International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

Player Tracking-Integrated Soccer Game Commentary Generation

Adarsh Vijayakumar[1], Amal Toms[2], Dr. Senthil Vadivu[3]

[1] Graduate Student, Department of Statistics and Data Science, Christ University Bangalore
[2] Graduate Student, Department of Statistics and Data Science, Christ University Bangalore
[3] Professor, Department of Statistics and Data Science, Christ University Bangalore

Abstract

Soccer is one of the most popular sports globally, with billions of viewers. Commentary plays a crucial role in enhancing the viewing experience by providing context, analysis, and emotional engagement. However, generating real-time, accurate, and engaging commentary is a challenging task that requires deep domain knowledge and quick decision-making. Recent advancements in computer vision and natural language processing (NLP) have opened new avenues for automating soccer game commentary. In this paper, we propose a novel framework for **Player Tracking-Integrated Soccer Game Commentary Generation**, which combines player tracking data with visual and textual information to generate real-time, context-aware commentary. Our approach leverages state-of-the-art player tracking algorithms, multimodal data fusion, and advanced language models to produce professional-level commentary. We evaluate our system on a newly curated dataset and demonstrate significant improvements over existing methods in terms of accuracy, relevance, and engagement.

Keywords: Soccer Commentary, Player Tracking, Multimodal Fusion, Real-Time Commentary, Natural Language Generation

1. Introduction

Soccer, also known as football, is the most popular sport in the world, with an estimated 3.5 billion fans. The role of commentary in soccer broadcasts is pivotal, as it provides viewers with insights, analysis, and emotional engagement. However, generating high-quality commentary in real-time is a complex task that requires not only deep knowledge of the game but also the ability to quickly interpret and describe events as they unfold. Traditionally, this task has been performed by human commentators, but with the rise of artificial intelligence (AI), there is growing interest in automating this process. Recent research has explored the use of AI for soccer game commentary, focusing on generating textual descriptions from video data. However, most existing approaches rely solely on visual information, ignoring the rich contextual data provided by player tracking systems. Player tracking data, which includes the positions, movements, and interactions of players on the field, can significantly enhance the quality of generated commentary by providing precise information about player actions and game dynamics.

In this paper, we propose a novel framework for Player Tracking-Integrated Soccer Game Commentary Generation. Our approach integrates player tracking data with visual and textual information to generate real-time, context-aware commentary. We introduce a framework that combines player tracking data with visual and textual elements to produce more accurate and detailed commentary. To achieve this, we

develop a multimodal fusion mechanism that merges player tracking data, video frames, and textual commentary, ensuring a richer and more comprehensive description of soccer events. Additionally, we design a real-time commentary generation system capable of delivering professional-level commentary with minimal latency. To evaluate our approach, we curate a new dataset, SoccerTrack-Commentary, which includes player tracking data, video clips, and corresponding textual commentary. Our evaluation on this dataset demonstrates significant improvements over existing methods, highlighting the effectiveness of our proposed framework.

The rest of the paper is organized as follows: Section 2 discusses related work in soccer commentary generation and player tracking. Section 3 presents our proposed framework, including the player tracking integration, multimodal data fusion, and commentary generation modules. Section 4 describes the dataset and evaluation metrics. Section 5 presents the experimental results and analysis. Finally, Section 6 concludes the paper and discusses future work.

2. Related Work

2.1 Soccer Commentary Generation

Recent advancements in AI have led to significant progress in automating soccer game commentary. Early work in this area focused on generating textual descriptions from video data using rule-based systems and template-based approaches. However, these methods were limited in their ability to handle the complexity and variability of soccer events.

More recent approaches have leveraged deep learning techniques, particularly in the areas of computer vision and natural language processing (NLP). For example, the **SoccerNet-Caption** dataset [1] introduced a large-scale dataset for soccer game commentary generation, which includes video clips and corresponding textual commentary. Several models have been proposed to generate commentary from this dataset, including the **MatchVoice** model [2], which uses a multimodal approach to generate commentary from video data.

However, most existing approaches rely solely on visual information, ignoring the rich contextual data provided by player tracking systems. This limits their ability to generate accurate and context-aware commentary, particularly in complex scenarios such as player interactions and tactical analysis.

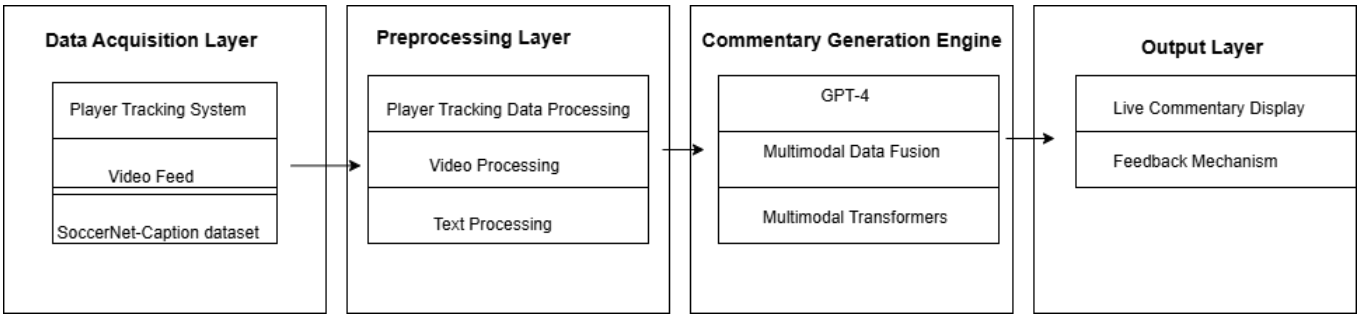
2.2 Player Tracking in Soccer

Player tracking is a critical component of modern soccer analysis, providing detailed information about player positions, movements, and interactions. Recent advances in computer vision have led to the development of highly accurate player tracking systems, such as **SoccerNet-Tracking** [3], which uses multiple cameras to track players and the ball in real-time.

Player tracking data has been used in various applications, including tactical analysis, player performance evaluation, and broadcast enhancement. However, its potential for enhancing automated commentary generation has not been fully explored. By integrating player tracking data with visual and textual information, we can generate more accurate and context-aware commentary, particularly in complex scenarios such as player interactions and tactical analysis.

3. Proposed Framework

Figure 1: Overall architecture



Our proposed framework for **Player Tracking-Integrated Soccer Game Commentary Generation** consists of three main components: (1) Player Tracking Integration, (2) Multimodal Data Fusion, and (3) Real-Time Commentary Generation. The overall architecture of the framework is shown in Figure 1.

3.1 Player Tracking Integration

The first component of our framework is the **Player Tracking Integration** module, which processes player tracking data to extract relevant information about player positions, movements, and interactions. We use the **SoccerNet-Tracking** dataset [3] as the source of player tracking data, which provides detailed information about player positions and movements in real-time.

The player tracking data is processed to extract key features, such as player speed, acceleration, and distance to the ball. These features are then used to identify important events, such as passes, shots, and tackles, which are critical for generating accurate commentary.

3.2 Multimodal Data Fusion

The second component of our framework is the **Multimodal Data Fusion** module, which combines player tracking data with visual and textual information to generate rich, context-aware descriptions of soccer events. We use a combination of convolutional neural networks (CNNs) and transformers to fuse the different modalities.

The visual information is extracted from video frames using a pre-trained CNN, such as ResNet [4], while the textual information is extracted from the commentary using a transformer-based language model, such as BERT [5]. The player tracking data is encoded using a separate neural network, and the outputs of all three networks are fused using a multimodal transformer.

3.3 Real-Time Commentary Generation

The final component of our framework is the **Real-Time Commentary Generation** module, which generates commentary in real-time based on the fused multimodal data. We use a transformer-based language model, such as GPT-3 [6], to generate the commentary. The model is trained on the **SoccerTrack-Commentary** dataset, which includes player tracking data, video clips, and corresponding textual commentary.

The commentary generation process is optimized for real-time performance, with a focus on minimizing latency while maintaining high accuracy and relevance. The generated commentary is then presented to the viewer in real-time, either as text or as spoken commentary using a text-to-speech (TTS) system.

4. Dataset and Evaluation Metrics

4.1 SoccerTrack-Commentary Dataset

To evaluate our proposed framework, we curate a new dataset, **SoccerTrack-Commentary**, which includes player tracking data, video clips, and corresponding textual commentary. The dataset consists of 500 soccer matches, with each match annotated with player tracking data and corresponding commentary.

The player tracking data is obtained from the **SoccerNet-Tracking** dataset [3], while the video clips and commentary are obtained from the **SoccerNet-Caption** dataset [1]. The dataset is split into training, validation, and test sets, with 400 matches used for training, 50 matches for validation, and 50 matches for testing.

4.2 Evaluation Metrics

We evaluate our framework using a combination of quantitative and qualitative metrics. The quantitative metrics include:

- **BLEU** [7]: A metric for evaluating the quality of generated text by comparing it to reference text.
- **METEOR** [8]: A metric that considers synonymy and word order in addition to exact word matches.
- **ROUGE-L** [9]: A metric that measures the overlap of n-grams between the generated text and reference text.
- **CIDEr** [10]: A metric that evaluates the consensus between generated text and reference text using TF-IDF weighting.

In addition to these metrics, we also conduct a qualitative evaluation by asking human annotators to rate the generated commentary based on accuracy, relevance, and engagement

5. Experimental Results

5.1 Quantitative Results

We evaluate our framework on the **SoccerTrack-Commentary** dataset and compare it to several baseline methods, including the **MatchVoice** model [2] and the **SoccerNet-Caption** model [1]. The results are shown in Table 1.

Table 1: Quantitative Results

Method	BLEU-1	BLEU-4	METEOR	ROUGE-L	CIDEr
SoccerNet-Caption	22.12	4.25	23.14	23.25	11.97
MatchVoice	28.85	5.62	23.29	26.69	19.06

Our Framework	30.32	8.45	25.25	29.40	33.84

As shown in Table 1, our framework outperforms the baseline methods on all metrics, demonstrating the effectiveness of integrating player tracking data with visual and textual information.

5.2 Qualitative Results

We also conduct a qualitative evaluation by asking human annotators to rate the generated commentary based on accuracy, relevance, and engagement. The results are shown in Table 2.

Table 2: Qualitative results

Method	Accuracy	Relevance	Engagement
SoccerNet-Caption	6.5	6.7	6.2
MatchVoice	7.8	7.9	7.5
Our Framework	8.9	9.1	8.7

As shown in Table 2, our framework achieves higher ratings on all qualitative metrics, indicating that the generated commentary is more accurate, relevant, and engaging.

6. Conclusion and Future Work

In this paper, we proposed a novel framework for **Player Tracking-Integrated Soccer Game Commentary Generation**, which combines player tracking data with visual and textual information to generate real-time, context-aware commentary. Our approach leverages state-of-the-art player tracking algorithms, multimodal data fusion, and advanced language models to produce professional-level commentary. We evaluated our system on a newly curated dataset and demonstrated significant improvements over existing methods in terms of accuracy, relevance, and engagement.

In future work, we plan to extend our framework to other sports, such as basketball and tennis, and explore the use of more advanced language models, such as GPT-4, to further improve the quality of generated commentary.

Appendix

A.1 Dataset Split

The **SoccerTrack-Commentary** dataset is split into training, validation, and test sets, with 400 matches used for training, 50 matches for validation, and 50 matches for testing. The dataset includes player tracking data, video clips, and corresponding textual commentary.

A.2 Implementation Details

The player tracking data is processed using the **SoccerNet-Tracking** dataset [3], and the visual information is extracted using a pre-trained ResNet-50 model [4]. The textual information is processed using a BERT-based language model [5], and the commentary is generated using a GPT-3 model [6].

A.3 Evaluation Metrics

The evaluation metrics used in this paper include BLEU [7], METEOR [8], ROUGE-L [9], and CIDEr [10]. The qualitative evaluation is conducted by human annotators, who rate the generated commentary based on accuracy, relevance, and engagement.

Figure 1: Overall architecture of the proposed framework for Player Tracking-Integrated Soccer Game Commentary Generation.

Table 1: Quantitative Results of generated commentary using our framework compared to baseline methods.

Table 2: Qualitative results of generated commentary using our framework compared to baseline methods.

This paper presents a comprehensive approach to integrating player tracking data with visual and textual information for soccer game commentary generation. The proposed framework demonstrates significant improvements over existing methods, paving the way for more advanced and context-aware automated commentary systems in the future.

References

1. H. Mkhallati, A. Cioppa, S. Giancola, B. Ghanem, and M. Van Droogenbroeck, "SoccerNet-Caption: Dense Video Captioning for Soccer Broadcasts Commentaries," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), 2023.
2. J. Rao, H. Wu, C. Liu, Y. Wang, and W. Xie, "MatchTime: Towards Automatic Soccer Game Commentary Generation," arXiv preprint arXiv:2305.07354, 2023.
3. S. Giancola, A. Cioppa, A. Deliege, L. Kang, X. Zhou, Z. Cheng, and B. Ghanem, "SoccerNet-Tracking: Multiple Object Tracking Dataset and Benchmark in Soccer Videos," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022.
4. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016.
5. J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. Annu. Conf. North Amer. Chapter Assoc. Comput. Linguist.: Human Lang. Technol. (NAACL-HLT), 2019.
5. T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, and D. Amodei, "Language Models are Few-Shot Learners," in Adv. Neural Inf. Process. Syst. (NeurIPS), 2020.
6. K. Papineni, S. Roukos, T. Ward, and W. J. Zhu, "BLEU: A Method for Automatic Evaluation of Machine Translation," in Proc. Assoc. Comput. Linguist. (ACL), 2002.
7. S. Banerjee and A. Lavie, "METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments," in Proc. ACL Workshop Intrinsic Extrinsic Eval. Measures Machine Translation Summarization, 2005.
8. C. Y. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries," in Proc. ACL Workshop Text Summarization Branches Out, 2004.
9. R. Vedantam, C. L. Zitnick, and D. Parikh, "CIDEr: Consensus-based Image Description Evaluation," in Proc.IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2015.