

(AI-based) Video Analysis and Processing: Master Thesis Project Description

Eirik Eggset

Purpose and Research Questions

The primary objective of this thesis is to leverage machine learning techniques for the processing of sports event video data, with a focus on enhancing efficiency and accuracy. A sports event contains several different event types, including but not limited to goal detection, penalties, big chances or interesting play. To detect these types of events the raw data is usually sent through a pipeline which breaks down the footage and analyzes it with different techniques, over multiple modalities [1].

The main benefit behind video analysis of this type of data is that a lot of the processing steps of going from raw video and sound data to sharing the processed data is manual work performed by humans [1]. The potential benefits of automating these processes are substantial, promising a faster and more cost-effective method of handling data. However, the effectiveness of such automated pipelines must meet a certain threshold. Addressing this, the research questions guiding this project include:

How can processing steps, traditionally performed by humans, be automated? This question can be scoped down to the individual steps of the pipeline. The goal is to improve the pipeline to give more accurate and low latency results.

Which of the steps of a video analysis pipeline for sports events pose the biggest bottleneck for accuracy? Considering the manipulation and analysis of data across various modalities and employing different algorithms/models, identifying critical steps impacting accuracy is crucial [2]. Detecting which step this can be very beneficial for improving the overall output of such a model.

By addressing these research questions, this seeks to contribute to the ongoing research in AI video analysis for sports content and offer practical insights for the development of more efficient and accurate processing pipelines.

State of video analysis today

In the current landscape of video analysis for sports content, several key areas are crucial for effective results. Different users have different interests and different platforms might have

different needs. Before being distributed the raw video data must be preprocessed in a variety of ways. What preprocessing steps which are to be used depends on the platform it will be distributed on as well as the user needs. For example, if the goal is to make a short summary of the whole match or just to display an interesting highlight. One example of such a pipeline is proposed in [1], leveraging a variety of different techniques.

Event detection

One of the most important steps when creating a video of a sports event is to decide which parts are the most interesting and relevant. Some highlights include for example goals, a player dribbling, or a penalty kicks. A challenge with automatic event detection today is that techniques often suffer from high latency and low accuracy [1].

Clipping

Clipping the video to extract the highlights is one of the most expensive operations in video distribution for sports events [1]. This is because this is often done manually by a human. By having a pipeline to automatically clip videos, while still maintaining the quality a human produce, a significant amount of time and expenses can be saved. There are also benefits for the quality of the output, as cropping can help simplifying the complexity of the data, by isolating regions of interest [2].

Cropping

There are multiple ways to share and show a videoclip of a sports event. Different platforms might have different standards or required formats. When choosing how to adapt the cropping from one aspect ratio to other leads to an increased end-user quality of experience [2]. When cropping it is important to preserve the essence of the original video and keep the most important points of interest.

Cropping the data can also be beneficial for the accuracy of the pipeline [2]. This is because it lowers the complexity of the data by isolating areas of interest. This is especially important for sports events where the objects move at a fast pace as well as rapid camera changes.

Thumbnail generation

To interest and engage viewers a thumbnail which captures the essence of the video is expected [1]. Different techniques can be used for this, including logo detection, close-up shot detection and image quality analysis.

Summarization

Creating a summary of the video content is something which can be useful for users. To manage this, you must analyze the video and output text. This can be achieved by relying on a multimodal scene understanding, utilizing multiple modality specific encoders [1]. There are

multiple different factors to take into consideration when deciding on which modalities and techniques to consider.

Thesis scope and objectives

Building on the current foundation laid by the current state of video analysis for sports events, the initial step is to detect which of the components of the current video analysis pipelines for event detection poses the biggest bottleneck. Once this has been identified the objective is to propose solutions or improvements to enhance the overall results.

These improvements could involve experimenting with different models/techniques. Given the complexity of the fast-paced sports content, selecting a suitable processing algorithm/model is crucial for achieving optimal results. Sometimes trying a different type of model could be beneficial for capturing the most significant aspects of the footage. Alternatively, in cases where the model is effective, fine-tuning or adjusting hyperparameters to match the specific type of data may result in a more efficient outcome.

Bibliography

- [1] C. Midoglu, S. S. Sabet, M. H. Sarkhoosh, S. M. M. Dorcheh, S. Gautam, T. Kupka and P. Halvorsen, "AI-Based Sports Highlight Generation for Social Media".
- [2] S. M. M. Dorcheh, M. H. Sarkhoosh, C. Midoglu, S. S. Sabet, T. Kupka, M. A. Riegler, D. Johansen and P. Halvorsen, "SmartCrop: AI-Based Cropping of Soccer Videos".