AI4TV 2020: 2nd International Workshop on AI for Smart TV Content Production, Access and Delivery

Raphaël Troncy
raphael.troncy@eurecom.fr
EURECOM
Sophia Antipolis, France

Jorma Laaksonen
jorma.laaksonen@aalto.fi
Aalto University
Espoo, Finland

Hamed R. Tavakoli
hamed.rezazadegan_tavakoli@nokia.com
Nokia Technologies
Espoo, Finland

Lyndon Nixon
lyndon.nixon@modul.ac.at
MODUL Technology
Vienna, Austria

Vasileios Mezaris
bmezaris@iti.gr
CERTH-ITI
Thermi-Thessaloniki, Greece

Mohammad Hosseini
mohammad_hosseini@comcast.com
Comcast
Chicago, IL, USA

ABSTRACT

Technological developments in comprehensive video understanding – detecting and identifying visual elements of a scene, combined with audio understanding (music, speech), as well as aligned with textual information such as captions, subtitles, etc. and background knowledge – have been undergoing a significant revolution during recent years. The workshop brings together experts from academia and industry in order to discuss the latest progress in artificial intelligence research in topics related to multimodal information analysis, and in particular, semantic analysis of video, audio, and textual information for smart digital TV content production, access and delivery.

CCS CONCEPTS

• **Information systems** → **Multimedia information systems**; **Multimedia databases**; **Multimedia content creation**; • **Computing methodologies** → **Artificial intelligence**.

KEYWORDS

Artificial Intelligence; TV content production; TV content delivery; TV content analysis; TV content annotation; intelligent multimedia

1 INTRODUCTION

New scientific breakthroughs in video understanding through the application of AI techniques along with the increase in the volume of multimedia content and more computational power have led to significant improvements in automated video description and have opened fresh avenues for the seamless combination of multiple modalities' analysis. The main goal of the workshop is to promote AI techniques for multimedia analysis to enable smarter content production, access and delivery with the emphasis on large TV and radio program archives.

2 WORKSHOP SCOPE

This workshop, the 2nd of the AI4TV workshop series [6], aims to bring together experts from academia and industry in order to discuss the latest research progresses on topics related to multimodal information analysis, and in particular, semantic analysis of video, audio, and textual information for intelligent digital TV content production, compliance, access and delivery. Such topics include, but are not limited to, the following multimedia analysis techniques for broadcasted TV and radio programs as well as large TV archives:

- Multimodal content analysis: scene segmentation, person recognition, object detection, speaker gender recognition, speaker diarization, topic identification using video, audio and metadata
- Multimodal embeddings for multimedia (audio, visual, text, Knowledge Graph)
- Automatic multimedia summarization
- Automatic deep captioning
- Automatic content description
- Interactive multimodal search in archives
- Hyperlinking and enrichment of TV content
- Anomaly and violation detection in TV media contents
- Automated TV content and camera compliance (emotion detection, fire detection, etc.)
- Media-rich fake news detection
- Breaking the language barrier of TV content using multimodal translation
- Gender studies on TV and radio programs

3 WORKSHOP PROGRAMME

The workshop programme includes two keynote talks and six full papers.

The two keynote talks are:

(1) *AI in the Media Spotlight*, delivered by Alexandre Rouxel (EBU, Switzerland).
(2) *And, Action! Towards Leveraging Multimodal Patterns for Storytelling and Content Analysis*, delivered by Prof. Natalie Parde (University of Illinois, USA).

The oral presentation sessions include six full papers:

(1) *Named Entity Recognition for Spoken Finnish* presents to this end a Bidirectional LSTM neural network with a Conditional Random Field layer on top, which utilizes word, character and morph embeddings. To overcome the lack of annotated training corpora for low-resource languages like Finnish, this paper examines a knowledge transfer technique to transfer tags from an Estonian dataset. [3]

(2) *Avoid Crowding in the Battlefield: Semantic Placement of Social Messages in Entertainment Programs* proposes a method for placing public announcements, in the form of text messages, in relevant locations within a video. For this purpose, this paper exploits semantic annotations of the video, and performs spatio-temporal querying on these annotations to find candidate locations for message placement; then chooses the final locations by also considering parameters such as the spacing and the length of the messages. [4]

(3) *Realistic Video Summarization through VISIOCITY: A New Benchmark and Evaluation Framework* deals with various aspects of video summarization. It introduces a new dataset that comprises longer videos, compared to the current popular video summarization datasets, and ground-truth concept annotations. It also presents an approach for automatically generating multiple reference summaries from this kind of annotations. Finally, it proposes a simple recipe for enhancing an existing video summarization model. [1]

(4) *Neural Style Transfer Based Voice Mimicking for Personalized Audio Stories* examines how computer-based storytelling can be turned into a personalized experience for children. It applies CNN-based neural style transfer on audio by asking users (i.e., parents) to record a few sentences, so that it can learn to mimic their voice. The user audio recordings are converted to spectrograms, the style of which is transferred to the spectrogram of a base voice narrating any one of a number of different stories. [5]

(5) *Predicting your future audience's popular topics to optimize TV content marketing success* proposes the use of AI-based predictive analytics for identifying the topics that will be popular among future audiences of TV programs. These predictions can then be used in the digital content marketing strategy of media organisations, i.e. for optimizing the distribution of content across digital channels based on which topics of content will potentially be most successful by channel and time in the future. [2]

(6) *Video Analysis for Interactive Story Creation: The Sandmännchen Showcase* builds a story generation application around the well-known children's programme "Unser Sandmännchen". The paper applies video analysis techniques to a pool of originally-broadcast Sandmännchen cartoon videos; then presents a smart speaker application that interacts with the user (i.e., child) for selecting the desired segments of Sandmännchen episodes and combines them to generate a new video that is compatible with the user requests. [7]

4 WORKSHOP COMMITTEES

ACKNOWLEDGEMENTS

REFERENCES

[1] Vishal Kaushal, Suraj Kothawade, Rishabh Iyer, and Ganesh Ramakrishnan. 2020. Realistic Video Summarization through VISIOCITY: A New Benchmark and Evaluation Framework. In *Proc. Int. Workshop on AI for Smart TV Content Production, Access and Delivery (AI4TV 2020) at ACM Multimedia 2020*. ACM, Seattle, WA, USA.

[2] Lyndon Nixon. 2020. Predicting your future audience's popular topics to optimize TV content marketing success. In *Proc. Int. Workshop on AI for Smart TV Content Production, Access and Delivery (AI4TV 2020) at ACM Multimedia 2020*. ACM, Seattle, WA, USA.

[3] Dejan Porjazovski, Juho Leinonen, and Mikko Kurimo. 2020. Named Entity Recognition for Spoken Finnish. In *Proc. Int. Workshop on AI for Smart TV Content Production, Access and Delivery (AI4TV 2020) at ACM Multimedia 2020*. ACM, Seattle, WA, USA.

[4] Yashaswi Rauthan, Vatsala Singh, Rishabh Agrawal, Satej Kadlay, Niranjan Pedanekar, Shirish Karande, Iaphi Tariang, and Manasi Malik. 2020. Avoid Crowding in the Battlefield: Semantic Placement of Social Messages in Entertainment Programs. In *Proc. Int. Workshop on AI for Smart TV Content Production, Access and Delivery (AI4TV 2020) at ACM Multimedia 2020*. ACM, Seattle, WA, USA.

[5] Syeda Maryam Fatima Taqvi, Marina Shehzad, and Sami Murtaza. 2020. Neural Style Transfer Based Voice Mimicking for Personalized Audio Stories. In *Proc. Int. Workshop on AI for Smart TV Content Production, Access and Delivery (AI4TV 2020) at ACM Multimedia 2020*. ACM, Seattle, WA, USA.

[6] Raphaël Troncy, Jorma Laaksonen, Hamed Tavakoli, Lyndon Nixon, and Vasileios Mezaris. 2019. AI4TV 2019: 1st International Workshop on AI for Smart TV Content Production, Access and Delivery. In *Proc. ACM Multimedia 2019*. ACM, Nice, France.

[7] Miggi Zwicklbauer, Willy Lamm, Martin Gordon, Konstantinos Apostolidis, Basil Philipp, and Vasileios Mezaris. 2020. Video Analysis for Interactive Story Creation: The Sandmännchen Showcase. In *Proc. Int. Workshop on AI for Smart TV Content Production, Access and Delivery (AI4TV 2020) at ACM Multimedia 2020*. ACM, Seattle, WA, USA.