Video Analysis for Interactive Story Creation: The Sandmännchen Showcase

Migli Zwicklbauer¹, Willy Lamm¹, Martin Gordon¹, Konstantinos Apostolidis², Basil Philipp³, Vasileios Mezaris²

¹ Rundfunk Berlin-Brandenburg, Berlin, Germany
² CERTH-ITI, Thessaloniki, Greece
³ Genistat, Zürich, Switzerland

This work was supported by the H2020 project ReTV (grant agreement No 780656)
Introduction

- The days of a passive public depending upon a handful of selected broadcasters for their information and entertainment are long gone.
- Thanks to the internet, professional content creators and owners can create new, or reinvent existing, broadcast channels to successfully find an audience for their content.
Introduction

- In the ReTV project we have intensively explored and researched how end users can benefit from AI-based recommendation and user profiling systems
- We present a method to interactively create a new Sandmännchen story
- Sandmännchen is a well-known children’s programme from Rundfunk Berlin-Brandenburg (rbb)
- Seven-minute show broadcasted daily
- Targeted at pre-school children and accompanies them to bed with a bedtime story at 18:00
Introduction

- We built a smart speaker application which:
  - Interacts with the user
  - Selects appropriate segments from a database of episodes
  - Combines them to generate a new story

Broadcasted episodes → Video Analysis → DB → "I want to see an episode of Sandmännchen arriving on a scooter to tell a story about Jan and Henry" → Episode created on demand
Video Analysis framework

- To be able to create customized Sandmännchen episodes, we constructed a video analysis framework.
- The goal:
  - Fragment a Sandmännchen episode taking into consideration the peculiarities of the application domain
  - Annotate the main story part with the main involved character
Video Analysis - Temporal Segmentation

- Each Sandmännchen episode has three parts:
  - The introductory part
  - The main part of the episode
  - The closing part
Video Analysis - Temporal Segmentation

- In order to segment an episode to its three parts, we must detect:
  - The intro transition (i.e., transition from the introductory part to the main story)
  - The outro transition (i.e., transition from the main story to the closing part of the episode)
Video Analysis - Temporal Segmentation

- In most cases, the frames around the intro and outro transitions contain a characteristic camera zooming in and out from a screen, respectively.
- The screen is different every time, sometimes being a TV screen, other times being just a projection on wall.
- The zooming is accompanied with a fading transition, where in most cases the camera zooming fades out to a white frame.
Video Analysis - Temporal Segmentation

- We trained a Random Forest classifier on a set of 5 simple frame features that are able to capture the variations of the sought transitions:
  - Edge Change Ratio (ECR)
  - Homogeneity
  - Blackness/Whiteness
  - Blurriness
- We also implemented a DCNN-based method to segment the video to shots by adopting and extending a method of the literature
Video Analysis - Temporal Segmentation

- To detect the three parts of a Sandmännchen episode we employ the Random Forest classifier model for classifying the video frames into two classes: “normal frame” and “transition frame” (frame-level inference)
Video Analysis - Temporal Segmentation

- Taking a further step and not relying solely on the frame-level inference we incorporate the results of shot segmentation for making a video-level prediction.
- This is accomplished by employing the following simple domain rules:
  - For a frame to be considered a “transition frame”, it must belong to either the first or the last 1/3 of the video.
  - For a frame to be considered a “transition frame”, it must additionally have a temporal distance of no more than four seconds from a shot boundary.
Video Analysis - Character Annotation

- The main story deals with a different protagonist each time
- The protagonist can be:
  - a single character (Kalli - a blonde boy) or
  - a character set, which will always appear together (e.g., Rita und das Krokodil - Rita and her very hungry friend, Crocodile)
Video Analysis - Character Annotation

- We selected a subset of 11 out of the total 30 characters/characters set, as can be seen in the table in this slide

<table>
<thead>
<tr>
<th>Characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Herr Fuchs und Frau Elster</td>
</tr>
<tr>
<td>Jan und Henry</td>
</tr>
<tr>
<td>Kalli</td>
</tr>
<tr>
<td>Der kleine König</td>
</tr>
<tr>
<td>Der kleine Rabe Socke</td>
</tr>
<tr>
<td>Die Moffels</td>
</tr>
<tr>
<td>Meine Schmusedecke</td>
</tr>
<tr>
<td>Pittiplatsch, Schnatterinchen und Moppi</td>
</tr>
<tr>
<td>Plumps</td>
</tr>
<tr>
<td>Pondorondo</td>
</tr>
<tr>
<td>Rita und das Krokodil</td>
</tr>
</tbody>
</table>
Video Analysis - Character Annotation

- We decided to employ a DCNN model of the EfficientNet state-of-the-art architecture.
- We utilized the weights of an ImageNet pre-trained EfficientNet instance as the initial weights of our model and then fine-tuned it to detect the character of the main story.
Video Analysis - Character Annotation

- Our model annotates each frame of an input video with the detection score for each one of the 11 characters/character sets (frame-level character inference)
- Video-level prediction: performing a majority voting over the frame-level predictions, since a Sandmännchen episode deals with the same character/characters set in its entire main story part.
Video Analysis Service

- The discussed video analysis techniques have been incorporated into a video analysis component.
- This component is deployed as a REST service that:
  - retrieves a video file
  - performs the temporal segmentation of a Sandmännchen episode
  - analyzes the main part to identify the main character, and
  - stores the results in a JSON-structured file which can be downloaded using a specific type of call
Evaluation of Temporal Segmentation

- Using the Random Forest classifier for the detection of the transitions (frame-level inference) → 88.5% F-score
- Employing the additional domain rules, for the prediction of transitions (video-level inference) → 91.7% F-score
Evaluation of Character Annotation

- Using the DCNN model for the frame-level predictions, we observed classes that perform very well but also classes with noticeably bad performance. This is due to:
  - Varying difficulty of detecting each character/characters set due to its specific characteristics
  - The main character/characters set are not necessarily depicted in all analyzed frames
- After employing majority voting to infer video-level predictions → 100% accuracy for all classes
Abendgrüß Application

- The Abendgrüß is designed for the use with smart speakers with display
- The first prototype was developed as an action for Google Assistant, focusing on the Google Nest Hub
- We use Google's Dialogflow, a chatbot framework integrated with the Google Assistant
Abendgruß Application Voice Commands

- When a user speaks to the Abendgruß application on the Google Assistant, their commands are sent to Dialogflow and mapped to API calls.
- Those calls are then sent to the Abendgruß API, which either returns options for the user to choose from, or the customized video in the final step.
Abendgruß Application

- To start Abendgruß, the user has to say “OK, Google, speak to Abendgruß”
- The Nest Hub answers: “All right, I'm starting the test version of Abendgruß”
- The application opens
- The user sees the start screen and gets a welcome combined with a call to action: “Hello! To watch your own Abendgruß, say the word ‘Abendgruß’”
Abendgruß Application

Auf dem Roller

Zu Fuß
Abendgruß Application

- First, the user can choose how the Sandmännchen should arrive
- Two options are shown for example, “On the scooter or by foot?”
- These two options are selected randomly each time the application is used
Abendgruß Application
Abendgrüß Application

- Secondly, the user determines her/his main story by answering the question “And what story do you want to see today?”
- Again two options are presented for example “Jan and Henry or the Pittiplatsch and Schnatterinchen?”
- These two options are selected randomly each time the application is used
The Abendgrüß application finally shows an automatically-generated Sandmännchen video.
Thank you!

Contact:
Vasileios Mezaris, bmezaris@iti.gr

This work was supported by the H2020 project ReTV (grant agreement No 780656)