Automated System for Indoor Sport Video Production

Sebastian Purtak[1], Jagoda Lazarek[1,2], Paweł Gora[1,3] and Łukasz Gąsiorowski[1]

[1] *Stellis sp. z o.o., Piotrkowska 157A/10.03, 90-270 Lodz, Poland*

[2] *Polish-Japanese Academy of Information Technology, Koszykowa 86, 02-008 Warsaw, Poland*

[3] *University of Warsaw, Faculty of Mathematics, Computer Science and Mechanics, Banacha 2, 02-097 Warsaw, Poland*

Keywords: Computer Vision, Object Detection, Object Tracking, Handball, Futsal, Artificial Intelligence, Sport, Deep Learning, Expert Systems.

Abstract: Information is the key to success in every area of life, including professional and amateur sports. This is why the recording of matches is so important for sports clubs and organisations, as it allows them to gain an information advantage. In this paper, we present the prototype of the ISVP.AI system, which uses Artificial Intelligence technology to automate the production process of sports recordings in handball and futsal. Our solution involves the use of YOLO family algorithms to detect objects such as players or a ball for each frame of the recording. This information is then used by the expert system, whose task is to detect events relevant to the course of the game. The representation of the state of the game built in this way allows for the automatic production of full sports recordings, match highlights, calculation of statistics and adding effects that improve the attractiveness of the produced recordings. Thanks to this, our system allows for potentially significant reduction of costs associated with recording handball and futsal sports games.

1 INTRODUCTION

Recordings of indoor sports games are an important source of information for participating teams. The analysis of the match recording allows for a detailed assessment of the game of both teams, the performance of individual players or the effectiveness of the tactics used. In result teams can identify their strengths and weaknesses, as well as use information from recordings of opponents' matches to modify their strategy. In addition to tactical and training values, the registration of sports events is also of marketing importance. A properly recorded match recording is in itself an entertainment product that can be addressed to fans of a given club. Match highlights are the perfect type of content for all kinds of social media channels. High-quality video materials can be used not only to attract sponsors or new viewers but also to promote the entire sports discipline.

The correct preparation of the recording of an indoor sports match is a complex task that requires the involvement of financial, technical, and personal resources. During regular matches in indoor sports, image recording is usually carried out using a single camera, placed in the middle of the field, behind the field line. The camera is operated by an operator who controls it in such a way that it follows the action on the field. The recorded material must then be processed to produce match highlights and calculate the statistics. These tasks require specialist knowledge in the field of image registration and processing, which significantly increases the cost of the entire project. For this reason, sports clubs that do not have budgets often give up the practice of recording matches, and thus also all the benefits that come with it.

To solve these problems, there were several attempts to build systems for supporting and automating the process of producing sports videos (Ariki et al., 2006). The available solutions allow for partial automation of the production of sports recordings, but at the same time, they have several significant limitations. First of all, the most advanced systems are dedicated to the most popular sports disciplines (Quiroga et al., 2020). There are no systems on the market dedicated to futsal and handball, which would cover the characteristics of these sports disciplines with their specification. Moreover, none of these systems offer full automation of the production process of visually attractive recording. Thus, their use still involves additional staff. Other significant limitations of the solutions available on the market include the need to purchase dedicated equipment, limited possibilities for the user to interfere in the recording production process, or the inability to combine match highlights.

All this means that the use of systems supporting the production of match recordings, although it can be

Purtak, S., Lazarek, J., Gora, P. and Gąsiorowski, Ł.
Automated System for Indoor Sport Video Production.
DOI: 10.5220/0012159300003587
In *Proceedings of the 11th International Conference on Sport Sciences Research and Technology Support (icSPORTS 2023)*, pages 23-33
ISBN: 978-989-758-673-6; ISSN: 2184-3201

helpful in the implementation of large projects, usually involves a significant increase in costs. Therefore, the possibility of using such solutions remains poorly available to small organisations that do not have significant financial resources. The lack of products dedicated to futsal and handball additionally results in the low popularity of these systems in these sports.

In this paper, a prototype of the ISVP AI system, which enables full automation of match recording production, is presented. The solution is dedicated to futsal and handball, and its main assumption is the use of artificial intelligence algorithms to significantly reduce the cost of the entire project. ISVP AI uses advanced computer vision algorithms and an expert system, which builds a data structure describing the state of the game for each frame of the recording. Such a form of representation then allows for the implementation of more advanced functions, such as determining the visibility area or generating match highlights.

The prototype does not require the purchase of a dedicated camera and can process the image recorded with any equipment, as long as it is recorded in 4K resolution and 30 FPS. In addition, all ISVP AI functionalities have been implemented in such a way as to ensure the greatest possible user intervention in the production process. All of the solutions used are aimed at building a product that will significantly increase the availability of match recording practice among small and medium futsal and handball clubs.

2 STATE OF THE ART

2.1 Systems for the Production of Sports Recordings

Systems offering full automation of sports match recordings have been a subject of research by companies in the sports technology industry. A prime example is the prototype developed by Genius Sport, which aims to automate basketball game recordings (Quiroga et al., 2020). This system is constructed from a series of components responsible for object detection within matches, court modelling, action recognition, and game state tracking. The object detection component employs the YOLOv3 algorithm and is capable of identifying players, the ball, and referees. The remaining components of the system are tailored to the specifics of basketball as a discipline. This system is exclusively focused on automating the production of complete sports recordings, thereby omitting the creation of highlights or streaming capabilities.

Nonetheless, it stands as a solution that, despite its prototype nature, demonstrates the potential of utilising contemporary machine learning techniques for achieving full automation in sports recordings.

One of such solutions is the offer of the Hudl company, which includes an extensive platform for sports data analysis and a set of cameras that allow for the automation of the match recording process (Agile Sports Technologies, 2021). These cameras come in three variants, designed to record outdoor games (Focus Outdoor), indoor sports matches (Focus Indoor), and football matches (Focus Flex). The image is recorded in 1080p resolution. The cameras can automatically follow the action on the field and have the support of the AI system to track match objects. Integration with the Hudl software allows recording to start automatically, according to the club's schedule, and the cameras can simultaneously stream the recorded video. The software includes an application to control Hudl Focus cameras and several additional programs. All these products are fully compatible with each other and create one ecosystem whose main task is to support sports clubs in the field of recording, processing, and sharing sports recordings. At the moment, the system operates in five sports disciplines, i.e. football, basketball, volleyball, rugby, and ice hockey. This solution is addressed primarily to large sports clubs that produce and analyse large amounts of match recordings. The Hudl platform does not support futsal and handball, and for its proper operation, it is necessary to purchase dedicated equipment. It is also a system that primarily supports the production of sports recordings, not its full automation.

Systems like those offered by Hudl are aimed at larger clubs that can afford to invest in such advanced technical solutions. The segment of smaller sports recording systems can be represented by a set of VEO sports cameras (VEO, 2021). The system is designed to automate the streaming of football matches. It consists of a specialised Veo Cam 2 camera, equipped with hardware AI support, and software in the form of a mobile application and Veo Editor program with extensions. The camera is weatherproof and able to automatically follow the action on the field. The system can automatically detect match events and allows users to interfere with the recording by applying markings and notes or adding events themselves. The main limitations of the system come down to the need to purchase dedicated hardware and the fact that it is mainly designed to operate in outdoor conditions. For this reason, the manufacturer indicates that its product is primarily optimised for football games. The cost of the entire solution is additionally increased by the need to purchase software along with paid addons that are necessary to use the full capabilities of the system.

To sum up, currently available systems for the production of match recordings on the market successfully use computer vision technologies to partially automate this process. However, these are solutions that require the purchase of specialised equipment and offer automation only as a supporting function. Moreover, there are currently no such solutions dedicated to handball and futsal. Thus, in order to produce visually attractive recordings of matches in these disciplines, it is still necessary to engage financial and personal resources that are beyond the reach of small clubs and sports organisations.

2.2 Object Detection

The key technology in building systems to support the production of match recordings is computer vision. A special role in this regard is played by object detection algorithms, as they allow for the identification and collection of information about individual match participants. The significant progress that has been made in this field in recent years is the result of the use of Deep Learning.

Modern object detection algorithms can be divided into two main groups: one-stage and two-stage detectors. Among the two-stage detectors, the dominant role is played by the R-CNN algorithm (Girshick et al., 2013) and its subsequent versions, including Fast R-CNN (Girshick, 2015) and Faster R-CNN (Ren et al., 2015). The operation of these algorithms consists of the first step in determining the areas in the image where the object may potentially be located (Region Proposal Method). In the second step, localisation and classification are carried out in the detected areas in order to determine whether the sought object actually exists, what class it belongs to, and where exactly it is located. A characteristic feature of these solutions is high detection efficiency, which is paid for by a relatively long inference time.

Single-stage detectors perform detection at once over the entire image. By omitting the step of detecting potentially significant image areas, the inference time is significantly reduced and the algorithm learning procedure is simplified. These solutions perform worse in terms of detection efficiency than two-stage detectors, but their fast operation time makes them suitable for real-time object detection. The most important group of such solutions are YOLO algorithms (Redmon, 2016). It is this family of algorithms that best meets the technical requirements of systems for automating the production of match recordings, as it allows not only to significantly reduce production time, but can also be used in live streams.

The most advanced versions of the YOLO algorithm are YOLOv4, YOLOv5, YOLOv7, and YOLOR. It should be mentioned that the development of these algorithms is carried out today by many different teams, which leads to a lack of consistency in the naming. The original development line consists of versions from one to four, and it is YOLOv4 that is today the latest original implementation of this algorithm (Bochkovskiy et al., 2016). YOLOv5 is the first unofficial implementation that performs slightly better than YOLOv4 and is very popular due to its easier learning and implementation procedure (Horvat and Gledec, 2022). Another unofficial implementation, which is YOLOv7, introduces new techniques for reparameterisation and model scaling. It also allows for a significant reduction in the inference time and slightly higher precision compared to other models (Wang et al., 2022). YOLOR achieves results similar to YOLOv4, with a much shorter inference time, and allows the use of implicit and explicit knowledge in the learning process (Wang et al., 2021). All of these models, in their basic version, are trained using popular datasets, which are benchmarks in the field of object detection. The most important are COCO (Lin et al., 2014), PASCAL (Everingham et al., 2009), and ImageNet (Deng et al., 2009). These sets are also used to compare and evaluate the aforementioned models.

Finally, a new type of detectors should be mentioned that use the transformer architecture known from large language models (LLM) (Carion et al., 2020). These algorithms achieve very promising results, which should be considered as a direction for further development of object detection technology. However, at the moment the inference times of these models do not allow them to operate in real time, and thus they do not apply to video streaming systems. What's more, transformers in their current version are not very good at detecting small objects. This may be important in the context of ball detection, and thus affect the effectiveness of such a solution in the system for the automatic production of match recordings. On the other hand, the YOLO family of algorithms, as a mature solution, offers several innovations to improve their performance in this area. For this reason, for the purposes of this paper, experiments were carried out using the most advanced versions of YOLO.

3 APPLICATION

3.1 Features

The ISVP AI system is a solution entirely dedicated to the automation of match recordings in futsal and handball. This means that all the technologies used in it have been adapted to the specificity of these sports disciplines. This applies primarily to the machine learning models used, which were subjected to the learning process using training data sets covering only handball or futsal match recordings.

Specialisation allowed us to build a solution enabling full automation of match recording production. This means that ISVP AI, after proper configuration, can produce a visually attractive match recording without any human intervention. The final product of our system's operation is a recording covering the entire field, with a virtual camera that steers the field of view following the action. This mechanism simulates the work of the camera operator by selecting the area of the frame to be visible, moving it, and making close-ups depending on the situation on the field.

To complete this task, it is necessary to correctly recognize the state of the game for each frame. The system must be able to determine where the action is happening at a given moment to correctly indicate the area of the recording that is visible to the viewer. ISVP AI uses object detection algorithms to determine and track the position of all visible match objects. The detection of more complex match events is carried out using an expert system. In this way, a representation of the state of the game is created for each frame, which allows the implementation of the main functionalities of the system.

Match objects are understood as all persons who, under the rules of a given discipline, are directly involved in playing the match. So there will be both players and goalkeepers of both teams, as well as referees. In addition, the props necessary for the match should also be considered as match objects. In the case of futsal and handball, these will be goals and a ball.

In order to perform its task correctly, the system must be able to independently detect the location of all the mentioned match objects on each frame of the recording. In addition, due to the sequential nature of the processed data, it is necessary to track the displacement of individual match objects on subsequent frames of the recording. Finally, in order to correctly determine the situation on the field, it is necessary to identify and classify individual match objects. In other words, the system must not only be able to determine that there is a match object in a given area of the presented frame, but also that it is a player and to which team this player belongs.

The information collected in this way allows for precise determination of the location of the action on the recording covering the area of the entire field. Thanks to this, there is no need to record the video from several different points and it is possible to use a single camera. There is also no need to purchase specialised equipment, because the algorithms used can work with any camera whose parameters fall within a broad technical framework.

Information about detected match objects is then transferred to the expert system whose task is to detect match events. This term should be understood as situations that are important for the course of the game, in which the match objects are involved. A prime example of a match event is the scoring of a goal. The system should therefore be able to combine information from the sequence of presented images, and then classify them correctly.

Precise determination of the state of the game for individual frames allows you to determine which of them should be included in the final recording. The detected and classified match events will enable the production of highlights of both the actions and the entire match. The same data can be used to calculate match statistics automatically. The technologies used will also allow for a much greater intervention of the end user in the entire recording production process. Detecting the position of individual match objects will allow, among other things, to add AR effects without the need to edit the recording using separate software and without the need for specialist knowledge.

All of these operations, including the production of a full match recording, take less than 4 hours. The achieved results make ISVP AI one of the fastest systems of this type currently. At the same time, it is the only solution that is fully dedicated to futsal and handball, thus becoming a significant innovation in the field of technical support for these indoor sports disciplines. This is especially important in the context of low system implementation costs. ISVP AI is currently the only system to support and automate the production of sports recordings, which does not require a dedicated camera to operate. The system can accept and process data in the form of a video file or stream, in 4k resolution, at 30 FPS, and in HEVC encoding. This allows potential users to use any cameras that can record video with such parameters, which further reduces the costs of the entire project.

3.2 Architecture

The ISVP AI system consists of two main components: the ISVP APP web application and the Machine Observer service. The system is controlled from the web application, while the Machine Observer is responsible for performing calculations using machine learning algorithms. Both main components have a modular structure, which facilitates the potential scaling and expansion of the system (Fig. 1). The backbone of the entire solution is the Google Cloud infrastructure.

ISVP APP, apart from its standard logic, distinguishes four main modules: Game State, Cameraman, Event Detector, and Object Highlighting. Each of these modules contains logic that implements one of the key functionalities of the application. Game State is responsible for saving and storing information about the state of the game on individual frames that were obtained in the course of video processing. Event Detector includes an expert system whose task is to detect match events based on data on detected match objects. The cameraman also uses match object data to add a virtual camera effect to the video. Finally, Object Highlighting is the module responsible for generating match highlights, for which it uses a full set of information about both objects and match events.

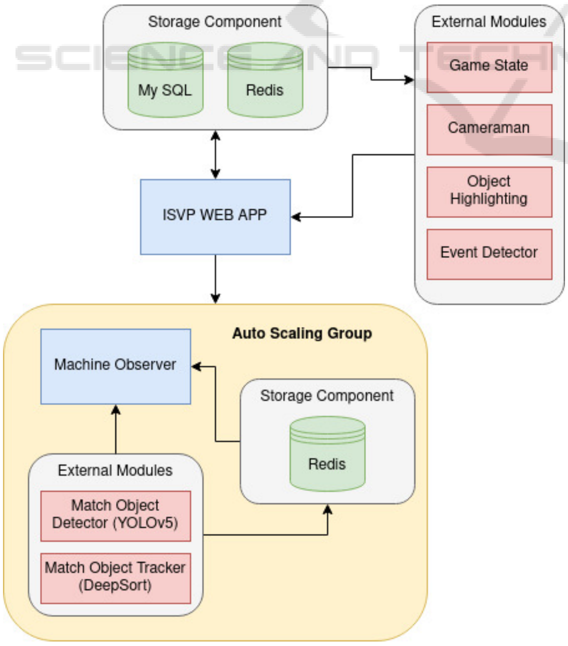


Figure 1: Diagram of the high-level architecture of the ISVP AI system.

All of the listed modules are containerised and their operation can be moved to a separate virtual machine, if necessary. A similar solution was used in Machine Observer. In this case, two modules can be distinguished, the first of which is responsible for the implementation of the object detection model, and the second implements the object tracking model. The data flow in the system starts with the user uploading the unprocessed version of the video. It is then stored in Google Storage and the user can choose to start the processing. In such a situation, the ISVP APP establishes communication with the Machine Observer and leases a virtual machine to perform the calculations. Machine Observer performs the detection process using both detection and tracking and then sends the results back to the application. Here they are transferred to the Game State module, which saves them, and control is returned to the user. The latter can decide to start the next stage of processing or decide to intervene on his own by adding AR effects or manually marking match objects. Depending on the user's decision, one of the application modules is launched and the data collected in the Game State module is then transferred to it. The entire system was made in Python technology with the use of Postgresql and Redis databases.

4 EXPERIMENTS

4.1 Detection and Tracking of Match Objects

In the area of match object detection, we conducted a series of experiments to select the best possible detector. In addition to the average precision (mAP), which is a classic metric of object detection efficiency, the inference time also played an important role in the evaluation of the models. This is due to the need for real-time detection in the case of match streaming, which was one of the factors determining the choice of algorithms from the YOLO family as the primary detectors.

For the purposes of the experiments, there were two datasets prepared, separately for futsal and handball (Tab. 1). Each set had the same set of five classes of match objects: player, goalkeeper, referee, goal, and ball. In order to collect data, a total of 32 matches (16 for futsal and 16 for handball) were recorded in 5 sports halls. The footage covering 1,372 minutes of futsal matches and 1,555 minutes of handball matches was collected. The videos were recorded in 4K resolution, at 30 FPS, and in HEVC encoding.

The first algorithm that was tested was YOLOv5. This is an unofficial version of the main development line of YOLO algorithms, which differs only slightly

Table 1: Characteristics of data sets.

	Futsal	Handball
number of frames	48942	49638
player	198419	295235
goalkeeper	48263	48443
referee	49341	49051
goal	48956	42560
ball	22195	21340

from YOLOv4 in terms of architecture. However, the implementation is completely different, which in this version allows for much easier customisation of the model. There are four main variants of YOLOv5, which are essentially scaled-down versions of the same architecture. The starting point for our experiments was the YOLOv5x variant, which is the largest version of this detector. Such a solution, although computationally expensive, allowed for determining the upper limits of the unmodified version of YOLOv5 in detecting match objects.

The review and analysis of the data then allowed identifying potential directions for modifying the algorithm. Among all the classes we detect, there are two main groups: people and props. They differ fundamentally in terms of the proportions and sizes of bounding boxes (Fig. 2). While the bounding boxes for players, referees, and goalkeepers are relatively similar in size, the bounding boxes for different props can vary significantly. There are two classes of props in our datasets: ball and goal. They differ not only in size but also in specificity, because the goal is a stationary object, always remaining in the same place on the field, while the ball often moves at high speed. Therefore, it can be assumed that the goals will be relatively easy to detect objects, while the ball can be a challenge for the model.

With this in mind, it was decided to modify the model's architecture to improve not only its inference time but also its ability to detect small objects. Usually, it is appropriate to use a concatenation-based network as a backbone (Benjumea et al., 2021), such as DenseNet. These are very effective feature extractors, the use of which is computationally expensive and can negatively affect the inference time. Instead, we modified the neck, in which we used the Feature Pyramid Network. It is a component specialised in detecting objects at different scales, which turns out to be particularly useful when detecting small objects (Lin et al., 2017). At the same time, it is a simpler solution than the standard neck construction in YOLOv5, which should allow us to reduce the inference time while maintaining efficiency in detecting small objects.

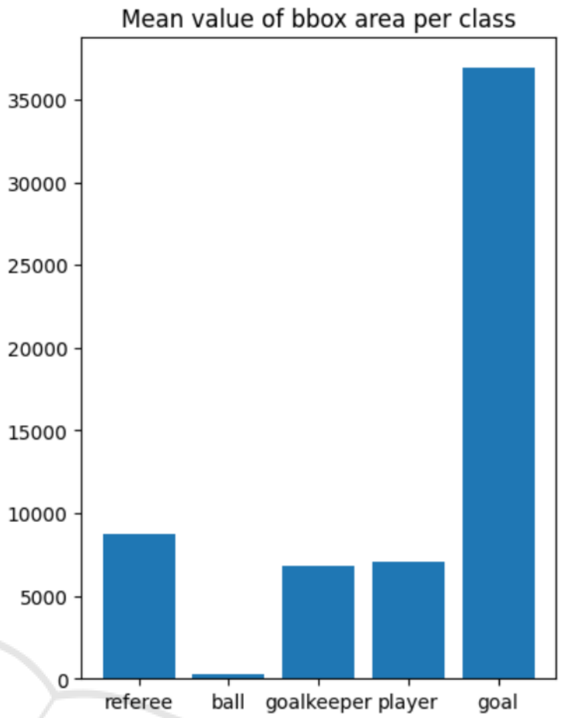
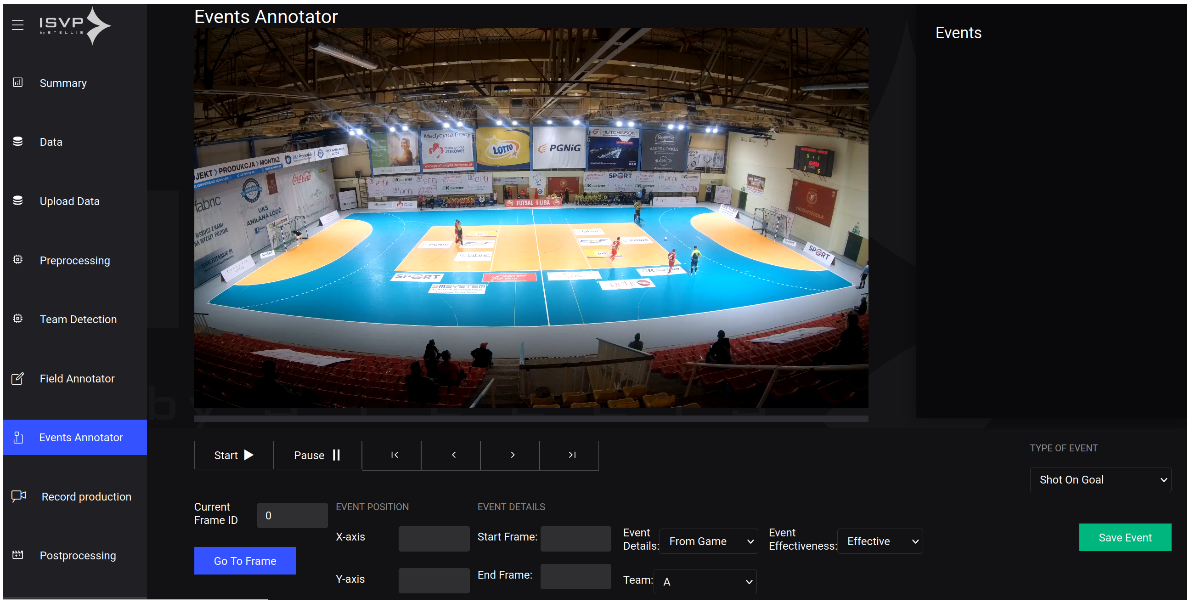


Figure 2: The average value of bbox area in the futsal dataset.

Finally, the last version of the algorithm that was used in the experiments was YOLOv7. This decision resulted from the specific architecture of this model. The YOLOv7 backbone is built on the E-ELAN network, which is an architecture based on the concatenation mechanism, with all its consequences, both in terms of detection efficiency and inference time. On the other hand, the authors of YOLOv7 present several solutions in their work (Wang et al., 2022) aimed at optimising this type of architecture. As a result, YOLOv7 achieves results similar to YOLOv5 with a much shorter inference time.

4.2 Events Detection

Information about detected match objects is the basis for detecting more complex phenomena such as match events. These are then used to produce match highlights and calculate statistics. A separate Event Detector module is responsible for detecting match events, which accepts data in the form of XML or MOT files. The data in the files is an ordered sequence that stores information specific to each frame. The user can intervene in the event detection process using a separate panel in the web application (Fig. 3). It can manually indicate the frames that are the beginning and end of individual events, as well as mark the field lines.



Events Annotator

Events

TYPE OF EVENT: Shot On Goal

Current Frame ID: 0

EVENT POSITION
X-axis:
Y-axis:

EVENT DETAILS
Start Frame:
End Frame:
Event Details: From Game
Event Effectiveness: Effective
Team: A

Figure 3: The Event Editor window in the ISVP AI application.

The Event Detector module is built as an expert system based on rules that allow the detection of complex match situations. A separate detector is responsible for the detection of each type of match events, which marks single frames or pairs of frames that are the beginning and end of a given event. The parameters and configuration of the detectors were selected through experiments. Currently, the module can detect different types of shots (on target, defended, and missed), restart after a break in the game, and count ball possession statistics.

Event detection is performed sequentially, in accordance with the order of recorded frames (Fig. 4). Information about the frames is transferred to individual detectors, and based on them, they update their internal state and save their results in the object responsible for storing information about the state of the game. Some types of events are spread over time, which means that the corresponding detectors analyse not only individual frames but also the saved state of the game for previous frames.

The operation of a single detector can be demonstrated in the example of the Shot Detector for futsal, which is responsible for detecting the moment of starting the shot. It has the character of a finite automaton, which occurs in two states: waiting and investigation.

After receiving information about the frame, the detector checks if the acceleration of the ball has exceeded the value of 7.5 units:

$$\frac{p}{t^2} > 7.5 \tag{1}$$

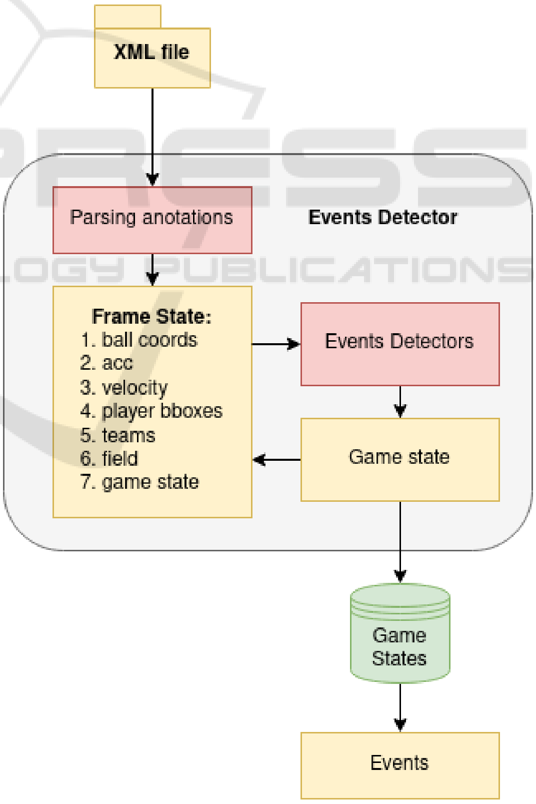


Figure 4: Diagram of the Event Detector data flow.

where p is the number of pixels and t is the frame duration.

If this threshold is exceeded and the ball is within the field of play and within 150 pixels of the player's bounding box, a shot factor is calculated.

$$s = \alpha\sqrt{c} \tag{2}$$

where α is the angle between the direction of the ball and the centre of the goal, and c is the distance of the ball from the centre of the goal.

If the shot factor is less than or equal to 15, the detector enters the investigation state, in which it checks whether there has been a change in the direction of the ball flight. The shot is detected if such a change has not occurred in the next three frames of the recording.

4.3 Final Recording

The final step in the production of a match recording is adding the effect of the camera following the action. For this purpose, the system generates an operator's path, within which it sets a point for each frame that is the centre of the simulated image. This point is calculated as a weighted average for the players' positions and the ball. In the vertical axis, the height of the camera is only calculated once during the match. It is the average between the player's highest and lowest position during the entire match, increased by 150 pixels. The finished path is the basis for simulating the camera movement. The maximum size of the viewport is 1600 x 900. Any camera shake is compensated for by the smoothing function in the camera module.

Regardless of the production of the final recording, it is possible to generate match highlights. The process is similar and includes operator path generation, camera movement simulation, and postprocessing. It is possible to generate match highlights for all, selected or random match events. When two events occur close to each other on the timeline, they are combined into one highlight. When preparing the operator's path, the system selects the recording speed and zoom level. The process of preparing the action summary proceeds in a similar way to the production of the final recording. After its completion, the postprocessing stage is also available, during which it is possible to add effects such as slowing down time.

The user has the option of interfering both in the production process of the final recording and in generating match highlights. In each of these cases, it is possible to manually determine the operator's path and save it to a JSON file, which can be used in the next stage of the system's operation. When generating action highlights, the user can also select start and end frames and adjust the video speed.

5 RESULTS

All of the YOLO model architectures we tested achieved satisfactory mAP results, both for futsal and handball. For futsal, the best results were achieved by the v5-FPN and v7 variants, between which there were slight differences for most classes (Tab. 2). The YOLOv5-FPN gained an advantage in detecting the ball, which seems to confirm the effectiveness of this architecture in the task of detecting small objects. The basic version of YOLOv5x obtained a slightly lower mAP than the other models, but the main problem of this variant was the long inference time. YOLOv7 was the best in this aspect, with an inference time of 19.4 ms. While the v5-FPN variant needed less than 7 ms more, the v5x performed calculations in almost twice as long (Tab. 4). In the case of handball, the v5x model achieved a mAP level similar to the other models, although in the case of a ball, its detections were clearly better (Tab. 3). However, this version of YOLO was still characterised by twice as long inference time.

Thus, it was the v5-FPN and v7 versions that turned out to be the most promising for use in the ISVP AI system. The large size of the v5x model resulted in a high computational cost, which translated into a long inference time and made it difficult to use this solution in the case of stream. At the same time, it was the large size of this model that allowed it to achieve better results in the field of ball detection in handball. This is probably due to the specificity of this discipline, in which the ball is often poorly visible on the recording. In these situations, the greater number of parameters in v5x allowed this model to better deal with partially obscured objects. The fast inference time, high mAP scores, and relatively simple v5-FPN architecture allowed for further experiments with the event detection module.

Since Event Detector is an expert system, it was necessary to gather and use the knowledge of specialists in the field of sports recording production to assess its operation. For this purpose, experts marked frame intervals on selected videos with various types of events. Then, for the same videos, the detection was carried out by individual detectors, and their results were compared with the answers of experts (hereinafter referred to as ground truth). It was accepted as True Positive for detections that had a common part with ground truth, greater than the assumed threshold value (20%). If there were more such detections, the first one was considered True Positive. The remaining detections were classified as False Positive. Then, for each detector, the Correctness (P) was computed:

$$P = \frac{Precision + Recall}{2} \tag{3}$$

In order to properly assess the effectiveness of the Event Detector module, an experiment using two sets

Table 2: Match object detection results for futsal (in mAP).

	total	players	goalkeepers	goal	referee	ball
YOLOv5x	83	87	70	99	65	85
YOLOv5-FPN	86	91	76	99	76	87
YOLOv7	85	92	82	99	72	82

Table 3: Match object detection results for handball (in mAP).

	total	players	goalkeepers	goal	referee	ball
YOLOv5x	91	95	96	99	88	75
YOLOv5-FPN	89	96	97	99	92	63
YOLOv7	90	96	98	99	92	64

Table 4: Comparison of inferential times (in ms).

	Futsal	Handball
YOLOv5x	54.0	54.5
YOLOv5-FPN	26.2	26.7
YOLOv7	19.4	17.7

of match object data: manually prepared by a human and performed by the YOLO algorithm was conducted. In result, it was possible to evaluate both the operation of the entire ISVP AI system and the match event detectors operating separately.

Table 5: Futsal results with manually marked match objects.

Detector	Recall	Precision	Corr.
Shot on target	0,82	0,69	0,76
Shot off target	0,62	0,48	0,55
Shot defended	0,69	0,28	0,48
Penalty	1	0,5	0,75
Game restart	0,85	0,92	0,88
Game restart	0,85	0,92	0,88

The obtained results are characterised by high variability. However, higher detection accuracy can be observed in cases where event detectors operate on manually prepared data (Tab. 5, 7). This means that even high results in terms of automatic detection of match objects (on average above 0.8 mAP) leave significant room for improvement (Tab. 6, 8). This is especially true for ball detection, which results in poorer performance in detecting different types of shots.

Table 6: Futsal results with automatically detected match objects.

Detector	Recall	Precision	Corr.
Shot on target	0,36	0,44	0,40
Shot off target	0,62	0,18	0,40
Shot defended	0	0	0
Penalty	1	0,33	0,67
Game restart	0,77	0,83	0,8
Ball ownership	0,58	0,68	0,63

Table 7: Handball results at manually marked match objects.

Detector	Recall	Precision	Corr.
Shot on target	0,72	0,32	0,52
Shot off target	0,37	0,15	0,26
Shot defended	0,34	0,21	0,28
Penalty	1	0,57	0,79
Game restart	0,83	0,41	0,62
Ball ownership	0,46	0,65	0,55

Nevertheless, the obtained results turned out to be sufficient enough to build a prototype system capable of fully automatic production of match recordings. The level of effectiveness of the match object detection allows for adding the effect of "camera follows the action". As a result,it is possible to automatically produce a full recording. The operation of the event detector, although it requires further development, already at this stage allows for the automatic generation of match highlights of some actions. It should also

Table 8: Handball results with automatically detected match objects.

Detector	Recall	Precision	Corr.
Shot on target	0,28	0,16	0,22
Shot off target	0,17	0,04	0,11
Shot defended	0	0	0
Penalty	0,63	0,71	0,67
Game restart	0,86	0,40	0,63
Ball ownership	0,41	0,75	0,58

be emphasised that the ISVP AI prototype does not require the construction of a dedicated camera, which in the future may significantly reduce the costs of production and purchase of such a system.

6 DISCUSSION AND FURTHER RESEARCH

The prototype of the ISVP AI system proves the effectiveness of using modern methods of detecting objects in systems to support the production of sports recordings. The obtained results also show that it is possible to fully automate this process. Nevertheless, the key result that was achieved is a potentially significant reduction in the cost of production of match recordings.

Thanks to the technology used, a system does not require the employment of additional employees or the purchase of specialised equipment. It is this potential difference in the cost of using ISVP AI that may have a positive impact on the practice of recording futsal and handball matches. Compared to other modern systems, the ISVP AI system manages to achieve better results in the match object detection metric, as well as create a completely new rule-based system for futsal and handball event detection.

Nevertheless, further development of the ideas presented here is still necessary. The obtained results indicate that even relatively high results of match object detection do not ensure high efficiency in detecting more complex events. However, this is crucial in building a virtual representation of the game state, which is necessary to implement the basic functionalities of the system. Hence, further research in this area should focus primarily on improving methods of combining the work of image processing algorithms with expert systems.

REFERENCES

Agile Sports Technologies, I. (2021). Efficient analysis: Record and upload automatically. https://www.hudl.com/en_gb/products/focus.

Ariki, Y., Kubota, S., and Kumano, M. (2006). Automatic production system of soccer sports video by digital camera work based on situation recognition. In *Eighth IEEE International Symposium on Multimedia (ISM'06)*, pages 851–860. IEEE.

Benjumea, A., Teeti, I., Cuzzolin, F., and Bradley, A. (2021). Yolo-z: Improving small object detection in yolov5 for autonomous vehicles.

Bochkovskiy, A., Wang, C. Y., and Liao, H. (2016). Optimal speed and accuracy of object detection.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S. (2020). End-to-end object detection with transformers. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, pages 213–229. Springer International Publishing.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255.

Everingham, M., Gool, L. V., Williams, C. K., Winn, J., and Zisserman, A. (2009). The pascal visual object classes (voc) challenge. In *International journal of computer vision*, pages 303–308.

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2013). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587.

Horvat, M. and Gledec, G. (2022). A comparative study of yolov5 models performance for image localization and classification. In *Central European Conference on Information and Intelligent Systems*, pages 349–356. Faculty of Organization and Informatics Varazdin.

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer International Publishing.

Quiroga, J., Maldonado, H., Ruiz, E., and Zapata, L. (2020). As seen on tv: Automatic basketball video production using gaussian-based actionness and game states recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 894–895.

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, page 28.

VEO (2021). Revolutionizing sports recording… again. https://www.veo.co/record.

Wang, C. Y., Bochkovskiy, A., and Liao, H. (2022). Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors.

Wang, C. Y., Yeh, I. H., and Liao, H. (2021). You only learn one representation: Unified network for multiple tasks.