Automatic Assessment of Skill and Performance in Fencing Footwork

Filip Malawski[a], Marek Krupa and Ksawery Kapela
Faculty of Computer Science, AGH University of Science and Technology, Krakow, Poland

Keywords: Action Recognition, Pattern Recognition, Sports Support, Fencing, Assistive Computer Vision.

Abstract: Typically human action recognition methods focus on the detection and classification of actions. In this work, we consider qualitative evaluation of sports actions, namely in fencing footwork, including technical skill and physical performance. In cooperation with fencing coaches, we designed, recorded, and labeled an extensive dataset including 28 variants of incorrect executions of fencing footwork actions as well as corresponding correct variants. Moreover, the dataset contains action sequences for action recognition tasks. This is the most extensive fencing action dataset collected to date. We propose and evaluate an expert system, based on pose estimation in video data, for measuring relevant motion parameters and distinguishing between correct and incorrect executions of actions. Additionally, we validate a method for temporal segmentation and classification of actions in sequences. The obtained results indicate that the proposed solution can provide relevant feedback in fencing training.

1 INTRODUCTION

Human action recognition (HAR) has applications in multiple areas such as tracking daily activities, human-computer interfaces, or rehabilitation support (Kong and Fu, 2022). One prominent application is tracking and analyzing actions in sports in order to provide valuable information for athletes and coaches (Wu et al., 2022). Supporting sports training requires not only identifying when and which actions occur but also measuring how well an action was performed. There are two relevant aspects for assessing the execution of a sports action. The first is physical performance in terms of speed, force, acceleration, time, range, etc. In many sports disciplines achieving better physical performance (e.g. speed) gives an advantage over the opponent or even is the main goal by itself. The second aspect is technical correctness, which depends on the athlete's skill and can be understood as precision of movement. Those two aspects often stand in opposition - the faster or stronger the movement the less precise it becomes. E.g. in volleyball, it is beneficial to hit the ball with a high force so it would go faster and therefore be more challenging for the opposing team, however, there is little gain from the high speed if the ball falls outside of the court.

In this work, we consider assessing actions in fencing, particularly footwork actions, which constitute a large part of the training routine in this discipline. Bladework, while equally important, is out of the scope of this work. Fencing requires both high physical performance and high technical skills in order for the athletes to be effective. Instructions and feedback from a coach are crucial for improving fencers' level. However, usually a coach has to split their focus to the entire group, rather than train each person individually. We propose an automated system that will provide personalized feedback for footwork exercises, without the supervision of a coach. It is worth mentioning that the purpose of our proposed solution is not to substitute the coach but rather to provide a complementary means of training. We expect that automated assessment should facilitate obtaining a reasonable level of correctness in performing basic actions as well as help eliminate typical errors. Therefore fencers' time with a coach can be better spent on more advanced exercises.

A relevant limitation in introducing automatic assessment of sports actions is the cost and availability of the employed solutions. While professional motion capture systems provide high accuracy in tracking human motion, those are usually affordable only for professional teams. On the contrary, the goal of this work was to design a low-cost, widely available system. To that end, our solution employs only RGB video data, that can be acquired with a typical smartphone. Moreover, the entire data processing framework can also

[a] https://orcid.org/0000-0003-0796-1253

Malawski, F., Krupa, M. and Kapela, K.
Automatic Assessment of Skill and Performance in Fencing Footwork.
DOI: 10.5220/0012577600003660

In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 4: VISAPP*, pages 637-644
ISBN: 978-989-758-679-8; ISSN: 2184-4321

run on a mobile device. Our contribution includes the acquisition of an extensive, novel dataset tailored for analysis of technical correctness and performance of fencing footwork, as well as designing, implementing, and evaluating methods for assessing technical skill and performance of footwork actions. We also validate a method for temporal segmentation and action classification. Our work was conducted in cooperation with fencing coaches in order to ensure that the proposed solution has practical value in fencing training. To the best of our knowledge, this is the first work to address qualitative assessment of a wide range of motion parameters in fencing.

2 RELATED WORK

In general, HAR can include different data modalities such as RGB, depth, infrared, inertial, or even audio or radar signals (Sun et al., 2022). In this work, we focus solely on RGB video modality, which by itself covers a wide range of methods and applications (Beddiar et al., 2020). Early approaches focused on finding relevant points of interest in videos (Laptev, 2005) or estimating motion trajectories (Wang and Schmid, 2013). With the extensive development of deep learning methods, methods based on neural networks have become popular, in particular employing convolutional neural networks (CNN) (Karpathy et al., 2014; Feichtenhofer et al., 2017). An alternative approach is to first create an intermediate representation of the human pose by detecting and tracking relevant joints of persons present in the visual data. This approach has gained much popularity with the release of the Kinect sensor, which employed depth maps to provide so-called skeleton modality (Han et al., 2013). Later works focused on estimating pose from RGB data (Munea et al., 2020). Due to employing deep learning techniques several effective RGB pose estimation methods were proposed (Kendall et al., 2015; Toshev and Szegedy, 2014). Recent solutions can run effectively on mobile devices (Bazarevsky et al., 2020) and are available as ready-to-use components in mobile frameworks (Google, 2023). Pose estimation is particularly useful in sports analysis as it provides detailed information regarding the movement of different body parts.

Action recognition in sports differs from other applications, as it covers a wide range of sports disciplines, each with specific actions (Host and Ivašić-Kos, 2022) and therefore specific datasets (Wu et al., 2022). Some works focus on the classification of sports disciplines, starting with small datasets of 10 disciplines (Soomro and Zamir, 2015) and then expanding to large-scale datasets including hundreds of classes (Kong et al., 2017). Another approach is to define tasks for tracking elements specific to particular disciplines. In team sports player tracking was investigated (Manafifard et al., 2017; Fu et al., 2020). Ball trajectory tracking is a relevant problem for many disciplines, such as tennis (Zhou et al., 2014). Methods for qualitative assessment of action are less common and usually even more specific to particular disciplines. Authors of (Wang et al., 2019) propose a system for detecting incorrect poses in skiing. Golf swings are analyzed in (Ko and Pan, 2021) using CNNs and LSTM networks. Quality of gymnastic actions is assessed in (Zahan et al., 2023) using sparse temporal video mapping.

Several works address action recognition in fencing. Blade action classification was performed using either EMG data (Frère et al., 2010; Klempous et al., 2021), or motion capture systems (Mantovani et al., 2010). Basic fencing footwork actions were classified based on visual and inertial data using local trace images (Malawski and Kwolek, 2018), joint motion history context (Malawski and Kwolek, 2019), and temporal convolutional networks (Zhu et al., 2022). Methods for temporal segmentation of footwork actions were proposed in (Malawski and Krupa, 2023). Our current work advances this research area even further, by addressing the problem of qualitative analysis of actions. In cooperation with fencing experts, we identify key parameters in fencing exercises, that can be measured and evaluated using RGB videos and automatic motion analysis methods. We have recorded a novel, dedicated dataset including multiple variants of correct and incorrect executions of footwork actions. We propose and evaluate a solution for analyzing fencing skills and performance and providing relevant feedback.

3 FENCING FOOTWORK

Fencers typically start in a basic ready stance (called 'on guard') with the sword hand and front foot directed towards the opponent, see Fig. 1 (top-left). Step forward is performed by moving the front leg and then then the back leg, without crossing them. Step backward is analogous but starts with the back leg. Offensive actions are most often performed with a lunge - a dynamic motion in which the fencer reaches out with the front foot while straightening the back leg, see Fig. 1 (top-right). For defense, other than parry with the blade, fencers can perform dodge actions, the two most common being dodge down and dodge back. During dodge down action, the fencer

Figure 1: Selected fencing footwork actions: on guard position (top-left), lunge (top-right), dodge down (bottom-left), dodge back (bottom-right).

lifts both feet and bends the knees in order to fall on the feet with acceleration gained from gravity, see Fig. 1 (bottom-left). Dodge back is performed by dynamically moving the front leg towards the back leg and finishing in a tilted position Fig. 1 (bottom-right).

The effectiveness of each footwork action depends heavily on physical performance, particularly speed and range, as well as technical skill. In steps, proper distance between feet must be maintained, proper hip level (due to bent knees) should be kept, also the front knee should be fully straightened in step forward. The performance of steps is measured with the speed of moving. Technical aspects of the lunge include straightening the front knee in the initial phase, then straightening the back knee in the final phase, as well as maintaining proper front knee angle and body tilt at the end of the action. The range is relevant for the lunge performance. In dodge down it is important to lift the feet in order to take advantage of the gravitational acceleration, but without jumping up (lifting hips), which would slow down the action. In dodge back proper body tilt must be performed at the end and also relative timing of hand and body motion is relevant. All such aspects must be taken into consideration when assessing footwork actions.

4 METHODS

The main goal of this work is an automatic assessment of the quality of footwork actions, however, it required multi-stage development. First of all, we needed to obtain a dedicated dataset, that would include not only sequences of footwork actions but also multiple variants of each considered action, with correct and incorrect executions. Secondly, RGB pose

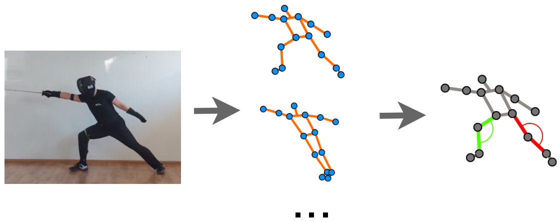

Figure 2: Architecture of the proposed system. Starting with RGB data, pose estimation and action recognition are performed, followed by analysis of selected parameters, such as knee angles.

estimation was performed using state-of-the-art methods. Then, temporal segmentation and classification of actions were performed on sequences of footwork exercises in order to evaluate recognizing actions during footwork practice. Finally, we developed an expert system for the assessment of technical skill and performance measured by 15 relevant parameters. The architecture of the system is presented in Fig. 2.

4.1 Data Acquisition

Obtaining a proper dataset was one of the crucial aspects of this work. First, we conducted consultations with fencing experts in order to prepare a list of footwork actions and parameters describing technical correctness and performance. Secondly, initial recordings were made, including different variants of actions performed by an experienced fencer. Acquired material was analyzed both by a computer vision specialist and fencing coach and jointly a list of final parameters to be measured was prepared, as listed in Table 1. The acquisition plan for each recorded person was as follows. First, three sequences of continuous fencing footwork were recorded, each with at least three repetitions of each action, performed in random order. This part was collected for the development and evaluation of temporal segmentation and action classification methods. Next, the fencers were asked to perform specific variants of each action, corresponding to correct and incorrect execution for each considered parameter. E.g. they would perform a correct lunge then a lunge with not fully straightened front knee, then a lunge with incorrect body tilt, etc. Each variant was performed only once, due to time constraints in recording sessions, however, a fencing expert supervised the process and asked to repeat an action if it was not representative of a given parameter. It is however worth mentioning, that while some executions strongly emphasized specific variants, others may differ only slightly from the correct variant, depending on the action and performing person. The

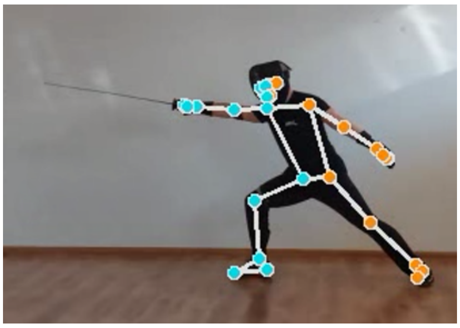

Figure 3: Pose estimation of fencing footwork action.

data was acquired using a mobile device (Samsung A52s smartphone) using a custom video recording application with 60 frames per second. Simultaneously inertial data from five inertial sensors mounted on the fencer were recorded. While the inertial data is out of the scope of this work, it allows for multimodal action recognition in future works. All recordings were manually labeled by a fencing expert.

4.2 Pose Estimation

Our method relies on RGB pose estimation, for which we employ the Blazepose algorithm (Bazarevsky et al., 2020), available in the Mediapipe library (Google, 2023). This implementation can also be easily used on mobile devices. A total of 33 keypoints are tracked, however not all are relevant for this work, e.g. face landmarks or positions of index fingers and thumbs are not used in our scenario. An example of pose estimation for the lunge action is presented in Fig. 3. Additionally, the positions of joints are filtered with a moving average filter (window size = 9) to remove glitches in pose estimation. Since the employed state-of-the-art algorithm was trained on a large dataset, it is robust to variance in environmental conditions, such as lighting or background, as well as variance regarding tracked persons, such as different clothes or different proportions of body parts. Therefore, our motion analysis models can focus on capturing variance in the performance of actions.

4.3 Action Recognition

While action recognition is not the primary goal of this work, we adapt previously proposed methods for temporal segmentation and action classification on the sequence recordings of our dataset. We extend the approach proposed in (Malawski and Krupa, 2023) by replacing handcrafted features with an automated framework for feature extraction and selection, namely TSFEL (Barandas et al., 2020). Initial data includes a time series of positions of all joints tracked by the Mediapipe library. All features available in the framework were extracted in time windows of size 50, and then feature selection based on a decision tree ensemble (ExtraTree) was employed to automatically select the most relevant features. Each frame is classified based on its context (time window) using the XGBoost classifier, and then neighboring frames of the same class are combined as segments corresponding to actions. Outlier segments, with less than 15 frames of specific action are reclassified to match the closest larger action segment. While we consider five top-level actions for qualitative analysis, action recognition includes a total of nine classes in order to cover the full spectrum of actions. Additional classes include return from lunge, return from dodge down, return from dodge back, and 'other', which corresponds to all untypical actions that sometimes occur, such as jumping steps.

4.4 Skill and Performance Assessment

The key difficulty in this work was identifying how to measure relevant parameters of motions in order to assess the quality of actions. Table 1 presents a list of technical (T) and performance (P) parameters of footwork actions, created in cooperation with fencing experts. While other aspects of motion may be relevant as well, we include only parameters that were identified as common sources of errors in exercises, while at the same time being measurable in video recordings. The table includes a short description of expected correct and incorrect executions as well as the corresponding parameter measured with the proposed method, based on automatic pose estimation. Depending on the parameter, its value is measured in different phases of the action, which is also included in the table. In the moving phase, the pose dynamically changes and usually min. or max. value of a parameter is relevant and in the resting phase the pose is stable for a short time (e.g. after lunge) and static parameters such as knee angle are considered. Some parameters are measured in both phases, e.g. hip height in steps.

For step forward and step backward actions most parameters are similar, except for straightening of the front knee, which is relevant only in the step forward. Feet distance is measured relative to shoulder distance, which is estimated based on the entire recording, as in some frames it is not well visible. Measuring speed is tricky, as observed changes in position depend on the distance to the camera. Moreover, slow, normal, and fast movements may be different for each fencer. Therefore, we normalize this value

Table 1: Fencing footwork parameters.

Action param.	Correct	Incorrect	Phase	Type	Measured param.
Step forward (SF) / step backward (SB)					
Feet distance	Similar as shoulder width	Feet too close or too wide	Resting	T	Ratio of min. and max. ankle dist. to shoulder dist.
Knees angle	Knees bent moderately	Knees almost straight	Both	T	Mean knee angle (both knees)
Hips height stability	Small variance	Large variance	Both	T	Variance of hip to ankle distance
Front knee straight. (SF)	Fully straightened	Not fully straightened	Moving	T	Max. front knee angle
Speed	-	-	Moving	P	Mean velocity of hip joints
Lunge					
Front knee straightened	Fully straightened	Not fully straightened	Moving	T	Max. front knee angle
Back knee straightened	Fully straightened	Not fully straightened	Resting	T	Max. back knee angle
Front knee position	Above ankle	Above middle or end of the foot	Resting	T	Horizontal dist. between the front ankle and front knee
Body tilt	Medium tilt	No tilt or too much tilt	Resting	T	Hip to shoulder line angle relative to the ground
Arm straightening timing	Before body movement	After body movement	Moving	T	Elbow angle after ten frames
Range	-	-	Both	P	Horizontal distance between hip position before and after the lunge
Dodge down					
Feet lifted	Feet slightly lifted	Feet not lifted	Moving	T	Max. vertical distance of ankles
Hips not lifted	Hips not lifted	Hips lifted	Moving	T	Max. vertical distance of hips
Dodge back					
Body tilt	Approx. 45 degrees	Too small or too much	Resting	T	Hip to shoulder line angle relative to the ground
Arm straightening timing	Approx. at the same time as body	Too soon or too late	Moving	T	Elbow angle after 15 frames

Table 2: Results of action assessment (F1 score).

Variant	XGB	SVM
Step forward		
Feet too narrow	0.952	0.947
Feet too wide	0.875	0.824
Knees bent too little	0.706	0.778
Speed slow	0.800	0.800
Speed fast	0.875	0.875
Front knee not str.	0.824	0.857
Hips unstable	0.778	0.800
Step backward		
Feet too narrow	0.941	0.941
Feet too wide	0.800	0.875
Knees bent too little	0.667	0.533
Speed slow	0.875	0.778
Speed fast	0.933	0.933
Hips unstable	0.900	0.900
Lunge		
Front knee not str. S	0.842	0.889
Front knee not str. L	0.941	1.000
Back knee bent	0.727	0.762
Range short	0.917	0.957
Range long	0.900	0.842
Front knee too far S	0.917	0.917
Front knee too far L	0.963	1.000
Body tilt too small	0.800	0.800
Body tilt too much	0.875	0.800
Arm after body	0.909	1.000
Dodge down		
Feet not lifted	0.833	0.769
Hips lifted	0.952	0.842
Dodge back		
Body tilt too small	0.900	0.900
Body tilt too much	0.957	0.900
Arm after body	0.947	0.947
Mean	0.868	0.864

using mean leg length (hip to knee to ankle distance) from the entire recording. Similarly, the range of lunge action is also normalized using mean leg length.

5 EXPERIMENTS

5.1 Dataset and Action Recognition

Our dataset includes recordings acquired with 8 experienced fencers, 5 male and 3 female. Sequences include from 9 to approx. 50 repetitions of each action per person, depending on the action (less common are dodging actions, most common are steps). A total of 40 variants of actions were recorded with each person, including correct and incorrect executions in the context of different action parameters. As mentioned before, temporal segmentation and action classification was not a primary objective of this work and therefore we do not include a comparative study for this part. The method described in Section 4.3 was evaluated and actions were considered to be recognized correctly if the middle frame of predicted action was included in the ground truth segment. F1 score = 95.5% was obtained. Further experiments, considering the qualitative analysis of actions, were performed using recorded variants of actions.

5.2 Skill and Performance Assessment

For each variant, a separate binary classifier (correct vs incorrect) was trained and evaluated using leave-one-person-out cross-validation (eight folds, one test person in each fold). Both XGBoost and SVM classifiers were evaluated. Table 2 presents F1 scores obtained for each considered variant. Only variants with incorrect execution are listed, as they were internally compared to the correct variant, using the binary classifiers. In the case of performance parameters (speed, range) there are no correct or incorrect executions, however, specific variants (e.g. slow, fast speed) were compared against typical execution (e.g. normal speed). Some variants were recorded separately with small (S) and large (L) differences from correct execution.

Results indicate that for most parameters it is possible to efficiently distinguish between correct and incorrect variants. Out of 28 variants of incorrect executions only 4 have F1 score lower than 0.8, while 13 have F1 score at least 0.9. Particularly difficult to measure are parameters regarding bending the knees, both in steps and in lunges. On the other hand, some very important errors in executions, such as keeping the feet too close in steps or not straightening the front knee in a lunge are very well recognized. The mean F1 score is 0.868 for the XGBoost classifier and 0.864 for SVM. While on average both classifiers perform very similarly, there are some significant differences in specific parameters. This indicates that more data points would be beneficial to obtain more statistical information. Alternatively, thresholds between correct and incorrect executions could be determined by a human expert.

We analyzed several cases in which the classification of correct and incorrect variants failed, in order to investigate possible reasons. One relevant source of errors is how each variant was performed by each person. In some cases, the difference between incorrect and correct action performance was relatively small. While human experts are able to distinguish these variants, the differences in executions were not always captured by the automatic system. For some parameters, we expect that additional per-user calibration may be needed, as inter-person differences may be higher than differences between correct and incorrect variants. Occasionally, problems stemmed from incorrect pose estimation. Mediapipe is robust to varying environmental conditions, moreover employing moving average filters out most outliers. However, some errors in pose estimation still occur, which may be particularly problematic for parameters based on minimum or maximum values.

6 CONCLUSIONS

In this work, we addressed the problem of qualitative evaluation of actions in fencing footwork, including assessment of technical skill and physical performance. The goal was to provide relevant information for fencing training evaluation.

In cooperation with fencing experts, we designed and recorded a novel dataset including sequences of fencing footwork practice as well as 40 variants of actions per person (28 with incorrect execution variants to be recognized). The dataset includes manual labels for actions and variants, provided by fencing experts. To the best of our knowledge, this is currently by far the most detailed dataset of fencing actions. The employed method for temporal segmentation and action classification is sufficiently effective to be used in practical applications. We designed and evaluated specific methods for measuring motion parameters relevant to each variant of incorrect execution. Results indicate that in most cases, our system can provide relevant feedback for fencers.

In future work, we intend to focus on improving the recognition of several action variants by including user-specific calibration or adaptation mechanisms. Moreover, we plan to investigate the idea of using expert-based thresholds instead of automatic ones. Finally, the proposed solution is currently being implemented in a mobile application. Therefore, we expect to validate our approach during fencing training sessions and gather valuable feedback from fencers and coaches.

ACKNOWLEDGEMENTS

The research presented in this paper was supported by the National Centre for Research and Development (NCBiR) under Grant No. LIDER/37/0198/L-12/20/NCBR/2021. Data acquisition and expert consultations were carried out in cooperation with Aramis Fencing School (aramis.pl).

REFERENCES

Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., Liu, H., Schultz, T., and Gamboa, H. (2020). Tsfel: Time series feature extraction library. *SoftwareX*, 11:100456.

Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., and Grundmann, M. (2020). Blazepose: On-device real-time body pose tracking. *arXiv preprint arXiv:2006.10204*.

Beddiar, D. R., Nini, B., Sabokrou, M., and Hadid, A. (2020). Vision-based human activity recognition: a survey. *Multimedia Tools and Applications*, 79(41-42):30509–30555.

Feichtenhofer, C., Pinz, A., and Wildes, R. P. (2017). Spatiotemporal multiplier networks for video action recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4768–4777.

Frère, J., Göpfert, B., Nüesch, C., Huber, C., Fischer, M., Wirz, D., and Friederich, N. (2010). Kinematical and emg-classifications of a fencing attack. *International journal of sports medicine*, pages 28–34.

Fu, X., Zhang, K., Wang, C., and Fan, C. (2020). Multiple player tracking in basketball court videos. *Journal of Real-Time Image Processing*, 17:1811–1828.

Google (2023). Mediapipe.

Han, J., Shao, L., Xu, D., and Shotton, J. (2013). Enhanced computer vision with microsoft kinect sensor: A review. *IEEE Transactions on Cybernetics*, 43(5):1318–1334.

Host, K. and Ivašić-Kos, M. (2022). An overview of human action recognition in sports based on computer vision. *Heliyon*.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., and Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725–1732.

Kendall, A., Grimes, M., and Cipolla, R. (2015). Posenet: A convolutional network for real-time 6-dof camera relocalization. In *Proceedings of the IEEE international conference on computer vision*, pages 2938–2946.

Klempous, R., Kluwak, K., Atsushi, I., Górski, T., Nikodem, J., Bożejko, W., Chaczko, Z., Borowik, G., Rozenblit, J., and Kulbacki, M. (2021). Neural networks classification for training of five german longsword mastercuts-a novel application of motion capture: analysis of performance of sword fencing in the historical european martial arts (hema) domain. In *2021 IEEE 21st International Symposium on Computational Intelligence and Informatics (CINTI)*, pages 000137–000142. IEEE.

Ko, K.-R. and Pan, S. B. (2021). CNN and bi-LSTM based 3D golf swing analysis by frontal swing sequence images. *Multimedia Tools and Applications*, 80:8957–8972.

Kong, Y. and Fu, Y. (2022). Human action recognition and prediction: A survey. *International Journal of Computer Vision*, 130(5):1366–1401.

Kong, Y., Tao, Z., and Fu, Y. (2017). Deep sequential context networks for action prediction. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1473–1481.

Laptev, I. (2005). On space-time interest points. *International journal of computer vision*, 64:107–123.

Malawski, F. and Krupa, M. (2023). Temporal Segmentation of Actions in Fencing Footwork Training. *Computer Science Research Notes*, 3301:241–248.

Malawski, F. and Kwolek, B. (2018). Recognition of action dynamics in fencing using multimodal cues. *Image and Vision Computing*, 75:1–10.

Malawski, F. and Kwolek, B. (2019). Improving multimodal action representation with joint motion history context. *Journal of Visual Communication and Image Representation*, 61:198–208.

Manafifard, M., Ebadi, H., and Moghaddam, H. A. (2017). A survey on player tracking in soccer videos. *Computer Vision and Image Understanding*, 159:19–46.

Mantovani, G., Ravaschio, A., Piaggi, P., and Landi, A. (2010). Fine classification of complex motion pattern in fencing. *Procedia Engineering*, 2(2):3423–3428.

Munea, T. L., Jembre, Y. Z., Weldegebriel, H. T., Chen, L., Huang, C., and Yang, C. (2020). The progress of human pose estimation: A survey and taxonomy of models applied in 2d human pose estimation. *IEEE Access*, 8:133330–133348.

Soomro, K. and Zamir, A. R. (2015). Action recognition in realistic sports videos. In *Computer vision in sports*, pages 181–208. Springer.

Sun, Z., Ke, Q., Rahmani, H., Bennamoun, M., Wang, G., and Liu, J. (2022). Human action recognition from various data modalities: A review. *IEEE transactions on pattern analysis and machine intelligence*.

Toshev, A. and Szegedy, C. (2014). Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1653–1660.

Wang, H. and Schmid, C. (2013). Action recognition with improved trajectories. In *Proceedings of the IEEE international conference on computer vision*, pages 3551–3558.

Wang, J., Qiu, K., Peng, H., Fu, J., and Zhu, J. (2019). Ai coach: Deep human pose estimation and analysis for personalized athletic training assistance. In *Proceedings of the 27th ACM international conference on multimedia*, pages 374–382.

Wu, F., Wang, Q., Bian, J., Ding, N., Lu, F., Cheng, J., Dou, D., and Xiong, H. (2022). A survey on video action recognition in sports: Datasets, methods and applications. *IEEE Transactions on Multimedia*.

Zahan, S., Hassan, G. M., and Mian, A. (2023). Learning sparse temporal video mapping for action quality assessment in floor gymnastics. *arXiv preprint arXiv:2301.06103*.

Zhou, X., Xie, L., Huang, Q., Cox, S. J., and Zhang, Y. (2014). Tennis ball tracking using a two-layered data association approach. *IEEE Transactions on Multimedia*, 17(2):145–156.

Zhu, K., Wong, A., and McPhee, J. (2022). Fencenet: Fine-grained footwork recognition in fencing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3589–3598.