Towards an Automated Assessment of Musculoskeletal Insufficiencies

Danny Dressler[1], Pavlo Liapota[2], and Welf Löwe[3]

[1] AIMO AB, https://www.aimo-health.com
[2] Softwerk AB, https://softwerk.se
[3] DISA, Linnaeus University, Welf.Lowe@lnu.se, https://lnu.se/disa

Abstract. The paper suggests a *quantitative* assessment of human movements using inexpensive 3D sensor technology and *evaluates* its accuracy by comparing it with human expert assessments. The two assessment methods show a high agreement. To achieve this, a *novel sequence alignment* algorithm was developed that works for arbitrary time series.

1 Introduction

Healthcare is in the middle of fundamental changes from fee-for-service to value-centred systems. Approaches for payment based on patient value (best possible health achieved) and system value (effective treatments at efficient costs) need to be able to measure clinical outcomes. This capability, until recently, was not part of most hospital, health, or enterprise-resource-planning systems [4]. With inexpensive sensor technologies and data analytics becoming increasingly available, it is nowadays possible to collect data on the clinical activities of healthcare, the health status of a patient and the change in this status after treatment. Our work contributes to changes in healthcare with an automated objective assessment of physical health of the human musculoskeletal system to help diagnose, predict or prevent related pain, injuries and chronicle diseases.

Like other approaches,[4,5] our solution supports the diagnosis of musculoskeletal issues based on inexpensive 3D motion sensing devices, e.g., the Kinect.[6] These approaches are tools for physical therapists and add little value to the caring or nursing process. Our approach localizes issues by identifying *weak links* and quantifies their severity. It assesses an overhead deep squat exercise standardized by the National Academy of Sports Medicine (NASM)[7] that comes with movement execution and scoring specifications. It is medically validated. A low NASM score is an indicator of mobility and stability insufficiencies that, in turn, indicate current or future problems with the musculoskeletal system [9].

Quite a few studies research the recognition of human movements using commodity 3D sensor technology [18, 17, 20, 21, 16, 5, 11, 14]. While these approaches

---

[4] https://www.qinematic.com
[5] https://kinetisense.com
[6] https://en.wikipedia.org/wiki/Kinect
[7] https://en.wikipedia.org/wiki/National_Academy_of_Sports_Medicine

are similar to ours in their feature extraction and preprocessing steps, their goal is the classification of different movements not their quality assessment. The same technology has also been used in movement quality assessment [19, 6, 10, 15]. However, these studies aim at *qualitative* assessments or at identifying different abnormal movement patterns rather than at *quantitatively* scoring the quality of a movement. For some of the studies [10, 15], preprocessing transformations such as dimensionality reduction using manifold learning [3] enable a fast online assessment but make it impossible to localize the impairments. Pirsiavash et al. suggest a learning framework for training models able to quantitatively assess the quality of human movements from 2D videos [12]. Their approach trains a regression model from spatiotemporal pose features to scores obtained from expert judges. Features are extracted using unsupervised feature learning directly from 2D video data. Therefore, a localization of issues is not possible.

Matching equal phases of different movement sequences is an important subproblem of human movement recognition and assessment. Approaches based on Dynamic Time Wrapping (DTW) [13] and Hidden Markov Models (HMM) [2] are arguably predominant in addressing the matching problem for general sequences and also for movement sequences. DTW approaches are used, e.g., in [17, 19, 6, 16], HMM approaches in [20, 15]. None of the two approaches can ignore irrelevant leading, intermediate or trailing subsequences. Cutting out these subsequences, if done at all, is handled in (manual) preprocessing steps. While DTW based techniques have been improved in robustness [7] and performance [6, 8], our *Sequence Alignment* algorithm is, to the best of our knowledge, the first DTW generalization able to find an optimal matching of *relevant subsequences*.

The paper contributes with (i) an automated, quantitative and localizable assessment of issues with the human musculoskeletal system and with (ii) an evaluation of its accuracy by comparing with human expert assessments. Feature extraction and preprocessing steps include (iii) a novel sequence alignment algorithm matching user and reference exercises. It is generally applicable to any time series, not limited to human movement recordings. Section 2 introduces the NASM assessment system and basic definitions. Section 3 introduces our assessment approach including our sequence alignment algorithm. Section 4 evaluates the approach. Section 5 concludes the paper and shows directions of future work.

2 Setting the Scene

The AIMO assessment and scoring algorithm is a digitalization of the NASM overhead deep squat assessment and scoring, i.e., it implements the NASM scoring specification. The NASM suggests assessing different potential weak links of a body from feet to head and to score them individually. An overall NASM score is then set based on the scores for these weak links contributing to the score with different weight factors. The potential weak links are **heels rise** contributing with a high weight factor to the overall score, **knees move inwards** or **outwards**, **excessive forward lean** and **asymmetrical weight shift** contributing with medium weight factors, and **arms fall forward**, **forward head**,

and **shoulder elevation** contributing with low weight factors. The weak links **lower back arches** or **rounds** are dropped due to the low camera sensitivity.

A *sequence* is a recording of a human movement, e.g., an overhead deep squat. It is a list of frames. It is called a *master* sequence if the movement is executed correctly, e.g., is a correct deep squat according to the NASM execution standard. Any sequence that is to be scored is called a *user* sequence. Each *frame* is a part of a sequence and describes the body posture at a specific point in time during the recorded movement. It is a record of features. Each *feature* is a part of a frame and describes an aspect of the body posture at specific point in time. A feature is called *direct* if it is directly measured by the 3D camera or *indirect* if it is computed from the direct or other indirect features. The direct features include the x, y, and z coordinates of skeleton joints. The Kinect identifies 24 joints; their x, y, and z positions are given in a 3D coordinate system with the camera in its origin and a horizontal x, z plane with z the depth coordinate.

NASM features are special indirect features that are used to assess user sequences according to the NASM standard. For deciding whether or not an overhead deep squat is executed well enough to be scored, we compute the angles between the tights projected to the y, z plane and the z axis, between the tights and lower legs projected to the y, z plane, and between the legs (vectors $Hip \rightarrow Ankle$) projected to the x, y plane and the y axis. The former features are used to decide whether or not the squat was deep enough. The latter feature is used to decide whether or not the legs were too widely open. NASM features also include features for indicating weak links and eventually computing an overall AIMO score. More specifically, we compute for **heels rise**: the angle between the lower leg and the z axis. Note that a direct assessment of the foot to lower leg angle or the foot to z axis angle did not work since the camera does not provide reliable foot positions; for **knees move inwards or outwards**: the angle between the lower leg projected to x, y plane and the y axis; for **excessive forward lean**: the angle between the vector $Spine\ Base \rightarrow Spine\ Shoulder$ projected to y, z plane and the y axis; for **asymmetrical weight shift**: the angle between the vector $Spine\ Base \rightarrow Spine\ Shoulder$ projected to x, y plane and the y axis; for **arms fall forward**: the angle between the limb $Shoulder \rightarrow Elbow$ and the vector $Spine\ Base \rightarrow Spine\ Shoulder$ projected to the y, z plane; for **forward head**: the angle between the limb $Neck \rightarrow Head$ projected to y, z plane and the y axis, for **shoulder elevation**: the angle between limb $Spine\ Shoulder \rightarrow Shoulder$ projected to the x, y plane and x axis.

An *aggregated sequence* aggregates one or more sequences into one. It is a list of aggregated frames. An *aggregated frame* aggregates two or more frames into one. It is a vector of sample distributions of the feature values of each feature. For scoring a user sequence, we compare it with an aggregated master sequence.

3   The Assessment Algorithm

The assessment and scoring algorithm consists of four steps described in the subsections below. (1) *Building the aggregated master sequence* is only performed

once while the following steps are performed for each user sequence. (2) *Preparing the sequences* mitigates noisy feature values due to random camera errors and makes sure that a user sequence is scoreable. (3) *Matching* aligns the user sequence with the aggregated master sequence. (4) *Scoring* computes indicators for the individual weak links and the overall AIMO score.

3.1 Building the Aggregated Master Sequence

Assume to have a set of master sequences. We first build an initial aggregated master sequence. If the set contains more than one master sequence, the initial aggregated master sequence is aggregated with the other master sequences.

Building an initial aggregated master sequence All but the first two steps are done automatically. (1.1) Select the best master sequence. It should have a constant movement speed without any delays and stops. (1.2) Cut off leading and trailing frames of postures that do not belong to the movement. (1.3) Prepare the master sequence, i.e., apply all steps before matching and all but the first step after matching, cf. Section 3.2. (1.4) Group subsequent frames: each group contributes to a separate aggregated frame. Frames are added to a group based on the average of angles between the tights projected to the y, z plane and the z axis. We define 31 groups: 15 groups where these angles are decreasing, one group around the deep down position, and 15 groups where the angles are increasing. (1.5) Separately for each group and for each feature, compute a numerical sample distribution of the feature values. (1.6) Depending on the angles between the tights projected to the y, z plane and the z axis, some aggregated frames are *skippable*, i.e., they may be, but do not necessarily need to be matched, cf. Section 3.3. If for an aggregated frame the average of these two angles (for the left and the right leg) is smaller than $-0.2rad \approx -11.5°$, i.e., the hips are lower than the knees, then it is *skippable*.

Extending the aggregated master sequence The following steps are executed only if there are any master sequences remaining and, if so, for each of them in arbitrary order: (1.7) Prepare the master sequence, cf. Section 3.2 **before matching**. (1.8) Match the master sequence with current aggregated master sequence, cf. Section 3.3. Matching maps each frame of the master sequence to an aggregated frame of the aggregated master sequence. (1.9) Prepare the master sequence, cf. Section 3.2 **after matching**. (1.10) For each frame and each feature of the master sequence, add the feature value to the sample distribution of the respective feature of the matched aggregated frame.

3.2 Preparing the Sequences

Before matching (2.1) *Floor clip plane alignment*: for each frame, the joint position vectors are rotated such that the floor clip plane is parallel to x, z plane. (2.2) *Smoothening*: for all *direct* features, a sliding average of feature values is computed. The window length is 11 frames with 2 outlier frames on both sides. More specifically, feature values of 11 consecutive frames are sorted, the 2 smallest and 2 largest values are removed, then the average of the remaining

values is calculated. (2.3) *Transforming to Spine Base*: compute a translation transformation that moves to the *Spine Base* joint of the first frame of the sequence to the mean *Spine Base* joint position of the first aggregated frame of the aggregated frame sequence. Then apply this translation to all other frames of the sequence. Note that this translation has no influence on the scoring and is done just to make visual debugging easier. 2.4) *Interpolate*: if a joint was not visible for less than $k = 3$ sequential frames, then its position is interpolated. If a joint was not visible for k or more sequential frames, then the joint is considered not tracked in the sequence. If the joint is important to calculate an NASM feature, the sequence is not further processed or scored and an error is reported. **After matching** The following transformations require the sequence to be matched to an aggregated master sequence: (2.5) *Cut* leading and trailing (unmatched) frames of postures that do not belong to the movement. (2.6) *Scaling*: compute a scaling transformation (using a so-called procrustes analysis) that moves each joint of the first frame of the sequence to the corresponding mean joint position of the first aggregated frame of the aggregated frame sequence. Then apply this scaling to all other frames of the sequence. This is an optional transformation as it has no effect on the NASM features (all angles). (2.7) *Hip rotation*: compute a rotation transformation that rotates the *Spine Base* → *Hip* limbs of the first frame of the sequence to let it skeletons "look" towards the camera. This means, the following two angles should be equal: the angle between *Spine Base* → *Left Hip* limbs projected to x, z plane and the z axis and the angle between the *Spine Base* → *Right Hip* limbs projected to x, z plane and the z axis. Then apply this rotation to all other frames of the sequence. (2.8) *Fixing the ankle*: we observed that the ankle positions received from sensor are considerably flickering in the middle of the deep squat while the ankles are almost impossible to be moved during the exercise. Therefore, the ankle positions in all frames of a sequence are replaced with ankle positions of the first frame. (2.9) *Compute* the indirect features, e.g., NASM features, from the direct features for each frame. (2.10) *Reject* a sequence if the deep squat is not deep enough or the legs are too widely open.

3.3   Matching—The Sequence Alignment Algorithm

Let N be the number of aggregated master sequence frames and M be the number of user sequence frames. A *matching* $\mathcal{M}$ is a relation $\subseteq [1 \ldots N] \times [1 \ldots M]$. A matching $\mathcal{M}$ is *correct iff*

(i) $\forall n \in [1 \ldots N] : \begin{cases} (n, \_) \in \mathcal{M} \vee & \text{(matched)} \\ n \in skippable & \text{(unmatched)} \end{cases}$

(ii) $\forall m \in [1 \ldots M] : \begin{cases} (\_, m) \in \mathcal{M} \vee & \text{(matched)} \\ \forall m' \leq m : (\_, m') \notin \mathcal{M} \vee & \text{(unmatched leading)} \\ \forall m' \geq m : (\_, m') \notin \mathcal{M} & \text{(unmatched trailing)} \end{cases}$

(iii) $(n, m) \in \mathcal{M} \Rightarrow \nexists (n', m') \in \mathcal{M} \wedge n' < n \wedge m' > m$

For a matching to be correct, all aggregated master sequence frames are matched or *skippable* (i), all user sequence frames are matched, leading or trailing (ii), and the matching must obey the order of frames in the sequences (iii).

Known algorithms for sequence alignment such as DTW- or HMM-based algorithms are not capable of coping with all three conditions, e.g., they cannot ignore skippable, leading and trailing frames. We therefore developed our own matching algorithm called *Sequence Alignment* that can be understood as a generalization of DTW. *Sequence Alignment* finds a matching with minimum costs among all *correct* matchings.

The costs of a matching are defined as the deviations of the aggregated master sequence frames and the matched user sequence frames. More specifically, the deviation of a feature value v of a user sequence frame from a distribution D of a the corresponding feature values in an aggregated master sequence frame with mean μ and standard deviation σ is the z score of v in D, i.e., $z = \frac{v - \mu}{\sigma}$.

Data: $N, M : \mathbb{N}$; $dev : [1 \dots N, 1 \dots M]$ *of* $\mathbb{R}$; $skippable : [1 \dots N]$ *of* $\mathbb{B}$
Result: $costs : \mathbb{R}$
1 $costs = \text{MAX\_VALUE}$
2 **for** $k \in 1 \dots M$ **do**
3 $\quad$ $costs = min(costs, Seq\_Align(0, 0, k)/k)$
4 **end**
5 **return** $costs$
6 **function** $Seq\_Align(n, m, k)$**:**
7 $\quad$ **if** $k == 0$ **then**
8 $\quad\quad$ **return** $n == N$? 0 : MAX_VALUE
9 $\quad$ **end**
10 $\quad$ $costs = \text{MAX\_VALUE}$
11 $\quad$ **if** $n < N \wedge m < M$ **then**
12 $\quad\quad$ $costs = min(costs, dev[n, m] + Seq\_Align(n + 1, m + 1, k - 1))$
13 $\quad$ **if** $n > 0 \wedge m < M$ **then**
14 $\quad\quad$ $costs = min(costs, dev[n - 1, m] + Seq\_Align(n, m + 1, k - 1))$
15 $\quad$ **if** $m > 0 \wedge n < N$ **then**
16 $\quad\quad$ $costs = min(costs, dev[n, m - 1] + Seq\_Align(n + 1, m, k - 1))$
17 $\quad$ **if** $n == 0 \wedge m < M$ **then**
18 $\quad\quad$ $costs = min(costs, Seq\_Align(n, m + 1, k))$
19 $\quad$ **if** $n < N \wedge skippable[n]$ **then**
20 $\quad\quad$ $costs = min(costs, Seq\_Align(n + 1, m, k))$
21 $\quad$ **end**
22 **return** $costs$

Algorithm 1: Minimum matching costs

The deviation of a user *sequence* to an aggregated master sequence is the average deviation of the matched user and aggregated frames. The deviation of a single user *frame* to the matched aggregated master sequence frame is the average of the standard deviations of all contained features. Let $dev[n, m]$ be the deviations of the m-th frame in a user sequence matched to the n-th aggregated frame in

a master sequence, $1 \leq n \leq N, 1 \leq m \leq M$ with M and N the length of the user and the aggregated master sequences, resp. Let $skippable[n]$ be true if the aggregated master sequence frame n is skippable. Algorithm 1 computes the minimum cost of a matching. The recursive function Seq_Align computes the minimum costs for an aggregated master sequence starting at frame n and a user sequence starting at frame m computing a matching of cardinality k, i.e., the matching contains k pairs of aggregated master and user frames. If applicable, the algorithm tries to match the aggregated fame n with the user frame m (line 12), match the aggregated fame $n-1$ with the user frame m (line 14), match the aggregated fame n with the user frame $m-1$ (line 16), skip a leading user frame m (line 18), skip a $skippable$ aggregated frame n (line 20), and selects whatever leads to minimum costs. The matching is derived as a side effect of this selection.

Dynamic programming avoids multiple recursive invocations of Seq_Align with the same parameters. While still trying all correct matches, it keeps book of the minimum costs of a matching resulting from calls to $Seq\_Align(n, m, k)$. This leads to a cubic algorithm that is fast in practice. With an Oracle JDK 8 on an Intel Core i7, 2.5 GHz, matching takes about 0.5 seconds for a user sequence of 150 frames, i.e., ca. 5 seconds of recording.

For the NASM deep squat, matching is computed based on the angles between the tights projected to the y, z plane and the z axis. We assume that these two features, for the left and the right leg, respectively, follow the same distribution and have the same standard deviation σ. As a consequence, we simplify the deviation calculation for the sake of computing performance without changing the optimum matching and use $z' = v - \mu$ instead of the z-score.

3.4   Scoring

Offline and once for the aggregated master sequence, we calculate weights w between 0 and 1 for its frames. Weights are set in such a way that the relaxed positions at the beginning and at the end of the deep squat exercise are ignored ($w = 0$) while the demanding deep down position is regarded as it is ($w = 1$). A quartic interpolation of weights in between makes them grow rapidly to almost 1. As for matching, weights are calculated based on the (average of the) angles between the tights projected to the y, z plane and the z axis. For the aggregated master sequence, let the maximum angle be a_{max} ($\approx 90°$) at the beginning and at the end of the deep squat and let a_{min} ($\approx -30°$) be the minimum angle in the deep down position. The weight w of an aggregated frame with the average a of these angles is set to $w = 1 - \left(\frac{a - a_{min}}{a_{max} - a_{min}} \right)^4$.

Scoring is based on the NASM features for assessing the weak links. Recall that all these features are angles, cf. Section 2. For each such weak link angle, a weighted averaged angle difference d is calculated as follows. For each matched user sequence frame, the difference of the angles of this frame with the mean μ of corresponding feature of the matched aggregated master sequence frame is calculated. Then the angle difference is multiplied by the aggregated frame

weight w. From all user frames the three largest weighted angle difference values are selected. Finally, set d to the average of these three weighted differences.

For each weak link, d is expected to be 0, but a deviation $d_0 > 0$ may still be ignorable. Offline and once for each weak link, we define d_0 along with an angle d_1 showing a clear deviation from the expectation. These threshold angles are different for the different weak links. For each actual angle difference d, a weak link indicator $wi(d)$ is computed as follows: d is linearly mapped to the indicator $wi(d)$ such that an indicator of 0 corresponds to the ignorable deviation threshold d_0 and an indicator of 1 corresponds to a clear deviation threshold d_1: $wi(d) = \frac{d - d_0}{d_1 - d_0}$. Then $wi(d)$ is updated such that negative values are set to 0. Then $wi(d)$ is softened: $wi(d) := wi(d)^2$ if $wi(d) < 1$ and $wi(d) := \sqrt{wi(d)}$, otherwise. Finally, $wi(d)$ is updated such that values above 1.5 are cut off.

The AIMO score AS is a weighted sum of the weak link indicators. We pick the maximum of the left-right and the in-outwards symmetric indicators. For the weak link with a high weight factor recommendation (heels rise) the weight is 4. For each weak link with a medium weight factor recommendation (knees move inwards or outwards, excessive forward lean, asymmetrical weight shift) the weight is 2. For each weak link with a low weight factor recommendation (arms fall forward, forward head, shoulder elevation) apply a weight of 1. Let WI be the weighted sum of the weak link indicators wi. Then $AS = \max(10 - WI, 0)$.

4 Evaluation

For assessing the agreement of the AIMO and the NASM scoring methods, we let human NASM experts score a sample of 81 user sequences videos. The sample is not representative for the whole population. We evaluated the AIMO score against the NASM score and analyzed both the correlation of the two score variables, cf. Fig. 1 (left), and their differences (right).

The Pearson correlation coefficient is $r = 0.86$ indicating a high correlation of the two variables. However, Altman and Bland argue that a high correlation is necessary but not sufficient for a good agreement between two methods and suggest an additional analysis of differences [1]. The mean of the differences of the two methods is 0.63 score points at a confidence level of $p = 0.01$, i.e., the AIMO score systematically gives marginally higher scores, which can be neglected. However, 95% of the differences between the two methods are between 4.5 and -3.3 score points, which is quite a high difference interval.

Analyzing the root cause of the differences of the two methods, we found that inconsistencies in the NASM expert scoring was a major factor, especially, inconsistencies between different experts and between the assessment of the first user sequences compared to later assessments of one and the same expert. In order to exclude these human errors, we ask the most qualified expert to carefully reassess the user sequences and to put extra effort on scoring all exercises consistently. We limited this assessment to one uninterrupted session in order to reduce the differences between the first and the last assessments, which reduced the sample to 34 user sequence videos. The expert was neither exposed to the
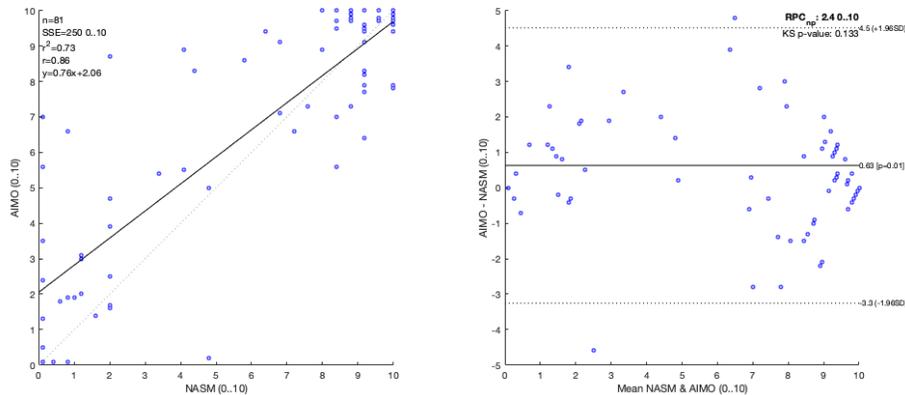
Fig. 1. AIMO vs NASM deep squat scores

AIMO scores nor to his earlier NASM scores nor was he given any further instructions to change the NASM scoring method. For this reassessed sample, the Pearson correlation coefficient increases to $r = \mathbf{0.96}$. The average of the differences of the two methods almost remains the same at **0.64**; the 95% boundaries of differences reduce considerably and are now **0.99** and $-\mathbf{2.3}$ score points.

5 Conclusion and Future Work

The paper describes and evaluates the digitalization of a standardized deep squat assessment. It shows that an automated assessment based on commodity 3D sensor technology is both effective and efficient. It is therefore applicable in systems providing inexpensive and objective decision-support for the assessment of musculoskeletal insufficiencies. In order to achieve this, the paper introduces a number of preprocessing steps for extracting the features and for dealing with their high statistical variance due to shortcomings of the 3D sensors. Preprocessing includes also a novel approach to the alignment of general time series capable of ignoring leading, intermediate and trailing subsequences.

The assessment method was validated against a ground truth from human experts, but validation needs to be strengthened. Larger and representative samples need to be selected. In order to reduce threads to validity due to subjective expert scoring, experts need to be trained in the NASM standard such that subjective inconsistencies can largely be excluded. It is an open research question if this desired assessment stability of human experts can ever be achieved.

Finally, the approach should be generalized to assess other human movements exposing, e.g., specific musculoskeletal issues once a weak link was localized, and to incorporate multimodal sensor technologies, e.g., for pain assessment as an additional information source for more accurate decision support. Therefore, the dependency of the approach to deep squat exercise specifics needs to be reduced.

References

1. Altman, D.G., Bland, J.M.: Measurement in medicine: the analysis of method comparison studies. The Statistician **32** (1983)
2. Baum, L.E., Petrie, T.: Statistical inference for probabilistic functions of finite state markov chains. Ann. Math. Stat. **37**(6) (1966)
3. Elgammal, A.M., Lee, C.S.: The role of manifold learning in human motion analysis. In: Human Motion – Understanding, Modelling, Capture, and Animation. Springer (2008)
4. Elton, J., O'Riordan, A.: Healthcare Disrupted: Next Generation Business Models and Strategies. Wiley (2016)
5. Jiang, M., Kong, J., Bebis, G., Huo, H.: Informative joints based human action recognition using skeleton contexts. Signal Processing: Image Communication **33** (2015)
6. Khan, N.M., Lin, S., Guan, L., Guo, B.: A visual evaluation framework for in-home physical rehabilitation. In: IEEE Int. Symp. Multimedia (2014)
7. Lu, C., Mandal, M.: A robust technique for motion-based video sequences temporal alignment. IEEE Trans. Multimedia **15**(1) (2013)
8. Martins, R.M., Kerren, A.: Efficient dynamic time warping for big data streams. In: IEEE Int. Conf. Big Data: Workshop Real-time and Stream Analytics in Big Data and Stream Data Management (2018)
9. National Academy of Sports Medicine: NASM Essentials of Personal Fitness Training. NASM, 6th edn. (2017)
10. Paiement, A., Tao, L., Camplani, M., Hannuna, S., Damen, D., Mirmehdi, M.: Online quality assessment of human motion from skeleton data. In: British Machine Vision Conference. BMVA Press (2014)
11. Pazhoumand-Dar, H., Lam, C.P., Masek, M.: Joint movement similarities for robust 3d action recognition using skeletal data. Visual Communication and Image Representation **30** (2015)
12. Pirsiavash, H., Vondrick, C., Torralba, A.: Assessing the quality of actions. In: Euro. Conf. Computer Vision. Springer (2014)
13. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. Acoustics, Speech, and Signal Processing **26**(1) (1978)
14. Slama, R., Wannous, H., Daoudi, M., Srivastava, A.: Accurate 3d action recognition using learning on the grassmann manifold. Pattern Recognition **48**(2) (2015)
15. Tao, L., Paiement, A., Damen, D., Mirmehdi, M., Hannuna, S., Camplani, M., Burghardt, T., Craddock, I.: A comparative study of pose representation and dynamics modelling for online motion quality assessment. Computer Vision and Image Understanding **148** (2016)
16. Vemulapalli, R., Arrate, F., Chellappa, R.: Human action recognition by representing 3d skeletons as points in a lie group. In: IEEE Conf. Computer Vision and Pattern Recognition (2014)
17. Wang, J., Zheng, H.: View-robust action recognition based on temporal self-similarities and dynamic time warping. In: IEEE Int. Conf. Computer Science and Automation Engineering (CSAE). vol. 2, (2012)
18. Wang, J., Liu, Z., Wu, Y., Yuan, J.: Mining actionlet ensemble for action recognition with depth cameras. In: IEEE Conf. Computer Vision and Pattern Recognition (2012)
19. Wang, R., Medioni, G., Winstein, C.J., Blanco, C.: Home monitoring musculo-skeletal disorders with a single 3d sensor. In: IEEE Conf. on Computer Vision and Pattern Recognition (Workshops). (2013)
20. Xia, L., Chen, C.C., Aggarwal, J.K.: View invariant human action recognition using histograms of 3d joints. In: CVPR Workshops (2012)
21. Yao, A., Gall, J., Van Gool, L.: Coupled action recognition and pose estimation from multiple views. Int. J. Computer Vision **100**(1) (2012)