2023 IEEE 31th Telecommunications Forum TELFOR Serbia, Belgrade, November 21-22, 2023

Soccer Artificial Intelligence Commentary Service on the Base of Video Analytic and Large Language Models

Roman V. Pavlovich
AtFrame DOO
Belgrade, Republic of Serbia
prv@atframe.rs

Evgeniya A. Tsybulko
COS&HT
Dolgoprudnii, Russian Federation
jen@cos.ru

Konstantin N. Zhigunov
AtFrame DOO
Belgrade, Republic of Serbia
kst@atframe.rs

Aleksandr V. Khelvas
COS&HT
Dolgoprudnii, Russian Federation
hel@cos.ru

Aleksandr A. Gilya-Zetinov
COS&HT
Dolgoprudnii, Russian Federation
entuser@yandex.ru

Illya V. Tykhonov
IPMCE
Moscow, Russian Federation
tixons.i.t@gmail.com

***Abstract*—Our article describes an approach to creating automatic commentaries for soccer games. The soccer match is logged by an artificial intelligence video processing component. This log is used to generate a football games commentary on the base of Large Language Models. The time limits of the generated comment are controlled by a separate neural network that ensures the ranking of episodes. The generated commentary text is spoken using the text-to-speak solution.**

Keywords: machine vision, soccer analytic, artificial intelligence, large language models

I. Introduction

The modern state of video analytic for soccer has seen significant advancements in recent years. Video analytic in soccer involves using computer vision and machine learning techniques to analyze video footage of matches and extract valuable insights and data.

Overall, video analytic has revolutionized soccer analysis by providing objective and data-driven insights into player performance, team strategies, and match events. It offers coaches, analysts, and fans a new level of understanding and appreciation for the game.

It's worth mentioning that video analytic in soccer is a rapidly evolving field, and there is still ongoing research and development to further improve the accuracy, speed, and sophistication of analysis techniques. As technology continues to advance, we can expect video analytic to play an even more prominent role in soccer, benefiting players, teams, and fans alike.

Top level of Large Language Models (LLM) and Artificial Intelligence (AI) success expertise permit us to analyze and interpret live sports events and provide commentary on them. This technology leverages natural language processing (NLP), LLM, machine vision (MV) and other AI techniques to recognize and interpret what's happening during the game and then convey that information to the audience in real-time as a set of voice messages.

There are several potential benefits to AI sport commentary. AI can provide more in-depth analysis and insights into the game. There are a few examples of companies that have attempted to use AI for sport commentary. One is IBM's Watson, which has been used to provide commentary for the Wimbledon tennis championships.

However, there are also some challenges associated with AI sport commentary. For example, it can be difficult for AI to capture the emotions, drama, and human-interest stories that make sports events so compelling to watch.

Overall, AI sport commentary is an interesting application of artificial intelligence technology, but it remains to be seen how effective it will be in replacing human commentators, who bring their own unique insights, personalities, and entertainment value to the broadcast.

II. Literature Review

The video analytic technologies state-of-the-art survey is presented in the article [1]. The automatic recognition of important events in soccer broadcast videos plays a crucial role in enhancing the analysis and viewing experience for soccer fans and some details about event recognition in broadcast soccer videos are discussed in [2].

Automatic event recognition in soccer broadcast videos democratizes access to insights and enhances the viewing experience. It saves time by pinpointing crucial moments

979-8-3503-0313-1/23/$31.00 ©2023 IEEE

and providing a summary of the game's most significant events. This technology benefits various stakeholders, including coaches, analysts, broadcasters, and fans.

However, it's worth mentioning that some challenges still exist. Variations in camera quantities, angles, video quality, occlusions, and the complexity of soccer events make accurate event recognition a difficult task. Nonetheless, ongoing research and advancements in computer vision techniques, deep learning models, and large-scale annotated datasets continue to improve the effectiveness and reliability of automatic event recognition systems.

The best and really popular book from the Professor of Applied Mathematics at the University of Uppsala David Sumpter [3] presents the most simple and effective metrics and approaches in soccer analytic.

The possibility of soccer game summarizing using audio commentary, metadata, and captions is proposed in the [4].

The goal of this work is to create an automated soccer game summarizing pipeline using AI. The focus is on generating complete game summaries in continuous text format while adhering to length constraints. This is done by leveraging various AI techniques, such as NLP tools and heuristics, along with available game metadata and captions.

The most famous text-to-speach solution is Amazone Web Services (AWS) service [5].

The ElevenLabs text-to-voice service https://elevenlabs.io is an other example of application for generation a realistic-sounding voice rather than a synthetic robotic voice.

In [6] there is discussed the copyright problem in context of AI generated content including the AI generated sport commentary. It seems like there are some complex issues being raised regarding copyright, attribution, and compensation for the use of original creations in training data and the resulting output of generative AI systems.

Following FIFA's 2015 approval of electronics performance and tracking system during games, performance data of a single player or the entire team becomes the base of traing process for the soccer industry.

The paper [7] presents video analytics, examines recent state-of-the-art literature in elite soccer, and summarizes existing real-time video analytics algorithms.

III. Methods and Implementation

In this section the total pipeline of AI commentary service and some details of this pipeline implementation are presented.

A. The pipeline first stage

The scheme of the pipeline first stage is presented on fig.1.

a) Preproccessing of camera stream: includes RAW2RGB conversion, Debayer and rough quality improvement.

b) Estimation of reciprocal cameras locations: Cameras on the field can move on rails and make rotations, which allows to get a complete picture of the stadium. This can help us obtain and recognize the boundaries of the field. Therefore, we have two coordinate systems: first one is associated with

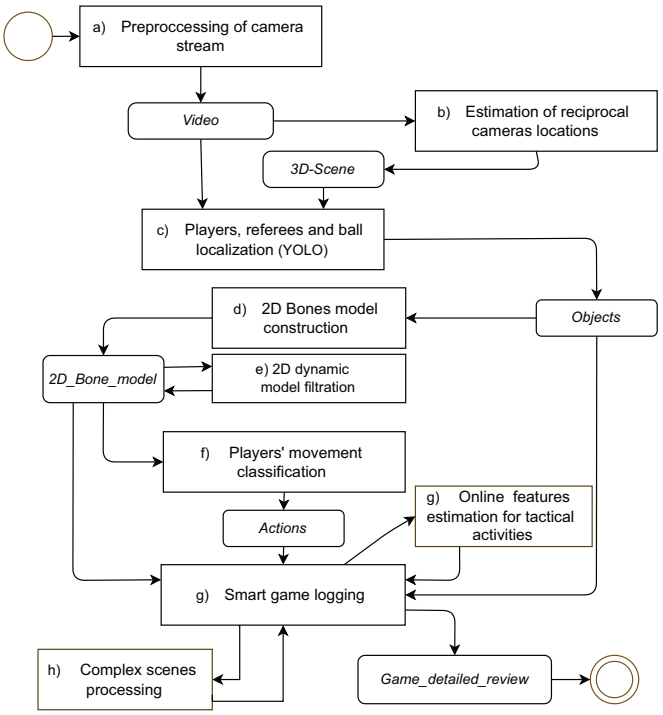

Fig. 1. Scheme for the first stage of pipeline

the field and its boundaries, second one is associated with the cameras and their positions. This will allow us to get the coordinates of the players and other objects in both coordinates systems.

c) Players, referees and ball localization: The next step is to find players and referees on the field for the further implementation of our work. To do this, we restricted the recognition area for the neural network to field area.

Then we used YOLO7 neural network [8] to identify the players and referees. As the result we got the coordinates of desired objects. Bounding boxes for players and referees are presented on fig.2.

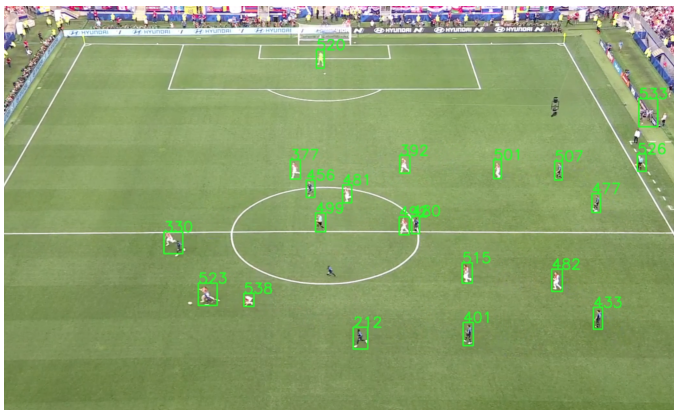

Fig. 2. Detection of players and ball on the tactical camera view

Recognition of the ball on the field is crucial for players'

actions tracking and smart match logging. To do this, we use a neural network that will monitor the location of the ball relatively to the field and players.

d) 2D Bones Model Construction: This step reduces complexity for future behaviour analysis. In order to do this we implement expensive DL algorithms for pose extraction.

The object obtained at the previous step is transformed into a "skeleton" using the solution developed by iPi soft (https://ipisoft.com/).

The "skeleton" consists of twenty points representing human's main joints and a polyline connecting these points.

As an output we want to get a 2D Skeleton Model in the camera frame coordinate system for further analysis of object's actions.

e) 2D dynamic model filtration: 2D dynamic model filtration is a way to avoid possible measurement errors by analyzing joint movements and to estimate the Improved 2D Dynamic Skeleton Model.

For example, if we get a physically impossible location of a joint point, we get rid of it or predict and use new the most likely location.

f) Players' movement classification: The previously obtained model is used for rough recognition the actions of players and interactions with the ball.

By changing the coordinates of each point individually and the coordinates in the aggregate, we can draw conclusions about the object movement (e.g. standing, running, slowdown, etc.). The combination of different actions helps to make important conclusions about a person's physical condition.

For a complete analysis of what is happening on the field we also need to distinguish and identify players.

To do this we can recognize the numbers of players, taking into account their movements from one camera to another. Anthropometric players data can also be used in future. After the player has been detected we can trace him for a set of frames (until the player can be recognized by given camera).

g) Smart match logging: Collecting features such as tracking the trajectory of objects, their identification, the ability to classify their actions, we can track what is happening on the field. This allows us to convert this information into a text that will accompany the match in real time (see fig.3). This simplest structured game description consists time stamp for events, events type and subtype, player ID and X/Y coordinates in percentages from the soccer field sizes.

So as a result we get the smart structured description for Game.

h) Online features estimation for tactical activities: Completeness of features listing for tactical activities allows us to most accurately transmit information about the game.

We want to recognize such tactical actions as acceleration of each point independently as well as of the body as a whole, movement harshness, rate of position change, hitting the ball, passing, hitting an opponent, kick the ball with a head, ball handling, etc.

To estimate tactical activities of players we estimate various parameters of "skeletons", for example acceleration,

9:45	PASS		Player4	83	1
9:49	BALL LOST	INTERCEPTION	Player8	97	28
9:50	RECOVERY	INTERCEPTION	Player21	93	39
9:51	PASS		Player21	91	35
9:52	PASS		Player19	83	31
9:53	PASS		Player16	91	35
9:55	PASS		Player21	80	12
9:56	PASS		Player18	83	13
9:59	BALL LOST		Player24	58	4
10:02	RECOVERY		Player2	33	13
10:02	PASS		Player2	33	13
10:06	BALL LOST	INTERCEPTION	Player11	16	52
10:09	CHALLENGE	AERIAL-WON	Player17	54	84
10:09	RECOVERY	INTERCEPTION	Player17	54	84
10:09	BALL LOST	HEAD	Player17	54	84

Fig. 3. Soccer Game structured description example

movement harshness, rate of position change, instead of row image data. This is a good opportunity for faster analysis.

B. The second stage of pipeline

Scheme on fig.4 presents the second stage of pipeline.

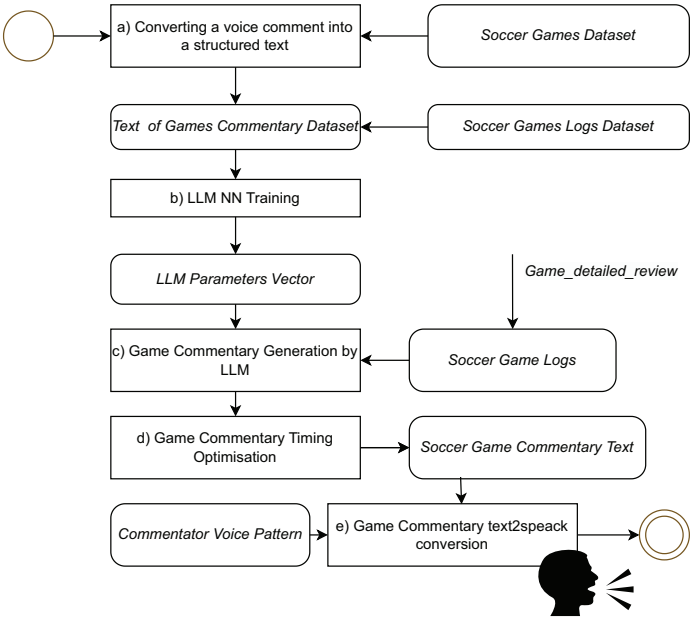

Fig. 4. Scheme for second stage of pipeline

a) Converting a voice comment into a structured text: First task is converting a voice commented soccer games dataset into a structured text comments with timestamps superimposed on the comments. This texts dataset are used for further LLM LLAMA [9] training and now is about 100 games.

This dataset was used for further neural network training.

b) LLM Neural Network training: The purpose of the second step is the conversion of a structured machine-readable commentary dataset in combination with games structured descriptions into LLM model parameters vector.

c) Game commentary text generation by LLM: The important step of proposed pipeline is generation of the game text comment optimized for a short time slot for each commentary element. This transformation time delay lead us to commentary delay and so we have to delay the game translation for several decades of seconds.

d) Game Commentary Timing Optimisation: Soccer game text commentary on this step is optimised for game timing. The structure of game timing is used for commentary elements texts size optimisation.

e) Game Commentary text2speak conversion: This step is implemented on the base of any existing text2speak services. The result is voice commentary for soccer game.

It is worth saying that a significant improvement in the quality of the proposed service is possible by integrating it with VAR solutions used by football referees. This will make it possible to more accurately determine formally recorded events on the football field: goals, violations, outs, offsides, etc.

IV. Conclusions

The proposed approach to soccer commentary generation on the base of AI/LLM technologies now is in the proove-of-concept state. But this approach lead us to the set of possible but not implemented advantages listed below:

1. Speed and Efficiency: AI can process video from several cameras and provide real-time analysis much faster than existing human commentators.

2. Unbiased Analysis: AI is not influenced by personal biases or emotions and can provide objective analysis based purely on data and facts.

3. Consistency: AI commentary does not get tired or lose focus, ensuring that the audience receives a consistent quality of analysis and insights from the beginning to the end of the match.

4. Enhanced Insights: AI can provide in-depth analysis by incorporating historical data, player statistics, and other relevant information.

5. Multilingual Capabilities: Our current efforts based on English datasets but AI has the potential to provide simultaneous commentary in multiple languages, making sports events accessible to a wider audience globally.

Implementing AI soccer games commentary service by telecom companies can enhance the sports viewing experience. Here are a few ways for telecom companies to integrate AI commentary:

Streaming Platforms: Telecom streaming platforms can use AI to provide real-time commentary alongside live soccer broadcasts. By integrating AI algorithms into their platforms, telecom companies can offer viewers the option to switch between human and AI-generated commentary including the analysis results.

Personalization: AI can analyze or ask user preferences to deliver personalized sports commentary.

Interactive Features: Telecom and broadcasting companies can incorporate interactive features into their soccer translation systems. For instance, viewers could ask questions or request specific insights during a live game, and the AI system could provide instant responses or additional statistics to enhance their understanding of the game.

Language Localization: Telecom companies can leverage AI to provide localized commentary in different languages. By using natural language processing, AI can automatically translate and generate commentary in real-time, allowing fans to enjoy sports events in their native or preferred language.

Second-Screen Experience: Telecom companies can develop companion apps or platforms that sync with the live soccer game and provide AI-generated commentary on users' mobile devices or smart TVs. This would enable viewers to access additional insights, statistics, and analysis while watching the game.

Social Media Integration: AI commentary can be integrated with social media platforms like Instagram, YouTube, TicTok etc., allowing viewers to share AI-generated insights, highlights, and analysis with their friends and followers in real-time. This would enhance the social aspect of sports viewing and foster engagement among fans.

By incorporating AI commentary into their services, telecom companies can revolutionize the way sports content is delivered, making it more personalized, interactive, and accessible to an audience. It would provide viewers with a unique content, while also opening up new revenue streams for telecom companies through enhanced services and user engagement.

References

[1] S. Akan and S. Varlı, "Use of deep learning in soccer videos analysis: Survey," *Multimedia Syst.*, vol. 29, no. 3, p. 897–915, dec 2022. [Online]. Available: https://doi.org/10.1007/s00530-022-01027-0

[2] H. Saraogi, R. A. Sharma, and V. Kumar, "Event recognition in broadcast soccer videos," in *Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing*, ser. ICVGIP '16. New York, NY, USA: Association for Computing Machinery, 2016. [Online]. Available: https://doi.org/10.1145/3009977.3010074

[3] D. Sumpter, *Soccermatics: Mathematical Adventures in the Beautiful Game*, ser. Bloomsbury sigma series. Bloomsbury Publishing Plc, 2016. [Online]. Available: https://books.google.ru/books?id=IwW0jwEACAAJ

[4] S. Gautam, C. Midoglu, S. Shafiee Sabet, D. B. Kshatri, and P. Halvorsen, "Soccer game summarization using audio commentary, metadata, and captions," in *Proceedings of the 1st Workshop on User-Centric Narrative Summarization of Long Videos*, ser. NarSUM '22. New York, NY, USA: Association for Computing Machinery, 2022, p. 13–22. [Online]. Available: https://doi.org/10.1145/3552463.3557019

[5] E. Amazon, "Spot instances https://aws.amazon.com/ec2/spot," *Last accessed*, vol. 24, 2022.

[6] P. Samuelson, "Generative ai meets copyright," *Science*, vol. 381, no. 6654, pp. 158–161, 2023. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.adi0656

[7] D. Jha, A. Rauniyar, H. D. Johansen, D. Johansen, M. A. Riegler, P. Halvorsen, and U. Bagci, "Video analytics in elite soccer: A distributed computing perspective," in *2022 IEEE 12th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, 2022, pp. 221–225.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2015, cite arxiv:1506.02640. [Online]. Available: http://arxiv.org/abs/1506.02640

[9] H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozière, N. Goyal, E. Hambro, F. Azhar *et al.*, "Llama: Open and efficient foundation language models," *arXiv preprint arXiv:2302.13971*, 2023.