CVSPORTS CVPR2017

IBM

Automatic Curation of Golf Highlights using Multimodal Excitement Features

Michele Merler   Dhiraj Joshi   Quoc-Bao Nguyen   Stephen Hammer   John Kent   John R. Smith   Rogerio S. Feris

IBM H5 Masters Highlights

MASTERS

TOTAL CLIPS PROCESSED: 866   HOURS OF COVERAGE: 114

0.78 OVERALL EXCITEMENT LEVEL

0.14 COMMENTATOR EXCITEMENT

1.00 ACTION RECOGNITION

1.00 CROWD CHEERING

CURRENT TIME: 1:34 PM   CLIP TIME: 2:15 PM APRIL 9, 2017

HOLES 15 & 16: LOUIS OOSTHUIZEN m HOLE

COMMENTARY:

2:15 PM APRIL 9, 2017

HOLES 15 & 16

0.78

EXCITEMENT LEVEL

2:11 PM APRIL 9, 2017

BROADCAST

0.18

EXCITEMENT LEVEL

2:11 PM APRIL 9, 2017

BROADCAST

0.24

EXCITEMENT LEVEL

2:11 PM APRIL 9, 2017

BROADCAST

0.51

EXCITEMENT LEVEL

The team

Michele Merler

Rogerio S. Feris

Dhiraj Joshi

John R. Smith

Stephen Hammer

John Kent

Quoc-Bao Nguyen

Main Contributions

First-of-kind system for automatically extracting golf highlights

uniquely fuse multimodal excitement measures from the player, spectators, and commentator

Novel techniques for learning multimodal classifiers

without costly manual training data annotation

Self-supervised learning framework for player recognition

build rich feature representations without manually annotated training examples

Live demonstration of the system at a major golf tournament

processing live streams and extracting highlights from 4 channels during 4 consecutive days

IBM

HIGHlights
From
Intelligent
Video
Engine

Framework
Multimodal
Excitement
Markers

Experiments and
Evaluation

Conclusions

90 players

18 holes

4 days

100s

hours of footage

IBM

IBM

HIGHlights
From
Intelligent
Video
Engine

Motivation

Framework
Multimodal
Excitement
Markers

Highlights Detection

H5

Self-Supervised Player Learning

Experiments and Evaluation

Demo

Conclusions

Proposed Framework

Input Video → Multimodal Marker Detection → Highlight Start/End Frame Detection → Highlight Video Segments

Commentator

Spectators

Player

Audio Analysis

Crowd Cheering

Commentator Voice Tone

Speech2Text

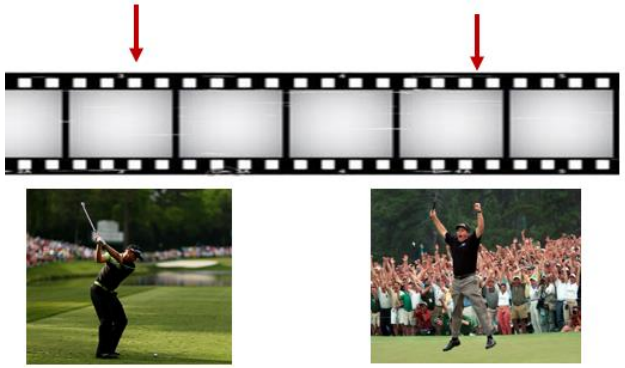
Visual Analysis

Player Action Recognition

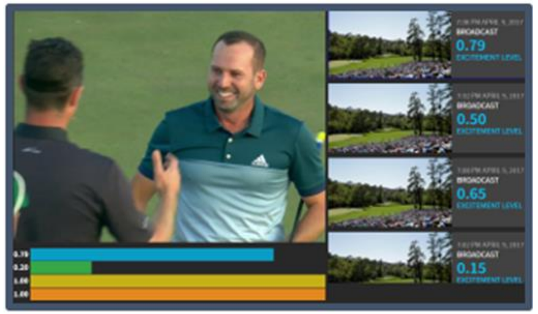
TV Graphics Detection

OCR

Shot-boundary Detection



Highlights with excitement score and other metadata



Overall Excitement

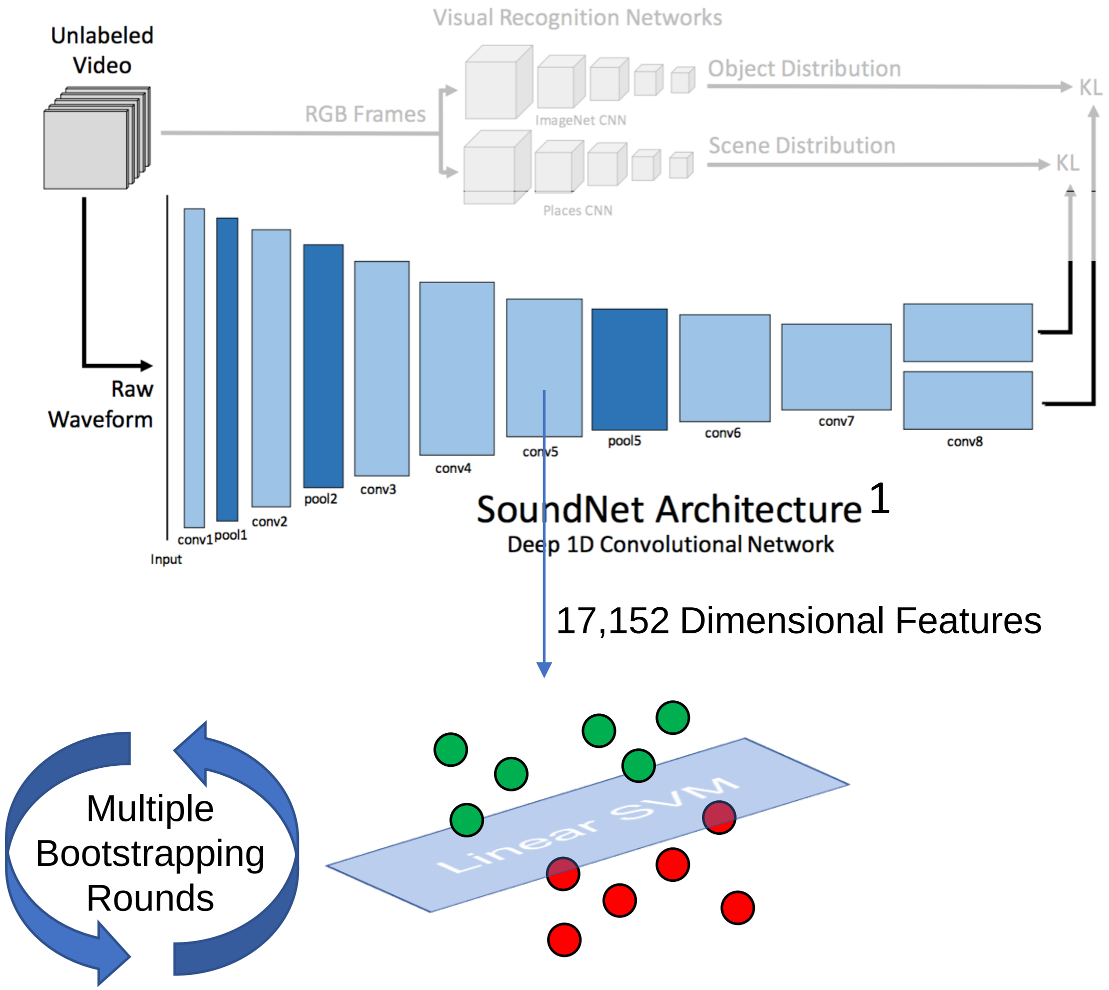
Commentator Excitement

Action Recognition

Crowd Cheering

IBM

Audio Based Markers for Excitement Detection



Crowd Cheering

Commentator Tone

- ✓ 6 seconds segments
- ✓ 16 bit PCM encoding at rate 22,050Hz
- ✓ Bootstrapping of Linear SVM on top of Deep features
- ✓ Training data from 2016 Masters + Youtube
- ✓ Leave One Out Cross-validation Accuracy
  - Crowd cheering **99.4%**
  - Commentator tone **81.3%**

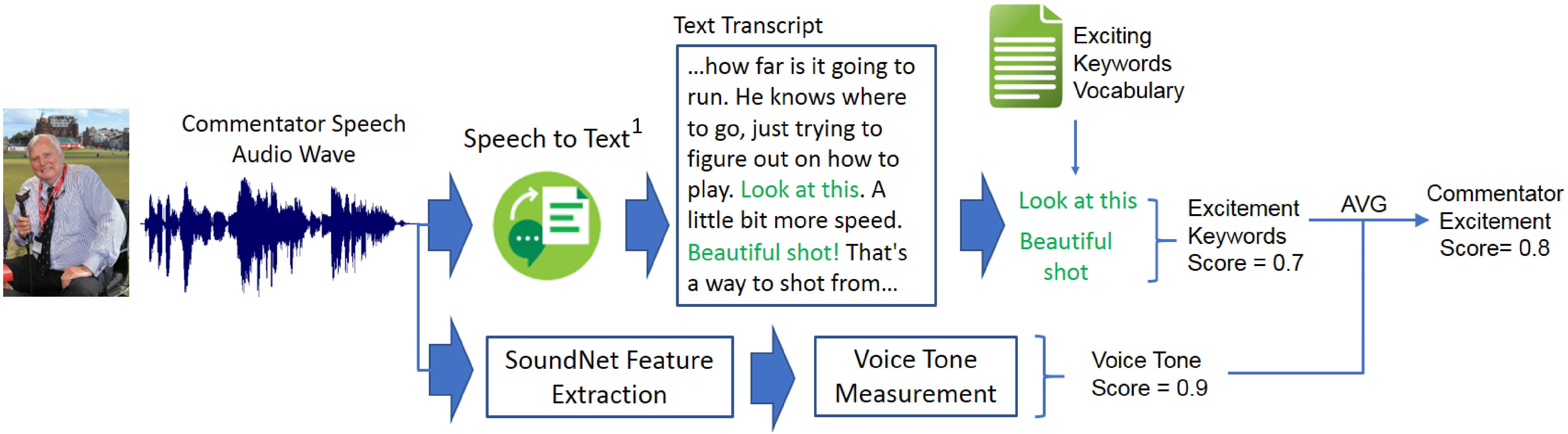
1. Y. Aytar, C. Vondrick, and A. Torralba. Soundnet: Learning sound representations from unlabeled video. In *NIPS*, 2016.

Audio Based Markers for Excitement Detection

Commentator Excitement

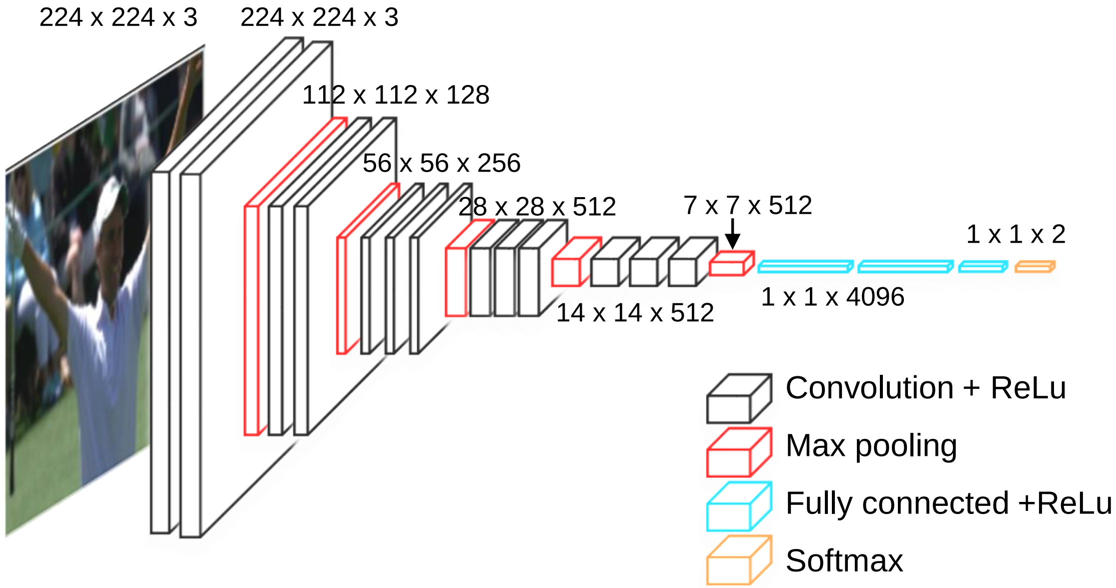
Combination of Audio based Commentator Tone Model + Recognized Excitement Keywords Model

Audio based Model drives selection of which segments to analyze



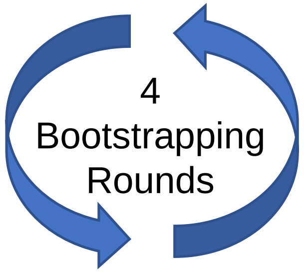
1. https://www.ibm.com/watson/developercloud/speech-to-text.html

Visual Based Markers for Excitement Detection



Player Action of Celebration

- ✓ Audio-based classifiers drive selection of video segments to annotate
- ✓ VGG-16[1] model pretrained on ImageNet
- ✓ Initial training set: **~1K images**



Hard negative mining over ~50 hours of video (Youtube + 2016 Masters)

- ✓ Final training set: **~10K images**
- ✓ **88%** accuracy on a ~1,300 images test set

1. K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv* preprint arXiv:1409.1556, 2014

IBM

HIGHlights
From
Intelligent
Video
Engine

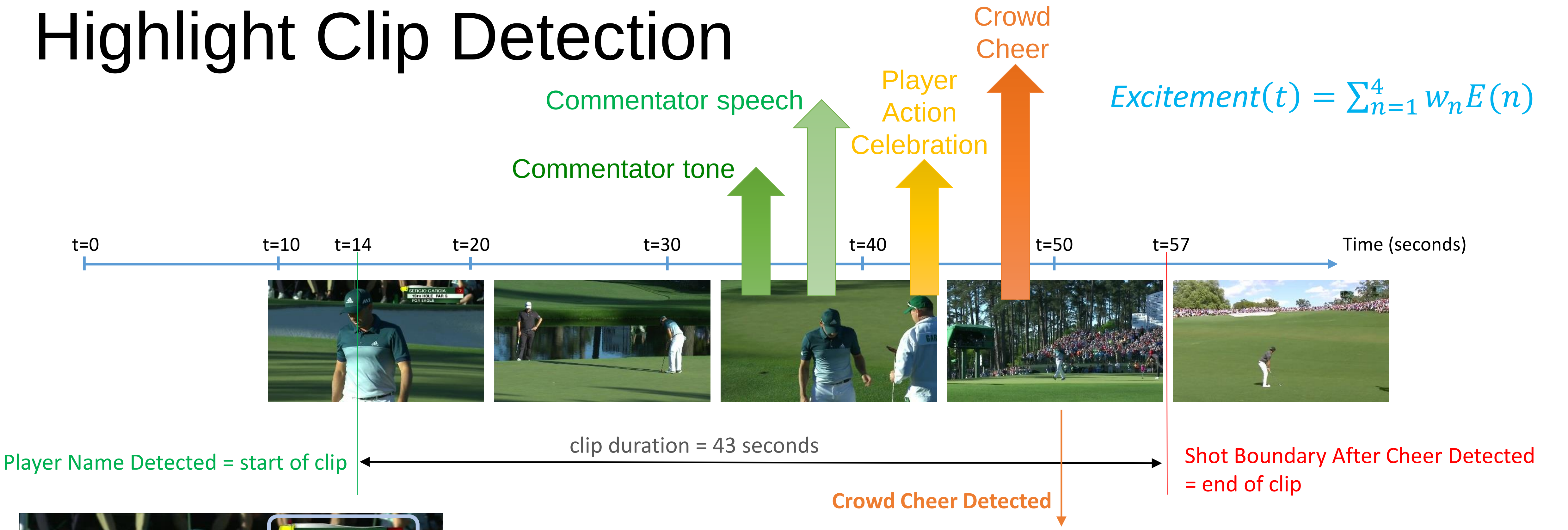
Framework
Multimodal
Excitement
Markers

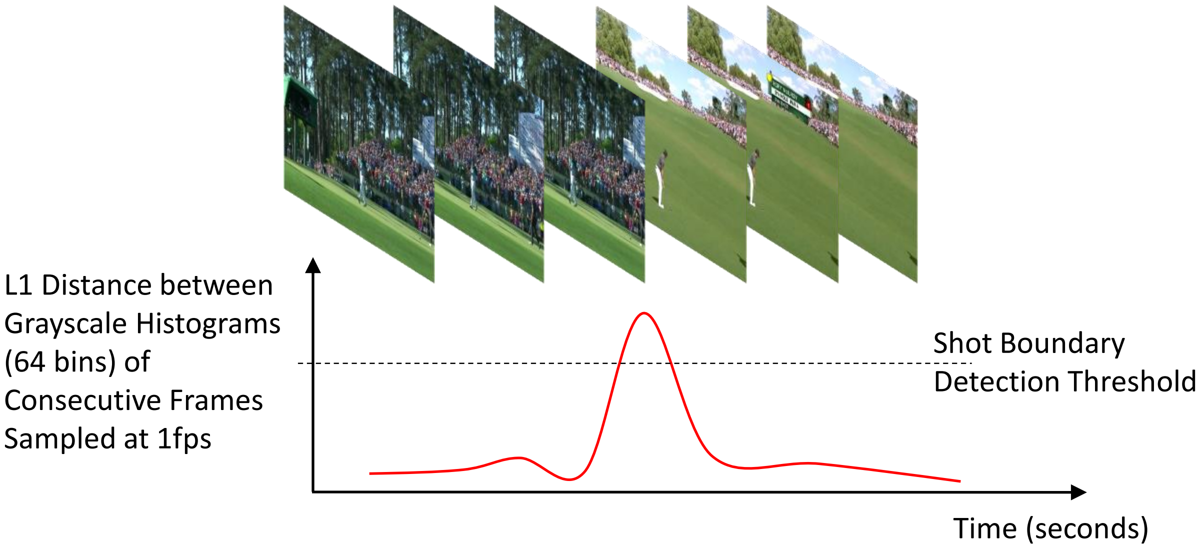
H5

Experiments and
Evaluation

Conclusions

Highlight Clip Detection





- Graphics + Text Detection (based on logo color and position)
- OCR Engine applied to text region
  Recognized text: SERGIO GARCIA 15TH HOLE PAR 5 FOR EAGLE

IBM

HIGHlights
From
Intelligent
Video
Engine

Framework
Multimodal
Excitement
Markers

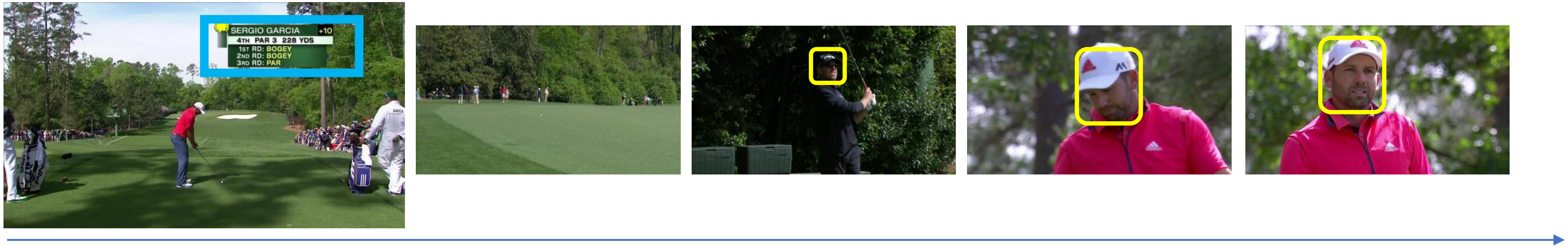
H5

Experiments and
Evaluation

Conclusions

Self-Supervised Player Face Learning

Face of a player very likely to appear (temporally) close to graphics with his name



time

- ✓ Collect faces detected by faster RCNN[1] model
- ✓ Perform 2-class clustering on top of VGG16-Face[2] fc7 features to eliminate noise
- ✓ Automatically collect hundreds of face images per player **NO SUPERVISION REQUIRED**

1. S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In NIPS, 2015
2. O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In BMVC, 2015.

IBM

HIGHlights
From
Intelligent
Video
Engine

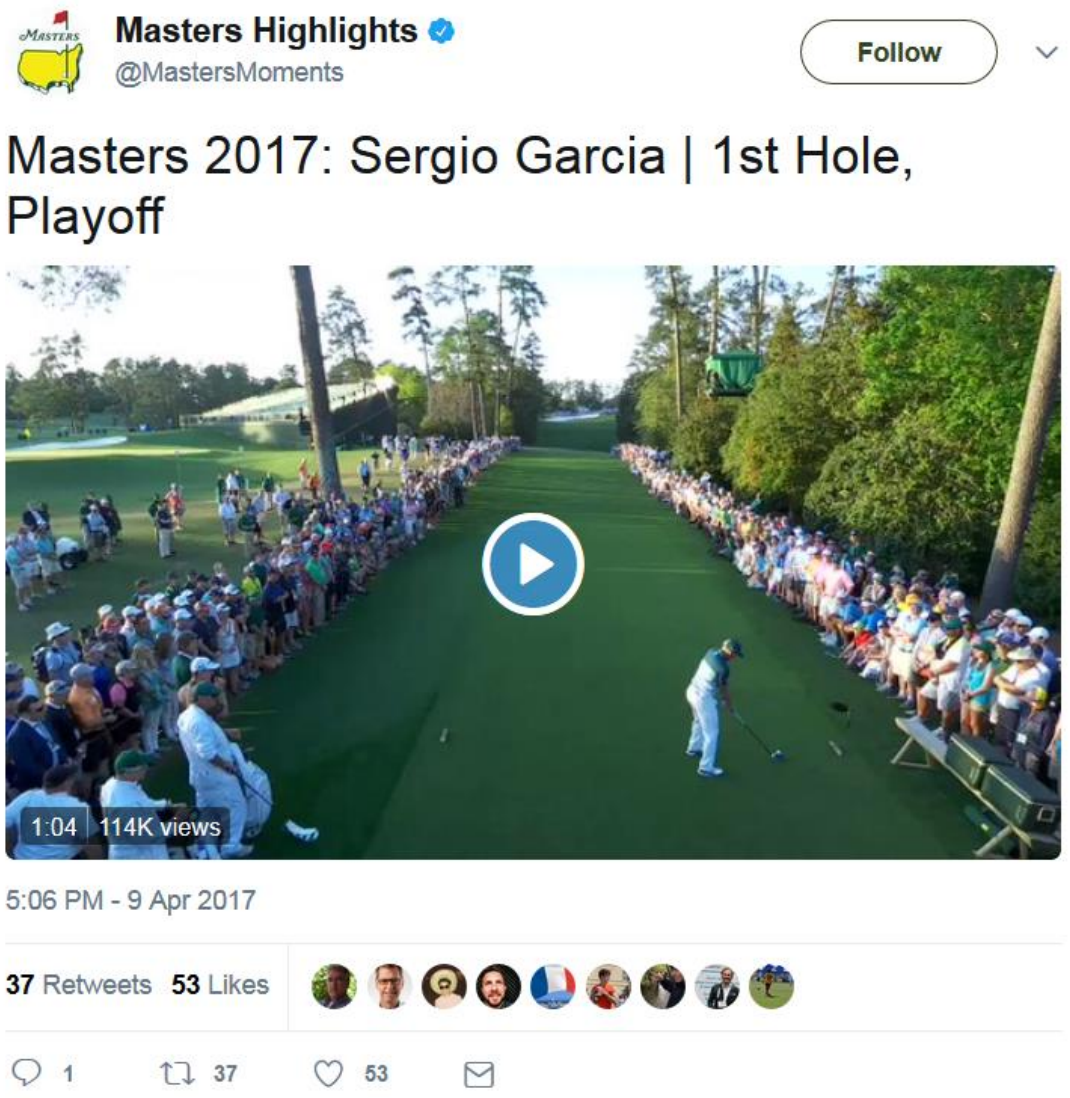
Framework
Multimodal
Excitement
Markers

H5

Experiments and
Evaluation

Conclusions

Highlights Evaluation

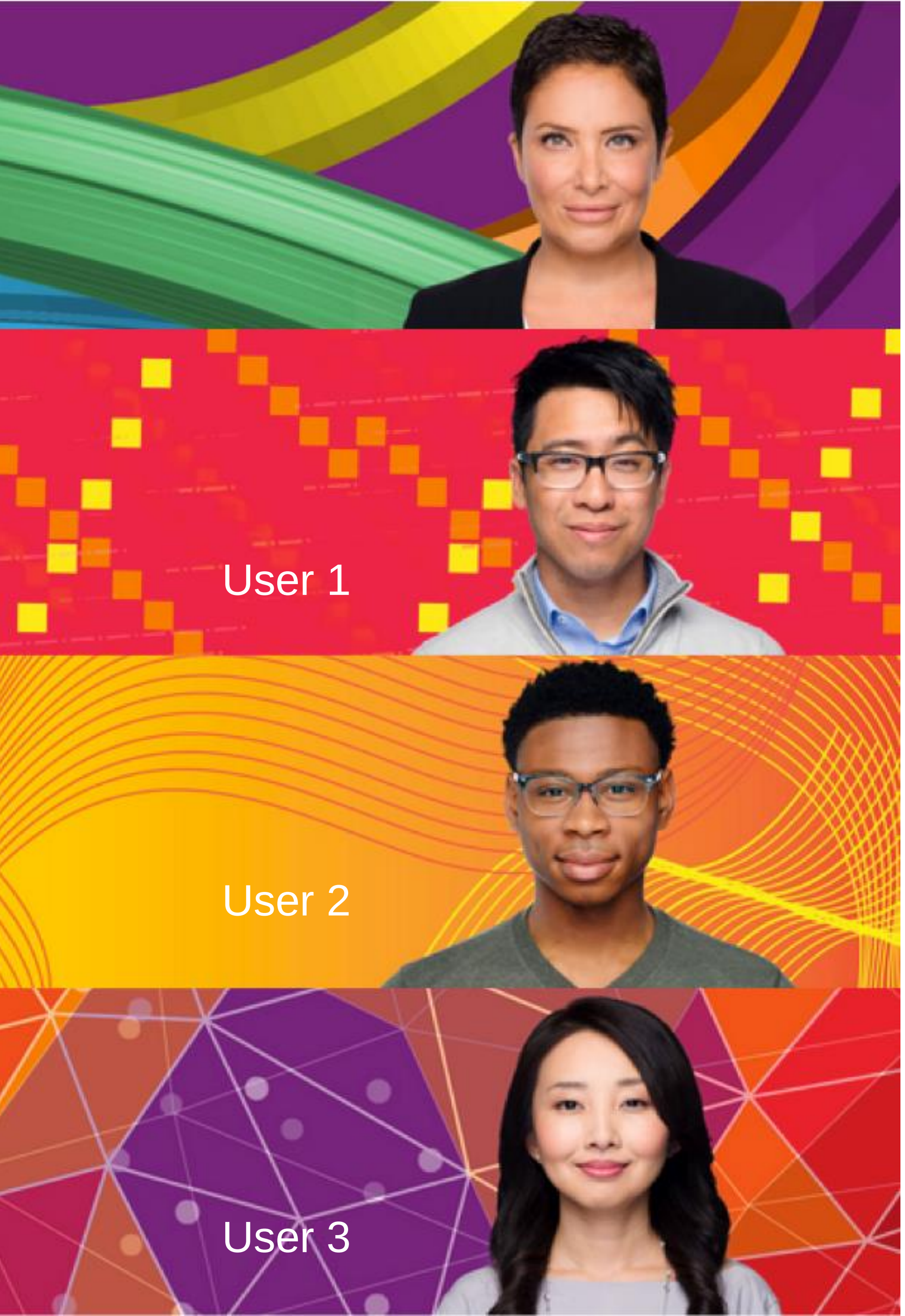


Human Evaluation of Highlights Rankings

Comparison with Official Master Highlights

Highlights Evaluation 1

Human Evaluation on Highlights Rankings



How Exciting is this clip?

5 Extremely exciting

4 Very exciting

3 Exciting

2 Moderately exciting

1 Highlight of wrong player

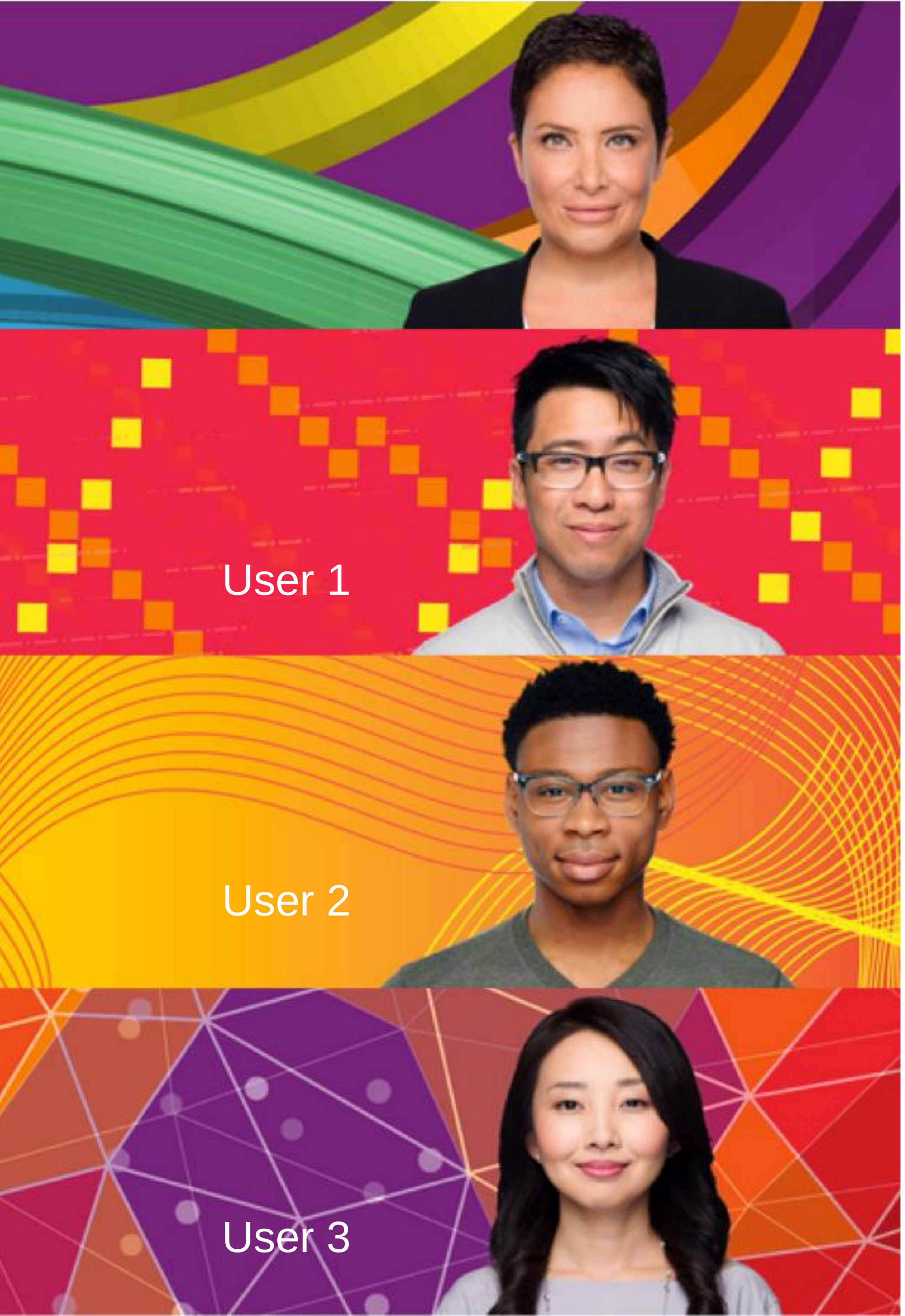
0 Not a highlight/ wrong clip

5

4

4

4.33



Highlights Evaluation 1

Human Evaluation on Highlights Rankings

Research Questions

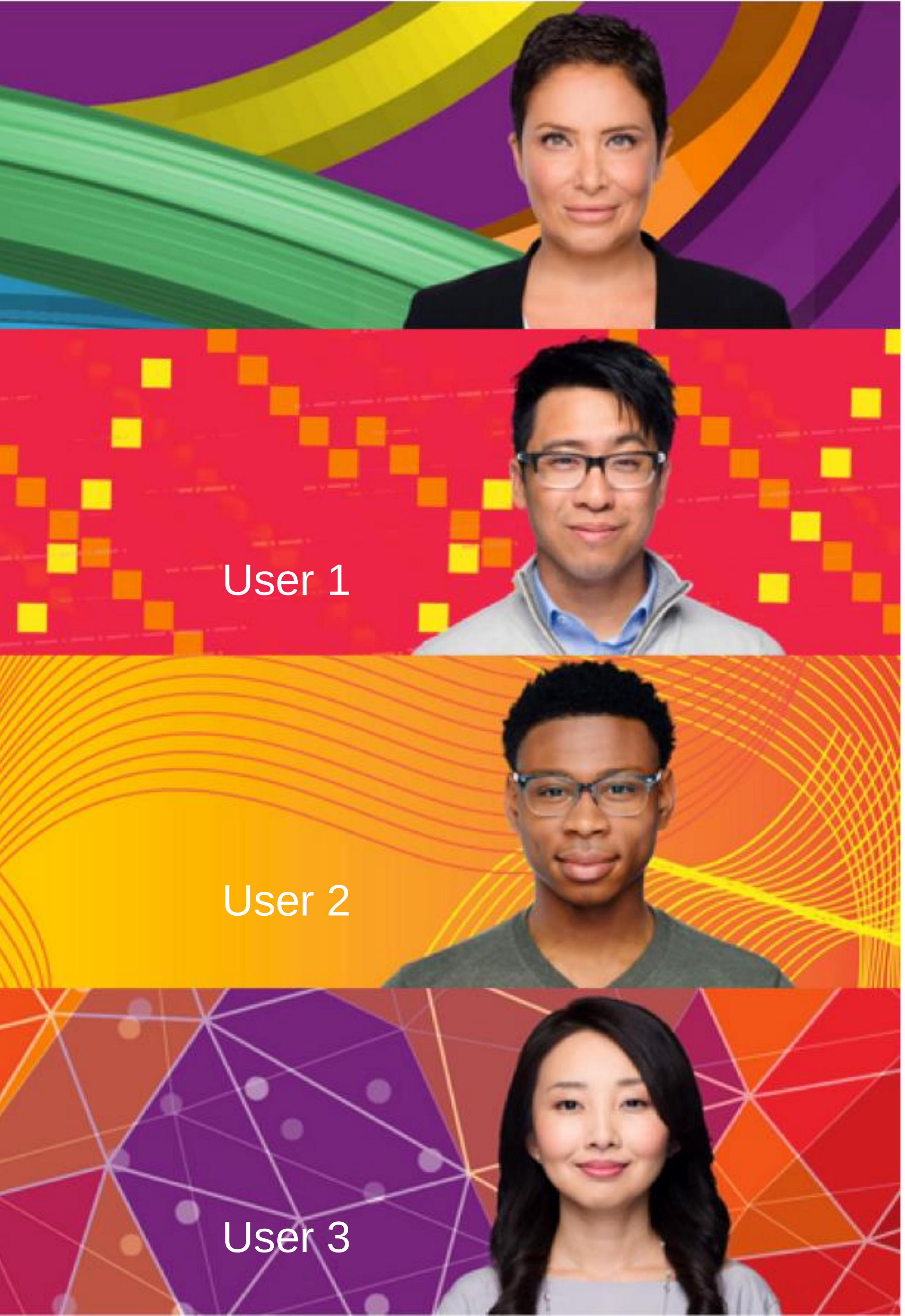
- ✓ What is the perceived quality of the clips produced by the system?
- ✓ What is the relative importance of each component?

Experiments Setup

- ✓ Top 120 clips produced by the system on Day 4 of the Masters
- ✓ 3 Users asked to rate clips on a scale of 0 to 5
- ✓ *nDCG* between users ranking and ranking of each modality + fusion

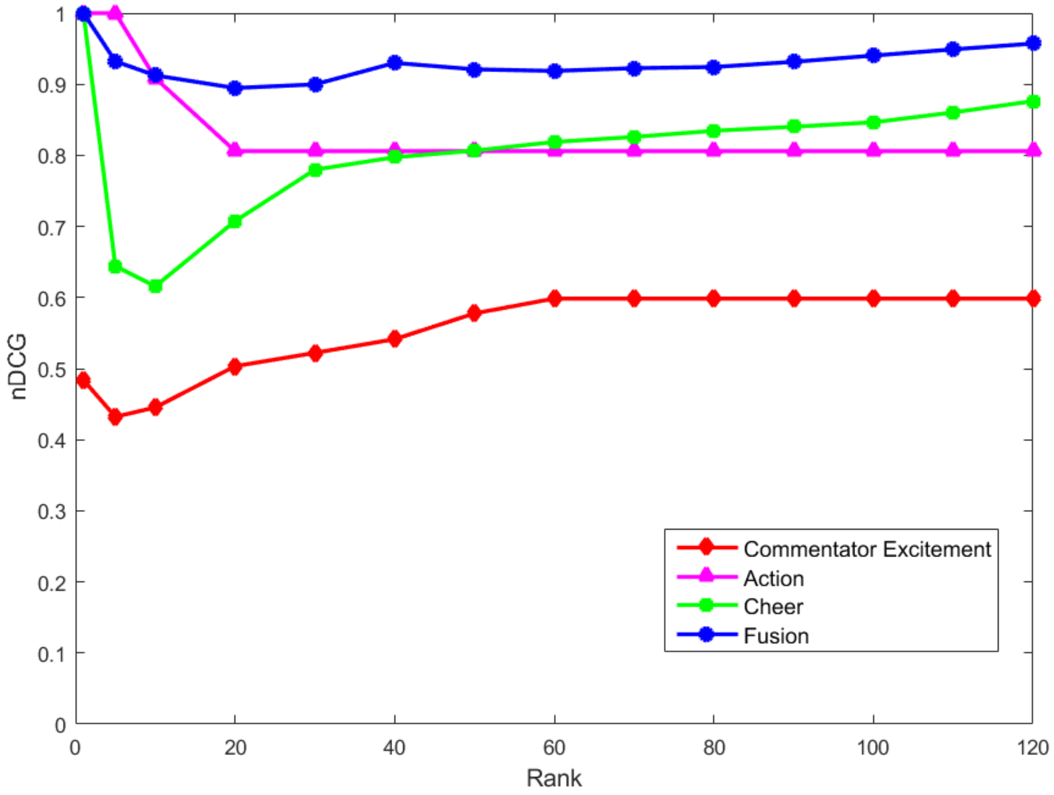
Results

- ✓ **92.7%** of clips scored 2 or above
- ✓ Accordance on most exciting highlight clip
- ✓ Player reaction important for top 10 clips, after fusion shows benefits



Highlights Evaluation 1

Human Evaluation on Highlights Rankings



<u>Results</u>

- ✓ **92.7%** of clips scored 2 or above
- ✓ Accordance on most exciting highlight clip
- ✓ Player reaction important for top 10 clips, after fusion shows benefits

Masters Highlights
@MastersMoments
Follow

Masters 2017: Sergio Garcia | 1st Hole, Playoff

5:06 PM - 9 Apr 2017

37 Retweets 53 Likes

1 37 53

Highlights Evaluation 2

Comparison with Official Master Highlights

<u>Research Questions</u>

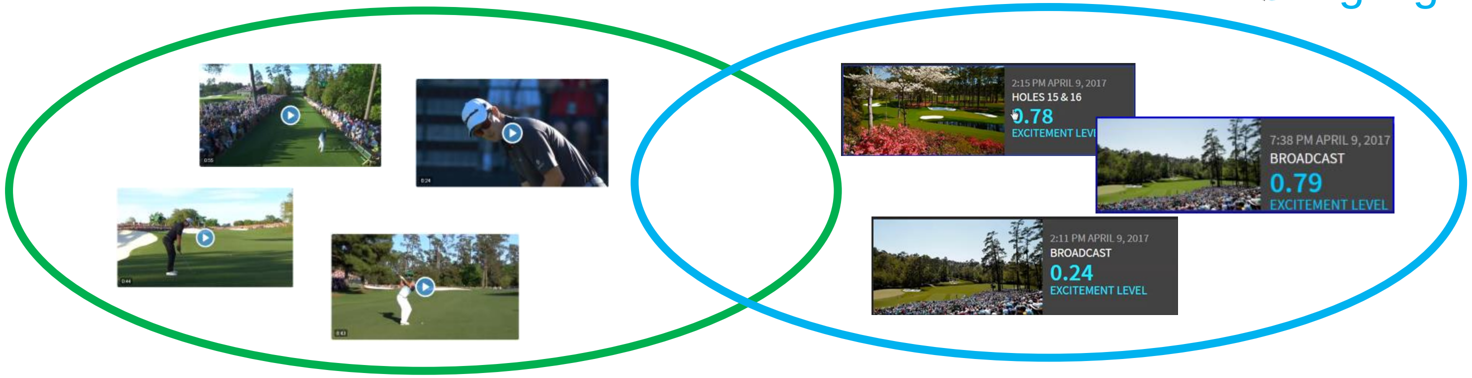
- ✓ Does H5 get all clips deemed important by professional editors?
- ✓ How does the quality of H5 clips compare to professionally edited ones?

90 Highlights from Day 4 at *Official Masters Highlights*

https://twitter.com/mastersmoments

Professional Highlights

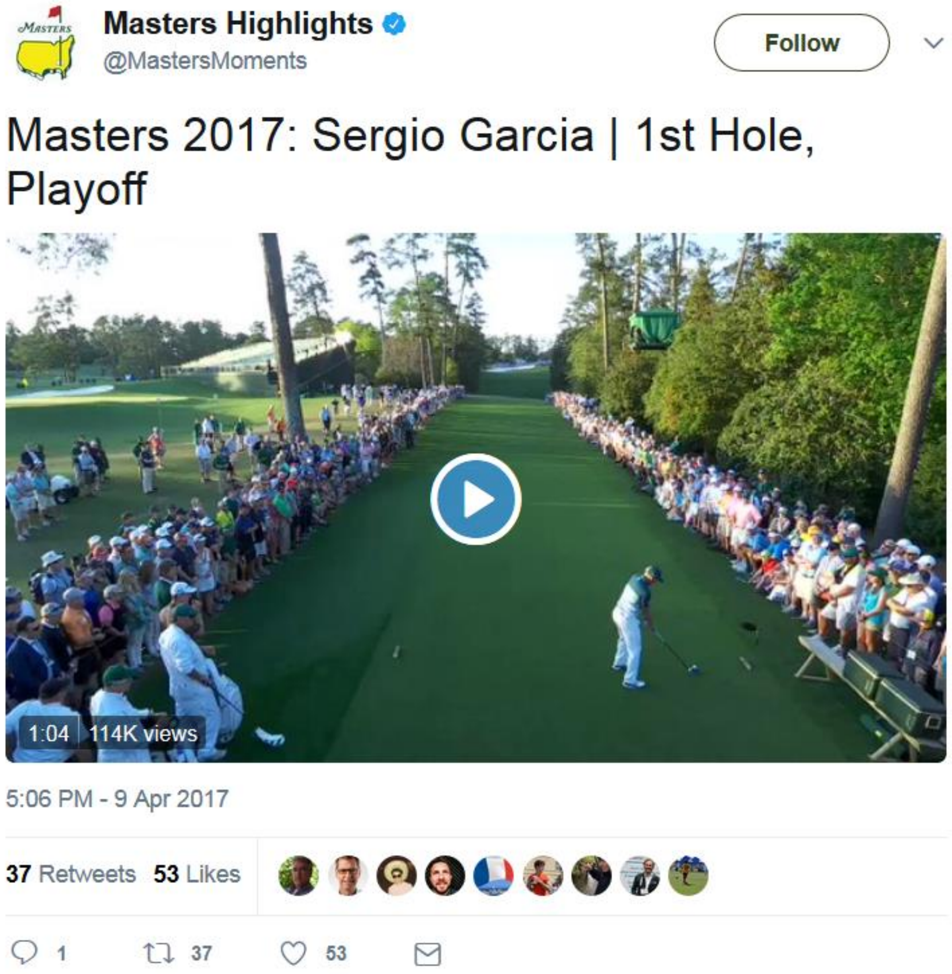
H5 Highlights



NOTE: 1 to many relationship

Highlights Evaluation 2

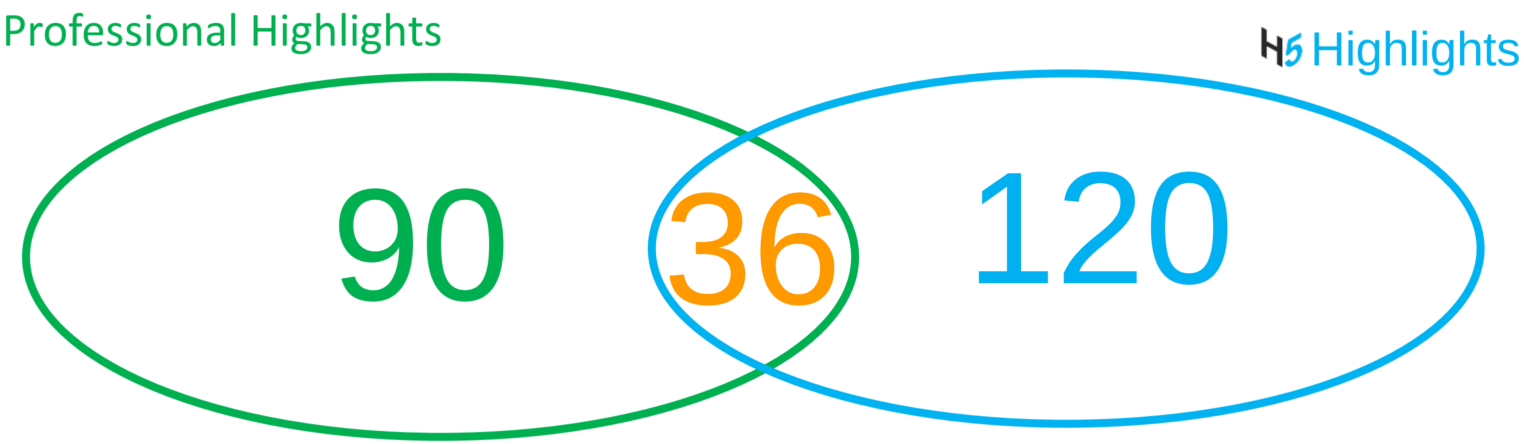
Comparison with Official Master Highlights



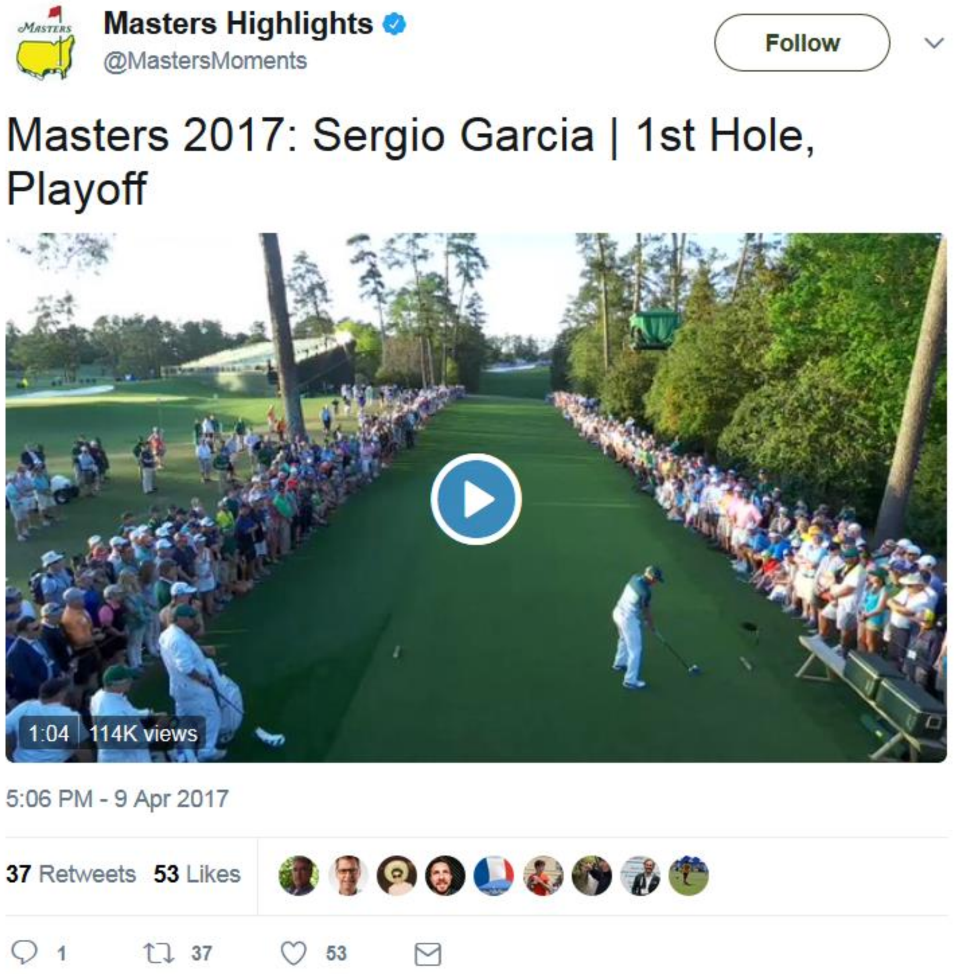
Research Questions

- ✓ Does H5 get all clips deemed important by professional editors?
- ✓ How does the quality of H5 clips compare to professionally edited ones?

90 Highlights from Day 4 at *Official Masters Highlights*
https://twitter.com/mastersmoments



NOTE: 1 to many relationship



Depth	120	500
Precision	0.54	0.35
Recall	0.4	0.9

Highlights Evaluation 2

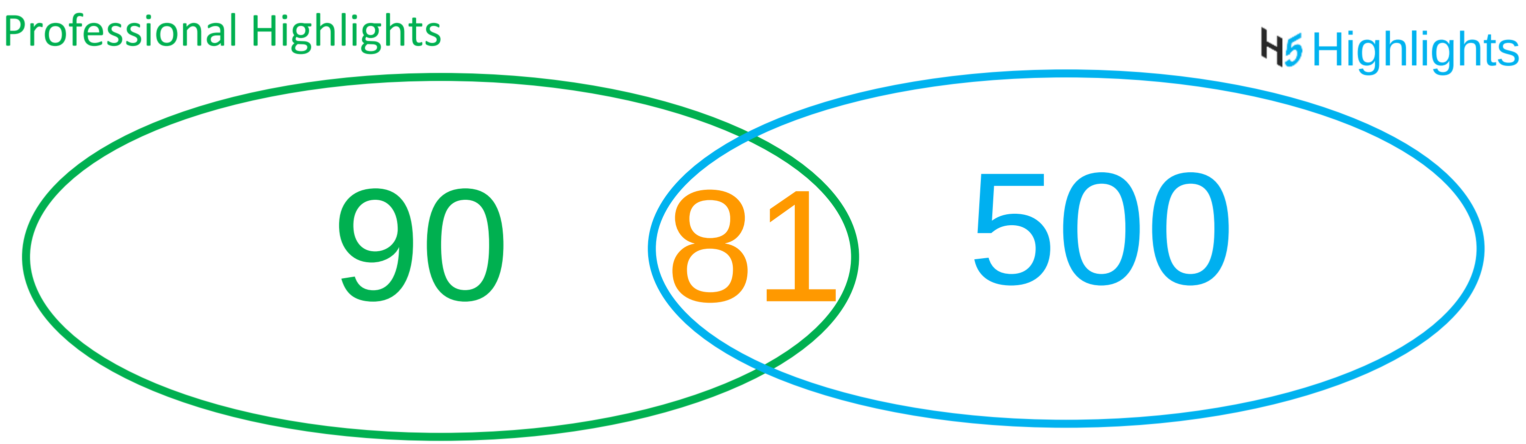
Comparison with Official Master Highlights

Research Questions

- ✓ Does H5 get all clips deemed important by professional editors?
- ✓ How does the quality of H5 clips compare to professionally edited ones?

90 Highlights from Day 4 at *Official Masters Highlights*

https://twitter.com/mastersmoments



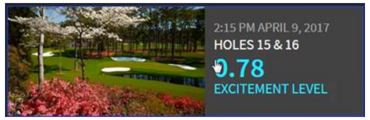
NOTE: 1 to many relationship

Highlights Evaluation 2

Comparison with Official Master Highlights

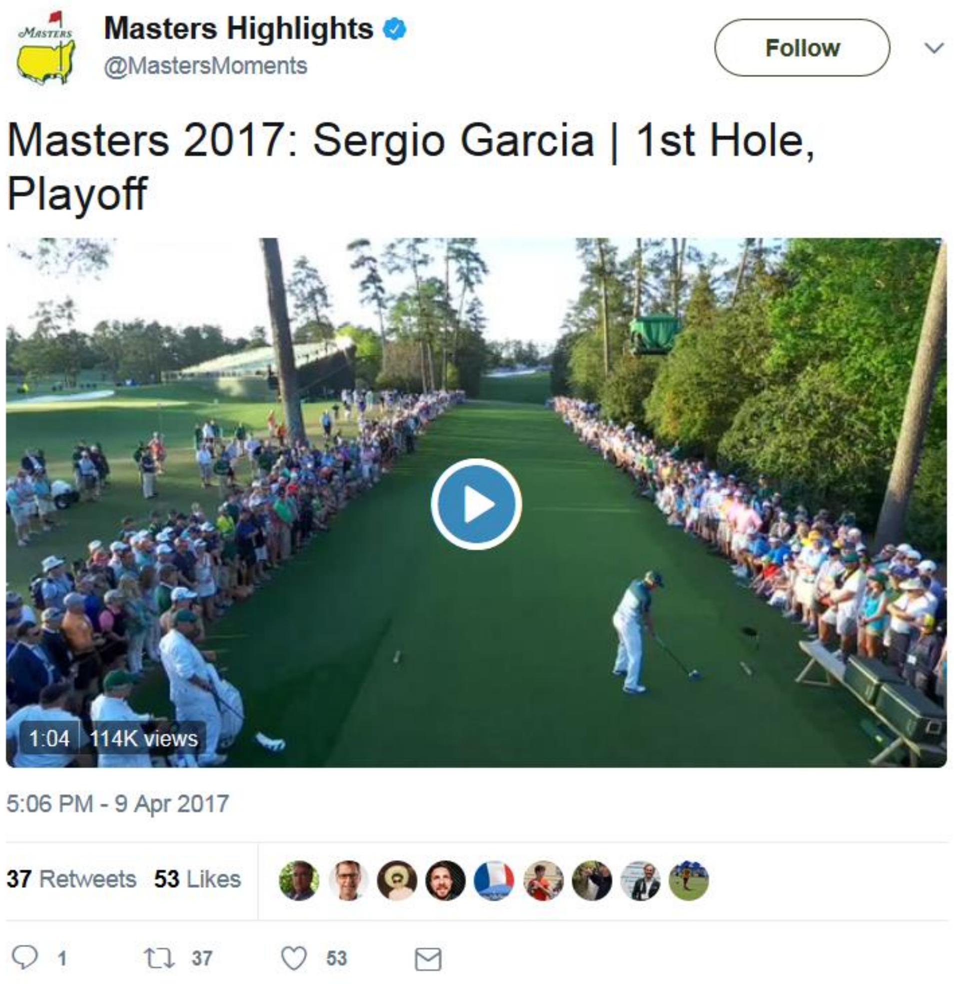
Matching Highlights A

Non-Matching Highlights B



Depth	120	500
Precision	0.54	0.35
Recall	0.4	0.9
Matching Highlights Preference	0.57	-
Non-Matching Highlights Preference	0.33	-
Equivalent	0.10	-

- A better than B
- B better than A
- They are Equivalent



Depth	120	500
Precision	0.54	0.35
Recall	0.4	0.9
Matching Highlights Preference	0.57	-
Non-Matching Highlights Preference	0.33	-
Equivalent	0.10	-

Highlights Evaluation 2

Comparison with Official Master Highlights

<u>Research Questions</u>

✓ Does H5 get all clips deemed important by professional editors?

✓ How does the quality of H5 clips compare to professionally edited ones?

90 Highlights from Day 4 at *Official Masters Highlights*

https://twitter.com/mastersmoments

<u>Results</u>

✓ H5 can find **most highlights** clips selected by humans

✓ The perceived **quality** of H5 highlights is **approximately the same** as professionally produced clips

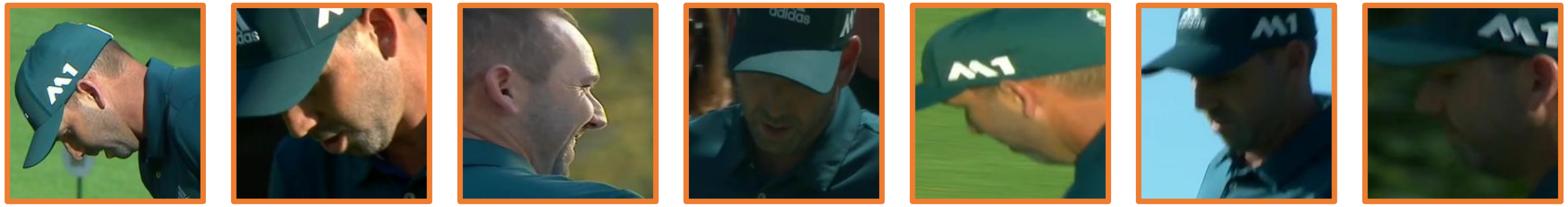
Self-Supervised Player Face Learning Evaluation

- ✓ 10 players who participated in both 2016 and 2017 Golf Masters
  - 2016 Training
  - 2017 Test
- ✓ 10 class model built by fine-tuning VGG16-Face model on 2016 training data

Number of Players	10
Number of Training Images	2,806
Training Clusters Purity	94.26%
Number of Test Images	1,181
Random Guess	10.00%
Classifier Alone Accuracy	66.47%
Classifier + Clustering Accuracy	**81.12%**

True Positives

False Negatives



False Positives

HIGHlights
From
Intelligent
Video
Engine

Framework
Multimodal
Excitement
Markers

Experiments and
Evaluation

Conclusions

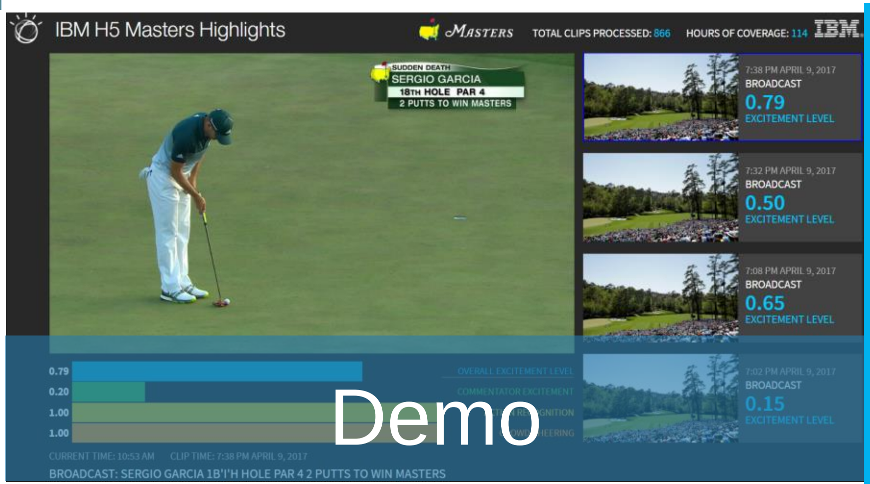
IBM

HIGHlights
From
Intelligent
Video
Engine

Framework
Multimodal
Excitement
Markers

H5

Experiments and
Evaluation



Conclusions

Conclusion and Future Directions

Multimodal markers for excitement detection

Correlation among modalities used to learn models with reduced manual annotation

Quality of fully automatic highlights comparable to professionally produced ones

Fully Working Multimodal System
for Excitement Based Highlights Production

Demonstrated at 2017
Golf Masters Tournament

Adapt System to Tennis (Wimbledon)
to produce Official Highlights

Conclusion and Future Directions

Fully Working Multimodal System
for Excitement Based Highlights Production

Demonstrated at 2017
Golf Masters Tournament

Adapt System to Tennis (Wimbledon)
to produce Official Highlights

…and the Most Exciting Highlights of All

Excitement Level = 1