International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2024)
S. Hasegawa, N. Sakata and V. Sundstedt (Editors)

Analysis of Tennis Forehand Technique using Machine Learning

P. Kán[1], G. Gerstweiler[2], A. Sebernegg[2], and H. Kaufmann[1]

[1]Institute of Visual Computing and Human-Centered Technology, TU Wien, Austria
[2]VR Motion Learning GmbH & Co KG, Austria

Figure 1: *Visual feedback of the trainee's motion analysis by our method. The left panel shows correctness of performing individual training rules and the right window plays a 3D replay of the captured trainee's motion (blue) in comparison to a professional motion (green). These two motions are temporally aligned using dynamic time warping. When a user selects a specific training rule from the left window, a textual explanation of this rule is shown to help the user improve this specific aspect of the technique.*

Abstract

Analysis of human motion is instrumental in many areas including sports, arts, and rehabilitation. This paper presents a novel method for human motion analysis with the focus on tennis training and forehand technique assessment. We address the problems of automatic motion analysis and incorrect technique identification by a machine learning approach. We utilize the concept of training rules that are used to individually assess specific aspects of a given type of motion. Our method for motion analysis is based on insights from professional trainers and our training rules are co-designed with them. The presented method is evaluated quantitatively using recorded dataset of tennis forehand motions. This evaluation compares two variants of sport technique correctness classification: informed and uninformed learning. Both learning variants fall into the category of supervised learning, but informed learning additionally utilizes motion features and motion phases derived from tennis training methodology. Our experiments suggest that informed learning leads to higher accuracy and faster speed of the algorithm. Finally, we studied our method in a qualitative expert study.

CCS Concepts

• ***Computing methodologies*** → ***Motion processing;*** *Machine learning approaches;* • ***Human-centered computing*** → *HCI;*

1. Introduction

Knowledge of correct sport technique plays an important role for injury prevention and performance improvement in sports both for beginners and professionals. While the best way to practice sport technique is with a professional trainer on-site, this option is often inaccessible due to the time and cost constraints. Therefore, methods for automatic assessment of sport technique and guidance have been emerging in recent years [MAKD17, HGH*18, SNT18, AKF*24]. While these methods enable computer-aided motion analysis, they often focus on motion type classification and do not analyze if the motion is performed biomechanically correct with respect to the technique of a given sport and with relation to its training methodology.

DOI: 10.2312/egve.20241363

We address the task of motion analysis, for the case of tennis forehand technique, by a data-driven approach utilizing machine learning and apriori information from trainers. This apriori information consists of the definition of training rules and motion phases for a tennis forehand stroke. Furthermore, specific motion features such as distance ratios are created in addition to positions, rotations, and velocities to augment the motion data with richer information for technique correctness classification.

While certain sport rules can be defined by simple thresholds on motion features for specific body joints in specific motion phases, some advanced rules require a very complex motion model including many joints. Therefore, we introduce rules that can be evaluated in a data-driven fashion in comparison to a reference motion, performed by a professional. Our methodology contains both types of rules: stand alone rules and comparison rules.

The problem of player motion comparison with a reference motion is the requirement of perfect alignment of these two motions in time. We address this problem by utilizing a dynamic time warping algorithm [SC78]. This way, the optimal mapping between frames of player's motion and reference motion is found. Once the two motions are aligned in time, the motion phases can be identified and motion features can be extracted for technique correctness assessment by machine learning classification.

We demonstrate the applicability of the proposed methodology in a tennis training scenario. Training rules for tennis motion assessment were derived with insights from professional trainers. A dataset of tennis motion capture data was recorded using an optical tracking system. This dataset was split to a training set and a test set of motions while test motions were only used to calculate the accuracy of classification. We compared two learning approaches, informed and uninformed learning. Informed learning used motion phases and motion features suggested by professional trainers while uniformed learning utilized all available motion features on the whole motion duration. Both learning approaches belong to the category of supervised learning because the ground truth classification of each motion was known during training. This investigation seeks for an answer to an interesting research question, whether additional information from sport methodology helps to improve accuracy or speed of motion assessment. Our results suggest that informed learning achieves higher accuracy and speed than uninformed learning. Finally, we studied our method in a qualitative expert study to reveal insights of professional tennis trainers and to investigate potential application in virtual reality (VR) training.

The proposed methodology is applicable in forehand performance assessment and computer-aided tennis training. The results of our investigation indicate that knowledge of sport technique from professional trainers brings invaluable insights into the machine learning for motion data classification.

2. Related Work

Large body of research has been conducted on the topics of human motion analysis and training support systems. Serious games have been presented to increase motivation for physical exercises and training [GGdNGRNA18]. Additionally, various approaches for guiding users during sports were presented [OSC22, OIMK18, IHK18]. Previous methods focused on different sub-problems in computer-aided sport training, including motion analysis, motion classification and temporal alignment.

2.1. Human Motion Analysis

Analysis of human motion both in 2D video sequences and 3D motion sequences plays an important role in numerous application areas. Therefore, many methods were presented in the past to address this problem. A rule-based system for analysis of athlete motions in the area of body building was presented by Örücü and Selek [OS20]. This system assessed the correctness of two bodybuilding exercises using 12 rules. A rule-based approach was also examined and shown to be useful in case of rehabilitation exercises [ZREL17]. We also utilize the concept of training rules and we additionally combine it with a machine learning approach to classify correctness of execution of every rule.

Several methods focused on analysis of tennis motion for the purposes of training and motion learning. Oshita et al. [OIMK18] captured motion of a trainee and visualized its differences from an expert motion. These differences were calculated using spatial, rotational, and temporal motion features. Morel et al. [MAKD17] proposed a methodology for calculation of spatial and temporal errors between trainee's and expert's motion. The authors used the dynamic time warping (DTW) algorithm [SC78] to achieve temporal alignment between two motions and then calculate the error between them. They evaluated their results on tennis serve and karate motions. Our method is also investigated with tennis motions. Similarly to Morel et al., we utilize DTW to achieve temporal alignment between two motions but we additionally use machine learning to train motion error classifiers from the labelled data.

Previous research studied motion analysis for various other human motions including analysis of rowing technique [BHS*19], salsa dance analysis related to music [SNT18], golf swing analysis using neural networks [LHK22], Tai Chi motion analysis by classification of motion errors [HGH*18], and analysis of rehabilitation exercises [RWHH19]. The latter compared a rule-based approach and a machine learning approach. We utilize a combination of rules and machine learning to analyze fast motion data in an automatic fashion while taking sport methodology into account.

Sport learning in VR or augmented reality (AR) setups was also previously addressed. Farzinnejad et al. [FRKG23] studied the impact of practicing table tennis in AR exergame. Their results suggest significant improvement of stroke skills when using the exergame in comparison to a traditional training. The authors did not use the methodology of motion analysis and consequent feedback during training. AR was also used for 3D visualization in golf training [IHK18]. Oagaz et al. [OSC22] proposed a VR system for training table tennis with multi-modal feedback on the trainee's posture. Additionally, Saito et al. used a VR environment to analyze differences between motions of experts and novices in tennis [SMS*18]. The results indicated differences in preparation phase, leg movement, take-back returns and degree of spine twist between two groups of participants with distinct skill levels. While our system provides on-screen feedback, it can be extended in future to a VR scenario to increase immersion. A survey of research methods using virtual environments for sports training was presented by Miles et al. [MPW*12]. Bloomfield et al. [BPO04] proposed a taxonomy for time-motion analysis with unified descriptions of directions, intensities, turning categories, and other properties. This taxonomy can be used in manual and automated analysis.

2.2. Motion Classification by Machine Learning

Research in the fields of computer vision and signal processing focused on the task of motion classification to a given set of classes. Such classification can be used in sport motion analysis to identify types of errors according to sport theory or to identify specific part of the motion in time. Komori et al. [KIM*23] presented a method for sport motion classification using motion history image. They suggested using time-weighting motion history image and analysed its performance on the task of classification of tennis shot direction. Zhang et al. [ZYL*18] utilized long short-term memory network for classification of tennis motions from vision data and inertial motion data. Their system classified tennis motion into five different action classes: forehand topspin, forehand slice, backhand topspin, backhand slice, and serve. Swing type classification for racket sports, based on convolutional neural network, was proposed by Anand et al. [ASS*17]. Sport activity classification based on convolutional neural network, using inertial motion data, was also discussed by Hsu et al. [HCC19]. Additionally, training activity classification based on discrete wavelet transform and random forest classifier was presented by Ahmadi et al. [AMR*15]. The authors used motion data from inertial sensors to initially classify the type of training activity and then identify potential injuries. Support vector machines (SVMs) and hidden Markov model were used in the past research to classify full-body motion capture data into various actions from a set of military maneuvers [SS05]. A survey of action recognition methods from sport videos was presented by Wu et al. [WWB*23]. Our method also classifies motion performance of a trainee as either correct or incorrect with respect to each specific training rule using SVMs. The connection of classifiers with the concept of training rules enables individual assessment of every aspect of player's technique for a defined motion.

2.3. Temporal Alignment

Temporal alignment of two motions is typically required if a part of motion assessment compares trainee's motion with pre-recorded reference motion of a professional or a trainer. Such an alignment can be reached by finding a non-linear mapping between the frames of two motions sequences. Many previous methods are based on the DTW algorithm [SC78] that uses dynamic programming to efficiently calculate such mapping. An important parameter of this algorithm is a distance function between two frames of the motion (i.e. two body poses in the case of human motion alignment). Numerous features can be used to calculate this distance, including joint angles, positions, velocities, accelerations, etc. Several improvements of DTW were proposed to increase the speed of the calculation, including limitation of warping range (i.e. each frame can be only warped to its neighborhood frames in a given radius) [Ita75, SC78], iterative multi-scale calculation with coarse-to-fine refinement [ZM06], and using piecewise aggregate approximation to downscale input resolution and thus reduce complexity [KP00]. We also utilize DTW to perform temporal alignment of trainee's motion with reference tennis forehand motion of a professional player. This alignment not only enables direct comparison of body poses across the whole motion sequence, but it also enables correct subdivision of trainee's motion to the motion phases according to the sport methodology (Section 3.2.1).

3. Tennis Motion Analysis

Our proposed method for tennis forehand analysis is centered around the concept of training rules. Training rules are specific subsets of sport motion methodology that focus on individual aspects of a motion (e.g. tennis players should stand in the basic position before the swing, having their arms and elbows in front of their body). Subdivision of sport motion methodology to individual training rules provides good granularity for assessing each individual aspect of the motion separately. Therefore, our method evaluates individual training rules in an automated fashion given the defined set of motion features and motion parameters for each rule.

Our method assesses the player's motion performance in a post hoc fashion and it consists of the following steps: The initial step includes temporal alignment of the trainee's recorded motion with the reference motion of a professional player. This alignment is needed because some rules evaluate the correctness of motion by comparison to reference motion. Temporal alignment results in warped frames of trainee's motion in a way that they closely match the frames of reference motion. In the second step, each motion rule is evaluated individually by calculating motion distances using its defined motion features. This step is described more in detail in Section 3.3. In the third step, the results of the rules evaluation are fed into machine learning classifiers to perform binary classification of each rule to be played either correctly or incorrectly (Section 3.4). We use Support Vector Machines (SVMs) [Vap00] for this classification and we pre-train them on the training set of labeled tennis forehand data. Finally, once the motion analysis is calculated, feedback for the user about their performance should be displayed. We use two types of feedback (Figure 1): (1) summary of motion rules and their correctness assessment, and (2) synchronized playback of two virtual avatars in 3D, one representing the reference motion and one the actual trainee's motion. When a user clicks on assessment of any motion rule, its description is displayed to indicate what can be improved in next trials.

3.1. 3D Motion Data Acquisition

Our dataset consists of Motion Capture (MoCap) data of correct and incorrect forehand groundstrokes performed by a tennis coach, that serve as ground-truth data for our method. For the recording of incorrect forehand strokes, the subject was instructed to perform the motion while deliberately violating a specific coaching rule. As an illustration, the subject was asked to violate rule *BP3*, described in Table 1, by keeping the knees extended or overly bent in the basic position. The recorded data was then labeled respectively. The motion data was captured in an indoor tennis court using SIMI Motion[†], an optical, markerless 3D motion capture system. The

[†] http://www.simi.com/

motions were recorded at the frame rate of 200 frames per second using a setup with 8 cameras. The multi-camera setup requires calibration before recording and manual postprocessing in the software module *Simi Shape 3D* to export the captured motion data.

The resulting dataset contains 104 recorded movements that are stored in the BVH (BioVision Hierarchy) file format, of which 26 represent correct forehand strokes and 78 various coaching rule violations. One of the correct motions was used as a reference motion of professional player for all the calculations. The remaining 103 motions were divided into a training set of 72 motions and a test set of 31 motions. The skeletal model, depicted in Figure 2, represents the initial pose and hierarchical structure of joints within our MoCap data. Each joint's 3D rotation and translation relative to its parent joint within the hierarchy is encoded as a time series in the motion file format. Additionally, the number of frames and the sampling rate of the motion are stored. This information can be used to derive other motion features, such as global joint data, average velocities, ratios of distances to the center of mass, height differences, and joint angles. Figure 2 right illustrates examples of motion features that are calculated in our method (Section 3.3.1).

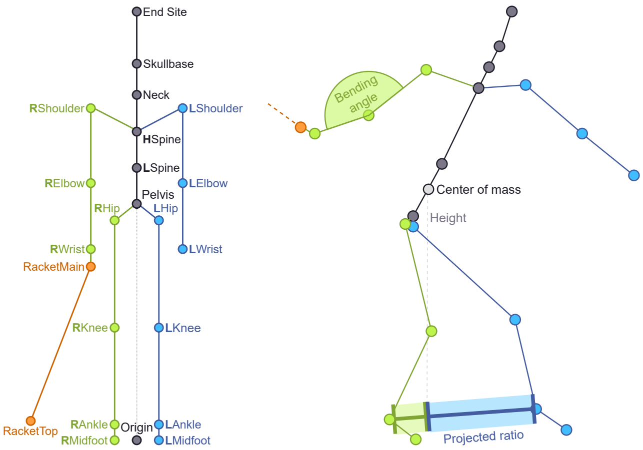

Figure 2: *Skeleton model and motion features. The left skeleton model represents the initial pose and joint hierarchy of our MoCap data. The skeleton model on the right depicts a pose in a tennis forehand motion and illustrates motion features.*

3.2. Temporal Alignment

We use DTW to perform non-linear temporal alignment of actual trainee's motion with the reference motion of the professional. Utilizing this algorithm, the frames of trainee's motion are warped to match the frames of reference motion with the minimum pose distance between them. The monotonicity of the warping is preserved to preserve the order of frames in the trainee's motion. Velocity data is calculated before time warping and it is assigned to the specific frames to avoid distortion by the non-linear frame shifts.

During the training of machine learning classifiers, we used an optimized DTW implementation by constraining the maximum warping range to 600 frames because DTW needs to be performed for each training sample. The selected maximum warping range was set empirically. This constraint was not used during trainee's motion assessment because DTW needs to be calculated only once per each new motion.

3.2.1. Automatic Detection of Motion Phases

Tennis forehand motion can be subdivided to the defined motion phases according to tennis methodology. Our method enables reliable identification of these motion phases and only requires their accurate labeling in the reference motion. Once the start and end frames of each motion phase are correctly labeled for the reference motion, these frames can be easily identified also in any new motion of a trainee after warping its frames to match the frames of reference motion by DTW [SC78]. Thus, the DTW warping path is used to find a mapping of the start and end frame of each motion phase between known reference motion and unknown trainee's motion and identify motion phases in the latter.

In tennis, separate motion phases are defined for upper and lower body. The phases of upper body motion as they appear in the right forehand swing are: Basic position, swing out phase, strike phase, hit, follow through. The motion phases for the lower body motion are: Basic position, split step, movement towards the ball, hit, hit pose. The phases for lower and upper body have distinct start points and durations.

3.3. Training Rules

Training rules are the core of our method for sport motion analysis because every rule represents a specific part of the sport motion methodology that can be evaluated individually using specific motion features in specific phase of the motion. In order to enable high representativeness of training rules, each rule has the following properties defined in addition to its name and description: (1) Motion features to be used for evaluation of trainee's motion according to this rule, (2) set of biases, one for each feature, to be subtracted from the feature distance to influence its threshold, (3) set of body joints to be involved in motion assessment by this rule, (4) set of binary indicators if the motion features should be evaluated only in the first frame of the motion phase, or aggregated during the whole motion phase, and (5) motion phases in which the rule should be evaluated. For evaluation of each rule, only the motion phases that are indicated in this rule are used to calculate a pose distance vector in feature space.

3.3.1. Motion Features

Every motion assessment rule can use a set of motion features to calculate a distance vector (from the ideal motion) that is used to classify motion correctness. Each feature calculates this distance vector either by comparison to the reference motion or by comparison to a predefined value (angular, spatial, or other). In the investigated use-case of tennis forehand, we used the following motion features for the calculation of pose distances:

- **Global positions**: This feature calculates distance between two body poses in a given frame by calculating spatial distance of each specified joint to its counterpart in the reference motion in the global coordinate frame. The global coordinate frames of two compared motions are aligned by the position of torso to avoid bias introduced by the spatial offset of the player.

- **Global rotations**: Calculate the angular distances between rotation of each selected bone and its counterpart in the reference motion in the global coordinate frame. We represent two rotations as quaternions and this angular distance is a scalar angle between them.
- **Local rotations**: They also represent angular distances between two joint rotations of trainee's pose and reference pose in each time frame of the motion phase. However, for local rotations, the local coordinate frame of the parent joint is considered. E.g. if the angle of the wrist join with respect to elbow is the same between two motions, no matter the global arm rotations, the distance will be 0 because the rotations are the same with respect to their parent joint.
- **Velocity magnitudes**: Distance between two velocity magnitudes of two motions for a given joint is calculated as an absolute value of a scalar difference between them.
- **Height**: This feature is a scalar value that enables distance calculation with the focus on the height of a specific body joint in comparison to a predefined threshold (set as a bias). This feature is calculated only on the trainee's motion and it does not compare two motions. It is used for specific use-cases, e.g. for detection of small jumps by monitoring height of player's ankles.
- **Height comparison of two joints**: This pose distance feature evaluates only trainee's motion without comparison to reference motion. It focuses on comparing heights of two joints. The distance between heights is calculated as their difference. Such a feature can be used for direct evaluation of the player's pose, e.g. if the wrist joint is higher than elbow.
- **Projected ratio of joint distances to the center of mass**: This feature helps with assessment of correct weight distribution between two feet in a given frame. It calculates the ratio of distances of two selected joints (typically two ankles) to the center of mass in the horizontal plane. A distance of this calculated ratio to a defined constant ratio is calculated by an absolute value of their difference. For the tennis use-case, we typically used ratio of 80% : 20%. This feature calculates pose distance from the defined ideal weight distribution ratio and, therefore, does not need reference motion. Projection of center of mass in the horizontal plane was calculated by projecting the lower spine joint to this plane.
- **Comparison of projected ratios of joint distances to the center of mass with the reference motion**: It is similar to previous feature but it compares the calculated weight distribution ratio of the trainee's motion to the reference motion for each defined frame of the rule's motion phase.
- **Bending angle of a joint**: Calculates the angle of a given joint with its parent joint. This evaluates the amount of joint bending and it is used without comparison to the reference motion.
- **Angle of joint with vertical axis**: This feature represents a scalar angular distance of the current joint rotation to the vertical axis of a coordinate system. This feature only requires trainee's motion.
- **Frontal distance of a joint in the coordinate space of upper spine**: This features evaluates trainee's motion by calculating distance of a given joint from the upper spine joint in a frontal direction. This is achieved by first transforming the joint position into the coordinate space of the upper spine and then reading the coordinate along the front axis. This feature does not require the reference motion.

3.3.2. Motion Distance and Pose Distance Calculation

Pose distance (i.e. difference between a desired pose and a recorded pose) can be calculated using the above-mentioned features for each frame of the given motion phase. However, the defined training rules require assessment of the motion sequences rather than individual frames. Therefore, we calculate aggregated feature distance values instead of values of individual frames to be able to later assess each rule and classify its correctness. For each defined motion feature in each training rule, we calculate the maximum, minimum and average values of given feature distances across defined motion phases. This data then forms a feature vector that is used as an input to our SVM classifier. Some rules define the motion distance only in one frame (e.g. start frame of a given motion phase) and for these rules the feature values of this frame are used.

3.3.3. Tennis Motion Rules

Using the above-described motion features and motion phases, we created a set of tennis training rules. These rules were created together with professional tennis trainers and they are based on training methodology of tennis technique. These defined rules are used in our method for automated assessment of tennis motion. The designed rules and their respective motion phases and motion distance features are described in Table 1. These rules were designed for the forehand swing motion analysis of the right-handed players.

3.4. Classification

Once the aggregated motion feature distances have been calculated for all training rules, we classify execution of each rule in a given motion either as correct or as incorrect. We employ Support Vector Machines (SVM)s for this task. We trained an SVM classifier for each motion rule on our dataset of labelled tennis motions. Our SVM classifiers use Gaussian kernel [SSB*97] to enable nonlinear classification. We also used data augmentation to increase the size of the training set of motions and to increase data variability. For each motion in the training set we generated two additional motions by augmentation. The augmentation was done by perturbing each local joint rotation by a random angle in a range of +-10° in each of the three Euler axes. As we wanted to avoid jittery motion caused by random perturbations of joint rotations, we smoothed the random angular offset by using 98% of angular perturbation from last frame and only 2% of new random angular perturbation. In addition to the angular offset, we augmented each motion also by temporal perturbation of each frame's position in time (with preserving the order of frames). Finally, the results of rule correctness classification are displayed to a user using visual feedback (Figure 1).

4. Evaluation and Results

4.1. Comparison of Informed and Uninformed Classification

In our research, we investigated whether the addition of training rules, that inform about utilization of specific motion features in defined motion phases, increases accuracy of classification of labeled

Table 1: *Training rules, used in our tennis forehand motion assessment. Each rule contains its name, description, list of joints involved in the assessment, list of features to be used for distance calculation from optimal motion, motion phases in which the assessment should be calculated, and an indicator if only the first frame of the motion phase is used instead of whole motion phase (FF Only).*

Rule	Description	Involved Joints	Distance Features	Motion Phases	FF Only
BP1	Arms and elbows are in front of the body.	Elbow right, Elbow left, Wrist right, Wrist left	Frontal join distance with respect to upper spine	Upper body basic position	*X*
BP3	Knees are in a slightly bent position.	Knee left, Knee right	Local rotations	Lower body basic position	*X*
BP8	Hip and shoulder axes are parallel.	Spine low, Spine high, Pelvis, Shoulder left, Shoulder right	Global rotations, Global positions	Lower body basic position, Upper body basic position	*X*
H3	In the point of racket hit with the ball, the arm should not be fully extended (i.e. the elbow joint should not be straight).	Elbow right	Bending angle of a joint	Lower body hit pose	✓
H4	The right foot is positioned in the back and the body weight distribution is 80% on the left foot and 20% on the right foot.	Ankle left, Ankle right	Projected ratio of joint distances to the center of mass	Lower body hit pose	✓
H7	Upper body and head position should be upright during the hit.	Skullbase, Neck, Spine high, Spine low, Torso	Angle of each joint with vertical axis	Lower body hit pose	✓
SP2	At the beginning of "the movement towards the ball" phase, the weight is shifted to the right foot if possible. The ratio of 80% : 20% is again used but this time towards the right foot.	Ankle right, Ankle left	Projected ratio of joint distances to the center of mass	Lower body movement towards ball	✓
SP10	In the swing out phase the racket head is always above the wrist.	Racket top, Racket main	Height comparison of two joints	Upper body swing out phase	*X*
SP12	Immediately before the forward movement in the strike phase, the right-handed person puts weight on the right foot and pushes off to the left foot for a slightly larger stride.	Ankle right, Ankle left	Comparison of projected ratios of joint distances to center of mass with the reference motion	Upper body strike phase	*X*
SS1	The knees are in a slightly bent position again after finishing the split step, at least never straightened.	Knee left, Knee right	Local rotations	Lower body split step	*X*
SS2	The split step should be executed as flat as possible. There should be just a slight elevation of player's position but not a real jump.	Ankle left, Ankle right	Height	Lower body split step	*X*
UT4	Elbows should be in one plane with the upper body during a unit turn. Upper body should not be actively moved upwards but by "centrifugal force"	Shoulder right, Spine high	Local rotations, Velocity magnitudes	Upper body swing out phase, upper body strike phase	*X*

data. For this purpose we compared two conditions: (1) Informed learning that contains a defined set of features, joints and motion phases per each rule (Section 3.3), and (2) Uninformed learning that uses all above-described motion features across all frames for all body joints. We hypothesize that informed learning achieves higher accuracy in comparison to uninformed learning due to better guidance of motion distance calculation. Additionally, we expected informed learning to also achieve higher speed due to the smaller number of frames to be evaluated.

We trained an SVM classifier for each rule separately for informed and uninformed learning using our training set of motions. We then evaluated the trained classifiers on the test set separately for informed and uninformed approach. The results of this evaluation can be seen in Figure 3. The test accuracy of informed learning was higher for all training rules except the rule UT4. This result suggests that the added information about tennis technique from trainers has a positive impact on the performance of classification.

4.2. Computational Time

We evaluated computational time of training of our SVM classifiers as well as runtime of user's motion assessment. We also compared training times and motion assessment times between informed and uninformed learning. The training of SVM classifier for one rule took in average 1 minute and 20 seconds for informed learning and 1 minute and 47 seconds for uninformed learning. The total training time for training all 12 SVM classifiers was 16 minutes and 5 seconds for informed learning and 21 minutes and 22 seconds for uninformed learning. The main reason for difference in training time between informed and uninformed learning was the fact that informed learning used only a selected subset of features in a selected range of frames for calculation of motion distance while uninformed learning needed to calculate all features for all frames during training of each rule. Eventually, the motion features could be precalculated for both approaches trading time complexity for memory complexity.

An important aspect of sport motion analysis for training is also its runtime speed. Therefore, the time of motion analysis for new recorded motion was evaluated for both informed and uninformed machine learning approaches. Additionally, the break-down of computational time to individual steps of our method was performed. The results of this runtime analysis can be seen in Figure 4. The average duration of motion analysis with an informed approach was 4.74 seconds while motion analysis with uninformed approach took in average 11.14 seconds. Dynamic time warping algorithm during individual motion assessment did not use maximum warping range constraint. Both training of classifiers and runtime speed analysis was conducted on a laptop with Intel i7-8750H CPU, 16GB RAM, and NVIDIA GeForce GTX 1070 GPU. Our application for motion analysis was implemented in Unity game engine.

4.3. Expert Study

Additionally to accuracy measurements, we studied our method for tennis forehand motion assessment in an expert study. We used qualitative methodology to investigate trainers' views on our method regarding usability, correctness of motion assessment, applicability to training practice, motivation, potential for tennis technique improvement or injury prevention, visual feedback, possibility of integration to virtual reality, and applicability to other sports. The study was conducted in the form of a showcase of the application for tennis training using our method, followed by a semi-structured interview. Three professional tennis trainers were recruited to participate in the study.

4.3.1. Qualitative analysis

The answers of the trainers to our open questions were analyzed using open coding methodology. The resulting codes (bold font) and their descriptions are listed as follows:

- **Training vs. Match**: The trainers indicated that our motion analysis is suitable for a single type of motion that can be repeated during training, but it does not cover a wide variety of different situations that can occur during a tennis match. The motion analysis would be also highly suitable for serve because serve motion has fixed rules also during the match.
- **Automation**: The trainers highlighted the high level of automation in our system that is a positive contribution in comparison to usually used manual motion analysis (e.g. from a video).
- **Technique improvement**: The trainers expect the proposed motion analysis to lead to the improvement of trainees' tennis technique.
- **Motivation**: The motion analysis has a potential to increase motivation for some people by enabling them to observe their progress. This effect can be the strongest for the people who are starting to play tennis.
- **Individual training**: Individual training with the motion analysis without the presence of trainers was not recommended to avoid misinterpretation of parts of the methodology and to properly prioritize learning of individual rules.
- **Skill level**: The trainers commented on the appropriate skill level of players for which the proposed motion analysis would be applicable in training. They suggested to use such analysis for intermediate players because the basic technique for beginners is typically easily assessed by trainers themselves and professional technique may differ for each player.
- **Visual feedback**: If two players are shown in parallel, they should not overlap. While the visualization of rules correctness was rated positively, the trainers also noticed tracking errors visible in the replay animation. Real-time visualization of motion phases during replay was appreciated.
- **Mobile application**: It would be beneficial for usability if the motion analysis could be deployed to a mobile phone (including tracking).
- **Virtual reality integration**: Trainers considered VR integration of motion analysis as not prospective for the training, because contact with the real world is necessary during training. That includes both visual sense and perception of forces that affect the trainee's body and the racket. Additionally, tennis players might prefer to rather play the real tennis than a VR tennis. One trainer sees a potential in remote VR tennis coaching that would enable trainee and a coach to be in the same virtual environment for remote feedback and virtual replay. Personalized avatar could be used in VR for better relation of a player with the avatar.

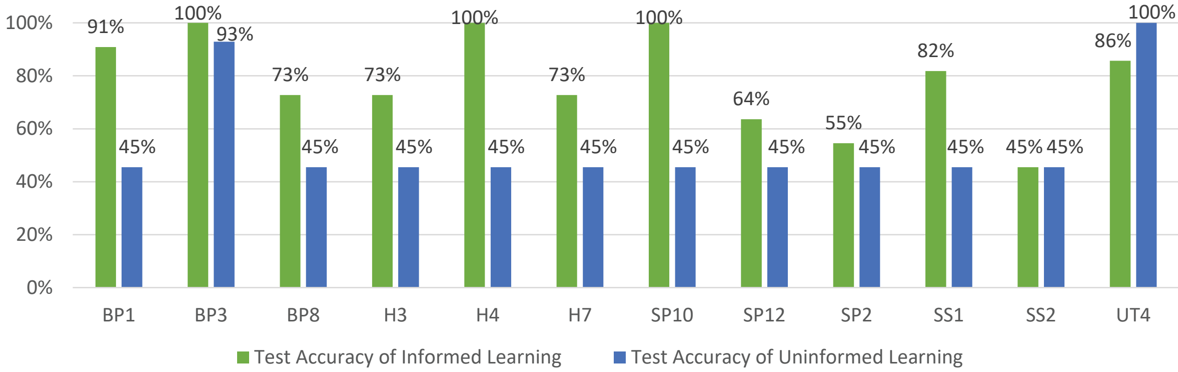

Figure 3: *Test accuracy of trained classifiers for each motion rule. Accuracy of informed and uninformed learning is compared.*

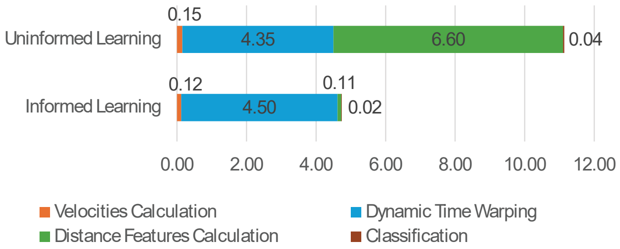

Figure 4: *Analysis of computational time of our method. The time is stated in seconds. The bars show stacked duration values of individual steps of our pipeline. Motion assessment took 4.74 seconds for informed and 11.14 seconds for uninformed method.*

5. Discussion

In our research we hypothesized that informed learning achieves higher accuracy and faster speed than uninformed learning. Our hypotheses were partially supported by the accuracy, measured on the test set of motions (Sec. 4.1), and by the comparison of computational time between the two approaches (Sec. 4.2). Only one rule (UT4) out of 12 achieved higher accuracy with uninformed learning and the informed learning dominated in the remaining 11 rules. After detailed analysis of UT4, we discovered that extending the informed approach with additional joints (elbows, ankles, and racket) helped to achieve the same accuracy as with uninformed approach.

Our expert study reveals the opinions of professional tennis trainers on our method. In summary, the trainers perceived the motion analysis positively, with the rules being correctly evaluated by the system and the application providing a potential for improvement of players' tennis technique. A drawback of integration of the proposed methodology to training practice was seen to be the lack of time for computer-aided training during the training session. On the other hand, the motion analysis was identified as prospective for increasing motivation of players, specifically at the beginner's level. The trainers also mentioned the high applicability of the proposed methodology to other sports including athletics, golf, baseball, basketball, and others.

5.1. Limitations and Future Work

While our proposed method achieved high accuracy of motion error classification in tennis technique and was supported by professional tennis trainers, it still contains many open challenges and avenues for future research. One of the limitations of our accuracy investigation was the limited set of labeled motions that all belong to one tennis player. In future, our method can be extended by containing multiple reference motions and by selecting the one that has the closest match to the physiological properties of a trainee.

Our method utilizes data from motion distance calculation as an input to SVM classifiers in an aggregated form. While this setup enables fixed-sized input, that is independent of the frame count, it significantly lowers dimensionality of the input data and thereby also restricts its representativeness. A possible avenue for future research is to utilize more information from distance calculation.

We believe that in addition to tennis, the proposed methodology can be used in many other sport disciplines. Validation on motion data from these disciplines remain as future work. If our method is applied to other sport, the feature set needs to be extended to cover specific aspects of the desired motion. Finally, more tennis rules and stroke types can be analyzed in future work.

6. Conclusion

This paper has presented a novel method for automated analysis of tennis technique, that combines the concept of training rules, temporal alignment by DTW, and machine learning classification. We demonstrated the applicability of the presented method on the use case of tennis forehand training. The database of tennis motions was captured and used for training SVM classifiers that can predict motion correctness for each training rule on a new recorded motion. We also compared the machine learning approach using apriory information from professional trainers (informed learning) to the machine learning without this information (uninformed learning). Our results suggest that informed learning achieves higher accuracy of classification and faster speed. We conducted an expert study to investigate trainers' judgement of our method. The results reveal insights about applicability of our method to practice and highlight trainers' positive view on the automated motion analysis.

References

[AKF*24] Adolf J., Kán P., Feuchtner T., Adolfová B., Doležal J., Lhotská L.: Offistretch: camera-based real-time feedback for daily stretching exercises. *The Visual Computer* (May 2024). doi:10.1007/s00371-024-03450-y. 1

[AMR*15] Ahmadi A., Mitchell E., Richter C., Destelle F., Gowing M., O'Connor N. E., Moran K.: Toward automatic activity classification and movement assessment during a sports training session. *IEEE Internet of Things Journal 2*, 1 (2015), 23–32. doi:10.1109/JIOT.2014.2377238. 3

[ASS*17] Anand A., Sharma M., Srivastava R., Kaligounder L., Prakash D.: Wearable motion sensor based analysis of swing sports. In *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)* (2017), pp. 261–267. doi:10.1109/ICMLA.2017.0-149. 3

[BHS*19] Becker A., Herrebrøden H., Sánchez V. E. G., Nymoen K., Freitas C. M. D. S., Torresen J., Jensenius A. R.: Functional data analysis of rowing technique using motion capture data. In *Proceedings of the 6th International Conference on Movement and Computing* (New York, NY, USA, 2019), MOCO '19, Association for Computing Machinery. URL: https://doi.org/10.1145/3347122.3347135, doi:10.1145/3347122.3347135. 2

[BPO04] Bloomfield J., Polman R., O'Donoghue P.: The 'bloomfield movement classification': Motion analysis of individual players in dynamic movement sports. *International Journal of Performance Analysis in Sport 4*, 2 (2004), 20–31. URL: https://doi.org/10.1080/24748668.2004.11868300, arXiv:https://doi.org/10.1080/24748668.2004.11868300, doi:10.1080/24748668.2004.11868300. 3

[FRKG23] Farzinnejad F., Rasti J., Khezrian N., Grubert J.: The effect of an exergame on the shadow play skill based on muscle memory for young female participants: The case of forehand drive in table tennis. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (Los Alamitos, CA, USA, oct 2023), IEEE Computer Society, pp. 1152–1160. URL: https://doi.ieeecomputersociety.org/10.1109/ISMAR59233.2023.00132, doi:10.1109/ISMAR59233.2023.00132. 2

[GGdNGRNA18] González-González C., del Nazaret Gómez Río, Navarro-Adelantado V.: Exploring the benefits of using gamification and videogames for physical exercise: a review of state of art. *International Journal of Interactive Multimedia and Artificial Intelligence 5*, 2 (09/2018 2018), 46–52. URL: http://www.ijimai.org/journal/sites/default/files/files/2018/03/ijimai_5_2_6_pdf_43532.pdf, doi:10.9781/ijimai.2018.03.005. 2

[HCC19] Hsu Y.-L., Chang H.-C., Chiu Y.-J.: Wearable sport activity classification based on deep convolutional neural network. *IEEE Access 7* (2019), 170199–170212. doi:10.1109/ACCESS.2019.2955545. 3

[HGH*18] Hülsmann F., Göpfert J. P., Hammer B., Kopp S., Botsch M.: Classification of motor errors to provide real-time feedback for sports coaching in virtual reality — a case study in squats and tai chi pushes. *Computers & Graphics 76* (2018), 47–59. URL: https://www.sciencedirect.com/science/article/pii/S0097849318301304, doi:https://doi.org/10.1016/j.cag.2018.08.003. 1, 2

[IHK18] Ikeda A., Hwang D. H., Koike H.: AR based Self-sports Learning System using Decayed Dynamic TimeWarping Algorithm. In *ICAT-EGVE 2018 - International Conference on Artificial Reality and Telexistence and Eurographics Symposium on Virtual Environments* (2018), Bruder G., Yoshimoto S., Cobb S., (Eds.), The Eurographics Association. doi:10.2312/egve.20181330. 2

[Ita75] Itakura F.: Minimum prediction residual principle applied to speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing 23*, 1 (1975), 67–72. doi:10.1109/TASSP.1975.1162641. 3

[KIM*23] Komori H., Isogawa M., Mikami D., Nagai T., Aoki Y.: Time-weighted motion history image for human activity classification in sports. *Sports Engineering 26*, 1 (Oct 2023), 45. URL: https://doi.org/10.1007/s12283-023-00437-1, doi:10.1007/s12283-023-00437-1. 3

[KP00] Keogh E. J., Pazzani M. J.: Scaling up dynamic time warping for datamining applications. In *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2000), KDD '00, Association for Computing Machinery, p. 285–289. URL: https://doi.org/10.1145/347090.347153, doi:10.1145/347090.347153. 3

[LHK22] Liao C.-C., Hwang D.-H., Koike H.: Ai golf: Golf swing analysis tool for self-training. *IEEE Access 10* (2022), 106286–106295. doi:10.1109/ACCESS.2022.3210261. 2

[MAKD17] Morel M., Achard C., Kulpa R., Dubuisson S.: Automatic evaluation of sports motion: A generic computation of spatial and temporal errors. *Image and Vision Computing 64* (2017), 67–78. URL: https://www.sciencedirect.com/science/article/pii/S0262885617300951, doi:https://doi.org/10.1016/j.imavis.2017.05.008. 1, 2

[MPW*12] Miles H. C., Pop S. R., Watt S. J., Lawrence G. P., John N. W.: A review of virtual environments for training in ball sports. *Computers & Graphics 36*, 6 (2012), 714–726. 2011 Joint Symposium on Computational Aesthetics (CAe), Non-Photorealistic Animation and Rendering (NPAR), and Sketch-Based Interfaces and Modeling (SBIM). URL: https://www.sciencedirect.com/science/article/pii/S0097849312000957, doi:https://doi.org/10.1016/j.cag.2012.04.007. 3

[OIMK18] Oshita M., Inao T., Mukai T., Kuriyama S.: Self-training system for tennis shots with motion feature assessment and visualization. In *2018 International Conference on Cyberworlds (CW)* (2018), pp. 82–89. doi:10.1109/CW.2018.00025. 2

[OS20] Örücü S., Selek M.: Design and validation of rule-based expert system by using kinect v2 for real-time athlete support. *Applied Sciences 10*, 2 (2020). URL: https://www.mdpi.com/2076-3417/10/2/611, doi:10.3390/app10020611. 2

[OSC22] Oagaz H., Schoun B., Choi M.-H.: Real-time posture feedback for effective motor learning in table tennis in virtual reality. *Int. J. Hum.-Comput. Stud. 158*, C (feb 2022). URL: https://doi.org/10.1016/j.ijhcs.2021.102731, doi:10.1016/j.ijhcs.2021.102731. 2

[RWHH19] Richter J., Wiede C., Heinkel U., Hirtz G.: Motion evaluation of therapy exercises by means of skeleton normalisation, incremental dynamic time warping and machine learning: A comparison of a rule-based and a machine-learning-based approach. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2019, Volume 4: VISAPP, Prague, Czech Republic, February 25-27, 2019* (2019), Trémeau A., Farinella G. M., Braz J., (Eds.), SciTePress, pp. 497–504. URL: https://doi.org/10.5220/0007260904970504, doi:10.5220/0007260904970504. 2

[SC78] Sakoe H., Chiba S.: Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing 26*, 1 (1978), 43–49. doi:10.1109/TASSP.1978.1163055. 2, 3, 4

[SMS*18] Saito K., Masai K., Sugiura Y., Kimura T., Sugimoto M.: Development of a virtual environment for motion analysis of tennis service returns. In *Proceedings of the 1st International Workshop on Multimedia Content Analysis in Sports* (New York, NY, USA, 2018), MMSports'18, Association for Computing Machinery, p. 59–66. URL: https://doi.org/10.1145/3265845.3265854, doi:10.1145/3265845.3265854. 2

[SNT18] Senecal S., Nijdam N. A., Thalmann N. M.: Motion analysis and classification of salsa dance using music-related motion features. MIG '18, Association for Computing Machinery.

URL: https://doi.org/10.1145/3274247.3274514, doi:10.1145/3274247.3274514. 1, 2

[SS05] SUKTHANKAR G., SYCARA K.: A cost minimization approach to human behavior recognition. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems* (New York, NY, USA, 2005), AAMAS '05, Association for Computing Machinery, p. 1067–1074. URL: https://doi.org/10.1145/1082473.1082635, doi:10.1145/1082473.1082635. 3

[SSB*97] SCHOLKOPF B., SUNG K.-K., BURGES C., GIROSI F., NIYOGI P., POGGIO T., VAPNIK V.: Comparing support vector machines with gaussian kernels to radial basis function classifiers. *IEEE Transactions on Signal Processing 45*, 11 (1997), 2758–2765. doi:10.1109/78.650102. 5

[Vap00] VAPNIK V. N.: *The nature of statistical learning theory*, 2. ed. ed. Statistics for engineering and information science. Springer, New York, NY, 2000. 3

[WWB*23] WU F., WANG Q., BIAN J., DING N., LU F., CHENG J., DOU D., XIONG H.: A survey on video action recognition in sports: Datasets, methods and applications. *IEEE Transactions on Multimedia 25* (2023), 7943–7966. doi:10.1109/TMM.2022.3232034. 3

[ZM06] ZINKE A., MAYER D.: *Iterative Multi Scale Dynamic Time Warping*. Techreport CG-2006-1, Universität Bonn, 2006. 3

[ZREL17] ZHAO W., REINTHAL M. A., ESPY D. D., LUO X.: Rule-based human motion tracking for rehabilitation exercises: Realtime assessment, feedback, and guidance. *IEEE Access 5* (2017), 21382–21394. doi:10.1109/ACCESS.2017.2759801. 2

[ZYL*18] ZHANG C., YANG F., LI G., ZHAI Q., JIANG Y., XUAN D.: Mv-sports: A motion and vision sensor integration-based sports analysis system. In *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications* (2018), pp. 1070–1078. doi:10.1109/INFOCOM.2018.8485910. 3