Highlight Detection in Soccer Matches

Chiotis Nikolaos, Kyriazopoulos Christos, Stavrianoudakis Vassileios

MTN2221, MTN2207, MTN2215

MSc Artificial Intelligence

June 30, 2023

Contents

Current Section

Requirements

- pytorch
- librosa
- pyAudioAnalysis
- moviepy
- sklearn
- yt dlp
- opencv

Problem Definition

- Extract highlights from footbal matches and create summary
  - Make use of 3 modalities
    - Visual
    - Audio
    - Text

- Assumptions made
  - Resonable Assumption –> Distance based(outlier detection) approach
  - Light Assumption –> highlight duration is X seconds (10s in our case)
  - Strong Assumption –> k highlights per match

Current Section

Methodology Visual

- First attempt –> Video transformer model(didn't work)
- Extract 1 frame/10s
- Use pretrained model (resnet18) to extract features
- Use pretrained model to extract field coverage feature
- Combine feature vectors
- Calculate cosine distance of feature vectors
- Calculate score based on distance and rank segments

Figure 1: Field Coverage Filter

Methodology Audio

- Audio segmentation (10s segments) with ffmpeg
- Feature extraction with pyAudioAnalysis
- Feature selection using a variance thresholding –> Keep features with higher variance
- Calculate cosine distance of feature vectors
- Calculate score based on distance and rank segments
- Further improvements showed in experiment section

Methodology Text

- Extract transcription from audio segments using pretrained model(whisper)
- Extract speech rate for each segment
- Use pretrained model to extract sentiment scores from text segments

  Joy, Sadness, Anger, Disgust,
  Neutral, Surpise, Fear

- Combine features in one vector
- Calculate cosine distance of feature vectors
- Calculate score based on distance and rank importance of segments

Current Section

Experiment 1

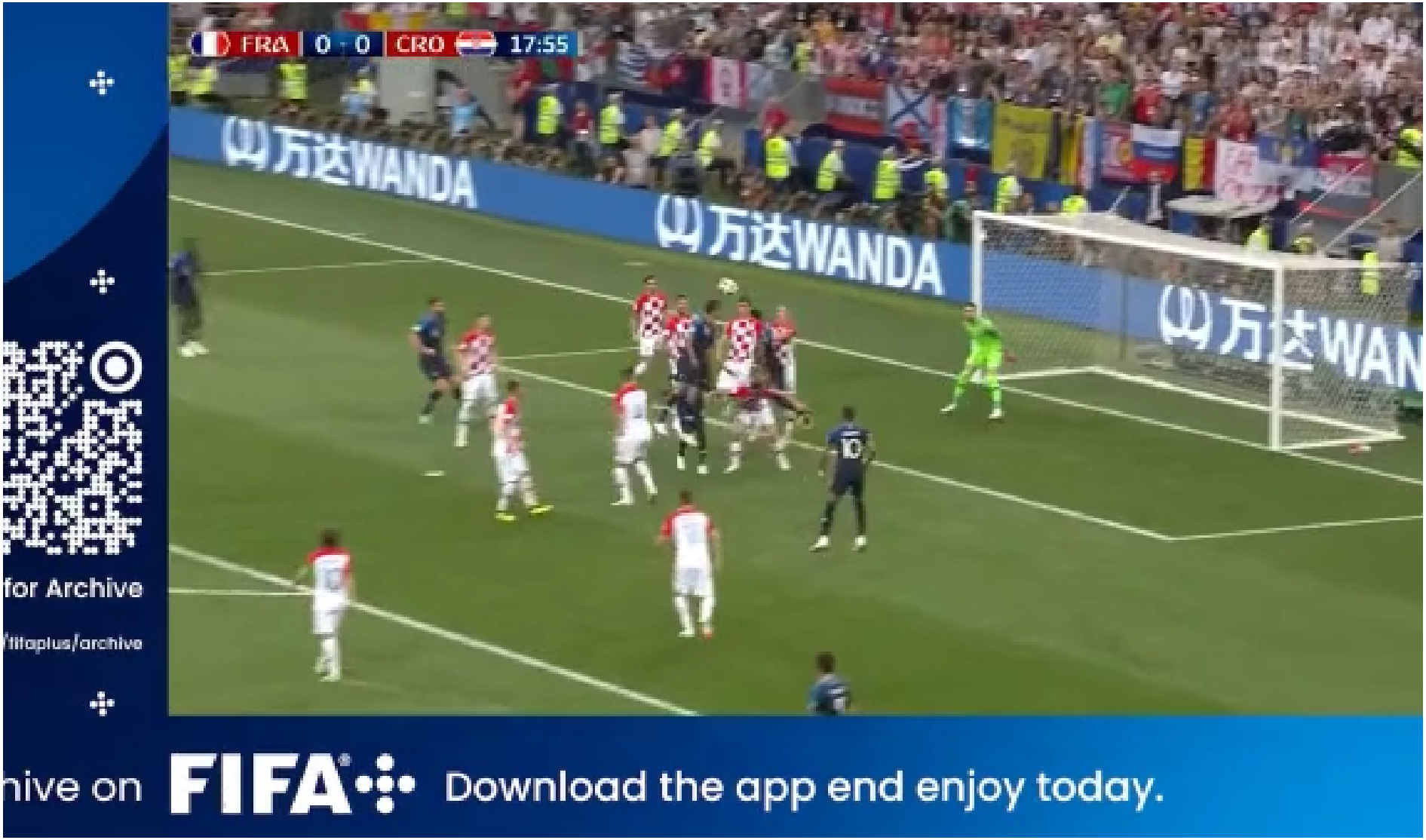
- Get scores calculated from image and audio
- Calculate mean score for each segment
- Select k most distant segments
- for k=10 approximately 5 out of 10 where actual highlights

Experiment 1 Results

Experiment 2

- Get scores calculated from image and audio
- Thresholding based on mean audio amplitude –>
  - calculate mean amplitude of all segments
  - relevant segments –>
    mean segments amplitude >30% mean overall amplitude –>
    dummy silent segment removal
- Calculate mean score for each relevant segment
- Select k most distant segments
- for k=10 approximately 7 out of 10 where actual highlights

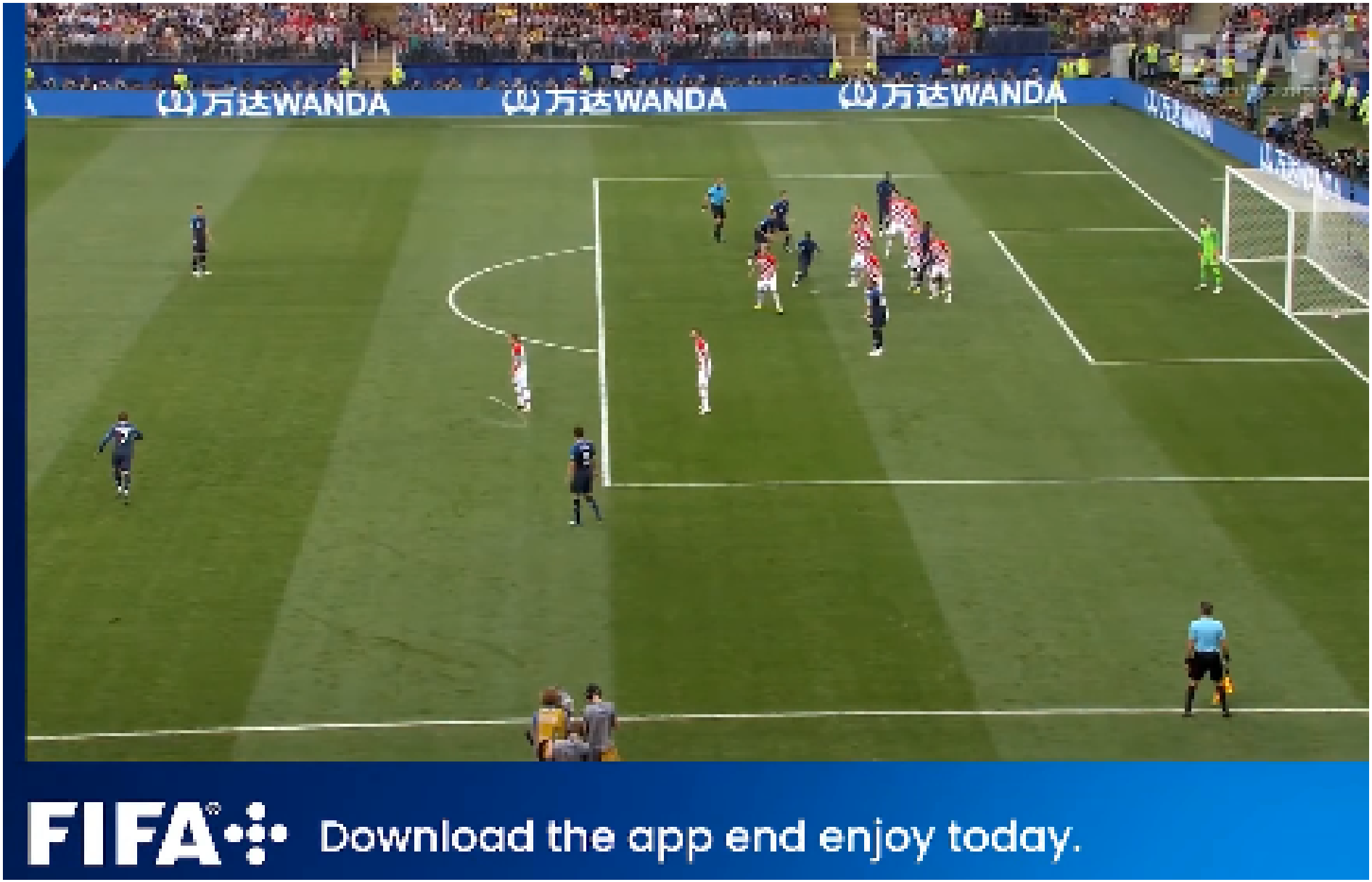
Experiment 2 Results



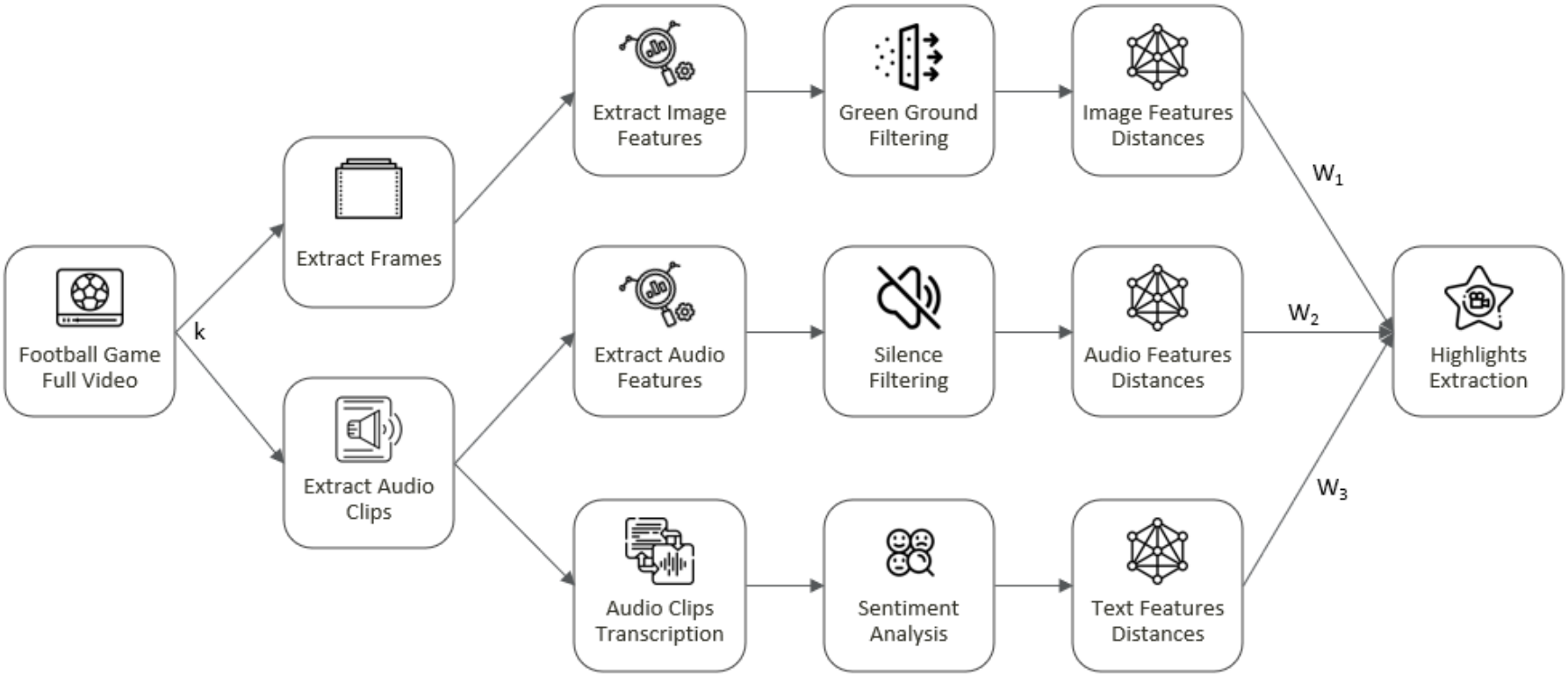
Experiment 3

- Get scores calculated from image, audio and **text**
- More strict thresholding with mean audio amplitude
- Calculate mean score for each segment
- Select k most distant segments
- for k=10 approximately 8 out of 10 where actual highlights

Experiment 3 Results



Final Pipeline



Evaluation

- Selected 5 matches
- Benchmarking with actual highlights compared to our highlights
- Used google form

	Game1	Game2	Game3	Game4	Game5
Score	3	4	3.5	2.5	3

Current Section

Most Important Results + Future Work

Conclusions

- Pros
  - Successful use of completely unsupervised technique.
  - Zero training needed.
  - Simple hand-crafted rules for highlight selection
  - Inference time 20min (5 min without text modality)
  - k and X parameter are costumizable
- Cons
  - Wave-handy thresholding selection
  - highlight duration parameter somewhat restrictive
  - k parameter is very restrictive

Future Work

- More consise and explainable thresholding definition
- Enforce thresholding using other modalities except audio
- Make use of domain knowledge to extract most meaningfull features for each modality

Thank You