaluates the proposed method's accuracy and reliability by comparing the results with photography-based goniometry.

**Materials And Methods**

**Participants**

Thirty healthy young adults (20 males, 10 females, 22-35 years old), with no claim of medical history nor impairment in the upper limbs participated in this study. This study was approved by the institutional review board of our institution (2021-387). Estimated sample size was calculated by PASS software (version 15.0) using equivalence test for the difference between two means. With a type one error (a) of 0.05, power (1-b) of 0.95, equivalence limit of 10 degree, and standard deviation of 10, that a minimum of 27 samples would be required. All subjects were given full explanations about the motion tasks. After that, written consent was obtained for the use of their images for research purposes.

**Measurement setup**

The measurement environment is shown in Supplementary Figure1A. Three commercial digital cameras (HIKIVISION DS-2CD3T56FWDV2-I3) were positioned around the field (one in the front and two in the sides). The height of the cameras was 1.5m and the distance between the camera and subjects was 3m. To ensure the consistency of the participant placement, feet markers were placed 3m away from the cameras.

**Motion tasks and parameters extracting**

We designed a 6-task procedure including shoulder abduction, shoulder elevation, elbow flexion, elbow extension, wrist flexion and extension (Shown in Supplementary Figure1B). To control the impact introduced by rotation, all the interest angles in our design were fully presented in either sagittal or coronal view. Participants were asked to stand in the field and perform the motion tasks one after another. To ensure their performances were the same as we recommended, we set a screen in front of participants with word and video instructions. Moreover, their motion videos were real-time displayed on that screen as well. All photographs were taken from the anterior side, except the elbow flexion was taken from the lateral side (one for each side).
Automatic measurement

The landmarks of each joint were estimated by the Openpose Human Pose Estimation library (version 1.5.0)(16). The coordinates for landmarks of joints were further extracted, and skeleton models were rebuilding accordingly. Then, the joint angle was calculated by corresponding coordinates.

Digital photography-based measurement

After the automatic measurement, the photography-based measurements were conducted by using the same images. The angle of joints was measured by two hand surgeons individually, applying a screen goniometer software to the images displayed on the computer screen (The main reason of screen-goniometry was to make sure the posture present to measurement system and human researchers were identical. The validity of this method have been previously confirmed(20, 21)). To minimize the uncertainty of manual assessment, these images were reassessed by the same researchers at an interval of one week. The landmarks included the center of the shoulder, elbow and wrist, axis along the center of the upper arm and forearm, and central axis along the metacarpals. During the measurement, observers were free to locate the landmarks after reading the instruction. During this procedure, observers were not allowed to see the results of automatic measurement or another observer's report.

Data processing and statistical analysis

The mean values of the four measurements (2 researchers * 2 round) were considered as the standard results for comparison. All measurements are presented as mean±standard deviation (means ± sd). The deviation between the automatic assessment and standard results and the 95% confidence interval (CI) were calculated to assess the accuracy. The intra-class correlation coefficient (ICC) was also performed between the standard and the proposed measurement for assessing the agreement. Next, the results were analyzed using Bland and Altman analysis(22). The upper limits of agreement (LOA) were considered reference values to judge if the proposed measurement could be a reliable method for upper limb ROM. Since it has been confirmed that the results were in complete agreement when using Openpose to analyze the same image twice, the repeatability of the automatic methods was not assessed. In comparison, the repeatability of manual measurement was evaluated by comparing the test-retest results. In addition, to confirm the validity, linear regression analyses were conducted to compare the manual and system measurement data. R-square was calculated to evaluate the correlation between different methods.

Statistical analysis was performed by SPSS 22.0 (Armonk, NY: IBM Corp) and R software 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria). Results with p<0.05 were considered statistically significant. Interpretation of ICC value was as follows: <0.20: unacceptable, 0.20-0.40: questionable, 0.41-0.60: good, 0.61-0.80 very good, 0.81-1.00: excellent. The correlation coefficient, 1 indicates a total positive linear correlation, 0 means no linear correlation, and -1 shows a total negative linear correlation.

Results
The measuring results in the shoulder, elbow and wrist measured by two observers and the human tracking algorithm are summarized in Table 1 and Figure 1. The example of automatic measurements result is shown in Supplementary Figure 2.

### Table 1

Summary of the measurement results of different methods

<table>
<thead>
<tr>
<th>Motion Tasks</th>
<th>Sys(`x±sd ) (degree)</th>
<th>Doc1(`x±sd ) (degree)</th>
<th>Doc2(`x±sd ) (degree)</th>
<th>Doc_mean(`x±sd ) (degree)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shoulder abduction</td>
<td>94.2±5.13</td>
<td>92±4.93</td>
<td>95.38±4.5</td>
<td>93.69±5</td>
</tr>
<tr>
<td>Shoulder elevation</td>
<td>174.15±4.82</td>
<td>177.86±4.91</td>
<td>176.17±5.08</td>
<td>177.02±5.06</td>
</tr>
<tr>
<td>Elbow flexion</td>
<td>36.47±4.33</td>
<td>34.77±4.29</td>
<td>39.1±3.73</td>
<td>36.94±4.56</td>
</tr>
<tr>
<td>Elbow extension</td>
<td>-3.21±7.51</td>
<td>-3.91±7.92</td>
<td>-3.82±8.73</td>
<td>-3.87±8.32</td>
</tr>
<tr>
<td>Wrist extension</td>
<td>105.84±10.22</td>
<td>104.63±6.92</td>
<td>105.4±6.3</td>
<td>105.02±6.62</td>
</tr>
<tr>
<td>Wrist flexion</td>
<td>138.28±14.57</td>
<td>128.95±12.44</td>
<td>129.68±12.67</td>
<td>129.32±12.54</td>
</tr>
</tbody>
</table>

Sys: The automatic measuring system; Doc: Doctor; sd: standard deviation;

### Pose estimation

The poses of participants were successfully estimated in all but two images, and both were because of the person detection failure (The reason of error was due to these pictures included more than one person and the angle calculation was performed on the wrong target). The success rate was 99.44% (358/360).

### Difference between observers

The results of the inter and intra-observer comparison are presented in Table 2 and Table 3. There was excellent agreement between observers, with mean difference ranging from 0.08 to 4.33 and ICC value ranging from 0.897 to 0.951. The intra-observer comparison also indicates a good consistency, the mean differences between test and re-test measurements were less than 5 degrees.

### Difference between observer and machine

As shown in Table 2, the observer-system differences were comparable to the inter and intra-observer difference. The most significant difference was found in wrist flexion (8.96±12.71; 95%CI: -12.24--5.68). In the other 5 motions, the 95% confidence intervals of the mean differences between manual and automatic assessment were less than 5 degrees. Similarly, the Bland-Altman plots also indicate acceptable agreements for the shoulder and elbow motions. In comparison, the conformity for wrist motions is relatively poor (Figure 2), as the credible intervals were more than 10 degrees. Then, the consistency was further evaluated by ICC values. The results
suggested a good to excellent agreement (ICC>0.60) in all motions (Table 3). The lowest consistency was found in shoulder elevation and wrist extension (ICC=0.620), while the best was found in elbow extension (ICC=0.831). Additionally, linear correlations between system and observer measurement were also demonstrated (R ranges from 0.45 to 0.71, p<0.001 Figure 3).

Table 2
Summary of the inter-group differences

<table>
<thead>
<tr>
<th>Motion Tasks</th>
<th>Sys vs Doc_mean (degree)</th>
<th>Doc1_1 vs Doc1_2 (degree)</th>
<th>Doc2_1 vs Doc2_2 (degree)</th>
<th>Doc1 vs Doc2 (degree)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>`x±sd</td>
<td>95%CI</td>
<td>`x±sd</td>
<td>95%CI</td>
</tr>
<tr>
<td>Shoulder abduction</td>
<td>-0.51±4.38 (1.64-0.62)</td>
<td>2.93±4.83 (1.68-4.18)</td>
<td>-1.35±4.61 (2.54-0.16)</td>
<td>-3.38±2.08 (3.92-2.85)</td>
</tr>
<tr>
<td>Shoulder elevation</td>
<td>2.87±4.84 (1.62-4.11)</td>
<td>1.35±4.35 (0.23-2.48)</td>
<td>1.22±3.99 (0.19-2.26)</td>
<td>1.68±2.16 (1.13-2.24)</td>
</tr>
<tr>
<td>Elbow flexion</td>
<td>0.38±3.27 (-0.33-1.09)</td>
<td>3.28±4.91 (2.00-4.58)</td>
<td>-0.85±4.38 (-2.00-0.30)</td>
<td>-4.33±1.88 (-4.82-3.84)</td>
</tr>
<tr>
<td>Elbow extension</td>
<td>-0.65±5.48 (-2.07-0.76)</td>
<td>-1.76±8.25 (-3.89-0.38)</td>
<td>-1.84±8.92 (-4.14-0.47)</td>
<td>-0.08±3.57 (-1.01-0.84)</td>
</tr>
<tr>
<td>Wrist extension</td>
<td>-0.78±8.40 (-0.55-1.39)</td>
<td>-1.47±8.29 (-3.62-0.67)</td>
<td>-1.46±7.11 (-3.29-0.38)</td>
<td>-0.77±2.61 (-1.44-0.09)</td>
</tr>
<tr>
<td>Wrist flexion</td>
<td>-8.96±12.71 (-12.24-5.68)</td>
<td>-3.33±16.70 (-7.64-0.99)</td>
<td>-2.24±16.86 (-6.60-2.11)</td>
<td>-0.73±4.01 (-1.77-0.30)</td>
</tr>
</tbody>
</table>

Sys: The automatic measuring system; Doc1_1: The first measurement of the first doctor, the rest are in the same manner; Doc_mean: The average measuring value of doctors; sd: standard deviation; CI: confidence interval

Table 3
Summary of the intra-class correlation coefficients
The range of motion (ROM) of the upper limb is an important clinical parameter to various functional evaluations before and after treatment. Conventionally, ROM was assessed manually using the standard goniometer. This procedure is time-consuming and requires expertise. However, various reasons such as financial, geographic, or busy schedule could prevent patients from clinic visiting. Therefore, telemedicine has become popular as a method of patient evaluation. Photographs are easily obtained and disseminated in our daily life. Getting movement parameters from remote photographs has potential to decrease the cost of physical evaluation. Human pose tracking algorithms can automatic calculate joint angles from RGB images and provide a new option for the remote evaluation. However, the reliability of this method is extremely important before the using in clinical settings.

This study sought to evaluated the reliability of an automatic goniometry method. In our analysis, we found that the algorithm-based method has acceptable reliability compared to human observers. The results indicate that the differences between the proposed method and the average value of observers are less than 5 degrees in shoulder and elbow motions, comparable to the inter and intra-observer differences. Compared to that reported in previous studies, these differences are notably more minor than that of visual estimation and are comparable to inertial sensors and depth camera. Therefore, the proposed method may have great accuracy and reliability in measuring ROMs of the shoulder and elbow.

In this study, the greatest observer-machine difference was found in wrist flexion, and the mean value was 8.96 degrees. However, this reliability is still competitive compared to other image-based applications. Nevertheless, as seen in the Bland-Altman plots, we found the angle was over-estimated by the system in most cases. Thus, we speculate this might be a systematic error that could be correct when a larger sample size is available.

It is difficult for participants to keep their posture still during measuring, as previous studies indicated. According to the literature, several methods were employed to minimize this problem. Cook et al. used a wooden triangle with fixed internal angles to support the joints of interest during assessing. In comparison, Chang et

<table>
<thead>
<tr>
<th>Motion Tasks</th>
<th>Sys vs. Doc_mean</th>
<th>Doc1 vs.Doc2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ICC</td>
<td>95%CI</td>
</tr>
<tr>
<td>Shoulder abduction</td>
<td>0.693</td>
<td>0.485-0.816</td>
</tr>
<tr>
<td>Shoulder elevation</td>
<td>0.620</td>
<td>0.363-0.773</td>
</tr>
<tr>
<td>Elbow flexion</td>
<td>0.787</td>
<td>0.639-0.874</td>
</tr>
<tr>
<td>Elbow extension</td>
<td>0.831</td>
<td>0.717-0.899</td>
</tr>
<tr>
<td>Wrist extension</td>
<td>0.620</td>
<td>0.363-0.773</td>
</tr>
<tr>
<td>Wrist flexion</td>
<td>0.623</td>
<td>0.369-0.775</td>
</tr>
</tbody>
</table>

Sys: The automatic measuring system; Doc_mean: The average measuring value of doctors;

Discussion

The range of motion (ROM) of the upper limb is an important clinical parameter to various functional evaluations before and after treatment. Conventionally, ROM was assessed manually using the standard goniometer. This procedure is time-consuming and requires expertise. However, various reasons such as financial, geographic, or busy schedule could prevent patients from clinic visiting. Therefore, telemedicine has become popular as a method of patient evaluation. Photographs are easily obtained and disseminated in our daily life. Getting movement parameters from remote photographs has potential to decrease the cost of physical evaluation. Human pose tracking algorithms can automatic calculate joint angles from RGB images and provide a new option for the remote evaluation. However, the reliability of this method is extremely important before the using in clinical settings.

This study sought to evaluated the reliability of an automatic goniometry method. In our analysis, we found that the algorithm-based method has acceptable reliability compared to human observers. The results indicate that the differences between the proposed method and the average value of observers are less than 5 degrees in shoulder and elbow motions, comparable to the inter and intra-observer differences. Compared to that reported in previous studies, these differences are notably more minor than that of visual estimation and are comparable to inertial sensors and depth camera. Therefore, the proposed method may have great accuracy and reliability in measuring ROMs of the shoulder and elbow.

In this study, the greatest observer-machine difference was found in wrist flexion, and the mean value was 8.96 degrees. However, this reliability is still competitive compared to other image-based applications. Nevertheless, as seen in the Bland-Altman plots, we found the angle was over-estimated by the system in most cases. Thus, we speculate this might be a systematic error that could be correct when a larger sample size is available.

It is difficult for participants to keep their posture still during measuring, as previous studies indicated. According to the literature, several methods were employed to minimize this problem. Cook et al. used a wooden triangle with fixed internal angles to support the joints of interest during assessing. In comparison, Chang et
al. adopted a glass plate as hand support to reduce movement during the 3D scanning process (32). More commonly, many studies choose the 3D motion capture system to achieve data collecting (33–35) simultaneously and thus minimize the differences caused by involuntary posture changing. Our study compared the results of the automatic system and human observers by measuring the same image individually. In this way, we can conclude the actual differences between the two methods without impacting the inconsistency of motions. The concept of obtaining joint ROM from photographs is not new. Previous studies have indicated that it is accurate and reliable compared with conventional clinical goniometry (20, 21). Additionally, the results of image-based goniometry could be more consistent than that of the conventional way in some cases (21). This present study also proved the value of screen goniometry as a reliable alternative for measuring, with slight inter and intra-observer differences.

There are still some limitations of our study: Firstly, the participants were limited to young, healthy persons, and did not include the elderly nor the patients, making the results statistically less robust and lessening the generalizability of the proposed method. Secondly, motions with rotation were not assessed because it was hard to estimate 3D motions through 2D images. Although it could be an inevitable technical error, this issue will be the aim of our future studies. In addition, angle of joints may contain the movements of several joints (For example, the angle of shoulder joint includes the movement of the scapula, thorax, and thoracic spine) which lead to inaccurate of measurement, but we believe that is still good enough for telemedicine system. Another drawback is that the accuracy of our method depends on the compliance and cooperation of participants to some extent. If the subject cannot properly understand our purposes, the results can exhibit deviation.

**Conclusions**

This study demonstrates a reliable and valid method to measure joint ROM of the upper limb using RGB photographs. It provides an exciting alternative to remote evaluation, adding objective and accurate information on upper limb mobility. Besides, the proposed method is a fully automatic technique. Users can obtain reliable kinematics parameters personally without travel to clinical centers. However, it would be interesting to implement a study with a larger sample of patients or the elders with movement disorders and study more motions.

**Abbreviations**

ROM: range of motion; LOA: limits of agreement; ICC: intra-class correlation coefficient; CI: confidence interval;

**Declarations**

**Ethics approval and consent to participate**

The study was approved by the Ethics Committee of the First Affiliated Hospital of Sun Yat-sen University. All participants provided informed consent prior to participation in this study (reference number: 2021-387) AND publication of identifying information/images in an online open-access publication. We confirm that all methods were performed in accordance with the relevant guidelines and regulations.

**Consent for publication**
Availability of data and materials

The data analyzed during this study are not publicly available due to containing identifying information (the accuracy of the algorithm would be impacted if the eyes/facial region was blurring) but are available from the corresponding author on reasonable request.

Competing interests

The authors declared no competing interests

Funding

None

Authors’ contributions


Acknowledgements

We want to thank Prof. Bai Leng, Prof. Zhenguo Lao and Jianwen Zheng for technical help, writing assistance, and general support.

References

1. Norvell DC. Musculoskeletal Outcomes Measures and Instruments, 2 Vols.


**Figures**

**Figure 1**

Comparison of the measurement results of the 6 motions.
Sys: The automatic measuring system; Doc1_1: The first measurement of the first doctor, the rest are in the same manner;

Figure 2

Bland-Altman plots for inter-rater agreement.

Sys: The automatic measuring system; Doc: Doctor; Doc_mean: The average measuring value of doctors;

This plot compares the individual measurement result with the average value of doctors. The x-axis represents the mean value; the y axis represents the inter-rater difference. The dotted lines represent the limit of differences.

Figure 3

The linear correlation between raters

Sys: The automatic measuring system; Doc_mean: The average measuring value of doctors;

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure1.tif
- SupplementaryFigure2.tif